Lexicon immigration service:

Prolegomena to a theory of loanword integration
Lexicon immigration service:
Prolegomena to a theory of loanword integration

De immigratiedienst van het lexicon:
Prolegomena tot een theorie over leenwoordintegratie

( *met een samenvatting in het Nederlands*)

Imigracijska služba leksikona:
Prolegomena za teoriju integracije tuđica

( *sa sažetkom na srpskohrvatskom*)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op donderdag 28 mei 2015
des middags te 12.45 uur

door

Marko Simonović

geboren op 17 september 1983 te Zaječar, Joegoslavië
Promotoren:  Prof. dr. R.W.J. Kager  
               Prof. dr. W. Zonneveld
## Contents

Acknowledgements ................................................................................................. III
Preface ..................................................................................................................... 1
Chapter 1: Bilinguals and borrowing ................................................................. 5
  1.1. Bilingual communities and borrowing .................................................. 6
  1.2. Categorical restrictions on code-switching .......................................... 7
  1.3. Structural restrictions: code-switching vs. borrowing ....................... 11
  1.4. Summary of the chapter and further development ............................. 14
Chapter 2: Borrowing research ........................................................................... 15
  2.1. Introduction ......................................................................................... 15
  2.2. Rule-based approaches to borrowing ............................................... 17
  2.3. Delineation research: The Phonology/Perception debate .................. 37
  2.4. Grammar-based research .................................................................. 49
  2.5. Conclusion of the chapter and further development .......................... 69
Chapter 3: Borrowedness research .................................................................... 71
  3.1. Core-periphery relations .................................................................... 71
  3.2. A brief history of formalised borrowedness ....................................... 72
  3.3. Onion models of lexical stratification ................................................. 73
  3.4. The Reality of Lexical Strata .............................................................. 79
  3.5. A ‘too elegant’ alternative: Underspecification .................................. 80
  3.6. Conclusion of the chapter and further development ........................... 85
Chapter 4: Present Model: Ingredients & Representations ............................... 87
  4.1. The ingredients ................................................................................... 92
  4.2. Representations and being faithful to them ....................................... 88
  4.3. Conclusion of the chapter and further development .......................... 103
Chapter 5: Morphology and morphosyntax of integration ............................... 105
  5.1. Morphosyntactic assignment ............................................................. 106
  5.2. The initial form is void of morphological and morphosyntactic features ................................................................. 110
  5.3. The initial form is a surface form ...................................................... 119
  5.4. The initial form has a (specific) meaning .......................................... 127
  5.5. Conclusions of the chapter and further development ....................... 129
Chapter 6 .............................................................................................................. 133
  6.1. Introduction ......................................................................................... 133
  6.2. Inter-language mappings cover what LC misses ............................. 139
  6.2. What is a mapping? ........................................................................... 141
  6.3. Mappings are conventional ............................................................. 146
  6.4. Dynamics of inter-language mappings: The LC factor .................... 150
  6.5. Doing without loanword-specific Markedness .................................. 155
  6.6. Conclusions ....................................................................................... 160
Acknowledgements

Personne ne mentionne ton nom dans l'introduction.
A pronunciation exercise in my French book

I. Why acknowledgements are never quite satisfying

There are great many reasons why acknowledgements should be abolished, together with many other antediluvian institutions which have lost credibility or purpose somewhere along the way, for instance, marriage and national borders and a third thing which I forgot. (I could contact my external-memory paranymph and find out what I forgot, but she’s traveling the world at the moment and it may very well be that I just meant the stripkenkaart, which is already abolished, so the whole contacting endeavour would be for nothing.) Some of these reasons are:

✓ Many people use acknowledgements to thank people for existing or for doing their job. While I agree that both existing and doing one’s job are quite nice, I believe that neither should be reasons for thanking people publicly.

✓ Many people use acknowledgements to show that they actually have a life. On the other hand, acknowledgements are quite useful if you need to know the author’s relationship status. On a third hand, why not just ask them?

✓ Some people use acknowledgements to make their audience believe that they should have become writers. There are some indications that I’m one of those people.

✓ Most importantly, most acknowledgements are not even really written by the same person as the rest of the book. The chapters of the book are usually written by a panicking PhD student, while the acknowledgements are virtually always put together by a person who already knows that their manuscript has been approved by the committee. (I know of two exceptions, and I would love to name them, but I forgot how many h’s again.) I have observed, both in what is commonly referred to as ‘me’, and in what is usually termed ‘other people’, that there is an enormous difference between these two persons in character, priorities in life and ideas about why we are here, and by here I mean in this book-reading situation (which is all a fascinating topic I would like to recommend for further research).
There is, however, one merit of this section which I am not sure how to replace in the world after the anti-acknowledgement revolution, and which even makes me think that it is a good idea to call off the whole revolution. This merit lies in the fact that this section partially allows the author to shake off the shaky construct called the author.

II. Something about the main character of this book

I would like to start by acknowledging the most important character in this story, the real author. That would, of course, be — you. No, really, you, Anonymous Reader. Thanks for reading this book in whatever form and for whatever reason. Nothing is more meaningful in respect of this book (or any book) than what you are doing right now, blessed be your eye movements. Especially if you manage to spend some minutes of focused reading with it: as far as I know, that’s as close as things get to heaven.

No need to pretend, dear reader, just think about it: you are probably the only person reading this book at this moment. So, in that sense, you are the life of this book here and now. You are the world interacting with all the years we have spent making this book perfect for you, obsessing about the mental states we are creating and images we are manipulating in your mind, worrying about your general knowledge, your interest in the topic and your attention span, hoping you would like our tone and not feel that we are too stuck-up or too informal.

So here you are, reading, perfectly in line with the primary purpose of the book, doing some serious sense-giving by the very fact of showing up. Thanks, dear reader, let’s cherish the moment for a moment.

III. Something about who we are

I have just spent a long walk home thinking about what it means that a book was written by me. I decided that, in common understanding, I wrote every book that my hands have typed out and which no one else claims to have written or read or to have been able to read before I typed it out (and I’m not sure what happens if someone claims to have dictated it into my ear). In that sense, this book has been written by me. Apart from that, I believe that we all know that no one has ever written a book: this book, as every book, grew in an undefinable and unnamable in-between shared with all kinds of supervisors, colleagues, friends, parents, paranymphs, parasites and whoever else showed up.

a. The supervision

It is always risky to make claims about things one hasn’t tested empirically, but there is a good chance that I have been the object of the best and most democratic supervision in the history of the universe (including the future). It could be that my brain is making everything prettier for the occasion, but I
actually recall zero meetings with my supervisors where there was no substantial laughing and as many meetings after which I was not extremely positive about the whole PhD situation. It was in those laughter-ful and optimism-inducing meetings that I learned all I know about how to be academically nice and how to write for people who are not in my head.

b. The friendship and family ship

I have never really understood the difference between these two ships, which I believe has to do with the friendly, honest and unassuming parent ship I have had. Speaking of which, 75% of the case study chapters in this book are (even officially) co-authored. During this co-authoring enterprise, I was observing how all the aspects of doing research became immensely more interesting and pleasant than when done alone. The same is true of co-organising all those workshops in Split, Stockholm, and the future and co-presenting papers in Barcelona, Geneva, Utrecht, Niš, Stockholm, Dubrovnik (remind me not to try that again), as well as co-preparing papers for publication on GDrive. This experience determined, how I believe I want to spend the rest of my academic life: in co-authorship.

For the rest, there was the rest of life. There was working and saving the world in that huge office at Jankskerkhof (where Rita met the suitcase), in 2.34 (also known as the loanword department), in 0.43a (where I haven’t even been) and 0.45 (where I’m not even sitting at the moment). There was thousands of hours of teaching Dutch, Serbo-Croatian, Slovene and German. There was speaking English, Dutch, Serbo-Croatian, Slovene, Italian, German, rudimentary Slovak and imaginary Scandinavian. There was proofreading my writings before publishing them here or on Peščanik or on Dutch review (the process in which I finally learned the Serbo-Croatian quotation marks and the difference between a dash and the other thing). There was planning the most improbable outfits and outcomes with my paranymphs. There was good sleeping in Boeroehof, Čupićeva, that warm christian house with guinea pigs, Dickeislaan, Dinoland, Dubrovačka, Gospodara Vukića, Hoogstraat, Laura&Thijs’, Omladinskih brigada, Opaalweg, Predicasa, Vjekoslava Kovača and Zahumska. There was being part of an abishop, an academic brotherhood/sisterhood, a Baradclub, a canmship, a council that never decided anything, a department, a Dragović/Grašoship, an institute, an intervision group, a kibbutz, a komesarijat, two paranymphships, a reading group that has only read one article, a research group, a Sander&Pamelaship and a zbigship.

And that was all us and this is all you as well. And I would love to say that this book is dedicated to you, but I’ve already said that it was by you (and that I was just the guy who took the minutes), and self-dedication is not really a thing something I want to google again. But don’t be sad, dear friends, we will always have Ariel’s voice, Barcelona, besmislennoe, Bologna, Centrala, Coffee Dream (incl. Story reading), Constanța, Donji Čehi (not to be confused with České Budějovice, which we will also always have), Drvce, Fribourg, Lund, Luxembourg, that monastery in the middle of nowhere,
Milan, Rome, Padova, Paris, Petnica, Platamon, Split, the razgalica, the sneak, the svadba (even if we don’t have the brak), Tartu, Tutu, Vapiano, Verona, Wrocław, Žuljana and all the rest. We will always have more than an equal share. Which reminds me.

IV. Dedication

As everything, this book excludes something, and it is haunted by what it excludes. In order to acknowledge that (and, as Donna Haraway would have it, ‘stay with the trouble’), I dedicate this book to Bica and Deljana. If you are not a netizen of Yugosphere, you probably don’t know Bica and Deljana. Well, let me tell you the story.

One of the most popular funny video clips in Yugosphere shows an interview from some local TV station with a female inhabitant of a Roma settlement somewhere in Serbia. When asked whether her children go to school, she replies that only her son does. The interviewer informs the woman that primary education is obligatory, to which she responds that she intends to send two more children to school ‘as of tomorrow’, pointing at the elected ones (which the viewer never gets to see) and naming them: Bica and Deljana. This clip produced immensely popular memes in Yugosphere. The three versions of this clip I have consulted have more than 2 million views on Youtube and there are 36,200 Google hits for “Bica * Deljana”. (“Marko Simonović” almost gets there, but that is because of that famous basketball player who’s called that way, which is also the reason why no one has ever found me on Google.) Indeed: the clip is really funny.

There is nothing wrong with laughing. Yet after we are done laughing, more stays with us than just the joke. Bica and Deljana are the only living people I know by name who have never got any school grades, so they couldn’t possibly get good grades and have their brain drained to a richer country, where they could eventually stand in front of a committee in gender-appropriate attire (Oh, so it wasn’t the strippenkaart after all!) and defend a PhD thesis. No matter how bright they were, already at the age of 7, their life was determined to be something else than what brings me here.

Dear Bica and Deljana, this book has nothing to do with you. It(s author) only gets nervous from thinking of you: it was written by and for people who have CV’s, ID’s, PC’s, standard languages and something we could call ‘a certain outlook’ acquired at institutions you have almost certainly had no access to. This book is not even about you: for methodological reasons, we needed to restrict our discussion to language contact situations in which there is a clear and clearly perceived boundary between languages, presumably not the linguistic reality you live in. And yet, it is all dedicated to you, with my hope that, by some miracle which I have done nothing to make happen, you went to school.
Preface

Languages seem to have a universal potential to take over elements from other languages and incorporate them into their structure. The field of loanword research aims to provide answers to questions which arise whenever language contact leads to borrowing of words or word-like units, i.e. whenever some pronounceable linguistic content is imported into a language. (As a consequence, all cases in which no lexical matter is imported – e.g. calquing, grammatical shift, sound shift, etc. – fall outside the explanandum of loanword research.)

It is uncontroversial that loanwords generally get introduced by (broadly defined) bilinguals, initially as code switches, drawn from the lexicon of the source language (henceforth SL) and incorporated into the recipient language (RL). However, it is not obvious when a word turns from a code switch into a loanword, i.e. when the word becomes represented in the RL lexicon. This distinction is one of the most debated issues in the modern literature on code-switching (Chapter 1) and any attempt to make it in a coherent way forces us to ask further important questions: for whom does this word become a loanword? (an individual speaker? the whole community?) and is this an irreversible process or can a loanword be borrowed, forgotten, and borrowed all over again? (and, again, by whom?). These questions have received quite some attention in the literature and the appropriate answer seems to be: all of the above. In other words, rather than lending themselves to a simplistic theory, loanwords require one to stay aware of the various dimensions along which they exist and matter: individual and communal, synchronic and diachronic, etc.

This is not to say that any discussion of loanwords has to fade into hopeless relativism because we are unable to even tell where the loanword domain begins. Researchers have established quite a few litmus tests for distinguishing code switches from borrowings. One of the most prominent characteristics of loanwords (with limited diagnostic use, however – see Poplack 1980) is phonological integration into the RL. It seems universally true that words which enter a language are getting modified in ways which can be explained as a result of the agency of the borrowing system. This is the first area in which there is a sizable corpus of modern generative research: the area usually termed loanword adaptation. Phonological integration seems to go hand in hand with morphological integration, and the latter has even been shown to be triggered by the former (see e.g. Jurgec 2014). However, morphological integration of loanwords has, unlike its phonological counterpart, not yet been transformed in an established research area.
Many factors are entangled here and many perspectives are possible, so it comes as no surprise that loanwords are somewhat of a shared asset within the field of linguistics, cutting across boundaries between fields, programmes and ideologies. On the one hand, they are traditionally the object of the field of language contact research, which is considered part of sociolinguistics. In this field, loanwords are situated within a continuum of related phenomena, such as code-switching, language shift, language mixing, creolisation, etc. (Thomason 2001, Matras 2009). This field is dominated by the so-called usage-based approaches to linguistics (see e.g. Backus 2012). On the other hand, loanwords are the object of extensive research in phonology, which is dominated by formal approaches. As the overview of this field in Chapter 2 will show, loanwords have never really been situated in this tradition, and it is still rather implicit where loanword processes are to be imagined in the architecture of the linguistic faculty. In other words, for formal phonologists, loanwords are an interesting puzzle which should be solved using familiar tools.

The goal of this dissertation is to empower the field of formal loanword research by (a) taking insights from sociolinguistic research seriously and plugging them into formal models and (b) incorporating morphological (and morphosyntactic) integration into the picture of loanword treatment, now dominated by data on phonological integration. The method employed here is one of re-reading the proposed models through and with each other, in order to arrive at prolegomena to an integral theory of loanwords. While building the model, the inquiry will be guided by some of the classical questions in modern linguistics:

a) Is the process in question synchronic or diachronic? If it is synchronic, is the process part of the synchronic grammar or another module?

b) Is the process in question a reflection of the speakers’ knowledge or emergent in the community? If it is a reflection of the speakers’ knowledge, how is the process mentally represented?

As a result, a new model will be proposed in order to account for the interface between languages. This model will cover data from phonological, morphological and morphosyntactic integration, since we shall see that all these types of integration are entangled aspects of a single broad process: lexicalisation, i.e. the creation of a new lexical entry based on a foreign surface form. The presented model will in this sense be characterised by a certain telicity, not unlike the existing adaptation models (Chapter 2).
However, while these latter models see loanword processes as moving towards becoming indistinguishable from native items, the integration model will have as its endpoint the creation of a fully functional RL lexical entry (sometimes very distinguishably non-native). Since loanwords display processes of all the types mentioned above, the model, which will be presented in Chapters 4, 5 and 6, will be comprised of two different apparatuses able to capture different aspects of loanword behaviour without losing sight of what they exclude.

The more diachronic apparatus of the model will concentrate on the ways in which properties of the initial code switch are preserved in the process of integration into the lexicon (shared by the language community), which involves the creation of paradigms, the assignment of morphosyntactic features, etc. We will present strong evidence that borrowing is to be seen as lexicalisation based on a surface form, guided by a force which militates against the introduction of new versions of the incoming form – Lexical Conservatism. The more synchronic apparatus will be more suitable for viewing the regularities which are part of borrowers’ knowledge: the inter-language mappings, which emerge within the community and which contain instructions for converting SL structures into RL structures.

The rest of this dissertation is organised as follows. Chapter 1 presents the most important findings of sociolinguistic research into loanwords. Chapter 2 reviews research done by generative phonologists in the field usually termed loanword adaptation. In Chapter 3, research into lexical stratification is reviewed. In Chapter 4, the main ingredients of the model proposed in this dissertation will be discussed. Chapter 5 will consider the cases of morphosyntactic integration. In Chapter 6 the inter-language mappings are introduced and discussed. Chapter 7 brings an interim summary and announces the four subsequent chapters, which bring four case studies, in which the proposed model is put to use to account for larger data sets. Chapter 8 presents an account of consonant gemination in loanwords. Chapter 9 discusses *a*-epenthesis in Serbo-Croatian from the contact perspective. Chapter 10 brings an account of verb borrowing and aspect in Serbo-Croatian. In Chapter 11, the Latinate nominalisations in Serbo-Croatian are analysed from the perspective of our model. Chapter 12 concludes this dissertation.
Chapter 1
Bilinguals and code-switching: lessons from sociolinguists

This chapter will review the relevant findings of research into bilingual production as considered by sociolinguistic literature. It is important to make it clear that, within this field, matter borrowing does not constitute an isolable object, since it is part of a continuum of related phenomena such as code-switching, language shift, language mixing, piginisation, creolisation, etc. This chapter therefore presents a selection of findings which can most productively be combined with those of phonological research, which typically treats loanwords without any reference to other contact phenomena. Most prominently, we will examine code-switching and nonce borrowing as places where borrowing originates.¹

This field is as contentious as that of loanword phonology and the issues under dispute also have a bearing on our model of loanword integration. Still, the objective in this chapter will not be to take a stance on the open issues, but to present the most substantiated findings and their consequences for loanword research. For this reason, the discussion will strongly rely on research conducted on big corpora of spontaneous bilingual production, conducted for a substantial part by Shana Poplack and colleagues (for more general overviews, see Muysken 2000, Myers-Scotton 2002, Poplack 2001, 2004).

Finally, it is important to note that the researchers in this field rarely consider phonological data per se. Although phonological integration is recognised by the scholars in this area as a good diagnostic for establishing that an item is a borrowing (and not a code switch), collecting valid data on the phonology of code-switching seems to be an extremely difficult ordeal for methodological reasons. The main confounding factor is that bilingual speakers often have a foreign accent in one of the languages in question, at least as a part of what they actively use or are able to understand, i.e. of their verbal repertoire (see e.g. Poplack et al. 1988). This variety is usually put into action in situations where code-switching is common (intercultural discourse). Now when features of language A appear on the embedded words from language B, they can either come from A or from the A-coloured accent in B. In order to disentangle the influences here, the researcher would have to decide which part of the verbal repertoire serves as the source for...

¹ The term nonce borrowing (introduced by Weinrich 1953) is commonly used to refer to items which show symptoms of being borrowed into the recipient language, but occur only once in the investigated corpus.
embedded elements, a decision which cannot easily be motivated and which seems to gloss over the very heart of the phenomenon — that bilingual speakers are more than 2-in-1, i.e. that they are often able to shift between a multiplicity of linguistic identities. As a consequence, in Chapter 4, arguments will be presented for a more complex conceptualisation of the input to loanword integration.

This chapter starts by discussing the literature pertaining to the relation between the properties of bilingual communities and code-switching/borrowing in §1.1. The main finding is that there is a clear correlation between the duration of contact and the stabilisation of inter-language mappings. §1.2. discusses a number of well-attested patterns regarding different categories’ susceptibility for code-switching and their consequences for the input to loanword integration. First, an asymmetry between lexical and functional categories will be shown to shape code-switching patterns in ways which are informative for the input of borrowing: since functional items are not available in the embedded language (which eventually plays the role of the SL) only a single paradigm member will typically play the role of ‘the input’. Also, nouns will be shown to be special, also in the sense of often being the only category into which novel items can be introduced. Finally, §1.3. discusses the structural properties of code-switched elements and the proposed ways to determine their loanword/code switch status.

1.1. Bilingual communities and borrowing

Most scholars describing bilingual communities devote some attention to the conventions concerning overt flagging of switched items, such as using metalinguistic comments and pauses (as opposed to smooth code switch described for, for instance, the New York Puerto Rican community in Poplack 1980) and the amount of foreign material which gets embedded (for a classification see Poplack 2004). However, very few claims are made regarding the relation between type of bilingual community and the patterns expected to emerge in the recipient language as the result of borrowing.

The most influential statements in this area were made by Einar Haugen (1950, 1956, 1969), more than half a century ago. Haugen differentiates between three types of bilingual communities with different types of borrowing. During the pre-bilingual period, loanwords are introduced “by a relatively small group of bilinguals and spread widely among the monolingual majority; the words show (almost) complete native substitution, with great irregularity in the phonetic results” (Haugen
The period of adult bilingualism introduces more regular adaptations and some sounds from the donor language. Finally, in the period of childhood bilingualism, loanwords are created systematically and with many non-adaptations.

It should be noted that Haugen’s generalisations were made based on the developments in immigrant languages in the US, in communities which eventually switched to English. There seem to be two axes along which Haugen’s scenario proceeds: the duration of contact (which brings regularity\(^2\)) and the proficiency in SL (which brings more non-adaptations). These two aspects are inseparable in the communities observed by Haugen, but their conceptual separation is essential if we are to compare most of the contact situations in the modern world, characterised by the existence of long-term pre-bilingual communities. In such cases speakers establish what Haugen calls “interlingual identification” (Haugen 1956:44) – one-to-one mappings between SL and RL elements, which provide a steady integration strategy for each SL structure. This concept, termed *inter-language mappings*, will be crucial in the analysis of loanword integration presented in this dissertation and we will return to it in Chapter 6. Although based on observations in a limited number of similar communities, Haugen’s data do receive confirmation in more recent literature. The study of Spanish loanwords in Nahuatl (a Uto-Aztecan language spoken in Mexico) presented in San Giacomo & Peperkamp (2008) shows that a higher occurrence of non-adaptations correlates with more advanced community bilingualism, not with individual bilingualism. In other words, bilinguals perform less adaptation if their interlocutors are more proficient in the SL. Paradis & LaCharité (2009) arrive at the same conclusion in a study of the influence of English on Quebec French. Furthermore, Paradis & LaCharité (2009) also found that during the initial period of language contact there were more irregular adaptations (false analogies in their terms).

1.2. **Categorical restrictions on code-switching**

It is commonly acknowledged in sociolinguistic literature that not all categories are equally borrowable, a fact which nicely matches the observation that not all categories are available for (insertional) code-

\(^2\) A terminological caveat is in place here. The irregularity contingent on the initial period of group bilingualism should not be confused with the variation which characterises the early life of borrowings (often systematic variation between non-adaptation and different degrees of adaptation). Haugen (e.g. 1956: 55) argues that the latter type of irregularity stems from the speakers’ awareness of the origin of the borrowed item, which wears out with time.
Chapter 1

switching. In this section we restrict our attention to the main findings on
the categorical restrictions on code-switching, which will inform our
borrowing model in important ways in Chapter 4.

1.2.1. Lexical vs. functional morphemes
One of the most influential findings of the field of language contact research
is that lexical and functional categories are not treated on a par in code-
switching. Generally speaking, functional morphemes from only one
language are allowed in a sequence (often assumed to be minimally a
complementiser phrase, CP). In the Matrix Language Frame (MLF) model
developed by Carol Myers-Scotton and colleagues (Myers-Scotton 1993, 2002)
this fact is seen as central to code-switching. The language which provides
the grammatical morphemes (or system morphemes, in the terms used by
Myers-Scotton 2002) is the Matrix language of the sequence, whereas the
contribution of the Embedded language is limited to providing lexical
material (or content morphemes). Depending on the relative acceptability of
content morphemes in isolation, the grammatical morphemes of the
Embedded language can either be omitted or lose their grammatical
meaning. An example comes from Malinche Mexicano, a variety of Nahuatl,
with the final adjective in Spanish.

(1) In nonāntzin, poderoso.
   ‘As for my mother, she is powerful.’
   (Hill&Hill 1986, via Myers-Scotton 2002)

The Spanish form poderoso is morphologically complex – it contains the
masculine gender marker -o (cf. poderosa ‘powerful-feminine’). However,
masculine forms seem to serve as default forms in code-switching for
Malinche Mexicano speakers. This shows that inflectional morphology is
suspended in one of the languages, namely the embedded language, an
observation which has important consequences. The content morpheme
poderos- never occurs in isolation in Spanish.

The observation that no novel, isolated realisations of bound morphemes
are found in code-switching can be interpreted, as it will be here, as a
confirmation of the fact that no morphological analysis is performed on
elements of the embedded language. Whatever the contribution of the
embedded language may be, this contribution can only consist of
morphologically opaque chunks. However, this does not mean that it is
equally impossible to borrow anything other than free content morphemes.
Rather, borrowing of grammatical/bound morphemes is restricted in
expected ways. The literature on language typology and language contact
shows that “no non-lexical property can be borrowed unless the borrowing language already includes borrowed lexical items from the same source language” (Moravcsik 1978: 110). This provides a clue of how less regular (derivational) morphology can get ‘smuggled in’ without bringing inflection with it. For instance, if Malinche Mexicano borrows the word poderoso ‘powerful’ and the word poder ‘power’ and a number of other adjective/noun pairs with the same morphemic structure, monolingual speakers will have good opportunities of analysing -oso as an adjectiviser. This scenario seems plausible for the massive import of loan derivational morphology in languages such as Dutch (Booij 2002), English or Serbo-Croatian (for a discussion, see §5.4.1. and Chapter 11). However, it remains a fact that whole items have to enter the language without internal morphological structure first.

These examples, along with many others quoted by Myers-Scotton (2002), illustrate the general point that, independently of their command of the languages involved, speakers who perform code-switching will limit the role of the embedded language to providing lexical material.

Another important consequence of the asymmetry between lexical and functional categories is that functional categories within a domain are predicted to always belong to the same code – the Matrix language. Based on this prediction, another important principle of code switching was formulated and substantiated: the Functional Head Constraint (Belazi et al. 1994). This constraint blocks code-switching between functional heads and their complements, for instance between a complementiser and the clause it introduces.3 This essentially means that, since functional heads switch the Matrix language as soon as they are introduced, they can hardly occur in a context which would make them eligible for borrowing. This can be an explanation for the fact that content morphemes are borrowed far more frequently.

Researchers of this phenomenon have offered two types of explanations for discrepancies between functional and lexical categories in code-switching. Myers-Scotton (1993, 2003) assumes that the source of these differences lies in the fact that these two types of morphemes are universally produced/processed in different ways. Muysken (2000), on the other hand, argues that function morphemes vary more across languages (e.g. a morpheme can be bound in one language, but its correspondent in the other language can be a free morpheme), which makes it harder for the speakers to

3 The reader can verify this by code-switching to her favourite language in the two indicated positions in the sentence He said that [1] he did not agree and [2] ran away. The code-switch after the conjunction (position 2) should sound much smoother than the one after the complementiser (position 1).
establish equivalence between them. Clearly, these two explanations could very well be part of the same account, which will not be pursued here. What is important for our modelling purposes is the fact that lexical/functional asymmetries are real and determining for code-switching.

Focusing for a moment on the consequences of lexical/functional asymmetries for loanword research, a number of preliminary conclusions can be drawn, which restrict the possible inputs (and consequently outcomes) of loanword integration. Since lexical items can introduce maximally one (morphologically unanalysed) form into the new language and no full paradigms with all allomorphs are available during the process (because the opposite case would mean activating function morphemes), it might very well be the case that the underlying representations on which these allomorphs are based are unavailable as well. Speakers, when in the insertional code-switching mode, simply have access to the surface form of the one form which they use. This still does not mean that we know exactly how much phonetic detail enters the representation accessible to the speaker. In §2.3, various approaches to the input of loanword phonology will be discussed. The same issue will be discussed in the framework of this dissertation in Chapter 4.

1.2.2. Nouns are special

Single code-switched nouns are by far the most common items in bilingual production corpora (e.g. Poplack 2004). This matches the fact that most loanwords in any language contact belong to this category (for an overview of borrowability hierarchies, see Matras 2009).

Seemingly these facts have to do with the semantic/referential properties of nouns. However, there is ample evidence that languages often borrow other categories, turning them into nominals. The most discussed phenomenon are bilingual verbs (term from Muysken 2000, see also Myers Scotton 2003). In many languages the class of verbs is not accessible to borrowings and foreign verbs are adopted as the construction “light verb (meaning ‘to do’) + foreign verb”, which can then be reanalysed as a nominal complement of the light verb. Turkish is such a language, which has shown the same pattern in contact with Dutch, German, Norwegian and Bulgarian (for data and discussion, see Muysken 2000 and references therein). Whereas foreign nouns are readily inflected in immigrant Turkish, verbs get adopted using the light verb yapmak. In (2) are two examples from Dutch-Turkish and Norwegian-Turkish code-switching. The parts in the embedded language are italicized.

---

4 Friesner (2009) is being followed here in avoiding the original term *lone* item in order to avoid confusion with the homophone *loan*. 
(2) Ben seninkisini lenen yapmak istedim, toen had ik ze al.
  I yours borrow-INF do wanted, then had I them already
  ‘I wanted to borrow yours, but then I had them already.’
  (Turkish/Dutch, from Backus 1996, via Myers-Scotton 2003)

Studere yapmayı isterdim.
  study-INF do I would like
  ‘I would like to study.’
  (Turkish/Norwegian, from Türker 1993, via Muysken 2000)

In both cases, the infinitive form from the source language is used as the
object of the light verb. In other words, if verbs are turned into nouns, it is
done in syntax. In Korean, on the other hand, both adjectives and verbs
enter the class of nouns and only then derive the form that matches their
original class using productive verbalising and adjectivising suffixes. In (3)
examples from Sohn (2001) are given.

<table>
<thead>
<tr>
<th>(3)</th>
<th>Verb</th>
<th>Adjective</th>
<th>Adverb</th>
</tr>
</thead>
<tbody>
<tr>
<td>miry</td>
<td>‘insensitivity’</td>
<td>miry-an-ha-ta</td>
<td>miry-an-ha-n</td>
</tr>
<tr>
<td>soran</td>
<td>‘tumult’</td>
<td>soran-ha-ta</td>
<td>soran-ha-n</td>
</tr>
<tr>
<td>nice</td>
<td></td>
<td>nais-ha-ta</td>
<td>nais-ha-n</td>
</tr>
<tr>
<td>kind</td>
<td></td>
<td>k-haint-ha-ta</td>
<td>k-haint-ha-n</td>
</tr>
</tbody>
</table>

Comparing nice to miry shows that English adjectival stems are treated on
a par with native nouns, i.e. adjectives have to undergo a class switch in
order to be integrated. The fact that code-switching (and consequently
borrowing) is often restricted to certain classes of words has far-reaching
consequences for the internal dynamics of the lexicon fed by borrowing: it
is often the case that a language actually only adds to the part of the lexicon
containing nouns or only to the one containing content words. This is
especially important in conjunction with the fact that the categories to
which borrowing is often restricted (nouns, content words) are the ones
which have independently been indicated as categories tolerating the
most phonologically marked structures in languages (Smith 1997). We
return to the issue of core/periphery differences in Chapter 3.

1.3. Structural restrictions: code-switching vs. borrowing
As pointed out in the preface, the researchers of bilingual production who
worked on corpora of spontaneous speech encountered a major classification problem: how to distinguish a code-switched element from a nonce borrowing. Poplack & Meechen (1998) and Poplack (2004) give an overview of studies which show that this issue arises in most cases of mixed discourse, given the fact that single-word switches are by far the most common instantiations of it. We have already mentioned that phonological criteria for distinguishing code-switching from borrowing are notoriously difficult to establish in such cases. That is why obligatory morphology was widely used as a diagnostic (for methodology, see Poplack & Meechan 1998; for empirical data see, e.g., Adalar & Tagliamonte 1998 for Turkish-English, Eze 1998 for Igbo-English, Mustafawi 2002 for Arabic-English, Budzhak-Jones & Poplack 1997 for Ukrainian-English, Sankoff et al. 1990 for Tamil-English, and Poplack et al. 1989 for Finnish-English code-switching). In order to illustrate this phenomenon, (4) gives an example from Budzhak-Jones & Poplack (1997). In this sentence, the case endings, obligatory in Ukrainian, are attached to the English stem *car*, and the Ukrainian perfective verb was derived from the stem *start*.

(4)

\[
\begin{array}{ll}
\text{chetverta hodyna,} & \text{chy pjata hodyna,} \\
\text{koly vin} & \\
\text{kinchaje praczju.} & \text{maje podibnu rich} \\
\text{finishes work-F.sg.Acc has similar-F.sg.Acc thing-F.sg.Acc again:} & \text{do cary,} \\
\text{zastarutuje caru,} & \text{ide dodomu,} \\
\text{starts car-F.sg.Acc. drives home.} & \\
\end{array}
\]

The main result of the studies focusing on the morphological integration of single code-switched items is that these items follow the pattern of established loanwords (and native items) in the recipient language. From this fact, the authors draw two conclusions: first, that such items should be analysed as borrowings and not code switches, and, second, that nonce borrowings (i.e. the ones which occur rarely in the corpus) show little difference from integrated borrowings. Interestingly, the authors postulate phonological, morphological and syntactic integration as differential features of borrowings (e.g. Poplack and Meechan 1995), but they do not make any claims about the items’ representation in the lexicon of the recipient language. In other words, they do not state the ultimate consequence of their conclusion: if single code-switched items are borrowings, they receive a (temporary) representation in the recipient language. Given what is known about classes of words which can be code-switched/borrowed, this means that minimally all nouns from one language
can receive such a representation in the other language. This actually seems to be the underlying assumption of these authors, as can be read in the following quote from Sankoff et al. (1990:98).

That nonce borrowing has access to all the content words in the lexicon of the donor language, in contrast to the restricted stock of established loans, does not mean that it is an intermediate process situated somewhere between code-switching and borrowing in the more traditional sense. The morphological and syntactic roles of nonce loans are identical to those of established loans, and as processes, the two of them contrast sharply from code-switching.

This picture can also be read off the graph in (5), quoted from Poplack et al. (1989). Note that all the content words are eligible for nonce borrowings.

There have been different views on the status of nonce borrowings as well. Muysken (2000) claims that single items (instantiations of insertional code-mixing, in his terminology) are different in that they are not “listed” in the sense of being “part of a memorized list which has gained acceptance within a particular speech community” (p. 71). This distinction, which takes the interlocutor into account, should be taken seriously. However, there seems to be evidence that bilinguals adapt their nonce borrowings to the level acceptable in the community (San Giacomo & Peperkamp 2008) so that in
nonce borrowing the interlocutor is already assumed to have the used item among her “listed” items.

1.4. **Summary of the chapter and further development**

This chapter has reviewed the existing sociolinguistic literature on bilingual production. In §1.1., it was shown that sociolinguistic literature points to two aspects of community bilingualism which may influence the type of borrowings encountered in the recipient language. First, higher proficiency increases the number of non-adapted items both in bilinguals’ production and in the resulting corpus of borrowings. Second, the correspondences between the SL and RL items become more regular as the contact lasts longer. These findings will play an important role in the development of the proposed model of loanword integration, presented in Chapter 4. In §1.2., it was shown that code-switching processes restrict the input to loanword phonology in a number of relevant ways. First, content words, especially nouns, are the privileged categories in this process, whereas function words and morphemes are the cross-linguistically dispreferred loci of code-switching. Second, code-switching does not tolerate morphologically complex embedded elements and inflections usually get reanalysed as part of the stems they are attached to. The fact that paradigms of embedded items are not available provides an important clue as to the nature of the input for loanwords adaptation. §1.3. presented evidence from research on single embedded items in bilingual discourse that nonce borrowings show very similar patterns to the matrix language items, indicating that the occurrence of single embedded words leads to fast borrowing and that in bilinguals whole classes of words can be eligible for this process.
2.1. Introduction

This chapter provides an overview of the phonological approaches to borrowing. The reader familiar with what is commonly termed *loanword phonology* will be aware that loanwords in generative research are dealt with in two different (and often disparate) ways. In the tradition which usually goes by the name of *loanword adaptation*, borrowing processes are considered. In this field, the word from the source language is taken as the input to the process, and the new (usually modified) version of the same word in the borrowing language is seen as the output. For instance, if the English word *thanks* is entering a language which allows no interdental fricatives and no complex codas, it is an interesting issue whether the [θ] will be replaced and by which segment ([s], [t] or something else) and whether the complex coda will be repaired and in which way (by epenthesis e.g. [tænækæsæ] or by deletion e.g. [tæn]). In this line of research, the most preferred cases (for methodological reasons) are those of total nativisation. Only in this type of cases is it uncontroversial to claim that the recipient language is 'kept constant', because the recipient grammar allows the same structures to surface as before contact. In all other scenarios, things are rather complicated. In terms of our example, in a case where the interdental fricative is imported into the system, but the complex coda is repaired by deletion (so the new word is [θen]) RL is changing on the one hand (in the sense of the segment inventory), but stays the same on the other (in the sense of syllable structure). In such a case, one can never be entirely sure whether deletion is caused by the old RL (the one which bans [θ]), or the new RL (the one which allows [θ]). For this methodological reason, a huge majority of data in the loanword adaptation literature avoids discussing the cases where loanwords *import new structures into RL*. However, since importation is as common as nativisation, it has also become of interest to phonologists, but in a very different type of research: one on special treatment of loanwords in synchronic grammars. In our example, if the recipient language imports complex codas (so the new word is [tænks]), but still keeps applying deletion to native forms which would surface with a complex coda, then it is interesting to see how complex codas in loanwords are protected. This line of research is thus mainly interested in the synchronic situation which is a result of borrowing. In could therefore be
called borrowedness research, because there is little interest in the source and consequently also in the borrowing process.

Both borrowing and borrowedness research are couched in generative frameworks and therefore offer synchronic views of single languages, the difference being that in the former the synchronic system is seen in action, operating on novel items, whereas in the latter the synchrony is one of post-contact type, where foreign items have been accommodated for and protected by language-internal mechanisms. Given that there have been very few attempts at integrating the two approaches into a single model, they will be discussed in separate chapters. The current chapter is therefore devoted to borrowing research, whereas borrowedness research is discussed in the next chapter.

Borrowing research has introduced a rather different framing of both the goals and the scope of research from the sociolinguistic approaches presented in the previous chapter. In what follows, an attempt will be made to point out the main differences.

In the sociolinguistic tradition dealt with in the previous chapter, the object was the contact itself, the in-between entity which cannot be properly assigned to either of the languages in contact. In this tradition, contact phenomena are interesting by definition and if a specific contact fact happens to provide a window on the grammar of one of the languages in contact, this is a welcome addition, but it is by no means part of the raison d'être of the field. Moreover, windows of this type are not frequently observed in contact research. It seems that the focus on the in-between and the speakers of the in-between (i.e. the bilingual speakers) makes it difficult to consider separate languages as relevant (or, possibly, even real) objects of interest.

In the formal approaches, there is no such obvious justification for grappling with contact data. Moreover, in the generative tradition, in which these formal approaches have been couched, not every language-related fact which shows some regularity is immediately recognised as legitimate object of linguistic research. The so-called performance effects (opposed to the competence effects) such as regularities related to frequency and low-level phonetics are systematically seen as real, yet not relevant to grammar proper. All this makes it a non-trivial task not only to motivate the introduction of loanwords (and other contact facts) into generative research, but also to define what qualifies as a loanword problem and which conditions a solution of this problem has to meet.

The issue of delineating the object of interest is complicated by the fact that in this field, the theories which are applied to loanwords are the ones which have been developed for grammars of single languages (often idealised
to be spoken by a homogenous community). As we have seen in the previous chapter, contact researchers have had good reasons to view borrowing as a purely diachronic process (see Backus 2010). In the generative rendition of the borrowing facts, an attempt can be seen to view the process as synchronic. There are precursors for this move in the language contact-literature. For instance, Haugen (1950:213) points out that “borrowing as here defined is strictly a process and not a state, yet most of the terms used in discussing it are ordinarily descriptive of its results rather than of the process itself”. Also, Weinreich (1953) shows that behind all contact effects there is a synchronic contact at the micro-level (i.e. individual level). He states that “two or more languages will be said to be IN CONTACT if they are used alternately by the same persons. The language-using individuals are thus the locus of contact” (p.1; emphasis in the original). These considerations seem to have been taken into account by generative linguists, but the result is different: what is called loan phonology or loanword adaptation extends to other phenomena, the ones which are more easily viewed as synchronic, psychologically ‘real’ and reflecting the competence of the native speakers, such as L2 acquisition, perception of novel sounds, foreign accents, etc.

Based on such considerations, it will be our goal to cover at least three aspects of the formal models of loanwords. First, an attempt will be made to make clear what kinds of facts the specific models are focusing on and what kind of speakers are considered. Translated into the terms of languages in contact, it will be asked whether the model focuses on the in-between of the language contact or on one of the languages involved. Second, we shall try to make as explicit as possible what a loanword problem is in a formal approach. Finally, the question will be asked what it takes to solve this problem.

2.2. How loanwords became formal data: Rule-based approaches to borrowing

This section deals specifically with the formal analyses of loanword phenomena before the advent of constraint-based approaches in the early 1990s. These early analyses are important both in the sense that they have framed the discourse of formal accounts of loanwords and that they still contain important conclusions pertaining to present-day discussions.

An important feature of the early analyses is that they make serious attempts to provide justification for taking loanwords as the object of linguistic interest. In that sense, for this period it is easier to answer the question of what a/the loanword problem looks like and what the solution to
Chapter 2

This problem is than it is for later approaches. However, in this pioneering work, the problem of the definition of the object is much more pronounced and we shall see that the earliest theories have been claiming to cover phenomena as different from each others as foreign accents and loanword integration.

The most important common methodological property of this period is that the authors were thinking in terms of rules (i.e. mappings between levels of representation) and that loanwords were seen as so-called external evidence for the psychological reality of rules. So, unlike in the subsequent era of constraint-based approaches (to be discussed in §2.4.), where processes in loanwords are ideally accounted for by the constraints which are already present in the language, the initial idea of rule-based theories was a more modest one: to see if (at least some of) the postulated rules apply to loanwords and use this as external evidence for their reality. In what follows we will present four examples of loanword accounts proposed in this era.

2.2.1. Loanwords as justification of phonological grammars: Hyman (1970)
The account of Hyman (1970) stands at the very beginning of generativists’ modelling of borrowing. Not only is his article telling of the generative discourse of the hour, but it has also established many aspects of the way in which loanwords are still discussed by generativists. For Hyman, the issue is not that loanwords should not remain unaccounted for, but that there are so many plausible accounts of linguistic data and so few ways of testing them empirically. Explaining the need for criteria which can be used “in a more direct, perhaps more experimental way [...] to verify the psychological reality of a given proposal” (p. 3), Hyman quotes Kiparsky’s (1968) diagnosis that “what we really need is a window on the form of linguistic competence that is not obscured by factors like performance” (p.174). The importance of not being “obscured” by performance is determinant for the way in which loanwords are presented in this article. First, Hyman occasionally presents no surface form, but an intermediate form to which some additional rules apply, but in ways which are considered phonologically irrelevant (e.g. “Whenever the predictable labialization or palatalization from LR/PR is not crucial to the example, the unassimilated consonant will be given.”, p. 16). Second, “borrowing and (monolingual) cross-linguistic perception” are “treated as one”. Third, Hyman only covers the cases in which “the forms resulting from both forms of contact (borrowing and foreign language) are possible lexical items in the first language”, even excluding those lexical items which would require the feature [+foreign]. Finally, this author is explicitly not interested in variation among speakers and “cases where a particularly talented speaker of the interpreting language accurately
perceives and reproduces foreign sounds in the process of foreign language acquisition” (p. 7). All these characteristics have influenced the history of generative approaches to loanwords, mostly in the sense of becoming implicit criteria for what is considered to be loanword data. Especially the final point is important: it not only implies that the monolingual situation (and L2 acquisition) is the norm, but it also excludes from the object of interest the cases where this norm is violated because of the learning skills of specific speakers. This is an important and necessary step in marrying borrowing (for which the terms “lexicalisation”, “nativisation” and “adaptation” are also used) and foreign sound perception, because the latter, unlike the former, is virtually absent in bilingual communities.

This extension of the object of interest is important for the definition of the task of research proposed by Hyman: he states that “if we have a theory of borrowing (encompassing the perception of foreign sounds in other contact situations), then by analyzing the occurring borrowed forms and/or running the necessary tests on foreign sound perception, various aspects of the internalized phonology can be determined” (p. 8). In this context, the most important question is “if a foreign word is totally assimilated, what determines its lexicalized shape?” (p. 8) Answering this question, Hyman starts one of the most important dichotomies in generative loanword research: that between the phonetic and the phonemic/phonological approach. He further distinguishes between the phonological stance and the phonemic stance, where the latter would be reduced to “taxonomic phonemics” which only uses distributional facts to determine what is a phoneme. In order to illustrate the two stances, Hyman selects a number of early precursors, whose theories are elaborated in modern terms, to an extent (re)creating a history of ideas in loanword phonology. First, he illustrates the approach termed “phonetic approximation” using a quote of the Neo-grammarian Hermann Paul (1966).

Um eine fremde Sprache exakt sprechen zu lernen, ist eine Einübung ganz neuer Bewegungsgefühle erforderlich. So lange diese nicht vorgenommen ist, wird der Sprechende immer mit denselben Bewegungsgefühlen operieren, mit denen er seine Muttersprache hervorbringt. Er wird daher in der Regel statt der fremden Laute die nächstverwandten seiner Muttersprache einsetzen und, wo er den Versuch macht, Laute, die in derselben nicht vorkommen, zu erzeugen, wird er zunächst fehlgreifen. (p. 394)

This historical reference was replicated recently, for instance, in Calabrese (2009). Importantly, Paul is actually dealing with phenomena of L2 learning.
Hyman falsifies the “phonetic approximation” stance using an example from L2 production: the fact that the English interdental fricatives [θ] and [ð] are rendered as [t] and [d] in Serbo-Croatian, but as [s] and [z] in French, whereas both Serbo-Croatian and French have [t,d,s,z]. Hyman shows how a system of phonetic features alone would fail to predict the selection of these values, since it would predict both languages to select the same substitution.

In order to get closer to a viable account of the process, Hyman touches upon the concept of Sprachgefühl, which he characterises as “intangible”, but still useful to show that “foreign sound adaptation is mental in nature”. He also quotes the following statement from Haugen (1950) as evidence that “phonetic approximation was insufficient to phonemicists”.

Neither the speaker himself nor the linguist who studies his behaviour is always certain as to just what sound in his native tongue is most nearly related to the model. Only a complete analysis of the sound system and the sequences in which sounds appear could give us grounds for predicting which sounds a speaker would be likely to substitute in each given case. (p. 215)

Although these considerations are useful for motivating the phonological approach, Hyman is very clear that he “would disagree with much of this statement of Haugen’s”, most importantly because “to a certain extent a speaker of a language can predict how a foreign sound will come into his language” because this “may well be part of the speaker's tacit knowledge of his language” (footnote 7 on p. 13). This is actually illustrative of a major difference between the sociolinguistic approach and the formal approaches. Whereas sociolinguists are satisfied with a list of probable outcomes of language contact, recognising that one of them will be conventionalised in the late stage of bilingualism, generativists are trying to squeeze the temporal dimension (and with it the historical contingencies) out of language contact and produce a theory which is able to predict the outcome based on knowledge of the two systems (or even one system). For this reason, Hyman finds his real early precursor in a Prague functionalist, Polivanov (1931) who, according to Hyman “sums up the phonological relevance to borrowing”.

En entendant un mot inconnu étranger [...] nous tâchons d’y retrouver un complexe de nos représentations phonologiques, de le décomposer en des phonèmes propres à notre langue maternelle, et même en conformité avec nos lois de groupement des phonèmes. (p. 80)
This statement by Polivanov has become the favourite early precursor quote, figuring in later work from Lovins (1975) to Boersma & Hamann (2009). In order to appreciate its importance, the perspective of the monolingual (and L2 learner) is necessary: in a community of bilinguals, speaking of 'our' representations, 'our' language and 'our' laws is not very insightful, just like the foregrounding of the scene when one 'hears an unfamiliar foreign word'. This is, however, entirely in line with Hyman’s restricted interest in loanwords: only insofar they are data which shed light on the representations and rules of the borrowing language. In this sense, for Hyman there is no loanword problem: loanwords can only prove more or less useful on the general test for linguistic data: helping evaluate different linguistic grammars. In this vein, the main part of the article is devoted to “determining to just what extent 'foreign sound adaptation' can be utilized to justify phonological grammars” (p.6).

The final choice for the phonological over the phonemic approach becomes clear in Hyman’s analysis of Nupe, where the changes in “lexicalisation” and “the Nupe accent” are considered the same thing. In his account, Hyman proposes and proves correct three hypotheses, which, as we shall see, underpin the phonological stance.

a) Foreign sounds are perceived in terms of underlying forms (of RL).

b) Foreign segments equivalent to native segments derived by rule are lexicalised as the corresponding underlying forms.

c) When a foreign segment appears in an environment in which the equivalent native derived segment does not appear, then the form of the incoming foreign word is modified so that the structural description of that rule is met and the segment in question is then derived in the appropriate environment.

The most concrete (and least controversial) example to illustrate these generalisations comes from the distribution and nativisation of sibilants. In Nupe, [s] and [ʃ] are in near complementary distribution and [ʃ] can analysed as the allophone of /s/ before front vowels: /sī/ becomes [ʃi] ‘to buy’. However, there are cases where the sequence [si] surfaces, but this is always in reduplications. For instance, /REDsī/ surfaces as [ʃiʃi] ‘buying’, but /REDsá/ surfaces as [ʃiʃa]. This is one of the reasons why the strictly “phonemic stance” would count /ʃ/ as a separate phoneme, whereas a phonological stance would have a way of excluding it from underlying representations (by ordering the palatalisation before reduplication). To summarise, both [si] and [ʃi] surface in Nupe and /ʃ/ is not a phoneme for the phonological stance, it is one to the phonemic stance. This is where loanwords kick in. The fact that the Yoruba [sísì] ‘sixpence’ becomes the Nupe [ʃiʃi] is taken as evidence that the Yoruba surface form is stored as a Nupe underlying form /sísì/, to which
all the Nupe rules apply. The lexicalisation part is crucial here. As Hyman puts it, “a Nupe can pronounce [si], but only under certain phonological conditions. A phonetic or phonemic analysis fails to perceive this” (footnote 10 on p.19). This is how the phonemic stance is falsified.

Evidence for the other two generalisations can be illustrated by the Hausa [ŋù:ɡábá] ‘leader’ becoming the Nupe [jìgàbá]. In accordance with (b), [ʃ] is lexicalised as the underlying /s/, but in order not to end up with the surface form *[sùgàbá] the following vowel [u] had to be turned into a front vowel [i], because only the latter triggers palatalisation. This is, in its turn, in accordance with (c).

An obvious question is why the same kind of repair does not apply to preserve the incoming [u], in which case *[sùgàbá] would have surfaced. Also, it remains unclear where in the proposed architecture of grammar this repair applies. Hyman’s answer is that the attested result “must mean that the palatalization of the initial sound has caused the Nupe speaker to perceive the vowel [u] as [i], the corresponding front vowel that would make palatalization possible” (p. 30). It remains unclear whether perception is invoked as a separate process preceding lexicalisation or not. It is at any event an unexpected way out in a model where Nupe speakers are able to ‘recognise’ phonemes which never surface in their language. For instance, Hyman argues that Nupe has phonemes /ɛ/ and /ɔ/ which are always surface neutralised to [a]. Consequently, the fact that the Yoruba word [tɔre] ‘gift’ is borrowed into Nupe as [tʰɔrɛ] is evidence for three rules of Nupe: (1) palatalisation before [ʃ], (2) labialisation before [ʒ], and (3) neutralisation of both [ʃ] and [ʒ] to [a].

Importantly, there is another missing piece in Hyman’s model: it simply does not cover the probably most common situation: the one where the sound from SL is entirely foreign to the RL, i.e. when it is neither a surface segment nor a phoneme. Actually, the example of the substitution of the English [θ] and [ð] in French and Serbo-Croatian is a case at hand. Hyman’s methodology was applied to this case by Janda et al. (1994) in order to produce what they characterise the most “devastating” argument against Hyman’s model. According to them, if Hyman is right, one can claim that “French [s, z] derive from underlying (English-like) /θ, ð/ because many French speakers substitute [s, z] for /θ, ð/ when they speak English.”

It seems more probable that accounting for these examples is simply not part of Hyman’s programme of loanword research, because cases like this can never be evidence for any process of the recipient system and are in this sense simply uninteresting. We should once more bear in mind that although Hyman does propose dealing with loanwords, he by no means argues that this should be done for loanwords’ sake. Alternatively, the French
substitution would indeed be evidence in favour of a theory which derives [s, z] from /θ, ð/, but it would by no means be sufficient reason to propose such a theory. Once again, Hyman is very clear on loanwords being welcome external evidence for phonological grammars, which is very different from saying that loanwords should be the place to start in constructing such grammars.

While much of Hyman’s article is devoted to proving the existence of phonemes which never surface and mechanisms which are presently considered obsolete, the ingenuity of Hyman’s proposal is elsewhere: in viewing borrowing as lexicalisation. The idea of borrowing as lexicalisation is one which has not received enough credit or enough application in subsequent work. Because this idea will be one of the main ingredients of the model presented in Chapter 4, consider once more how this idea is deployed in Hyman’s contribution.

It is important to bear in mind, however, that although I make constant reference to attested borrowed forms, all of the processes illustrated are still productive and are observable in situations where Nupes speak a foreign language with a ‘Nupe accent’. Thus it is not the case that a handful of borrowed words suffice in themselves to motivate the claims made in this paper, but rather these borrowed cases should be seen as an institutionalized reflex of a more general on-going process, that of perceiving and reproducing foreign sounds and sequences. In other words, both the borrowed words which have been lexicalized and have become part of the Nupe language, and the way Nupes ‘repeat’ non-Nupe words provide verification of the reality of both the morpheme structure conditions and the phonological rules in a grammar. (p.27)

Hyman’s equalisation of lexicalisation with “Nupe accent” is the most problematic aspect of the data presented here: the accented speech is assumed to involve lexical representation in the L1 lexicon of the speaker. There are passages where this assumption is clearly present, e.g. “if a Nupe were to say the Hausa word âkwí:yà ‘goat’ with a Nupe accent (or, equivalently, borrow and lexicalize it), one can predict that it would come out as underlying /àkúyà/ with a phonetic realization of [âkwúyá]” (p.32).

The theoretical value of Hyman’s model is as large as the influence it has had in conceptualising loanword research, but the applicability of his analyses is somewhat obscured by two aspects of the model: its exclusive focus on monolingual situations (and late L2 acquisition) and the fact that a
plethora of terms and concepts from distinct contact phenomena are used as equivalent. As will be shown in the remainder of this chapter, the task of disentangling the contact processes has been taken up by many, but a certain level of ambiguity concerning the borders of loanword research will remain.

2.2.2. Loanwords as evidence for innate processes (Lovins 1975)

Julie Lovins’ (1973, 1975) dissertation on Western loanwords in Japanese (accessed here through its 1975 slightly revised version) stands out for its richness both in the sense of the presented data and of the theoretical discussions. Lovins frames her intervention as a result of her dissatisfaction with treatments of loanwords from the perspective of lexical features and especially “such notions as ‘distance from the lexicon’ and ‘lexical feature’ in a Sound Pattern of English framework" because such treatments provide “no clear correlate to ‘degree of assimilation’” (p.1). Even more importantly, Lovins explicitly argues for a dynamic view of loanword treatment because “the structure of a lexicon is the product of the phonological system whose workings are so nicely demonstrated in the process of borrowing, and the understanding of loan phonology is therefore logically prior” (p.1). After having reviewed the loanword literature in this and the following chapter, we shall be able to conclude that the impossibility of taking the perspective of both grammar and the lexicon into consideration when analysing loanwords is one of the hallmarks of all loanword research. This is one of the main problems we are trying to address in the chapters in which the new model is presented.

Although Lovins’ dissertation is extremely rich, her approach is also very characteristically restrictive: she considers and theorises data from very different sources, while still drawing a clear line around what she considers relevant. In order to understand how this separation works, it should be taken into account that Lovins works within a Natural Phonology framework (Stampe 1969). Natural phonologists’ engagement with loanwords focuses on processes, which are innate and widely present in L1 acquisition rather than on rules, which are more language-specific and not motivated in any universal way. In such a view, the static picture of the recipient language system (viewed through constructs such as the lexicon or the segment inventory) is discarded and what loanword research is supposed to provide is evidence of innate substitution processes. As Lovins states, “the content and organization of segment systems cannot be analyzed per se, but only as a

---

5 The Sound Pattern of English: Chomsky & Halle (1968), the early standard theory of generative phonology.
manifestation of a process system. Implicational laws governing phoneme inventories are directly derivable from innate processes” (p. 21). In this framework, loanwords become evidence not only (or not at all) for specific analyses of specific languages, but also (or only) for UG. Furthermore, the concept of rules is used more conservatively and loanwords are becoming a litmus test for the productivity of rules. As Lovins puts it, the common response to the fact that more and more rules were not applicable to loanwords “has unfortunately been more in the direction of ‘there’s something strange about loanwords’ than ‘there’s something strange about the rules we’re writing’.” (p. 32)

As far as the data considered by Lovins are concerned, there appear to be at least three filters which eliminate large portions of contact data from the data set. First, “the most interesting interference phenomena attested by loanwords come to light when the speakers of [RL] who borrow from [SL] are nearly monolingual, or when these mediators are imitated by monolingual speakers of [RL] with no attempt to adjust their speech habits to the phonology of [SL]” (p.6). She carefully discusses references by Haugen, Weinreich and others mentioned above and adds an important quote by Deroy (1956: 247).

Quand il existe plusieurs termes empruntés notoirement à une langue déterminée, le locuteur peut reconnaître certaines correspondances qu’il étend ultérieurement à d’autres emprunts.

All this is registered and carefully discussed by Lovins and she crucially does not exclude such cases because they are less real, but because they are less informative for the specific approach she is taking. In doing so, she conforms to Stampe (1979: 68-69) who states:

Ideally, the native speaker hears foreign words in terms of his native system, treating foreign pronunciation as if it were derived through native processes from forms which conform to the constraints this system imposes on underlying representations.

In this view, knowledge of more than one language is always a methodological complication. The same holds for the awareness of the source and any kind of concept as to how loanwords should be adapted. As Lovins claims, “all these preconceptions may distort [the speaker’s] adaptation of borrowings from [SL] from what it might otherwise be” (39).

The second important filter imposed by the theory used by Lovins is that of the medium of contact: Lovins points out that “weeding out” spelling pronunciations is one of the most important problems on compiling the data
set (p. 47). In this case as well, spelling seems to create too much deviation from the ideal borrowing situation and the elements which come from spelling are not saved by the fact that they lead to pronounced forms.

The final filter is probably the most interesting one for our purposes. Lovins does mention lexicalisation, but she explicitly does not deal with morphological facts. The importance of morphology is discussed in just one footnote, where she, on the one hand, admits that “in terms of the overall picture of borrowing phenomena, one must of course recognise that phonological interference is only one determinant of the eventual shape of [RL] forms containing borrowed material”, but also remains assured that “for a purely phonological investigation, the relevance of e.g. morphological complications is that they may be misleading” (p. 119). This is once again a useful reminder that in this period the goal of loanword research is not to predict the future lexical item of RL, but to shed additional light on (RL) grammar, or in the case of Natural Phonology, on the innate grammar. Fortunately for the reader who is less restricted in terms of theoretical frameworks, Lovins often violates her own filters in order to present the full picture. So, in her discussion of consonant gemination in Japanese (e.g. in *bukku* ‘book’ and *pabbu* ‘pub’), she accepts the explanation proposed by Ohso (1971).

The basic syllable pattern of English is CVC, while that of Japanese is CV, there is a close tie between VC of English CVC. If we change CVC to CV-CV in borrowing, the auditory impression of the original word is spoiled to a great extent, Thus we insert the phoneme Q to retain this close tie.

This account is not integrated in the bigger picture of Natural Phonology in an explicit way and there is no mention of its being a natural process. Probably for exactly this reason, Lovins offers interesting observations on data exactly from the domains which have been excluded. So, for instance, she discusses the fact that the versions of some borrowings with degeminated voiced consonants (e.g. *pabu* ‘pub’) are considered “modern and, despite their lesser degree of conformity with pronunciation of the source forms, more foreign-sounding than their geminated counterparts”. In accounting for this case of what could be dismissed as “speakers’ preconception”, she invokes both the native ban on voiced geminates and the spelling. She then concludes that “it seems that linguistic potential (the tendency to avoid violating constraints of the language) is actualized in cooperation with social factors”. Here, the entanglement of the included and excluded factors is well illustrated in passing. Furthermore, Lovins provides a partially
Generative approaches to borrowing

morphological explanation for some cases of gemination. So, for borrowings such as atatt’imeNto ‘attachment’ and uittonesu ‘witness’ where gemination occurs stem-finally, she states that “stress on a stem-final syllable before a derivational suffix in English assures that the stem-final consonant belongs to the stem, phonetically; therefore it is geminated in borrowing”. Also this entanglement of morphology (or rather, morpheme identity preservation) and what is claimed to be phonetics is not further commented on. It is important to note, however, that this does not necessarily present a problem for Lovins’ view, since no morphological awareness on the side of the borrowers is presumed.

In order to illustrate the cases in which more formal claims are made, we replicate the account of the epenthetic [o], which often occurs after [t] and [d] (baketto ‘bucket’). The reason for the emergence of [o]-epenthesis seems to be that the general epenthetic vowel /u/ (pronounced as the unrounded [u] and also often voiceless [u]) causes sibilisation (e.g. baketsu, which is also attested). Lovins uses this example to additionally qualify Hymans third generalisation on loanword lexicalisation, which we repeat below.

When a foreign segment appears in an environment in which the equivalent native derived segment does not appear, then the form of the incoming foreign word is modified so that the structural description of that rule is met and the segment in question is then derived in the appropriate environment. (p.40)

The initial step is: perception of “[t_u] for /t/”, in which, “since silence and [u] are most closely related perceptually, the former is interpreted as the latter” (108-109). Then “lexicalisation begins by running backwards through the process (a sort of null process, but still valid)”. The process is the one in (6).

\[
(6) \quad t \rightarrow t / \_V
\]

\[
[-hi]
\]

This leads to the analysis of [tu_u] as [to], i.e. “the environment is altered (by Hyman’s third principle) to allow [t] to remain unchanged” (p.109). However, since [to] “is not an acceptable underlying sequence”, “it is further backwards derived to /to/ by the vowel-devoicing process” (p.109). This means that devoicing has to extend to [o], which is new, because, as Lovins mentions before “non-high vowels are not normally devoiced at an average speaking tempo”. But this is exactly the point for Lovins: that here the devoicing process has been extended to a vowel which violates its original structural description “because /u/ has been eliminated from consideration by the
previous process” (p.109). The revision of Hyman’s rule comes in a quite baroque disguise.

Given dominant \( a'\rightarrow A \), and \( a \rightarrow a'/\_b \), and \([a'c][SL]\), then if \([a'c][SL]\) is interpreted as a sequence (i.e., without reference to the dominant process), \{b\} will be generalised in perception to \{b,c\} and \([a'c]\) lexicalised (or semi-lexicalised) as /ab/ if this contradicts no prior process. If it does, /ac/ may be chosen. (p.109)

In this case, the important process is the creation of the environment \{b,c\} which Lovins describes as a “generalisation operation”. In other words, the vowel devoicing process has extended to both environments here (a result of which we no evidence in actual Japanese). The point is that “the environment is generalised to cover c as well as b, not ‘changed’ from the latter to the former” (p.109).

There are at least two important features of this account which are worth taking into consideration here. First, on the whole there is indeed a generalisation, but the derivation with many steps invisible at the surface obscures the simple version of the generalisation: that the English dental stops are always lexicalised as – Japanese dental stops, and that the epenthetic vowels, which correspond to nothing in English, are modified to make this possible. Second, and related to the first point, by making the epenthetic vowel part of the initial step, Lovins precludes an explanation of why [t] becomes the focus of preservation, rather than [ŋ]. In other words, insisting on the speaker’s naiveté about the form in the source language, she actually limits the explanatory apparatus to whatever can be construed as a native process. This complicates the unification of accounts for the phenomenon we have just seen and, for instance, that of consonant gemination where, as we have seen, some meta-awareness seems to be involved.

In sum, Lovins has provided a very rich resource for loanword research which is still relevant today. It appears, however, that the most interesting ideas emerge where there is no pressure from the rather limiting theoretical framework to which Lovins attempts to contribute.

2.2.3. Rules across languages: Hyperforeignisms as evidence for linguistic rules (Janda et al. 1994)

Janda et al.’s approach to language contact makes them a clear outlier in the history of generative approaches to borrowing. This article never became an influential reference in loanword phonology, perhaps precisely because it conceptualises the object of phonological interest in loanwords in a radically
innovative way, which leaves little room for interaction with the main line of research. It not only brings about a radical change of perspective, but also does so by strictly applying the very generative apparatus and looking for all the mappings which have to be represented in actual speakers.

Janda et al. focus on hyperforeignisms – cases in which loanwords receive features not present in the original word, added because they are associated with the source language by the recipient language speakers. A case in point is the American English rendition of the French *linger*[i] as *linger*[ej]. The final stressed [ej] is the French-sounding element (as in *passé*, *blasé*, *papier-mâché*, etc.) added to the French word which ended in the not especially French-sounding [i].

In is worth noting that this article was published in the 1990’s, when rules had on the one hand become much more concrete than those from Hyman’s and even Lovins’ frameworks, but on the other hand also started being seriously questioned by the proponents of constraint-based theories. In this specific context, hyperforeignisation is invoked to substantiate two specific claims: (a) that rules are real because hyperforeignisms “often show speakers display systematicity and productivity in the application of general patterns to novel non-native context” and (b) that “many (if not most) rules are actually ‘local generalisations’ which typically range over lexically quite limited sets of data, rather than vast, more globally defined linguistic domains” (p.67).

The authors argue that hyperforeignisms constitute a case of overgeneralisation which “results from acting on an organized perception about a pattern that is thought to hold generally, and hence in particular instances is also assumed to be the case: i.e., a rule” (p. 72). The set of hyperforeignisms ranges from word-specific over-generalisations of the type *linger*[ej] to cases in which segments and prosodic patterns are “boosted” because they are associated with certain source languages. So, for instance, they discuss the fact that “speakers of English tend to treat the palatal(-alveolar) fricatives /ʃ,ʒ/ as generic foreign (non-English) consonants – apparently à la Pseudo-German and Pseudo-French, respectively” (p.80). Examples of some hyperforeignisms of this type are listed in (7).

(7)  Original  Hyperforeignisation
      Swedish:  [s]morgasbord  [ʃ]morgåsbord
      French:  parme[z]an  parme[ʒ]an
      Mandarin:  Bei[tɕ]ing  Bei[ʒ]ing
      ((tɕ] is typically perceived as [dʒ])
It is important to note that [ʃ] and [ʒ] are not simply inserted at any place or substituted for any segment in order to increase foreignness, they always emerge at places where a similar more English-sounding segment ([s] for [ʃ] and [z,dʒ] for [ʒ]) is assumed to have substituted the original [ʃ] and [ʒ]. In this sense, the form Aʃerbaiʒan hinges on the existence of the form A[z]erbai[dʒ]an, plus the awareness of the fact that some speakers would nativise a [ʒ] in this context as a [z]/[dʒ]. These substitutions are therefore real cases of undoing what other speakers are assumed to have done, i.e. real hypercorrection. Hypercorrection, crucially, presumes not only meta-analysis, but meta-analysis of how others speak, plus the intervention i.e. the undoing of others’ wrong analysis.

There are, however, examples of hyperforeignisms where the relevant feature of the original word never makes it to the borrowing language and a regular “autochthonous” pattern is established. A case at hand is the English treatment of Japanese prosody (in words ending in a vowel): Japanese loanwords get borrowed into English with penultimate stress, regardless of the original stress pattern. Janda et al. (1994) conjecture that this pattern is extended from Spanish/Italian words, which often end in a vowel as well and typically have this stress pattern.

(8) Japanese English
   Natsuko Tsujimura Natsuko Tsujimura
   Hiroshima Hiroshima
   sake sake

Discussing these examples, the authors move beyond their original purpose of justifying rules and argue against the borrowing metaphor as the default way of conceptualising transfer of linguistic material.

Nevertheless, both "borrowing" and "stealing" imply that what is taken into an L1 from an L2 must necessarily cease to exist in L2, whereas "loanwords" ("thefts") clearly need not disappear from their donor languages. Hyperforeignization, though, gives the coup de grâce to such proprietary metaphors, since one obviously cannot either borrow or steal what doesn't exist. Rather, hyperforeignisms demonstrate that biological analogies like "replication" or "cloning" are much more apt for lexical interchanges between languages. Indeed, there exists a terminology along these lines which goes all the way back to the work of Haugen (1950) and Weinreich (1953): in particular, the target of copying from one language into another can be considered as a model, and the result of such copying as a replica.
And, just as mutations can potentially occur in the course of biological replication or cloning, so one can misperceive or even, in a sense, hallucinate one's linguistic model, thereby creating either a deformed copy or a copy of nothing — a hyperforeignism. (74 - 75)

For the discussions of the hour, this was an considerable jump from only “intralectal” external evidence (i.e. within one language) to “interlectal” evidence (across languages) and it could seem that this article constructs a linguistic problem where grammar rules stop and the vague area of surface generalisations begins. However, Janda et al. argue in a rather persuasive way that the process underlying hyperforeignisation – overgeneralisation through hypercorrection – may be relevant for a larger part of linguistic behaviour than we are used to thinking. They also carefully navigate the generative discourse by building the arguments around intuitions of monolingual (American English) speakers, which, and this is exactly the point, are as real as any intuitions within the “native” system.

Most importantly for our purposes, Janda et al. discuss a range of phenomena which all crucially involve what both Hyman and Lovins exclude from their object of research: the awareness of the SL of the word (correct or misperceived) and a theory of what SL words are like determine the treatment of the structures encountered in the word. Janda et al. also correctly conclude that hyperforeignisms are an instance a widespread, yet under-theorised phenomenon of hypercorrection. After a discussion of Labov’s (1989) data on $\theta$-tensing in Philadelphia English, which show that many native forms seem to be computed as “self-conscious hypercorrections”, they state:

We must conclude, then, that hypercorrection is a potentially pervasive phenomena which must always be taken into account in situations where there is contact between speakers of different language-varieties. And so, even in the case of exotic-seeming hyperforeignisms which may involve interactions between genetically unrelated languages, we are never really far from the garden-variety hypercorrection that occurs so commonly back home, in our first, native language. (p. 77)

This is an important similarity between Lovins and Janda et al.: both claim that loanwords can help us make universal claims. In this case, the awareness (or a theory) of the language of others and normativity feature as inherent properties of linguistic behaviour.
It is important to note that Janda et al., at least as far as borrowing research is concerned, are proposing a counter-theory (or a series of counter-examples followed by a hint of a meta-theory) and are in that sense not comparable to more complete models. However, they make the crucial point that even in the most formal approaches it would not be correct to assume that speakers can ignore the existence of other speakers and avoid any kind of construction of a theory of how others speak. In that sense, the example of the prosodic pattern applied to Japanese words (in English and many other European languages) remains non-theorisable for most approaches to borrowing.

2.2.4. Multiple scansions: Silverman (1992)

Silverman (1992) set the stage for the introduction of constraint-based approaches into loanword research. Moreover, this account had a crucial role in establishing goals and limits for this new brand of modern loanword phonology, which will be predicated on interest for loanwords per se. Silverman defines the goals of loanword phonology thusly:

In loanword phonology we seek to uncover the processes by which the speakers possessing one phonological system perceive, apply native representational constraints on, and ultimately produce forms which have been generated by a different phonological system. (p. 289)

This seems to be the first statement in generative literature in which loanwords are presented as interesting *qua loanwords*. Instead of having loanwords serve as external evidence for something within language(s), this new version of the loanword research agenda uses all the other processes (perception, grammar, etc.) to compile the explanatory apparatus whose goal will be to account for loanwords.

There is, however, a crucial aspect of Silverman’s approach in which it is still very much in line with Hyman and Lovins: Silverman seems to describe an automatic process which takes place within any speaker who is using a foreign word in a native context *without a known precedent*, thereby creating a good native-like word. In other words, since all the work is done by the first-case bilinguals, it is not only true that other speakers are doing nothing interesting for loanword phonology, they are also doing nothing at all with loanwords and their possible awareness of the foreign origin of some words is epiphenomenal. The methodology used in the paper confirms that this indeed reflects Silverman’s conception of loanword phonology: he uses elicited forced-choice preference responses as equally informative as existing loanwords.

Having restricted loanword phonology in this way, Silverman discusses
two crucial issues related to it: the nature of the input and the nature of the levels (or scansions, neither of which is reducible to the RL grammar) through which every input is taken in order to arrive at the resultant form. This distinction is an important achievement in Silverman, often lost in subsequent research. The discussion of the general issues has one important methodological shortcoming, however, which is by no means endemic to this article: Silverman uses data from only a single language contact, English loanwords in Cantonese. He does point out that Cantonese is suitable for this kind of consideration because there are very few native alternations, so the grammar (conceived of as a collection of rules visible from alternations) can clearly be distinguished from what happens to English words.

Discussing the input for the loanword phonology, Silverman makes a case for the assumption that “loanwords do not come equipped with their own phonological representation”, but rather as a “superficial non-linguistic signal” (p.289). The distinction, for instance, between aspirated and non-aspirated plosives, non-phonemic in English but phonemic in Cantonese, remains preserved in the relevant words: tie becomes [t³ay], but motor becomes [mo.ta]. This fact necessitates a representation of the input to loanword phonology richer than the underlying representation in English monolinguals. An important question left unanswered by Silverman is the exact source of these forms and whether the speakers base their pronunciation on their own English forms or some other source. In his discussion of the treatment of the English /r/, he mentions that “Cantonese forms are based upon British pronunciation” (p.297). Although this does not take us much further in constructing an explicit theory of input to loanword phonology, it seems to suffice for Silverman’s purposes, since all phonetic detail (which is not contrastive in RL) assumed in the input will get pruned away by the Cantonese representational restrictions which are part of the first scansion.

The surface SL form goes through two levels/scansions, which transform it in order to produce a native-like output. The first stop is made at the Perceptual Level (at which Scansion One applies). At this level, the incoming words are changed so as to conform to the constraints which follow from the “native segment and tonal inventory” (p. 293). Next, the words proceed to the Operative Level (Scansion Two) and the “native phonotactic constraints and preferences” have their say. In order to make plausible that multiple scansions are necessary, Silverman quotes pairs such as [pʰi.lin] ‘print’ ~ [pʰɪn.ta] ‘printer’ in which the same acoustic material [pʰɪnt] has different correspondents. He assumes that at the Perceptual Level, the initial cluster of both words is interpreted as [pʰl] and that the further destiny of the liquid is decided at the Operative Level, where a prosodic tendency – the fact that
Cantonese seems to prefer disyllabic words - decides between deletion and epenthesis. It could be argued that the division of tasks between the levels in Silverman’s model is directly falsified by the results of subsequent research showing that segmental perception can be influenced by native phonotactics (Hallé et al. 1999; Dupoux et al. 1999). However, since the exact source of the SL forms is not clear, the response could be that words such as *print* and *printer* are bound to have a common representation or related representations in proficient bilinguals, who are assumed to be the single agents of borrowing in Silverman's model.

Silverman’s intervention has been determinant for subsequent work in several important ways. The terminological solutions used for the two levels have had consequences for the further development of the theory, especially the less well-motivated term ‘perceptual’. Silverman quotes a study on perception in bilinguals (Elman et al. 1977) showing that Spanish/English speakers’ perception of a *p/b* continuum is influenced by the language in which the experiment is set up. This leads him to conclude that perception of foreign material is “language-set dependent” (p. 296). There seems to be an additional stipulation, however, that segmental and tonal sets fall under this level, whereas, say, the inventory of syllable onsets plays a role on the higher level. Be this as it may, Silverman does not provide any evidence for his prediction that bilinguals actually merge, say, English [i] and [l] in *perception* (even of their own English produced forms) only when they are speaking Cantonese. Moreover, many formulations in the text show that the processes relegated to the Perceptual Level have more to do with representation than with perception. So, for instance, in the definition of the principles on the Perceptual Level, Silverman states the following.
Perceptual Uniformity Hypothesis
At the Perceptual Level, the native segment inventory constrains segmental representation in a uniform fashion, regardless of string position. (p. 297)

This formulation seems to imply that Silverman’s Perceptual Level is actually conceived as adjustment to the native segmental system, rather than as a module that needs to correspond to cross-linguistic perception. Furthermore, discussing the merger of the two liquids, Silverman states that “speakers are only equipped to represent the native feature matrix which most closely approximates the system’s sole liquid” (p. 297). Clearly, it is safe to conclude that the term ‘perceptual’ was selected somewhat arbitrarily and nothing in Silverman’s theory depends on it.

The term ‘operative’, on the other hand, was selected to emphasise that the processes on this level are grammar-like (because they can be formalised in terms of rules), but crucially extra-grammatical in that native words never undergo them. In other words, for Silverman, the Operative Level functions as a loanword-specific grammar. For instance, whereas in English fricatives occur in both onset and coda positions, in Cantonese they are restricted to the syllable onset. Silverman then assumes that on the Perceptual Level, the speakers perceive and represent all the fricatives as fricatives and that it is only at the Operative Level that the syllabic position plays a role. Here if a fricative is associated to an Onset position, it remains as such (e.g. film becomes [fey lm]), otherwise it gets changed into a stop (10). Importantly, nothing of this kind can be registered in native words.

(10) Input        shaft    lift
    Perceptual Level [æf ]  [lif]
    Operative Level  [æp ]  [lip]

It is noteworthy that the second consonant of the coda cluster in these examples never makes it even to the Perceptual Level. Later in the text Silverman argues that “these obstruents are never represented by Cantonese speakers, due to their insufficient phonetic salience” (p. 325). This seems an important complication of the model in that it introduces a description in terms of syllable positions at a level which comes long before the one where we expect syllabic positions to start mattering. Silverman realises this and posits this deletion even before the Perceptual Level. In order to do so, he re-qualifies the Perceptual Uniformity Hypothesis by adding the following caveat: “input whose acoustic phonetic properties cannot be discerned due to its presence in an impoverished context (a context to be determined on a
language-specific basis) is not supplied representation on Scansion One of the loanword phonology.” (p. 325) In addition to complicating Silverman’s system, this addendum supports our claim that Scansion One should not be equated with cross-linguistic perception, as it seems to operate on the output of yet another level.

In sum, Silverman establishes the agenda of loanword adaptation phonology and answers some crucial questions. The input to the loanword phonology is the phonetic signal, which goes through two scansions: a perceptual one and an operative one, after which the new native-like form is attained and the new word can be integrated into the lexicon of the recipient language without further ado.

The research agenda established by Silverman was followed in almost all subsequent research in loanword adaptation, in which researchers have made attempts to reduce the two-layered picture of loanword adaptation. Silverman’s intervention had the important consequence that it defined multiple domains, whose boundaries and mutual relations were not necessarily self-evident. The existence of multiple levels begs the question of how much each one is related to whatever we are looking for in loanword data (e.g., the classical answer from the previous period – evidence for grammars). Furthermore, since the domains are connected serially, it becomes difficult to simply exclude a domain from consideration (by labelling it pre-linguistic or non-linguistic), because it crucially serves as an input for the following level. The most important practical consequence of this subdivision is reflected in the fact that perception becomes a level which is not easily (or self-evidently) conflated with adaptation, as was the case in Hyman, and its correct positioning is becoming an important issue.

As a consequence of the proposed serialism on the level of domains, a clear delineation became part of the task of phonologists after Silverman (1992) in the sense that they needed to explicate both the kind of representations involved in the account and the nature of processes which handle these representations. The issue of delineation has become central to a large body of research, which has made it a central goal to locate loanword adaptation either within phonological grammar or outside of it, in perception. These accounts will be discussed in §2.3. For a large number of researchers, however, situating loanword adaptation between phonological grammar and perception was not the main goal in dealing with loanwords. These researchers, although concerned with creating a theory of loanwords, are also dedicated to a specific theory of grammar, for which loanwords are not necessarily additional evidence, but rather a domain to which the
mechanisms of the specific theory can be extended in an insightful way. §2.4. will present an overview of these, grammar-based approaches.

2.3. Delineation research: The Phonology/Perception debate

The discourse of loanword accounts discussed in this section has been marked by different stances on a single question: whether the mechanisms of loanword adaptation are phonological or phonetic. Two opposing views have been defined, which share the assumption that the whole process of loanword adaptation is determined by the nature of the input to loanword adaptation (i.e. that identifying the nature of the input offers a key to accounting for the whole process). This move turns the issue of the input into the central issue.

On the one side, following and extending Silverman’s assumptions, Perception-based models (§2.3.1.) see the input to loanword adaptation as so phonetically concrete that the only possible process it can go through is perception, in which all the relevant modifications occur. On the other side, the Phonemic categorisation-based approaches (§2.3.2.) claim the input to be so phonologically abstract that the processes it goes through involve abstract-to-abstract transformations, guided by preservation considerations, which cannot always be related to the processes ‘regular’ abstract inputs go through in RL phonology.

The decision to put the two extreme stances in the same category and contrast them with the milder views, which are more tied to specific theories of grammar, might seem somewhat arbitrary. However, the two approaches converge in a crucial aspect: they participate in the same dispute, whose underlying assumption is that there is one basic and natural way loanwords can get adapted, while the other scenarios are derived, secondary and uninformative. The main goal of loanword research conceptualised this way is then finding out what this basic and natural procedure is.

An inspection of the discourse in the field shows that authors from both stances massively discuss work by the authors of the opposing tradition, reanalysing their best examples as results of confounding factors. By now, the two approaches have developed default explanations of the counterexamples to their own theories. So, if there is a residue of phonological mapping in the corpus analysed by perception-based researchers, it is readily explained as orthographic influences (see e.g. Iverson 2005, Vendelin & Peperkamp 2006). On the other side of the scale, perceptual effects, termed “naïve phonetic approximations” in the phonemic categorisation-based literature, are seen as rare attempts of monolingual speakers to perform the type of adaptation normally performed by bilinguals
(the only competent borrowers), when the latter insert non-adapted SL items into the RL discourse (Paradis & LaCharité 2009).

Finally, both approaches carefully select language data which they explore in order to confirm their hypotheses. So, for example the Cantonese sub-phonemic substitution of English voiceless obstruents (discussed by Silverman 1992, above) is the hobbyhorse of the perception-based authors, whereas the Mandarin substitution, in which the aspiration is irrelevant, is discussed by phonemic categorisation-based authors (see Paradis & Tremblay 2009).

2.3.1. Perception-based models

The perception-based model of loanword adaptation (Peperkamp & Dupoux 2003, Peperkamp 2005, Vendelin & Peperkamp 2006, Peperkamp et al. 2008, Dupoux et al. 2011) is usually labelled ‘psycholinguistic’ by its authors. This model makes two important assumptions:

1. that the input to loanword adaptation is unclassified phonetic signal and
2. that most relevant distortions of the foreign elements occur during “phonetic decoding” (= perception), before there is any mapping to phonological representations, which means that loanwords adaptation it extragrammatical.

Figure (11) gives a schematic representation of this model from Peperkamp et al. (2008).

(11) Perception-based model of loanword adaptation
In this model, loanword adaptation lacks any overlap with phonological (production) grammar. This predicts that languages can treat the same structures in foreign inputs differently than they do in domestic underlying representations (contra standard assumptions). Indeed, such patterns are used as one type of evidence for this model: for instance, Peperkamp et al. (2008) quote examples from seven languages.

A second type of evidence presented in support of this approach shows, focusing on the nature of the input, that adaptation processes can make sub-phonemic differences in SL categorical in RL (contra phonemic categorisation-based approach). So, to add another example, Korean consistently adapts instantiations of English [s] as either lax /sl/ or tense /s\'l/ depending on its relative length in different positions (Kim 1999, Kim & Curtis 2002, Iverson & Lee 2006): [s] in a cluster with an obstruent, being shorter, gets adapted as a lax /sl/ (stop > [sit\textsuperscript{b}op], fast > [p\textsuperscript{b}as\textsuperscript{t}\textsuperscript{i}]), whereas in other positions it becomes a tense /s\'l/ (single > [s\textsuperscript{'i}ng\textsuperscript{i}], bus > [pus\textsuperscript{'}i], dance > [tens\textsuperscript{'i}]).

Another important type of evidence is collected from experiments on cross-linguistic perception, which show that speakers actually “hear” repairs typically assigned to loan phonology (contra all “phonological approaches”). There is ample evidence for so-called perceptual assimilation. Dupoux et al. (2008) found that even proficient L2 speakers of Spanish with French as L1 are unable to perceive position of stress in disyllabic words. The same is true for the insertion of elements into the input – Dupoux et al. (1999) and Dehaene-Lambertz et al. (2000) show that Japanese listeners are unable to distinguish between forms with consonant clusters and the ones with inserted vowel (e.g. ebzo vs. ebuzo). Dupoux et al. (2011) show that Japanese and Brazilian Portuguese speakers “perceive” epenthetic vowels when none are present in the input and even when the co-articulation cues of other vowels are present.

Peperkamp et al. (2008) tried to pit a phonetic and a phonological explanation of a phenomenon against each other. In Japanese, English /n/ in coda position gets adopted as a moraic nasal consonant – pen becomes /peN/. French coda-/n/, however, gets resolved by gemination and epenthesis Cannes [k\textsuperscript{a}n] > [kan\textsuperscript{nu}], whereas nasal vowels become combinations of vowel and a moraic nasal consonant · Caen [k\textsuperscript{a}] > [k\textsuperscript{a}N]. Now Peperkamp et al. test two explanations:

(1) Different adaptations are due to the fact that French makes the distinction between oral and nasal vowels phonemically – a phonological account.

(2) French final [n] is longer and more often released than the English final [n], which is why it gets adapted differently – a phonetic account.
In order to pit the two accounts against each other, native speakers of Japanese were asked to perform a forced-choice identification task – they had to match nonce words pronounced by American English and French speakers with a number of options, each containing either a final moraic nasal or a sequence of a nasal stop and the vowel u. Since the speakers applied perceptual epenthesis significantly more often when exposed to English inputs, the authors conclude that the phonetic explanation was proved correct.

There are a number of possible objections to these conclusions, which the authors did not address in their discussion of the results. First, the experiment puts the subjects in the situation which never occurs in actual borrowing – since all the possible solutions in the identification task had either a moraic nasal consonant or an epenthetic vowel, the subjects’ attention was focused on the word final elements (so the options for [d repreh] were: repreh - repreh - repreh - repreh - repreh - repreh and, not surprisingly, they found that there are two types of inputs and did the matching. Real-life borrowers are never confronted with such a task. Second, an important argument for the phonological explanation of the real-life adaptation pattern has been glossed over – the distinction between oral and final nasals is a crucial aspect of speaking French as a second language, since it does not only encode the difference between words, but also plays an important role in the grammar (cf. [bɔ̃] ‘good, masculine’ and [bon] ‘good, feminine’). This increases the chances that the distinction between oral and nasal vowels in French will constitute a part of the language training, which can make the bilingual speakers, who introduce the first instantiation of the borrowing into Japanese, additionally sensitive to this difference.

Accounting for proposed cases of phonemic correspondences in loanword adaptation, these authors have recognised another potential source of adaptation patterns – orthographic information. There has been experimental research which shows that presentation of written sources can influence the outcome of online adaptation (Vendelin & Peperkamp 2006, Kaneko & Iverson 2009). These results are then used to reanalyse the cases in which phonemic correspondences were claimed to play a role. For example, Spanish speakers have been argued to have the inclination, due to VOT values, to hear both voiced and voiceless-aspirated stops in English as Spanish voiceless stops, a prediction which is borne out in early learners of English with Spanish as L1 (see LaCharité & Paradis 2002, 2005 and references therein). However, English borrowings into Spanish always have voiced correspondents for English voiced obstruents (baseball > [besbol] and not *[pespöl]). LaCharité & Paradis conclude that there has to be a mapping between the two phonemic systems. In response to this, Iverson (2005)...
develops an argument which attempts to falsify LaCharité & Paradis’s prediction on theoretical grounds and relegate the patterns in question to orthographic correspondences which are internalised by proficient learners of English.

Although providing a backup source of adaptations might seem like a peculiar addition to the otherwise straight-forward theory, it actually does cut ice in the present discourse of the field: second language acquisition is inseparable from literacy in the modern world. This last fact limits the space in which the proponents of phonemic correspondences can find cases of phoneme-to-phoneme pairings in which the influence of orthography cannot be adduced.

A caveat is in place here about the arguments which disqualify orthography as a confounding factor which blurs the actual loanword integration. Whereas discarding orthography from the SL representation seems an informed choice (already Lovins presented excluding orthographic influences as a self-evident methodological intervention), it is not obvious how the graphic form can be disentangled from all the other representations available to borrowers. Moreover, excluding borrowing situations where the orthographic information available from the data sets may lead into essentialism, where real-life contact situations play an ever-smaller role. What is at stake here is not only the idea that certain representations which are clearly available to borrowers can be excluded, but also the more general notion underlying a large part of phonological modelling of loanword treatment that there needs to be a single form which can serve as the input to the model of language contact.

Finally, although it aims to offer a general theory of loanword treatment, there are a number of issues which this approach does not address. The authors seem to be reluctant to translate the central tenet (“loanword adaptation is L1 perception”) into actual contact situations, so there are very few statements about the agents of adaptation in the specific community. This seems to be symptomatic hiatus, since, as already argued, language contact situations are not quite about cross-linguistic perception. In this sense, the Perception-based model still seems to rely on the assumption (derived from Silverman’s model) that perception in SL/RL bilinguals is somehow equivalent to that of RL monolinguals once they are speaking RL. It should be noted that most of the data adduced in support for this approach come from contact situations in which most of the speakers of the RL have very limited command of the SL and little exposure to it as spoken by native speakers (Korean, Japanese, French), but this is not made explicit by the authors. This means that on the community level, the contact situation in these languages is closer to cross-linguistic perception – even if all bilingual
speakers maintain the native-like form of the new words, the rest of the language community will have all the chances of performing perceptual assimilations.

There is also no mention of the epenthetic bias – the cross-linguistic tendency to repair illegal structures by epenthesis rather than deletion of segments. It seems that the proponents of this theory could again resort to the secondary source of adaptations – in all the cases we know about, the speakers can be argued to have been biased to perceive and consequently preserve certain elements because they were present in the written input. As already mentioned above, this does not necessarily improve the theory, since this step is predicated on the assumption that the written form can be isolated from all the other available forms, but also that borrowings are based on a single representation.

Finally, Paperkamp et al. (2008) briefly address non-adaptations – instances of structures illegal in the recipient language which still get preserved, as a phenomenon which is are outside of the scope of their theory at the moment, but should be accounted for in the future. It seems that the perception-based model would have to give up on some of its central tenets in order to accommodate for non-adaptations. Especially hard to account for would be the variation and gradual assimilation to the native patterns: e.g. as Holden (1976:132) shows, in Russian, where unstressed /o/ is rendered as [a] “a recent, relatively unfamiliar borrowing [...] such as foksterjér ‘fox terrier’ will be less likely to reduce its un-stressed [o] than modél ‘model’, which is much more frequently used”. Crucially, these processes of post hoc adaptation show the same patterns as initial adaptation, yet they cannot be ascribed to perception and occur in items which do have an established phonological representation in the lexicon.

In the following section, we are turning to the other extreme – the phonemic categorisation-based accounts.

### 2.3.2. Phonemic categorisation-based approaches

Accounts based on phonemic categorisation (Paradis & LaCharité 2001, Paradis & Thibeault 2004, LaCharité & Paradis 2002, 2005) are built around the assumption that SL/RL bilinguals, the agents of borrowing, access the SL words in terms of abstract categories, which then get mapped onto RL categories, preserving as much feature information as possible, even when the result misses the phonetically closest target.

These authors include numerous references to sociolinguistic research (see Chapter 1), confirming that bilingual speakers introduce borrowings into the speech community. However, these references do not seem to contain any confirmation for the claim that bilinguals actually have the leading role in
the adaptation of the words they introduce. What one does find in the code-switching literature are methodological reservations in this respect, due to bilinguals’ verbal repertoire, i.e. due to the influence of their accents as a confounding factor.

In order to manipulate the abstract structures which serve as input, loanword phonology developed within this framework makes use of a number of formalisms which are almost exclusively applied to it. The emergent framework is that of the Theory of Constraints and Repair Strategies (TCRS, see Paradis & LaCharité 1993 and Paradis et al. 1993). (12) contains a synopsis of the central formalisms in this approach (based on formulations in Paradis & LaCharité 1997).

(12) Repair strategy
A universal, non-contextual phonological operation that is triggered by the violation of a phonological constraint, and which inserts or deletes content or structure to ensure conformity to the violated constraint.

Preservation Principle
Segmental information is maximally preserved within the limits of the Threshold Principle.

Threshold Hypothesis/Principle
(a) All languages have a tolerance threshold to the amount of repair needed to enforce segment preservation.
(b) This threshold is the same for all languages: two steps (or two repairs) within a given constraint domain.

Minimality Principle
(a) A repair strategy must apply at the lowest phonological level to which the violated constraint refers.
(b) Repair must involve as few strategies (steps) as possible.

Phonological Level Hierarchy
metrical level >> syllabic level >> skeletal level >> root node >> feature with a dependent >> feature without a dependent

Precedence Convention
In a situation involving two or more violated constraints, priority is given to that constraint referring to the highest phonological level of the PLH.

---

6 As noted by Sharon Peperkamp (p.c.) it can just as well be the case that bilinguals reproduce the SL form as accurately as possible, which then creates the cross-linguistic perception situation assumed in her (joint) work.
This approach is similar to Optimality Theory in that it is constraint-based, yet essentially different in that it lacks OT’s surface-oriented component. This is a consequence of the fact that this approach manipulates abstract units. Crucially, phonemes of one language become phonemes of the other language and only then can they be mapped onto surface forms. This assumed precedence makes loanword adaptation different from ‘regular’ phonology in this approach.

With the possible exception of the Threshold Principle, the formalisms just presented seem very reasonable assumptions to many modern phonologists, once it is accepted that there is a transfer of abstract units from one language to the other. It is in this sense that this approach is just as input-based as the one described in the previous section. Another important property of this approach is that its analyses are highly dependent on the assumed representations – note that in this case not only the languages in contact have to receive an adequate representation, but the model also has to give answers concerning the universal feature inventory necessary for a successful transmission. This may well be the reason why this line of research usually does not present its specific predictions of outcomes of loanword adaptation, which should be available just based on the two native phonologies. What these authors usually use as crucial arguments are counts of phonemic and phonetic repairs in various corpora of loanword adaptation (see e.g. LaCharité & Paradis 2005). (13) contains the relevant definitions for making the distinction between the two kinds of repairs from LaCharité & Paradis (2005).

The authors of this line of research have formulated two arguments for the superiority of their approach with respect to OT. First, Paradis (1996) argues that OT cannot replace the Threshold Principle. This indeed seems to be the case, since nothing in OT can count steps the way TS does. Note, however, that this principle is stipulated based on the data from two languages (Paradis & LaCharité 1997) and criticised on empirical grounds by Rose (1999). The second objection is that OT’s segmental Faithfulness is not able to account for cases in which deletion and epenthesis occur in the same language, depending on the relevant structure, since the ranking of the relevant constraints (Paradis 1996 discusses Parse segment and Fill) always favours one over the other. This objection, however, does not hold of fine-grained versions of Faithfulness, which can refer to particular features.
Loanword adaptation is generally based on the L2 (not the L1) referenced perception of L2 phoneme categories (because borrowers are bilingual and have simultaneous access to both codes at the time of the borrowing).

Difference between naive and intentional phonetic approximation: intentional phonetic approximation (importation/non-adaptation), which is characteristic of bilinguals, introduces L2 sounds and structures into L1 whereas naive phonetic approximation, which is characteristic of monolinguals, does not.

Criteria for identifying naive phonetic approximation
A. Phoneme non-perception (resulting in phoneme deletion): an L2 phoneme is deleted because it is not perceived by L1 borrowers.
B. Incorrect phoneme categorisation (phoneme mismatching)
1. perceptual confusion (due to perceptual proximity)
   a. licit phoneme-to-phoneme mismatching: an L2 phoneme that is permitted in L1 nonetheless changes phoneme category in loanword adaptation because the phonetic cues associated with the L2 phoneme relate to a different L1 phoneme;
   b. illicit phoneme-to-phoneme mismatching: an L2 phoneme that does not occur in L1 is adapted to the perceptually, rather than phonologically, closest L1 phoneme.
2. level confusion (mistaking variants for phonemes and vice versa)
   a. variant-to-phoneme mismatching: an L2 phonetic variant is identified as phonemic in L1;
   b. phoneme-to-variant mismatching: an L2 phoneme is identified as a phonetic variant of a different phoneme in L1.

Interestingly, these authors never report any problems or dilemmas in the implementation of these definitions and distinguishing between phonemes and variants. It seems that, given the ontological primacy of the phonemic adaptation in the theory, all the cases where either phonemic or phonetic adaptation can be assumed, phonemic mapping is preferred. Moreover, some of the examples of (possible or actual) phonetic approximation show that the term allophone is interpreted in a very unorthodox way. For instance, regional variants are considered to fall within this category, as can be seen in the examples in quoted from Paradis & LaCharité (2008).

Finally, an L2 phoneme might be perceived, but identified as a phonetic variant of a different phoneme in L1 (5B2b). For instance, the English
phonemic laryngeal /h/ might be identified as [h], which is a variant of palato-alveolar /ʃ/ or /ʒ/ in some Quebec French dialects. Speakers of those dialects may pronounce a word like changer ‘to change’ as [hãhe]. In fact, this type of confusion, though logically possible, does not occur in any of the Quebec French corpora. (p.89)

Furthermore, commenting on the English-Old Quebec French mapping light → [lɛːt], Paradis & LaCharité (2008) state the following.

This particular case may be an instance of a surface variant being taken at face value, since English /aj/ was often pronounced /e/ in early Canadian English (Ahrend 1934:138). However, this would still be a case of phonetic approximation, so it would not affect the overall rates of phonetic approximation in OQF. (footnote 8 on p.113)

Finally, while these authors readily propose unrealised possibilities for naïve phonetic approximation, the explanation for the phonemic pairing is often missing. So, for example Paradis & LaCharité (2008, 2009) make most of their generalisations based on English – Old Quebec French mapping of vowel systems without ever presenting the relevant feature hierarchy, so that it is never explicated why the mapping [æ] → [a] is phonemic, but [æ]→[e] is a phonetic approximation. Moreover, the conceptualisation of phonetic matching also does not seem to be entirely unproblematic, so it is actually not always straightforward which model predicts which outcome. For instance, the predicted phonetic approximation for LaCherité & Paradis (2005) is always based on comparisons of the sounds in isolation. However, Vendelin & Peperkamp (2005) show that French speakers’ perception of English vowels differs if they are presented in isolation and in CVC sequences.

Despite unresolved methodological issues, this research should be credited for fleshing out some examples which might pose a serious problem for perception-only accounts. One of the most convincing correspondences (and one of very few which are more immune to the simple orthographic explanation\(^8\)) is the one which connects Arabic uvular [ɣ] to French [g]

\(^8\) It can very well be the case that the bilinguals introducing these words were aware of the transliteration conventions for the two languages. However, it stands to reason to assume that these conventions will not influence the inter-language mapping as much as the use of the same grapheme does. Furthermore, unlike the sheer fact that many languages use Latin-based alphabets, these conventions have their source in some kind of pre-existent inter-language mapping established by bilinguals and, in this case, one based on a phonological criterion.
([azæl] → [gæzɛl], [maʁrib] → [magœrib]) although there is a phonetic correspondent [ʁ] in French (Paradis & LaCharité 2001:272). The explanation adduced by these authors is then based on the fact that [ʁ] does not function as a rhotic in Arabic, so it has a different feature specification from its French phonetic kin. Since bilinguals transmit features, the phonetic correspondent is too distant.

Authors in this approach regularly report on non-adaptations. Being a constraint-based approach, this line of research can make use of a core-periphery schema by assuming that some of the constraints (or parameter settings) are weaker and allow certain exceptions in non-core sectors of the lexicon. Such an implementation is made in Paradis & Lebel (1994) for Quebec French. As Paradis & Lebel show, some structures are tolerated to a certain extent in borrowings from English, whereas others are banned altogether. It is not clear, however, how the strata are defined in this case, i.e. whether every permitted foreign structure starts its own stratum, as on the graph in (14).

It seems that the implementation of lexical stratification might be problematic for some of the central tenets of the TCRS. Non-adaptations are essentially elements of the SL phonology which are not restricted to the discourse of bilingual speakers, as many examples from Quebec French show. Now, because these words are not restricted to bilinguals, their adaptation pattern is predicted to follow the track of naïve phonetic approximation. However, this does not seem to be the case any more than in the regular adaptation scenario.

(14)

Quebec French Core and Periphery
This line of research has restricted itself to segmental adaptations. An extension to other cases might call for some additional mechanisms, since, for instance, there is ample evidence that suprasegmental information is often preserved even when entirely predictable in the source language (e.g. the French word *attaché* has final stress in Serbo-Croatian, Slovenian, Dutch and many other languages). This opens new questions as to the nature of the input, left unaddressed by this approach.

Concluding the discussion of the phonemic categorisation-based approaches, it is important to address one less discussed aspect of this framework. A very strong prediction follows directly from the assumption that the input to the loanword adaptation is phonemic: reconstruction of SL underlying contrast which are neutralised in SL surface forms. Such “UR-reconstruction” should occur every time an SL segment has the underlying representation /p/ and the surface representation [p'] which can both be reflected faithfully in the RL segments [q] and [q'] respectively. Then the RL form should always be [q], because it reflects the actual input /p/. However this never seems to be the case. Examples of this type were adduced already by Kaye & Nykiel (1979). So, while Japanese forms [hitatsi] and [sasimi] are derived from the underlying /hitati/ and /sasimi/, English (and other languages in which /t, s, tʃ, and j/ are independent phonemes) has forms based on the surface ones - *hitachi* and *sashimi*.

Our discussion of the delineation approaches seems to have justified classifying them together. The moral is: once one type of input is assumed to be the ‘regular’ input to loanword integration and a theory is built around this assumption, the other patterns will necessarily appear exceptional and complicating. Since there are two such approaches in present-day loanword phonology research and every data set will be accounted for more readily using one of them, the general strategy seems to be finding more examples which fit the argued picture. In the present constellation, this means that there is no meeting halfway, but also no actual interaction possible in the current dispute. This impasse has been observed in the literature: introducing a volume devoted to loanword phonology, Calabrese & Wetzels (2009) claim that the two stances really discuss two different borrowing scenarios, which they term nativisation-through-perception (more probable if there are no bilinguals involved) and nativisation-through-production (in which bilinguals insert words in their recipient language discourse). If this is indeed true, then a resolution of the dispute is not to be attained by looking for the ultimate truth on the ‘real’ input to ‘genuine’ loanword adaptation, but in the creation of a complex model which would include both scenarios as
real and genuine. Such hope has been expressed by the proponents of the Phonemic Categorisation-based models in the same volume.

Analogously to the blind men touching different parts of the elephant, and falsely inducing the presence of a snake, a rope, a wall, etc., the two stances presented and advocated for in this volume is necessarily incomplete at this stage of our knowledge and may well turn out to be part of the same system, albeit with different raisons d’être” (Paradis & Tremblay 2009: 222)

In the following section, we turn to grammar-based approaches, some of which have similar assumptions about the input with the approaches just described, yet succeed at situating loanword research with respect to other aspects of the linguistic endeavour in a more explicit way.

2.4. Grammar-based research

In this section we discuss the approaches to loanwords which emerged in the wake of Optimality Theory (Prince & Smolensky 1993, Kager 1999) and have as a common denominator the claim that RL grammar (conceived as a ranking of violable constraints) serves as the main source for the explanation for loanword treatment. Optimality Theory (OT) remains the most widely used theory of phonology at the time this dissertation is written and its formalisms will be used in those parts of our account (presented in Chapter 4) where loanword treatment interfaces with the grammar. It is for this reason that our discussion will also be more chronological, making a clear distinction between the early OT, where loanwords were still used to show the virtues of the theory, and the late OT, in which the standard apparatus is expanded to account for various loanword phenomena. In this sense, to the extent that it belongs there, this dissertation can be read as adding to the tendency which is already present in the OT literature: seeing loanword treatment less and less as grammar proper and more and more as one of the grammar’s many interfaces.

It is first important to situate this line of research with respect to other approaches which may also claim the label ‘grammar-based’: the rule-based approaches (§2.2.) and the phonemic categorisation-based approaches (previous section). In order to do both, we need to explicate one of the main features of Optimality Theory and the linguistic discourses couched in it: since grammars are rankings of universal constraints on surface forms, OT grammars can be applied to any representation and they will turn it into a licit surface structure of the language. In other words, the core of an OT
Chapter 2

grammar is about what is allowed at the surface level, whereas the lexicon is of little importance: there are no restrictions on what can serve as input to the grammar. This last fact has been formalised as the famous principle of the Richness of the Base. In terms of our toy example from the introduction to this chapter, if in a language the English word [tæŋks] becomes [tæŋks], what makes this happen will not be a rule changing the interdental fricative to a dental stop, which would have to be postulated only for loanwords, and wouldn’t necessarily reveal anything about the grammar; rather, it would be a universal markedness constraint militating against interdentals (say *INTERDENTAL), plus a constraint ranking which makes this repair (fortition) more optimal than for instance one involving [s] or one involving deletion. These constraints would be available at any event, so, strictly speaking, nothing would need to be added to the grammar in order to accommodate for loanwords. Moreover, RL grammar would make straightforward predictions about loanword treatment.

Now we can situate the (OT) grammar-based approaches with respect to the other approaches which make use of grammar. First, this line of research shares its application of grammatical tools with the rule-based approaches. However, it is distinguished from them by its scope: while rule-based theories were happy to find some evidence for their independently proposed rules in the way loanwords are treated (but were not particularly interested in loanword-specific rules such as [θ] → [s] in our toy example), the OT grammar-based theories expect all the loanword phenomena to be in principle reducible to grammar, because the grammar already contains constraints against everything it does not tolerate. Second, these approaches share the foregrounding of phonological representations with phonemic categorisation-based approaches. However, while for the phonemic categorisation-based approaches loanword adaptation is very different from what takes place within languages (involving abstract repairs), for the (OT) grammar-based approaches, foreign material is seen just as one of the possible inputs to the same grammar.

It is because loanwords are just one of the possible inputs that in this period loanwords become a more and more common type of data, adduced casually in general discussions. For instance, Pater (1999) shows the virtues of OT by proposing an elegant account for NC conspiracies – constellations where sequences nasal + voiceless obstruent are avoided in different ways: the existence of a universal constraint *NC. Among other facts, Pater then quotes data from OshiKwanyama, a Niger–Congo language spoken in Angola and Namibia, where “loanwords are modified by voicing the post-nasal obstruent” (p. 330) and quotes the examples in (15). There is no further explanation of how foreign forms (and which forms exactly) become the input
of the native grammar. What is important is to show the constraint \*NC\_i at work.

(15) Postnasal voicing in OshiKwanyama loanwords

[sitamba] 'stamp'
[pelenda] 'print'
[øjinga] 'ink'

In the introduction to the volume “Optimality Theory – Phonology, Syntax, Acquisition” the first example quoted by Boersma et al. (2000) to illustrate constraint interaction comes from Dutch borrowings into Sinhalese, where Dutch complex onsets are repaired by epenthesis (16). Boersma et al. also provide the tableau replicated in (17) below.

(16) Dutch loanwords in Sinhalese

<table>
<thead>
<tr>
<th>Dutch loanword</th>
<th>Sinhalese</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan</td>
<td>pälâna</td>
<td>'plan'</td>
</tr>
<tr>
<td>procuratie</td>
<td>perakala:si-ya</td>
<td>'procura'</td>
</tr>
<tr>
<td>kraan</td>
<td>kara:ma-ya</td>
<td>'tap' (i.e. cock or faucet)</td>
</tr>
<tr>
<td>vrouw</td>
<td>porova</td>
<td>'queen' (in cards)</td>
</tr>
</tbody>
</table>

(17)

<table>
<thead>
<tr>
<th>/plan/</th>
<th>*COMPLEX ONSET</th>
<th>MAX-IO</th>
<th>DEP-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>[plän]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pän]</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>* pälâna</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

This tableau is interesting for our purposes because it shows the Dutch form (presented in the original Dutch orthography) as the input to the tableau. Again, the representations are not crucial, because whatever the exact representation may have been, the grammar turns it into a structure which is interesting to us: one with no complex onsets. Both examples illustrate an important methodological point: one part of OT grammars – the surface-oriented well-formedness constraints – can be discussed using loanwords without much clarity on what the exact representation of loanwords is and how they enter the grammar in the first place, since the effect of the constraint is there if a ‘repair’ is there. OT accounts, especially the early ones, which were still showing the suitability of OT grammars for handling loanwords, could therefore afford to assume that there might be a range of representations of the foreign input and safely focus on what all these incoming representations have in common: the structure which violates the
relevant constraint. However, the task of OT loanword models is more intricate, as they are expected to provide clarity on what gets transformed into what and what the status of the input is. The discourse of OT loanword research can be read as gradually moving from providing occasional accounts of loanwords for the purpose of demonstrating the virtues of the theory to providing models of loanword behaviour.

Because they are more frequently considered in general terms, loanwords are more than ever viewed as a type of data which falls under the jurisdiction of the phonological theory. An illustrative example of this new discourse can be found in Hall (2007). This author discusses the option of giving the language-specific segmental inventory an ontological status within OT, as an alternative to letting it emerge from the constraint ranking, as it works in the standard version of the theory. Thinking of possible objections, Hall notes: “The first objection is that the constraint grammar would no longer be either responsible for, or equipped to deal with, non-native input forms; an alternative account would therefore be needed for various loanword adaptation phenomena” (p. 30). This type of concern would hardly have arisen before the advent of constraint-based theories in phonology.

In the following section, we are turning to the literature which treats loanwords in a more systematic way, providing more global accounts.

2.4.1. The architecture of the OT grammars and the return of representations
As already hinted above, as long as native-like repairs (e.g., [nat+a] → [naθa], in a language with intervocalic spirantisation) and repairs to native-like structures ([naθa] → [nata] in a language with no interdentals) are encountered, the part of the OT architecture which we have already discussed – the markedness constraints – remains applicable and no special provisos need to be made for loanwords. It remains generally true that markedness constraints, in terms of their content and ranking, are assumed to be for loanwords the way they are for native words.9 If there were only cases of this kind, loanword research could safely be annexed by the general phonological theory, which would contain nothing loanword-specific.

However, as we shall see, from the very beginning, constraint-based researchers have been encountering cases where the native ranking alone does not seem to suffice and the solution was generally sought in the fact that incoming loanwords are clearly distinct from the representations that OT grammars generally apply to: underlying representations. This insight has been implemented in two different ways: one is elaboration on the role of

---

9One important exception to this generalisation is discussed in Chapter 6.
perception in loanword integration (aided largely by Silvermanian heritage, but also the perception-based models, which developed in the same period), and the other is redefining the set of Faithfulness constraints which are applicable to loanwords (prompted by research on Faithfulness between surface forms initiated in Benua 1997). In the remainder of this section we are giving an overview of the most important moves within the space just defined.

As already mentioned, chance would have it that Silverman published his 1992 article just a year before the advent of a theory which equated static regularities which hold within a language with those which are maintained by alternations. Optimality Theory, which assumes that all regularities in a language are a consequence of a ranked set of violable constraints on output forms, seemed to have the necessary mechanisms to wipe away Silverman’s Operative Level and relegate all its effects to RL grammar. This reanalysis was not long in coming. Yip’s (1993) analysis of Cantonese dispenses with Silverman’s Operative level and shows that all the changes which English words undergo when they enter Cantonese are a consequence of constraints which already apply to native words. Yip states that “[t]his account of loanword phonology is in fact minimalist: there is no such thing as loanword phonology, simply the phonology of the host language” (p. 287). Unlike in the rule-based world, the fact that some of these constraints are not seen ‘in action’, causing alternations in native words, does not take away the fact that they are active in phonology.

Yip already signaled some problems, which were actually inherited from Silverman. For instance, not all segments which should pass the Perceptual level survive in the output. Liquids sometimes get deleted (freezer → [fisa], place → [peysi]), whereas [s] always remains preserved (bus → [pasi]). This issue had already been addressed by Silverman and, as we have seen in §2.2.4., was the reason for restricting his Perceptual Uniformity Hypothesis. These facts led Yip to introduce a perception-based version of the Faithfulness constraint PARSE (and its counterpart FILL) illustrated in (18).

\[(18) \text{PARSE: Parse (salient) segments.}\]

Yip makes it clear that, in the classical OT architecture, “PARSE and FILL are part of FAITHFULNESS”, whereas in her loanword model they need to be separated because, for instance, “PARSE only cares about highly salient segments, like /s/” (p. 281). This is a peculiar step which ushers perceptual effects into grammar, especially given the fact that Yip still keeps the first
scan (Silverman’s Perceptual level) in the model (though outside of loanword phonology), where this scan deals with undesirable segments.

More recent approaches do have other ways of dealing with this issue: using more fine-grained versions of Faithfulness and/or formulating the first scan differently, as the results of experimental research show that the perception of segments is influenced by phonotactics (see e.g. Hallé et al. 1998, Moreton 2002, Berent et al. 2007). However, Yip’s Faithfulness split is symptomatic of the two issues with which the OT-based approaches which separate perception from grammar have been dealing until today:

1. How much of loanword adaptation is relegated to RL perception, that is, how phonetically concrete is the input to the grammar in the case of loanwords?

2. Which kind of Faithfulness constraints deal with the input to loanword phonology?

In answering the first question, researchers were aware that the more work is assumed to be done by perception (that is, the more concrete its output), the less comparable the treatment of loanwords by the grammar will become to the interactions encountered within the language. This is due to the way rankings are inferred in OT. If anything is known about Faithfulness in a language, it is generally from two sources: the inventories, because everything that surfaces testifies to some Faithfulness constraints which were ranked high enough to protect it, and the mappings of underlying forms to surface forms. The inventory information is not very enlightening for loanword research: it has been taken for granted that SL structures which already exist in RL will be preserved unchanged. Alternations, on the other hand, do help make predictions about loanword treatment. However, in languages such as Cantonese, where there are virtually no alternations and no mutual rankings of the Faithfulness constraints can be read off the native data, no conflict can be found between the native ranking and the ones necessitated for loanwords. This allowed Jacobs & Gussenhoven (2000:90) to make a unique step in the OT-based literature, assuming a “universally defined, fully specified phonological representation” as the input. This representation is the automatic output of the “universal segment parser”, which in combination with RL grammar is argued sufficient for replacing both Silverman’s Perceptual and Operative level. While this technically ‘solves the problem’ and reduces loanword phonology to RL phonology, RL grammar now contains many rankings which have to be established in a way native data do not and cannot give evidence for. This result begs the question of how these rankings are learned and how they spread through the community. This question has unfortunately not joined the standard corpus of questions addressed in OT-based loanword phonology (but cf. Davidson et
Generative approaches to borrowing

al. 2004, for a general discussion of hidden rankings, which emerge in nonce word and loanword treatment). The reason for this lies in the fact that this is clearly a diachronic question, whereas OT-based accounts of loanword treatment are crucially synchronic. As we will argue in Chapter 6, ignoring the reality of the existence of a bilingual community (contact speakers) and its shared knowledge leads these accounts to presenting the grammar of RL in contact with SL as RL grammar.

As the corpus of loanword data discussed in the literature grew, learnability did not long remain the only difficulty which models trying to reduce loanword phonology to RL phonology had to deal with. More empirical problems arose with the introduction of loanword data from languages which, unlike Cantonese, do show alternations in their native phonology: it turned out that Faithfulness for foreign inputs can have different effects from the general Faithfulness. Peperkamp (2005) and Peperkamp et al. (2008) quote many examples of languages where foreign inputs are repaired differently from native ones. For instance, in Lama, a Niger–Congo language spoken in Togo and Benin, coda-[n] is repaired by fronting in native words (/tin/ → [tin] ‘elephants’), but by epenthesis in French loanwords ([fiŋ] → [fiŋa]). The fact that the repair applied in loanwords seems to always preserve more elements of the input than the native repair led some of OT-based researchers to assume a special family of Faithfulness, perhaps with a universal ranking MAX >> DEP (see e.g. Gouskova 2001 for a discussion). Other authors proposed similar constraints. For instance, Yip (2006) uses MIMIC constraints. Other authors emphasise that the Faithfulness constraints in question refer to surface forms and call them Output-output constraints (Kenstowicz & Suchato 2006) and surface-to-surface constraints (distinct from OO-FAITH). Kang (2003) proposes the family BeSimilar, which “promote perceptual similarity between the input and the output strings of sounds” (p. 252). This inflation of Faithfulness constraints and the authors’ hesitance to use the constraints proposed by other authors are indicative of the insecurity about the exact nature of the representations these constraints compare, especially the input form. As we shall argue in Chapters 4 and 6, this hesitance actually reflects the intuition that the terms in which SL and RL structures correspond to each other are somewhat different in each contact situation.

The introduction of loanword-specific faithfulness constraints opened a new learnability problem (Broselow 2003): if they are different from the regular Faith and used in loanword adaptation, how can they be created and acquired? There was no hope that they might emerge by grace of some kind of universal bias for treating loanwords since cases exist in which two dialects of the same language systematically treat same structures in a
different way. As shown by Yip (2002), in the adaptation of English names Taiwan Mandarin has deletion where Mainland Mandarin has epenthesis.

(19) Mainland Taiwan
Friedman fu.lı.ta.man fu.lı.man
Burt pwo.thə pi
Spielberg s.phi.r.pwo.kə s.phi.pwo
Titanic thaj.ta.ni.kho the.ta.ni

Yip observes that the two dialects do not differ in any way which might influence the speaker’s perception of consonants. Interestingly, this leads Yip (2006) to the conclusion that “the loss must be phonological” (p. 955). Later on she concludes that the ranking of the MIMIC constraints is learned and used only when one “tackles loanwords” and “it is added on to the otherwise unchanged L1 grammar, which dictates such things as syllable structure and inventory of sounds” (p. 956). A similar assumption is made by Smith (2009, §2.4.3.). A remarkable property of these approaches is that they allow for an essential component of the grammar – Faithfulness – to be available to late acquisition and used only for relatively simple mappings, while it could in principle describe very intricate interactions. Furthermore, the initial observation that Faithfulness for loanwords always provides better protection to its input than the regular Faithfulness is then an emergent property of languages, whereas any other option should be learnable. In other words, loanword-specific Faithfulness predicts a much more extensive typology than the one that we actually encounter. The fact that all the cases we know of fall within a small (and relatively simple) slice of this predicted typology seems to have something to do with the nature of loanwords, but it remains unclear how this is to be formalised. Note also that, whatever interpretation of Faithfulness is selected, this still does not answer the question to which form this Faithfulness refers, i.e. what the actual input is, what its phonological properties are and how they are preserved. All these issues are under debate at the moment.

The fact that the use of default/production grammars has many unknowns, as well as the growing corpus of evidence for perceptual effects in loanword adaptations, led many researchers to abandon the production grammar as the only part of the process which can be modelled. This group of authors recognises perception as a domain in which OT formalism can be applied (Kenstowicz 2003, 2005, 2007, §2.4.1.) or, even more into the perception-based world, as the domain in which the actual adaptations occur (Broselow 2003, 2009, Boersma & Hamann 2009, §2.4.2.). However, these authors do not give up the assumption that RL grammar has a crucial role in
loanword adaptation. Rather, following a broader shift in the theory of grammar (ushered by Boersma 1998), they allocate perception to grammar.

Before turning to specific grammar-based approaches, we should stress the salubrious role of this line of research in the present loanword adaptation discourse. Being a principled approach with an established apparatus for testing hypotheses, this line of research was relatively quick in concluding that loanword adaptation cannot be based on exclusively phonological or exclusively phonetic categories (see e.g. Yip 1993, 2002, Kenstowicz 2003, 2005). It was from this line of thought that the crucial arguments against the ever-stronger perception-based model arrived. Shinohara (2000, 2003) showed, for instance, that in the accentuation of French borrowings into Japanese (whose original stress is overridden), epenthetic vowels are treated differently from the comparable ones present in the input. This would be problematic under a perception-based account, which would assume that epenthesis in loanwords is a pre-phonological hallucination. Uffmann (2006) shows using data from various languages that the quality of the epenthetic vowels used in loanwords is determined in phonology, by a complex interaction between vowel harmony, local assimilation to the preceding consonant and default insertion. Finally, Smith (2006) provides examples of earlier, 19th century Japanese borrowings which display deletion instead of the (currently pervasive) epenthesis as the repair strategy (see examples in (20)). This shows that the current epenthetic bias in Japanese contact with English is also a consequence of the way speakers now approach foreign inputs, based on the established borrowings in the standard language, and not a consequence of their inability to ignore consonantal cues and suppress perceptual epenthesis.

(20)  Earlier contact  Current contact  
ne.ba:.ma.in  ne.ba:.ma.in.do  ‘never mind’  
pu.riN  pu.din.gu  ‘pudding’  
o:.rai  o:. ru.rai.to  ‘all right’  

In the following three sections, three exponents will be briefly reviewed of the approaches emerging from the described research climate in the OT-based loanword adaptation. §2.4.2. will address one perception-based OT approach – Boersma & Hamann (2009). §2.4.3. discusses a correspondence-based one – Smith (2009). Finally, in §2.4.4., the model proposed by Dong (2012) is discussed, which combines the explicitness Boersma & Hamann (2009) and the awareness of different language contact situations of Smith (2009).
2.4.2. Perception (is) grammar: Functional OT-based model

Before turning to the model based on Functional OT (Boersma 1998), a brief overview will be given of the approaches which introduced OT perception grammars into loanword research. Some of these have already been touched upon in the previous section.

As we have seen above, due to the Silvermanian heritage, from the onset of OT-based research the influence of perception on loanword adaptation was recognised and taken seriously. However, this also means that Silverman’s serialism was inherited, as it was instantiated by the existence of two levels, neither of which, as we have seen, can be easily described as non-phonological. This serialism was reincarnated in the notion that the output of perception is the input to adaptation. This is not surprising since most OT models actually deal with production. However, there has been awareness that differences in perception are rooted in the properties of the native grammar and that, consequently, even if the perception-based models won the input dispute (that is, if there were a consensus that the input is a percept), loanword adaptations would still have to be explained through the interaction of native phonology and perception.

While the foundational work of Functional OT already contains a few references to loanwords, the earliest elaborate proposals including an OT perception grammar in loanword accounts were made by Kenstowicz (2003) and Broselow (2003). Kenstowicz essentially uses a reranking of production constraints to model the case where in Fon, a Niger–Congo language of Benin, the French word *poste* adapts as *posu*. Kenstowicz then resurrects serialism in assuming that the perception ranking gives the mapping [post] → [pos], whose result gets lexicalised; and the production ranking then leads to the mapping /pos/ → [posu]. Interestingly, the crucial ranking for this perceptual mapping is DEP-V >> *stop/obstruent_# >> MAX-C. Again, this resembles Silverman’s serialist approach in that perception appears more like a phonological level, rather than the decoding of an auditory input. It is hard to imagine how these constraints evaluate a mapping: the speaker has to hear everything accurately at some level, in order for the grammar to be able to decide what it wants the speaker to hear.

Broselow (2003, 2009) uses a specific version of perception grammar in order to obviate the need for loan-specific Faithfulness constraints (as proposed by Davison & Noyer 1997), which pose a learnability problem because, as we have seen above, their ranking cannot be read off native data by the learner. Just as in the case of Kenstowicz (2003), a complete picture of the assumed architecture of the grammar is lacking. However, the examples are insightful and the processes described by Broselow seem to be common in loanword adaptation which proceeds without developed bilingualism and
Generative approaches to borrowing

literacy. So, for example, in order to account for Spanish to Huave mappings of the type [ígado] > [ik], Broselow assumes a perception grammar constraint ASSUMEWORDEDGE-V C# (see Davison & Noyer 1997 for a Faith-based account, criticised by Broselow).

(21) ASSUMEWORDEDGE-V C#: In mapping the acoustic signal to phonological representations, assume a word edge following each consonant preceded by a stressed vowel.

This constraint, which never predicts incorrect segmentation in Huave, gives the underlying form /ig/, which can then be turned into [ik] by the production grammar. Broselow’s implementation of the perception grammar has been discussed by Boersma & Hamann (2009) from the perspective of Functional OT. The perception grammar proposed by Broselow is two-leveled, in that it maps the input (which Boersma & Hamann assume to be phonetic) directly onto an underlying form. This means the assumption of syllables, feet, etc. in the candidate underlying forms, all of which are generally assumed to be redundant in the lexicon (McCarthy 2003).

Functional OT (Boersma 1998) affords the possibility of concentrating all the relevant processes of the loanword adaptation in the perceptual module (which, in Boersma & Hamann’s view, is desirable) without giving up the virtues of OT grammars, since it considers perception to be just as grammatical as production.

The analyses in Boersma & Hamann (2009) implement the entire range of the Functional OT formalisms (developed for L1 grammar) and therefore have a more formalised concept of perception than those used by Kenstowicz or Broselow. As one can read in (22), perception is limited to a mapping between the phonetic form and the surface form, and the constraints which play a role are cue constraints and structural constraints.

(22)
Once the (RL!) surface form is found, the job of the phonology is finished and
the implementation of a novel underlying form typically involves a faithful
mapping. Again, crucially, the input is phonetic. The upside of this account is
that once the full phonetic detail of foreign outputs is taken into account,
domestic cue constraints and structural constraints can account for the
relevant modifications loanwords undergo. Cue constraints are those which
evaluate the mapping from the auditory-phonetic form to the phonological
surface form (e.g. [h]/-asp/, which protests when an aspiration cue is
ignored). Structural constraints, on the other hand, are output-oriented and
militate against marked structures (like regular markedness constraints). It
is by virtue of the latter that the native markedness ranking can influence
perception.

For instance, the fact that English coda obstruents (which are often
released, unlike their Korean counterparts) trigger epenthesis (again, unlike
their Korean counterparts) in English borrowings into Korean is accounted
using the crucial ranking *[burst]/C(.)/> *[ ]/h/, which essentially says that
the listener is better off hallucinating a vowel than perceiving a released
consonant as a coda. There is some support in the literature for the
assumptions about both English and Korean (see Kang 2003), so it makes
sense for this ranking to be internalised by speakers who have had
experience only with Korean phonetic forms and to produce the encountered
effects on English phonetic forms. The tableau in (23) illustrates Boersma &
Hamann’s analysis for the perception/adaptation of the English word tag.

(23)

```

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/C(.)/</td>
<td>/-asp/</td>
<td>/+asp/</td>
<td>/C(.)/</td>
<td>/+tense/</td>
<td>/V-tense(V)/</td>
<td>/i/</td>
</tr>
<tr>
<td>/tʰækʰ/</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/tʰækʰ/</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/tʰæ.kʰ/</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/tʰæ.kʰ/</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/tʰæ.kʰ/</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```

Although their programmatic statements may lead one to assuming
otherwise, Boersma & Hamann do formalise the loanword trajectory beyond
the creation of the phonological surface form. In their words, “this foreign-
language perception has to be followed by a process of lexical storage, which
can then lead to the adapter’s own productions of the borrowed word” (p. 33).
In this sense, given the approach defended in this dissertation, it is to
Boersma & Hamann’s credit that they shed light on the creation of an
underlying representation, recognising the fact that loanword integration can be discussed in terms of lexicalisation. Moreover, Boersma & Hamann show how grammar (and lexicon) can play a decisive role in the cases where the assignment of an underlying form to the surface form is not straightforward. For instance, a final surface /t/ can correspond to underlying |s|, |t|, |tʰ|, |tʃ| and |t’|. However, there seems to be a single solution selected, illustrated by the word shot, which is “perceived” as /sjat/ and gets the underlying form |sjas|, as shown the accusative form /sjasil/. In doing so, these loanwords follow a frequent class of Korean nouns, which has this distribution of stem-final consonants, e.g. ot~osil ‘clothes’. Boersma & Hamann account for this fact using the ranking of faithfulness and anti-faithfulness constraints *|-stri|/ /-stri(.)/ >> *|+stri|/ /-stri(.)/ in the production grammar, a ranking based on a speaker’s experience with existing words. The fact that these constraints are arbitrary is entirely in line with the principles of Functional OT (see Boersma 2008) because “our set of faithfulness constraints has to express arbitrary relations between underlying and surface form, just as cue constraints express arbitrary relations between surface form and sound” (p. 40).

One fact related to “lexical storage” seems to provide a more serious challenge for Boersma & Hamann’s account. As shown by Kang (2003), epenthesis is more common after coronals than after labials and dorsals, counter to what we would predict from the frequency of their released realisations in English. Kang takes this to be a consequence of paradigm uniformity, which saves the loanwords from having diverse allomorphs (so the English word cut becomes [kʰʌtʰi]~[kʰʌti]kʰa] rather than [kʰi]~[kʰas]). Boersma & Hamann say: “Kang’s proposal is plausible, but we do not attempt to give a formalization of this paradigm uniformity effect here” (p. 40). It seems that if they had done so, this would have led to serious modifications of their proposal because it would have to include an undesirable look-ahead effect: a suboptimal candidate from the perception tableaux is favoured by a process which concerns paradigm formation. Note that deviations from paradigm uniformity are not ungrammatical in Korean (recall [ot]~[osil] ‘clothes’) and in the presented model each loanword receives a lexical representation as the result of the adaptation process. It is therefore unclear how a later step in the life of the word (paradigm formation) influences the outcome of perception. Whatever is done to accommodate for these facts, it seems as if the central assumption that loanword adaptation is about L1 perception holds only if languages have no paradigm-related preferences.

Unlike the perception-only accounts, Boersma & Hamann seem to be more open to the fact that bilinguals can play a role in loanword adaptation.
For them, the experience of bilingual speakers can actually help restrict the hypothesis space in finding the good surface form. For instance, they argue that “bilinguals may analyse English as having only lax plosives such as /p/ and aspirated plosives such as /pʰ/, and therefore as lacking /p'/. For the labial plosive in spy they would then have two options […]” (p. 51). Although it is recognised as occurring, the analysis by bilinguals seems to be yet another aspect of loanword adaptation which affects perception in a way that seems hard to integrate in the model presented here while keeping the main tenet: loanword phonology = RL perception.

As for the epenthesis bias, these authors claim that it “has a direct explanation in terms of high-ranked cue constraints for positive auditory cues (and concomitant low-ranked cue constraints against inserting “illusory” phonological material)” (p. 45). If this is generally true, one may wonder what could be the difference between the two Mandarin dialects illustrated in (19) or between the two generations of Japanese loanwords quoted in the previous section in terms of this model.

Finally, there is no mention of non-adaptations in Boersma & Hamann (2009). Here, unlike the case of input-based approaches, the fact that the gradual assimilation of non-adaptations follows the same trajectory as instant adaptations is an argument in support of the model, since the same structural constraints can play a role in both cases. This, however, would be yet another aspect of the theory which defies its essentialist labelling as an RL perception-based account.

We are now turning to a model which goes the other way in dealing with the specialness of loanwords, making crucial use of OT's notion of Correspondence in order to describe the interactions between the perceived SL form and the word entering the language.

### 2.4.3. A surface-to-surface correspondence-based model: Smith (2009)

The model presented in this section capitalises on the assumption that loanword phonology has to have an input and an output form, which has marked loanword phonology since Haugen (1950). In the version of the model recently presented by Smith (2009), it is proposed to put an end to the disputes concerning the input form and the role of perception by introducing into the model “the borrower’s posited representation of the source-language form” (p.155), which Smith terms the pLS representation. This form can be influenced by perception, orthography, explicit knowledge of the SL grammar, etc. Since this form can have different origins, Smith assumes that this form (written between bars “||”) is neither phonological nor phonetic and that it is similar to the form of the morphological base (Benua 1997). The assumption of one representation on which all the influences converge has to
be recognised as attractive, although it may have some drawbacks. First, as shown in the previous section, a less optimal form can be favoured by the process of lexicalisation and become the winner, so keeping more potential forms 'in the pool' of possible lexicalisations may turn out to be closer to real-life adaptation processes. Second, selecting a single perceived form may be seen as solving a problem which does not really arise in OT, since, as already mentioned, OT grammars are perfectly able to map variable input onto the same output form. To be entirely fair, while no evaluations with multiple inputs are presented, nothing in Smith’s theory militates explicitly against multiple forms which could comprise a “unified repository for the Lb speaker’s knowledge of the Ls form” (p.156). The concept of knowledge is crucial here: while it is true that in different contacts different features of what bilinguals know are taken into account and ignored in loanword treatment, it is not safe to assume that bilinguals are unaware of any properties of the SL forms just because there features do not play any role in loanword integration. As we will argue in Chapter 6, there is a process of convergence on what is taken in consideration in the creation of the new RL form. The convention which is the result of this process is therefore not informative about what bilinguals know about SL.

The constraints referring to pLs form, connecting the perceived source form to the borrowing – SBcorr constraints (from Source language to Borrowing language) – “are added to a speaker’s grammar only if the speaker establishes an SB correspondence relation, preparing to establish an Ls form” (p.172). Smith further agrees that “adaptation strategies are often variable for early loanwords, later becoming conventionalised (Haugen 1950)”, but she claims that “whether a speaker chooses to preserve certain Ls characteristics or learns particular source-similarity conventions, the relevant SBcorr constraints will be ranked above any competing constraints” (p. 173).

A number of objections may be raised against this separation of SBcorr from the general FAITH(I-O). A general problem for any approach assuming that Faithfulness to foreign representations is entirely independent of Faithfulness to the domestic underlying representations is that such an approach cannot account for the fact that foreign inputs are faithfully mapped in all the aspects in which the domestic underlying representation are, plus some additional ones. If SBcorr is left freely rerankable, the factorial typology predicts, for instance, the existence of languages with deletion of the elements of the foreign inputs which are preserved in the native inputs, an unattested situation.

Moreover, since languages ‘support’ their native segments in representations from other languages, the whole family of IDENT
constraints would have to be duplicated in the SBcorr hierarchy, but this would not be predicted by the theory. Some other cross-linguistic tendencies are also left unaccounted for. Foreign structures which can be produced recombining existing native structures get introduced more easily than the ones which require introduction of novel features. For instance, there is evidence that non-local phonotactic restrictions are more easily violated in loanwords (Ussishkin & Wedel 2003), presumably because the speakers already produce the structures in question in word combinations. Also introduction of novel distinctive features (and gestural molecules, see Ussishkin & Wedel 2003) is more problematic than reusing the existing ones in new combinations (see Rose 1999 for a discussion). This all points to the conclusion that the version of Faithfulness responsible for loanwords has to, at least initially, recapitulate the ranking of the more general Faithfulness, an issue not directly addressed by Smith.

Smith sets out to show that general tendencies in the ranking of the Faithfulness constraints responsible for loanword adaptation are not absolute. In order to show that BS Faithfulness can display a ranking different from the usually assumed MAX >> DEP, she quotes examples in which stridents, which are universally salient, are deleted from the loanword input. In what follows we will try to show that these cases do not necessarily have to be seen as cases of phonological deletion.

Smith discusses three languages, out of which Finnish most extensively. In old borrowings from Swedish and Slavic, Finnish deletes stridents in initial sC sequences. The same process has been attested in the adaptation of English words in United States Finnish (as described in Kolehmainen 1937).

(24) Finnish US Finnish
    puola < spole (Swedish) ‘spool’ toori ‘store’
    tuoli < stol (Swedish) ‘chair’ touvi ‘stove’

A crucial fact is that this deletion occurs in the initial position only. Smith notes that, given the fact that Finnish has initial main stress, epenthesis would lead to a stressed epenthetic vowel. This enables her to assume that candidates with epenthesis are excluded by the constraint HEAD-DEP, which requires the prosodic head to have a correspondent in the input, i.e. not to be an epenthetic vowel. The relevant tableau is in (25).
Generative approaches to borrowing

Note, however, that forms such as [ˈsitoori] or [ˈistoori] are also harder to ‘hallucinate’ given the English [straː]. Of course, this is still insufficient evidence to assume that $s$-deletion happened in perception. However, there are more relevant facts which point at an extra-phonological solution. Fenzvesi & Zsigri (2006) report on massive deletion of initial unstressed syllables in US Finnish. 18% of the words in their corpus display this type of deletion, the vast majority of which had a light unstressed onsetless initial syllable (68% of such words lost their initial syllables).

(26) US Finnish

| komoteitata  | ‘accommodate’ |
| kriimentti   | ‘agreement’   |
| meersensi    | ‘emergency’   |
| praispaarti  | ‘surprise party’ |

This seems to indicate that, in cases like toori, the initial $s$ may have been perceived, but not as part of the word. The fact that there are cases in which $s_C$ sequences are preserved in US Finnish – Smith quotes only skeptikko ‘skeptic’ – cannot be used as evidence for the speakers’ ability to perceive word boundaries in every case and certainly not as evidence for their ability to perceive word boundaries of English words which were the originals of long-integrated borrowings. This issue of different kinds of borrowers ushers another problem for the assumption of pLs representations as special entities visible to the grammar – it cannot be decided in advance where their influence ends, i.e. when a word is sufficiently lexicalised so that regular FAITH can cater to it. Finally, we do not know whether a pLs representation can be ‘updated’, e.g. because the RL community is increasingly bilingual and whether it can then influence the adaptation process. There are indications that this is not the case – Broselow (2003) shows historical layers in the
lexicon of Huave, as spoken by the contemporary community, which is bilingual. These layers (not to be confused with lexical strata discussed in Chapter 3) show effects of adaptation procedures which (for Broselow) reflect different perception grammars of the Huave speakers at the time of adaptation. Contemporary speakers could update forms like *ik* (< *ígado*) or *garabát* (< *garabáto*), since more faithful forms would be permitted by the Huave grammar now. This does not happen – the pLs form is only relevant until the new word gets lexicalised. Moreover, since we know that loanwords enjoy extra protection long after they have been adopted, we may need three types of competing Faithfulness – BS-corrrespondence, FAITH(Loan) (see Chapter 3) and regular FAITH. This rich set of Faithfulness constraints, whose ranking has to be acquired separately, seems a very daunting task for the learner and predicts the existence of very complicated grammars, which are unattested.

Turning back to the deletion patterns, which are Smith’s crucial support for establishing the new family of constraints, we have one more language which restricts *s*-deletion to the initial preconsonantal position – Sranan. We give some examples with English-origin words in (27). Smith does note that Sranan is a creole, but argues that the new constraints can also cover the relations with the perceived words of the lexifier language in creole formation.

(27) Sranan
    piki  ‘speak’
    tranga  ‘strong’
    tori  ‘story’

If these facts are to be considered as a case of loanword adaptation, the same ‘perceptual’ account can apply as for Finnish. The third case advanced by Smith is more daunting – White Hmong, a strictly CV language, is the most oft-quoted case of a language with pervasive deletion in loanword adaptation. In this language, *s*-deletion is restricted to syllable codas.

(28) White Hmong
    khį.mă  ‘Christmas’
    mbă  ‘bus’
    pṳ.lĩ  ‘police’

Deletion also occurs word-medially and this seems a very good case for the ranking DEP>>MAX in the family of FAITH responsible for adaptation. However, there are facts which might again point to some explanation à la
Broselow (2003). The authors who originally reported on these facts emphasise that “all Hmong roots and words (except compounds) are monosyllabic” (Golston & Yang 2001). So it seems that this case of coda deletion could be due to the constraints on word size or lexical storage, such that it does not have to be performed by phonology, or at least it need not be due to a Faithfulness ranking.

In sum, no persuasive evidence has been found that deletion can become the pervasive procedure in loanword adaptation. More generally, the separation of the family of FAITH responsible for loanword adaptation seems an insufficiently supported step. As will be argued in Chapter 3, the necessary distinctions can be made using the type of Faithfulness used for lexical stratification, which seems to be a universally available. Smith seems to avoid this option.

If an etymological loanword was nativized during adaptation, it has become an Lb lexeme that conforms to Lb phonotactics. Subsequent Lb learners of this lexeme (not from Ls speakers) have no motivation to assume an Ls-like form and recapitulate the adaptation process. Alternatively, if an etymological loanword was not fully nativized during adaptation, then it has become an Lb lexeme with “foreign” characteristics. Such lexemes may cause Lb learners to develop a stratified lexicon (Fukazawa, Kitahara, and Ota, 1998; Ito and Mester, 1999; Ota, 2004; Pater, 2005) or to reanalyze the core Lb grammar (Rice, 2006), but these consequences are distinct from the original process of adapting loanwords. (p. 176)

We can conclude that the correspondence-based approach to loanword adaptation initiates an interesting line of research elaborating the idea of special representations for loanwords. However, the strong form of this assumption predicts a very vast variety of grammars, for which little persuasive evidence seems to exist.

In the following section, a model will be addressed which takes the insights from both Functional OT-based and Correspondence-based approaches seriously, in producing a comprehensive account, which also clarifies the limits of a loanword theory.

2.4.4. Getting real: Dong (2012)
Although empirically restricted to borrowing from English into Mandarin Chinese, Xiaoli Dong’s 2012 dissertation provides an important innovation in the discourse of loanword research: it is the first extensive study of loanword integration which does not base its findings on a single type of borrowing
(based on a single type of input, performed by only one type of adaptors), assuming that all the other types of borrowing are deviant and less informative for the theory of loanword treatment. Already in her online adaptation study, Dong carefully distinguishes between seven borrowing scenarios. She considers as factors: (a) input type (auditory, orthographic and the combination of both) and (b) adapter type (monolinguals, semi-advanced learners of English and proficient bilinguals). Each type of input is matched with each adapter type, but proficient bilinguals were shown to perform the same type of adaptation since they know both relevant representations. One of the generalisations, which matches the insights from sociolinguistic literature, is that, for each input type, adaptations by monolingual speakers were more variable than those performed by semi-advanced learners. This already sets the stage for the long-awaited deconstruction of what we have been calling essentialism: foregrounding one type of adaptation scenario and drawing universal conclusions from it. In Dong’s words, “inspecting some previous studies, it can be noticed that they only accounted, or in fact only tried to account, for sub-domains of the complete domain of loanword adaptation” (p. 250). Dong goes further, showing that some other usual assumptions on loanword treatment suffer from unjustified generalisations. First, her study shows that online adaptations can have different outcomes from what is encountered in the “large corpus of loanwords borrowed from English into MC during a variety of stages of the 20th century” (p. 2), so that none of the scenarios can be safely equalised with the pattern displayed in the corpus. Second, Dong shows that “perception is not a unified phenomenon: the perception of English sounds of early semi-advanced MC learners of English was less accurate in the adaptation experiment than in the perception experiment, even though the two experiments had the same design” (p. 251). This all is not to say that Dong introduces more relativism into loanword research than there was before. Rather, her study shows that in each scenario, there is a dominant grammar which guides adaptation, but “depending on input and/or adapter types, loanword adaptation is driven by one or more of the following grammars: the native RL phonological production grammar, the native RL perception grammar, the RL/SL interlanguage phonological production grammar and the RL/SL interlanguage perception grammar” (p. 2-3).

In order to disentangle the different driving forces in adaptation, Dong performs a shakeup of essentialist and generalising discourses on yet another level by showing the impact of a number of apparently contingent, yet relentlessly real aspects of Mandarin Chinese adaptation processes. These are the specific aspects of the MC linguistic community which determine the outcomes of loanword adaptation in ways which should not be
ignored. Newly integrated words need to receive an orthographic representation, which will comprise characters, which encode syllables. Characters already have a meaning and sometimes that meaning can be incompatible with the meaning of the loanword to such an extent that the character cannot be used. Moreover, not all phonotactically possible syllables have an existing character and no new characters are introduced for loanwords. Finally, MC has its own romanisation (Pinyin) which the speakers are aware of and which enables even monolingual speakers to make a guess about the pronunciation of foreign words. These generalisations about the RL community have been formalised by Dong in that “before analyzing adaptation patterns in a phonology/phonetic fashion, two approaches have to be taken” (p. 247). The first is the application of “chaos filters, filtering out the output resulting from extra-linguistic factors such as misinterpreted orthography and semantic association” (p. 3). The second is the addition of an undominated constraint against introducing syllables which do not already exist and have an orthographic representation. For instance, on phonological/phonetic grounds one can claim that “[f]or English [s], MC [s] is optimal but [s] and [c] are suboptimal” (p. 247). However, this predicts that the optimal output for the English [si] would be Mandarin Chinese [si], which is an in principle possible, but unattested sequence, and therefore never can become the output – [ci] is selected instead. Dong formalises this effect using the constraint USE-LISTED-SYLLABLE, which effectively promotes a suboptimal candidate which is orthographically representable. Although this may seem like blurring a phonological evaluation with non-phonological factors, it is not only realistic – nothing will become an MC word if it is unrepresentable in characters but this approach also reveals the criterion for the selection of phonologically-suboptimal-yet-optimal candidates.

The final feature of Dong’s approach seems most innovative of all: she asks what it takes for an output of phonological (inter-language) grammar to be able to become an RL word. As we shall show in Chapter 4, posing this question in languages with more morphological complexity than MC will lead to an important reorientation of the agenda of loanword research.

2.5. Conclusion of the chapter and further development

In this chapter we have reviewed various approaches to borrowing. As has been shown, probably the most prominent property of the modern research climate and the main reason for very little cross-fertilisation among different approaches lies in the fact that the researchers have very different
objectives. While the input-based approaches are involved in a dispute on the most common input to loanword phonology, OT-based accounts try to fit loanword adaptation into the existing theory of grammar using as few loan-specific mechanisms as possible.

As shown in the discussion of various approaches, quite a few phenomena remain either uncharted or untheorised for most approaches. The knowledge internalised by the bilinguals in ongoing contact is only occasionally taken into account and is often considered to be a confounding factor. The stratification of synchronic lexica, evidently a consequence of loanword (non)adaptation, remains out of the picture in all approaches (pace potentially correspondence-based accounts). Furthermore, due to the type of data employed (before/after pairs) the dynamics of assimilation of initially unadapted forms (gradual nativisation as registered in Holden 1976) are also unaddressed.

Finally, it should be noted that this overview has covered the history of loanword research taking into account the data which have been discussed – segmental adaptations presented as before/after pairs. Since morphological and suprasegmental facts have not received systematic attention, references to work on these phenomena will be discussed in the chapters in which the new model is proposed. In the following section, we turn to the research into synchronic lexica, which, as shall be clear, suffers from a lack of coherence with the borrowing theories.
Chapter 3
Generative research into borrowedness: Synchronic stratified lexica

This chapter addresses the main achievements of research in the special status of (already lexicalised) loanwords in synchronic lexica, termed borrowedness research in this dissertation. As will become clear by the end of the chapter, the models of lexical stratification, which describe the final stage of loanword adaptation, provide crucial insights which ought to be taken into account in building the theory of loanword phonology defended here.

3.1. Core-periphery relations

Relatively independently from research presented in the previous chapter, the line of research concerned with lexical stratification produces synchronic descriptions of lexicons, as well as grammars which are able to make reference to lexical strata. The core approach was presented by Ito & Mester (1994, 1999, 2001, 2002), by whom a number of languages were discussed, lending strong support to a stratified model of the lexicon. In such a model only one stratum, the core, acts as 'completely native', characterised by the strongest restrictions on the surface structures (in OT terms: all the markedness constraints active in the language are active in this core stratum), while other strata, in which subsets of the available structures are observed, constitute the periphery. Note that this is not just an instantiation of the long-established observation that language is “a system of principles with parameters fixed, along with a periphery of marked exceptions” (Chomsky 1986:150), but it also shows that the periphery displays systematicity in its being marked and borrowed.
Chapter 3

Approaches couched in OT afford the possibility to formalise the core-periphery relations by keeping the ranking of Markedness constant and separating Faithfulness for each stratum. Similar generalisations and formalisations have been made for various grammatical classes: nouns seem to be universally most liberal in terms of the permitted structures (Smith 2001), and stems allow more marked structures than affixes (Beckman 1998).

The following section contains a brief overview of the pre-OT predecessors of the stratification models, which often recognised the core-periphery relations but did not produce a model which would formalise the subset relations of the structures permitted in the strata.

3.2. A brief history of formalised borrowedness

Borrowedness research already has relevant predecessors amongst structuralists. Pike & Fries (1949) described a number of phonological features in Mazateco which show that certain structures, most prominently sequences nasal + voiceless obstruent, occur only in Spanish loanwords, for instance *siento* ‘one hundred’. Based on these data Pike & Fries argue for an extension of the common assumption that every language can be described in terms of a single system since “[t]he speech of monolingual natives of some languages is comprised of more than one phonemic system; the simultaneously existing systems operate partly in harmony and partly in conflict” (p.29). Pike & Fries occasionally tag the non-native system as “non-uniform”, “fragmentary material” and “unassimilated sounds”, thereby foreshadowing the observation that the non-native strata do not obey some specific grammar of their own, but rather contain elements of another system added to the periphery of the lexicon. This matches the conclusions of the literature on code-switching, which shows that SL grammar is ‘suspended’ in code-switched items.

Despite the basic insight of these early observations, an acknowledgement of the core-periphery distinction was very long in coming: research in rule-based generative grammar allowed for complicated relations between highly abstract underlying forms and varieties of intermediate and surface forms. The grammars of particular languages could therefore (and often did) recapitulate not only their own history, but also the histories of the languages they borrowed from. We can illustrate this with an example from a moderately abstract ‘historian’ Saciuk (1969).
Thus, for example, a synchronic description of Spanish has to account for the forms:

a. lluvia ‘rain’
b. pluvial ‘rainy’
c. chubasco ‘heavy shower’

all three derived from the underlying form /plub/. Lluvia is the [+Native] form, since it undergoes the rule Cl > ll, characteristic of the [+Native] component; chubasco is borrowed from Portuguese and so the initial cluster undergoes a rule found in this language; pluvial is a “learned” counterpart, i.e. it is just [-Native], because it does not undergo the Cl > ll rule. (p. 484)

In order to be fair to Saciuk and other authors from this tradition (of which Saciuk also gives an overview), it should be emphasised that in addition to many purely etymologically based features this author introduces a number of features which should be seriously considered in attempting to understand the representation of lexical strata. First, Saciuk claims that the first universal divide within every lexicon is based on the feature [Homogeneous]. Rules and URs which are native or related to the native ones are [+Homogenous], whereas the remaining, more scattered rules and URs are [-Homogenous]. The homogeneous stratum roughly corresponds to what would be considered actually productive patterns in most modern approaches. Another notable feature is [Learned]. Although not very explicitly defined, this feature seems to have been introduced to distinguish a class of lexical entries which are usually acquired after the core system, probably during formal education. It seems that the recognition of this class of ‘learned’ items, which are more accessible for meta-linguistic consideration, is essential for a realistic theory of lexical strata. For instance, we need this distinction to account for the special status of the cases in which standard languages import some of the inflectional morphology of the SL (such as the Latin plurals in English cactus~cacti, alga~algae, etc.), all of which are apparent exceptions to the generalisation (on which it will be elaborated in the next chapter) that languages borrow single forms based on which they build paradigms.

3.3. Onion models of lexical stratification

A formalisation of the observation that loanwords typically allow more marked structures than native words has come only with the advent of constraint-based theories. As already mentioned, research into synchronic
lexica couched within Optimality Theory has developed a theory of lexical strata, defined by stratum-specific Faithfulness constraints (Fukazawa et al. 1998, Ito & Mester op.cit). Faithfulness constraints indexed to loanwords (henceforth FAITH(LOAN)) then have the function of protecting certain marked structures in loanwords from the effects of general markedness constraints, which have full effect in the native stratum. The crucial ranking is then FAITH(LOAN) >> Markedness >> FAITH.

This formalisation was first proposed for the Japanese lexicon (see Ito & Mester 1994) and Japanese is still the only language for which various phenomena have been discussed and an extensive inventory of stratal constraints was proposed. As for other languages, only limited sets of data were discussed for German (Ito & Mester 2001), Slovenian (Jurgec 2008, 2010), Thai (Kenstowicz & Suchato 2006), Quebec French (Paradis & Lebel 1994), Cantonese (Yip 2006), Hebrew (Becker 2003), Hungarian (Kertész 2003) and Serbo-Croatian (Zec 2002, Simonović 2009ab). Since there is extensive work on Japanese, our discussion here will largely use examples from this language.

Japanese strata are defined by various marked structures permitted in subsets of the lexicon, so that all of them occur in the most peripheral stratum and none of them in the core stratum (called the Yamato stratum). The table in (31) (from Ito & Mester 1999) shows that three out of four markedness constraints are respected in only some subset of strata. Note that the constraints on syllable structure (such as the ban on complex syllable margins and codas with a place specification) are never violated, and their satisfaction is a condition on being considered a Japanese word.

---

10 The first version of the lexical stratification model (Ito & Mester 1994), which followed the traditional descriptions of Japanese, also contained a Mimetic stratum, which was not based on language-contact induced stratification. The structures of this stratum did not constitute a proper subset of those of any other stratum. The Mimetic stratum vanished from the stratification picture in Ito & Mester (1999, 2001, 2002). This seems a reasonable step because mimetic words are a part of a different stratification of the lexicon. Moreover, these words do not nativise and do not absorb other words. However, since this exclusion was never made explicit, various authors continued criticising Ito & Mester for this inconsistency. Crawford (2007) and Inkelas & Zoll (2007) show that Mimetic and Sino-Japanese strata exhibit a “markedness reversal” allowing disjunctive sets of structures. While the mimetic stratum prohibits sequences of nasals and voiceless obstruents, but permits singleton p (so the constraint ranking is *NT>>*P), the Sino-Japanese stratum does just the opposite (*P>>*NT). As shown by Fukazawa et al. (1998), even this can be accounted for using only indexed Faith, since the two structures are licensed by different types of Faithfulness – the ranking of IDENT[lab] does not depend on the ranking of IDENT[voice] for each stratum.
Ito & Mester propose that between every consecutive pair of markedness constraints there is ranked a Faithfulness constraint which licences the structures in question in its part of the lexicon.

\[ (\leftarrow \text{FAITH}_5) \]

\( \text{SYLL-STRUC} \)

\[ (\leftarrow \text{FAITH}_4) \]

\( (= \text{FAITH/UNASSIMILATED FOREIGN}) \)

\( \text{NO-DD} \)

\[ (\leftarrow \text{FAITH}_3) \]

\( (= \text{FAITH/ASSIMILATED FOREIGN}) \)

\( \text{NO-P} \)

\[ (\leftarrow \text{FAITH}_2) \]

\( (= \text{FAITH/SINO-JAPANESE}) \)

\( \text{NO-NT} \)

\[ (\leftarrow \text{FAITH}_1) \]

\( (= \text{FAITH/YAMATO}) \)

In the evaluation of candidates by the grammar, the faithfulness constraints indexed to separate strata will evaluate only the words which belong to the particular stratum, as demonstrated in the tableau in (32).

\[ (\text{SINO-JAPANESE}: \, \text{šin-tai} \, \text{‘body’} \, (\text{violation of NO-NT}) \]

\[ (\text{YAMATO}: \, \text{šin-de} \, \text{‘die-GERUND’} \, \text{(no violation of NO-NT)} \]

One of the favourite empirical arguments for this model comes from so-called impossible nativisations (Ito & Mester 1999, 2001, 2002). The German adaptation of the English word *story* shows clearly set priorities. The original form [stɔɪ] has two structures forbidden in German native words – the
initial combination [st] and the approximant [ɾ]. Speakers will repair the prohibited structures to different extents, which leads to variation from total preservation of the SL structures [ʃɾi] to total assimilation [ʃɾi]. However, if only one structure is repaired, it is always the [ɾ], so that *[ʃɾi] is an impossible adaptation. The reason is that markedness constraints are mutually ranked as shown in (33) and the relevant tableaux are replicated in (34).

(33)

\[
\begin{array}{c}
\text{Faith-C (unassimilated)} \iff \text{[ʃɾi]} \\
\times[ɾ] \\
\text{Faith-B (partially assimilated)} \iff \text{[ʃɾi]} \\
\times[s\text{C}] \\
\text{Faith-A (native, fully assimilated)} \iff \text{[ʃɾi]} \\
\end{array}
\]

(34)

<table>
<thead>
<tr>
<th>'story'</th>
<th>Faith-C</th>
<th>*₁</th>
<th>Faith-B</th>
<th>*₂[s\text{C}]</th>
<th>Faith-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. /ʃɾi/ A</td>
<td>[ʃɾi]</td>
<td>d.n.a.</td>
<td>d.n.a.</td>
<td>*!</td>
<td>*!</td>
</tr>
<tr>
<td>(most nativized)</td>
<td>stɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>stɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>jɾɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b. /ʃɾi/ B</td>
<td>[ʃɾi]</td>
<td>d.n.a.</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
</tr>
<tr>
<td></td>
<td>stɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>stɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>jɾɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c. /ʃɾi/ C</td>
<td>[ʃɾi]</td>
<td>***!</td>
<td>d.n.a.</td>
<td>*!</td>
<td>*!</td>
</tr>
<tr>
<td>(least nativized)</td>
<td>stɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>stɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>jɾɾi</td>
<td>*!</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
</tr>
</tbody>
</table>

Ito & Mester’s nativisation tableaux are one of the rare examples of an interaction of what we term borrowedness (the static synchronic lexicon which contains loanwords) and borrowing (the process which involves more than one language). However, the situation becomes more complex once the nativisation logic is implemented in the stratification model. (35) below depicts a more complicated example from Ito & Mester (2001). Unlike the
Generative approaches to borrowedness

In the general case, in which a very limited number of strata are proposed, now we see a proliferation of indexed Faithfulness constraints, which can fit between any two markedness constraints.

(35)

<table>
<thead>
<tr>
<th>Jongleur ‘juggler’</th>
<th>[s]</th>
<th>[z]</th>
<th>[ε:]</th>
<th>[ə]</th>
<th>number of changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. joglaur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>b. *joglaur</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>c. *joglaur</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>d. *joglaur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>e. *joglaur</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>f. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>g. joglaur</td>
<td>yes</td>
<td></td>
<td>yes</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>h. *joglaur</td>
<td>yes</td>
<td></td>
<td>yes</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>i. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>j. *joglaur</td>
<td>yes</td>
<td></td>
<td>yes</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>k. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>l. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>m. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>n. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>o. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>p. *joglaur</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

In (35), just like in (34), the stratum defined by the most liberal Faithfulness reflects the SL pronunciation and every markedness ranking on the way from the outermost to the core stratum opens up a possible location for a new stratum and a new set of Faithfulness constraints. This makes the Faithfulness constraints which are involved in borrowing a rather different animal from the Faithfulness constraints we usually see in synchronic analyses. Ito & Mester do not address the issue of constraint proliferation in borrowing explicitly. There is, however, one proposed restriction: the ranking of specific types of constraints with respect to each other should not vary per stratum, which is repeated below.

(36)

Ranking Consistency

Let $F$ and $G$ be two types of faithfulness constraints (IDENT, MAX, etc.), and $A$ and $B$ types of correspondence (input-output, output-output, base-redundant, base-truncatum, etc.). If $F_A \gg G_A$ for faithfulness dimension $A$, then there is no faithfulness dimension $B$ such that $G_B \gg F_B$:

$$\exists \text{AB} (F_A \gg G_A \land G_B \gg F_B)$$

Most theoretical criticism of this approach originated from Ito & Mester’s use of the traditional labels for the Japanese strata, without clarifying the relationship between their lexical strata and the diachronic strata. This issue
arises, for instance, in words with no clear stratal cues: the foreign words such as *samba* or *medaru* ‘metal’ could just as well be Yamato. Furthermore, many of the reliable stratal cues are purely static, that is, there are no alternations which enforce them and if speakers assign the words containing them to a ‘wrong’ stratum, they will never produce wrong forms. Palatals, which almost exclusively occur in Sino-Japanese words, but are not actively repaired in other strata, are such a case. Since Ito & Mester did not make it clear how these cases are treated in the acquisition of the strata, they were often accused of presupposing that the speakers know the etymology of words (see Inkelas et al. 1997).

Various implementations of the stratification model arose. Fukazawa et al. (2002) and Ota (2004) propose a weak stratification model (see Gelbart & Kawahara 2007, for a discussion), in which only the structures which are actively repaired and give rise to alternations qualify to enforce stratum formation during acquisition. On the other hand, authors who worked on stratal perception proposed a strong stratification model, according to which words can enter strata even if they contain no stratal cues. As discussed in the following subsection, these authors argue that Japanese speakers have distributional reasons to treat the foreign word *samba* differently from, for instance, the Yamato word *tombo* ‘dragonfly’.

It has been shown by Pater (2005, 2007) that stratified OT grammars which capture static regularities by indexed Faith are learnable. Pater assumes that the learner, upon encountering a marked surface structure, does not immediately promote general Faithfulness, but assumes for the time being that that is the only word permitting such a structure by virtue of a special Faithfulness constraint. Then, as the list of morphemes indexed to a special Faith grows, the learner might make the constraint general.11

A number of studies argued that empirical facts necessitate extensions of the basic model of lexical stratification, to the effect that loanwords can end up in the core stratum. Jurgec (2008, 2010) claims that Slovenian loanword adaptation processes necessitate markedness constraints indexed to loanwords. Gelbart (2005) presents an analysis of Latvian, according to which native words have more marked phonology (and in this sense became peripheral in this language). We address these arguments in Chapter 6, after having presented the relevant apparatus of the current model.

In sum, Ito & Mester focus on strata as they are defined by the distribution of segments. A theory of stratification based purely on segmental

---

11 Here Pater actually follows the scenario introduced by Rice (1997) as a *reductio ad absurdum* of lexical stratification, according to which the learner of English has two different grammars for the words *fond* and *font*. 
distributional patterns would not be particularly persuasive for at least one reason – if the words grouped in a stratum only have certain marked structures in common and do not constitute a class visible to the grammar in any other way, then the segmental structures in question could be assumed to have a special representation, instead of labelling whole words as members of lexical strata. Let us now briefly turn to non-phonological evidence for the reality of strata, which has brought additional evidence for the strong lexical stratification model.

### 3.4. The reality of lexical strata

Experimental evidence has been adduced that speakers use their knowledge of lexical strata in perception. For example, Japanese speakers have been shown to shift their perceptual boundaries according to the stratal affiliation of the item they are listening to. When exposed to nonce items containing foreignness cues, they extend the perceptual area which they categorise as a long final *a* (allowed in foreign words, Moreton & Amano 1999). The same effect was found for existing loanwords which contained no foreignness cues (Gelbart 2005, Gelbart & Kawahara 2007). These data can be explained in terms of the specific sociolinguistic situation in Japanese, where literate speakers need to know the stratal affiliation of the words in order to be able to write them down, since different scripts are used for different strata. Because of this fact, metalinguistic labels referring to lexical strata are commonly used during formal education and they are present in common Japanese discourse.

Moreton & Amano (1999) mention another crucial fact of stratification in Japanese: strata are morpheme co-occurrence classes. Many processes in Japanese morphology have been shown to be sensitive to the stratal affiliation of morphemes (see Kurisu 2000). Stratal affiliation has also been shown to play a role in the derivational morphology of English and Dutch (Booij 2003). For instance, a huge majority of the adjectives to which the Latinate nominalising suffixes *-ity* and *-iteit* can apply are themselves Latinate: English Latinate *obscure* and Dutch Latinate *obsuur* have nominalisation *obscurity* and *obscuriteit*, but Germanic adjectives *dark* and *duister* ‘dark’ have none (*darkity*, *duisteriteit*). Moreover, as we have seen in Chapter 1, loanwords tend to be restricted to certain parts of speech. This is another fact which will facilitate the acquisition of stratal affiliations.

Given these facts, it is not surprising that it has been found that speakers are aware of stratal cues in languages where literacy and common discourse do not require any etymological competence. Jurgec (2008) found that Slovenian speakers use stratal cues (such as the segments *f* and *dž* [dʒ]) in
their judgements of non-existing words. The same result has been replicated for Serbo-Croatian complex codas (Simonović 2009b). Jurgec’s subjects were also able to single out loanwords which lacked any foreignness cues. The same study also found that Slovenian speakers are even able to group non-native items by the period in which they entered the language – the oldest borrowings were judged foreign less often than the newer ones or and far less often than the newest ones. This seems to be the consequence of phonological adaptation, but also of the integration into derivational families. Morphological orphanhood (i.e. the fact that certain words have no derivates in a language) was reported by Schwarzwald (1998) to function as a foreignness cue in Hebrew. As we will show in Chapter 11, the scenario according to which words gradually build derivational families, from which the duration of their presence in the language and their level of integration can be predicted, follows from the fact that the introduction of novel allomorphs is discouraged in the lexicon.

### 3.5. A ‘too elegant’ alternative: Underspecification

While most researchers of lexicons which appear to tolerate certain structures only in certain subsets of words use Faithfulness constraints indexed to loanwords, there has been one alternative, representational account of the core-periphery effects, proposed by Inkelas et al. (1994, 1997).

The division between accounts using diacritics (such as stratal features) and the ones using representational means for encoding exceptional patterns is an extension of two lines of research which existed in the rule-based era of generative phonology (see Pater 2005, 2010 for a general discussion). Inkelas et al. (1994, 1997) argue more generally that “static distributional subregularities, including lexical exceptions” (p. 21) should be separated from the regularities which show alternations. For the former, prespecification in the lexicon is proposed, whereas for cases which display exceptional patterns attestable by alternations, cophonologies are used – minimal rerankings of the main, master ranking, which apply to subclasses of the lexicon. One such subclass, analysed using cophonologies, is the Turkish stress pattern for place names – Sezer stress. \(^{12}\) We begin by giving an example in which this approach shows all its virtues. The analysis is based on Inkelas et al. (1997), with some details modified for the sake of clarity.

\(^{12}\) Inkelas & Zoll (2007) do mention lexical strata as one of the classes which show “subgrammatical phonological patterns”. In this paper, they argue that such patterns should be analysed using cophonologies. Though their arguments might hold of other cases, for loanwords they show no good reasons to give up on the more restrictive indexed FAITH-based approach.
In Turkish there are three classes of obstruent-final noun stems: (a) the ones with non-alternating voiceless obstruent (devlet), (b) the ones which show alternation between a voiced and voiceless obstruent, conditioned by the position in the syllable (kanat) and finally (c) the stems which have a voiced obstruent in all forms (etüd). In (37) we give examples of each group.

(37)

\[
\begin{array}{ccc}
\text{a.} & \text{devlet} & \text{devletler} & \text{devleti} \\
& \text{‘state’} & \text{‘states’} & \text{‘state-accusative’} \\
\text{b.} & \text{kanat} & \text{kanatlar} & \text{kanadı} \\
& \text{‘wing’} & \text{‘wings’} & \text{‘wing-accusative’} \\
\text{c.} & \text{etüd} & \text{etüdler} & \text{etüdi} \\
& \text{‘study’} & \text{‘studies’} & \text{‘study-accusative’} \\
\end{array}
\]

In (37a) and (37b), the language may appear like a standard final devoicing language (such as Dutch, German or Russian). However, the class of loanwords illustrated by (c) contradicts this assumption.

A stratification-based account in this case would make use of the fact that the words which follow the pattern in (c) are loanwords and assume that both kanat and etüd have a voiced obstruent in the underlying form, but the one in /etyd/ is protected by \text{FAITH(LOAN)} and therefore surfaces in all the forms. /kanadı/, on the other hand, gets devoiced whenever it ends up in the coda position. The associated ranking is given in (38).

(38) IDENT-IOVOICE(loan) $\gg$ *VOICEDCODA $\gg$ IDENT-IOVOICE

The alternative underspecification analysis offered by Inkelas et al. (1997) is to leave the alternating stems underspecified for voicing, so the words in (37) would get the representation in (39).

(39) /devlet/ /kanaT/ /etyd/

\[
\begin{array}{c|c|c}
[-\text{voice}] & [0 \text{ voice}] & [+\text{voice}] \\
\end{array}
\]

Now the specified obstruents retain their voicing in all positions, due to undominated IDENT-IO-VOICE, whereas the underspecified obstruents get their value depending on the position they surface in. The analysis requires a markedness constraint which imposes intervocalic voicing – *VČV. In order to get voiceless codas, one can either use *VOICEDCODA or a general
constraint *VOICE. The latter is used in the tableaux below, in which we show the behaviour of the alternating stem /kanaT/.

\[(40)\]

<table>
<thead>
<tr>
<th>/kanaT/</th>
<th>IDENT-IO[voice]</th>
<th>* VCV</th>
<th>*VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ≠ kanat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.   kanad</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/kanaT+uu/</th>
<th>IDENT-IO[voice]</th>
<th>* VCV</th>
<th>*VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.    kanatu</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. ≠ kanadu</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

This approach appears elegant as long as we stay within the realm of phonotactic regularities, since it does not introduce any diacritics, such as tags for different classes of loanwords to which faithfulness constraints can be indexed. It also appears to be superior in cases where features of more strata co-occur within a word. Such cases have been attested in various languages (see Inkelas et al. 1997 for Turkish examples). Ito & Mester (1999) mention stems such as Sino-Japanese *teŋka `empire`, which contains a violation of *NT, but still undergoes Rendaku-voicing, normally restricted to Yamato words, in compound forms such as onna-*deŋka `petticoat government`. Such a form poses no problem for the underspecification account, since its alternating part can be underspecified while the non-alternating part has a full representation. The representation would then be /TeNkə/ and nothing in the system would mind its unusual combination of an underspecified stem-initial obstruent and a specified postnasal one.

Ito & Mester (2001) address this underspecification alternative, emphasising that it is unable to exclude the “impossible adaptations”. Assuming that if underspecification is permitted it should be permitted in all positions, they take the remarkable step of putting various combinations of underspecified segments in “nativisation tableaux”. This then shows that an impossible nativisation of the English words *story* (!) would emerge.
Ito & Mester argue that the problem lies in the fact that [Stɔᵣi] is an input to begin with and that underspecificationists will “need to come up with a general way of removing such ‘wrong’ inputs from, i.e., for the case at hand, declaring /Stɔᵣi/ once and for all verboten for speakers of German” (p.288). Obviously, this would not be easy since “any notion of “impossible input” violates the principle of Richness of the Base, a cornerstone of standard OT, which accepts literally all universally viable linguistic representations as possible inputs” (p. 288).

This argumentation against Inkelas et al.’s proposal seems less than convincing in two ways. First, the claim that the sequence [ʃtɔᵣi] is an illegal nativisation of /stɔᵣi/ does not mean that it is an illegal nativisation of anything. As a matter of fact, the output [ʃtɔᵣi] is even predicted by Ito & Mester’s own grammar Faith(C) >> *sC >> ., once it is fed the right input – /ʃtɔᵣi/. The fact that such a form would be illegal in English should not matter if the Richness of the Base is the limit. Moreover, there actually exist English words which allow testing this hypothesis – words such as shrink or shriek have the right shape for this. Below we present how the most liberal German grammar from Ito & Mester (2001) treats the input shrink.
The tableau in (42) shows that structures like [C] and [i] can coexist within a word in the most liberal stratum of German.

Second, it seems like Ito & Mester gloss over an important property of their model: the integration tableaux like the one (41) presume an input form which has to come from another language. In this sense, it can hardly be used as a counter-argument to the proposal by Inkelas et al., which makes no claims about borrowing.

In sum, the criticism presented by Ito & Mester seems to have backfired because they stayed within a field in which they cannot win: as long as the strata are defined solely on segmental distributional grounds, the representational solution will be more economical. However, there seem to be domains where prespecification would mean introducing contrasts into the lexicon which are not necessitated on other grounds. So the fact that complex codas are tolerated only in loanwords in Serbo-Croatian (Simonović 2009b, Chapter 9) would necessitate prespecifying syllable positions in the lexicon, a famously unpopular move. Other non-segmental contrasts have also been shown to be problematic: Kager (2008) shows that quantity alternations cannot be handled by underspecification. Moreover, as soon as we introduce the arguments from morphology and categorical restrictions on stratal affiliation, the representational approach, unable to keep track of more or less underspecified lexical entries and group them together, has to capitulate.

This brings us to an important observation concerning the input representations in loanword integration: as will be discussed in the following chapter, loanwords will always arrive to the RL as fully specified structures and are in that sense not informative in RotB discussions. Their full specification might actually serve as one of the foreignness cues, helped by the fact that some of these structures are banned from the rest of the language. This matches the fact that loanwords add new structures to the language (see e.g. Rice 2006). Further support comes from the fact that loanwords have reduced paradigms and few related words in the language, which all enforces a specified representation. Note that even when whole derivational families from another language enter a new language, the alternations which occur in them will hardly point to an analysis which
assumes a common underlying representation (e.g. English velar fronting in syllabi[k]~syllabi[s]ity).

Given all the evidence for the existence of strata and morphology’s sensitivity to them, we cannot agree that underspecification can fully replace stratal Faithfulness. However, accounts which make use of more detailed representation should be taken seriously and a balance should be found, so that stratal effects are correctly captured, but not overgenerated. It seems that an optimal way of bridging this gap is to make stratification sensitive to the level of underspecification – underspecified structures would then be considered part of the core stratum of the lexicon.

3.6. Conclusion of the chapter and further development

In this chapter, we have reviewed the literature on lexicon stratification. The most important insights, which will also be relevant for our model, can be summarised as follows. First, there are good phonological and morphological reasons to integrate the stratification of the lexicon, as well as its formalisation in OT-based research – FAITH(LOAN) – into our loanword model. However, the representational alternative to the stratification model also introduces some important insights: loanword integration cannot be considered analogous to regular OT grammars in that its input is always fully specified. In the following two chapters we will give further evidence that the fact that loanwords have specified structure is a consequence of the design of the mental lexicon and the one optimally used by the grammar.
Chapter 4
The present model: ingredients and representations

The three previous chapters summarised the most important insights from the three research areas which take loanwords as their object: language contact research, formal research of borrowing, and formal research of borrowedness. We have seen that all three areas have developed very specific discourses which constrain what can be recognised as data and what can be a possible account of these data. Simplifying to an extent, we could say that there are two main dichotomies. The first dichotomy pertains to what is perceived as the goal of loanword research. This dichotomy separates the area of (sociolinguistic) contact research, less interested in pertaining to general research into the language faculty, from the two latter areas, which are only ever interested in relating loanword data to constructs of general linguistic theories (prevalently based on monolingual situations). The second dichotomy is between the two theories with general linguistic ambitions: whereas the theories of borrowing try to matter to the general theory of grammar by showing how loanwords are handled by either the grammar or a module strongly influenced by the grammar, the borrowedness models attempt to say something about the architecture of the lexicon and lexical representations. This separation is also evidenced by the nature of Optimality Theory models that these two subfields have produced: while OT models of borrowing mainly deal with the application of already justified Markedness constraints to new inputs (in other words: with what gets changed and how these changes relate to the native grammar) assuming special representations only when necessary, the borrowedness models mainly deal with what gets exceptionally preserved in loanwords and, by consequence, special representations and the special Faithfulness constraints indexed to these representations.

The core aim of the current approach is to overcome the two dichotomies in an attempt at integrating the insights of all three approaches into a simple, coherent model. This will be achieved by theorising two aspects of borrowing which are generally left out of the formal models: the bilingual competences and morphological integration. Including these two aspects of language contact into the model will prove helpful in overcoming both dichotomies. To begin with, in order to overcome the first dichotomy, it will become one of the tasks of our model to theorise what bilinguals know, i.e. what is represented in the bilinguals’ competence. As we will discuss in Chapter 6, data considered in the borrowing literature usually consist of SL/RL word pairs in which SL structures are shown to always map to the
same RL structures. This actually presupposes the existence of what we will call *inter-language mappings*, SL/RL correspondences established in the bilingual community and productively applied to new items. The concept of inter-language mappings will help us account for the role of bilinguals in language contact, not as an idiosyncrasy, but as a crucial element of systematic borrowing. Next, in order to overcome the second dichotomy, our data will crucially include the domain which shows the entanglement between grammar and lexicon: morphological integration, a type of facts conspicuously absent from general formal theories of borrowing. As a result, borrowing will be seen as the creation of a lexical representation dictated by the need to make a new item compatible with the language’s morphological and syntactic structure.

In sum, the core feature of our approach will be viewing borrowing as lexicalisation of a novel item (crucially including morphological and morphosyntactic integration). This process is always mediated by what borrowers already know about the RL and SL structures and how they are related, as represented in their inter-language mappings.

In this chapter, the main conceptual preliminaries of the current model will be discussed. This is done using data which are predominantly phonological and would be considered part of the explanandum of the models discussed in the two previous chapters. In the two subsequent chapters, we will focus on the data and accounts which are becoming visible due to design properties of the present model. In Chapter 5, we address the processes of morphosyntactic integration and in Chapter 6, we turn to inter-language mappings.

The remainder of this chapter is organised as follows. In §4.1, the input and output of the borrowing process will be considered. We will show that the difference between the two is ontological and greater than any model proposed so far has taken into account, and that the process of integration involves a radical change of representation. In §4.2, the levels of representations which are necessary for representing the trajectory of a loanword entering RL are considered, as well as the forces which guide the transformations that the representation of the word undergoes.

### 4.2. Representations and being faithful to them

Since the task of the loanword integration process is to turn a morphologically simplex surface string into a lexical entry, it is important to determine which specific levels of representation are involved as this transformation takes place. As will be shown in this section, considering the
present representations (and constraints referring to them) will bring us to a coherent model of loanword integration.

Our main claim is that there are only two representations which play a role. At the first stage, before the word has received a representation in the host lexicon, the relevant representation is the initial surface form. At the second stage, the relevant representation it is the new underlying representation.

The crucial question is how the presence of a representation is to be established. The existence of the initial surface form is uncontroversial from the very beginning (since there needs to be code-switching). Also, it is already clear that it would be misleading to conceptualise the switch from the initial surface form to a UR in the RL lexicon as a categorical one on the community level, since the same word can have a different status for different speakers. The same is arguably true on the individual level: words can be briefly lexicalised and then forgotten, and borrowed again. For this reason, it is important not to postulate a UR too early in the process. We therefore propose as a methodological strategy that the initial surface form should be assumed as long as possible.

While the presence of the initial surface form at the beginning of the process is unproblematic as such, its characterisation is far from a trivial issue. Crucially, in our approach, the presence of specific phonological features in either of the two forms is not to be taken for granted, but evidenced from the fact that these specific features are being preserved. So, for instance, since Dutch has no aspirated stops, it is not unproblematic to assume that the initial surface forms of the borrowings such as *picalilly*, *coach*, *claim*, *poster*, *team* necessarily have the aspirated initial consonants just because (monolingual) English has them. It can very well be the case that the aspiration is not a feature of the English of the relevant bilinguals or that it is conventionalised in this community not to implement it in code switches.

There is a good reason to focus on representations and faithfulness to them when describing the loanword trajectory. This reason lies in the fact that there are always different types of speakers involved in language contact. For this reason, what is usually described as loanword processes (for instance, and quite trivially, loss of aspiration in English borrowings into Dutch) defies a simple classification into either the synchronic or the diachronic kind. As already briefly discussed at the beginning of this chapter, one of the insights we are trying to incorporate in the current model is the fact that bilinguals are always aware of correspondences between RL structures and SL structures, which can have various sources, including being the result of factors playing a role at older stages of contact. In stable
language contact, structures are identified from existing word pairs and no word is really entirely new. This more ‘synchronic’ aspect of our approach will be explicitly addressed in Chapter 6. For the moment, as a good approximation, we are focusing on the more ‘diachronic’ aspect of loanword treatment, for which we use the concept of the loanword trajectory, following the transformations a word undergoes during lexicalisation.

As already mentioned, we hypothesise that the initial surface form is the form which surfaces at the moment when the new word is used in the RL discourse. The issue of the source code of the word, i.e. from exactly which lexicon the word is coming, is also not a trivial one. The fact that we generally cannot tell RL-accented SL from phonological integration into RL, has resulted in a shift of attention in the literature away from this stage of interaction between languages in the code-switching and language contact literature (Penelope Gardner-Chlores, p.c., see also Poplack 1980). The present model indeed forces the researcher to keep aware of the array of forms which can surface in the RL discourse. Importantly, this lack of certainty as to which exact representation is at play is less problematic than it might appear because all the relevant forms may very well share the feature under consideration.

The initial surface form (or indeed array of forms, which should not be confused with the paradigm of the SL word) influences the emergent RL paradigm in a way much comparable to that attested in creation of new allomorphs within a single lexicon – by enforcing the copying of salient surface features, such as long and stressed vowels, syllabification, final consonants, etc. The mechanism in charge here was termed Lexical Conservatism by Steriade (1997, 1998, 1999, 2000). Lexical Conservatism (henceforth LC) crucially relies on the concept of listedness: an allomorph is listed if it is attested in an established form. In paradigm formation, LC constraints then require every new allomorph to copy a phonological property of the listed allomorph. In (49), we provide the formal definition of LC in the form of constraints (from Steriade 1997).

(49) The form of Lexical Conservatism conditions: Lex (P)

Let $T(\mu)$ be the allomorph of $\mu$ appearing in a form under evaluation.
Let $L(\mu)$ be a listed allomorph of $\mu$.
Let $P$ be a phonological property.

$T(\mu)$ is characterized by $P$ only if some $L(\mu)$ is characterized by $P$.

Our implementation of the Lexical Conservatism constraint family as a property of loanword integration will be equivalent to that in Steriade. As it turns out, the concept of listedness is very well suited for research into the
integration of loanwords since there is no doubt which allomorph was listed at the outset: it is always the initial surface form.

The reader versed in Optimality Theory will note that Lexical Conservatism is similar to the mechanism of output-output Faithfulness (first introduced by Benua 1997). There are two reasons to maintain Steriade’s term and the proposed implementation. First, it is important to make a clear conceptual and terminological distinction between this preservation of surface features and input-output Faithfulness, which covers the preservation of features of underlying representations in surface forms. This is in order to avoid presuming unity where no unity is proven. Second, and related to the first point, unlike all types of Faithfulness, which are always seen as part of synchronic grammar, LC guides a process which is not unambiguously synchronic. LC-guided lexicalisation is better seen as a type of parallel diachronic change in multiple speakers, which then leads to lexicalisation of the new word across the community. This lexicalisation process then automatically brings the new word under the auspices of RL grammar by creating an RL underlying representation.

Once a new RL underlying representation has been created, its link to the corresponding SL word may continue to exist (and contribute to the inter-language mappings), but the representation of the RL word relies on the RL lexicon. This means that most of the features of the initial surface form preserved by LC are at this point invested in the new lexical representation, which is inevitably protected by input-output Faithfulness. Of, course, as illustrated by many cases of importation of foreign structures (the bread and butter of the approaches discussed in Chapter 3), LC sometimes preserves features which are not allowed in RL native words. When this happens, a fraction of LC effects is protected by a special kind of Faithfulness, the one indexed to loanwords. In sum, while LC is assumed here to be active during the process of paradigm formation, when lexical items have to be formed, Faithfulness is part of the synchronic grammar and plays a role in the computation of the surface form.

By taking the initial surface form as the point of departure, we are implicitly excluding the issue of the source of the initial surface form from our account. As we have seen in the discussion of the borrowing models, various properties of the initial form can have various sources, such as perception and the orthographic form, which can all contribute pronounceable units to the initial surface form. The selection of specific properties is often mistakenly taken to indicate what bilinguals have access to. For instance, if a contrast is neutralised in borrowing, the speakers are assumed to be unable to perceive it. We are pursuing a different approach by integrating the insight that, at least in stable contact situations, the
selection of properties which will appear in the initial surface form is regulated by inter-language mappings. This is especially conspicuous in cases where ‘changes’ undergone by loanwords cannot be ‘justified’ synchronically (see Chapter 6 for examples and further discussion). In order not to replicate the error committed by the models which privilege a single input type, our model is not aiming at identifying the exact source of the initial surface form, which is the starting point in the formation of the novel lexical entry. Instead, we propose that loanword integration research of the input form should focus on the elements of representation this form can have. As already mentioned above, evidence for the representation is then to be sought in the process of lexicalisation, during which some (surface) properties of the initial form are preserved.

In sum, rather than attempting to identify a single source of the loanwords’ representation, our model capitalises on the fact that the loanword trajectory starts with an initial surface form and ends with a full-fledged lexical item. In this sense, loanword integration is viewed as a lexicalisation process, during which LC enforces the preservation of (often non-contrastive) features of the initial surface form by militating against the emergence of new allomorphs.

In the following two sections two cases will be presented which justify the distinction between the two proposed representations, the initial surface form and the underlying RL representation (§4.2.1.) and the two mechanisms of feature preservation, LC and input-output Faithfulness (§4.2.2.).

4.1. The ingredients

In this subsection, the main ingredients of the novel model of loanword integration are presented. This is done step by step, moving conceptually from the endpoints – the input and the output of the integration process, tracing the path between the two representations.

4.1.1. Input

There are three important features of the input of loanword integration, which typically remain implicit in discussions of loanwords, and which it will be useful to highlight here:

a) Loanwords typically arrive in RL as single morphologically opaque forms – they cannot take their SL paradigms into RL.

b) Loanwords arrive to the RL lexicon as surface forms, a fact which makes them representationally crucially different from native underlying representations.
c) Loanwords have a specific meaning, typically a subset of the meaning of the lexeme in the source language. Each of these points has been taken into account in the literature to some level of explicitness. Some confirmation of the first claim can be found in the common assumption that borrowing is initiated in code-switching. As discussed in Chapter 1, one of the most robust findings of research into code-switching is that only one language is fully activated at a time. The fully activated language (termed the matrix language) then provides all functional morphemes, whereas the contribution of the other embedded language is limited to providing lexical material. In light of the fact that code-switching is the first stop on the borrowing trajectory, we can see how this insight from code-switching research has important consequences for borrowing: words enter the new language with a single form, which then serves as the base for the novel paradigm.

So far we have established that the initial single form is the one which is used in the bilingual discourse. In this pronounced/pronounceable form, all the surface features of the SL are present to the extent that they are present in code-switched forms. As we will see below, there is an independently necessitated mechanism, Lexical Conservatism, which enforces the preservation of these surface properties of the initial form in the integration process.

Finally, there is evidence that words are not borrowed with their entire networks of meanings; rather, their use in code-switching and as established loanwords demonstrates that only specific subsets of their meanings make it into RL (for an explicit discussion of semantic specificity in code switching, see Backus 2001). Examples of ‘English’ nouns in Dutch used with only a subset of their full meaning potential are easy to come by: in (43), some examples are provided.

(43) Meanings lacking form Dutch
Meaning in both D and E

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>game</td>
<td>‘video game’, ‘game in tennis’</td>
<td>‘way of playing’</td>
</tr>
<tr>
<td>goal</td>
<td>in sports</td>
<td>‘aim’</td>
</tr>
<tr>
<td>date</td>
<td>‘romantic appointment’</td>
<td>in the calendar, ‘appointment’</td>
</tr>
<tr>
<td>board</td>
<td>‘surfing board’, ‘council’</td>
<td>‘plank in general’</td>
</tr>
</tbody>
</table>

For each of the words in (43), there is a native Dutch word with a more general meaning, which is in no way ‘endangered’ by the presence of the borrowing, since the borrowing is covering a very small subset of its meaning. The general counterparts of the words in (43) are *spel* ‘game’, *doel/doelpunt* ‘goal’, *datum* ‘date in the calendar’ or *afspraak* ‘appointment’ and *plank* ‘board’.
We have now briefly introduced three crucial features of the input to borrowing: the single, morphologically opaque form, with its surface properties and a specific meaning. In order to illustrate the way in which these features of the input matter to the borrowing process, we present data from Slovenian/Serbo-Croatian contact.\(^\text{13}\) Serbo-Croatian and Slovenian are two closely related Slavic languages with comparable morphological systems. Virtually all Slovenian speakers have some command of Serbo-Croatian.\(^\text{14}\)

Serbo-Croatian has a large group of masculine gender nouns which have a mobile, invariably penultimate stress. When such words have cognates in Slovenian, they typically have an invariable stem-final stress (44a). Of course, there are many words which have this prosodic shape, but no cognate in the other language (44b).

\[
\begin{array}{lll}
\text{Serbo-Croatian} & \text{Slovenian} & \text{Meaning} \\
\text{(a)} & \text{vidîkNOM – vidîka\text{\textsc{gen}}} & \text{vidîkNOM - vidîka\text{\textsc{gen}}} & \text{‘view’} \\
& \text{delfinNOM – delfina\text{\textsc{gen}}} & \text{delfinNOM – delfina\text{\textsc{gen}}} & \text{‘dolphin’} \\
& \text{PârizNOM – Paríza\text{\textsc{gen}}} & \text{ParízNOM – Paríza\text{\textsc{gen}}} & \text{‘Paris’} \\
& \text{sîmbolNOM – simbóla\text{\textsc{gen}}} & \text{simbólNOM – simbóla\text{\textsc{gen}}} & \text{‘symbol’} \\
\text{(b)} & \text{lâžovNOM – lažóva\text{\textsc{gen}}} & \text{(no cognate)} & \text{‘liar’} \\
& \text{(no cognate)} & \text{skopúh\text{\textsc{nom}} – skopúh\text{\textsc{gen}}} & \text{‘miser’} \\
\end{array}
\]

Interestingly, when Serbo-Croatian words of the mobile-stress type enter Slovenian (45), they invariably maintain the stress of the original bare form (i.e. nominative singular form) throughout their paradigms.

\[
\begin{array}{lll}
\text{Serbo-Croatian borrowings in Slovenian} & \\
\text{Serbo-Croatian} & \text{Slovenian} \\
pásuljNOM – pasúlja\text{\textsc{gen}} & \text{pásuljNOM – pásulja\text{\textsc{gen}}} & \text{‘beans’} \\
rôštiljNOM – rôštilja\text{\textsc{gen}} & \text{rôštiljNOM – rôštilja\text{\textsc{gen}}} & \text{‘grill’} \\
štápićNOM – štapić\text{\textsc{agen}} & \text{štápićNOM – štapić\text{\textsc{agen}}} & \text{‘stick’} \\
rážnjićNOM – rážnjič\text{\textsc{agen}} & \text{rážnjićNOM – rážnjič\text{\textsc{agen}}} & \text{‘stick’} \\
Bihàc\text{\textsc{nom}} – Bihàč\text{\textsc{agen}} & \text{Bihàc\text{\textsc{nom}} – Bihàč\text{\textsc{gen}}} & \text{‘Bihać’} \\
\end{array}
\]

\(^{13}\) I’m grateful to Marko Hladnik and Mili Gabrovšek for all the help with the Slovenian data.

\(^{14}\) In the Eurobarometer report “Europeans and their Languages” Serbo-Croatian surfaces under the name Croatian as the strongest foreign language of Slovenia: 59% of the speakers report speaking it “well enough in order to be able to have a conversation”. This number would probably be much higher if the other names for the same language (Serbian, Bosnian, BCS, Serbo-Croatian) were taken into account. The report only mentions the top 3 languages.
There is no doubt that Slovenian speakers are able to perceive Serbo-Croatian mobile stress and implement it when they speak Serbo-Croatian. Moreover, Slovenian also has itself a small group of nouns with mobile stress (e.g. \( \text{rázlog}_{\text{NOM}} - \text{razlóga}_{\text{GEN}} \) ‘reason’). It is therefore not immediately clear why Serbo-Croatian mobile paradigms do not become Slovenian mobile paradigms.

The tendency for a single language to be activated which is observed in the code-switching literature (as discussed in Chapter 1) offers a clue here. Importantly, these loanwords were initially Serbo-Croatian code switches in Slovenian discourse. In terms of the Matrix Language Frame (MLF) (Myers-Scotton 1993, 2002), they originate from the discourse in which Slovenian plays the role of the fully activated language – the matrix language, whereas Serbo-Croatian is restricted to providing lexical material. Taking this initial stage of borrowing into account helps us provide an explanation for the fixed stress pattern: Serbo-Croatian paradigms (which are required in order to establish mobile stress) are not available and all the forms in the emergent Slovenian paradigm are based on the bare Serbo-Croatian form, which is also the citation form.

In sum, while Slovenian speakers have full paradigms when they speak Serbo-Croatian, they do not use these paradigms when Serbo-Croatian is the embedded language in code-switching, so that the new Slovenian paradigm is based entirely on the morphologically simplex nominative singular form (e.g. \( \text{pásulj} \)). This in turn gives rise to invariable stress. The limitations of code-switching offer an explanation for the fact that the only Serbo-Croatian form which makes it into Slovenian is the morphologically opaque citation form.

Similarly to the examples in (43), if we consider the meaning of the common nouns in (45), we see that the meaning of the newly established Slovenian word is always much more specific than that of the Serbo-Croatian original. In other words, in each case, there is a specific meaning which is now shared between the two languages and a general meaning, which is only available in Serbo-Croatian.

\[(46) \begin{array}{ll}
\text{Shared S-C&SL meaning} & \text{General meaning available only in S-C} \\
pásulj ‘meal made of beans’ & \text{plant, food product, etc.} \\
štápić ‘stick for grilling meat’ & \text{Chinese sticks, ice-cream sticks, etc.} \\
rážnjić ‘stick for grilling meat’ & \text{diminutive of ražanj ‘spit’} \\
róštilj ‘grilled meat’ & \text{grill, barbecue} \\
\end{array}\]

Here also, for the general meaning, RL uses another word. In the first three cases there is a native Slovenian item (\( \text{fižol} \), \( \text{palčka} \) and \( \text{raženjček} \)).
respectively), whereas in the fourth case next to the native žar, English borrowings (barbecue, barbekju) are common.

In sum, the examples from Slovenian/Serbo-Croatian contact show that Serbo-Croatian items entering Slovenian are not complete lexical items in at least two ways. First, on the side of the form, only the bare form is taken into account. Second, on the side of the meaning, only a very specific sub-meaning of the Serbo-Croatian noun is transferred into Slovenian.

In this subsection we have defined the input to loanword adaptation: it is a morphologically simplex surface string with a specific meaning which serves as the base for a paradigm, if one is created. At this point already, we can see that the common metaphor of word borrowing can obscure the nature of the object of transfer. This object is neither a ‘word’ with its paradigm nor the full meaning of the ‘word’ in SL.

4.1.2. Output

The output of loanword integration is a lexical entry. Crucially, this amounts to different things in different languages. The differences between recipient lexicons pass under the radar if morphological and syntactic properties of the lexical entries are not taken into consideration, as in most purely phonological accounts we have discussed in Chapter 2.

Importantly, being a lexical entry does not necessarily imply having a native-like paradigm. For instance, in Russian, loanwords like metro and kupe ‘compartment’ (and other foreign items ending in -o and -e) are indeclinable, as can be confirmed from the comparison between the paradigms of the native noun pero ‘feather’ and metro in (47). In this, loanwords join a small class of names (e.g. Ukrainian surnames in -enko) which do not decline. Even so, nouns like metro are readily used in all syntactic contexts in which any Russian noun would be used. In other words, the nouns like metro nouns do have full paradigms, which are special only for the fact that all the slots are filled by the same form.

(48) The declensions of metro and pero (from Wiktionary.org)

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominative</td>
<td>peró</td>
<td>pér'ja</td>
<td>metró</td>
<td>metró</td>
</tr>
<tr>
<td>genitive</td>
<td>perá</td>
<td>pér'jev</td>
<td>metró</td>
<td>metró</td>
</tr>
<tr>
<td>dative</td>
<td>perú</td>
<td>pér'jam</td>
<td>metró</td>
<td>metró</td>
</tr>
<tr>
<td>accusative</td>
<td>peró</td>
<td>pér'ja</td>
<td>metró</td>
<td>metró</td>
</tr>
<tr>
<td>instrumental</td>
<td>peróm</td>
<td>pér'jami</td>
<td>metró</td>
<td>metró</td>
</tr>
<tr>
<td>locative</td>
<td>peré</td>
<td>pér'jah</td>
<td>metró</td>
<td>metró</td>
</tr>
</tbody>
</table>
The creation of a fully functional lexical item is the natural result of loanword integration. We are not aware of any descriptions of languages in which words systematically stop midway into lexicalisation, for instance by receiving incomplete paradigms.

4.1.3. Between the input and the output
Loanword integration is the process which turns an initial surface form ('injected' into RL by bilinguals), which is not a lexical entry, into a lexical entry in the RL lexicon. The term integration does justice to the actual process, which crucially involves a change of representation, but not necessarily in the direction of what is common (or even attested) in native words. This approach makes comparison between language contacts a much more precarious business than was the case in models which assume adaptation (turning into native like structures), but it produces valid generalisations and an explicit loanword theory situated with respect to other linguistic research areas, able to inform the general theory of the architecture of grammar. For instance, what can already be read off the Slovenian and Russian paradigms above is the tendency for loanwords to avoid joining paradigms in which the stress shifts in different forms. In the following sections, we shall see that this tendency can be derived using mechanisms which have already been proposed for relations within languages, but it will also become clear from our model how the situation of a loanword is different from patterns encountered within languages.

In the next section, the stages of the loanword integration trajectory will be discussed, as well as the representations attested at these stages.

4.2.1. Representations: Serbo-Croatian and Slovenian derived environment effects
In at least some languages which have relatively complex nominal paradigms, certain foreign segments can surface only in morphologically simplex, non-derived contexts. Consider the Serbo-Croatian sentences in (50), in which the English name Patrick is used as a code-switch.

(50)  
  a. On se zove [pʰætɹɪk].
  b. On se zove [patrik]/[petrik].
  ‘His name is Patrick(nominative).’
  c. * Razgovarao sam sa [pʰætɹɪkəm].
  d. Razgovarao sam sa [patrikəm]/[petrikəm].
  ‘I talked to Patrick(instrumental)’
The acceptability of (50a) shows that, for the speakers of Serbo-Croatian who master the pronunciation of the English segments [pʰ], [æ], [ɹ] and [ɪ], which are not present in Serbo-Croatian, these segments are possible in the bare form code switches. However, these segments are banned from derived forms (50c). Similar Slovenian and Catalan data have been presented by Jurgec (2010). This author terms such effects long-distance derived-environment effects (DEE) and accounts for them using local conjunction between Align-R(stem, PrWd) and markedness constraints militating against non-native segments. While this account does work, it uses a new type of constraint interaction, the one which allows for long-distance derived-environment effects, which seem to be attested only in loanword treatment. In being purely formal, this proposal then seems to miss the contact nature of the phenomenon.

Our interpretation of these data is informed by a property already mentioned and verified with speakers of Serbo-Croatian and Slovenian: as a rule, non-native segments in non-derived environments surface in speakers who have some command of the source language. This being so, it is more realistic to analyse words containing foreign segments as code switches. This in turn means that it is safe to keep the generalisation that Serbo-Croatian and Slovenian do not tolerate the foreign segments [pʰ], [æ], [ɹ] and [ɪ]. This matches the fact that, as soon as paradigms are formed, i.e. as soon as the word starts receiving a representation in the host lexicon, non-native segments become nativised. In other words, RL Markedness has no say over code switches (insertion of surface forms from another language), because they do not pass a native evaluation. However, Markedness has its say in all forms which are based on RL representations and inflected/derived forms are necessarily of that kind. In this sense, the grammar of Serbo-Croatian (or Slovenian), which is responsible for dealing with the representations coming from the Serbo-Croatian (or Slovenian) lexicon, does not change under the influence of loanwords. Arguments in favour of this analysis can be found in Jurgec’s data as well. Consider, for instance, the following English words in Slovenian from Jurgec (2012).

(51)  English     Slovenian
      wʃɓnktan     uʃɓnktan-a
‘Washington’  ‘Washington.gen.sg’
      wilson       vilson-tʃok
‘Wilson’      ‘Wilson.diminutive’

Although these words are used by Jurgec to illustrate the occurrence of w only in non-derived environments, they also show a number of other
substitutions: [ɔ] becomes [a] and shwas are replaced by [o]. None of these substitutions can be explained by the ranking in the Slovenian OT-style grammar, in which shwa alternates with Ø, [a], [e] and [ɛ], whereas [ɔ] alternates with [o]. Rather, the resulting sounds clearly come from the written form. From this we see that the form in the bare context does not serve as the only base from which the inflected forms are derived. As a matter of fact, the substitutions in derived forms follow an older pattern of adaptation, which relied on orthographic information to a greater extent. Nowadays, [ɔ] is consistently borrowed as [ɔ] or [o] (e.g. face wash is borrowed as [fejs ɔj] or [fejs ʊj]). Even though the borrowings quoted by Jurgec are relatively old, the bare form remains the environment in which the proficient bilingual has the option of inserting a code-switch.\footnote{15}

As already mentioned, after the paradigm has been formed, the bilingual speaker is free to code switch in bare contexts, but she does not have to, i.e. the bare forms with all native segments (in our examples: [patrik]/[petrik], [vaʃɪŋktɔn] and [vɪlson]) are also possible. According to my native speaker informants, this is true of both Slovenian and Serbo-Croatian.

In sum, Jurgec’ approach is overly grammatical/phonological and misses some of the aspects specific to loanwords. When these are taken into account, the result is more loanword-specific and calls into question the very ‘grammatical’ nature of the phenomenon. Moreover, our approach enables us to situate the relevant phenomenon with respect to grammar by identifying the point when a word comes under the auspices of RL grammar.

The Slovenian examples in (51) are telling as such, but they represent a specific situation in which language contact has lasted long enough for several integration patterns to have been active. In the prototypical case, the code switch does show the features from which the properties of the derived forms can be arrived at.

The generalisation that bare forms create an environment in which code-switching is possible implies an interesting typology concerning the availability of the two forms, the code switch and the integrated loanword. On the one hand, in cases where inflection is not required, the original pronunciation is longer available and/or possible. A case at hand are newly borrowed English adjectives in Serbo-Croatian (illustrated in 52), where the original pronunciation is available for proficient speakers.

\footnote{15} The option of inserting a code-switch where there is an established loanword seems particularly acceptable in names. This is arguably related to the fact that, while most loanwords have a different semantic content from the SL original (due to semantic specificity discussed above), in names, this difference between SL and RL does not arise.
100

Chapter 4

(52)

\textit{ta glamurozn-a/jef\'{i}n-a ode\'{c}a}

\textit{ta GLAMOROUS_/CHEAP_ ode\'{c}a}

‘those glamorous/cheap clothes’

On the other hand, in languages where there are no bare (citation) forms (as is for instance Lithuanian), uninflected SL forms can only be encountered in larger code switches, such as clausal and sentential code-switch. This means that since Lithuanian also does not tolerate non-native segments in inflected form (Jolanta Šinkūnienė, p.c.), the only environment in which non-native segments can surface is constituted by longer code switches.

While we are arguing that the Slovenian effects described by Jurgec do not necessitate a formalisation in terms of derived-environment effects, it is important to stress that non-derived forms do have a special status in language contact. They impose themselves as candidates par excellence for single-word code switches, and therefore also for initial surface forms. This special status has been recognised in L2 acquisition research: Eckman et al. (2001) have found that target language contrasts absent from L1 are first acquired in non-derived environments. This may very well be related to the fact that non-derived forms contain a single expressed morpheme and therefore impose less processing burden on the speaker and that, cross-linguistically, they tend to be the citation form.

In this section we have discussed the nature of the relevant representations along the loanword trajectory. In the following subsection, we turn to the question of how the mechanisms which are in charge of preservation of phonological features of along the loanword trajectory are put to use.

4.2.2. LC and FAITH – Serbo-Croatian, Macedonian and American Italian

This section focuses on the preservation mechanisms involved in maintaining the features of the two representations discussed above: the initial surface form and the underlying RL representation. As we have already established, the presence of the initial surface form is unproblematic at the onset of the integration process, as is the mechanism which protects the features of this form, LC. The main question is then how and at which point Faithfulness takes over from LC. In what follows, we will consider three possible scenarios of loanword integration which illustrate three different ways in which the results of LC are maintained after lexicalisation has taken place.
4.2.2.1 Serbo-Croatian a-epenthesis
Serbo-Croatian native words never contain complex codas. Obligatory a-epenthesis prevents such structures from surfacing: underlying /visk/ maps onto the surface form visak ‘pendulum’ (cf. visk-a ‘pendulum-genitive’).

However, complex codas are attested in thousands of frequent loanwords, such as Bask ‘Basque’ and disk ‘disc’ (but also kobalt ‘cobalt’, ofsajd ‘offside’, jamb ‘iamb’, marketiŋg). Given the fact that there is a native process which is blocked in loanwords, we have to assume that these words, after having been inserted as code switches, cannot receive the identical underlying representation as the native words such as visak – they have to be flagged as loanwords in the lexicon and Faithfulness indexed to loanwords has to be invoked. The crucial ranking of constraints after lexicalisation is then DEP(LOAN) >> *COMPLEXCODA >> DEP (while at lexicalisation point, it was LC >> *COMPLEXCODA >> DEP).

Note that the stratification approaches (see Chapter 3) correctly predict that novel loanwords such as helpdesk or task will be integrated without epenthesis, because they enter the periphery of the lexicon and automatically come under the auspices of FAITH(LOAN). In sum, once LC has introduced certain structures to the lexicon, FAITH(LOAN) can take over and preserve them and LC need not play any role in the synchronic grammar. LC is only relevant when the lexical entry is under (re)construction; when there is no instability, FAITH suffices.

4.2.2.2. Macedonian lexical stress
Native Macedonian words have a regular stress pattern – antepenultimate if the word is trisyllabic or longer, otherwise initial (53a). However, there are a number of lexically stressed words with either a stressed penult or ultima (Hammond 1989). As it turns out, all words quoted by Hammond are nouns borrowed from Russian and French. These nouns have stem-bound stress, that is the stress of the source word copied to all the other paradigm members, as long as the stress stays within the three final syllables. In cases where stress would have left the three-syllable window, the native stress pattern is followed (53b-c).

Similarly, if the initial surface form of a loanword has antepenultimate stress (53d), in all longer derived forms, the trisyllabic window would be violated if stem-based stress were to be maintained, so the native-like pattern surfaces immediately. This is the reason why the borrowing audíciija behaves on a par with the native vodénica. This means, in other words, that if the source form has antepenultimate stress, the new lexical item automatically becomes native-like.
(53) bare form  definite singular  definite plural
(a)  grad ‘town’  grádot  gradóvite
    ódnos ‘relation’  ódnosot  odnósíte
    góspodar ‘owner’  gospódarot  gospódárite
    vodénica ‘mill’  vodenícata  vodenícite
(b)  restorán ‘restaurant’  restoránot  restoránite
(c)  konzumátor ‘consumer’  konzumátorot  konzumatórite
    (*konzumátorite)
(d)  audícija ‘audition’  audícijata  audícijite

In this case, our account also assumes that the integration trajectory starts with a code switch, introducing a word with the non-native stress pattern into RL discourse, which consequently spreads its stress pattern to the rest of the paradigm, as far as is allowed by undominated Markedness – in this case the constraint ranking which maintains the three-syllable window (see Kager 2012 for a formal OT analysis). Now, at the point in time where speakers create a representation in their Macedonian lexicon, there is no reason to invoke indexed Faithfulness: general FAITH(STRESS) can do as well. As a consequence, the newly introduced items can be treated as the only Macedonian words with lexical stress. Note that this does not immediately predict that lexical stress will ‘spread’ to native words.

In this scenario, the effects of LC – the preservation of the original stress pattern, which needs to be assumed anyways – can be maintained using general Faithfulness, so there is no need for stratum formation. Note that the distinction between this scenario and the previous one cannot be made in a model which assumes stratification as the all-round solution for loanwords (see Chapter 3).

4.2.2.3. American Italian stem-final gemination

In languages which allow geminates, there is a general pattern to integrate foreign singleton word-final consonants as geminates (see Chapter 8). A case in point is American Italian, in which geminates are common in native words (cf. setta ‘sect’ and seta ‘silk’). In this variety, the English word coal gets adapted as [kol.le] and not as [ko.le], which is phonotactically as licit (for further examples and references for other languages see Repetti 2009).

In different accounts proposed in the literature, this gemination has been interpreted as a morphologically conditioned effect (Italian, see Repetti 2009), a phonotactically conditioned effect (Japanese, Shinohara 2003) or even as a case of Retreat-to-the-Unmarked (Hungarian, see Kenstowicz 2005).
In the present model, the account of gemination can be unified. Gemination is then viewed as a consequence of the fact that the syllabification of the initial surface form is better preserved in forms with a geminate: in *kol.le* better than in *ko.le*. Furthermore, even in languages which allow both singleton and geminate coda consonants and have bare citation forms, the original syllabification is better preserved throughout paradigms if the SL coda is reproduced as a geminate. A case in point is Hungarian (Nádasdy 1989). For instance, the fact that, this language, *match* is reproduced as *meccs* [mɛtʃː] does not seem particularly beneficial if we only consider the bare form: *meccs* [mɛtʃː] is not any better at replicating the original syllabification than *mecs*[mɛʃː]. However, already in the plural form *meccsek* [mɛʃː.ɛk], it becomes clear that gemination guarantees preserving the original syllabification throughout the paradigm.

In this scenario, the new items look like any of the native items whose stem ends in a geminate. This being so, there is no need for either the grammar or the structure of the lexicon to change in any way in order to accommodate these newcomers. The facts of this type, together with the Slovenian data from §4.1., falsify the assumption discussed in §4.2. that loanword modification is by definition a matter of the RL grammar. The processes illustrated by this scenario have little to do with native processes. In native words, lexical items are never derived from a single surface string, so that what happens to loanwords can only be captured if we recognise that we are dealing with something conceptually different from native evaluations.

Compared to the two scenarios discussed in the previous two sections, this is one in which all work is done by LC. In the first scenario both RL grammar and RL lexicon were active, in the second it was only RL grammar that was invoked. In this third scenario, LC is basically the only relevant player and the native grammar and the lexicon are delivered a native-like item.

### 4.3. Conclusion of the chapter and further development

In this chapter, we have addressed the issue of the nature of the two endpoints of the loanword trajectory, the input, the initial surface form, and the output, the RL lexical item. After that we have addressed the nature of the processes which preserve these two representations, and identified the mechanism which drives the transformation from one to the other: Lexical Conservatism.
The next chapter will be concerned with the morphological and morphosyntactic part of the borrowing process, which becomes visible once the new model is put to use. Before that, in a brief intermezzo, the possible scope and mission of the new model will be discussed.
Chapter 5
Morphology and morphosyntax of loanword integration

In the previous chapters we have seen how phonological modifications which occur in loanword integration can be accounted for if the process is viewed as the creation of a new lexical item based on incoming foreign surface material. This has helped provide a reanalysis of the phonological data which have been used as evidence by the phonological accounts discussed in Chapters 2 and 3, and situate different approaches in these research traditions as complementary rather than opposed.

However, the move to a lexicalisation perspective of loanword treatment has far-reaching consequences for both the theoretical embedding and the empirical coverage of loanword research. On the theoretical side, the present model is focused on the question of how loanword integration is positioned with respect to RL grammar and lexicon. As we have seen in Chapters 2 and 3, the goal of most previous models was to account for the modifications undergone by loanwords by showing how the model predicts the good result of the process using as few loanword-specific mechanisms as possible. These approaches were constructed around the question of whether loanword integration is some sort of RL grammar, assuming a yes–no answer. The present model avoids such an assumption by viewing loanwords from the perspective of integration into the lexicon. Viewed this way, loanwords, which are assumed to lack virtually all features present in native lexical entries (paradigms, morphosyntactic features, etc.) at the outset of the integration process, are unmistakably different from native words (at any level of their representation). Therefore, the process of loanword integration is not equivalent to anything encountered in grammar. The goal of loanword research is then not to predict the correct output in the simple mapping from one string to another but to restrict the space of possible interactions in the complex process in which bilingual knowledge, lexicon and grammar are entangled in determining the phonological, morphological, syntactic and semantic features of the new lexical item.

On the empirical side, this repositioning has important consequences for what a loanword model is supposed to account for. If loanword integration is lexicalisation, phonological modifications lose both their primacy and their ontological status as independent objects of explanation: loanwords do not enter the phonology of a language, they can only ever enter the lexicon, and what guides their ‘phonological’ integration is their integration into the lexicon. As a result, phonological facts lose their role in the selection of data and the processes of morphological and morphosyntactic integration are
legitimate objects of loanword models even if they do not affect phonological modifications.

This chapter will focus on the part of the borrowing process which involves morphological and morphosyntactic integration, the part which has generally been excluded from phonology-dominated models. In order to do this, we will present a (by no means exhaustive) catalogue of effects of the relevant processes, showing how the proposed model restricts the possible interactions between the lexicon and grammar. In other words, the necessary formal apparatus presented in the previous chapter will, without any substantial extensions, be put to use to account for new types for data.

As in the previous chapter, we still assume the ‘diachronic’ perspective of loanword integration in the sense that we are observing the loanword trajectory, with the implicit assumption that a single idealised speaker is integrating a word into her lexicon. The discussion of the ‘synchronic’ aspect of loanword integration – the regular mappings internalised by the integrating speakers/bilingual community and their role in the model – is postponed until Chapter 6.

The remainder of this chapter is organised as follows. In §5.1. a brief overview is made of assignment of morphosyntactic properties. In §5.2. through §5.4. case studies are presented, each highlighting a feature of the initial surface form and showing how this feature influences the process of morphosyntactic assignment. §5.5. concludes this chapter.

5.1. Morphosyntactic assignment: Lexical requirements and morphosyntactic interpretation

The accounts presented in this chapter assume the view of loanword trajectory taken in Chapter 4, which takes the initial surface form (the surface form that can be inserted into RL discourse) as its starting point. There are three properties of the initial surface form presented there, and slightly reformulated here. First, the initial surface form is typically lacking any morphological or morphosyntactic features (with the possible exception of the specification for the part of speech). Second, it is a relentlessly surface form, which can mean that it contains cues for interpretation in terms of morphological and morphosyntactic properties of the recipient language. Third, it has a specific meaning.

Morphosyntactic integration of loanwords can then be recapitulated as two broad, partially overlapping processes: one is the consequence of what the initial form lacks, the other a consequence of what it has. First, because the initial surface form has no morphosyntactic specification, it needs to acquire a specification for all the morphosyntactic features which a lexical entry in the borrowing language has to contain in order to be fully functional.
Second, being full-fledged surface forms and having a meaning, the initial forms can be ‘morphosyntactically interpreted’ in terms of RL categories, based on generalisations holding within RL. In other words, there are two types of generalisations loanwords depend on in their integration: on the one hand, there are lexical requirements, which have to be met in order for the new lexical item to be created, and on the other hand, there are surface generalisations and meaning generalisations, which relate certain chunks of linguistic material or certain semantic information to morphosyntactic categories, which hold within RL and naturally extend to incoming forms.

As already mentioned, there can be an overlap between the two types of processes because morphosyntactic interpretation triggered by the surface form can either be in terms of categories for which morphosyntactic assignment is required or it can be in terms of categories which are not obligatory properties of every lexical item. In the first case, lexical requirements are met with the help of surface-true phonological and morphological generalisation, for instance when the surface form or the meaning trigger certain gender assignment in a language where all nouns are specified for grammatical gender. In the second case, for instance, when a language interprets certain loanwords as morphologically complex due to segmental or prosodic cues, interpretation is not related to meeting lexical requirements.

A typical and very common case of morphosyntactic assignment is that of paradigms/inflectional classes. The proposed view of the loanword trajectory enables making specific predictions here. Given the fact that loanwords typically start their career as morphologically simplex strings whose surface properties tend to be copied as much as possible (due to Lexical Conservatism) we can make the prediction in (54).

(54)

In every language, all other things being equal, the paradigm/inflectional class which involves the least allomorphy will be the host for loanwords.

This is indeed true of many languages, some of which we have already seen: the Slovenian examples in §4.1. illustrate a case where loanwords join an existing inflectional class, and that is the one with the least allomorphy in the language. There are also cases in which RL lacks a ‘stable’ paradigm which could host loanwords, so the influx of loanwords causes the creation of a novel paradigm. The Macedonian facts from §4.2.2.2. illustrate a case where loanwords in fact start a new type of paradigm – one with stem-bound stress, so that the number of prosodically different allomorphs is limited as much as RL constraints allow. In Simonović (2009a) further examples are provided from Slovenian, Hebrew, Polish and Turkish.
In all these cases, the inflectional classes which host loanwords are the ones with stem-based stress. This is not surprising, given the fact that the initial form typically coincides with the new stem. Frequent borrowing into paradigms with stem-based stress can even determine the most common stress pattern in the language. For instance, Stojanov (1983) describes the following situation in Bulgarian. The definite ending -ăt can attract the stress if attached to monosyllabic nouns: *glas* - *glasăt* 'voice'. The corpus used by Stojanov contained 510 monosyllabic masculine nouns, out of which 236 were Slavic and 274 were borrowed. Among the native words a small majority of 133 had a stem-bound stress in the definite form, whereas 103 of them had a stressed definiteness marker. Loanwords are, expectedly, generally non-alternating: almost all of them (269) have a stem-bound stress (*tank, talent, tampon*), whereas only five borrowed nouns form an exception to this regularity (*klas, stil, tel, tim* and *tip*). Consequently, loanwords make up more than two thirds of the class of nouns with the most common, stem-bound stress (269 compared to 133 native nouns). \( \chi^2=103.35, p\leq0.0001 \)

So far, we have only seen cases where the assignment of paradigms (and the assignment of other morphosyntactic categories) can occur without violations of LC constraints because the initial form coincides with the new stem. However, there are cases in which loanwords trigger interpretation in terms of the RL grammar, which proceeds through splitting the morphologically opaque initial form into two morphemes. This necessarily occurs to the detriment of LC constraints. For instance, in Serbo-Croatian virtually every item ending in -a in the nominative form is analysed as *stem+a* and assigned to the feminine inflectional class. Among loans, the Indonesian town’s name *Kuta* will automatically be turned into *Kut-a* (*Kúta*, accusative singular *Kútu*). In this case, a surface phonological property is the trigger for interpretation and the assignment of internal morphological structure (*stem+a*), inflectional paradigms and morphosyntactic features (in this case, the feminine gender). This can condition further modifications, which may appear phonological at first sight. For instance, the fact that the -a is never stressed (due to a general ban on final stress on desinences) can cause a stress shift only in this specific class. So, whereas Serbo-Croatian tolerates final stress in foreign names such as *Pastër* (Pasteur, accusative *Pastër-a*) and *Fukó* (Foucault, accusative *Fukó-a*) in which the new stem coincides with the initial form, it often moves the stress in items such as *Deríd-a* (Derrida, accusative *Deríd-u*) and *Dima* (Dumas, *Dím-u*), where preserving the original stress pattern would lead to a different morpheme being stressed in each form\(^{16}\). In sum, what appears to be a simple case of

\(^{16}\) An alternative way of dealing with the final stressed -a in male names is transforming the initial form into an a-final stem, as in Fransoá (accusative Fransoáa). As expected, no stress shift occurs in such items.
accommodation to the prosodic constraints of Serbo-Croatian, turns out to be a local pattern which depends on morphosyntactic interpretation.

In other cases, assignment of a paradigm is not influenced by the structural features of the surface form, but by its meaning. For instance, the only class of indeclinable nouns in Serbo-Croatian consists of nouns which have female referents and do not end in -a (lejdi, Jane/Džejn, Femke, Lulu, Lo, Lilly/Lili). The female referent is the only feature by which these nouns differ from all other borrowed nouns ending in a consonant or a vowel other than -a. The latter type of nouns all become declinable by joining the masculine declension class, where the new stem generally coincides with the initial form. There are even minimal pairs to illustrate this distinction. For instance, the indeclinable female name Lilly/Lili is homonymous with the masculine noun lilly/lili (a perfumery chain, also a shop in that chain), which has the paradigm liliNom, lilijaGen, lilijuDat, etc. The reason for the creation of this island of indeclinability in a language in which all nouns decline lies in the fact that in the nouns of the type Lilly/Lili and lejdi (and only in them) there is a mismatch between the gender assigned based on the natural gender of the referents and the unavailability of an inflectional class which would host feminine-gender nouns not ending in -a in S-C. This results in items of the type Lilly/Lili and lejdi remaining indeclinable.

In a small number of cases (which seem to be restricted to standard languages with normative tradition) morphosyntactic information can be exceptionally borrowed, which can then lead to blocking the development of a native-like paradigm. In Czech, the word aroma was borrowed with neuter specification (presumably from German). As Kavitskaya (2003) put it “if aroma were to be borrowed as a feminine noun and analysed as having a morphological ending -a, it would decline according to the feminine declension. However, it is neuter and thus indeclinable since no neuter noun in Czech can end in -a” (p. 268). The only way in which aroma is different from the Serbo-Croatian items like Lilly/Lili and lejdi is the source of the morphosyntactic feature: while aroma has kept its gender from the source language, Lilly/Lili and lejdi are feminine due to the natural gender of their referents.

In between indeclinability and full declinability, we are aware of one case in which LC allows declining only when the initial form stays together. Naughton (2005) shows that in Czech a small number of borrowed items ending in -le, -re and -te (e.g. finále), are indeclinable in all cases except the instrumental singular (the form is then finálem). As it turns out, the instrumental singular is the only form other than the citation form, in which, were finale to be declined like the native pole ‘field’, the final -e would survive, by virtue of being part of the ending.
In this section we have seen that the integration of loanwords is determined by the fact that (a) loanwords have to receive certain morphosyntactic features (including paradigms) in order to become RL lexical items, and (b) loanwords receive some additional morphosyntactic features (such as internal morphological structure) due to the fact that they are fully specified in their form and their meaning. Note that the distinction between the required and the additional morphosyntactic features is not ontological, but simply used in order to cover the full diversity of phenomena of morphosyntactic integration.

In the remainder of this chapter, it will be shown how lexicon and grammar interact in morphosyntactic integration and how the apparently unconstrained diversity of phenomena is restricted by the principles outlined in our model: Lexical Conservatism and the properties of the initial form (morphological opacity, the surface nature and the specific meaning). The case studies are presented in three sections, each of which features one of the characteristics of the initial form presented at the beginning of this section (and Chapter 4): the fact that it lacks morphological and morphosyntactic features (§5.2.), the fact that it is a surface form (§5.3.) and the fact that it has a specific meaning (§5.4.).

5.2. The initial form is void of morphological and morphosyntactic features

This section focuses on the consequences of the observed regularity that the initial form lacks morphological and morphosyntactic features, or at least fails to transfer them from the SL into the RL.

In §5.2.1., English verbs currently entering Serbo-Croatian will be used to exemplify how LC-effects, in combination with the lack of a morphosyntactic specification (in this case aspect), can lead to the assignment of exceptional morphosyntactic features to loanwords (in this case biaspectuality). As will be shown, the exact combination of the presence of some features (word class) and the absence of others (aspect) ushers a specific type of treatment, which is practically restricted to English verb stems. In §5.2.2. the pattern for borrowing English band names will be used to illustrate a case where the input form is not simplex from the perspective of the source language because the simplex form is either not available or not prominent (as in the English rock band names). This will be shown to lead to a type of hyper-regular paradigm formation. §5.2.3. presents a case from Maltese where, apparently counter to our predictions, Italian allomorphy seems to get borrowed into Maltese. As will be shown, this transfer of allomorphy is actually preferred by the relevant LC constraints.
5.2.1. Integration of English verbs into Serbo-Croatian
In this section we present a sketch of the current integration pattern for English verbs into Serbo-Croatian (e.g. rifreš(ov)ati ‘to refresh’), which is also the object of a diachronic analysis in §6.4.2. and a more detailed discussion in Chapter 10. As we will see, the borrowing pattern can only be accounted for if the trajectory specific to borrowed verbs is taken into account. Here, our focus is on the emergence of bispectuality and on the fact that bispectuality seems reserved for verbs derived from foreign verbs, while other productive patterns of verb derivation always yield single-aspect verbs. Additionally, this case will show the virtues of the conceptual separation between RL grammar and the borrowing pattern, as the way English verbs are treated is strictly distinct from any regularity in the native domain.

Serbo-Croatian has the general Slavic distinction between perfective and imperfective verbs. The ‘pairs’ usually have a shared stem (i.e. kup-itiPF vs kup-ovi-taIMP ‘to buy’, ses-itiPF vs sed-e-itiIMP ‘to sit’).

The examples in (55) illustrate the typical situation, where multiple aspectual pairs involve the same stem. In such verbal chains, it is both true that perfective verbs are derived from imperfective ones (a, b) and vice versa (c).

(55) 6 verbs sharing the stem kop
  a. kop-a-tiIMP ‘dig’                        is-kop-a-tiPF ‘dig up’
  b. is-kop-av-a-tiIMP ‘dig up continuously’ po-is-kop-av-a-tiPF ‘dig up all’
  c. za-kop-av-a-tiIMP ‘bury’                za-kop-a-tiPF ‘bury’

Finally, the pairs in (55) represent one of the possible ways of dividing the ‘chain’ into ‘pairs’. Derivationally, four verbs from (55) could have been used to construct two other pairs.

(56) Two alternative pairs
  a. is-kop-av-a-tiIMP ‘dig up continuously’ is-kop-a-tiPF ‘dig up’
  b. kop-a-tiIMP ‘dig’                        za-kop-a-tiPF ‘bury’

Importantly, from the perspective of synchronic grammar, there is no reason to favour any of the analyses into pairs. The aspectual ‘pairs’ then appear to be a construct which emerges from the comparison of Slavic with the Indo-European languages which do not mark aspect in this way.

Finally, the examples in (55) and (56) also give an illustration of the fact that aspectual pairs, rather than being minimal pairs with a single distinct feature, aspect, display other semantic distinctions as well. It then comes as no surprise that it is impossible to identify a single default procedure for either perfectivisation or imperfectivisation.
Although virtually all Serbo-Croatian verbs have aspect, biaspectuality was well-attested in the language before the current contact with English. Contemporary Serbo-Croatian has and has had long before the contact with English started:

(a) a very limited class of biaspectual native verbs (e.g. *videti* ‘to see’, *čuti* ‘to hear’, *većerati* ‘to have dinner’), and
(b) a numerous class of biaspectual borrowed verbs (from Greek, German, French, etc.) derived using the following suffixes:

• Greek-origin –ISA–: *fotografisati* ‘photograph’, *definisati* ‘define’, *rezervisati* ‘reserve’, *tolerisati* ‘tolerate’,
• German-origin –IRA–: *delegirati* ‘delegate’, *diplomirati* ‘receive a bachelor’s degree’, *formirati* ‘form’, *parkirati* ‘park’, *distancirati* ‘distance’,
• native –OVA–: *protestovati* ‘protest’, *kandidovati* ‘to put forward as a candidate’, *kritikovati* ‘criticise’.

The examples in (b) show that at least 3 suffixes were employed for integration of borrowed verbs in the contact(s) preceding that with English. Based on this picture, we may expect a competition between these suffixes in the present-day contact with English. However, our elicitation data from an experiment in which borrowing is pitted against creative derivation within Serbo-Croatian (see Chapter 10), an existing glossary of new anglicisms (Vasić et al. 2001), as well as Google searches, show a very stable integration pattern in the present contact with English. First, new verbs take the same suffix: –OVA– in Eastern Serbo-Croatian (roughly equivalent to what is presently referred to as ‘Bosnian’, ‘Montenegrin’ and ‘Serbian’) and –A– in Western Serbo-Croatian (roughly equivalent to what is presently referred to as ‘Croatian’). In (57), a few examples are provided.

(57) anzip(ov)ati ‘unzip’, apdejt(ov)ati ‘update’, ben(ov)ati ‘ban’,
    daunlod(ov)ati ‘download’, ed(ov)ati ‘add’, pejst(ov)ati ‘paste’,
    pouk(ov)ati ‘poke’, post(ov)ati ‘post’, restart(ov)ati ‘restart’,
    riplaj(ov)ati ‘reply’, sejv(ov)ati ‘save’

Second, morphosyntactically, verbs newly borrowed from English are always biaspectual unless there is a semantic reason to have a single aspect (for instance, *surf(ov)ati* IMP, which always implies duration).

A comparison of the integration pattern for English verbs to the other patterns which display signs of productivity in Serbo-Croatian shows that both biaspectuality and the consistent use of the same suffix are endemic to this integration pattern. In order to illustrate this, in (58) we give a handful of verbs originating from the same contact, but derived from English-origin nouns and adjectives.
(58) a. verbs derived from borrowed English-origin nouns: šopingovati IMP ‘shop’ (from šopingN), sitingovati IMP ‘to sit’ (from sitingsN), keširati IMP – is-keširati IMP ‘cash, pay’ (from kešN), masterirati IMP – is-masterirati IMP ‘to obtain a master’s degree’ (from masterN), printati IMP – is-printati IMP ‘print’ (from printsN), kliknuti IMP ‘click’ (from klikN), guglati IMP – is-guglati IMP (from GuglN).

b. verbs derived from English-origin adjectives: kulirati IMP – iskulirati IMP (from kulA), fejkovati IMP – isfejkovati IMP (from fejkA).

The examples in (58) show that in the case of verbs derived from English nouns and adjectives, a variety of suffixes are employed and the new verbs always have a single aspect. So the question is now: How can we account for this discrepancy between verbs derived from verbs and those derived from non-verbs?

The first step in accounting for the difference between the two types of lexical enrichment is acknowledging that they emerge under different circumstances. While the verbs based on English verbs are truly borrowed, the formation of the verbs derived from nouns and adjectives took place after the corresponding noun or adjective had been borrowed. This claim is confirmed by our elicitation data (presented in Chapter 10), where Serbo-Croatian speakers were asked to make nonce verbs out of English verbs and Serbo-Croatian nouns. In the case of nouns, the speakers generally produced verbs with a single aspect, using a wide variety of suffixes, whereas in the case of English verbs, they display the uniform pattern described above. This being so, the verbs in (58) are not different from any other verbs derived within the Serbo-Croatian lexicon.

When these data are viewed from the perspective of Lexical Conservatism, an explanation emerges. In the case of English verbs, the initial surface form corresponds to a single incoming verbal stem (for instance refresh), which is not yet present in the Serbo-Croatian lexicon and which can either be split into two lexical items in the process of integration or left as is. It is Lexical Conservatism that prefers the unity of the original lexical item, preserved by the formation of a biaspectual verb. In sum, given the choice between a single lexical item and a pair (or even a chain) of lexical items, LC chooses the more conservative solution, thereby keeping the interaction of the new stem with the native morphology at a minimum. This seems especially beneficial since the examples in (58) show that the most probable perfectivising pattern would involve the prefix iz-, probably yielding forms such as *izrifrešovati PERF, with both prefixation and suffixation, making the original stem harder to recover.

None of this applies to verb derivations within the Serbo-Croatian lexicon. When a new verb is made from a noun or adjective, new lexical items of the verbal kind have to be formed anyway. Given the fact that aspect does
not have an unmarked value (or exponent), lexical entries with a single verbal aspect are formed depending on the semantic/pragmatic context.

The difference between the two types of derivation can be expressed in terms of what is present and what needs to be added. The situation of a noun or adjective serving as the stem of a verb (for instance, \textit{masterirati}_{\textsc{verb}}) differs from that of an English verb turning into a Serbo-Croatian verb (for instance, \textit{rifrešovati}) in two important ways. First, whereas in the case of a nominal or adjectival base, there is a representation in the Serbo-Croatian lexicon, but no verbal stem, in the case of an English verb there is no representation in the Serbo-Croatian lexicon, but the incoming stem is already verbal. By virtue of already being a verbal item but having no aspect specification, the borrowed verbal stem will automatically turn into a biaspectual verb in order to avoid splitting into two lexical items. In the case where for the purpose of a particular verb phrase (which already has an aspect) a new item has to be created based on existing noun/adjective, there is no particular advantage in creating a biaspectual item. Second, the semantics of the borrowed verb is always determined by the semantics of the English original (and therefore usually biaspectual). Whereas the semantics of the nonce verb derived from the noun/adjective is created in the context of a specific VP and may easily have an aspect and trigger the use one of the meaningful prefixes/suffixes.

The important consequence of this analysis is that it allows us to conceptually separate the derivation taking place within the lexicon from the integration of foreign material. As we have seen, these two processes have quite different outcomes in Serbo-Croatian and only one of them – borrowing – displays a unified solution for all stems. Moreover, the patterns described here show another case where the current approach makes a clear distinction between the borrowing mechanism and what is encoded in the synchronic grammar – the introduction of the biaspectual verb is a contact phenomenon which involves foreign verbs which, naturally, are not specified for Serbo-Croatian aspect. Note that in our analysis there is no reason to stipulate a stratum of foreign biaspectual verbs, although biaspectuality is ‘peripheral’ in Serbo-Croatian and almost always coincides with foreignness. This is because given the architecture of the Serbo-Croatian lexicon – most importantly the absence of a ‘regular’ (im)perfectivisation pattern – members of ‘aspectual pairs’ are separate but related lexical entries. In other words, there is no need to include the LC-guided lexicalisation process into the synchronic grammar.
5.2.2. English plural forms as stems in Serbo-Croatian

As we have seen in the cases so far, the input for loanword integration is typically a morphologically simplex base allomorph to which LC constraints refer. This form is typically bare in the SL, i.e. it involves a null morpheme – a morpheme which is realised as a phonologically null affix. However, there are cases in which the bare form (from the perspective of the SL) is not available as an independent form and, as a consequence, a complex form (again, from the perspective of SL) is imported. In some cases, this leads to the integration of an entire complex form as the new RL stem. Serbo-Croatian, for instance, has integrated the English plurals *cornflakes* and *chips* as singular forms *kornfleks* and *čips* with the plural forms *kornfleksi* and *čipovičipsevi*, respectively.

Note that the emergence of paradigms of this kind does not falsify one of the central assumptions our theory makes – that the initial surface form is void of morphosyntactic features. While the English plural morpheme is indeed ‘smuggled’ into Serbo-Croatian, the initial surface form is still crucially morphologically opaque from the perspective of RL. The scenario which our model truly excludes is the one where the plural morpheme appears only in the plural forms (e.g. the unattested paradigm *čip* *čipsevi*).

While initial forms such as *kornfleks* and *čips* can be interpreted as singular forms due to the fact that mass nouns are typically singular in Serbo-Croatian, there are cases in which semantic/pragmatic factors enforce the plural interpretation of the initial form. One such case, also attested in Japanese (Ito & Mester 2006) is the integration of the names of sports clubs and music bands. In such cases, Serbo-Croatian attaches the native plural suffix *es*, to the plural form from English (59). The backformations which mean ‘a single member of the club/band’ keep the whole initial form, for instance, *Bitls* ‘a Beatle’, as illustrated in (60).

(59) Bitlsi   ‘The Beatles’
     Kilersi   ‘The Killers’
     Pepersi   ‘The (Red Hot Chilly) Peppers’
     Lejkersi  ‘The (LA) Lakers’

(60) Video sam Bitls-e/Kilers-e/Pepers-e/Lejkers-e u Beogradu. Ali jedan Bitls/Kilers/Pepers/Lejkers se nije pojavio.
     ‘I saw the Beatlers/Killers/Peppers/Lakers in Belgrade. But one Beatle/Killer/Pepper/Laker did not show up.’
formation of especially LC-friendly paradigms, which, unlike all the other incoming nouns from English, display only one prosodic allomorph of the stem.

Serbo-Croatian has the general integration pattern, which covers all the other loanwords from the present contact with English, to add the augment -out-ov- to monosyllabic bases of nouns in all plural forms (e.g. *teg* - *tegovi* ‘tag(s)’, *skejt* - *skejtovi* ‘skate(s)’, *bedž* - *bedževi* ‘badge(s)’, etc.). The augment always surfaces with a high tone (H), which leads to all plural forms having rising accents (i.e. a tonal span over two syllables), unlike the singular forms, which have a falling accent (i.e. only one H). This is illustrated in (61) using three case forms of both singular and plural of the nouns *tenk* ‘tank’ and *hard* ‘hard disc’.

(Capitals are used for all the syllables which have a High tone. The traditional falling accent is the High on a single syllable, whereas the rising tone is a tonal span over two adjacent syllables. Doubled vowels stand for long vowels.)

(61)  

<table>
<thead>
<tr>
<th></th>
<th>Sg</th>
<th>Pl</th>
<th>Sg</th>
<th>Pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td>tEnk</td>
<td>tEnkOvi</td>
<td>hAArd</td>
<td>hAArdOvi</td>
</tr>
<tr>
<td>Gen</td>
<td>tEnka</td>
<td>tEnkOOvaa</td>
<td>hAArda</td>
<td>hAArdOOvaa</td>
</tr>
<tr>
<td>Dat</td>
<td>tEnku</td>
<td>tEnkOvima</td>
<td>hAArdu</td>
<td>hAArdOvima</td>
</tr>
</tbody>
</table>

The only monosyllabic stems from the present contact with English which avoid the addition of the augment are band/team names which have a monosyllabic stem. We illustrate theses on the three case forms of the nouns *Gans* ‘a member of Guns N’ Roses’ and *Stons* ‘a member of the Rolling Stones’.

(62)  

<table>
<thead>
<tr>
<th></th>
<th>Sg</th>
<th>Pl</th>
<th>Sg</th>
<th>Pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td>gAns</td>
<td>gAnsi</td>
<td>stOOns</td>
<td>stOOnsi</td>
</tr>
<tr>
<td>Gen</td>
<td>gAnsa</td>
<td>gAnsa</td>
<td>stOOnsa</td>
<td>stOOnsaa</td>
</tr>
<tr>
<td>Dat</td>
<td>gAnsu</td>
<td>gAnsimaa</td>
<td>stOOnsu</td>
<td>stOOnsima</td>
</tr>
</tbody>
</table>

The pattern into which *gAns* and *stOOns* are integrated is surprisingly conservative compared to other loanwords: the same falling accent is maintained throughout the paradigm and no length differences appear in the paradigm. The pattern above is also more conservative than what we encounter in the small class of native monosyllables in which do not add the augment in plural forms (e.g. *GRk* ‘Greek’ and *dAAn* ‘day’), as these nouns always have at least one more allomorph in their paradigm (cf. the genitive plural forms *gRRkaa* and *dAAnAA*).
Recall that, since all other monosyllabic nouns from the current contact with English get the toned augment -o-/-e- in plural, they can never have the same accent in all forms. The question is now why only team and band names are allowed to have the extremely conservative paradigms.

Again, Lexical Conservatism can help us make sense of these facts. In the case of a ‘regular’ monosyllabic noun with the singular interpretation (e.g. *hArd), the initial form will be assigned a falling tone, since there is only one syllable. All the other singular case forms will follow this pattern (e.g. hArd, hArd, hArdo) all perfectly satisfying LC. In the plural forms, then, LC(accent) is violated (hArdOvi, hArdOOva, hArdOvima), presumably due to the fact that there are no (native or borrowed) nouns with the shape *hArdovi. This seems to be tolerable for LC because there are many prominent forms (i.e. all the singular forms) which replicate the accent of the initial surface form.

In nouns in which the initial surface form first becomes the stem of the plural form (for instance, stOOns) the situation is quite different. If the augment were to get applied to stOOns, the new NomPl form would be *StOOnsOvi and the falling accent of the initial surface form would disappear from the emergent paradigm. So in this case LC intervenes to save the accent of the initial form in at least some form, and this leads to the formation of the perfectly regular paradigm of the type represented in (62). In sum, the main difference between the two types of nouns (hard and Stons) lies in the fact that the latter, unlike the former goes through a phase in which it only has plural forms and it is important to LC to preserve the accent pattern of the initial surface form at this stage.

The generalisation emerging from these data is that LC can enforce quite unusual morphological integration in order to avoid total loss of surface features, in the sense that they are not preserved anywhere in the paradigm. This is not surprising, given a functional interpretation of LC – that of recoverability. Thus in each case it is important to observe the size of the paradigm and how invasive the available affixes are for the surface properties of the initial form.

In the following subsection we turn to a case where SL truly provides a pair of allomorphs which will present an apparent counterexample to our claim that the initial form is a single base allomorph.

5.2.3. Italian verbal augment in Maltese: Lexical Conservatism Integrated

In this section a case is presented where, instead of a single allomorph, a pair of allomorphs from SL participate in the formation of the new RL paradigm. This appears to be an obvious counterexample to our claim that LC, as an anti-allomorphy mechanism, guides morphological integration.
However, as our analysis will reveal, it is actually Lexical Conservatism that prefers borrowing both allomorphs in this case.

One class of Italian verbs ending in -ire, illustrated by the verb suggerire in (63) below, has the augment -isc-, which is added to its stem (e.g. sugger-) in all the forms where the stress would otherwise fall on the stem. So, unlike the verbs which lack the augment (e.g. sentire), the verbs with the augment have no allomorph with the stressed stem (cf. the first person singular present tense forms suggerisco and sénto). This phenomenon has been termed “metrical consistency” by DiFabio (1990). In our terms, the distribution of the augment in the SL paradigm is guided by Lexical Conservatism constraints which keep the stem stressless.

Maltese has elaborate patterns for integrating Italian verbs. Quite spectacularly, in the case of the verbs of the type suggerire Maltese has not only borrowed the pair of allomorphs, issuğger(i) and issuğgerixx (x standing for [j]) but also the distribution which avoids the stressed stem. As Hoberman & Aronoff (2003) state, word stress in Maltese is phonologically conditioned, “falling on the final syllable if it is extraheavy (CVCC or CV:C), otherwise on the penultimate syllable” (p. 75). This means that (in the concatenative sector of the Maltese verbal morphology) stress in principle falls on the verbal stem in all forms where the suffixes are monosyllabic. These are exactly the contexts in which the augment is invoked. Considering the paradigms in (63), we see that, for instance, the present/imperfect form for the first person singular nissuggerixxi is actually a way of avoiding the form with the stressed stem, *nissuggeréri, just as its Italian counterpart suggerisco avoids the form *súggero/*suggéro.

(63)  | Italian | Maltese |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perfect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sg 1</td>
<td>sentű</td>
<td>suggerũ</td>
</tr>
<tr>
<td>sg 2</td>
<td>sentísti</td>
<td>suggerísti</td>
</tr>
<tr>
<td>sg 3m</td>
<td>sentí</td>
<td>suggerí</td>
</tr>
<tr>
<td>sg 3f</td>
<td>sentí</td>
<td>suggerí</td>
</tr>
<tr>
<td>pl 1</td>
<td>sentímmo</td>
<td>suggerímmo</td>
</tr>
<tr>
<td>pl 2</td>
<td>sentístete</td>
<td>suggerístete</td>
</tr>
<tr>
<td>pl 3</td>
<td>sentírono</td>
<td>suggerírono</td>
</tr>
</tbody>
</table>
Here again what at first sight appears to be an exception to the principles of our model, which crucially assumes that a single form is introduced into the new language, not whole paradigms or pieces of paradigms, turns out to be a case of extreme Lexical Conservatism. Maltese borrowed an ordered pair of allomorphs (implemented in the sense of Kager 2008), which is still far less than the whole paradigm, and just enough to maintain the existence of a single version of the stem – the unstressed one. For instance, the Italian [isk]/[i] allomorphy is not borrowed, and neither are any paradigm members, as that would not have improved the situation LC-wise.

In this section we have seen that the fact that loanwords lack morphological and morphosyntactic features leads to seemingly uninspected assignment of paradigms. Importantly, these paradigms are new and unexpected only with respect to what we find within RL. This illustrates a way in which our theory diverges from the theories in loanword behaviour we have discussed in chapters 2 and 3. As Kang (2011) observes, most research is based on the claim that loanword integration can be considered “a real-life Wug test (Berko 1958) which can enable us to probe into the grammatical knowledge of speakers in ways that native data alone cannot” (p. 2258). As confirmed by all the cases considered here, the words entering the lexicon are crucially foreign material. For this reason, the type of the incoming material is constrained in an important way: it does not take its morphosyntactic information along.

5.3. The initial form is a surface form

In this section the second important feature of the initial form is discussed: the fact that it is a relentlessly surface form, which, even though it is void of SL morphosyntactic information, often contains various cues for parsing in terms of RL morphological and morphosyntactic properties. In §5.3.1. we will
see a case in which, due to LC and native ineffability, the initial surface form gets 'reincarnated' and becomes an allomorph of the stem. In §5.3.2. we deal with cases in which morphological complexity is assigned to incoming words based on generalisations which have held in the language prior to language contact, thereby producing new types of complex words.

5.3.1. Re-emergence of the initial surface form in the integration of ō-final items into Serbo-Croatian
The morphological integration of foreign nouns ending in a short unstressed ō displays quite some variation and occasional ineffability. This is due to the fact that ō is the desinence of quite a few declension classes, none of which provides a safe destination for all nouns which end in short unstressed ō. The foreign nouns which in which the final ō is long (metrō) or stressed (rokokó) always lexicalise the whole initial surface form as the new stem and add suffixes to it. As the only open declension with the null desinence is the masculine one, all these nouns become masculine. This integration is clearly expected based on Lexical Conservatism: long and stressed vowels are more salient and therefore introducing an allomorph without them is violating more LC constraints than in the comparable situation where a short unstressed vowel is parsed away.

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Voc</td>
<td>métrō-Ø</td>
<td>metró-i</td>
</tr>
<tr>
<td>Gen</td>
<td>metró-a</td>
<td>metró-a</td>
</tr>
<tr>
<td>Acc</td>
<td>métrō-Ø</td>
<td>metró-e</td>
</tr>
<tr>
<td>Dat/Loc</td>
<td>metró-u</td>
<td>metró-ima</td>
</tr>
<tr>
<td>Ins</td>
<td>metró-om</td>
<td>metró-ima</td>
</tr>
</tbody>
</table>

This pattern, with a stem ending in a vowel still only occurs in borrowings in Serbo-Croatian and it typically does not extend to words in -ō. The only noun we have been able to identify in which this is common is loto 'lotto' (although here also, the author has a strong preference for considering -ō a desinence).

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Voc</td>
<td>loto-Ø</td>
<td>loto-i</td>
</tr>
<tr>
<td>Gen</td>
<td>loto-a</td>
<td>loto-a</td>
</tr>
<tr>
<td>Acc</td>
<td>loto-Ø</td>
<td>loto-e</td>
</tr>
<tr>
<td>Dat/Loc</td>
<td>loto-u</td>
<td>loto-ima</td>
</tr>
<tr>
<td>Ins</td>
<td>loto-om</td>
<td>loto-ima</td>
</tr>
</tbody>
</table>
The remaining nouns tend to have their short unstressed final o reanalysed as a desinence. There are two paradigms which these items seem to model their forms on. One is the masculine paradigm in (66), which hosts virtually only names and has no plural forms. Note that the unavailability of the plural forms is not an artifact of being a name, since all the other names pluralise (e.g. Mark-ovi 'Mark.NomPl').

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Voc</td>
<td>Mark-o</td>
<td>/</td>
</tr>
<tr>
<td>Gen/Acc</td>
<td>Mark-a</td>
<td>/</td>
</tr>
<tr>
<td>Dat/Loc</td>
<td>Mark-u</td>
<td>/</td>
</tr>
<tr>
<td>Ins</td>
<td>Mark-om</td>
<td>/</td>
</tr>
</tbody>
</table>

The other declension which plays a role in the integration pattern is the neuter declension illustrated in (67) by the noun zlato 'gold'. While this class may seem ideal for accommodating items with a short unstressed -ő, its singular is actually closed to loanwords. This may be related to the fact that this class still displays heavy phonotactic restrictions on what can precede the desinence -ő (and the Instrumental ending -őm), so that it is unable to host items in which the -ő is preceded by a palatal consonant (e.g. poncho). In plural, however, this declension has exactly the same endings as the complementary declension for palatal stems, exemplified by polje 'field' in (68), so that the same plural endings apply to all stems.

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Acc/Voc</td>
<td>zlat-o</td>
<td>zlat-a</td>
</tr>
<tr>
<td>Gen</td>
<td>zlat-a</td>
<td>zlat-a</td>
</tr>
<tr>
<td>Dat/Loc</td>
<td>zlat-u</td>
<td>zlat-ima</td>
</tr>
<tr>
<td>Ins</td>
<td>zlat-om</td>
<td>zlat-ima</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Acc/Voc</td>
<td>polj-e</td>
<td>polj-a</td>
</tr>
<tr>
<td>Gen</td>
<td>polj-a</td>
<td>polj-a</td>
</tr>
<tr>
<td>Dat/Loc</td>
<td>polj-u</td>
<td>polj-ima</td>
</tr>
<tr>
<td>Ins</td>
<td>polj-em</td>
<td>polj-ima</td>
</tr>
</tbody>
</table>

This strange combination of available paradigms leads to the unexpected situation that, in most nouns with an inanimate referent the singular form is masculine whereas the plural form is neuter.
This type of declension introduces the first inanimate items in desinence -ő into the masculine declension. Note that the fact that the accusative case has the same form as the nominative case (unlike in the declension of Marko) is due to the general masculine distinction between animate (Acc=Gen) and inanimate (Acc=Nom) syncretisms.

The nouns with an animate referent have both the singular and plural forms according to the masculine declension.

These forms are also new, since there are no native masculine nouns in -o with a full paradigm. As can be verified from a comparison with the paradigms in (65), the plural endings are those of the mainstream masculine declensions.

Now, this picture is additionally complicated by the fact that two endings which figure in the plural forms – the masculine NOMPL ending -i and the masculine/neuter DATINSLOCPL ending -ima – obligatorily change the stem-final velars k, g and h into c (=ʃ), z and s. This leads to occasional re-emergence of the initial surface form in some plural forms, a loanword-specific phenomenon.

The generalisation that the endings -i and -ima obligatorily change the velars has virtually no exceptions in Serbo-Croatian. For instance, the new English borrowings of the type marketing automatically get marketinizzi as the NomPl and marketinzima in DatInsLocPl. However, this alternation also has no confirmation in the native declensions with the -o desinence. The type Marko has no plural forms, also not when the stem ends in a non-velar (e.g. Danilo). The type zlato has two very frequent nouns whose stem ends in a velar: mleko ‘milk’ and blago ‘treasure’. While the ending does not apply (the NomPl forms are mleka and blaga), the speakers report total ineffability of the DatInsLocPl forms of these nouns. They do report having computed
the forms *mlekima/mlecima* and *blagima/blazima*, but none of the forms are considered ‘good’. In sum, there are no velar-final native stems which could serve as a model for incoming *o*-nouns. This leads to massive variation within which two innovative strategies are noticed: one is the reintroduction of the initial surface form, while the other amounts to introducing a new allomorph which avoids the alternation. Both will be illustrated by the results of a small-scale data collection experiment. In the first experiment 76 native speakers were asked to choose between the offered case forms of five nouns of the type *ego* (velar-final stem) and five of the type *geto*. The tested forms were the LOC/SG (representative for the singular declension, where there are no suffixes which cause alternation), the NOM/PL (where the non-alternating –a should prevail) and the DAT/INSLOC/PL. The results for these two nouns (in numbers of tokens) are displayed in (71).

(71)
The most surprising fact is that while there is no sign of the stem allomorphs [ego] and [geto] in the singular forms, they become by far the most common solution in the case where the alternation cannot be avoided: the DatInsLocPl form of the type ego. However, this type of allomorph seems to affect the other cells of the nouns of the type ego, and, more surprisingly, even the nouns of the type geto.

The emergence of the allomorphs [ego] and [geto] is clearly an alternation-avoiding pattern and in that sense certainly LC-guided, but it is also entirely loanword-specific. The native nouns mleko ‘milk’ and blago ‘treasure’ do not have the option of resorting to the stem allomorphs [mleko] and [blago] and get the DatInsLocPl forms mlekoima and blagoima because the strings [mleko] and [blago] have no ontological status other than concatenations of stems and desinences. This is not the case in words of the type ego and geto in which these strings still have the status of the initial surface form, which can re-emerge to save the form from ineffability.

5.3.2. Pseudo-complex words in Finnish and Serbo-Croatian

In this section cases from Finnish and Serbo-Croatian are discussed in which morphological integration is triggered by prosodic cues. As we shall see, it will remain an open question whether in these cases LC is sacrificed due to a strong surface generalisation in the native lexicon, because it is strictly speaking not true that new allomorphs of the stem are produced. This discussion will shed some light on the issue of the assignment of structure to loanwords due to structural cues they carry.

Kiparsky (2003) shows that in Finnish some polysyllabic borrowed nouns have lexical secondary stress in addition to the general primary initial stress. These nouns, termed fixed polysyllables by Kiparsky, have a fixed stress pattern, and contrast with other polysyllables, termed movable polysyllables, which have rhythmic secondary stress. As Kiparsky states:

“Movable polysyllables have a rhythmic secondary stress which oscillates between the third or fourth syllable, normally according to the weight of those syllables, in line with the LH effect. Fixed polysyllables have a lexical secondary stress which is invariant on a given syllable of the stem. Although the nominative singulars of movable and fixed stems have the

---

17 The LH effect is explained in the following way. “Finnish stress is assigned by laying down binary feet from left to right. Final syllables are not stressed if they are light, and only optionally if they are heavy. An important phenomenon is the LH EFFECT: when the left-to-right scansion encounters a Light-Heavy sequence, the light syllable is skipped, with the result that a ternary foot is formed.” (p.111)
same output stress pattern, their inflected forms (Inessive and Ablative
Singulars, in these examples) diverge as follows.” (p. 113)

(72) Movable stress (L'H effect in inflection) (judgements from Kiparsky)
   a. Káleva  ?*Kálevalassa Kálevalassa  'Kalevala'
b. Ámerikka  ?*Ámerikassa Ámerikasssa  'America'
c. ártikkeli ?  ?*ártikkielissa ártikkielissa  'article'
d. ápteekkari  ?*ápteekkarilla ápteekkarilla  'pharmacist'

(73) Fixed stress (no L'H effect) (judgements from Kiparsky)
   a. Álabàma  Álabàmassa  ?*Álabamàssa  'Alabama'
b. pálsternàkka  pálsternàkassa  ?*pálsternakkàssa  'parsnip'
c. ésplanàdi  ésplanàdilla  ?*ésplanadìlla  'esplanade'

Kiparsky also shows that nouns with a lexical secondary stress allow for two
different vowel harmony patterns: one which is based on the whole word and
another which is otherwise encountered only in native compounds, in which
the harmony domain starts from the syllable with the lexical secondary
stress. For example, a word like pyramid-i can either be treated as a whole
root with mixed vowels, in which case the back variant of the elative suffix is
used: pyramìdi-ista, but it can also be analysed as a compound, in which case
the ending harmonised with its “second part” and the front ending is
selected: pyramid-ìstä.

As Kiparsky puts it, “the double stress merely invites the compound
reanalysis, it does not force it. Fixed polysyllables can very well be treated as
single prosodic words, as their harmonic variation confirms” (p. 115).

(74) The two vowel harmony patterns (judgements form Kiparsky)
   a. Lexical stress often initiates a vowel harmony domain:
      árkkitèhdila  árkkitèhdiltä  ‘architect’ (Ablat.Sg.)
      bólševikilla  bólševikillä  ‘bolshevik’ (Adess.Sg.)
      káramèlleja  káramèllejä  ‘candy’ (Prt.Pl.)
      pyramidista  pyramìdistä  ‘pyramid’ (Elat.Sg.)
   b. Rhythmic stress very seldom initiates a vowel harmony domain:
      ártikkeli  ?*ártikkeilä  ‘article’ (Adess.Sg.)
      prófessorèita  ?*professorèitä  ‘professor’ (Prt.Pl.)
      strátegisèsta  ?*strategisèstä  ‘strategic’ (Elat.Sg.)
      kálenterèsta  ?*kälenteristä  ‘calendar’ (Elat.Sg.)
Kiparsky concludes that words which allow for the compound pattern, also termed quasi-compounds, are “compounds from the prosodic point of view, though not from the morphological point of view” (p. 115).

Since there are no native words with a lexical secondary stress, and all the words in (74) are loans, it is safe to conclude that Finnish had developed an analysis which takes non-rhythmic secondary stress as a cue for the initial syllable of the second compound member. In that sense, it is exactly (inflectional) morphology that takes the “prosodic point of view, though not [...] the morphological point of view”. In other words, inflectional morphology has access to prosodic cues, but not to morphosyntactic structure. This apparent contradiction can be resolved if we reconsider the architecture of grammar, taking the loanword trajectory seriously and divorcing loanword treatment from grammar. As crucially surface forms, loanwords must interact with inflectional morphology in order to develop paradigms. In a language such as pre-contact Finnish, where compounds are prosodically cued, inflectional morphology never needs to take the semantic/syntactic structure of compounds into account: it finds all the necessary cues in the surface structure. However, loanwords, which bring their surface prosody from the source language, subvert this handshake between prosody and inflectional morphology and constitute a class of quasi-compounds. Importantly, the quasi-compound pattern is in variation with a new pattern, with lexical secondary stress which does not initiate a vowel harmony domain.

Serbo-Croatian has a pattern similar to Finnish. As mentioned in §5.2.2., Serbo-Croatian adds the augment -ov/-ev- to all monosyllabic bases of nouns in all plural forms (e.g. tegovi ‘tags’, skejtovi ‘skates’, bedževi ‘badges’ etc.) Serbo-Croatian also lacks nouns with final stress, except for a limited number of loanwords, which all satisfy a prosodic condition – they are all longer than two syllables and end in -nt (e.g. elemént, mastodónt). The only native polysyllables which have final stress are clear compounds which are usually spelled with a hyphen. These compounds have plural forms with the augment, implying that the second member of the compound is pluralised. (Nouns ending in -nt nouns have no augment e.g. plurals eleménti, mastodónti)

(75) áuto-pút ‘highway’ áuto-pút-ev-i ‘highways’
áuto-gól ‘own goal’ áuto-gól-ov-i ‘own goals’
bék-áp/bék-ap ‘back-up’ bék-áp-ov-i ‘backups’
Recent English loanwords preserve the position of stress of the source form, so that there is an influx of new nouns with final stress. All these nouns receive the augment in their plural forms.

(76) ripér-ov-i ‘repairs’ invájt-ov-i ‘invites’
ivént-ov-i ‘events’ ripláj-ev-i ‘replies’

Again the prosodic cue apparently makes inflectional morphology start a new prosodic domain, although semantically and syntactically there is none. The result of integration crucially depends on the fact that loanwords come in as surface forms which carry ‘misleading cues’ for morphosyntactic analysis. An important issue in the cases of the type under consideration is the status of the patterns of what used to be compound patterns before the influx of loanwords. As Kiparsky put it, the loanwords in these cases become “compounds from the prosodic point of view” and there is indeed no clear evidence that their subparts become stems of any kind. It is therefore important not to jump to the conclusion that this kind of parsing is malevolent to LC just because there is some kind of analysis of the original initial surface form. The question to ask is “What else could have happened?” For instance, it could be the case that in words of the type ripér ‘repair’, the options were either to modify the stress pattern in order to make the word declinable according to the polysyllabic pattern (*ríper-i) or to start the domain from the stressed syllable. If this is the case, parsing is actually in the service of LC: it helps the preservation of the stress pattern, without really introducing any new allomorphs of the stem.

5.4. The initial form has a (specific) meaning
As we have seen in Chapter 4, incoming loanwords have a specific meaning. In this section we are looking at the dynamics of a clear case of borrowing of morphologically complex words, which are analysed as complex in the recipient language – a process which the present model considers more marked than borrowing of simplex stems. This follows from two generalisations about the incoming initial form we have seen in Chapter 4: that it is simplex and that it has specific meaning. We show that borrowing of complex forms is indeed constrained in interesting ways under the present analysis.

5.4.1. Latinate abstract nouns in Serbo-Croatian (and Dutch)
Many Slavic and Germanic languages have developed a Latinate stratum of the lexicon which hosts numerous derivations, some of which clearly synchronically analysed as such. Serbo-Croatian has Latinate abstract nouns
derived from Latinate adjectives (or related to adjectives). The existence of a large class of such nouns seems especially puzzling, since Serbo-Croatian has a productive suffix -ost (roughly corresponding to English -ness) which can be attached to virtually any adjective to create an abstract noun.

A semantic analysis of the pairs of ost-nouns and the Latinate ones (Chapter 11) shows that only the ost-nouns, such as elokventnost, have the meaning transparently derived from that of the adjective, and that the ost-nouns are preferred in the contexts of the type ‘his eloquence was praised’, where the noun has the trope meaning, i.e. the semantics of ‘he is eloquent’. On the other hand, the Latinate nouns, such as elokvencija, usually have a much less transparent meaning and their semantics has to be lexically encoded.

(77)

<table>
<thead>
<tr>
<th>Adjective</th>
<th>ost-noun</th>
<th>Latinate noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>homoseksualni</td>
<td>homoseksualnost</td>
<td>homoseksualizam</td>
</tr>
<tr>
<td>elokventni</td>
<td>elokventnost</td>
<td>elokovencija</td>
</tr>
<tr>
<td>sterilni</td>
<td>sterilnost</td>
<td>sterilitet</td>
</tr>
</tbody>
</table>

In sum, even when it seems that a language is borrowing a word for a concept which can be covered by native derivation – in this case deadjectival abstract noun – it is more often the case that foreign complex words are not compositionally derived. So whereas homoseksualnost has the transparent semantics ‘X is homosexual’, homoseksualizam refers to all the other culturally conditioned meanings (the social phenomenon, the subculture, the movement, etc.). This also means that although there is no doubt that nouns such as elokvencija are related to adjectives such as elokventan, the meaning of the suffix -(c)ija is not comparable to that of -ost, since only the latter has straightforward semantics and produces transparent derivations. Importantly, as we will argue in Chapter 11, the many nouns in -izam, -cija, -itet and other Latinate suffixes have most probably re-established their link to the related adjectives only after both have been borrowed as simplex words with a specific meaning. These data clearly show that we need a more fine-grained view of morphological complexity and a clear distinction between productive and transparent derivations (which will be analysed in Chapter 11 as paradigm members) and non-transparent patterns and unproductive patterns. The fact that complex words sometimes get borrowed and even establish derivational patterns within the recipient language should not be taken as evidence that they derive identically to the native complex words.

A similar asymmetry can be observed in Dutch. Dutch also has Latinate nominalisations, which are much more widespread than those in Serbo-
Croatian. Booij (1999: 75) notes that the suffixes -heid (native) and -iteit (Latinate) are synonymous, but show an interesting asymmetry in distribution: whereas -iteit only combines with non-native bases, -heid can attach to both. His examples are *ab'surdheid and absurdi’teit (the base being *ab'surd) for non-native bases and ge’woonheid and *gewoni’teit (from ge’woon ‘ordinary’) for native bases. Interestingly, *ab'surdheid, although dispreferred by most speakers, will surface in contexts in which Serbo-Croatian has the noun with a trope meaning.

(78) de duidelijke absurdheid/*absurditeit van die regel
‘the clear absurdness/absurdity of that rule’

Also, for many other combinations in which heid-nominalisation is held impossible and absent from dictionaries, it actually surfaces in trope-contexts, in which the generally accepted nominalisation is blocked.

(79) De seksueelheid/*seksualiteit van hun relatie staat buiten kijf.
‘The sexualness/sexuality of their relationship is beyond dispute.’
Hij moet per se zijn speciaalheid/*specialiteit bewijzen.
‘He has to prove his specialness at any price.’

In sum, the facts which appear at first sight to be as a massive emergence of marked cases for our model (borrowing of morphologically complex items) turn out to support it: what appears to be productive morphology within the Latinate stratum has rather idiosyncratic semantics and works crucially differently from the native derivational morphology. Borrowing remains the domain of transfer of simplex units with specific semantics and the complexity emerges only in later analysis.

5.5. Conclusions of the chapter and further development

In this chapter we have considered various aspects the of morphosyntactic integration of loanwords, which is viewed as the assignment of morphosyntactic features to the initial surface form, which results in the formation of a full-fledged RL lexical item. As we have seen, this does not mean that all assignment of morphosyntactic features takes place in order to make a lexical item: loanwords undergo quite some assignment due to the fact that some of their features get interpreted in terms of RL surface generalisations.
During the process of assignment, Lexical Conservatism is active and aims to preserve as many features of the surface form as possible. Throughout this chapter we have seen that Lexical Conservatism interacts with all three main features of the initial surface form: its being void of morphosyntactic features (§5.2.), its surface nature (§5.3.) and its specific meaning (§5.4.). In each case, we have seen the benefits of considering loanword trajectory as conceptually distinct from RL grammar, which helped us spot patterns which are crucially different from what happens within RL.

Now that we have seen quite a few of these patterns we are in the position to see why loanword patterns are loanword-specific and not easily comparable to patterns within RL: if observed from the perspective of RL, all loanword processes are *diachronic*. Each process, if observed within the confines of RL, ushers a new word into the RL lexicon and when the process reaches its end there is no need for this trajectory to be memorised by the users of the new RL items. In (80) we give a preliminary picture of model exemplified on the integration of the word *whisky* into Serbo-Croatian.

(80)

Crucially, after the new Serbo-Croatian paradigm has been established, the speakers do not need to know the history of paradigm formation in order to use the word *viski*. In this sense loanword trajectory is rather similar to diachronic processes encountered within languages, such as morphological reanalysis guided by analogy, increase in paradigm uniformity, grammaticalisation, etc. However, there is a crucial difference which involves the part of the picture which we have been disregarding so far, as indicated by the question mark in (80). In cases of internal diachronic change, the
speakers have one version of a lexical item (or a class of lexical items) at one point in time, and then at some later point they have another version of the same item. In such cases, speakers are typically not in the position (or in need) to compare the two forms and extract the pattern which transforms the one form into the other. In the case of borrowing, however, the speakers who can be considered agents of all the processes discussed in this chapter can be reasonably assumed to be exposed to many examples of already borrowed items, which puts them in the position to extract the borrowing pattern from existing loanword. This means that all the processes we have seen so far have another dimension which remains unnoticed if the discussion is kept within the confines of RL (which is the reason for the question mark in the picture in (80)). This dimension is the knowledge of the conventionalised correspondences between SL and RL categories which gets accumulated in the bilingual community and is supported by every new borrowed word. It is crucial for our model that this dimension is not defined by being available to another type of speakers: although so far we were concerned with what happens upon the arrival of the initial surface form in RL, this should not be interpreted as the monolingual perspective, since everything discussed in this and the previous chapter is performed by the speakers who introduce borrowings – broadly defined bilinguals.

In the following chapter we turn to this other, conventionalised dimension of language contact: the inter-language mappings.
Chapter 6

Inter-language mappings

In this chapter, a final aspect of the present loanword model is discussed – the inter-language mappings, the correspondences between units of languages in contact internalised by bilinguals. The inter-language mappings will be argued not to belong to any of the languages in contact, but to the contact itself and will therefore be characterised as contact-specific. Moreover, these mappings will be argued to be highly conventionalised, in the sense that, rather than integrating every loanword as if it was the first one, bilingual speakers are always aware of the existing stock of loanwords and the generalisations which can be extracted from this stock.

6.1. Introduction

Until here, we have been zooming in on the recipient language, without paying much attention to the relation between SL and RL. We have done so by focusing on the loanword trajectory, which amounts to focusing on a single word going from being a code-switch, all the way to becoming an integrated loan. In order to address the inter-language mappings, we will now focus exactly on what was remaining in the background so far: the knowledge internalised by the bilingual community, which masters both languages and picks up on the fact that structures are integrated in a uniform way across words. In (81), we are repeating the graphic representation of our model in order to introduce a discussion on its scope and the relation between its parts.

So far, we have been focusing on the fact that a word, once it has emerged in another language, needs to receive all the morphosyntactic properties required in order to obtain the status of an RL lexical item. It is quite natural that this was the perspective from which we first observed the loanword data, since it is the one that always applies. Even if a single lexical item is taken from one language to another, it will need to go through some kind of loanword trajectory. However, loanword data as discussed by linguists never feature a single item or a handful of accidental borrowings: loanword data virtually always consist of well-established mappings between SL and RL categories which readily apply to not yet borrowed items. In this sense, loanword data actually always presuppose the existence of inter-language mappings.
Applied to the example in (81), this means that until now we have been focusing on the fact that the word *whisky* has been integrated into Serbo-Croatian and that its integration has followed the loanword trajectory: at some point in time there was no word *viski* in Serbo-Croatian and at a later point there was. This implied the creation of a paradigm under the supervision of Lexical Conservatism. Now our focus is shifting to the fact that the English word *whisky* instantiates many structures which are always mapped to the same Serbo-Croatian structure across different words. For instance the English [w] always corresponds with the Serbo-Croatian [v], the English [ɪ] always corresponds to the Serbo-Croatian [i] and all nouns which end in an –i (and do not have a human female referent) join this declension (recall §5.1.). We will argue that these mappings between RL and SL structures are highly conventionalised and cannot be predicted based on RL grammar. So, for instance, during its contact with Turkish, which ended in the 19th century, Serbo-Croatian had a totally different, yet equally consistent integration strategy for items ending in –i: for instance *cimri* ‘miser’ and *deli* ‘devotee’ was adapted as *džimrija* and *delija* and joined the feminine declension. In sum, so far we have been focusing on the fact that each word enters RL at a certain moment in time and passes the loanword trajectory, but in this chapter we will be focusing on the fact that in stable language contact incoming words generally instantiate familiar structures and these structures are treated in the same way across many words. In this sense, we could say that while traces of LC are visible in the way integration takes place, in stable language contact most words could be said to take a free ride on the existing integration pattern fixed in inter-language mappings. In this sense, while it still may have been possible that for a word like *whisky* some other integration options were considered by some Serbo-
Croatian speakers, the 2014 borrowing *selfie* has certainly taken a free ride on the well-established integration pattern: it is hard to imagine a Serbo-Croatian speaker who was in the position to integrate *selfie* without first having established the generalisation based on the words like *sherry, brandy, martini, floppy, proxy, Harry, Billy* and many others. It is therefore safe to claim that the integration of *selfie* has been influenced by LC only in the sense that an LC effect had already been invested into the inter-language mapping.

This brief discussion hopefully helps reveal an important property of our model: its two main ingredients – LC and inter-language mappings – also introduce two complementary perspectives on loanword processes. And while there is no doubt that LC effects and the creation of inter-language mappings are concomitant in every language contact, LC and inter-language mappings cannot be observed in action simultaneously. In order to make the this two-perspective structure of the model comprehensible before turning to a detailed discussion of inter-language mappings, we will first make a brief intermezzo and look at some of the limits of a theory of loanwords.

### 6.1.1. What can a loanword theory be about?

In our theorising of the loanword behaviour so far, the main proposed innovation was removing the grammar bias – the tendency to observe changes in loanwords as an instantiation of grammar by definition – from our approach to loanwords. This can be seen as going against the grain of the research tradition reviewed in Kang (2011), that claims that loanword integration can be considered “a real-life Wug test (Berko 1958) which can enable us to probe into the grammatical knowledge of speakers in ways that native data alone cannot” (p. 2258). Our critique of this type of comparisons hinges on the fact that in a Wug test a speaker of a language is supposed to believe that a nonce word is an actual word of her language, so she can apply her native (and monolingual) grammar to it. In other words, it is crucial for the speaker to work with the form without activating the two other elements of the semiotic triangle (Clark 1998, Croft 2009) – the meaning (because the speaker only learns about the corresponding concept from the experiment) and the community in which the meaning is conventional (because there is none). In other words, the Wug test is suitable for isolating the formal aspect of the linguistic competence. With loanwords this option is excluded already at the first step. The bilingual speakers who introduce a new word are aware of its meaning in SL (minimally the specific meaning with which the word is introduced into RL) and the use of loanwords is always socially interpretable, that is, it always defines some kind of community. Given the fact that loanwords are relentlessly meaningful, both in the sense of
Chapter 6

semantic/conceptual content and in the sense of defining a community, it is fair to ask: Can formal linguistics, which has oftentimes been accused of the “imperialism of the signifier” (Lecercle 2002), interact with loanword integration in an insightful way? In other words, shouldn’t removing the grammar bias from loanword research lead to a separation between loanword research and formal linguistics?

We believe that the answer is that formal linguistics actually needs to be involved, since there is a part of the loanword integration process in which the formal side of the phenomena needs to be recognised and considered. This part is the ‘bridge’ between two lexicons which has been discussed in the literature on code switching (recall also §4.1.): there is no transfer of whole lexical entries, typically the only thing that can be transferred is a single pronounceable string with a specific meaning which then serves as the basis for a new lexical entry in the recipient lexicon. In our graphic representation, only the shaded forms are the ones which actually participate in the transfer and the modifications which the SL form undergoes are determined by inter-language mappings. As we have seen throughout this dissertation, there is ample evidence that whenever linguistic material is transferred from one code to another, the ‘bridge’ gets activated and the recurrent structures get modified in the same way across many tokens, typically without reference to social or any other type of meaning. This aspect of the process indeed allows us to maintain the imperialism of the signifier. The loanword model, which will be completed in this chapter, is a theory of the ‘bridge’.

However, the formal approach to the ‘bridge’ has to remain restricted in both an empirical and theoretical sense. There are two important empirical restrictions, both implicit to what has been said on loanwords so far and determinant for the empirical scope of this dissertation. First, there can only be loanword integration if there is actual transfer of linguistic material. This restricts our focus to what has been termed “replication of linguistic matter” (Matras 2009) and excludes cases where only the structure or meaning has been transferred (e.g., calques, semantic shifts, but also phonological shift under foreign influence). For instance, we will have nothing to say about the fact that many European languages have a calqued correspondent of the English word *skyscraper* – for instance, Dutch *wolkenkrabber* of Serbo-Croatian *neboder*, or about the contact-induced spreading of the uvular realisations of the *r* in Western and Northern Europe. Second, in order for transfer to take place, there need to be two distinguishable codes with separate lexicons. As many examples in Matras (2009) show, there are cases in which the distinction between the two codes is blurred, so that what appears to be permanent code-mixing is actually a code in its own right,
perceived as code-mixing only by linguists who know the history. Such cases seem to show different phenomena as well. A famous instance of whole paradigms from two different languages being apparently present within the same code is Romani in Bulgaria, in which many Turkish verbs are conjugated using exclusively Turkish affixes (p. 182-184). However, the boundary between the two languages might be something of a linguists’ construct. As Matras reports, “When asked in the majority language, Bulgarian, to answer a question in ciganski (i.e. ‘the Gypsy language’), consultants often responded in Turkish, rather than in Romani” (p. 184). For these speakers, there is no real distinction between Turkish and Romani items, since both are part of the in-group code opposed to the out-group code of Bulgarian. In such cases, it is hard to claim that there is a strict separation between the Romani and Turkish lexicon, which would be a necessary condition for speaking of a transfer.

As for the theoretical side, the formal approach to loanwords has to make two important distinctions when it comes to the incorporation of loanword facts into the linguistic theory. The first one is between the two perspectives featured by our model, as discussed above. The first is the perspective of what we have been calling the loanword trajectory: the integration of a single item by a single speaker, guided by the restrictions such as Lexical Conservatism and the Matrix Language Principle. The other aspect concerns generalised and conventionalised integration patterns, which arise within a community and can be expressed in the form of inter-language mappings. In both cases formal approaches have the task of making predictions as to what a possible loanword trajectory/mapping is. Second, in the wake of our effort to liberate loanword research from the grammar bias, the task emerges of situating the patterns that we find productively applying in loanword treatment with respect to grammar. For this purpose, a careful distinction needs to be made between the processes which can be expressed in formal categories comparable to those used for language-internal surface generalisations (phonemes, prosodic units, and so on) and those processes which can be directly attributed to grammar. As will be shown in this chapter, inter-language mappings are often arbitrary from the perspective of the synchronic grammar of either language. In this sense, there is no motivation for taking these mappings to be part of the Chomskyan I-language in any sense other than possibly using the abstract/formal representations.

This last point requires some further clarification. In order to illustrate processes whose description requires some abstract units which would not be taken to be grammatical by most linguists, we will give an example of a language-internal surface generalisation leading to a productive pattern.
Recently, there has been a minor social-network hype in Serbia caused by the *Polumenta generator* (available from http://polumenta.zardina.org/), named after the oriental music singers Šako (Sakib) and Dado (Damir) Polumenta. Both names display the pattern which in Serbia gets perceived as typically Bosnian/Herzegovinian/Montenegrin: a CVVC stem + desinence -o, and the rising accent on the first syllable, so šAAkO, dAAdO, but also mUUjO (from Muhamed), hAAsO (from Hasan), sUUljO (from Sulejman). *Polumenta generator* generates random CVVC strings and concatenates them with the family name Polumenta, implying that the name on the screen is a possible name of a member of the Polumenta family. This pattern was recognised by the speakers and produced a comic effect. In order to describe this phenomenon, we need to use quite a few abstract units: consonant, vowel, vowel length, rising accent, desinence, yet the pattern can hardly be considered grammatical, in the sense that not many linguists would agree with the existence of a *Polumenta morpheme*. As we will show later in this chapter, inter-language mappings seem to be based on surface generalisation and therefore display a similar level of abstractness.

The study of inter-language mappings is the place where the formal approach can profit from a strategic alliance with usage-based approaches – since mappings tend to emerge and be maintained by occurring frequently and being noticed and adopted by borrowers, formal approaches can profit greatly from the usage-based idea of entrenchment (see Backus & Verschik 2012) without committing to the usage-based view of language as a whole. In fact, it is a general advantage of divorcing loanword treatment from grammar that loanword research is now able to use the virtues of various linguistic theories without committing to their views on grammar.

In this chapter, the separation of loanword patterns from those attributable to RL grammar will be completed, enabling us to position the two with respect to each other. As already mentioned, proceeding will require shifting the perspective from which loanword facts are viewed. A crucial aspect of the data presented in this dissertation, which is often taken for granted (also in our discussion so far), but will be foregrounded here, is the striking regularity with which the same structure gets integrated across different words. In our model presented so far, there has been nothing that enforced this uniformity across words. What this kind of presentation of data seems to imply is that there is something ‘predestined’ about the outcomes of language contact, which can be read off either the combination of the two grammars or the RL grammar. As will be shown below, this is not the case at any stage of language contact: while the initial phase of language contact shows an abundance of integration patterns, typically one of them survives and becomes conventional. Moreover, in the course of language contact, it
does not only get negotiated how certain structures will get adopted, but also what will be a borrowable unit, both in the sense that a universal concept of what is a word is far from self-evident, and in the sense that not all SL elements are (equally) borrowable or covered by inter-language mappings.

In sum, this chapter completes the process of divorcing loanword research from grammar by using the concept of inter-language mappings, a non-grammatical tool par excellence. Inter-language mappings are necessary to account for the uniformity with which the same structures are treated in the same way across words, and for the productive application of the same generalisation to new words. But unlike (OT) grammars, integration patterns are not equipped to deal with unfamiliar structures as their input. There is no reasonable application of concept of the Richness of the Base to the domain of the inter-language mappings. Inter-language mappings are mere generalisations extracted from the existing data by bilingual speakers and as such they only cover the structures which are encountered in the actual languages in contact. The synchronic nature of inter-language mappings is in the focus of the next section.

6.2. The synchronic side: Inter-language mappings cover what LC misses

So far all the processes discussed in Chapters 4 and 5, as well as in this chapter, have been viewed as LC-guided lexicalisation, for which we used the metaphor of the loanword trajectory. Such a perspective emphasises the diachronic, once-only side of the process. Now, in order to be able to position our model of loanword behaviour with respect to generative grammar, which is inherently synchronic, we need to make the final step. We need to ask the question: What is internalised in speakers? Is there a ‘competence of the language mixer’? There is a clear intuition in language contact researchers that experienced borrowers/code-switchers simply “know what to do” with a certain structure (Pieter Muysken p.c.). This has been expressed already by one of the founders of language contact research, Einar Haugen: “When languages are in contact, there is a strong tendency for speakers to equate items of the one with items of the other. Some items thereby acquire associations different from those they have for monolinguals without such experience” (Haugen 1956: 44-45). Inter-language mappings are extracted out of the speakers’ experience with related words. Once the mappings are in place, they are productively applied to any words which fit their description, so that speakers sometimes do not perceive the difference between repeating an existing borrowing and introducing a new one. For instance, Hualde (2006) says, about the facts we shall see in §6.3.3.: “A Basque speaker who
uses a word [...] will not necessarily know whether s/he is repeating a Basque word which s/he has learned of whether s/he has in fact adapted it spontaneously from his/her Spanish lexicon.” (p.353).

We will call this stage of language contact, where a considerable number of words support inter-language mappings and make certain words potentially borrowed, the state of language entanglement. As in every entanglement, this does not prevent us from saying anything about the properties of the entangled elements (languages) or their position with respect to each other (Basque structures probably are not as integratable into Spanish), it just prevents us from considering them as totally separate objects with a clear boundary. From the perspective of the proposed model as a whole, we are now in a position to show how LC effects of the kinds presented in the previous chapters get propagated by becoming part of inter-language mappings. LC, as we have seen it in Chapter 5, generally requires the preservation of all the properties of the initial form in the process of lexicalisation. Crucially, LC pays no attention to structures available in the recipient language or to the processes imposed by the grammar of the recipient language: it just selects its favourite among the options offered in the lexicalisation process. In other words, in cases of nonce lexicalisation (when a single word gets introduced into a language on a single occasion), there are many ways in which various representations and factors can interact and yield different results for different speakers. However, since words usually come together and are usually introduced by a bilingual community, the effects of LC are mediated and propagated by the mappings which become conventionalised over time. Once a mapping has been established, the LC constraints which are inherently violated by the mapping will not play any palpable role in the lexicalisation processes (just like the universal OT constraint *CODA can appear not to play any role in the languages which allow codas if it is ranked very low), but it will re-emerge as soon as there is a change in the mappings (e.g. due to a new source language). In terms of Kiparsky (2006), the present model can be called amphichronic, with LC as its only universal. As Kiparsky states, “universals should constrain any kind of change, and they should define ‘attractors’ on which a variety of diachronic paths converge” (p. 221). Again, this does not mean that we have built yet another amphichronic model of grammar, which is then applied to loanwords, as will be clear from the discussion of mappings below.
6.2. What is a mapping?

A mapping is a conventionalised correspondence between elements of different languages, established on a contact-specific basis. Since mappings are generalisations, the existence of a mapping always enforces productivity: it always pushes the pattern to be used on new items. In other words, the existence of a mapping makes sets in the source lexicon in principle borrowable and, as we have seen, possibly-already-borrowed. In this section, we shall address the two features of inter-language mappings broached above: the uniformity across words and the ability to fix a certain implementation of the notion of a borrowable unit. Finally, we will briefly turn to another type of inter-language generalisations which seem to make use of the same mechanism.

6.2.1. Some familiar mappings

Since a certain amount of regularity is a necessary precondition for the data to be even visible to linguists, all the data sets we have seen so far are also instantiations of mappings. As we shall see, some of the mappings will be complete instructions for the formation of a lexical entry, whereas others will be more context-free, referring to phonological structures, for instance. For instance, the English→Serbo-Croatian mapping for the integration of verbs, whereby *download* becomes *daunloud(ov)ati* (recall §5.2.1.) is shown in (82). This mapping has a lexical item as its output.

(82) English→Serbo-Croatian for verbs

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_v</td>
<td>X(ov)atiV.BIASPECTUAL (and the whole paradigm)</td>
</tr>
</tbody>
</table>

The same general mapping can be part of lexicalisation mappings with different levels of specificity in different contacts. As we shown in §5.3.3., stem-final gemination is conventionalised in the borrowing from English to American Italian (for instance *coal* becomes *[kol.le]*).

(83) English→American Italian

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>-VC</td>
<td>w</td>
</tr>
</tbody>
</table>

As we shall see in Chapter 8, the mirror-image phenomenon, stem-initial gemination, attested in Maltese, is only generalised in verbs because this is the only context in which LC profits from this relation. Below we are giving
some examples of other mappings illustrated by data sets in Chapter 5. The mapping in (84) is what takes the Beatles to Bitlsi in Serbo-Croatian.

(84) English→Serbo-Croatian plural nouns

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>XsNOUN</td>
<td>XsİNOUN.NOMINATIVE.PLURAL (and the whole paradigm)</td>
</tr>
</tbody>
</table>

Note that the input form is not specified as plural. The assumption here is that the extra-linguistics factors (the speakers’ awareness of the fact that the referent is plural) are the ones which enforce the plural interpretation of the initial surface form. Finally, in (85), we see the mapping which takes the Italian verb suggerire to the Maltese issuğgerejt, nissuğgerixxi, etc.

(85) Italian→Maltese

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXireVERB</td>
<td>iCiCiXejt1stPERSON.SINGULAR.PERFECT, niCiCiXixxi1stPERSON.SINGULAR.IMPERFECT (and the rest of the paradigm)</td>
</tr>
</tbody>
</table>

Now we have seen illustrations of some mappings, they can be compared to LC. Not surprisingly, the mappings contain exactly what LC constraints abstract over. LC is concerned with the exponents of the incoming morpheme and it looks at how many allomorphs are created and by which features they differ from the initial surface form. LC is not interested in the recipient language categories. These are covered by the inter-language mappings, which are aimed at giving as unequivocal instructions for integration as possible.

Already this brief comparison between LC and inter-language mappings makes it clear that the very concept of mappings necessitates explicit statements about the nature of transferable units and their status in the contact. As the next section will show, how “linguistic matter” gets transferred is for a big part established in the specific contact situation rather than being predetermined in some universal way.

### 6.2.2. Units of transfer

In this section we reconsider the notion of borrowable unit in language entanglement by looking at various verbal mappings.

First, there are cases in which words can be isolated, but nevertheless mappings cannot be established because of the different structure of lexical entries. Hebrew is a well-known example of a language which has no productive mappings for English verbs because the lexical representations
are not compatible – whereas in English vowels and consonants are all part of the lexical representation, in Hebrew vowels are part of the inflection and consonants are restricted both with respect to their number and combinations. Second, in certain cases, the word status is unclear. One such case is that of the English complex verbs. Even for speakers of languages which have stable mappings for English, such as Dutch and Serbo-Croatian, verbs such as *set out*, *get round* and *hit up* are hard to integrate into L1 sentences. In Serbo-Croatian the options are to either make the preposition a part of the new stem (*setaut(ov)ati*) or integrate only the verb (*set(ov)ati* out). However, both are seen as very hard to process. In Dutch, there is a complicated network of options, also related to the fact that Dutch does have separable verbs, but with different positions of elements, and that there are many mutually related representations in the two languages. So, for instance, *uitzetten* ‘to turn off’ consists of the cогnate subparts of *set out*. There are indications, however, that this area is uncharted by inter-language mappings and that speakers avoid it. De Decker & Vandekerckhove (2013) find, in a heavily code-switched corpus of chat messages produced by Flemish adolescents in their late teens, with 29,3% of all code switches being verbs, that there is no single loan complex verb integrated into Dutch clauses. Tellingly, the existing loan verb *checken* is reported to cover for the meaning ‘to check someone/something out’ thereby avoiding the direct integration of the verb *to check out*. This is illustrated in the sentence in (86).

(86) Voor meer info, check onze website.
“For more info, check out our website.”

Staying with verbal mappings, in a third type of cases, the word status is at stake. In Turkish spoken by the immigrant population in the Netherlands, verbs are generally adapted using the light verb construction with the verb *yapmak* which is combined with the infinitive. So for instance *studerren* *yaptım* ‘I studied’ is used where the inflected Dutch form would be *ik studeerdelik heb gestudeerd*. Furthermore, this often means that a whole VP is taken from Dutch. In (87) we show an example from Backus (2010). Here, the whole phrase, which refers to a culture-specific concept, is integrated into a Turkish sentence. Importantly, the speakers have sufficient command to be aware of the fact that these are separate words, perhaps even borrow them as such.
‘I’m going to live on my own.’

So the mapping for Turkish speakers in the Netherlands is that shown below.

(88) Dutch to Turkish

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>XVP/V</td>
<td>X yapmak (and the whole paradigm)</td>
</tr>
</tbody>
</table>

While complex verbs with the light verb yapmak are common in Turkish, there are no constructions where yapmak is complemented with a verb. This construction is an artefact of language contact. The lesson to draw from this example is that linguistic matter is always transferred in units which are established during contact. What is a borrowable unit is an issue negotiated in the bilingual community during the process of establishing mappings.

Another telling case is that of two large Dutch-Moroccan Arabic corpora from 1986 (Utrecht, collected by Mustapha Lkoundi) and 1991 (Nijmegen/Tilburg, collected by Youssef Azghari). While in one the light verb construction was conventionalised, the other one barely had any verb incorporations of this type (Jacomine Nortier, p.c.). This result need not imply that there are differences in either Dutch or Moroccan Arabic as spoken in either of the places, which may have led to this variation. What is more probable is that the two communities conventionalised different outcomes from the original pool of possibilities.

In this section, we have seen what mappings are: contact-specific, conventionalised correspondences between units of different languages. This might sound like giving a theoretical status to parochial generalisations, in the sense that inter-language mappings might appear as a very specific mechanism, present only in very fluent bilinguals, and very different in every community. However, there are many comparable inter-language mechanisms which are often glossed over due to the monolingual bias in linguistics – the general tendency to be interested in the competence of native speaker, not the knowledge shared by communities of ‘native switchers/borrowers’. For this reason, in the next section we will look at two comparable phenomena which are not inter-language mappings, but use the same underlying mechanisms. This will show how mappings are actually part of a very general mechanism of applying surface generalisations.
6.2.3. What is not a mapping (but runs on the same fuel)
The fact that speakers have linked representations between words in
different systems can hardly be surprising. The fact that speakers are able to
understand other registers, accents, dialects and related languages implies
some knowledge of the other variety (Benders 2011, Castella & Simonović
2012). This is a little-theorised aspect of the linguistic reality, but no one
appears to claim that speakers are unaware of this type of internalised
knowledge. There are, for instance, probably no speakers of standard Serbian
who are not aware of the conversions necessary to derive the related words in
standard Croatian. The same applies to regional and national varieties of say
Dutch (for instance, Belgian and Dutch), German (for instance, Austrian and
German) or Spanish (for instance, European and Colombian): speakers have
internalised knowledge of how others speak. Moreover, in §2.2.3. we have
seen some valuable arguments presented in Janda et al. (1994) that some
inter-language effects are best analysed as a special case of hypercorrection.
In this section, we show a case of overgeneralisations by L2 learners, which
is very similar to the inter-language mappings we have seen above. However,
a crucial difference between the two will be shown, with the reservation that
it might well be the case that this difference is due to cultural constructions
around the concept of language.

This section features a phenomenon in which Dutch and Serbo-Croatian
speakers act in similar ways. Both linguistic communities have clear
mappings for integrating English verbs into their language (except for the
phrasal verbs, as we have seen): the verbal inflection for Dutch (collapsen,
feeden, raisen) and the ‘biaspectualiser’ -o(va)- for Serbo-Croatian
(kopleps(ov)ati, fid(ov)ati, rejz(ov)ati). The three languages Dutch, Serbo-
Croatian and English, have interrelated Latinate verbs, between which
generalisations of different kinds obtain. So, both Dutch and Serbo-Croatian
speakers can generalise from the forms in (89) that the verbs ending in -eren
and -irati generally end in -ate in English.

(89) S-C Dutch English
iritirati irriteren irritate
rezonirati resoneren resonate
komplicirati compliceren complicate

That L2 speakers of English (with both Serbo-Croatian and Dutch as L1)
indeed do so is testified by the pseudo-English forms in (90).
Importantly, speakers may import these forms from their English into their L1, and in that case they would be following the well-established mappings (for instance, distantiaten in Dutch and distanšijejt(ov)ati in Serbo-Croatian).

The lesson we can draw from this is that although mappings between structures in different languages are numerous and active, at least in the cases of languages with clear boundaries and monolingual communities as the three above, these mappings are highly specialised and unidirectional. In this situation, introducing a new word into a language, which involves inter-language mappings and the creation of a new lexical item, is crucially different from guessing an L2 form. This distinction might be much blurrier in a situation where there are no national standard languages with monolingual communities and a clear concept of native speakers.

Having shown what mappings are (and aren’t) we will now turn to the most important feature of the inter-language mappings: their being conventionalised rather than ‘natural’.

### 6.3. Mappings are conventional

Recalling Chapters 2 and 3, most approaches to loanwords are making an attempt to reduce loanword treatment to a special case of grammar. As a consequence, most processes are seen as ‘natural’ in the sense of being guided by Markedness and other grounded mechanisms and acquired during the critical period. Our approach is quite different. Mappings between loanwords are not ‘naturally’ learned in early childhood and they need not be based on speech (which is generally true of processes within single languages). The fact that categories connected by mappings are generally similar in some way is seen as a consequence of the fact that mappings come into existence in the real world, but not as a necessary feature of inter-language mappings. In order to create a coherent model of inter-language mappings, it is important to disturb the idyllic picture in which all the outcomes are traceable back to a single source. For this reason, in this section, we are looking at three cases which show that sources of mappings should be sought at places different than those where grammatical processes originate.
6.3.1. Netherlandic vs. Belgian Dutch
A prominent difference between Netherlandic and Belgian Dutch (Flemish) is the way the English [æ] in words such as *badminton*, *laptop*, *stand* and *rap* is adapted.

\[
\begin{array}{lll}
\text{Belgian} & \text{Netherlandic} & \text{ pronunciation-based adaptation} \\
\text{b[æ]dminton} & \text{b[ɛ]dminton} & \text{’badminton’} \\
\text{l[æ]ptop} & \text{l[ɛ]ptop} & \text{’laptop’} \\
\text{st[æ]nd} & \text{st[ɛ]nd} & \text{’stand’} \\
\text{r[æ]ppen} & \text{r[ɛ]ppen} & \text{’to rap’}
\end{array}
\]

The native phonologies seem not to point in this direction in any clear way. Additionally, in L2 English the most common value for [æ] is [ɛ] in both varieties. Furthermore, Taalnet (2000) is a normative intervention which clearly states that [ɛ] would be the pronunciation-based adaptation. In sum, in most approaches, the Belgian values would be explained away as extra-linguistically conditioned spelling pronunciations, or possibly influences from French. The Netherlandic values, on the other hand, would be seen as pronunciation-based and therefore more representative. Not so for our approach: while all these causations are interesting in their own right, they are not full-fledged accounts: they actually only help us see which representations we have to take into account. However, they should not lead to jumping to conclusions about which speakers are exposed to which aspect of English. For both varieties, the input is not simply [æ], it is [æ] plus the graphemic value <a>, and this input is linked to the output [ɑ] and [ɛ] in Belgian and Netherlandic Dutch, respectively. In this situation, the Belgian speakers do not simply base their outcome on the spelling, they also take the pronunciation into account and lexicalise ‘against’ it. The same is true of Dutch speakers, who are well aware of the spelling. In sum, the differences are not based on the speakers knowing different things: it is simply the case that the communities have converged on different forms.

It is by now clear that there is nothing inherently preferable about mappings based on the pronunciation, the representation of mappings does not suffer from their different source. The mapping is then as shown in (92).
6.3.2. Syllabic \( r \) in English borrowings into Serbo-Croatian

The English vowel attested in words like *bird*, *word* and *fur* is typically mapped to the Serbo-Croatian syllabic \( r \) by the latter language’s speakers. This is a very prominent feature of the Serbo-Croatian accent in English and of spontaneous nativisations. It is also very obvious from the Serbo-Croatianised and phoneticised English emblematic and common in colloquial Serbo-Croatian. In the following example, taken from a blog, this accent is being ridiculed. The underlined letters are realisations of the vocalic \( r \).

These examples show that there is a clear awareness of the mapping from the English vowel to the syllabic \( r \) in Serbo-Croatian speakers. In standard Serbo-Croatian, however, there are no examples of syllabic \( r \) in English loans and the words containing this vowel are adopted with \( [Vr] \) sequences, whereby the vowel is usually taken from the graphical form e.g. *Berkeley*, *Berns*/*Burns* (Burns). Serbo-Croatian speakers seem to be aware of this aspect of the standard language. When asked to nativise the words *workshop* and *feature* the way they use them in a text message and the way they would use them in an academic article, they choose *vrkšop*/*fičr* in the former and *vorkšop*/fičer in the latter case. It is important to note that this

---

18 Source at
http://blog.b92.net/text/6039/%22WRAJTUJ%20KAO%20SHTO%20SPIKUJESH%2C%20RIDUJ%20KO%20SHTO%20JE%20WRAJTNUTO%22/
procedure is not entirely conscious in the speakers – they are not able to formulate a generalisation, and that integration strategies are not taught at any level of formal education. However, the speakers are able to support their choice with other examples of standard adaptations.

This example shows that even if the speakers have different intuitions about certain correspondences (which then get invested into substandard nativisations), they can acquire a prescribed pattern and use it actively. This also shows that speakers are able to access mappings as a whole and can deal with different mappings for different varieties of the same language.

Here again, Serbo-Croatian speakers pick up on the available patterns and apply them in further integrations, rather than developing their own intuitions on optimal correspondents in two languages. The available patterns can, as shown in this case, originate in standardising interventions and make crucial use of written form, far from the speakers’ internalised grammars of the two languages.

6.3.3. Spanish-Basque inter-language mappings
Hualde (2006) presents surprising correspondences which hold between Spanish nouns and their counterparts borrowed from Spanish into Basque. For instance, nouns which end in the sequence -on generally have -oi in their Basque correspondents.

(94)  | Spanish    | Basque    |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>camión</td>
<td>kamioi</td>
<td>‘truck’</td>
</tr>
<tr>
<td>electrón</td>
<td>elektroi</td>
<td>‘electron’</td>
</tr>
</tbody>
</table>

This outcome is surprising because very common words in Basque end in the ‘repaired’ Spanish sequence, for instance: gizon ‘man’ and on ‘good’. The correspondence has a historical explanation: the original Vulgar Latin sequence -one had originally been borrowed faithfully and consequently had different destinies in both languages (-one>-on in Spanish and -one>-oni>-oi in Basque), yet the mapping between the new forms has remained.

In sum, mappings do not have to be motivated in synchronic terms, and what we see on the output side are not strictly adaptations: mappings do not necessarily make foreign items more native-like. This does not take away from the fact that these conventionalised mappings have to refer to (meta)linguistically accessible units – they just do not have to relate them in a way which would seem phonetically natural. An interesting aspect is that the sequences in question (on and oi) are not morphemes in these words. However, equally arbitrary morpheme-to-morpheme correspondences can block this general correspondence. So the words which end in the Spanish
-ción/-sión get borrowed with a final -zio/-sio in Basque. Here as well it may be a matter of discussion whether the units -ción/-sión and -zio/-sio are real morphemes in all the words or just sequences related by mappings.

(95)  Spanish    Basque
      edición   edizio  ‘edition’
      conclusión  konklusio  ‘conclusion’

In sum, in this section we have seen that, depending on the available mappings, words of certain types (word classes, words containing certain structures) have the status of potentially already borrowed once their counterpart in the recipient lexicon is entirely predictable. The mappings by which this happens are conventionalised and not necessarily grounded in anything else than practice. As the last example shows, if languages are in contact for a long enough period, the corresponding structures change together with the rest of the language, which then necessarily leads to unnatural mappings. This is an important finding, because it implies that all the abundance of synchronic explanation that we find for most mappings has more to do with the fact that languages have been in contact for a relatively short period than with some universal feature of inter-language mappings.

6.4. Dynamics of inter-language mappings: The LC factor

While we see that the nature of mappings is quite unpredictable based on SL and RL only, the way the mappings develop seems to be under the crucial influence of the other element of our model: LC.

We have seen LC and inter-language mappings in action in various cases of contact, showing how stable integration patterns are established and maintained. Of course, just as languages are permanently in flux, stability in mappings is just an approximation and there is an ongoing reshaping of these patterns. In this section, we are looking at two cases in which different stages in the development of inter-language mappings can be distinguished. As we shall show, the direction of the change can be accounted for using LC. An important typological prediction will follow from the discussion of the data in this section: whenever mappings at different points in history can be compared with respect to the position of LC, LC can only ever be promoted. This is a direct consequence of the synchronic part of the model, more specifically from its general property that the most LC-friendly existing paradigm will host loanwords (recall §5.1.). In other words, once a new LC-
friendly paradigm has been created, all the new loanwords will enter this paradigm. We shall see such a case in §6.4.1., where a case of change in the integration pattern for -o-final nouns in Bulgarian will be discussed. However, already in §8.2., looking at Serbo-Croatian biaspectual verbs, familiar from §5.2.1 we shall see that at different stages of language history, different items are borrowable and that the format of mappings accommodates for the type of input delivered by the source language. As a consequence, we will conclude that comparability of different language contacts remains a relatively 'shady deal'.

6.4.1. Change in the integration pattern: Bulgarian neuters in -o
In Bulgarian, nouns are declined for number and definiteness. The paradigms of two native nouns selo ‘village’ and kopče ‘button’ are represented in (96).

(96)  indefinite  definite  indefinite  definite
      singular  sel-o  sel-o-to  kopč-e  kopč-e-to
      plural    sel-a  sel-a-ta  kopč-e-ta  kopč-e-ta-ta

Borrowed nouns in -o/-e also become neuter. These borrowed items had initially joined one of the common native patterns of plural formation. Specifically, the integration picture in Stojanov (1983) can be summarised as follows:

(a) Borrowed items in -e have the stress-preserving pluralisation in -ta: koliéSG – koliétaPL ‘necklace’, which was, not surprisingly, preferred to all the other options, which have a stress shift (e.g. imeSG – imenáPL).
(b) Since native items in -o have no stress-preserving pattern available and no patterns in which the -o remains unparsed by morphology as the singular suffix, the nouns in -o take the plural ending -a, which replaces the -o and causes a stress shift in disyllabic nouns: kínoSG – kináPL, librétoSG – librétapL, etc.
(c) Out of the two plural formation patterns, the more LC-friendly one – the one in koliéSG – koliétaPL is generalised to items ending in -i and -u, which had no native counterparts: randevu-ta, taksi-ta.

Bulgarian as described in Stojanov (1983) has not introduced any new pluralisation patterns, but recycled existing patterns to accommodate new items. As expected, for each structure, the most LC-friendly paradigm is activated. It is clear from this picture, however, that LC was frustrated by the destiny of the nouns in –o, which, unlike all the other nouns, fail to preserve both their final segment and their stress pattern in the plural form (kínoSG – kináPL).
It will then come as no surprise that there is an ongoing change in the language's integration pattern. Both the online generator http://slovored.com and the speakers which I have consulted confirm that more recent borrowings in -o have generalised the pluralising pattern in -ta. Crucially, this leaves the stress of the singular/initial form intact. Hence all my informants have *espréso-ta*, *bíngo-ta*, *avokádo-ta*, *démo-ta* and *lógo-ta* as the only plurals of these nouns.

So the novel, optimised pattern connects all the words ending in -o, -e, -i and -u to the inflectional type which has a stem-bound stress and -ta as the plural ending. Note that none of my sources has earlier borrowings (e.g. *kino*, *palto*, *bjuro*, *dynamo*, etc.) declined using -ta. These latter words have become native-like by the virtue of joining a native pattern when they entered the language and LC has no way of undoing this. This is due to the once-only nature of LC. Once a word has been lexicalised and the LC-violating allomorphs become listed, LC automatically stops being violated by them.

The conclusion of this discussion is that, as foreshadowed above, we cannot predict when a new pattern will emerge in morphological integration. However, we can predict that once such a pattern has been introduced, all the new words will follow it. As a result of change in the integration pattern, the mappings for integration into Bulgarian have simplified: now all final vowels except -a (which is necessarily analysed as the feminine declension ending, just like in Serbo-Croatian) are desinences for neuter gender and have plurals in -ta.

In (97) we present an overview of the two mapping sets for all nouns. The first two classes, not affected by the change discussed here, comprise items ending in a consonant, which are always masculine (e.g. *elementSG* → *elementiPL*) and items in -a, which always become feminine (e.g. *nokijaSG* → *nokiiPL*).

(97) Bulgarian 1                  Bulgarian 2
    input          output       input          output
    -CNOUN → -øSG, -iPL (masculine) -CNOUN → -øSG, -iPL (mascul.)
    -aNOUN → -aSG, -iPL (feminine) -aNOUN → -aSG, -iPL (fem.)
    -oNOUN → -øSG, -á/aPL (neuter) -oNOUN → -øSG, -taPL (neter)
    -en, -in, -un → -øSG, -taPL (neuter) -en, -on, -in, -un → -øSG, -taPL (n.)

As we can see, the input forms have no morphological sign on them. Typically the word class is determined by the code-switching slot in which the word occurs, but all the remaining properties get assigned. In Bulgarian this means receiving a gender and a paradigm specification. Furthermore,
the mappings in Bulgarian 2 are simpler than those in Bulgarian 1. Importantly, if we look at the constellation in Bulgarian 2, -ta is now covering all the new neuters and goes no further than that. In other words, gender is predictable from the shape of the incoming word and there is one inflectional class open for each gender.

It is also clear that the mappings are much simpler and straightforward than the complex correspondences which can be encountered within the native grammar. In the next section, we shall look at another set of diachronic data where, as we shall see, there are good reasons to assume that a change in the nature of the foreign input has led to the change in mappings.

6.4.1. The integration morpheme for verbs in Serbo-Croatian
Serbo-Croatian bispectral verbs have already been discussed in §5.2.1. In this section we focus on the change in the integration suffix used for foreign verbs. As presented in §5.2.1., borrowed bispectral verbs originating from language contacts before that with English (contacts with Turkish, Greek, German, French, etc.) were integrated using the following suffixes, applied to stems from various languages:

- Greek –ISA–: fotografisati, definisati, rezervisati, tolerisati,
- German –IRA–: delegirati, diplomirati, formirati, parkirati,
- native –OVA–: protestovati, kandidovati, deportovati, prezentovati.

From what we have seen in §5.2.1. it is clear that only –OVA– is now used for the integration of English verbs into Eastern Serbo-Croatian. In Western Serbo-Croatian, the same is true of the suffix –A–. In (98), we give examples of both integration patterns.

(98)  East  West
      daunlodovati  daunlodati  ‘to download’
      spemovati  spemati  ‘to spam’
      šerovati  šerati  ‘to share’

As we can see, there a piece of history here: –IRA– and –ISA– have lost the role of integration morphemes for verbs and this role has been taken over by a single suffix: –OVA– (East) and –A– (West). As we shall see in Chapter 10, both –IRA– and –ISA– are still used frequently when speakers are deriving nonce verbs from existing nouns, verbs and adjectives. Moreover, some of the modern creations from borrowed English nouns and adjectives in Serbo-Croatian still use the ending -irati: masterirati ‘to obtain a master’s degree (introduced after 2000)’ and kulirati ‘to ignore’ or ‘to relax’ (related to kul ‘cool’) are examples of this.
–IRA– seems to have been very strong at the very beginning of contact with English, since a few early loans from English have this suffix.

(99) džogIirAti ‘to jog’, parkIirAti ‘to park’, šutIirAti ‘to kick’, čekIirati ‘to check’

Note that in these cases the source of -irati could not have been German, because German has no -ieren in these cases. The same is true of the re latinised reciklierati ‘to recycle’ (related to ciklus ‘cycle’) and reanalysed overdozirati se ‘take overdose’ (related to the germanism dozirati ‘to dose’). There is even an example of across-the-board use of -irati with a German stem: the German verb föhnen ‘to dry hair with a hairdryer’ (related too Föhn ‘hairdryer’) corresponds to the Serbo-Croatian fenirati (related to fen).

From the perspective of our model, the question is: why have -ova- and -a- been preferred in the end over the well-established -ira- and -isa- (presumably in the early nineties of the twentieth century)? The answer comes from the realm of LC: both -ova- and -a- are typically toneless, whereas both -ira- and -isa- attract stress/tone, thereby erasing the stress of the stem. For instance, džogIirAti and fotografišati can only have this prosodic pattern, whereas ribUUtovati ‘to reboot’ and dAUunlou dovati ‘to download’ can preserve the stress of the source. Clearly, the new suffixes make it possible to preserve more contrast in verbs and that is why they are preferred by LC.

However, before we jump to the conclusion that LC has been promoted once again, it is important to notice that the repertoire offered by the source language also seems to play a role in the change. It is not strictly true that LC(STRESS) was being massively violated in the phase preceding the intense contact with English because there was little or nothing to preserve anyway. Whereas English verbs can be stressed on any syllable (a pattern which becomes additionally unpredictable once all reduced vowels are turned into full Serbo-Croatian vowels), previous main source languages – French and German – had predictable stress patterns: in French stress is always final (and quite a few of them stressed on the very -ire) and most German source verbs ended in the stressed -ieren. So, whereas in the previous contact there was no stress distinction in incoming verbal stems to begin with, in contact with English, there is one, and the integration pattern has accommodated for this fact.

In this case again, we see that the integration mapping is working much more specifically and concretely than (OT) grammar: it refers to
morphologically circumscribed inputs and outputs and it offers solutions for what occurs in the input.

In this section, an excursion has been taken to diachronic facts. The two key concepts of the present model have been shown to interact in a productive way: LC guides the change of inter-language mappings.

**6.5. A consequence: Doing without loanword-specific Markedness**

Now that we have completed the picture by situating the inter-language mappings, we can turn to the relation of our model with respect to the architecture of grammar, especially the architecture envisaged in Optimality Theory. In OT’s Markedness/Faithfulness architecture nothing directly clashes with our model in the sense that our model covers either a diachronic process (the LC part) or a strictly non-grammatical process (the inter-language mapping part). Furthermore, Faithfulness as such never plays a role in loanword integration – it is always preceded by LC constraints. However, general Markedness is needed to filter out the structures in the input which are not tolerated in RL.

Our model predicts that loanword-specific Faithfulness will take over from LC in all cases where LC protected features of the initial surface form which are actively repaired in RL. However, there is no way for loanword-specific Markedness to arise. In principle, all the cases in which loanword-specific markedness may seem at work – the cases where structures are repaired in loanwords, but allowed to surface faithfully in native words – should be analysed as results of inter-language mappings, where due to the existence of a mapping, the integration “misses” a more faithful target (just like in the Basque examples in 6.3.3.). This is why now that we have introduced the inter-language mappings, we revisit an example of proposals which make use of loanword-specific markedness – that presented in Jurgec (2010). As will be shown, Jurgec’ data contain compelling evidence that for a mapping analysis.

**6.5.1. Ljubljana Slovenian mid vowels**

As shown by many examples in Jurgec (2008, 2010, 2012), in Ljubljana Slovenian the RTR mid vowels $\varepsilon$ and $\sigma$ are avoided in borrowing from English and replaced by the ATR counterparts $e$ and $o$ respectively. In native words, on the other hand, both $\varepsilon$ and $\sigma$ are well attested (e.g. in $[\sigma]\check{\varepsilon}$ father’ and $[\varepsilon]n\varepsilon$ ‘one’). Most loanwords go through a brief initial phase when they are used as code switches with $\varepsilon$ and $\sigma$ before they are replaced by the process just alluded to.
Jurgec (2010) also notes that these words change their RTR vowel as soon as they get an overt case ending (DEE, compare our discussion in §4.2.1.), so that the only allowed genitive forms (with the ending -a) for the nouns in (100) are fleša, ekstazija, voka and potkasta. This tendency is very strong in present-day Slovenian adaptations and seems to point to a markedness constraint specific to loanwords. Such an analysis is presented in Jurgec (2010). In this analysis the loanword-specific markedness constraints *ɛ/ɔ are put to use to guide the integration process. Although this account technically works, there are many aspects of the Slovenian facts which point at a different solution.

Jurgec does not mention that in colloquial Slovenian there are many frequent older borrowings from German which do contain ɛ and ɔ. Some examples are in (101), with their standard language correspondents.

All these words are banned from standard Slovenian and for each of them there is an unrelated native correspondent. This need not be a problem for Jurgec’ account, given the fact that these words are not part of the current contact with English and need not be encoded in the lexicon as loanwords anyway. However, the same dichotomy is true of the few words of English origin which preserved their RTR mid vowels even in case forms when declined.
The word *bluf* is the only apparent exception, but it is felt as a loanword, even being spelled as *bluf*. The fact that the change of ε/ɔ to e/o (for instance, *fleš* → *fleš*) is related to the properties of the standard language has been noted by Jurgec (2008). There he states that many of the English words in question have a colloquial version with an RTR vowel, but “as soon as they start being used publicly, frequently and/or formally, their mid vowels get replaced by ATR” (p.40). However, this observation is not formalised in Jurgec’ model. In sum, the change of ε/ɔ to e/o could be interpreted as an inter-language mapping, an effect of standardisation, comparable to the Serbo-Croatian phenomena described in §7.4.2. We will argue that this is indeed the case in a minority of speakers – those who have no contrastive tone in their standard Slovenian. In the majority of speakers, including the speakers of the prestigious dialect of the capital Ljubljana, the described change will be shown to be a consequence of an interaction of Lexical Conservatism and the native constraints on paradigms and the distribution of tone.

Towards this goal, consider the following. All the words in which the change ε/ɔ → e/o has happened are declinable nouns. Jurgec (2008) also points out that in Slovenian speakers who have tonal distinctions, the vowel quality is “crucially dependent on the tone” (p.145), and one might argue that his indexed markedness constraints actually refer only to the tonal dialects. In that case it is peculiar that Jurgec’ 2010 article is about Standard Ljubljana Slovenian (p.2), which does have tones, although to a limited extent in some speakers (see also the discussion in Jurgec 2008). Moreover Jurgec (2011) claims that there are no standard Slovenian speakers who lack tonal distinctions.

Before proceeding to the discussion of paradigms and tone, it is important to distinguish between the two goals of this discussion. First, we are trying to make it clear that there are speakers who have the unmotivated mappings ε→e and ɔ→o which they apply when adapting English words, and these are minimally the non-tonal speakers and maximally all Slovenian speakers. This means that we are describing the same effect as Jurgec, but using different means to encode it (not loanword-specific Markedness, but inter-language mappings). Second, we are pointing out that even without the
concepts specific to our model, for most speakers an account is available which makes no use of loanword-specific markedness.

There are two lines of argument worth pursuing and they are both related to paradigms which the words under consideration can enter. As noted by Jurgec (2010), there are three declension patterns which novel consonant-final nouns can join:

(a) with stable e/o - smeh, smeha ‘laughter’
(b) with stable e/ɔ - gost, gosta ‘guest’
(c) with e~e and ɔ~o alternation - svet, sveta ‘council’, uhod, uhoda ‘entrance’

The first argument is valid for both tonal and non-tonal speakers: there are clear signs that the paradigm in (c) has many more members than the paradigm in (b). Toporišić (2000) quotes that in the case of a nominative form containing an e, in around 90% of the items there is an alternation with e (whereas the remaining items ‘can be learned’). This may mean that when speakers are exposed to a form like fleš, they just go with the statistical tendency and produce the genitive fleša. This would be a case of LC overruled by the statistics. However, the desire for paradigm uniformity then selects from the pool of available forms and chooses a paradigm which consistently has the ATR vowel (fleš). This analysis is possible (if the statistics are strong enough), but it goes against a tendency which we have seen over and over again: if LC can be satisfied, even a very peripheral (or even novel) paradigm type can become default for loanwords. However, there seems to be additional reasons for most (or all) standard speakers to avoid the paradigm in (b).

Joining a paradigm has consequences for the tone on the stressed vowel, which is distinctive. Tone can be high or low in Slovenian, but in loanwords it is almost invariably high. Discussing this phenomenon, Jurgec (2008) stresses that high tone better mimics the source stress in the languages from which Slovenian borrows most words (English, German, Italian). He also states that “there are paradigms with a high tone in all forms, whereas the opposite isn’t true” (p.95).

As it turns out, there are also no paradigms with e/ɔ and a consistently high tone. In Becker & Jurgec (forthcoming) we read “in paradigms with fixed stress […] e and ɔ show with predictable tone: Low tone on all members of the paradigm, except the NOM.SG and GEN.PL”.

That means that the nouns with an e/ɔ which enter Slovenian can keep their vowel only if they sacrifice both their paradigm uniformity with respect to the tone (i.e. if they introduce allomorphs with different tones) and, consequently, the resemblance to the original stress in many forms. In (103) we will use LC(tone) to encode this.
Inter-language mappings

(103)

<table>
<thead>
<tr>
<th></th>
<th>LC(tone)</th>
<th>LC(ATTR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. fleš, fleša</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. fleš, fleša</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Now we have to explain what blocks the candidate pattern fleš, fleša. As we have seen in the quote from Becker & Jurgec, words with e/ɔ have a low tone except in nominative singular and genitive plural. Becker & Jurgec assume the high tone in these forms to be due to a floating underlying tone associated with the case morphemes, which can override the lexical tone. Here again, the new arrival status of the loanword has a crucial role in our account – the word enters Slovenian as a bare form and with a high tone. Slovenian speakers have all the reasons to analyse the high tone as the exponent of the bare - nominative singular form, i.e. fleš gets analysed as fleš + O_H. Note that LC does not thwart this analysis because no new allomorphs are introduced. This is why when the time for the assignment of a paradigm comes, the introduction of the new paradigm fleš, fleša is not supported by LC and loses.

It has been pointed out to us by Peter Jurgec (p.c.) that Serbo-Croatian nouns which are perceived with a low tone show the same pattern, so that the name of the Belgrade park Kalemegdan is perceived Kalemegdan but is still adapted as Kalemédgan, Kalemégdana. All we need here is to assume that this word goes through a phase in which it is a bare form, analysed as a nominative singular form. In that case, the floating tone will override its ‘lexical’ tone and yield Kalemédgan, which then, when the time comes to receive a paradigm, goes the same way as fleš. This essentially means that the tone in the source language is irrelevant, since it gets levelled out by the Slovenian nominative singular morpheme. If this is so, Slovenian grammar enforces the emergence of a mapping in which e/ɔ will be neutralised to e/o.

In this section, we have shown that standard Slovenian facts contain no compelling evidence for loanword-specific Markedness and that the way the relevant mappings are arrived at can be accounted for using Lexical Conservatism. Moreover, since standard Slovenian has both tonal and non-tonal dialects, and it seems that the source of the ban on e/ɔ is in the tonal dialects, every account implies an answer to the question of what kind of entity spread from tonal to non-tonal dialects. While Jurgec’ account seems to imply it was a constraint ranking which somehow got acquired by non-tonal speakers, it follows from our account that it was a contact-specific mapping, which either spread from dialect to dialect or emerged on the bilingual-community level. In sum, our account seems to offer a more realistic scenario, in which the grammars of the two types of speakers are
not affected by the mapping, which is passed on to new contact speakers due
to the simple fact that a sufficient number or existing loans contains
evidence for it. More generally, it appears to be quite a daunting task to
imagine a scenario in which loanword-specific Markedness would
spontaneously spread through a community and, since every such case can
be described in terms of inter-language mappings, it seems that the
mappings accounts are generally preferable.

6.6. Conclusions
This chapter has presented the aspect of our loanword model which focuses
on language entanglement and, rather than extending the logic from the
regularities which hold within single languages to contact situations, focuses
on what is specific to language contact situations: the inter-language
mappings. The inter-language mappings were shown to be contact-specific,
conventionalised correspondences between units of the languages in contact.

This was the final step in our theoretical intervention, which conceptually
divorces the discussion of language contact from that of processes within
languages, or at least allows for such a separation. The emergent agenda of
loanword research should then focus on alignment between languages
without thinking in terms of adaptation, but rather in terms of matching
multiple systems in order to make material transferable. One important
issue in such an agenda is what the possible representations and levels of
analysis accessible to inter-language mappings are and how they relate to
representations within the language. The issues discussed in this
contribution also invite a reconsideration of the boundaries of prototypical
languages. When something becomes a word in Dutch, Serbo-Croatian or
Basque is not obvious at all and the concept of inter-language mappings
shows why it shouldn’t be. Inter-language mappings are as real as the
grammars of any of these languages and crucial for accounting for the
predictability of the outcome of the borrowing of massive subsets of SL in RL.

We have also seen that inter-language mappings can change over time
and that this is exactly the place where LC effects can be identified:
mappings always change in the direction of satisfying more LC
requirements.
Chapter 7
Interim conclusions

In the previous three chapters we have presented our model of loanword integration. The goal of this model is to cater to the need we had repeatedly observed in the three chapters before that, in which we assessed the existing language contact and loanword literature: the need for a theory which introduces the lexicalisation perspective into the modelling of the borrowing process. Such a theory needed to incorporate the important insight that the object of borrowing is neither a sound nor a lexeme, since there is no transfer of whole lexical entries. Typically the only thing that can be transferred is a single pronounceable and morphologically opaque string with a specific meaning, which then serves as the basis for a new lexical entry in the recipient lexicon. Most processes which are usually discussed under the rubric of loanword adaptation can actually be viewed as part of a lexicalisation process in which the new lexical item receives all the properties required by RL, most importantly an RL paradigm. In this short chapter, we briefly summarise the presented model and situate it within the existing literature.

In Chapter 4 we foregrounded the three characteristics of the SL surface form which gets inserted into the RL discourse – the initial surface form: the fact that it lacks morphological and morphosyntactic features, the fact that it is a surface form, and the fact that it has a specific meaning. This form, we argued, cannot serve as an input for RL as is. As Chapter 5 showed, morphosyntactic integration can be seen as the trajectory which takes the initial surface form to whatever being a fully functional lexical item in RL entails. In data set after data set, we saw how this process is constrained by a single force, Lexical Conservatism, which requires the preservation of all the phonological features of the initial surface form in the newly formed paradigm. At the end of the trajectory, when the new lexical item is ready, its properties which cannot be protected by the native Faithfulness constraints have the option of coming under the protection of Faith(loan) – a family of constraints indexed to loanwords.

While this perspective of loanword integration is insightful, it only reveals one dimension along which loanwords matter. This is the perspective in which a single word enters the lexicon of a single speaker, which, in a certain sense, is always the case. And yet loanword data discussed by linguists never feature a single item or a handful of accidental borrowings: loanword data sets virtually always illustrate well-established mappings between SL and RL categories which readily apply to not yet borrowed items.
In this sense, loanword data sets actually always presuppose the existence of some level of familiarity and the existence of internalised correspondences between RL and SL structures. We termed such conventionalised correspondence between elements of different languages, established on a contact-specific basis, the \textit{inter-language mappings}. Since mappings are generalisations over existing word pairs, the existence of a mapping always enforces productivity: it always pushes the pattern towards being used on new items. In other words, the existence of a mapping makes sets in the source lexicon in principle borrowable and, as we have seen, possibly-already-borrowed. We will call this stage of language contact, where a considerable number of words support inter-language mappings and make certain words potentially borrowed, the state of \textit{language entanglement}.

This leaves us with a model whose two main ingredients – LC and inter-language mappings – also introduce two complementary perspectives on loanword processes. The LC perspective is more RL-centered and focuses on the essentially diachronic change, i.e. a word that was absent from a lexicon joins that lexicon. The mapping perspective focuses on the link between SL and RL and on what is synchronically present in the agents of borrowing. While there is no doubt that LC effects and the creation of inter-language mappings are concomitant in every language contact, LC and inter-language mappings cannot be observed in action simultaneously. In terms of Kiparsky (2006), the present model can be called ‘amphichronic’, with LC as its only universal. As Kiparsky states, “universals should constrain any kind of change, and they should define ‘attractors’ on which a variety of diachronic paths converge” (p.3).

Now we can look back at the two formal fields of loanword research and position the type of accounts they produce with respect to the ones we have proposed. The first area, discussed in Chapter 2 and commonly termed loanword adaptation, can offer many insights into how a mapping comes into existence (in more naturalistic studies) and into what a mapping can relate. In this sense, it is all the more important that recent studies, such as Dong (2012) (discussed in \S2.4.4), have deconstructed the ideal of the normal or regular borrowing situation by showing that borrowing proceeds differently and has different outcomes in different types of contact. As for the approaches to stratified post-contact lexicons discussed in Chapter 3, we recognise them as descriptively adequate for what we had termed \textit{borrowedness}: the status of items marked as loanwords and protected by a special type of Faithfulness \texttt{- Faith(loan)}. Our model is mainly concerned with what precedes this stage in language contact and is in that sense complementary to this field.
Although our model is complementary to most loanword models proposed so far, the accounts our model offers do raise different questions and target different data. In order to illustrate this, the following four chapters present four case studies in which the presented model is applied. In these case studies we shall account for broader phenomena and data sets than those discussed so far. Chapter 8 discusses the emergence of geminated consonants at stem edges in a variety of languages. This case study is unique in tackling a cross-linguistically attested phenomenon, showing how it is always guided by the same force: Lexical Conservatism. Moreover, it will offer an additional argument for separating loanword patterns from the ones we encounter within languages. Chapter 9 provides an analysis of two related consonant cluster-disrupting alternations in Serbo-Croatian, among which the $a$-epenthesis has already been discussed in §4.2.2.1. This case study focuses on phenomena which take place within the language after loanwords which violate native patterns have been integrated. This chapter will show that loanword-related LC can turn out to be the tip of the iceberg of the native LC processes. In Chapter 10, the current mappings for integration of English verbs into Serbo-Croatian are discussed, as well as other patterns of borrowing which still show some signs of productivity. While this case study focuses on Eastern Serbo-Croatian (where the integration suffix is -ova-), the existence of a different standard dialect, Western Serbo-Croatian, where the integration suffix is -a- will prove an important point of comparison. Chapter 11 presents an analysis of the integration of Latinate nominalisations into Serbo-Croatian. As will be shown, Latinate nominalisations are not a case of massive borrowing of morphologically complex items (predicted to be marked by our model). Instead, they become analysed as derived only after being borrowed as simplex items with a specific meaning.
Chapter 8
Stem-edge consonant gemination in loanwords

In this chapter we zoom in on a phenomenon which we have already briefly discussed in §2.2.2. and §4.2.2.3.: the emergence of geminates at stem edges in borrowing. This phenomenon will serve as an important illustration of our model’s independence from the RL grammar – this process is only indirectly connected to any native process. This, we will argue, is due to the fact that the process protects a phonological property which is not a part of underlying representations: the syllabic position of the stem-final consonant.

This chapter is organised as follows. In §8.0., we summarise the facts and accounts already presented in previous chapters. §8.1. provides a more detailed cross-linguistic overview of the phenomenon and discusses the common feature of a group of proposed accounts, the alignment analysis. In §8.2. we present the LC analysis of the emergence of geminates in the stem-final position. In §8.3. the mirror-image phenomenon of stem-final gemination is discussed: the stem-initial gemination. Our account of this phenomenon enables us to formulate the relevant LC constraint more precisely. §8.4. concludes the chapter.

8.0. What we have seen so far
The English borrowings into Japanese of the type bukku ‘book’ and pubbu ‘pub’, quoted in §2.2.2., illustrate the phenomenon – at the end of the borrowed stem there is a geminate while the corresponding SL word has a singleton consonant in that position, and a RL word with a singleton in the same position would have been possible (and expected). In the same section we have quoted an early account from Ohso (1971) repeated in (104) below.

(104)

The basic syllable pattern of English is CVC, while that of Japanese is CV, there is a close tie between VC of English CVC. If we change CVC to CV-CV in borrowing, the auditory impression of the original word is spoiled to a great extent, Thus we insert the phoneme Q to retain this close tie. (p.30)

While this approach differs from ours in making reference to perception (whose usefulness for formalising stable language contact situation we have questioned throughout Chapter 2), we fully subscribe to it in assuming syllabification in the input to borrowing. Our account of this type of data was first presented in §4.2.2.3., where we briefly discussed American Italian
adaptations of the type [kol.le] ‘coal’ (where [ko.le], would have been as licit).

We argued that the syllabification of the initial surface form is preserved in forms with a geminate: in kol.le better than in [ko.le]. The important feature of this account is that it makes reference to syllabification. Syllabification is generally assumed not to be part of underlying representations (or at least not one of the parts that can be protected by Faithfulness). However it is one of the features which LC was proposed to protect. As Steriade (1998) states, “[t]he difference between Lex(P) constraints and McCarthy and Prince’s correspondence conditions involve the range of phonological properties for which correspondence is sought and the mode of evaluation of lex P conditions: P in Lex(P) may be any phonological property, including global properties like overall stress pattern, syllabic shape or global identity” (p. 7).

Our account is in this sense a formalisation of the intuition expressed by many authors, from Ohso (1971) quoted above to Bafile (2010), who argues that stem-final gemination in Italian is “a strategy of preservation of the syllabic structure by means of gemination of the final consonant, by virtue of which that consonant preserves its status of the final element of the syllable”.19

In the same section, we also mentioned the stem-final emergence of geminates in languages which allow both singleton and geminate coda consonants (Hungarian) and have bare citation forms. In these languages, there is no obvious reason to introduce gemination in the citation form. In such cases, for instance meccs [mɛt̠ː] ‘match’, it is only in the inflected forms that the advantage of a geminate becomes visible. While meccs [mɛt̠ː] is as good as the potential * mecs [mɛt̠] in replicating the syllabification of the initial surface form, it is only the plural form of the first one mecs.csek that does so (unlike * me.csek).

Concluding this brief discussion, we established that this pattern only has bearing on loanwords and that, in order to accommodate for it, there is no need for either the RL grammar or the structure of RL lexicon to change in any way. We characterised this pattern as one in which all work is done by LC.

Finally, in §6.2.1. we formalised the American Italian pattern in the form of an inter-language mapping, repeated here as (105).

---

19 “una strategia di conservazione della struttura sillabica grazie al raddoppiamento della consonante finale, attraverso il quale tale consonante conserva il suo status di elemento finale della sillaba”
Emergence of geminates

(105) English $\rightarrow$ American Italian

input output

\[-VC\] $\rightarrow$ -VC,C,V\]

It is clear that comparable inter-language mappings can be written for Japanese and Hungarian, as well as many other languages in which the emergence of stem-final geminates has been attested. In order to give an overarching account of the phenomenon, we first provide a more detailed overview and briefly turn to the accounts proposed in the literature.

8.1. Overview of the phenomenon and proposed accounts

In order to provide a better typology of the phenomenon, we are quoting some additional data with English as the source language in each case. In (106), we repeat data quoted in Karvonen (2009). Note that all these languages have a singleton/geminate contrast, and that in these words, a singleton had been possible where the geminate arose.

(106)

a. Finnish

pop $\rightarrow$ poppi sheriff $\rightarrow$ sheriffi picnic $\rightarrow$ piknikki

b. Japanese

whip $\rightarrow$ wippu snob $\rightarrow$ sunobbu kit $\rightarrow$ kitto
c. Hungarian

tip $\rightarrow$ tipp plus $\rightarrow$ pluss sweater $\rightarrow$ szvetter
d. Italian

chat+are $\rightarrow$ [tʃatta:re] ‘to chat’
freak+etto+ne $\rightarrow$ [frikketto:ne] 'freak'

These examples will be used to make a number of methodological points. First, since all accounts proposed so far deal only with stem-final geminates, we are for the moment restricting our attention to this edge of the word. In §7.3. we will expand our data set to include stem-initial gemination. Second, there are examples where the geminate does not occur stem-finally, as in the Hungarian example szvetter. In such cases, various accounts can be invoked, such as the augmentation of the stressed syllables (Steriade 2004). Since this kind of emergence of geminates seems to be far less systematic, we do not consider it part of the phenomenon we are accounting for here. Third, it is important to note that perception can be invoked for many cases of stem-final gemination, for instance when the realisation of the geminate in RL
shares some phonetic properties of the realisation of the singleton in SL (e.g. consonant release). However, we maintain that perception cannot serve as complete account for the various kinds of speakers and borrowing settings in which it is encountered, so that perception can rather be invoked as the reason why the gemination entered the pool of possible correspondences.

By and large, all the cases we have seen so far can be classified into three types, depending on the material that follows the stem.

(a) no ending in the citation form, vowel initial endings in inflected forms (the Hungarian example)

(b) epenthetic vowel (the Finnish and Japanese examples)

(c) various desinences in the citation form and derivational morphology (the Italian examples)

So far, accounts have been proposed for the cases which belong to the latter two types. Shinohara (2003), analysing Japanese, Repetti (2009) addressing American Italian, and Karvonen (2009), analysing Finnish, have all proposed the same formalisation of this phenomenon: an alignment constraint in the RL. This constraint typically has the form ALIGN(Stem, R, σ, R). Such a constraint requires the right edge of the stem to be aligned with the right edge of some syllable. Assuming that this constraint has its effect in loanword integration, it is possible to derive stem-final gemination. This analysis successfully captures the fact that, for instance, Finnish geminates the consonant before the –i which is added, as in piknikki, but it does not geminate the consonant in front of a vowel present in the original word as sirtaki (*sirtakki). Moreover, this analysis is close to ours in that it recognises that borrowing amounts to the creation of a new stem from a surface form.

All these alignment analyses point out that the constraint enforcing stem-final gemination interacts with the structural constraints, some of which are undominated, which limits the occurrence of geminates in different ways. For instance, because in Italian z has no geminate counterpart, the words ending in this segment will always be borrowed with a single z (e.g. American Italian bruza ‘bruise’). In this case the crucial ranking is *zz >> ALIGN(Stem, R, σ, R) (or any other constraint requiring stem-final gemination, see below). As for exact interactions of the constraints in specific languages, we have little to add to the existing analyses quoted above.20

20 The only possibly new phenomenon we have encountered is blocking on borrowing due to phonological constraints. In European Italian, verbs can be formed from English verbs using suffix -are and gemination is then common (e.g. settare ‘to set’). However, speakers report ineffability in words like buzz, because both *bə zarə and *bə zə zarə seem to violate undominated constraints. I am grateful to Marta Castella for pointing this out to me. Clearly, further research is necessary here.
Additionally, the fact that the alignment accounts are not dealing with cases of the type (a) above does not really pose a problem: as discussed in the previous section, taking whole paradigms into account suffices.

Our discussion is therefore mainly concerned with the nature of the constraint which enforces stem-final gemination, on which there seems to be a consensus in the literature. This constraint, we believe, can be reviewed in a more natural and insightful way if considered from the perspective of the model proposed in this dissertation.

Two conceptual and two empirical arguments can be raised against the alignment analysis. The first conceptual problem is a general feature of OT analyses of loanwords, that the SL forms are considered as inputs in the tableaux. We discussed this issue in §2.4.1. Second, invoking alignment constraints in this case seems to sidestep facts which need to be formalised under any circumstances. If edges of stems and syllables prefer to be aligned, all logically possible combinations are expected to occur and be undominated in languages – Align(Stem, R, R), Align(Stem, R, R), Align(Stem, L, L), Align(Stem, L, L) and Align(Stem, L, L). However, this does not capture the intuition behind this phenomenon. The foreign stem is embedded in the borrowing language structure and its syllabification, in particular that of the final consonant, is preserved by introducing the geminate. In sum, we should want to rule out Align(Stem, R, R, L) and Align(Stem, L, R, L) and rule in Align(Stem, R, R) and Align(Stem, L, L).

As for the empirical arguments, first, Karvonen claims that the constraint Align(Stem, R, R) is present in Finnish, but masked for the native noun stems, which always already end in a vowel. The assumption of a constraint with such an effect predicts the existence of a language in which all singletons get geminated across the board when a vowel-initial suffix is added. As far as we know, there is no such language. Symptomatically, the only example Karvonen uses to show the same constraint in action comes from Finnish truncations. Truncations are a good example of an area where two surface forms are related, one of which (the truncation) is always derived later. For this reason, truncations are an area where we also expect LC effects. So even if the effect of the alignment constraint can be shown in truncations, the question remains why there are no entire languages which show such relations between bare and suffixed forms. Moreover, the -ari/äri-truncations, which typically display a maximally heavy syllable based on the part preceding the second vowel of the base word (yöpuku ‘pajamas’ → yöppäri) are perfectly compatible with an analysis which invokes the augmentation of the stressed syllable (Steriade 2004), which Karvonen seems to accept for Hungarian and Japanese loanword gemination.
The second empirical argument is related to the data from American Italian, where the constraint does not apply for morphological reasons. In this language, English words get reanalysed using Italian derivational morphemes based on the many related words of the two languages (e.g. *furniture* is analysed as *forni+tura*). Repetti quotes such examples because gemination is blocked in words like [ku.po.ne] ‘coupon’ (*[ku.pon.ne]), due to the fact that -one is an American Italian derivational suffix, which gets ‘reconstructed’. Another case of reanalysis is [dʒin.dʒa.rel.la] ‘ginger ale’ (*ella is the diminutive suffix), where gemination does arise, but only because it is present in the representation of the suffix. However, if -ella and -one are suffixes in these words, then [kup] and [dʒindʒar] should be new stems. If this is the case, and ALIGN(Stem, R, o, R) has its say in loanwords, the forms *[dʒin.dʒar.rel.la] and *[kup.po.ne] would be expected, with the new ‘stems’ /dʒindʒar/ and /kup/ aligned with syllables’ right edges. This, however, is not what we encounter.

### 8.2. Present analysis

Since our solution is not a typical OT analysis, before turning to our version of the rationale behind the emergence of stem-final gemination, we need to clarify what the status of this rationale is in our theory.

Unlike the proponents of the alignment analysis, who see the alignment constraint as part of RL grammar, we maintain that the constraint in question is actually providing a diachronic explanation for the inter-language mappings which have been stabilised in the languages we are discussing. More importantly, this constraint provides an amphichronic explanation (in the sense discussed in §6.2.) because of the fact that mappings of this kind emerge time and again when the RL has geminates.

The rationale behind the emergence of stem-final gemination in terms of the present model is straightforward: the syllabification of the initial surface form is the phonological feature to which the relevant LC constraint – LC(syll) – refers. This constraint therefore prefers pik.nik.ki over pik.ni.ki because it preserves the prosodic structure of *pic.nic*.

Our account does not pose any of the problems of the alignment analysis. It does not assume surface forms to be inputs of RL tableaux, rather, it postulates that lexicalisation of loanwords is a process separable from native constraints and evaluations. It also predicts that the process will look similar in all languages and that it will always protect the right egdes of stem-final syllables (with a possible extension to the left edges, as discussed in §8.3.).

As for empirical criticisms, our account does not predict the grammar in which all stems are perfectly aligned with syllables described in the previous
section. There can, however, emerge ‘islands’ of a perfect alignment between stems and syllables under specific circumstances. The Arabic borrowings into Turkish in (107), quoted from Aktürk (2008), provide such a case.

(107)

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Turkish</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ma.hal ‘location’</td>
<td>ma.hal.l-i ‘location-ACC’</td>
</tr>
<tr>
<td>b. sur ‘secret’</td>
<td>sur.r-u ‘secret-ACC’</td>
</tr>
<tr>
<td>c. zam ‘price increase’</td>
<td>zam.m-u ‘price increase-ACC’</td>
</tr>
<tr>
<td>d. zan ‘suspicion’</td>
<td>zan.n-u ‘suspicion-ACC’</td>
</tr>
<tr>
<td>e. his ‘feeling’</td>
<td>his.s-i ‘feeling-ACC’</td>
</tr>
<tr>
<td>f. haz ‘enjoyment’</td>
<td>haz.z-u ‘enjoyment-ACC’</td>
</tr>
<tr>
<td>g. af ‘forgiveness’</td>
<td>af.f-u ‘forgiveness-ACC’</td>
</tr>
</tbody>
</table>

The accusative forms are representative for all contexts where the suffix begins in a vowel, whereas in all the cases where the suffix starts in a consonant (e.g. in the plural forms in -lar/-ler), the version of the stem with a singleton surfaces.

There are two empirical facts important for understanding these Turkish data. First, Turkish never allows geminates in syllable codas. Second, all these nouns have Arabic correspondents ending either with a geminate (maḥall, sirr) or with a consonant cluster which could have been perceived as a geminate (‘afw). The question is therefore not why there are no geminates in the bare (and pre-consonant) form, but rather why this optimal alignment is restricted to cases where the initial surface form had a geminate itself. In other words, why did the Arabic kitāb or English bar get integrated as kitap~kitabu and bar~baru respectively, and not as *kitap~kitabbu and *bar~barru? Our tentative answer is that, while Turkish has the option of lexicalising a word with a stem-final underlying geminate, this solution is only employed to avoid a situation where two skeletal slots are absent from the syllabic position they occupy in the initial surface form, which is what would have been the case if sirr had been integrated as *sur~sw.ru. We do not see how an alignment analysis would account for this difference between bar and sur.

The other unconfirmed prediction of the alignment analysis, surplus gemination in words of the type [dʒin.dʒa.rel.la] and [ku.p.on.nə], is also not made by our account. Since these words’ novel “stems” have no source language surface forms to ‘sponsor’ them, there can be no constraint requiring their independent syllabification to be preserved. On the other hand, it may appear that our model is unable to predict [ku.po.ne] instead of [ku.pon.nə]. This is not the case, since our model formalises these correspondences between English sequences and Italian morphemes which,
as Repetti states “outweigh any other phonetic, phonological, or morphological considerations that affect consonant length” (p. 227), as inter-language mappings. Once such a mapping is established, the relevant LC constraints no longer apply.

There is, however, one remaining imprecision in the way LC(syll) works: it is not clear how the notion of syllabification is implemented, or, more precisely what is exactly the constellation LC(syll) militates against. On the one hand, a structure like *pik.ni.ki could have been failed because the syllabic/prosodic structure of the original word has been changed, since not all segments have the same syllabic role as in the initial surface form – k is in an onset and not in a coda. On the other hand, the problem could be that the incoming form piknik is now not kept together in prosodic terms, i.e. a part of it has been syllabified together with the material which does not belong to it (*pik.ni.ki). The processes at the right edge do not provide any evidence which would enable us to distinguish between the two – gemination solves both problems simultaneously. There are, however, processes on the left edge which enable us to pit the two implementations of LC(syll) against each other. We turn to them in the next section.

8.3. Stem-initial gemination: Maltese “undigested” borrowed verbs

All the accounts in the literature discuss examples of gemination brought into action in order to protect the syllabification at the right edge of the borrowed stem. In this section we look at a case where there is a gemination process at the left edge: Maltese. These data will help us extend our account to the left edge and resolve the dilemma of the implementation of the present model discussed above.

Maltese has a productive process in which initial consonants get geminated.21 The latter process is active only in “undigested” borrowed verbs. The term “undigested” was introduced by Mifsud (1995) to denote those borrowed verbs which have not entered the Semitic templatic verbal system, i.e. no consonantal root gets extracted from them. As a consequence, these verbs do not show any alternations in their vowels and do not enter the derivational relations characteristic of the Semitic verbal patterns. This phenomenon of initial gemination has received quite some attention in the literature (see e.g. Mifsud 1995, Hoberman & Aronoff 2003, Hoberman 2007), with many admittedly unsatisfactory attempts at accounting for the gemination facts either synchronically or diachronically. (108) provides

21 The process has been extended to initial consonant clusters, but with certain restrictions (see Hoberman 2007).
examples of Italian/Sicilian-based (a) and English-based (b) verbs. The first column contains the related non-verb in which there is no gemination.

(108)
a. facilìtà ‘ease, facility’  jiffacilità ‘to facilitate’  
prova ‘proof’  jipprova ‘to prove’
b. żum ‘zoom’  jiżżumja ‘to zoom’  
bulxit  jibbulxitja ‘to bluff’

Importantly, not all borrowed verbs are “undigested”: there are many verbs of Romance origin which have entered the templatic system: their advantage was that they had combinations of vowels and number of consonants which allowed for an interpretation in terms of that system. Examples of such verbs are jkanta (cf. Italian cantare ‘to sing’) and jfallì (fallire ‘to be absent’).

“Undigested” verbs are generally better preserved, i.e., in terms of our model, Lexical Conservatism has its say in their form. Beside stem-initial gemination, there are at least two other characteristics of the undigested verbs which show a higher satisfaction of Lexical Conservatism in these verbs. First, the stress can be lexically determined and invariable throughout the paradigm (jikkoágula ‘coagulate’). The stress of Maltese verb paradigms is mobile and predictable, as we know from §5.2.3. Second, in many Romance and virtually all the English-based verbs, the verbs end in the derivational suffix –ja. This suffix allows for the syllabification of the original stem to be preserved. Compare the “digested” jfal.li and “undigested” jiddawnlowd.ja in this sense.

Bearing in mind these latter additional observations, we can turn to the initial gemination. Mifsud points out that its origins can be found in the optional word-initial and stem-initial gemination encountered in many Sicilian and Southern Italian dialects. However, none of these dialects distinguish between verbs and nouns as Maltese does. In this sense, this historical fact can only be the reason why the initial gemination was in the feature pool to begin with. A reason for the specialisation has to be found in the structure of Maltese.

As it turns out, the reason is straightforward: the nominal paradigms of Maltese are different from the verbal ones in one important way. The nominal paradigm contains no prefixes: plurals are either broken (forma – pl. forom ‘form, verbal theme’) or suffixed (čekk – pl. čekkijiet). Given such

---

22 The citation form is the third person singular imperfective form which has the prefix j(i)-. Most of the data quoted here and in other sources come from Mifsud’s dissertation.
paradigms, stem-initial geminates would not improve borrowed nouns in any way. On the other hand, the obligatory verbal paradigm contains the following consonantal prefixes: /n-/ /t-/ /f-/ . Introducing these prefixes as n(i)-, t(i)- and j(i)-, Mifsud explains that “the vowel in brackets occurs only when the stem of the verb begins with more than one consonant (e.g. n-kisser vs. ni-tkisser)” (p. 38). Now we can see what is achieved by gemination: in front of the geminate the epenthetic vowel has to be introduced because this vowel in its turn prevents the formation of stem-initial consonant clusters in which the prefix and the stem-initial consonant would be in the same syllable. In other word, by having all verbs act like tkisser, Maltese makes sure that the prefix ends up as the onset of a different syllable.

As a consequence of this conspiracy of gemination and epenthesis, the prosodic autonomy of the stem is preserved. We can illustrate this by comparing the three singular imperfect forms of the integrated verb ‘to sing’ – nkan.ta1, tkan.ta2, jkan.ta3MASCULINE – and the corresponding forms of the verb ‘to download’ – nid.dawn.lowd.ja1, tid.dawn.lowd.ja2, jid.dawn.lowd.ja3MASCULINE. The forms avoided by gemination are *ndawn.lowd.ja1, *tdawn.lowd.ja2, *jdawn.lowd.ja3MASCULINE. Again, since the entire paradigm is evaluated by LC, the new allomorph – the one with gemination – extends to forms where there is no prefix, e.g. the imperative singular form iddawnlowdja.

Now we can resolve the implementation dilemmas from the previous section. First, since the gemination process affects consonants in the onset position, we can confirm that it operates on syllables, not moras, as expected from the nature of the initial surface form. Second, the constraint LC(syll) is implemented so that it prevents the borrowed stem from being syllabified together with RL material, i.e. so that the prosodic autonomy of the original form is preserved in terms of syllables.

**8.4. Conclusions**

In this chapter, the stem-final and stem-initial emergence of geminates in loanwords is explained in terms of preservation of the prosodic autonomy of the initial form (which coincides with the new stem), by preventing its syllabification together with the recipient language morphemes. The mechanism which ensures this preservation of prosodic autonomy is a constraint from the LC family, LC(syll). The presented account makes an important typological prediction, which seems to be confirmed, namely that stem-initial and stem-final gemination are only attested in contexts where reference is made to syllabification of an initial surface form, and most prominently in borrowing. To take our Finnish example for one last time, a
native process taking *piknik* to *piknikki* could be interpreted in terms of moraic affixation, attested in many languages. In this case, nothing would necessitate all the vowel-initial suffixes in the language to add an extra mora and all the consonant-initial ones not to do so.

The goal of this chapter was to explain the force behind the stem-edge gemination. It is clear, however, that the process is encouraged in some contexts and mitigated in others. One important ally of the constraint we discussed here may be the tendency to preserve the stressed syllables better than the ones without stress. Steriade (1999) argues for a constraint which requires that “the entire contents of the stressed syllable must find a lexical precedent in some listed allomorph” (p. 249). This predicts that monosyllables create an environment in which gemination of the final singleton will be encouraged in all situations where it is possible. We hope that future research will bring more specific analyses of this and other relevant interplays which help establish geminating mapping.
Chapter 9
Lexicon, Markedness and Grammar in the Serbo-Croatian Wobbly $a^{23}$

This chapter focuses on two Serbo-Croatian (S-C) alternations which both involve the vowel $a$ which disrupts consonant clusters in certain environments. One of them, the $a$-epenthesis, has already been discussed in §4.2.2.1. We then argued that the S-C lexicon has a stratum of nouns indexed as loanwords, in which the normal native complex coda-disrupting $a$-epenthesis is blocked. All native and some borrowed S-C words behave like /visk/, which maps onto visak ‘pendulum’ in the bare nominative form, but shows the cluster in all the suffixed forms (visk-u ‘pendulum-dative’). And yet complex codas are attested in thousands of frequent loanwords, such as Bask ‘Basque’ and disk ‘disc’ (but also kobalt ‘cobalt’, ofsajd ‘offside’, jamb ‘iamb’, marketi[ŋ]). Our account was that the loanwords of the latter type, after having been inserted as code switches, could not receive the same representation as the native words of the type visak. Rather, since their initial form already revealed their special status, they were indexed as loanwords in the lexicon and the Faithfulness indexed to loanwords was invoked. The crucial ranking of constraints then looks like DEP(LOAN) >> *COMPLEXCODA >> DEP (previously LC >> *COMPLEXCODA >> DEP).

We also concluded that the stratification approaches (see Chapter 3) correctly predict that novel loanwords, such as helpdesk or task will be integrated without epenthesis, because they enter the periphery of the lexicon and automatically come under the auspices of DEP(LOAN). In sum, once LC has introduced certain structures to the lexicon, FAITH(LOAN) takes over after the words containing these structures are lexicalised, so LC need not play any role in the synchronic grammar. LC is only relevant when the lexical entry is under (re)construction; when there is no instability, FAITH suffices.

In what follows, a closer look will show that while this picture is correct, loanwords seem to have ushered quite some variation into S-C, because there are multiple related alternations which often fail to affect loanwords. This has led to the creation of paradigms which satisfy LC to different extents. This variation has, in turn, ushered in some new possibilities of reanalysis of native structures.

---

23 This case study is a slightly adapted version of the paper submitted to Arsenijević et al. (forthcoming), which I have co-authored with Dr. Antonio Baroni from the University of Padua.
In order to present the full picture of the interplay of LC, Dep(Loan) and other relevant forces, in this case study we focus on the synchronic situation which holds within S-C. The focus is restricted to the classes of nouns, which are the only class displaying $a$-epenthesis in loanwords.

This case study will allow us to see how LC operates in lexicalisation, but also how its status changes as the loan becomes a full-fledged lexical item. As will be shown, while it is very easy for LC to enforce the creation of a new item with minimal allomorphy, anti-allomorphy effects spread far less easily among already existing words, for which multiple allomorphs can be listed. Even when LC is absolutely undominated, as is the case in modern S-C, the lexicon can contain listed allomorphs which preserve outputs from previous stages of the grammar, and these allomorphs can then be preferred by the current rankings. In such cases what we call lexically sponsored unmarked allomorphs can surface and LC does not militate against them. In other words, while LC ushers in only minimal allomorphy in incoming items, it plays no role in simplifying the paradigms of existing lexical items.

This latter generalisation does not mean that LC plays no role in the dynamics of existing lexical items: if restructuring in the grammar makes a new type of paradigm possible, such a paradigm may never actually arise due to LC, which puts a serious limit to what can be lexicalised and, consequently, to what can surface. As this case study will show, in experimental settings, LC can be obviated and types of paradigms can be created which cannot lexicalise in any natural way. If speakers believe that they are guessing a new form of an existing word, they will not feel “responsible” to maintain the phonological features of the listed form (simply because there is none in their lexicon) and consequently produce paradigms which may in principle be allowed by the grammar, but cannot ever lexicalise given the actual sources of lexical enrichment. In sum, this case study will illustrate how LC militates against not-yet-lexicalised allomorphy, both in new items and in restructuring of existing items.

The rest of this chapter is organised as follows. §9.1 introduces the alternation and summarises the existing accounts. We show that there are two types of $a$ which shows in some forms of some nouns. One is epenthetic and surfaces only in bare forms of the native and nativised words, while the other is metathetic and emerges in GENPl of some nouns whose stem ends in a consonant cluster, where the cluster gets disrupted by the first half of the morpheme GENPl/aːː/ /aːː/. In §9.2, we introduce our OT account, which makes use of a faithfulness constraint indexed to loanwords Dep(Loan). While this suffices for accounting for the bare forms, in which an epenthetic $a$ surfaces, the other, metathetic $a$, which emerges in (the non-bare) GENPl of some nouns, requires an explanation which takes LC and the shape of the
paradigm into account. This is why the following three sections discuss the
three nominal paradigms in which the metathetic a emerges under slightly
different conditions. §9.6. addresses the restructuring of coronal ST clusters,
which seems to be directly attributable to the influx of loanwords with
complex codas and the undominated LC. We show experimental evidence
that these clusters, which act as inseparable in all existing words, could be
separated by modern speakers, but there is no reasonable lexicalisation path
which would allow such a paradigm to emerge. §9.7. concludes the chapter.

9.1. Introduction

Observed synchronically, there are at least two types of a in S-C, as
recognised in all traditional descriptions: one that surfaces in all forms of the
word (and can therefore safely be assumed to be underlying) and the other
which occurs only in some forms. The paradigms of the two nouns in (109),
both trivially borrowed, differ in that the stem of karat ‘carat’ contains only
undoubtedly underlying a’s, which surface in all forms in the paradigm. On
the other hand, in the paradigm of koverat ‘envelope’ the stem-internal a
surfaces only in two forms, NOMSG and GENPL, and it is absent from all the
other forms. This latter type of a is traditionally termed nepostojano a
‘fleeting, mobile a’ (for a discussion of the term and the concept, see
Mihaljević & Horvat 2008). The term we will be using for the a which
surfaces within the stem, but only in some forms, is wobbly a.

(109) NOMSG GenSG NomPL GenPL
(a) karat karata karati karātā
(b) koverat koverta koverti koverātā

What we see in (109b) can be summarised as the disruption of stem-final
consonant clusters in NOMSG and GENPL. However, as we already know, not
all nouns whose stems end in a consonant cluster exhibit this alternation.
(110) shows two nouns in which the zero:a alternation underapplies. In
koncert ‘concert’, it applies only in GENPL, whereas in jogurt ‘yogurt’ it does
not apply at all. Finally, there are no S-C nouns which display the pattern
opposite to that of koncert: in no noun with a full paradigm is it the case that
the alternation applies only in NOMSG.

24 The ā stands for a long a. The GENPL form has its own prosodic specificities. The
two final syllables of the GENPL form are long in all forms in which this form ends in
an ā. Also, in many cases the GENPL form has a different tonal pattern from the rest
of the paradigm. While recognising that the prosodic pattern may be necessary to
complete the picture, this paper ignores the prosodic information and focuses
exclusively on segmental alternations.
Alternations of the type presented in (109b) and (110a) are typically discussed under the rubric of Slavic *yer(s)* in the formal phonological literature (see e.g., Inkelas & Zec 1988, Halle & Nevins 2009). What these approaches have in common is partially replicating the historical development of the Slavic *yers*. The assumption is then that there are still abstract units, *yers*, in the underlying representations of the words which have a wobbly *a*, and grammar decides which of the two possible realisations – *a* or zero – it will have in each form. Such approaches make use of a modification of a vocalisation rule originally proposed by Lightner (1972), which boils down to vocalising all *yers* which are preceded by a *yer*, and deleting all other *yers*. Such an analysis can account for the conditions under which *yer* surfaces in NOMSG (assuming a *yer* as the NOMSG suffix). However, the occurrence of wobbly *a* in GENPL is more problematic: as can be seen in (109) and (110) the suffix is segmentally equal to that of GENSG, but it triggers the insertion of a wobbly *a*. The usual solution is then to assume that the derivation of GENPL is exceptional in that it involves an extra cycle, in which the *yer* is vocalised before the suffix is added (for instance because a *yer* is added at that cycle). Note that these approaches need to account for a rather peculiar distribution of these abstract *yers*: they never occur in any context other than between two last consonants of stems ending in a consonant cluster. There are, for instance, no nouns of the type /vrUk/ (*U* being the *yer*), which would have the NOMSG [vrak] and GENPL [vraka] but GENSG [vrka].

We propose an alternative analysis which does not make use of cycles or special underlying segments involved, and in which there is a structural reason why stems in which wobbly *a* surfaces always end in a consonant cluster. Different modules come into play to explain the wobbly *a* phenomenon: Lexicon, Markedness and Grammar. In our OT-based account, the introduction of a wobbly *a* prevents a marked structure – a consonant cluster and, as will be shown later, a hiatus – from surfacing.

As can already be read off the examples in (109) and (110), a merely phonological account is not possible, since among nouns with identical phonological structure some do display *a*:zero alternation in all contexts (109b), some do it only in a subset of contexts (110a), and others do not display it at all (110b). This information must be stored lexically. As will be

<table>
<thead>
<tr>
<th></th>
<th>NOMSG</th>
<th>GENSG</th>
<th>NOMPL</th>
<th>GENPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>koncert</td>
<td>koncerta</td>
<td>koncerti</td>
<td>konceratā</td>
</tr>
<tr>
<td>b.</td>
<td>jogurt</td>
<td>jogurta</td>
<td>jogurti</td>
<td>jogūrtā</td>
</tr>
<tr>
<td>c.</td>
<td>*[poncerat</td>
<td>poncerta</td>
<td>poncerti</td>
<td>poncērtā]</td>
</tr>
</tbody>
</table>
shown, both lexical representations and grammar are on the move in present-day S-C and the issue of lexical storage is a matter of an interaction of various factors at the grammar-lexicon interface.

9.2. Present Account

The crucial feature of our proposal is a conceptual separation between the wobbly a’s in the two contexts. What superficially looks like the same object, i.e. an epenthetic low vowel, is in fact the realisation of two different objects: an actual epenthetic vowel which occurs in certain nouns in the NomSG form – as in koverat, whose underlying form is /kovert/ – and a metathetic a, which is part of the discontinuous GENPL morpheme /aːaː/.

In both cases, it is Markedness that triggers the wobbly a’s surfacing within the stem. In the case of the epenthetic a (in forms like [koverat]), it is the Markedness constraints against complex codas that require the insertion of a vowel (thereby blocking forms of the type *kovert). In the case of the metathetic a, Markedness constraints block /aːaː/ from emerging faithfully and it therefore always spreads over two syllables, surfacing as [aːCaː] (where C = any consonant), e.g., [kovaːCaː].

The central piece of evidence for the existence of two types of wobbly a comes from their distribution. A ‘true’ epenthetic a displays a regular pattern, as it disrupts all complex codas in all native and quite a few borrowed nouns. The behaviour of metathetic a, on the contrary, appears to be the locus of massive variation and is highly dependent on the paradigm shape, morphological constituency, lexical frequency, etc. Moreover, the ‘unpronounceable’ shape of the GENPL ending helps us explain why in GENPL the wobbly a initially seems to disrupt a consonant cluster which occurs in the majority of the forms of the noun. In other words, GENPL is not less tolerant than GENSG when it comes to consonant clusters, it can simply afford to disrupt the consonant cluster by virtue of realising all the segmental material of the suffix.

9.2.1. OT Formalisation

The insertion of the ‘true’ epenthetic a can be formalised in OT considering the constraints in (111) and the ranking in (112).

(111)
\[
\text{MAX-STEM} \quad \text{Input segments belonging to the stem must have a correspondent in the output.}
\]
\[
\text{DEP} \quad \text{Output segments must have a correspondent in the input (No epenthesis).}
\]
Chapter 9

*COMPLEXCODA Complex codas are not allowed.
*[LOW] The feature [LOW] is banned from the output.
*[HIGH] The feature [HIGH] is banned from the output.

The crucial markedness constraint *COMPLEXCODA captures the markedness-reducing nature of the a-epenthesis in S-C: it serves the purpose of repairing a structure which would otherwise surface in many faithful outputs: complex codas.

(112) MAX > *COMPLEXCODA > DEP > *[HIGH] > *[LOW]

(113) Tableau: koverat, NOMSg

<table>
<thead>
<tr>
<th>Input: /kovert/</th>
<th>MAX-STEM</th>
<th>*COMPLEXCODA</th>
<th>DEP</th>
<th>*[HIGH]</th>
<th>*[LOW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [kovert]</td>
<td></td>
<td>*!</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b) [koverat]</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c) [kover]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) [kove]</td>
<td><em>!</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) [koverit]</td>
<td></td>
<td></td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
</tbody>
</table>

The tableau in (113) shows the evaluation through which candidate (b) emerges as the winner. The most faithful candidate, which is (a), is ruled out because it fatally violates *COMPLEXCODA, which is quite high in the hierarchy. Candidate (c) avoids the violation of *COMPLEXCODA by deleting the last consonant, but incurs a violation of MAX-STEM, the highest-ranking constraint. Candidate (d) dispenses with the coda cluster altogether, thus violating MAX-STEM twice. Both candidates (b) and (e) violate DEP, but while (b) inserts a low vowel in order to disrupt the cluster, (e) inserts a high one. Since *[HIGH] is ranked higher than *[LOW], candidate (b) is the winner.

As we have already seen, not all nouns which underlyingly end in two consonants behave like koverat: many loanwords avoid epenthesis. As already proposed in §4.2.2.1., our account is that a subclass of faithfulness constraints, FAITH(LOAN), which only apply to items marked in the lexicon as loanwords, allowing the consonant cluster to surface. In this specific case, we just need to rank DEP(LOAN) higher than *COMPLEXCODA.
(114) Tableau: *koncert ‘concert’ NOMSG

<table>
<thead>
<tr>
<th>Input: /konʃert/</th>
<th>DEP(Loan)</th>
<th>Max-Stem</th>
<th>*Complex Coda</th>
<th>DEP</th>
<th>*[Low]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [konʃert]</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [konʃerat]</td>
<td>!</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c) [konʃer]</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) [konʃe]</td>
<td>!*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All candidates except (a), which is the most faithful one, violate either DEP(Loan) or Max-Stem, therefore (a) is the only possible winner. As already argued before, DEP(Loan) is takes over from LC when the loanword becomes lexicalised.

As for metathetic *a, it can be analysed as the result of a ranking where *Hiatus (“Sequences of two vowels are disallowed”) is undominated and where Max-Suffix (“Do not delete input segment belonging to the suffix”) dominates Linearity (“No metathesis”).

(115) Tableau: *koverát‘ ‘envelope’ GENPL

<table>
<thead>
<tr>
<th>Input: /kover + aːa/</th>
<th>*Hiatus</th>
<th>Max-Stem</th>
<th>Max-Suffix</th>
<th>Lin</th>
<th>*[Low]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [kovaːtaː]</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [koveraːtaː]</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>c) [kovaːtaː]</td>
<td>!*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

In the tableau in (115) candidate (a) is the most faithful to the input, but maintaining the genitive plural morpheme as such incurs the fatal violation of *Hiatus, which is ranked higher than faithfulness constraints. Candidate (c) deletes one of the two long vowels, violating Max-Suffix. Candidate (b) is therefore the winner, since it only violates the low-ranked constraint Linearity.
The tableau in (116) shows the emergence of the GenPL of the loanword *koncert*, which in this case behaves exactly like native(-like) nouns and selects (b) as the winning candidate. As a matter of fact, since the wobbly *a* in *koncerata* is not epenthetic, its occurrence does not entail the violation of Dep(Loan).

Up to this point, we have accounted for all the forms which are derived by the grammar *stricto sensu*, defined as a ranking of universal constraints which evaluates the possible outputs and/or compares them to inputs which consist of (concatenations of) established underlying representations. Note that even in this picture, some lexical perspective is present: a class of Faithfulness constraints is indexed to a lexical stratum which allows more marked structures than the rest of the lexicon: that containing loanwords. This is also the domain which incoming loanwords naturally enter. However, our account still remains a domain of grammar in the sense that we recognise lexical stratification as a universal possibility which is reflected in various categories (stems, nouns, content words, etc.) allowing more marked structures than the rest of the lexicon, formalisable as faithfulness constraints indexed to these categories.

So far, we have sketched what would be the output of the grammar if the underlying representation were the only representation to be considered and if representations were not susceptible to restructuring. However, in order to extend our account to all nouns (among others, those in 110), we need to take into account that S-C nouns always appear in paradigms, and the interactions between the forms in the paradigm are decisive for what can surface in them. For this reason, from this point on, we are moving to considering separate paradigms. This will prove beneficial because, as we shall see, different factors will play a role in each case.
9.3. Masculine nouns

Since all the nouns used as examples so far belong to the class of masculine nouns which end in a consonant, which is also the only class with paradigms where both the epenthetic and the metathetic wobbly a surface, in this section we essentially complete our account of the rather complicated constellation of masculine nouns.

Considering paradigms, it is important to emphasise that, albeit conceptually distinct, the two wobbly a’s do introduce the same segmental allomorph when they surface. For instance, in the case of [koverat], with [a] occurring between [r] and [t], this allomorph can be both the result of epenthesis, as in the NomSing, and metathesis, as in the GenPl [koverata:]. Note that the GenPl suffix automatically makes the last two vowels of the case form long (e.g. in jogurtā), so it is safe to assume that the (prefinal) length is actually represented separately from the segmental content.

Our account is entirely Steriadian in the sense that this is exactly the kind of relations that Lexical Conservatism was devised to account for. LC is perfectly blind to the fact that [koverat] has different origins in NomSing and GenPl – it remains the same allomorph. Hence LC seems to be the reason why no native word in the class under consideration fails to disrupt their stem-final consonant clusters in both NomSing and GenPl: the allomorph with a wobbly a is ‘ushered in’ by the exceptionless epenthesis in native words and surfaces also in GenPl.

When it comes to loanwords, variation emerges, again as a result of LC constraints, which now play a rather different role in paradigm formation. This leaves us with the typology of nouns in this class repeated here as (117).

(117) NomSing GenSing NomPl GenPl
a. koverat koverta koverti kovarātā
   subjekat subjekta subjekti subjekātā
b. koncert koncert koncerti konsērtā
   delikt delikta delikti delikātā
c. jogurt jogurta jogurti jogūrtā
   impakt impakta impakti impiktā

d. *[poncerat poncerta poncerti ponsērtā]

The data in (117) indicate the existence of three subclasses within this class of masculine nouns. (117a) is an example of a group of nouns which exhibit epenthetic a in the NomSing and metathetic a in the GenPl. Other nouns, like koncert in (117b), only have metathetic a, and nouns behaving like jogurt, in
Chapter 9

(117c), do not display either epenthetic or metathetic a. Crucially, (117d) shows that among the four logical possibilities one is missing: there is no noun with epenthetic a but no metathetic one.

The model proposed so far accounts for the facts in (117a) and (117b), the difference between the two being that only koncert is under the auspices of FAITH(LOAN), although both words are etymologically loanwords. As for the gap in (117d), we propose to interpret this gap as evidence that the NomSG serves as the licenser, which also seems plausible given the loanword trajectory. This licensing relation has the important consequence that since epenthesis does not fail in any native items, metathesis is also omnipresent. This ‘ushering’ relation actually masks the fact that in S-C metathesis is often blocked in cases where it introduces a new allomorph without the support of NomSG. This division is already visible in the split between the types (117b) and (117c) in masculine nouns. Our notion of the licenser is similar in spirit with the notion of the base (Albright 2002). The main difference is that in the cases we are considering, especially in loanwords, the NomSG form pre-exists the GenPl since the former is also the citation form and the one which participates in inter-language mappings.

The only class which still needs to be accounted for is that of borrowings of the type (117c) jogurt. While the stem of this noun ends in the same cluster as koncert and koverat, it does not display either epenthetic or metathetic a.

(118) Tableau: jogurt ‘yogurt’ NomSG

<table>
<thead>
<tr>
<th>Input: /jogurt/</th>
<th>Dep(Loan)</th>
<th>Max-Stem</th>
<th>*Complex Coda</th>
<th>*[Low]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) jogurt</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b) [jogurat]</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c) [jogur]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) [jogu]</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

The NomSG of jogurt is [jogurt], entirely equivalently to koncert, since the insertion of a vowel would imply a violation of Dep(Loan) and the simplification of the complex cluster is not a viable solution given that Max-Stem is ranked higher than *ComplexCoda. Crucially, unlike koncert, jogurt allows no metathesis, which is still not predicted by our grammar, as the tableau in (119) shows.
(119) Tableau: jogūrtā ‘yogurt’ GENPL

<table>
<thead>
<tr>
<th>Input: /jogurt + aːa/</th>
<th>*HIATUS</th>
<th>DEP (LOAN)</th>
<th>MAX-SUFFIX</th>
<th>LIN</th>
<th>*[LOW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [jogurtaː:]</td>
<td>⬠</td>
<td></td>
<td></td>
<td>⬠</td>
<td></td>
</tr>
<tr>
<td>⬤ b) [joguraːtaː:]</td>
<td></td>
<td>⬠</td>
<td></td>
<td>⬠</td>
<td>**</td>
</tr>
<tr>
<td>⬤ c) [jogurtaː]</td>
<td></td>
<td></td>
<td>⬠</td>
<td>⬠</td>
<td>*</td>
</tr>
</tbody>
</table>

The current ranking predicts the victory of (b), jogurata, with metathetic a, but the actual output is jogurta, i.e. candidate (c). Note that the candidate (b) is not violated by DEP(LOAN) because the metathetic a is sponsored by input material.

Our solution to this incongruence between the current ranking and what actually surfaces is positing that in the paradigm formation of jogurt LC constraints played a crucial role and that LC still prevents the grammar from introducing the new allomorph [jogurat]. In assuming this, we remain committed to our claim that LC constraints only play a role when not all forms in the paradigm are established and a candidate form is about to introduce a new allomorph. In that sense, we are not claiming that LC for jogurt is somehow ranked higher than for koncert and koverat, but that at a certain point in the development of S-C, blocking of the introduction of a new wobbly allomorph became an option. From that point on, only the items which have a listed allomorph which contains a wobbly a were allowed to display metathesis in GENPL.25 This solution essentially means that the metathesis account we have proposed for the GENPL forms like konceratā (116) is actually recent history in one important sense: it is actually only possible for items which have a listed wobbly allomorph. All the other items are evaluated by the same ranking of constraints, but on top (or perhaps preceding) this ranking there is LC, which blocks all the candidates which introduce new segmental allomorphs. Since in present day S-C, new items never introduce extra allomorphs due to metathesis, we assume that all evaluations go through the lexical check by LC and that all items which allow metathesis have a listed allomorph, as shown in the tableau in (120). The bold line between LC and all the other constraints is used to indicate that LC might be ontologically quite different from the (other) grammar and,

25 Note that we implement the concept of listedness in the Steriadian spirit, so that listedness in the lexicon does not mean being part of the UR. For different implementations, see Kager (2008) and references therein.
since it seems to be undominated in all tableaux, so it may turn out that it should be conceived of as a filter rather than as a constraint.

(120) Tableau: koverätā, koncerätā and jogūrtā in Modern S-C

<table>
<thead>
<tr>
<th>Input: /kover + aː/</th>
<th>Listed allomorphs: [kover], [koverat]</th>
<th>LEXCON</th>
<th>*HIATUS</th>
<th>DEP (LOAN)</th>
<th>MAX-SUFFIX</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [koverat:aː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [koveraːtaː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) [koveraː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input: /kontsert + aː/</th>
<th>Listed allomorphs: [kontsert], [kontserat]</th>
<th>LEXCON</th>
<th>*HIATUS</th>
<th>MAX-SUFFIX</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [kontserat:aː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [kontseraːtaː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) [kontseraː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input: /jogurt + aː/</th>
<th>Listed allomorphs: [jogurt]</th>
<th>LEXCON</th>
<th>*HIATUS</th>
<th>MAX-SUFFIX</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [jogurtaːaː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [joguraːtaː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) [jogurtaː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that in the new system, there is no difference between koverat and koncert when it comes to GENPL. However, the second listed allomorph (the one with a wobbly a) is still licensed by the NOMSG for koverat, whereas for koncert, its listedness is a matter of encoding the output of an older state of the grammar in the lexical entry.

Importantly, the obligatory listedness of the wobbly allomorph does not mean that the wobbly a in GENPL is no longer metathetic. Metathesis still takes place and the normal faithfulness constraints still prevent the wobbly allomorph from surfacing in forms in which the ending is a single vowel (e.g. koncerta *koncerata 'yogurt-GENSG'), i.e. in context where the wobbly
allomorph would include a vowel absent from the UR. In other words, the fact that the allomorph [konšerat] is listed does not make it the new UR and it is only in some forms that such an allomorph is preferred. We term this allomorph in GENPL a **lexically sponsored unmarked allomorph**.

Also note that the analysis of the epenthetic wobbly *a* in *koverat* ‘envelope-NOMSG’ (113) and its absence in *koncerat* ‘concert-NOMSG’ (114) are not affected by listedness. Having listed the allomorph [konšerat] in the lexical entry by no means predicts that this allomorph will surface in NOMSG, since the NOMSG is crucially calculated based on the concatenation of the UR of the stem and a zero suffix, so that DEP(LOAN) still blocks epenthesis.

### 9.4. Neuter nouns ending in -o/-e

As already mentioned, masculine nouns ending in a consonant are the only class in which both the epenthetic and the metathetic *a* occurs. Taking only this class into account would have created the illusion that native words always allow metathesis in GENPL. This picture was complicated already by loanwords which enter this inflection class, as only some of the items allow metathesis.

In this section, we turn to neuter items, whose declensions are different from the masculine ones in very few paradigm cells, among which the citation form, which now always has a vowel ending: either -e or -o. Since the citation form of these nouns ends in a vowel, there is no context for epenthesis in the NOMSG. This noun class still exhibits metathetic *a* in the GENPL, but the wobbly *a* underapplies heavily, especially in infrequent forms. Examples are given in (121).

<table>
<thead>
<tr>
<th></th>
<th>NOMSG</th>
<th>GENSG</th>
<th>NOMPL</th>
<th>GENPL</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>koplje</td>
<td>koplja</td>
<td>koplja</td>
<td>kopǎljā</td>
<td>‘spear’</td>
</tr>
<tr>
<td>(b)</td>
<td>bezǔmlje</td>
<td>bezǔmlja</td>
<td>bezǔmlja</td>
<td>bezǔmljā</td>
<td>‘lack of reason’</td>
</tr>
<tr>
<td>(c)</td>
<td>staklo</td>
<td>stakla</td>
<td>stakla</td>
<td>stakālā</td>
<td>‘glass’</td>
</tr>
<tr>
<td>(d)</td>
<td>porēklo</td>
<td>porēkla</td>
<td>porēkla</td>
<td>porēklā</td>
<td>‘origin’</td>
</tr>
</tbody>
</table>

As can be seen, (121a) and (121c) show the emergence of metathetic *a*, whereas (121b) and (121d) do not. It appears that especially in this class, which consists only of native items, there is considerable inter-speaker variation and many cases with prevalent ineffability. For instance, in the author’s idiolect, *plātno* ‘linen’, *pisǎmce* ‘letter-DIM’ and *dno* ‘bottom’ have no
acceptable GENPL forms.\(^{26}\)

Our account is the same here as in the previous section: in present-day S-C, the ranking is the same for all nouns, and the type *stakalā* is possible because at some point in the history of the grammar the nowadays lexically sponsored unmarked allomorph *stakal* emerged and was listed, whereas the same did not occur to the unattested *porekal*.

(122) Tableau: *stakala* and *porekla* in Modern S-C

<table>
<thead>
<tr>
<th>Input: /stakl + aːaː/</th>
<th>Listed allomorph:</th>
<th>LEXCON</th>
<th>*HIATUS</th>
<th>MAX-SUFFIX</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>[stakl], [stakal]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) [staklaːaː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [stakaːlaː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c) [staklaː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input: /pore:kl + aːaː/</th>
<th>Listed allomorph:</th>
<th>LEXCON</th>
<th>*HIATUS</th>
<th>MAX-SUFFIX</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>[porekl]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) [pore:klaːaː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) [porekaːlaː]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) [pore:klaː]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

9.5. Feminine nouns ending in -a

In the two classes considered so far, the only available GENPL suffix was /aːaː/. The feminine nouns in -a also make use of this suffix, but display a radically different pattern due to the availability of another allomorph for GENPL, namely /iː/. This is also the only class where we register serious signs of restructuring with respect to the situation described in descriptive S-C grammars, which is why we first briefly describe the commonalities and specificities of the two stages. What has always been the case is a clear division, illustrated in (124),

\(^{26}\) For these specific items, the failed form with a wobbly a would also introduce a new tonal pattern, unattested elsewhere in the paradigm. The study of ineffability would profit much from including prosodic information and the LC constraints sensitive to it in the analysis.
between the nouns whose stems end in a consonant cluster (e.g., majk-a ‘mother’) and those whose stems do not (e.g., muk-a ‘trouble’ and bo-a ‘boa’). The difference is that in the latter group there is only one possible segmental exponent of the GENPL ending: -a, so the GENPL of reka and boa are segmentally identical to the NOMSG, while in the former group the stem can have an allomorph with a wobbly a (majaka-a ‘mother-GENPL’) and there is also the other version of the ending which always occurs with the a-less allomorph (majk-i ‘mother-GENPL’). What is specific to S-C grammars, but ungrammatical in modern usage (the author’s idiolect) is the occurrence of the GENPL ending: -a with the a-less allomorph on CC-final stems. For instance Barić et al. (1997) describe a system in which majka is (next to majaka and majki), a possible GENPL form of majka. In the modern usage, forms like majka (a-less stem + a) are entirely excluded for CC-final stems. In the overview in (123), the form which is not grammatical any longer is presented between brackets.

(123) NOMSG GENSG NOMPL GENPL Gloss
(a) muka mukē muke mūkā ‘trouble’
    boa boē boe bōā ‘boa’
(b) mājkā mājkē mājkke mājkākā (mājkē) ‘mother’

There also seems to be much more optionality in the older grammars that in modern usage: while majaka and majki both illustrate possible patterns, only the latter form is grammatical for this noun. On the other hand, devojka ‘girl’ only allows devojaka in GENPL (but no *devojki).

Given the fact that the situation described in older grammars might very well actually include several (dialect) grammars, we are focusing on accounting for the modern grammar. Note that unlike the previous cases of restructuring, the new stage of the grammar and the lexicon does not allow all the forms covered by the older version, although the form of some lexical entries can only be accounted for using the output of the old grammar as input to the modern learners.

An overview of the possible forms of GENPL in the modern usage is given in (124). As can be seen, in the new system, there are no GENPL forms of the type *tetkā/*čēkā.

(124) NOMSG GENSG NOMPL GENPL Gloss
a. tetka tetke tetke tetkā ‘battle’
b. čēkta čēkte čēekte čēkki ‘brush’
c. patka patke patke patkā/patkī ‘duck’
d. rēka rēka rēke rēkā *rēkī ‘river’
(124) shows that stems ending with a consonant cluster may select either /aːa:/ or /iː/ as GENPL allomorph, as in (124a-c), while other stems always select /aː/. Since /aːa:/ is the only option for stems which do not end in CC, it is safe to assume that /aːa:/ is generally preferred and the occurrence of /iː/ is a repair strategy which is only allowed when LC blocks metathesis. Crucially, unlike in the other classes, whenever /aːa:/ is used to disrupt consonant clusters, there is a wobbly a on the surface (*tētkā, *čētkā are not possible GENPL forms). Note that the total absence of [aː] as the exponent of GENPL (unlike masculine GENPL jogūrtā, neuter GENPL bezūmljā) is additional evidence in favour of the metathesis analysis.

In order to formalise the fact that the GENPL of reka is rek[aː] and the GENPL of četka is četki, we need to postulate that a complex constraint, MAX-SUFFIX&*NOCODA, is ranked higher than *[HIGH], where NOCODA stands for “Syllables should not have codas”. Complex constraints are the product of Local Conjunction, whose formal definition from (Smolensky 2006): is given in (125).

(125)

a. A constraint C in Con may be the local conjunction of two simpler constraints in Con, A and B: if C = A&B, then C is violated whenever A and B are both violated within a common domain D.

b. C may be viewed as implementing the conjunctive interaction of A and B.

c. Universally, the conjunction dominates its conjuncts: A&B >> { A, B }. (p. 68-73)

Even though some authors argue against the Local Conjunction between a faithfulness and a markedness constraint (e.g., Itô & Mester 1998, Fukuzawa & Miglio 1998), others have claimed that it is necessary (e.g., Baković 2000). We argue that MAX-SUFF&NC is a plausible combination because it is not likely to produce unnatural outputs, since both constraints are “independently necessary” (Crowhurst 2011:1486). Moreover, it seems to be the only reasonable formalisation of the fact that /iː/ emerges as the GENPL allomorph if (and only if) LC blocks metathesis and the stem ends in a consonant cluster.
Tableau: četki ‘brush’, GENPL

<table>
<thead>
<tr>
<th>Input: /tʃetk/ + /aːː/ or /iː/</th>
<th>Listed allomorph: [tʃetk]</th>
<th>LC</th>
<th>*HIATUS</th>
<th>MAX-SUFF&amp;NC</th>
<th>*[HIGH]</th>
<th>MAX-SUFF</th>
<th>NC</th>
<th>*[LOW]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a)</td>
<td>![</td>
<td>*</td>
<td></td>
<td>*</td>
<td>![</td>
<td>![</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b)</td>
<td>![</td>
<td>*</td>
<td></td>
<td>*</td>
<td>![</td>
<td>![</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c)</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d)</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
</tbody>
</table>

In the tableau in (126) candidate (c) is the winner because the selection of the non-default GENPL allomorph /iː/ is the only possibility to avoid the violation of HIATUS and MAX-SUFFIX&NC without creating a new allomorph. The creation of a *četak- form would violate LC, while candidate (d) is ruled out by the conjoined constraint.

Tableau: tetka ‘aunt’, GenPl

<table>
<thead>
<tr>
<th>Input: /tetk/ + /aːː/ or /iː/</th>
<th>Listed allomorphs: [tetk], [tetak]</th>
<th>LC</th>
<th>*HIATUS</th>
<th>MAX-SUFFIX&amp;NC</th>
<th>*[HIGH]</th>
<th>MAX-SUFF</th>
<th>NC</th>
<th>*[LOW]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a)</td>
<td>![</td>
<td>*</td>
<td></td>
<td>*</td>
<td>![</td>
<td>![</td>
</tr>
<tr>
<td></td>
<td></td>
<td>![b)</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c)</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d)</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
</tbody>
</table>

The ranking shown in the tableau in (127) is the same as that in (126), with the difference that this time an allomorph displaying wobbly a is listed in the lexicon. Therefore, the selection of the default GENPL allomorph /aːː/ proves to be the ideal solution. Candidate (b) avoids hiatus, does not delete any

\[²⁷\] Note that we are assuming that intervocalic clusters are syllabified by placing the syllable boundary between the two consonants. This can be formalised using the constraint *COMPLEX "Syllables do not have complex margins" (Kager 1999:288), which is ranked below MAX, so it does not cause any segment deletion, but it can still influence syllabification.
material and does not exhibit either a coda or a high vowel, unlike candidates (c) and (d).

(128) Tableau: reka ‘river’, GENPL

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [re:kaːaː:]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>b) [re:kiː:]</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) [re:kaː:]</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

The tableau in (128) is shown in order to explain why feminine nouns whose stem does not end with a consonant cluster do not select /iː/. Since *[HIGH] dominates Max-Suff, it is better not to have a front vowel in the output than to delete part of the suffix, and given that the stem /re:k-/ does not contain a cluster, the complex constraint does not play any role.

As already mentioned, while there has been restructuring which has led to different acceptability of forms, the nouns of the type tetka still carry traces of the old system in the sense that their lexical entry still contains the allomorph tetak which is not computed by the current grammar. Importantly, it is a simplification to assume that all of these allomorphs are solely maintained by the existence of the GENPL form: in the case of tetka, there actually does exist a related noun which contains this same allomorph derived by epenthesis: tetak ‘uncle’. However, there are many items with no support of this type: višnja ‘sour cherry’ has the GENPL form višanjā, but there is no related word with a wobbly a. It is an empirical question which we leave to further research in what part of cases the maintenance of the allomorph is “sponsored” by related words.

9.6. Residual issues: the touchability of the untouchables

Traditionally, SC clusters consisting of a sibilant followed by a coronal stop /st, zd, žt, ʒd/ (henceforth ST) are treated as ‘unsplittable’ in all ways. Namely, they are never disrupted by wobbly a and the GENPL allomorph in feminine nouns is always -a, just like in all stems which end in a single consonant.
This ‘unsplittability’ is not particularly surprising: cross-linguistically, ST clusters form the least separable extreme within the broader class of hardly separable SC clusters (for a cross-linguistic perspective see, Fleischhacker 2005). SC and ST clusters are therefore the cut-off point for insertion in many cases. For instance, Dutch loanwords containing consonant clusters are adapted in Sinhalese by inserting an epenthetic vowel between the two consonants, except in the case of SC clusters, where the epenthetic vowel precedes the sibilant (Boersma et al. 2000).

(129)  
<table>
<thead>
<tr>
<th>Dutch</th>
<th>Sinhalese</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan</td>
<td>päläna</td>
<td>‘plan’</td>
</tr>
<tr>
<td>vrouw</td>
<td>porova</td>
<td>‘queen (of cards)’</td>
</tr>
<tr>
<td>stal</td>
<td>istallaya</td>
<td>‘stall’</td>
</tr>
</tbody>
</table>

These loanword facts could be explained by perceptual distance: SVC is a more salient departure from SC than VSC, whereas CVR is closer to CR than VCR (where V = any vowel, R = any sonorant, cf. Steriade 2001).

The homorganicity of S and T makes ST clusters even more “unsplittable”. The additional unity of ST clusters is not surprising from a perceptual point of view either. These clusters are known to constitute a unique ‘perceptual bond’. Fleischhacker (2005) points out that breaking up such a cluster “result[s] in an extremely noticeable difference between the intact cluster and that cluster affected by insertion or deletion” (p.73). Olender (2013) points out that ST is different from other /s/+stop combinations because of its homorganicity. The latter cites the example of Smith’s (1973) study on his son Amahl, who was able to produce /st/ before other clusters.

Among S-C nouns, there are many (both native and borrowed) items of the type shown in (130a), but none of them follow the pattern shown in (130b). Moreover, the standard grammars only mention the feminine paradigms of the type shown in (130c), whereas those of the type shown in (130d) are never registered.

(130)  
<table>
<thead>
<tr>
<th>NOMSG</th>
<th>GENSG</th>
<th>NOMPL</th>
<th>GENPL</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. rāspust</td>
<td>rāspust</td>
<td>rāspusti</td>
<td>rāspūstū</td>
<td>‘vacation’</td>
</tr>
<tr>
<td>b. *[rapsust]</td>
<td>rapsusta</td>
<td>rāspusti</td>
<td>raspusātā</td>
<td></td>
</tr>
<tr>
<td>c. cesta</td>
<td>ceste</td>
<td>ceste</td>
<td>cēstā</td>
<td>‘road’</td>
</tr>
<tr>
<td>d. pasta</td>
<td>paste</td>
<td>paste</td>
<td>pasti/?pāstā</td>
<td>‘pasta’</td>
</tr>
</tbody>
</table>
In sum, in the picture presented in S-C grammars, ST sequences are uniquely inseparable in every way. However, there seem to be signs of a reanalysis of ST sequences into regular i.e. splittable clusters. First, for most recent loanwords the paradigms of the type shown in 130d) are common. Second, an experiment conducted by Simonović (2009b) showed that speakers exposed to nonce forms of the type natazad often create GENSG forms like natazda, forming exactly the unattested paradigm illustrated in (130b). In other words, speakers are willing to accept the hypothesis that wobbly a can disrupt a ST cluster since they derive the nonce form natazad from underlying /natazd/. However, this currently does not lead to any reanalysis in the treatment of the existing words, presumably due to LC.

9.6.1 The experiment

The experimental data on which we report here are relevant as qualitative evidence that S-C speakers approach ST sequences in a way that existing words in no way allow us to predict. The items we discuss were used as fillers in an experiment on perception of borrowedness in non-words. Among the 110 items which were used as stimuli, there were four items ending in ‘interrupted’ ST clusters [-sat, -zad, -ʃat, -ʒad]. These items were embedded in sentences as NOMSG forms and the speakers were asked to guess other case forms of this noun. It was expected that a native speaker exposed to NOMSG forms ending with [-sat, -zad, -ʃat, -ʒad] would never construct underlying forms with /-st, -zd, -ʃt, -ʒd/ because such an alternation is unattested in existing words.

Participants. 50 university students who are S-C native speakers were enrolled in the experiment, of both genders.

Materials. A list of nonce legal words, presented as rare S-C words, were presented to the participants in their (supposedly) NOMSG forms. The four items which we are focusing on here were natazad, nališat, pugusat and nogožad, each with an interruption of an uninterruptible cluster.

Procedure. Participants were exposed to the NOMSG forms and asked to produce other case forms of the nouns.

Results. A little less than half of the participants in the experiment deleted the a in a case form after having been exposed to NOMSG forms ending with a [SaT] sequence. For instance, the GENSG forms of natazad, nališat, pugusat, nogožad were produced as natazda, nališta, pugusta, nogožda.

This finding suggests that by now (at least some) speakers do not represent ST clusters as special in any way and treat them exactly like other consonant clusters. In such speakers, the pattern described in (130a) is still the only one that surfaces, but not because it is the only one allowed by the grammar. Rather, the pattern in (130a) used to be the only one allowed by
the grammar. Now the grammar does not block the pattern in (130b), but the
dynamics of the lexicon (LC) do not allow any items of the type (130b) to
lexicalise. Consequently, the items whose stem ends in an ST sequence are
represented in the native grammar as any other item which allows only one
allomorph. This is illustrated in (131).

The tableau in (131) shows that wobbly a is not allowed in raspust because
LC blocks it. The fact that an allomorph such as *raspusat has never been
produced by the older grammar is probably due to the perceptual unity of ST
clusters discussed earlier, which Boersma et al. (2000) formalise with a
DEP/S_C constraint. However, in the present-day S-C it is the existence of the
lexical item with a single listed allomorph that suffices to ensure no new
allomorphs are added, given the high ranking of LC. The same blocking
mechanism applies to new loanwords (e.g. kast ‘cast’, mast ‘must’), since they
enter the language with a single allomorph. In this respect, kast and mast
are much the same as recent loanwords in other consonant clusters, since
they are all assigned to paradigms in which the stem has only one allomorph,
as illustrated in (132). Forms like NomSg *konstrejnat or GenPl *impakata
are unattested.

The tableau in (131) shows that wobbly a is not allowed in raspust because
LC blocks it. The fact that an allomorph such as *raspusat has never been
produced by the older grammar is probably due to the perceptual unity of ST
clusters discussed earlier, which Boersma et al. (2000) formalise with a
DEP/S_C constraint. However, in the present-day S-C it is the existence of the
lexical item with a single listed allomorph that suffices to ensure no new
allomorphs are added, given the high ranking of LC. The same blocking
mechanism applies to new loanwords (e.g. kast ‘cast’, mast ‘must’), since they
enter the language with a single allomorph. In this respect, kast and mast
are much the same as recent loanwords in other consonant clusters, since
they are all assigned to paradigms in which the stem has only one allomorph,
as illustrated in (132). Forms like NomSg *konstrejnat or GenPl *impakata
are unattested.

<table>
<thead>
<tr>
<th>Input: /raspust/</th>
<th>Listed allomorph: raspust</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a</strong></td>
<td>[raspust]</td>
</tr>
</tbody>
</table>
| **b** | [raspusat] | *! | | | * | **
| **c** | [raspus] | *! | | | | *
| **d** | [raspu] | *!* | | | | *

(131) Tableau: raspust ‘vacation’ NomSg

The results of the experiment, in which native speakers allow ST clusters to be disrupted, can
therefore be accounted for by the fact that nonce words are not listed in the
lexicon of the speakers and therefore LC plays no role whatsoever, as shown in the tableau in (133). However, unlike in borrowing, in this experiment the speakers were led to believe that they were dealing with existing words which were not introduced into the language by them, so that the whole concept of the initial allomorph did not apply.

(133) Tableau: NomSg of nonce word natazad

<table>
<thead>
<tr>
<th></th>
<th>LEXCON MAX-STEM</th>
<th>*COMPLEX CODA</th>
<th>DEP</th>
<th>*[LOW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) [natazd]</td>
<td></td>
<td>*!</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>b) [natazad]</td>
<td>F</td>
<td>*</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>c) [nataz]</td>
<td>*!</td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

The crucial aspect of the experimental settings which is captured by the tableau (133) is that the speakers believe that they are guessing the forms of an existing word. A nonce word like natazad cannot be associated to existing allomorphs, so LC is vacuously satisfied. This explains the peculiar situation in a language which has undominated LC and Dep(Loan), so that restructuring will show only under very specific circumstances, such as guessing paradigms of nonce words and allowing the paradigms of the type (130d) for new items. This crucially means that, given the current grammar (of at least some speakers), the paradigms of the type (130b) are perfectly representable in S-C, and yet given the high ranking of LC and Dep(Loan) there is no plausible scenario which would lead to words with such paradigms being lexicalised.

9.7. Conclusions and goals for further research

The analysis of the zero:a alternation within S-C paradigms presented in this paper can be summarised in the following way.

(1) There are two different wobbling a’s in S-C. One is epenthetic and occurs only in forms with a null ending of nouns whose stem ends in a consonant cluster and it is enforced by the constraint that militates against complex codas. The other wobbling a is metathetic and surfaces only when the GENPL suffix /aːaː/ is applied. The metathesis is guided by a constraint that militates against hiatus, which also explains why metathesis only occurs in stems which end in a consonant cluster.

(2) There is massive underapplication of both types of wobbling a.
While the underapplication of the epenthetic $a$ is limited to loanwords, the metathetic $a$ shows much inter-speaker variation, even in native items. We account for the first type of underapplication using a special type of Faithfulness indexed to loanwords (DEP(LOAN)), while the other type is due to Lexical Conservatism.

(3) Although the two wobbly $a$’s surface for different reasons, they introduce the same segmental allomorph: stems ending in /-CC/ get an allomorph ending in [-CaC]. As a consequence, in the only class which has both types of wobbly $a$ in the same paradigm, the epenthetic $a$ licenses the metathetic $a$, making both $a$’s ubiquitous in all native words of this class.

This specific constellation makes the NOMSG forms appear as licensers of the GENPL forms in the sense that grammar guarantees the consistent presence of a wobbly $a$ in NOMSG, which makes an allomorph with a wobbly $a$ always listed and available for GENPL.

(4) In present-day S-C the class of words which allow wobbly $a$ acts like a closed class: new items never allow the introduction of the allomorph with a wobbly $a$. For this reason, we are proposing that in the current grammar LC is never violated, so that all paradigms of new words will display a single allomorph in all forms. Under such a grammar all the nouns which allow the metathetic $a$ have a listed allomorph in their lexical entry.

(5) The traditionally inseparable ST clusters (which were originally treated as ‘unsplittable’) show different signs of reanalysis into regular clusters. This reanalysis influences the results of an experiment in which the speakers are asked to produce the forms of nonce words which were presented as rare S-C words. Under these circumstances, ST clusters are interrupted by wobbly $a$ in a sizable number of items. However, this type of paradigm, while perfectly possible, is very unlikely to lexicalise, since all words enter S-C either as loanwords or as another type of neologism, always falling under the auspices of LC, which blocks the introduction of wobbly $a$.

Although the experimental results are based on a very limited number of items, it is an important finding that speakers produced paradigms which are not only unattested, but also seem impossible to lexicalise under the current circumstances. The crucial role of LC-guided lexicalisation in determining what is attested in S-C has important theoretical consequences. The standard OT concept of the Richness of the Base (which forces one to consider all the possible inputs to the grammar) does not suffice for
languages with complex paradigmatic relations and has to be complemented with a theory of grammar-lexicon interface.

Finally, cases like this are important for the general conceptualisation of the role of loanword facts in linguistic research. While loanwords are typically used as the supplier of new and unexpected inputs to the grammar, thereby revealing its hidden aspects and letting it ‘express itself’, cases like this show that the opposite can also be the case. In S-C, where virtually all borrowed nouns ending in a consonant get assigned to the masculine declension, loanwords, which always enter the language with a non-wobbly allomorph and stay that way, actually (even if only statistically) strengthen LC and thereby prevent the grammar, which already allows paradigms of the type in (130b), from “expressing itself”.
Chapter 10
Unspecified verb aspect in Serbo-Croatian borrowed verbs

This case-study chapter focuses on a phenomenon of which we have made multiple mentions of in previous chapters: the integration of borrowed verbs into Serbo-Croatian. Borrowing is a major source of what is commonly referred to as biaspectual verbs – verbs which, unlike a huge majority of Serbo-Croatian verbs, have no inherent verb aspect specification. In §5.2.1. we presented an overview of the well-established mapping which turns any English verb into a Serbo-Croatian verb by means of the suffix -(ov)at-. We saw that while this mapping produces only verbs with unspecified aspect, all the other ways of introducing new verbs, most prominently the use of a borrowed noun or adjective as a base, produce verbs with a single aspect. The explanation was that LC militates against splitting the incoming verb into two lexical items. In §5.2.1., we formulated the mapping which we repeat in (135).

(134) English→Serbo-Croatian for verbs

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xv</td>
<td>X(ov)atiV.BIASPECTUAL (and the whole paradigm)</td>
</tr>
</tbody>
</table>

In §6.2.2. we clarified that the unrestricted transfer of English verbs is blocked in particle verbs (for instance sign in), whose borrowing is complicated by the existence of the particle, so that such verbs cannot be considered as covered by the mapping. Finally, in §6.4.1. we presented evidence that the current integration suffix -ova- emerged as the fittest from a previous competition with two stress-attracting suffixes -isa- and -ira-. The explanation was that, a toneless/stressless integration suffix is preferred by LC, especially in contact with English, which allows stress in different positions.

In this chapter we will further refine the picture sketched so far, demonstrating that the influence of LC goes beyond the current inter-language mapping, as speakers will be shown to use LC-friendly suffixes in a nonce derivation experiments even if these suffixes are not part of their inter-language mappings. While further evidence will be presented for the

---

28 This case study is a slightly adapted version of the paper I am co-authoring with Dr. Tanja Samardžić from the University of Zürich, based on the handout quoted here as Simonović & Samardžić (2013).
absolute domination of the LC-friendly mapping in the current contact with English, we will use the verbs which for different reasons do not follow the mapping to show what other scenarios are possible for lexicalising a verbal stem and under which circumstances these scenarios apply. The rest of this chapter is organised as follows. §10.1. brings introductory remarks on the nature of verb aspect in Serbo-Croatian. In §10.2. we present a detailed overview of the current Eastern Serbo-Croatian mapping, which makes use of the suffix -ova-. In §10.3. we are turning to the borrowings from English and other languages which, for different reasons, do not participate in the current mappings. §10.4. concludes this chapter.

10.1. The nature of verb aspect in Serbo-Croatian

The inventory of Serbo-Croatian (henceforth S-C) verbs is organised in such a way that the same verbal stem can be (and typically is) used to express several verb aspect categories. Aspeсtual modifications are obtained through a series of lexical derivations where the basic form is combined with a number of prefixes and suffixes. We can distinguish two kinds of derivations: those involving prefixation (135) and those without prefixation (136). As shown in the parentheses in (135-6), all forms are assigned one of the two traditional Slavic verb aspect categories: perfective or imperfective. Perfective forms mostly express telic events and imperfective atelic events, but this correspondence does not hold for all verbs and events.

(135) Derivations with prefixes

a. basic_form
   kOp-A-ti (imperfective)
   dig-V-INF
   ‘dig’

b. prefix + basic_form
   is-kOp-A-ti (perfective)
   out-dig-V-INF
   ‘dig out’

c. prefix+basic_form+iterative_suffix
   is-kop-AA-vA-ti (imperfective)
   out-dig-V-IT.SUFF-INF
   ‘dig out continuously or repeatedly’

d. pref+pref+basic_form+iterative_suffix
   po-is-kop-AA-vA-ti (perfective)
   over-out-dig-V-IT.SUFF-INF
   ‘dig out all / everything’
As illustrated in (135-6), the basic form consists of a stem (expressing the root), a verbaliser (whose form varies depending on the inflectional class), and an inflectional morpheme. Apart from several exceptions (e.g. *kupiti*<sup>PERF</sup> ‘buy’, *baciti*<sup>PERF</sup> ‘throw’), the basic form is imperfective. The same root can form a perfective verb by adding a prefix (135b). The meaning of the prefixed form is resultative (Arsenijević 2007). Most S-C verb prefixes are also found as free prepositions (as can be read off the glosses). Combining the basic form with various prefixes is similar to attaching particles to English verbs: the prefix does not only change the aspect of the verb, but also its meaning (Milićević 2004). A suffix can be attached to the prefixed form, which then results in a new imperfective form (135c) whose meaning is ambiguous between progressive and iterative. Not all prefixed forms can enter further derivations. The question of why further derivations are not available for some prefixed forms has been in the center of the discussion on Slavic verb aspect (Arsenijević 2006, Žaucer 2010). Finally, one more prefix can be added to the new imperfective form (135d). The form consisting of two prefixes and a suffix is perfective and it expresses multiple telic events.

Generative approaches to Slavic aspect point out a crucial difference between prefixation and suffixation. Prefixation is considered closer to lexical derivation because it changes the meaning of the verb. Dickey (2012) found that, unlike Bulgarian and Slovenian, S-C has no prefixes which have lost their original special meaning, so that the S-C perfectivisation system as such “is largely limited to lexical prefixation and subsumption” (p.92).

Suffixation, on the other hand, is a more formal derivation because the semantic contribution of the suffixes does not change from verb to verb. In the baroque derivational system of S-C, there is one important generalisation which has no exceptions: while suffixes can lead both to imperfectivisation and perfectivisation, prefixes always have the effect of perfectivisation. Derivations with no prefixes involved (136) are simpler. A suffix attached directly to the basic form does not allow any further derivations.
It is important to understand the concept of derivation in the synchronic sense in which it is used here. In what we have seen so far there has been no introduction of novel items based on already present items. For instance, for the present-day speakers the ‘basic form’ $kOpati$ is not any ‘older’ than the more ‘derived’ forms $iskOpAti$ or $kOpnuti$. The concept of derivation has only been used to say something about related, synchronically present elements of the lexicon. These elements share a stem and are surely derivationally related, but not much is predictable about the exact selection of suffixes and prefixes. In many cases there are also segmental modifications inside the stem. For instance, an alternative form for (135c) is $iskAApAti$, where the ablaut $kop~kaap$ has the same function as suffixation. Moreover, the examples above show that the same stem shows up in quite a few prosodic shapes.

Given the way native verb stems participate in actual verbal lexemes, it will come as no surprise that borrowed verbal stems typically do not enter any of the two types of derivations. The typical pattern in which borrowed stems participate is illustrated by three well-integrated verbs in (137).

(137)

a. stem+IIrA+ti \hspace{1cm} evocIIrAti \ (unspecified aspect) \hspace{1cm} ‘evoke’  
b. stem+Isa+ti \hspace{1cm} fotogrAfsati \ (unspecified aspect) \hspace{1cm} ‘photograph’  
c. stem+ova+ti \hspace{1cm} AngAžovati \ (unspecified aspect) \hspace{1cm} ‘engage’  
d. stem+a+ti \hspace{1cm} gUglAti \ (unspecified aspect) \hspace{1cm} ‘google’

Borrowed verbs in S-C are typically associated with one of the suffixes -$ira$, -$isa$, -$ova$, -$a$. These suffixes are attached directly to the bare stem. Given this position, verbs formed with these suffixes could be analysed either as basic forms (135a) and (136a) or as derivations of the type illustrated in (137b-c). However, unlike the suffixes -$nu$ in (136b) and -$ka$ in (136c), -$isa$, -$ira$, -$a$ and -$ova$ do not have the capacity of changing verbs’ aspect. This means that verbs formed with these suffixes are in fact basic forms like (135a) and (136a) and that the function of these suffixes corresponds to the ‘verbaliser’ function of -$a$ in (135a) and (136a).

Nevertheless, there is an important difference between borrowed verbs and native basic forms: if the verbalisers in (137) were entirely equivalent to those in (135a) and (136a), borrowed verbs would all be assigned the imperfective aspect. However, this is not the case. Most borrowed verbs in S-C are not specified for aspect: the same verb form is used to express both telic and atelic events. The main goal of this chapter is to show that the difference between native and borrowed verbs with respect to encoding verb aspect is a consequence of features which are specific to loanword integration, particularly the loanword trajectory and LC.
Not all suffixes in (137) are currently part of inter-language mappings: both -ira- and -isa- are virtually absent from the currently active mappings. The remaining two suffixes, -a- and -ova- participate in the two mappings which are currently sovereign for incoming English verbs in the two different (though not clearly separable) standard dialects. While it is clear that each of the suffixes was dominant somewhere and at some point in time, since it is our goal to understand the nature of the borrowing process and the way the relevant mappings emerged, focusing on the current mapping(s) has important methodological advantages. First, we know what the actual initial surface form could have been in each case. Second, there are possibilities for testing hypotheses on actual speakers. For this reason, in the following section, we are presenting the currently productive mapping for the integration of verbs into (Eastern) S-C and experimental evidence for it. The other borrowed verbs will be discussed in the subsequent section, and in respect of this mapping.

10.2. The current mapping

Virtually any English verb can be converted into a S-C verb using the mapping shown in (134), which always produces a verb which can in principle serve as both imperfective and perfective. The discussion of this mapping is organised as follows. In §10.2.1 we discuss the behaviour of the suffixes involved in this mapping, both in native derivations and in the inter-language mapping. In §10.2.2, we turn to experimental evidence for the domination of the mapping in Eastern S-C and to the interpretation of its restriction to English verbs.

10.2.1 The integrating suffix -(ov)a-

The two integrating suffixes, the Western S-C -a- and the Eastern S-C -ova-, are very well attested in the native layer of the lexicon. They are also the two most frequent suffixes in the small group of the native verbs with unspecified aspect.29 However, since the number of native verbs with unspecified aspect does not exceed dozens, virtually all verbs in -a- and -ova- correspond to the pattern in (135a) and (136a), and are imperfective. In sum, while there is nothing close to a predominantly biaspectual suffix in the native stratum, -a- and -ova- are the best possible native candidates for suffixes that do not change verb aspect.

29 The number of native verbs with unspecified aspect is very small, especially if all tenses are taken into account. We were able to identify very few native verbs with unspecified aspect which do not end in either -ova- or -a-: jesti ‘to eat’, videti ‘to see’, razumeti ‘to understand’, čuti ‘to hear’.
As we have already seen in §6.4.1., -a- and -ova- seem not to have lexical tone and allow the initial surface form to keep its original prominence (cf. invAjt(ov)ati ‘to invite’ and hAjlAjt(ov)ati ‘to highlight’), which makes them LC-friendly. The same seems to hold in the native domain. The few verbs with unspecified aspect are faithful to the surface prosody of the related noun, as illustrated in (138).

(138)

\[
\begin{align*}
\text{dOrUč-k-a} & \quad \text{‘breakfast-GEN’} & \text{dOrUčkati}^{\text{IMP/PERF}} & \quad \text{‘have breakfast’} \\
\text{pRsten-a} & \quad \text{‘ring-GEN’} & \text{pRstenovati}^{\text{IMP/PERF}} & \quad \text{‘ring (a bird)’} \\
\text{rUUč-k-A} & \quad \text{‘lunch-GEN’} & \text{rUUčAti}^{\text{IMP/PERF}} & \quad \text{‘have lunch’} \\
\text{vEčEr-a} & \quad \text{‘dinner-NOM’} & \text{vEčErati}^{\text{IMP/PERF}} & \quad \text{‘have dinner’}
\end{align*}
\]

This is also true of most imperfective verbs, as illustrated in (139).

(139)

\[
\begin{align*}
\text{kOp-A} & \quad \text{‘excavation-GEN’} & \text{kOpAti}^{\text{IMP}} & \quad \text{‘dig’} \\
\text{rAt-a} & \quad \text{‘war-GEN’} & \text{rAtovati}^{\text{IMP}} & \quad \text{‘be at war’}
\end{align*}
\]

While the tonelessness of -ova- holds of virtually all examples we have found in Eastern S-C, in which -ova- is part of the inter-language mapping, we have found quite a few examples in which -ova- is stressed in Western S-C, which is most likely one of the reasons why it has not become part of the integration mapping (cf. e.g. Western kupOvAti^{IMP} vs. Eastern kUpovati^{IMP} ‘to buy’).

Another advantage of the two suffixes, from the perspective of LC, is that they do not cause any segmental alternations on the stems, also in their native use. Moreover, neither of the suffixes seem to impose any constraints on stems they can combine with, which makes them a suitable candidate for an ever-applicable mapping. This last fact seems to be a relatively recent innovation in the domain of -ova-. In native derivations, this suffix has two allomorphs, –ova– and –eva-, which are in a complementary distribution guided by the soft/hard distinction in the previous consonant (roughly corresponding to palatal/non-palatal), as illustrated in (140).

(140) -eva-/ -ova- allomorphy in the native lexicon

\[
\begin{align*}
\text{kralj-evati}^{\text{IMP}} & \quad \text{‘rule as a king’} & \text{bič-evati}^{\text{IMP}} & \quad \text{‘whip’} \\
\text{trg-ovati}^{\text{IMP}} & \quad \text{‘trade’} & \text{drug-ovati}^{\text{IMP}} & \quad \text{‘be friends’}
\end{align*}
\]

As the examples show, the ‘soft’ stems take -eva-, while the ‘hard’ stems take -ova-. Interestingly, the same distinction is still maintained in the plural augment in nouns, even in recent borrowings.
Aspect of borrowed verbs

(141) -ev-/-ov- allomorphy in the plural augment

<table>
<thead>
<tr>
<th>English Stems</th>
<th>S-C Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>rikvest-ov-</td>
<td>skrin-ov</td>
</tr>
<tr>
<td>riplaj-ev</td>
<td>bedž-ev</td>
</tr>
</tbody>
</table>

However, as can be seen from the data in (142), -ova- is insensitive when combined with English verbal stems.

(142) Generalised -ova- with English stems

<table>
<thead>
<tr>
<th>English Stems</th>
<th>S-C Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>rikvest-ova-IMP/PERF</td>
<td>skrin-ova-IMP/PERF</td>
</tr>
<tr>
<td>riplaj-ova-IMP/PERF</td>
<td>ingejdz-ova-IMP/PERF</td>
</tr>
</tbody>
</table>

The source of this difference between the verbs and the nouns arguably lies in the form onto which the initial surface form gets mapped: the plural augment is not part of the citation form, so that the pluralisation takes place ‘deeper’ within S-C.

As for the generalisation of the verbaliser -ova- to all phonological contexts in borrowing, we maintain that this innovation preceded the stabilisation of the inter-language mapping for English. The evidence is that -eva- did not combine with older borrowed verbal stems either. Moreover, we found at least five verbs from earlier contact with German (and/or French), in which -ova- is combined with a ‘soft’ base. Tellingly, all of these verbs but one are integrated with a suffix other than -ova- into Western S-C, which indicates that the versatility of -ova- is much less prominent in this dialect.

(143)

<table>
<thead>
<tr>
<th>East</th>
<th>West</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>klIc-ova-IMP</td>
<td>kibic-IIrAti IMP</td>
<td>kiebitzen</td>
</tr>
<tr>
<td>švErc-ova-IMP</td>
<td>švErc-atI IMP</td>
<td>schwärzen</td>
</tr>
<tr>
<td>šAc-ova-IMP</td>
<td>šAc-atI IMP</td>
<td>schätzen</td>
</tr>
<tr>
<td>Ilnč-ova-IMP</td>
<td>Ilnč-ova-IMP</td>
<td>lychen</td>
</tr>
<tr>
<td>AngAž-ova-IMP</td>
<td>angaž-IIrAti IMP/PERF</td>
<td>engagieren</td>
</tr>
</tbody>
</table>

10.2.2. Empirical evidence for the prevalence of -ova- in Eastern S-C

In the nonce verb derivation experiment which we will report here, we tested how strong the inter-language mapping is, and if it is strong as it appears to be, whether it also is some kind of default in native derivations.

Participants. 60 native speakers from Serbia, high-school and university students were enrolled in the experiment, of both genders.

Materials. The stems from which the verbs were to be derived were five English verbs (attach, forward, like, refresh, share) and five S-C nouns (dijamant ‘diamond’, koliba ‘hut’, lavabo ‘washbasin’, lutka ‘doll’, papir ‘paper’).

Procedure. Participants were asked to fill in a verb derived from a given stem
in 20 sentences (10 perfective and 10 imperfective contexts). Each stem was to be used in two sentences, one perfective and one imperfective. There were also 50 sentences with other tasks, which can be considered fillers.

Results. In the experiment, we obtained 449 nonce derived verbs based on English verbs and 552 nonce derived verbs based on S-C nouns. A comparison was performed between the verbs derived from English verbs and those derived from native nouns. This comparison was along two dimensions: the choice of the suffix and the aspect of the nonce derived verb. The distribution of suffixes used by the subjects in this task is shown in graphs in (144) and (145).

(144) Affixes in Serbo-Croatian verbs derived from English verbs (N=449).

(145) Affixes in Serbo-Croatian verbs derived from Serbo-Croatian nouns (N=452).
The two graphs show rather different distributions. The highly skewed distribution in the graph in (144) indicates that -ova- is virtually the only suffix used for English verbs. On the other hand, the rather even distribution in the graph in (145) testifies to the non-existence of a generalised pattern for the derivation of verbs within S-C. Consequently, -ova- is just one of the many suffixes used for derivation of verbs from Serbo-Croatian nouns.

In terms of our question, -ova- is clearly prevalent in the integration mapping, but it shows no signs of being the generalised verbaliser in S-C. Note also that the suffix which is part of the mapping in Western S-C, -a- (alone or combined with the agentive suffix -ar-) accounts for 29% of all native nonce derivations. This indicates that the LC-friendly properties of the suffix -a- are clearly available to Eastern S-C speakers as well. Still, our results show that S-C has no generalised default verbaliser.

As for the aspect of the derived verbs, in the whole data set there was one case of a prefixed verb based on an English verb – izatačovati\textsuperscript{per} ‘attach’, less than 1% of the whole set. The lack of prefixes in borrowed verbs leads to unspecified aspect. Recall that adding a prefix to the basic form changes verb aspect from the default imperfective to perfective, and that contrary to what happens in native verbs, borrowed verbs typically do not get prefixed. Without a prefix, the imperfective vs. perfective opposition cannot be established. Without this opposition, there is no need to specify the default aspect as imperfective. Therefore, aspect in borrowed verbs remains unspecified and the same form is used in both perfective and imperfective contexts.

As for the nonce derivations from S-C nouns, while the use of -ova- is not particularly common, this suffix is predominantly unspecified for aspect in these derivations as well. This is illustrated in graph in (146), which focuses on the 43 native -ova-derivations in perfective contexts, in which the form without a prefix is by far the most frequent.

(146) Serbo-Croatian verbs in -ova- from Serbo-Croatian nouns: Only perfective contexts (N=43).
At the other end of the spectrum, when the native bases were combined with one of the native suffixes (other than -ova-), prefixed forms occurred in 75% of the perfective sentences. This means that prefixation is very much alive, especially in the domain where native suffixes which play no role in borrowing apply. This means that while the suffix -ova- is undoubtedly a Slavic-origin suffix, the aspect of its derivations is importantly influenced by its role in borrowing.

In sum, we found a very strong domination of the described inter-language mapping in Eastern S-C. However, it should be noted that the other integrating suffix, -ira-, was used nine times, for different English items (refresh, share and forward) and by five different subjects. This suffix leads to unspecified aspect as well. Additionally, the form forwardirati seems to be lexicalised in Western S-C (Drljaća Margić 2011). This all points to the conclusion that -ira- is not entirely inactive in the current language contact and it cannot be disregarded in the dynamics of verbs with unspecified aspect in S-C. In §6.4.1. we argued that there has been a period when -ova- took over from -ira- because of its LC-friendly properties. In the following section, we are moving back in time, in the direction of this turning point, discussing the layer of verbs borrowed from English, partially obeying the previous mappings, which were mainly used in the context of verbs which have multiple source forms in different European languages.

10.3. Other integration suffixes: -ira-, -isa-, -a- and -nu-

In this section, we discuss the remaining four suffixes which emerge from Vasić et al. (2001), a dictionary of new anglicisms. Two are borrowed, German-origin -ira- and Greek-origin -isa-, while the other two are native: -a- and -nu-. The dictionary covers approximately 950 lemmata, collected from a corpus from 1998 and 1999. The authors state that this dictionary includes the words which have entered ‘Serbian’ in previous 30 years or so, or have acquired new meanings in this language. The distribution of suffixes in the 57 verbs in this dictionary is shown in (147).

Out of the 57 verbs in Vasić et al., only 38 have a clear corresponding English verbal base. The use of suffixes in these words is shown in (148).

(148) Suffixes in Vasić et al. (2001) – Conservative count  N=38

These two graphs show a less skewed distribution than that in (144), but still testify to a clear prevalence of -ova-. While the second position is still occupied by -ira-, there emerge three other suffixes, each represented by one

30 The rest of the verbs are those in which one part of the verb has been translated (e.g. pre-bukirati ‘overbook’), those which have a clearly nominal base (e.g. bebisitingovati ‘babysit’, piraterisati ‘pirate (copyrighted materials)’) or a clearly adjectival base (e.g. grogirati ‘stupefy’), as well as those whose meaning reveals that they were created within SC based on loanwords from English (e.g. bildovati ‘practice bodybuilding’).
item in the second graph. These three suffixes are -isa- (spondorisati\textsuperscript{IMP/PERF} \textit{sponsor}), -a- (printati\textsuperscript{IMP} \textit{print}) and -nu- (kliknuti\textsuperscript{PERF} \textit{click}). Importantly, for the latter three verbs it is possible that they were based on S-C borrowed nouns \textit{print}, \textit{sponzor} and \textit{klik}, rather than on English verbs. In sum, even these older and more extensive data indicate that the set of unquestionable integration suffixes is restricted to -ova- and -ira-.

In the following we consider each suffix, its properties and the level of productivity which it displayed in our nonce derivation experiment and other recent derivations. However, before we turn to individual suffixes, a brief note is due on the background against which early English borrowings with unspecified aspect were integrated into S-C. This background is formed by a rather populated layer of the lexicon which arose through a different type of borrowing. This layer contains verbs which are composed of international (typically Latinate) stems and the familiar integrating suffixes. These verbs also have predominantly unspecified aspect. Together with -ova-, the suffixes -isa- and -ira- are the common integrating suffixes of this layer of borrowed verbs. In (149), we repeat the examples of such verbs which we had given in §5.2.1.

(149)

–ISA–: formUlIsati \text{‘formulate’}, fotogrAfIsati \text{‘photograph’}, defInIsati \text{‘define’}, rezErIsati \text{‘reserve’}, tolErIsati \text{‘tolerate’},
–IRA–: anketIIrAti \text{‘survey’}, delegIIrAti \text{‘delegate’}, diplomIIrAti \text{‘receive a bachelor’s degree’}, formIIrAti \text{‘form’}, distancIIrAti \text{‘distance’},
–OVA–: orgAnIzovati \text{‘organise’}, prOtEstovati \text{‘protest’}, kAndIdovati \text{‘to put forward as a candidate’}, dlIrIgovati \text{‘conduct’}.

Note that these examples are representative for Eastern S-C. In Western S-C, the preferred forms are composed of the same stems and -ira- in all cases.

Identifying the source form of the verbs from this layer is not a trivial issue, since they are virtually always relatable to a number of sources: typically German and French, but it is safe to assume that the educated S-C speakers who were introducing them were familiar with Latin vocabulary as well. As Neikirk Schuler (1996: 174) stated, these verbs were borrowed in “a context in which there [was] no dominant-language interference”. The very few Eastern/Western S-C pairs which use different foreign stems reveal that the process of extracting the foreign stem from the ‘international’ input could have different outcomes. The Eastern \textit{kritik-ovati} vs. the Western \textit{kritiz-irati} is a good example. This same pair exemplifies the fact that both foreign
stems (kritik- and kritiz-) can be found elsewhere in common S-C vocabulary – in words like kritik-a ‘criticism’ and kritizertvo ‘petty criticism’.

The same process of lexical enrichment from this general European lexical pool has affected all other Slavic standard languages as well, with some of which (Slovenian, Macedonian, Russian, Czech) speakers of S-C have had some level of contact. It is therefore not surprising that verbs from this layer have cognates with unspecified aspects in various Slavic languages (Comrie & Corbett 1993 quote examples of such verbs from Slovenian, Bulgarian, Macedonian, Czech, Slovak, Polish, Belorussian and Ukrainian). Serbo-Croatian studirati ‘study’, telefonirati ‘make a phone call’ and aplicirati ‘apply’ have cognates in other Slavic (but also Germanic and Romance) languages. In sum, verbs of this layer cannot be analysed as borrowed from any one language and, as such, they do not strictly have a ‘source form’. Rather, they are parts of networks of verbs which can be established with virtually any modern Slavic, Germanic and Romance language. Notice that while there is a clear distinction between the verbs we have quoted in (149) and the very recent English of the type rikvest(ov)ati, some verbs cannot be easily classified because, while having an English base they might have entered S-C through another language (most plausibly German) and via S-C-internal formation. Example of such verbs are linčovati, džogirati, parkirati etc.

As for their aspect, all the verbs in (149) can be used both as perfective and as imperfective. However, for quite a few of them there are derivations with native prefixes which can only be perfective (e.g. pre-formulisati\textsuperscript{PERF} ‘rephrase’, etc.). This has been spotted already by Lazić (1976), who states that the “use of native prefixes with borrowed verbs is expanding rapidly, and contemporary dictionaries cannot keep abreast of the situation” (p.58). And yet, while prefixation is possible in many examples, we have found very little evidence that base verbs (stem+suffix) are acquiring a single aspect. So the prime examples used by Lazić anketI\textsuperscript{PERF}IrA\textsuperscript{ATI} and orgAnIzovati (frequently prefixed to izanketI\textsuperscript{PERF}IrA\textsuperscript{ATI} and izorgAnIzovati) are still both perfective and imperfective. Moreover, izanketI\textsuperscript{PERF}IrA\textsuperscript{ATI} and izorgAnIzovati still sound rather colloquial and dispreferred in standard S-C.

In sum, there are quite a few differences between this layer of international borrowed verbs and the one we have presented in the previous section. We summarise them in (150).

(150) English mapping International vocabulary
a. one dominant suffix three suffixes
b. aspect strictly unspecified aspect typically unspecified
c. prefixation blocked prefixation possible
d. falling accents frequent falling accents rare
(even more so in Western S-C)
The first feature can be analysed as the consequence of the type of contact – it is only in the case of language contact with a reconstructible initial surface form that a single specialised integration suffix needs to be introduced. As we have seen above, there is no such form, but rather an array of related forms for the “international” layer. The features (b) and (c) are clearly related. The fact that prefixation is virtually the only available mechanism for marking the perfective aspect means that the key to the impossibility of aspect marking is in the impossibility of prefixation. The main cause of the latter is, as we have argued, in LC. The feature (d), which we have not discussed yet, may play an important role in LC’s blocking of prefixation in novel items. As we have already pointed out in §6.4.1., -ira- and -isa- are always stress-attracting, so that they will always introduce H’s away from the initial syllable, which then leads to rising accents (as explained in the following chapter). This has one important consequence for LC: a rising accent implies that the prosodic pattern of the stem remains the same under prefixation (anketIrrati - izanketIrrati). This difference is almost categorical in Western S-C (where, as we have mentioned it is the difference between the stress-attracting -ira- vs. the recessive -a-). In Eastern S-C, the difference is in the fact that -isa- and -ira- are always stress-attracting, but also in the fact that new -ova-verbs rarely have rising accents for reasons which we will turn to in the next section.

In conclusion, the source of differences between the two mappings we have compared is that the more recent one, specialised for English items is, unlike the former, under more direct control of LC, which is expected, given the contact type.

10.3.1 -ova-

Since this suffix has been in the focus of our attention in §10.2., we have little to add to the analysis here. The only important fact which we have not yet addressed is the difference in the dominant prosodic pattern in the recent English borrowings as compared to that in the layer of international verbs. This difference is not a consequence of the properties of -ova- in the two mappings, but of the fact that English stems can have contrastive stress and a rather prominent difference in the average stem length in two mappings. We illustrate the possible prosodic patterns with monosyllabic and disyllabic bases in (151).

(151)    International mapping  English mapping
  a. o   rImovati ‘rhyme’  lOgovati ‘log’
  b. oo  dIrIgovati ‘conduct’  EdItovati ‘édit’
        rikvEstovati ‘requést’
Monosyllabic stems always have initial falling accents (151a). International disyllabic stems always have rising accents, with the stress on the first syllable, as do the English stems with initial stress. English disyllabic stems with final stress have a falling accent on the last syllable – a constellation is unattested in native and international stems. As argued in §6.4.1., the unattestedness of such a constellation in international stems should not be interpreted in terms of a ban on certain patterns, since the source languages (French verbs in -er and -ir and German verbs in -ieren) typically did not have stressed verb stems.

If we consider these data from the perspective of possible prefixation and LC, verbs like dIrIgovati and EdItovati have no prosodic reason not to combine with a prefix: their prefixed forms izdIrIgovati and izEdItovati do not change the stress position. The examples of the type rikvEstovati seem rather difficult to combine with prefixes, but the form *izrikvEstovati would probably have this prosodic pattern.

Finally, an important (statistical) role seems to be played by monosyllabic stems. Due to the general S-C ban on word-internal falling accents, when the prefix adds a syllable, these stems need to change the prosodic pattern and move stress to the initial syllable, as can be seen from forms like IzrInovatiperf ‘rhyme’ and IzlOgovatiperf ‘log out’. Note the stark contrast between the tolerated falling accent in rikvEstovati and the obligatorily repaired one in *izlOgovati, which matches our intuition that a prefixed form is less protectable by LC.

The fact that monosyllabic stems by definition introduce stress-shift when prefixed means that in a borrowing situation in which many stems are monosyllabic, LC will cause high resistance to prefixation. As it turns out, the English verbs which enter S-C are most frequently monosyllabic: out of the 26 -ova-verbs based on English verbs which we isolated from Vasić et al. (2001) in (148), 21 have monosyllabic stems.

10.3.2. -ira-

The suffix -ira- is the second integration suffix, and the fourth most frequent suffix in the native nonce derivation experiment (with 19%, see 145). As for prefixation, being virtually constrained to foreign stems and biaspectral verbs, -ira- outperforms even -ova- in defying prefixation in perfective contexts (albeit by only 2%, see 146).
A final note concerning the prosody of -ira-verbs is due. So far, in discussing why -ova- and -a- have taken over the integration mapping, it was important that -ira-, but not -ova- or -a-, deletes the prosodic pattern of the initial surface form/stem. While this indeed holds in all items, for completeness’ sake we should point out that -ira- has another shortcoming from the perspective of LC: it is accented in the infinitive form (e.g. *anketIIrAti* ‘survey’), but it is pre-accented in most forms in the present tense (*ankEtIIraam* ‘I survey’). This means that extra violations of LC constraints are incurred by the existence of different allomorphs. This fact may be useful in explaining why -ira- also lost in Western S-C (to -a-) where it was virtually the only integration suffix in the international stratum.

10.3.3. *-isa-*

The suffix *-isa-* has its origin in Greek, more precisely in “the aorist form of the Greek denominal verbs in -izō, the aorist in -isa” (Skok 1955: 37). This suffix is currently virtually inactive in verb borrowing (although it surfaces in recent creations from English-origin nouns, e.g. *didžEjIsati* ‘be a DJ’). It is, however, the second most productive suffix in the native derivations, where it accounts (alone or in combination with the agentive suffix -ar-) for 24% of all nonce native derivations. The inactivity of *-isa-* in verb borrowing seems to be reflected in the fact that this suffix is considerately more often co-occurs with prefixes in perfective contexts.
(153) Serbo-Croatian verbs in -isa- from Serbo-Croatian nouns: Only perfective contexts (N=52).

As Skok (1955) shows, -isa- was the main integration suffix in contact with Turkish (which predominantly affected Eastern S-C). Given the fact that Turkish verb stems do not have a stress contrast, it comes as no surprise that -isa- has dominant prosody: it is pre-accenting (e.g. didžEjIsati 'be a DJ', didžEjIšeem 'I am a DJ').

10.3.4. -a-
The suffix -a-, the prevalent suffix in verb borrowing in Western S-C, shows very few signs of being active in verb borrowing in Eastern S-C. However, as we have already pointed out in §10.2.2., it is the single most productive suffix in nonce native derivations in our experiment. Its inactivity in the domain of borrowing predicts that it will also not block prefixation, which is indeed the case, as shown in (154).

(154) Serbo-Croatian verbs in -a- from Serbo-Croatian nouns: Only perfective contexts (N=58)
As we already pointed out in §10.2., the high productivity that this suffix displays in nonce native derivation testifies to its LC-friendly recessive prosody. This also explains why some of the neologisms we quoted in §4.2.1. and analysed as based on nouns make use of this suffix (\textit{gUglAti}^{IMP/PERF} ‘google’, \textit{snIfati}^{IMP} ‘sniff (drugs)’).

\textbf{10.3.4. -nu-}

The suffix -\textit{nu-} is different from all the suffixes which we have discussed so far, in that its occurrence in descriptions and in our experiment is limited to very few items. Vasić et al. (2001) find only \textit{kliknuti}^{PERF} ‘click’ (related to the noun \textit{klik} ‘click’), Lazić (1976) finds only \textit{šutnuti}^{PERF} ‘to kick’ (related to the noun \textit{šut} ‘kick, shoot’) and \textit{maršnuti}^{PERF} ‘to put (someone) out at once’ (related to the exclamation \textit{marš} ‘get out’). (The latter item is marginal in the author’s idiolect; instead, the prefixed \textit{iz-marš-irati}^{PERF} is used). These three verbs have stable imperfective counterparts with familiar suffixes \textit{klik(ati)}^{IMP}, \textit{šutirati}^{IMP(East)}/\textit{šutati}^{IMP(West)}, \textit{marširati}^{IMP}. Our nonce derivation data show a single occurrence of this suffix - \textit{dijamantnuti}^{PERF}.

It may come as a surprise that the existence of a suffix which encodes the perfective aspect is not readily put to use by contact patterns. As a matter of fact, there is anecdotic evidence that this could have been the case in certain contact contexts. For instance, the title of the comically relexified text we have quoted in §6.3.2. is repeated here in (155), with all the English roots underlined.

\begin{enumerate}
\item[(155)] \textit{WRAJTUJ KAO SHTO SPIKUJESH, RIDUJ KO SHTO JE WRAJTNUTO} \textit{SC: Piši}^{IMP} kao što \textit{govoriš}^{IMP}, \textit{čitaj}^{IMP} k(a)o što je \textit{napisano}^{PERF}.
\item \textit{E: Write as you speak, read as is written.}
\end{enumerate}

The original S-C sentence is the most common formulation of the famous ‘phonetic principle’ on which the S-C spelling (especially the Cyrillic script) is based. It is important to note that the S-C original displays the aspect contrast between \textit{pisati}^{IMP} and \textit{napisati}^{PERF}. The nonce imperfective verbs created by the relexifier are \textit{wrajt-ovati}^{IMP}, \textit{spik-ovati}^{IMP} and \textit{rid-ovati}^{IMP}, while the perfective verb is \textit{wrajt-nuti}^{PERF}. This seems like the optimal solution to the relexification task to S-C speakers (including the author). In such a task, it is important to encode the aspect, as it is already encoded in the original, and a single unambiguous exponent of the perfective aspect, -\textit{nu-}, is preferred. However, borrowing is different from relexification, and the aspect contrast, even if perceived, is not the priority for encoding.

While this answers the question of why -\textit{nu-} is not part of an inter-language mapping, it still does not explain why -\textit{nu-} is virtually excluded from perfectivisation of borrowed verbs within S-C. The reason for this is the
specific meaning of this suffix – it is semelfactive, which means that, apart from being perfective, it is also atelic and punctual, or, as Lazić (1976) stated, “indicating a single or instantaneous action” (p. 51). This meaning was compatible with the meanings of a few specific verbs, such as ‘click’ and ‘kick’.

Due to its restricted meaning and few combinability restrictions, -nu- displays what Simonović & Samardžić (2013) termed ‘covert productivity’: many verbs with -nu- are attested in S-C, but unrecognised as existent verbs by speakers and avoided in contexts in which there is higher meta-linguistic awareness. Such forms are illustrated in (156) below. Most speakers we have consulted do not consider these ‘actual’ verbs when they are presented without a context. Moreover, if speakers who have acquired the concept of aspect in formal education are asked about the perfective counterpart of the verbs in (156a), they will readily offer prefixed forms napisatiPERF, odsviratiPERF and poljubitiPERF.

(156)

<table>
<thead>
<tr>
<th>a.</th>
<th>pis-a-tiIMP</th>
<th>svir-a-tiIMP</th>
<th>ljub-i-tiIMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>'write'</td>
<td>'play'</td>
<td>'kiss'</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>pis-nu-tiPERF</td>
<td>svir-nu-tiPERF</td>
<td>ljub-nu-tiPERF</td>
</tr>
<tr>
<td>'write briefly/once'</td>
<td>'sleep briefly/once'</td>
<td>'kiss briefly/once'</td>
<td></td>
</tr>
</tbody>
</table>

The relevance of these facts for loanword data is that there are signs of productivity of -nu- in loanwords as well, as examples in (157) show.

(157)

<table>
<thead>
<tr>
<th></th>
<th>esemesovatiIMP</th>
<th>hakovatiIMP</th>
<th>tvitovatiIMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>'text'</td>
<td>'hack'</td>
<td>'tweet'</td>
<td></td>
</tr>
<tr>
<td>esesnuni</td>
<td>haknuti</td>
<td>tvitnuti</td>
<td></td>
</tr>
<tr>
<td>'text briefly/once'</td>
<td>'hack briefly/once'</td>
<td>'tweet briefly/once'</td>
<td></td>
</tr>
</tbody>
</table>

In sum, while -nu- combines with a variety of verbal stems, due to its specific meaning and, paradoxically, its combinability with virtually any stem, speakers do not perceive -nu-affixation as lexical derivation.

**10.4. Conclusions**

In this chapter, we have considered the borrowing of verbs into S-C, with the main focus on the current mapping for English verbs, which yields verbs with unspecified aspect. We have seen that S-C integration suffixes which are used for incoming verb stems also display productivity in S-C-internal
derivation. This is why much of the discussion was about the extent to which S-C-internal derivation and borrowing are different.

We have found evidence that there is a clear distinction between S-C-internal derivation and verbalisation for borrowing purposes: it is only for the latter that a stable pattern has been established. In other words, the only truly productive and maximally predictable verbalising pattern is the borrowing pattern, whereas in S-C-internal derivations many patterns appear and interact. We interpreted this finding in terms of the necessity for a clear and general-purpose pattern for the inter-language mapping. And yet, while the two uses of suffixes (SC-internal derivation and borrowing) should be kept apart conceptually, we also found that the extent to which a suffix is active in borrowing predicts how often it will co-occur with prefixes, which are so repelled by LC: the integrating suffixes -ova- and -ira- almost never allow prefixation, -isa- allows it relatively often, while all other suffixes co-occur with prefixes in most perfective contexts.

Apart from the global conclusion, we have addressed at least two specific topics on which we hope more research will be done in the future: variation between borrowing mappings in polycentric standards and productivity.

The first issue needed to be addressed due to the encountered systematic variation in the two standard dialects of S-C. While our focus was on Eastern S-C, we found that the suffix used in the current mapping in Western S-C, -a-, is the one most frequently used in native nonce derivation by Eastern speakers (but virtually never in borrowing). We have also been able to identify at least two non-categorical differences between Eastern and Western S-C which made it more probable for the suffix -ova- to become part of the mapping in Eastern S-C only. The first is the fact that -ova- can be stressed (and therefore dispreferred by LC) in Western S-C and the second is the fact that the previous layer of international verbs was dominated by another suffix, -ira-, in Western S-C, which left -ova- not only largely constrained to native stems, but also to stems which end in the traditional 'hard' consonants. These factors made -a- a much stronger candidate.

Finally, we have only briefly addressed the atypical S-C suffix -nu-, characterised by covert productivity: being readily combined with different verbal stems, but somehow remaining under the radar of lexicalisation. This suffix remains out of the picture on the one hand (in the sense that it barely derived any actual lexemes), but is always already in the picture on the other: the existence of a verbal base virtually implies the possibility of a nu-derivation with a very specific meaning.
Chapter 11
The importance of not belonging: Paradigmaticity and Latinate nominalisations in Serbo-Croatian

In this chapter, the integration of Latinate nominalisations into Serbo-Croatian is considered. The main concern will be the status of the suffixes which seem to be borrowed into the language, in light of the fact that our model is based on the assumption that the initial surface form is morphologically opaque.

In §5.4.1. we have discussed the fact that in a number of Slavic and Germanic languages various Latinate derivational affixes and morphological patterns are relatively common. These affixes and patterns have effects as abstract as that of deriving event nouns from verbs and property nouns from adjectives. By comparing the Latinate nominalisation suffixes to their apparent native competitors in Serbo-Croatian and Dutch, it was shown that the former are not only much more restricted in their productivity, but also have much less transparent semantics. We concluded, based on this rather limited sample, that Latinate nominalisations do not contradict the generalisation at the core of our model that loanwords get borrowed as morphologically opaque forms with a specific meaning.

In this case study, we discuss the status of two Serbo-Croatian nominalising Latinate suffixes, -cija and -itet, which display relatively high values of productivity and frequency. As our analysis will show, these are not borrowed suffixes and derivational patterns in the sense that it is not the case that they were present in another language and consequently got copied into S-C. Rather, these suffixes and patterns emerged within S-C, and more specifically within the Latinate stratum of the S-C lexicon. In other words, we argue that members of the two classes of nominalisations were first borrowed as simplex words. Since sequences -cija and -itet display quantitative patterns characteristic of the productive native suffixes, we argue that these quantitative effects, in conspiracy with the shared semantic properties of the nouns ending in these sequences, have led to a reanalysis of the initial simplex borrowed forms as derived nouns involving a stem and the suffix -cija or -itet, respectively. While recognizing -cija or -itet as suffixes, we

---

31 This case study is an adapted version of the paper submitted to Arsenijević et al. (forthcoming), which I have co-authored with Dr. Boban Arsenijević from the Universities of Niš and Postdam. The parts of the original paper omitted here covered the quantitative measures of the productivity of the suffixes under inspection, especially when they are combined with native stems. While these aspects of the phenomena are important for understanding the synchronic picture of S-C, they are not directly relevant to this dissertation.
shall also show how the reconstructed lexical relatedness that these nouns have with verbs and adjectives is rather different from the relations between native nominalisations and their bases. In showing how Latinate nominalisations differ from the native ones, we will be making use of the concept of paradigm, defined as the domain of maximal productivity, phonological correspondence with the base and semantic transparency. While the domain of native nominalisations will be shown to be organised around paradigmatic nominalisations, Latinate nominalisations will be shown to defy joining paradigms of related verbs and adjectives. The emergent picture will be that of a layered lexicon in respect of morphological patterns, and complementary with but not equivalent to the stratified phonological lexicon discussed in Chapter 3. The periphery, which reconstructed morphological patterns join, is characterised by limited productivity, little semantic transparency and a lack of a straightforward phonological base shared by the nominalization and the related adjective/verb.

In sum, this final case study adds to the picture of the lexicon in contact by situating the morphological patterns established in the borrowed domain with respect to the rest of the lexicon. This case study is also important for introducing data from a contact situation in which there is no clear source language. Rather, the speakers which have introduced these words into standard S-C (established in the second half of the 19th century) based their representations on a network of related words from German, French and Latin (as well as in some cases Russian, Italian, English, etc.). In this property the Latinate nominalisations are similar to the Latinate verbs which were discussed in §10.3. and which, as Neikirk Schuler (1996) stated, were borrowed in “a context in which there is no dominant-language interference” (p. 174).

11.1. Abstract morphology is not typically subject to borrowing

It is a common observation in the literature on language contact that content words are much more easily borrowed than function words, and that free morphemes are much more easily borrowed than bound morphemes (Chapter 1, Backus & Versschik 2012). This is related to another (formulation of the same) observation, which we have foregrounded in Chapter 4: paradigms are generally not borrowable. As we have seen, in most language contact situations, there is a single form which serves as the base of the whole new lexical item in RL. This general observation also matches the finding that the initial stage of borrowing – code-switching – tends to target semantically
specific items (Backus 2001), also incorporated in our model. Since bound morphology typically has general and abstract meaning, it is expected not to be targeted by code-switching.

Finally, the general, abstract meanings expressed by bound morphology and their grammatical effects (e.g. of nominalising or adjectivising affixes) are typically covered by some native mechanism, so borrowing such bound morphemes would lead to multiple realisations of the same semantic content and the same set of syntactic features, taking place in the same general context. This makes the borrowing of such patterns additionally unexpected on grounds of both economy and introducing optionality into the system.

11.2. ... except everywhere

However, as we have seen, many if not all European languages seem to present blatant cases of exactly the situation we just described as excluded. Modern Slavic and Germanic languages have developed Latinate lexical strata with derivational patterns which hold within the stratum and which yield semantic effects as abstract as that of nominalization, and often in apparent competition with the native derivational pattern.

(158) Latinate and native nominalisation patterns in (apparent) competition

<table>
<thead>
<tr>
<th>Native</th>
<th>Latinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
</tr>
<tr>
<td>absurd-ness</td>
<td>absurd-ity</td>
</tr>
<tr>
<td>socialis-ing</td>
<td>socialis-ation</td>
</tr>
<tr>
<td>Dutch</td>
<td></td>
</tr>
<tr>
<td>absurd-heid</td>
<td>absurd-iteit</td>
</tr>
<tr>
<td>socialiser-ing</td>
<td>socialis-atie</td>
</tr>
</tbody>
</table>

The main goal of this chapter is to explore the status of Latinate nominalisations by focusing on their behaviour in Serbo-Croatian. The S-C reflexes of the Latin nominalising suffixes -itas (~ English -ity) and -tio (~ English -tion) are -itet and -cija respectively. Importantly, S-C nouns in -cija do not only come from the Latin suffix -tio, but also from the later Latin deadjectival nominalising suffixes -tia, -cia, and -ia which were combined with stems ending in -nt.

(159) Later Latin nominalisations in –aacia, -aatia and –(t)ia (Miller 2006: 36-37)

konspira-cija, from Lat. conspira-acia    arogan-cija, from Lat. arrogant-(t)ia
conspire-cija      arrogant-cija 
‘conspiracy’        ‘arrogance’
(cf. konspirativ-n-a, arogant-n-a
conspative-Adj-FemSg, arrogant-Adj-FemSg)
The question we are aiming to answer is whether S-C Latinate nominalisations instantiate the process of borrowing derivational suffixes as abstract morphological items and structural patterns, and, relatedly, whether, or to what extent Latinate derived nouns can be considered part of the paradigms of related adjectival and verbal stems.

We argue that although Latinate nominalisations are ‘real’ in the sense of being related to their bases, the Latinate patterns are not strictly borrowed, but emergent in the recipient lexicon. As we will show, lexical mechanisms prevent the emergent Latinate patterns from competing with the productive native nominalising suffixes by integrating the former as markedly non-paradigmatic.

Note that our use of the term paradigmatic is somewhat non-standard, although it has quite a few precedents. Haspelmath (1995: 47) identifies the forms derived from one stem by means of inflection with the paradigm of the stem. He argues that derivation results in the emergence of a new stem, which, if subject to inflection, generates its own paradigm. The properties that he lists as criteria for the identification of inflectional (i.e. paradigm-generating morphology) are regularity, generality and productivity. He does not delve much deeper into the characterisation of these three properties or into how they are tested, measured, and what their precise definitions are. In the same vein, Spencer (2013) presents an elaborate paradigm-driven theory of lexical relatedness. Both authors largely ignore prosody, which will crucial for our discussion.

In this chapter, we do not identify paradigms with inflection, but we do define them using similar properties. We also leave out the discussion of the usefulness of the distinction between inflection and derivation altogether due to space limitations. In a nutshell, our view is that the traditional boundary between inflection and derivation does not coincide with that between paradigmatic and non-paradigmatic relations: the paradigmatic space takes up all that is considered inflection and a good part of what is considered derivational morphology.

We borrow the definition of the paradigm from Arsenijević & Simonović (2013), where it is the set of all the words productively and compositionally derived from a certain stem, irrespective of the divide between inflection and derivation. This means that in order to be part of a certain paradigm (i.e. to be paradigmatic), an item needs to be formed from the exact stem that the paradigm is based on and be semantically transparent in respect of the contribution of the stem, the other morphemes present (typically affixes) and the structure of the derivation. In this view, for instance, a deverbal nominalisation pattern produces nominalisations which are members of the
paradigms of a class of verbs if these nominalisations are productively and compositionally derived from that class of verbs. On the other hand, a deverbal nominalisation is not a member of any verbal paradigm if it idiosyncratically targets arbitrary verbs or arbitrary members of a verb class, and/or if its semantics tends to undergo lexicalisation and appears idiomatic (a sketch is presented in §11.3., and Arsenijević & Simonović 2013 give a detailed overview of deadjectival nominalisation in this perspective). This view yields a hierarchical organisation of morphology in terms of paradigms: the entire set of declension forms of a deverbal nominalisation (inflection) is naturally part of the paradigm of the nominalisation itself, but it is also true that in case the nominalisation satisfies the above conditions (paradigmatic derivation), this same set is part of the paradigm of the verb (for illustration, see (165) below).

We draw our empirical material from Serbo-Croatian, which, already in the native stratum, displays various morphological and phonological reflexes of paradigmaticity (as illustrated in (160) and (161) below). The main diagnostic is the prosody of the two types of nominalisations, the paradigmatic and the non-paradigmatic ones. Paradigmatic nominalisations surface with the elsewhere attested allomorphs of the base, whereas in the non-paradigmatic nominalisations the base surfaces with a new allomorph, which appears to be imposed by the prosody of the suffix. As we will show, the same type of prosodic distinction obtains between the native and the Latinate nominalisations. The Latinate nominalising suffixes are always accented and dominant – they have a lexical tone which deletes whatever prosodic information the base may have contributed. The native nominalising suffixes, on the other hand, present a continuum whose most common (and most paradigmatic) exponents illustrate the other extreme: they have no prosody to impose on the base (i.e. they are unaccented and recessive), so that even when the base is toneless, the nominalisation displays the prosodic pattern with which the base surfaces on its own.

The rest of this chapter is organised as follows. §11.3. provides an analysis of the native nominalisation patterns in S-C. A clear contrast is observed between the productive suffixes, which have a transparent meaning and unaccented, recessive prosody – thus we view them as paradigmatic suffixes – and the suffixes which display limited productivity, induce lexicalised meanings and have dominant prosody: the non-paradigmatic suffixes. §11.4. discusses the structural flattening concomitant with the non-paradigmatic prosodic pattern. In §11.5. we show that Latinate nominalisations have joined the class of non-paradigmatic (idiomatic) derivations as expected, and make some initial speculations concerning the borrowing trajectory, arguing that the nominalisations in question got
borrowed as semantically idiomatic, morphologically simplex items. Their derivational relation with the base then emerged due to the fact that the number of borrowed event/property nouns sharing a final sequence had reached the critical mass, hence yielding a surface effect of a reconstructed (counterpart of the) source language derivational suffix in the recipient lexicon. Building on this idea, in §11.6. we explore the borrowing patterns for the word classes which serve as base words in nominalisations – adjectives and verbs – comparing them to that for nominalisations. We show that the way words are borrowed into S-C ensures that, on top of prosodic differences, there is a segmental mismatch between the base and the stem of the Latinate nominalisation, which effectively prevents the synchronic derivation of Latinate nominalisations from being transparent. As a consequence, Latinate nominalisations are prevented from joining the paradigms of S-C Latinate verbs and adjectives. In sum, Latinate items are allowed to establish some derivational relatedness, but that is never of the paradigmatic kind. In §11.7. some consequences for the general theory of grammar and lexicon are drawn.

11.3. Native nominalisations

The specificity of S-C is that, in both deadjectival and deverbal nominalisations, it displays a native contrast between:

(a) paradigmatic derivations, which strictly maintain the prosodic pattern of the base, and

(b) non-paradigmatic derivations, which diverge from the prosodic pattern of the base.

Recent morphological literature has shown that the split to inflectional and derivational morphology, where all derivational morphology takes place before all inflection morphology, and surfaces closer to the stem, cannot be maintained (ever since Di Sciullo & Williams 1987: 69-71, see also Haspelmath 1995). This has also brought into question the notion of a paradigm. We use the notion of paradigm in a different sense, independent of the above distinction. As specified in §11.2., we borrow the view from Arsenijević & Simonović (2013), where a paradigm is the set of all the forms which are compositionally derived by a productive morphological operation (typically affixation) from one and the same stem.

The asymmetry between paradigmatic and non-paradigmatic derivations is reflected in different prosodic shapes in S-C. Observe the examples in (160), where the same base adjective and the same suffix combine to derive one paradigmatic and one idiomatic noun. In each of the examples, the paradigmatic nominalisation surfaces with the prosodic pattern identical to
that of the base adjective (the suffix is prosodically vacuous, i.e. comes without tone or stress), while the non-paradigmatic version has a prosodic pattern which indicates the involvement of a dominant suffix.

(160) Native deadjectival nominalisations

<table>
<thead>
<tr>
<th>Base</th>
<th>Paradigmatic</th>
<th>Non-paradigmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>prIvaatn-a</td>
<td>prIvaatn-oost</td>
<td>privAAtn-OOst</td>
</tr>
<tr>
<td>privat-FSg</td>
<td>privat-oost</td>
<td>privat-oost</td>
</tr>
<tr>
<td>'private'</td>
<td>'privateness'</td>
<td>'privacy'</td>
</tr>
<tr>
<td>Opaasn-a</td>
<td>Opaasn-oost</td>
<td>opAAsn-OOst</td>
</tr>
<tr>
<td>dangerous-FSg</td>
<td>dangerous-oost</td>
<td>dangerous-oost</td>
</tr>
<tr>
<td>'dangerous'</td>
<td>'dangerousness'</td>
<td>'danger'</td>
</tr>
<tr>
<td>sEksuaaln-a</td>
<td>sEksuaaln-oost</td>
<td>seksuAAln-OOSt</td>
</tr>
<tr>
<td>sexual-FSg</td>
<td>sexual-oost</td>
<td>sexual-oost</td>
</tr>
<tr>
<td>'sexual'</td>
<td>'sexualness'</td>
<td>'sexuality'</td>
</tr>
</tbody>
</table>

Somewhat different behaviour is observed with verbs, where each verb derives only one nominalisation. S-C verbs are specified for grammatical aspect, bearing the values perfective or imperfective. The default nominalising suffix, closely matching the English -ing and the German -ung (Ignjatović 2013), is the suffix –anje/-enje (composed from the passive participle suffix -an/-en and the mass nominalising suffix -je, more precisely phonologically represented as -VV.je due to its lengthening of the final syllable of the stem – see Arsenijević 2007 for a semantic, and Simonović & Arsenijević 2014 for a phonological analysis). All imperfective and only imperfective verbs in S-C productively derive deverbal nominalisations, which remain within the paradigm in the sense adopted in this paper. A small, arbitrary subset of perfective verbs also derive -VV.je-nominalisations, but these always have a prosodic pattern unattested in the paradigm of the verb and typically come with a somewhat shifted or idiomatic semantics.

32 The following format is used to represent S-C prosody: capital letters mark all the syllables which get surface prominence (stress and tone) and double vowels mark long vowels. Single capitalised vowels stand for the falling accent (where tone and stress co-occur) and will be encountered only on the first syllable (in native words). Any two adjacent capitalised syllable nuclei stand for a rising span with the stress falling on the leftmost syllable of the span. In sum, the capital letters can be read as indicators of tone-bearing syllables, whereas the distribution of stress is predictable from that of tone (e.g. tatAtA stands for ta’tAtA, whereas tAtata stands for 'tAtata).
Native deverbal nominalisations

<table>
<thead>
<tr>
<th>Verb</th>
<th>Paradigmatic</th>
<th>Non-paradigmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>pridružIvA-ti\textsuperscript{IMP}</td>
<td>pridružIvAA-nje</td>
<td>none</td>
</tr>
<tr>
<td>conjoin-Inf</td>
<td>conjoin-N</td>
<td></td>
</tr>
<tr>
<td>‘to conjoin’ (imperf.)</td>
<td>‘conjoining’</td>
<td></td>
</tr>
<tr>
<td>pridrUUžI-ti\textsuperscript{PERF}</td>
<td>none</td>
<td>pridružEE-njE</td>
</tr>
<tr>
<td>conjoin-Inf</td>
<td></td>
<td>conjoin-nje</td>
</tr>
<tr>
<td>‘to conjoin’ (perf.)</td>
<td></td>
<td>‘accession’</td>
</tr>
<tr>
<td>rešAAvA-ti\textsuperscript{IMP}</td>
<td>rešAAvAA-nje</td>
<td>none</td>
</tr>
<tr>
<td>solve-Inf</td>
<td>solve-nje</td>
<td></td>
</tr>
<tr>
<td>‘to solve’ (imperf.)</td>
<td>‘solving’</td>
<td></td>
</tr>
<tr>
<td>rEEšI-ti\textsuperscript{PERF}</td>
<td>none</td>
<td>rešEE-njE</td>
</tr>
<tr>
<td>solve-Inf</td>
<td></td>
<td>solve-nje</td>
</tr>
<tr>
<td>‘to solve’ (perf.)</td>
<td></td>
<td>‘solution’</td>
</tr>
</tbody>
</table>

We can thus outline a local generalisation, elaborated more extensively in Simonović & Arsenijević (2014), where:

- imperfective verbs derive paradigmatic nominalisations with gerund semantics (pridružIvAA\textsuperscript{Anje} ‘conjoining’), whereas
- perfective verbs derive non-paradigmatic nominalisations (if any) with lexicalised semantics (pridružEE\textsuperscript{EnjE} ‘accession’).

While the source of the accent pattern in paradigmatic nominalisations is clearly the base, it is unclear where the accent paradigm of the non-paradigmatic nominalization comes from. Explaining this prosodic shape is the focus of the following section.

### 11.4. Forced Lexicalisation

The pattern which surfaces in non-paradigmatic deverbal nouns – a rising span over the two final syllables – sweeps away all prosodic information (stress, tone and vowel length) originating from the verb. While it is clear on functional grounds why such a constellation would emerge in a language with much interaction between morphosyntactic and prosodic structures, it still remains a question how this prosodic marking of the non-paradigmatic domain is exactly implemented. In other words, the question becomes whether S-C has a dedicated morpheme for marking non-paradigmaticity or if the non-paradigmatic pattern somehow emerges without actually having a dedicated representation in the lexicon. This question becomes even more urgent once we take into account that the right-aligned rising span is not seen as any kind of default in existing analyses of S-C.

In most generative analyses (Zec 1999, Becker 2007), S-C words which
have no prosodic specification in the lexicon (no lexical tone), receive a post-lexical prosodic pattern whereby the initial syllable bears both high tone and stress.

(162) /gledaan+VV.je/ → [gEdaanje], /kosovo/ → [kOsovo], /ostrvo/ → [Ostrvo]

There is, however, a peculiar gap in the repertoire of post-lexical patterns, which as far as we know has not been spotted in the literature. In the paradigms which have no bare, i.e. suffixless forms – the main feminine and neuter declensions in S-C – there are no morphologically simplex forms which have the post-lexical pattern with an unstressed long syllable. Simplex nouns without the high tone in the lexicon are always like KOsovo, never like *KOsovo, as illustrated in (163).

(163) Simplex H-less neuter nouns

kOsovo   ‘Kosovo’  no internal structure
/kosovo/
oloovo   ‘lead’     no internal structure
/lastovo/
ostrvo   ‘island’   no internal structure
/ostrvo/

All the nouns which have the surface prosodic configuration involving a high tone and stress on the initial syllable and a long syllable on any other syllable are morphologically derived and compositionally interpreted, as shown by the examples in (164).

(164) Derived H-less neuter nouns

glEdaanje   prEdiivo   sklAdiište
glEda+an + VV.je   prEd + iiv + o   sklAd + ište
look+ed + N_mass  weave + able + NeutSg  order + N_loc
‘looking’   ‘weft’    ‘storage’

In sum, the available surface forms give us no indication as to what output a simplex stem of the type /kosoo+v+o/ (or a verb like /gledaa+ti/) would have. The only clear fact is that such a pattern would be neutralised with another pattern (because it does not surface). From the existing accounts, it follows that this pattern neutralises with the one containing no tone and no lexical length, i.e. that phonology deletes the lexical length, so /kosoo+v+o/ and /kosov+o/ both surface as kOsovo. The fact that rising spans headed by a long stressed syllable are ubiquitous in S-C non-paradigmatic domain, may offer a clue to look in a different direction, and amend the theory of post-lexical
Chapter 11

prosody in S-C, to the effect that the pattern in pridružEEnjE and rešEEnjE (rising span over the two final syllables) would be post-lexical in cases where there is pre-final length and no derivational structure. In other words, a lexical item of the type /kosooov+o/ would then surface as kosOoOvO, which is also the optimal output for /kosooov+o, H/31

Consequently, we argue that pridružEEnjE has the representation /pridružeenj+e/ (which can also be reanalysed into /pridružeenj+e,H/). This in turn is a consequence of the fact that non-paradigmatic nominalisations undergo forced lexicalisation: they are forced to lose their internal structure and lexicalise as stems, which accounts for their subparts’ prosody being deleted. In other words, forced lexicalisation enables the formation of those nominalisations from perfective verbs which are good candidates to receive certain nominal meanings that are loosely, in a pragmatically reconstructed way, related to the semantics of the verbal stem. What then characterises non-paradigmatic nominalisations is not a special morpheme they all have in common, but rather the fact that they are all forced to lexicalise as separate items, which, together with the fact that they are all derived using the same suffix -VV,je, leads to the same prosodic shape – the rising span aligned with the right edge of the word and headed by a long syllable. This puts a prosodic stamp on them, marking that they sit at the top of the paradigm, i.e. that there is no bigger paradigm that they are part of. Effectively, this specifies their stems as separate simplex stems, and thus dissociates them from the paradigm of a semantically and derivationally related stem. In order to illustrate the hierarchical structure of paradigms, and the place of the two types of nominalisation in this context, in (165) we are showing samples of the respective paradigms and their relations for the perfective and imperfective verbal stems from the base uskrs- ‘resurrect’.

32 In this S-C representation, a floating high tone is associated to the stem. This representation is equivalent to the representation /violiin+a, H/ (surfacing as violIInA) discussed by Zec (1999).
(165)
a. The nominalisation of a perfective verb forms (i.e. sits at the top of) a
distinct paradigm

\[
\begin{align*}
\text{paradigm of } & \text{uskr}^H\text{smu}+ti/ \\
& \quad \text{UskRsmu-} \\
& \quad \text{uskrsmu}+e,H/ \\
& \quad \text{resurect}^{\text{PERF}} \\
& \text{uskrsmuU}U\text{c-} \\
& \text{resurection-} \\
& \text{uskrsmU}U\text{c-U} \\
& \text{resurection-DatSg} \\
& \text{uskrsmU}U\text{c-Em} \\
& \text{resurection-InstSg} \\
\end{align*}
\]
b. The nominalisation of an imperfective verb joins its paradigm

The paradigm of /uskrsavaH+ti, uskrsaaHva+m/

This view neatly matches the syntactic analysis of the asymmetry between paradigmatic and non-paradigmatic nominalisations in Arsenijević (2011). In this account paradigmatic nominalisations are nominalised predications, with the suffix merging with a complex structure and nominalising it, as in (166a), while non-paradigmatic nominalisations have a flat structure: a structurally simplex stem is merged with a structurally simplex suffix, as in (166b).

(166)

a. [-oost [postp [dp Jovan] ljubazan [Adj ljubazan]]] b. ljubzn-oost
-oost kind Adj kind Adj kind-oost

Jovanova ljubazn-oost
Jovan’s kind-oost
‘Jovan’s kindness’ (trop)

11.5. Latinate nominalisations

Latinate strata in English, Dutch, German, S-C, and probably a number of other languages, have one property in common: they are all characterised by dominant suffixes, as opposed to the native derivations which at least have
the option of involving an unstressed recessive suffix. This is illustrated by
the S-C examples in (167).

(167)  Unstressed native and stressed Latinate suffixes
<table>
<thead>
<tr>
<th>Base</th>
<th>Native</th>
<th>Latinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>stEriilna</td>
<td>stEriilnoost/sterIIlnOost</td>
<td>sterilItEEt</td>
</tr>
<tr>
<td>socijAlizovati</td>
<td>socijAlizovaanje</td>
<td>socijalizAAcIja</td>
</tr>
</tbody>
</table>

A careful reader is probably coming up with an objection: socijalizAAcIja
does not share the base of socijAlizovati. Exactly this observation, promoted
to a generalisation, creates the base of our main argument. Distinct stems
are the way for Latinate nominalisations to escape the paradigm that the
semantics and the shared portion of segments (analysable as a base)
associate it with. It is in this sense that patterns in Latinate nominalisations
uncover new levels of non-paradigmaticity. Such a constellation is very
improbable to arise systematically in the native stratum. Observe more such
examples in (168).

(168)  S-C Nominalisations
<table>
<thead>
<tr>
<th>Base</th>
<th>Paradigmatic</th>
<th>Non-paradigmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>spEcijaaln-a</td>
<td>spEcijaaln-oost</td>
<td>specijAAln-OOST</td>
</tr>
<tr>
<td>special-FSg</td>
<td>special-oost</td>
<td>special-oost</td>
</tr>
<tr>
<td>‘special’</td>
<td>‘specialness’</td>
<td>‘specialty’</td>
</tr>
<tr>
<td>elokvEntn-a</td>
<td>elokvEntn-oost</td>
<td>elokvEn-cIja</td>
</tr>
<tr>
<td>eloquent-FSg</td>
<td>eloquent-oost</td>
<td>eloquent-cija</td>
</tr>
<tr>
<td>‘eloquent’</td>
<td>‘eloquentness’</td>
<td>‘eloquence’</td>
</tr>
<tr>
<td>defInIsa-ti</td>
<td>defInIsaa-nje</td>
<td>definII-cija</td>
</tr>
<tr>
<td>devine-Inf</td>
<td>define-nje</td>
<td>define-cija</td>
</tr>
<tr>
<td>‘to define’</td>
<td>‘defining’</td>
<td>‘definition’</td>
</tr>
<tr>
<td>intonIIrA-ti</td>
<td>intonIIrAA-nje</td>
<td>intonAA-cIja</td>
</tr>
<tr>
<td>intone-Inf</td>
<td>intone-nje</td>
<td>intone-cija</td>
</tr>
<tr>
<td>‘to intone’</td>
<td>‘intoning’</td>
<td>‘intonation’</td>
</tr>
</tbody>
</table>

As already briefly discussed in §11.3. and elaborated in Arsenijević &
Simonović (2013), a special prosodic pattern for S-C non-paradigmatic
derivations is no big news. Since Latinate nominalisations enter the lexicon
from the outside (so, in a sense in addition to the always already present
paradigmatic and non-paradigmatic derivations), it is fully expected for them
to join the prosodic pattern of non-paradigmatic native derivations. However, 
Latinate nominalisations introduce a new level of non-paradigmaticity in a 
systematic way.

Due to the stratification of the lexicon, S-C derived words can be ordered 
respect of degree of non-belonging to the paradigm. Generally speaking, 
paradigmatic deadjectival and deverbal nominalisations share both the stem 
and the prosodic pattern of the source adjective or verb. Non-paradigmatic 
native nominalisations share the stem, but differ in the prosodic pattern. In 
Latinate nominalisations, we encounter a further degree: these deadjectival 
and deverbal derivations differ from their related adjectives and verbs both 
in the segmental content of the stem and in the prosodic pattern. This latter 
observation is illustrated by the underlined divergent stems in (169).

(169) Different stems in native and loan nominalisations

<table>
<thead>
<tr>
<th>Base</th>
<th>Paradigmatic</th>
<th>Non-paradigmatic/native</th>
<th>Non-paradigmatic/loan</th>
</tr>
</thead>
<tbody>
<tr>
<td>spEcijaahn-a</td>
<td>spEcijaahn-oost</td>
<td>specijAAlh-OOst</td>
<td>specijal-ItEEt</td>
</tr>
<tr>
<td>'special'</td>
<td>'specialness'</td>
<td>'specialty'</td>
<td>'culinary specialty'</td>
</tr>
<tr>
<td>eloKVEntn-a</td>
<td>eloKVEntn-oost</td>
<td>eloKVEn-clja</td>
<td>eloquence'</td>
</tr>
<tr>
<td>'eloquent'</td>
<td>'eloquentness'</td>
<td>'eloquence'</td>
<td>'eloquence'</td>
</tr>
<tr>
<td>defInIsa-ti</td>
<td>defInIsaa-nje</td>
<td>definII-cija</td>
<td>'definition'</td>
</tr>
<tr>
<td>'to define'</td>
<td>'defining'</td>
<td>'definition'</td>
<td>'definition'</td>
</tr>
<tr>
<td>intonIIrA-ti</td>
<td>intonIIrAA-nje</td>
<td>intonAA-clja</td>
<td>'intonation'</td>
</tr>
<tr>
<td>'to intone'</td>
<td>'intoning'</td>
<td>'intonation'</td>
<td>'intonation'</td>
</tr>
</tbody>
</table>

The ubiquitous difference in the stems between the native and the borrowed 
derivation patterns originates from the way borrowing into S-C takes place, 
i.e. from an obligatory morphological integration of the borrowed verbal and 
adjectival stems by means of integration suffixes. Borrowed adjectives are 
normally integrated into the S-C grammar by the adjectivising suffix -n 
(Simonović 2012), and borrowed verbs are integrated using the verbal 
suffixes -isa, -ova or -ira (see §5.2.1. and Chapter 10).

(170) Borrowed adjectives receive the native adjectivising suffix -n

<table>
<thead>
<tr>
<th>German Suffix</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>steril</td>
<td>-n</td>
</tr>
<tr>
<td>privat</td>
<td>-n</td>
</tr>
<tr>
<td>elementar</td>
<td>-n</td>
</tr>
</tbody>
</table>
Borrowed verbs receive the (native) verbalising suffixes -\textit{isa}, -\textit{ova} or -\textit{ira}.

\begin{tabular}{lll}
  German & Suffix & SC \\
  engagieren & -ova & angažovati \\
  jogg-en & -ira & džogirati \\
  definieren & -isa & definisati \\
\end{tabular}

The borrowing mechanism eliminates the possibility that a foreign adjectival/verbal stem directly becomes a S-C adjectival/verbal stem: it always adds native material first. As a consequence, Latinate nominalisations, consisting of a Latinate stem and a Latinate suffix, never share the stem of the native derivation from a Latinate base.

\begin{tabular}{lll}
  Base & S-C & Latinate \\
  spEcijaaln-a & spEcijaaln-oost, specijAAln-OOst & special-ItEEt \\
  'special' & 'specialness', 'specialty' & 'culinary specialty' \\
  elokvEntn-a & elokvEntn-oost & elokvEn-clja \\
  'eloquent' & 'eloquentness' & 'eloquence' \\
\end{tabular}

As was shown in Chapter 10, loan verbs typically enter the S-C lexicon as unspecified for aspect, and only later have the option of getting profiled as perfective or imperfective forms. Since only imperfective verbs productively derive deverbal nominalisations, it is predicted that borrowed deverbal nominalisations (including the Latinate ones) will be able to take on the semantic pattern typical of the perfective (non-paradigmatic) deverbal nominalisations, and will therefore block the emergence of native perfective (non-paradigmatic) nominalisations from borrowed verbs. This prediction is confirmed: perfective borrowed verbs never derive non-paradigmatic nominalisations using the native pattern (stress attracting -\textit{VVje}).

Borrowed verbs do not derive non-paradigmatic native nominalisations

\begin{tabular}{lll}
  Verb & Paradigmatic nom. & Non-paradigmatic nom. \\
  šutIrrAti\textsuperscript{IMP} & šutIrrAAnje & \\
  'to kick, to shoot' & 'kicking, shooting' & šUt *šutnUUčE \\
  šUtnti\textsuperscript{PERF} & štAampaanje & \\
  'to kick, to shoot' & 'kicking, shooting' & štAampa *odštampAAnjE \\
  štAAmpati\textsuperscript{BI} & štAAmpaanje & \\
  'to print' & 'printing' & \\
  OdštAAmpati\textsuperscript{PERF} & & \\
  'to print out' & 'press' & \\
\end{tabular}

The question now is how grammar ‘plans’ how to integrate the borrowed
nominalisations in such a neat way that paradigms are separated and memberships are marked. And the answer is: of course it does not. Rather than being of some 'paradigm separation mechanism', this optimal outcome emerges as a consequence of a number of independently justified processes, which are the focus of the next section.

11.6. The integration mechanism for nominalisations

We argue that in the initial stage of integration, Latinate nominalisations were borrowed as simplex words, with idiomatic semantics (as simplex words are supposed to have) and unrelated to their verbal or adjectival counterparts, even when these had also been subject to borrowing. However, the rise in the number of simplex items sharing a final sequence and a number of semantic and syntactic properties (e.g. -tion nominalisations being event nominals), while at the same time sharing a segmental sequence with a verb or adjective, led to a surface generalisation which eventually triggered the emergence of a set of suffixes limited in distribution to the domain of stems which do not participate in any native paradigm. The isolation from the native paradigms came as a consequence of the requirement that a borrowed stem be integrated by a native suffix in order to derive a paradigm with native affixes.

(174) Latinate nominalisations not part of the paradigm of the borrowed stem

<table>
<thead>
<tr>
<th></th>
<th>defInIsa-ti</th>
<th>defInII-cija</th>
<th>frustrIra-ti</th>
<th>frustrAA-ti</th>
<th>frUstrAA-cija</th>
</tr>
</thead>
<tbody>
<tr>
<td>define-Inf</td>
<td>‘to define’</td>
<td>‘definition’</td>
<td>‘to frustrate’</td>
<td>‘frustration’</td>
<td></td>
</tr>
<tr>
<td>intonIra-ti</td>
<td>intonAA-cija</td>
<td>evoluIra-ti</td>
<td>evoluUU-cija</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘to intone’</td>
<td>‘intonation’</td>
<td>‘to evolve’</td>
<td>‘evolution’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

That borrowed nominalisations were typically stressed on the suffix (i.e. that in Romance the respective endings are all stress-bearing) matched conveniently with the fact that in S-C non-paradigmatic derivations involve stressed suffixes. This has additionally supported the emergence of suffixes effectively (nearly) matching the suffixes from the language of origin.
(175) Borrowing by adding a native suffix and borrowing followed by suffix emergence

<table>
<thead>
<tr>
<th>borrowed</th>
<th>setting the stem</th>
<th>nominalisations mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>defin\textsubscript{V}</td>
<td>→ defin + isa\textsubscript{V}</td>
<td>→ definisa+nje ‘defining’</td>
</tr>
<tr>
<td>definicijan</td>
<td>→ defini←</td>
<td>→cijan</td>
</tr>
</tbody>
</table>

Recent decades have brought about further change in the status of the suffixes of the borrowed domain -cija and -itet – they have started to occur with native stems, in a limited, pragmatically restricted domain of the colloquial language: in nominalisations marked for a comical, vulgar or intimate character.

(176) Latinate suffixes with native stems

pretumba-ti pretumba-cija uživ-ati uživ-an-cija
scramble-Inf scramble-cija enjoy.Inf enjoy-an-cija
‘to scramble’ ‘scrambling’ ‘to enjoy’ ‘joy’
švaler(is)-a-ti švalera-cija za-jeb-ati zajeb-an-cija
have_affair-Inf have_affair-cija for-fuck.Inf for-fuck-an-cija
‘to have an affair’ ‘having affairs’ ‘to trick’ ‘fun’
zanima-ti zanima-cija opusti-ti opušt-en-cija
entertain-Inf entertain-cija relax-Inf relax-en-cija
‘to entertain’ ‘entertainment’ ‘to relax’ ‘relaxation’

There are however, no signs that these nouns may move further into the direction of paradigmatic nominalisations.

11.7. Conclusions

The emerging picture of the lexicon is that of a coconut-like structure in respect of productivity of morphological patterns. Its core contains ‘soft’ patterns which form domains of paradigmatic derivations, characterised by being: (a) maximally productive, (b) semantically transparent and (c) phonologically non-demanding. Next, the ‘medium’ layer involves the domains of constrained productivity, with idiosyncratic semantics and own patterns in the domain of prosody. Borrowed derivations form the ‘hardened’ outer layer of extreme non-belongers, with restricted compositionality and possible heavy phonological restrictions. This view of the lexicon in contact is compatible with the discussion of underspecification approaches in §3.5., where we concluded that underspecified structures are expected to be part of the core stratum of the lexicon, whereas more specification goes hand in hand with less relatedness to other words in the lexicon.

Latinate nominalisations belong to the outer, hard layer. They have not only special prosodic shapes, but also stems which differ from those of the related
verbs/adjectives. However, they do get reanalysed into stems and emerging suffixes, unlike the more recently borrowed English derivations of the type *rijaliti* 'reality show' *sekjuriti* 'security personnel', *selbriti* 'celebrity', *sešn* 'session' (only in fashion) and *fešn* 'fashion'. All these items are integrated as simplex stems (e.g. *rijaliti*-ja 'reality-GenSg', *rijaliti*-ju 'reality-DatSg') and no analysis into stems and suffixes is performed.

A relatively high productivity of what looks like borrowed Latinate nominalising suffixes is no blatant case of massive borrowing of bound morphology for general purposes. Loan nominalisations enter S-C with idiomatic semantics and limited possibilities of morphological relating with other items in the lexicon, and tend to remain that way.

These derivational patterns are not really imported – they rather reemerge in the recipient language. A number of constraints stand in the way of these patterns entering competition with the native ones, since they remain isolated by their different stems (forming separate paradigms), non-paradigmatic prosody (which specifies that these items should remain in isolation) and other restrictions, such as the pragmatic effects, as mentioned in §11.6.

In the following chapter, we summarise and conclude the dissertation.
Chapter 12
Conclusions

In this final chapter, we conclude this dissertation by summarising the main theoretical and conceptual innovations it has provided, the relations it has established and the hopes for future research which follow from it.

In the course of this dissertation, we have introduced quite a few conceptual and terminological innovations, especially in the chapters which present the proposed model. In what follows, the main notions deployed in the course of the dissertation will first be summarised. It is our hope that this overview will prove helpful to the reader who has read the previous chapters, as well as to the reader who tackled the dissertation by reading the conclusions first.

As pointed out in the Preface, our main goal was to create an empowered loanword theory by incorporating insights from two domains which are often underrepresented in formal theories of loanword treatment – sociolinguistic contact literature (especially that on code-switching) and research into morphological (and morphosyntactic) integration. In the Preface we already announced that the resulting model would encompass two complementary perspectives: the more *diachronic* perspective, which focuses on the process of lexicalisation of a new word and the process through which a new form acquires all the properties that a lexical item of the recipient lexicon needs to have, and the more *synchronic* perspective, which focuses on the bilinguals’ shared knowledge on how to transform words from the source language (SL) into those of the recipient language (RL).

In Chapter 1, which was devoted to the (sociolinguistic) literature on code-switching and bilinguals, one important insight emerged, paving the way to our *synchronic* perspective: bilingual communities develop what Haugen termed *interlingual identification* – stable correspondences between RL and SL structures which naturally extend to new words. Moreover, the literature reviewed in this chapter showed that there are rather formal restrictions on code-switching, including the fact that one code is fully activated, termed the *matrix code*, whereas the other language, the *embedded code*, is restricted to providing lexical material. The matrix code also provides function words and bound morphology. This insight played an important role in grounding the more *diachronic* perspective, since code-switching defines the input to borrowing. In Chapters 2 and 3, which discuss the existing phonological literature on loanwords, we established that there are two fields of loanword phonology, and separated the objects of the two
fields terminologically. There is, on the one hand, the literature on modifications which incoming loanwords undergo, which we termed *borrowing research*, and, on the other hand, there is the literature which focuses on the treatment of exceptionally preserved marked structures in loanwords in monolingual lexicons, which we termed *borrowedness research*. While recognising that the various approaches to *borrowing research* form the base for how we conceptualise loanwords today, we concluded that the discourse in this area has been somewhat constrained by the rather fixed objectives: on the one hand, the *input-based approaches* are involved in a dispute on the most common input to loanword phonology, while, on the other hand, *OT-based accounts* try to fit the loanword integration into the existing theory of grammar using as few loan-specific mechanisms as possible. And yet, while the most important dispute in the field was concluded to be an *essentialist* quest for ‘real’ loanword adaptation, our discussion of the most explicit borrowing models (such as Boersma & Hamann 2008 and Dong 2012) showed that these models are compatible with the bilingual and lexical perspectives, which our model capitalises on. Moreover, since our model uses some OT formalisms, we concluded that this dissertation can be read as adding to the tendency that has already been present in the OT literature: seeing loanword treatment less and less as grammar proper and more and more as one of the many grammar’s interfaces.

Also in the discussion of the models of the stratified lexicon presented in Chapter 3, important and valid insights have been found, which were incorporated in our model. As this field teaches, loanwords, allowing more marked structures than native words, constitute the periphery of the stratified lexicon, to which a special type of Faithfulness – FAITH(LOAN) is indexed. We concluded that the models developed in this field are descriptively adequate for borrowedness. However, we also concluded that a complementary theory of borrowing needs to avoid being overly grammatical, as simply assuming that FAITH(LOAN) can account for borrowing misses some crucial properties of borrowing.

Chapter 4, which presents the ingredients of our model, and Chapter 5, which discusses morphological and morphosyntactic integration were particularly rich in innovative concepts and terms. These two chapters were united by taking the diachronic perspective, which we termed the *loanword trajectory*: loanword integration was considered from the perspective of the lexicalisation of a single word entering RL. We first identified the initial and the final point of the trajectory – it starts from an *initial surface form* (inserted as a code switch) and ends with a full-fledged RL lexical item. While the endpoint means different things in different languages, the initial
surface form has a number of features independently of the RL: the fact that it lacks morphological and morphosyntactic features, the fact that it is a surface form and the fact that it has a specific meaning. In order for such a form to be transformed into a lexical item, its representation has to change considerably. We demonstrated that this process is controlled by the force which has been showed to protect surface features within lexicons – *Lexical Conservatism*. The phonological features which are successfully protected by LC can be passed to *Faith* at the endpoint of the trajectory if necessary. In Chapter 5, in which morphosyntactic integration was discussed, we showed that integration is always also morphological. Loanwords do not enter the phonology of a language: they only ever enter the *lexicon*, and what guides ‘phonological’ integration is the integration into the lexicon. We also concluded that the goal of loanword research is not to predict the correct output in the simple mapping from one string to another, but to restrict the space of possible interactions in the complex process in which bilingual knowledge, lexicon and grammar are entangled in determining the phonological, morphological, syntactic and semantic features of the new lexical item. In data set after data set, we saw how this process is constrained by a single force – *Lexical Conservatism*, which requires the preservation of all the phonological features of the initial surface form in the newly formed paradigm.

Presenting the ‘synchronic’ perspective of the proposed model, in Chapter 6 we argued that, the *inter-language mappings* – correspondences between units of RL and SL internalised by bilinguals – are as real as the SL and RL grammars. This second perspective completed the process of removing the *grammar bias* – the tendency to observe changes in loanwords as an instantiation of grammar by definition – from our model. While the mappings may appear very grammar-like at first sight, we argued that they are conventionalised, much simpler than correspondences encoded with grammars, and only equipped to deal with the inputs which occur in the actual contact. We maintained that both the LC-centered and the mapping-centered perspectives were equally crucial for the creation of a comprehensive loanword model. While the two perspectives cannot be taken at the same time – LC and inter-language mappings cannot be observed in action simultaneously – when inter-language mappings undergo change, it is in the direction of more LC-friendliness. In other words, LC can only ever get promoted.

In Chapter 7, which presented the interim summary of our model, the model was shown to enable us to explicitly describe the stage of language contact which most accounts of loanword phonology routinely presume – the stage in which stable inter-language mappings are established. We called
Chapter 12

At this stage, where a considerable number of words support inter-language mappings and make certain words borrowable and potentially-already-borrowed, the state of language entanglement. This chapter announced the four case-study chapters, which showed how the model proposed in this dissertation interacts with larger data sets, in each case contributing to what can be called a typology of the relations between loanword integration and RL grammar.

Our discussion of the gemination data in Chapter 8 presented a cross-linguistic catalogue of related phenomena with the same trigger – Lexical Conservatism. The findings of this chapter illustrated the constellation which, as argued throughout the dissertation, needs to be taken into account: a phonological modification which is not guided by RL phonology and due to which RL phonology does not get modified. In other words, Chapter 8 presented an empirical argument for the conceptual autonomy of loanword integration patterns.

The three remaining case-study chapters focused on a single language, Serbo-Croatian, revealing rather intricate relations between loanword integration and RL grammar. Chapter 9 showed how LC as applied to loanwords is just the tip of the iceberg of LC-related limitations on lexicalisation. LC turned out to play a decisive role in ways in which native words can be re-analysed. There are patterns which may be representable and allowed by the grammar, but they only occur in wug-test type experiments, since there is no lexicalisation path through which they may emerge. The verb borrowing data discussed in Chapter 10 demonstrated that, while the conceptual separation of the use of a suffix in integration from its RL-internal productivity is justified, the suffix' use in borrowing does influence its morphosyntax in RL-internal derivations. This chapter also brought the most detailed description of an active inter-language mapping in this dissertation. Chapter 11 showed how the predicted inflexibility and low productivity of the derivational morphology introduced by loanwords turns out to be a continuation of what was independently noticed in the Serbo-Croatian native morphology – the distinction between what we have termed the paradigmatic and the non-paradigmatic patterns.

The derivational patterns introduced by loanwords are conveniently added to the 'hard', non-paradigmatic layer of the lexicon, which implies low productivity and little semantic transparency.

In sum, while we have spent much of this dissertation arguing against the assumption that loanwords are an area in which RL grammar gets a chance to 'express itself', the case studies have demonstrated that once loanword integration patterns are granted the necessary conceptual autonomy, they can lead to further clarity on the reality of the native
patterns. In other words, instead of giving a universalising answer to the issue of the relation between loanword integration patterns and RL grammar, we can conclude that this relation can be of rather different types, and that taking this relation into account should be added to the agenda of descriptively adequate linguistic models, especially when such models are applied to languages which are in contact with other languages. What we maintain to be universal is that loanword integration is the creation of a lexical item based on a surface form under the control of Lexical Conservatism and that bilingual speakers have contact-specific knowledge of correspondences between RL and SL structures – the inter-language mappings.

The main goal of our endeavour was creating space for a conceptual separation of loanword treatment from RL grammar, or rather, defying their unification, which is self-evident to many current approaches. Once the separation was achieved, exploring the relation between the two could become part of the new research agenda. Moreover, the model presented here was created in an effort to relate the field of phonological loanword integration to its neighbouring fields – code-switching research and morphological integration – but also to general linguistic issues, such as the architecture of the grammar and the lexicon. The main consequence of our model which, we hope, will be hard to ignore, is that every pattern that can be analysed as a case of phonological integration goes hand in hand with various different integration processes which are part of a word becoming an RL lexical item. Such a pattern, rather than being a spontaneous reaction of a grammar to a new input, is typically a result of a conventionalisation within the bilingual community. In sum, our model matters not only to the way loanword phenomena should be explained, but also to how they should be described.

Due to the fact that it related quite autonomous areas of research, our model requires research(ers) to reach outside the original narrow field and establish new relations. While we hope to have put two types of relatedness which are design properties of our model on the agenda of loanword research, there still remains much for future research to relate. We hope that once liberated from the grammar bias, loanword research will look for the precedents of loanword patterns in the native domain. In this endeavour, which we have only occasionally initiated, there is much to learn from lexicon-internal lexicalisation processes (such as the lexicalisation of abbreviations) and language-internal mappings (such as interdialectal mappings). Moreover, we hope that future research will be able to relate language contact situations which have been explicitly excluded from our discussion (e.g. situations where the is no clear boundary between codes or
no replication of lexical material) to the concepts developed here (or account for the fact that they are not applicable to such situations).
Bibliography


Berent, Iris, Donca Steriade, Tracy Lennertz and Vered Vaknin. 2007. "What we know about what we have never heard: Evidence from perceptual illusions". *Cognition* 104(3):591-630.


Simonović, Marko and Boban Arsenijević. 2014. "Regular and honorary membership: On two kinds of deverbal nouns in Serbo-Croatian". Accepted for publication in *Lingua e Linguaggio*.


Samenvatting in het Nederlands

De immigratiedienst van het lexicon: Prolegomena tot een theorie over leenwoordintegratie

Talen hebben een universeel potentieel elementen uit andere talen over te nemen en in hun structuur te incorporeren. Het doel van leenwoordenonderzoek is antwoorden te vinden op vragen die zich aandienen wanneer taalcontact tot het ontlenen van woorden of woordachtige elementen leidt. Het belangrijkste theoretische doel van ons streven in dit proefschrift ligt in het creëren van ruimte voor een conceptuele scheiding van leenwoordbehandeling en grammatica, of liever gezegd, het tegenwerken van de in vele moderne benaderingen vanzelfsprekende conceptuele vereniging van beide. Zodra die scheiding compleet is, kan en zal het bestuderen van de relatie tussen leenwoordbehandeling en grammatica deel uitmaken van de nieuwe onderzoeksagenda. Bovendien is het hier voorgestelde model opgezet als een poging om het gebied van fonologische leenwoordintegratie betrekking te laten hebben op de omliggende velden – onderzoek naar codewisseling en morfologische integratie – maar ook op algemene taalkundige kwesties, zoals de architectuur van de grammatica en het lexicon.

Het is onomstreden dat leenwoorden in het algemeen door (breed gedefinieerde) tweetaligen geïntroduceerd worden (eerst als codewissel, dus opgehaald uit het lexicon van de brontaal) en dat ze aan het eind van het leenproces opgenomen zijn in het lexicon van de lenende taal. Het is echter geenszins duidelijk wanneer een woord ophoudt een codewissel te zijn en een geïntegreerd leenwoord wordt, d.w.z. vanaf welk moment een woord een representatie in het lexicon van de ontlenende taal heeft. Dit onderscheid is een van de meest omstreden onderwerpen in de moderne literatuur over codewisseling (zoals besproken in hoofdstuk 1) en elke poging om het op een coherente manier te maken leidt tot verdere belangrijke vragen: voor wie moet het woord in kwestie een leenwoord worden? (een individuele spreker? de hele taalgemeenschap?) en is dit een onomkeerbaar proces of kan een leenwoord geleend, vergeten, en vervolgens opnieuw geleend worden? (en wederom door wie?). Aan deze vragen is in de literatuur veel aandacht besteed en het juiste antwoord lijkt te zijn: al het bovenstaande. Met andere woorden, in plaats van zich te lenen voor een simplistische theorie, dwingen leenwoorden ons bewust te blijven van de verschillende dimensies waarlangs ze bestaan en ertoe doen: individuele en gemeenschappelijke, synchrone en diachrone etc.

Vele factoren zijn hier verstrengeld en verschillende perspectieven mogelijk. Het is daarom ook niet bepaald verbazingwekkend dat er in verschillende deelgebieden binnen de taalkunde interesse voor leenwoorden is ontstaan, dwars door de grenzen tussen velden, programma’s en ideologieën. Aan de
ene kant zijn leenwoorden traditioneel een thema binnen het gebied van taalcontactonderzoek, dat als onderdeel van de sociolinguïstiek wordt beschouwd. In dit deelgebied, dat door op taalgebruik gebaseerde (Engels: usage-based) benaderingen wordt gedomineerd (zie Backus 2012), worden leenwoorden als onderdeel van een continuüm aan gerelateerde verschijnselen behandeld, zoals o.a. codewisseling, taalverschuiving, talenmenging en creolisering. Anderzijds zijn leenwoorden het onderwerp van uitgebreid onderzoek in fonologie, die wordt gedomineerd door zogenoemde formele benaderingen. Zoals het uit het overzicht van dit gebied in hoofdstuk 2 blijkt, zijn leenwoorden nooit echt gepositioneerd binnen deze traditie, en het blijft in het huidige vertoog nog steeds vrij implicit waar leenwoordprocessen moeten worden geplaatst in de architectuur van het taalvermogen. Kortom, voor formele fonologen zijn leenwoorden een interessante puzzel die met behulp van vertrouwde middelen moet worden opgelost.

Het doel van het voorgestelde model is het gebied van formele leenwoordonderzoek te versterken door (a) inzichten uit sociolinguïstisch onderzoek toe te passen op formele modellen en (b) morfologische (en morfosyntactische) integratie in het beeld van leenwoordbehandeling op te nemen, dat nu gedomineerd woord door fonologische leenfenomenen. Tijdens het opbouwen van het model laten wij ons sturen door een aantal klassieke vragen van de moderne taalkunde: a) Is het proces in kwestie synchroon of diachroon? Als het synchroon is, behoort het proces tot de synchrone grammatica of tot een ander module? b) Is het proces in kwestie een weerspiegeling van de kennis van de sprekers of emergent in de gemeenschap? Als het een weerspiegeling is van de kennis van de sprekers, hoe is het proces mentaal gerepresenteerd? Als resultaat wordt een nieuw model voorgesteld dat inzichten biedt in het interface tussen talen. Dit model heeft betrekking op feiten van fonologische, morfologische en morfosyntactische leenwoordintegratie. Bekeken vanuit ons model blijken deze vormen van integratie dan ook verstrekkelde aspecten van eenzelfde brede proces: lexicalisatie, ofwel het creëren van een nieuw lexicaal item op basis van een oppervlaktetvorm uit de brontaal. Het voorgestelde model is in deze zin gekenmerkt door een zekere tenticiteit, net als de heersende adaptatiemodellen (hoofdstuk 2). Maar terwijl de adaptatiemodellen leenwoordprocessen als een beweging in de richting van inheemsheid ziet, heeft ons integratiemodel de totstandkoming van een volledig functioneel lexicaal item (soms zeer onderscheidbaar niet-inheems) als eindpunt.

Omdat bij leenwoordprocessen zowel synchroon als diachrone, en zowel individuele als gemeenschappelijke aspecten van belang zijn, bestaat ons model, dat in de hoofdstukken 4, 5 en 6 gepresenteerd wordt, uit twee verschillende perspectieven die ons in staat stellen om verschillende aspecten van het leenwoordgedrag te formaliseren en te verklaren, zonder uit het oog te verliezen wat op het moment buiten beschouwing gelaten wordt. Het meer diachrone perspectief van het model is vooral toegesneden op de manieren waarop de eigenschappen van de oorspronkelijke codewissel beschermd worden in het proces van integratie in het lexicon (gedeeld door
de hele taalgemeenschap), waaronder het creëren van een paradigma, de toekenning van morfosyntactische kenmerken etc. Het andere perspectief is meer geschikt voor het bestuderen van de regelmatigheden die onderdeel zijn van de kennis van de ontleners: de *interitalige mappings*, die binnen de tweetalige gemeenschap ontstaan en instructies bevatten voor het omzetten van elementen van de brontaal tot de corresponderende elementen van de ontlenende taal.

In wat volgt, wordt er een overzicht gegeven van de meest belangrijke inzichten van de hoofdstukken. Hoofdstuk 1 is gewijd aan de (sociolinguïstische) literatuur over codewisseling en tweetaligheid. Een inzicht dat een belangrijke rol speelt in het opzetten van ons synchone perspectief is dat tweetalige gemeenschappen de neiging hebben te ontwikkelen wat Haugen *interitalige identificatie* noemde – stabiele correspondenties tussen structuren van de brontaal en de ontlenende taal. Deze correspondenties hebben de tendens om productief toegepast te worden op nieuwe woorden. Bovendien blijkt uit het overzicht van de literatuur in dit hoofdstuk dat er belangrijke formele beperkingen op codewisseling zijn, waaronder het feit dat maar één taal volledig wordt geactiveerd (de zogenaamde *matrixcode*) terwijl de andere taal (de *ingebedde code*) beperkt is tot het verschaffen van lexicaal materiaal. De matrixcode levert dan ook de functiewoorden en gebonden morfemen. Dit inzicht speelt een belangrijke rol voor het meer diachrone perspectief, omdat codewisseling de input voor het leenproces bepaalt.

Hoofdstukken 2 en 3 bespreken de bestaande fonologische literatuur over leenwoorden. Daarin wordt eerst geconstateerd dat er twee gebieden van leenwoordfonologie zijn, die vervolgens ook terminologisch gescheiden worden. Er is, enerzijds, literatuur die modifikaties analyseert die inkomende leenwoorden ondergaan, en anderzijds is er literatuur die focust op de structuur van het postcontactlexicon, met name de formalisering van uitzonderlijk beschermde gemaakte structuren in leenwoorden. Het eerste deelgebied noemen we onteneningsonderzoek en bespreken het in hoofdstuk 2, terwijl het tweede deelgebied ontleendheidsonderzoek wordt genoemd en in hoofdstuk 3 besproken wordt. Onze discussie in hoofdstuk 2 erkent dat de benaderingen binnen het gebied van onteneningsonderzoek bepalend zijn voor de manier waarop we leenwoorden tegenwoordig conceptualiseren, doch we concluderen dat het discours op dit gebied enigszins beperkt is door op voorhand vastgestelde doelstellingen: aan de ene kant zijn er input-gebaseerde benaderingen die betrokken zijn bij een disput over de meest voorkomende/normale input voor de leenwoordfonologie, terwijl, aan de andere kant, de op Optimaliteitstheorie gebaseerde benaderingen leenwoordintegratie in bestaande modellen van grammatica proberen in te passen en daarbij zo weinig mogelijk leenwoord-specifieke mechanismen gebruiken. En terwijl het belangrijkste disput op het gebied op een essentialistische zoektocht naar de 'echte' leenwoordadaptatie uitloopt, toont onze bespreking van de meest expliciete onteneningsmodellen (zoals Boersma & Hamann 2008 en Dong 2012) aan dat
deze modellen compatibel zijn met de tweetalige en lexicale perspectieven, waarop ons model inspeelt. Aangezien ons model gebruikmaakt van een aantal formalismen van de Optimaliteitstheorie, concluderen we dat dit proefschrift kan worden gelezen als toevoeging aan de tendens die reeds aanwezig was in optimaliteitstheoretische literatuur: leenwoordbehandeling wordt steeds minder als grammatica strictu senso gezien en steeds meer als een van de vele interfaces van de grammatica. Ook in de bespreking van de onteendheidsmodellen van gestratificeerde lexica in hoofdstuk 3 zijn belangrijke en valide inzichten naar voren gekomen, die in ons model opgenomen zijn. Zoals dit gebied leert, vormen leenwoorden de periferie van het gestratificeerde lexicon, omdat er meer gemaakte structuren in kunnen voorkomen dan in de inheemse woorden. Daarvoor hebben de benaderingen in dit deelgebied een speciaal type RESPECT (Engels: FAITHFULNESS) ontwikkeld: RESPECT(LEENWOORD), dat voor leenwoorden geïndexeerd is. We tonen aan dat de modellen die in dit deelgebied ontwikkeld zijn descriptief adequaat zijn voor ontleendheid. We zijn echter ook tot de conclusie gekomen dat er een aanvullende theorie van het onteningsproces nodig is omdat een al te grammaticale, enkel op RESPECT(LEENWOORD) gebaseerde theorie van het onteningsproces een aantal cruciale eigenschappen daarvan mist.

In hoofdstuk 4, waarin de ingrediënten van ons model gepresenteerd worden, en hoofdstuk 5, waarin morfologische en morfosyntactische integratie besproken wordt, komen de meest belangrijke conceptuele en terminologische innovaties aan bod. Deze twee hoofdstukken zijn verenigd door het diachrone perspectief, dat we het leenwoordtraject noemen: leenwoordintegratie wordt vanuit het oogpunt van de lexicalisatie van een individueel inkomend woord gemodelleerd. In hoofdstuk 4 identificerden we eerst het begin- en eindpunt van het leenwoordtraject: het begin met een oppervlaktevorm (ingevoegd als een codewissel) en eindigt met een volwaardig lexicaal item in het lexicon van de lenende taal. Terwijl het eindpunt verschillende eigenschappen heeft in verschillende talen, heeft de beginvorm, onafhankelijk van de brontaal, een aantal cruciale eigenschappen: het is doorgaans een vorm zonder morfologische en morfosyntactische kenmerken, het is een oppervlaktevorm en heeft een specifieke betekenis. Om een dergelijke vorm naar een lexicaal item om te zetten moet er een degelijke verandering van representatie plaatsvinden. We voeren argumenten aan dat dit proces gecontroleerd wordt door de kracht die oppervlaktekenmerken aantoonbaar ook binnen lexica bescherm - Lexicaal Conservatisme (voortaan LC). De fonologische kenmerken die door LC succesvol zijn beschermd, kunnen aan het eindpunt van het traject indien nodig worden doorgegeven aan RESPECT (al dan niet geïndexeerd voor leenwoorden). In hoofdstuk 5, waarin morfosyntactische integratie wordt besproken, laten we zien dat leenwoordintegratie altijd ook morfologisch is. Leenwoorden komen namelijk niet in de fonologie van een taal terecht: ze worden altijd en vooral onderdeel van het lexicon van de lenende taal en wat 'fonologische' integratie stuurt is doorgaans de integratie in het lexicon. We hebben ook vastgesteld dat het niet de doelstelling van leenwoordenonderzoek
Samenvatting in het Nederlands

moet zijn om de uitkomst van een eenvoudige mapping tussen de vorm van een woord in de brontaal en de vorm van het corresponderende woord in de onttlenende taal te voorspellen, maar veeleer om de ruimte van mogelijke interacties in te perken in het complexe proces waarin tweetalige kennis, lexicon en grammatica verstrengeld zijn in het bepalen van fonologische, morfologische, syntactische en semantische kenmerken van het nieuwe lexicale item. In dataset na dataset wordt er gedemonstrerd hoe dit proces wordt beperkt door een enkele kracht: Lexicaal Conservatisme, dat het behoud van de fonologische kenmerken van de oorspronkelijke oppervlaktevorm in het nieuwgevormde paradigma vereist.

In hoofdstuk 6 komt het ‘synchrone’ perspectief van het voorgestelde model aan bod. We betogen daarin dat de intertalige mappings – de door tweetaligen geïnternaliseerde correspondenties tussen elementen van de brontaal en de onttlenende taal – net zo reëel zijn als de grammatica’s van beide talen. Dit tweede perspectief voltooit de eliminatie uit ons model van wat we grammaticale bias hebben genoemd – de neiging om veranderingen in leenwoorden als een instantiatie van grammatica te beschouwen. Terwijl de mappings op het eerste gezicht zeer grammatica-achtig lijken, betogen we dat ze geconventionaliseerd zijn, veel eenvoudiger dan verhoudingen die middels grammatica’s gecodeerd worden, en dat ze alleen de inputs aankunnen die zich in het feitelijke contact voordoen. Wij stellen dat het LC-gecentreerde perspectief en het mapping-gecentreerde perspectief even crucial zijn voor het opzetten van een integraal leenwoordmodel. Terwijl beide perspectieven niet tegelijk kunnen worden toegepast – LC en intertalige mappings kunnen niet tegelijkertijd in actie worden waargenomen – hebben we wel vastgesteld dat wanneer intertalige mappings verandering ondergaan, die verandering altijd in de richting van meer LC-vriendelijkheid is. Met andere woorden kan LC alleen maar bevorderd worden.

In hoofdstuk 7, waarin een interimaire samenvatting van ons model gepresenteerd wordt, demonstreren we dat het model ons in staat stelt een expliciete beschrijving te geven van het stadium van taalcontact dat de meeste benaderingen tot leenwoordfonologie routinematig veronderstellen: de fase waarin stabiele intertalige mappings bestaan. We noemen dit stadium, waarin een groot aantal woorden intertalige mappings ondersteunen, waardoor veel andere woorden leenbaar en potentieel-al-ontleend zijn, de staat van talenverstrengeling. Dit hoofdstuk kondigt de vier casushoofdstukken aan, die aantonen hoe het in dit proefschrift voorgestelde model met grotere datasets interageert. Tevens dragen deze hoofdstukken bij aan wat gekenschetst kan worden als een typologie van de relaties tussen leenwoordintegratie en grammatica. Onze bespreking van de medeklinkergeminatie in leenwoorden in hoofdstuk 8 brengt een cross-linguïstische catalogus van gerelateerde verschijnselen met dezelfde trigger – Lexicaal Conservatisme. De bevindingen van dit hoofdstuk illustreren een constellatie waarmee rekening dient te worden gehouden: een fonologische aanpassing die niet wordt gestuurd door de fonologie van de onttlenende taal en die niet tot een verandering van de fonologie van de onttlenende taal
leidt. Met andere woorden, presenteert hoofdstuk 8 een empirisch argument voor de conceptuele autonomie van leenwoordintegratiepatronen.

De drie overige casushoofdstukken spitsen zich toe op één taal, Servo-Kroatisch, en onthullen nogal ingewikkelde relaties tussen woordintegratie en de grammatica van de lenende taal.

Hoofdstuk 9, waarin α-epenthesis centraal staat, laat zien hoe LC zoals toegepast op leenwoorden slechts het topje van de ijsberg vormt van de bredere LC-grelateerde beperkingen op lexicalisatie. LC blijkt een beslissende rol te spelen in de manier waarop inheemse woorden heranalyse kunnen ondergaan. We treffen namelijk patronen aan die weliswaar representeerbaar en door de grammatica toegestaan zijn, maar die echter alleen in wug-testexperimenten voorkomen, omdat er geen enkel lexicalisatietraject is waardoor zij tot stand kunnen komen.

De leenwerkwoorddata die in hoofdstuk 10 besproken worden, tonen aan dat, terwijl de conceptuele scheiding van het gebruik van een suffix in leenwoordintegratie en de productiviteit van datzelfde suffix binnen de lenende taal gerechtvaardigd is, het gebruik van het suffix in het leenproces wel invloed kan hebben op zijn morfosyntaxis in afleidingen binnen de taal. Dit hoofdstuk brengt ook de meest gedetailleerde beschrijving van een actieve intertalige mapping in dit proefschrift.

Hoofdstuk 11 laat zien hoe de voorspelde inflexibiliteit en lage productiviteit van de derivationele morfologie die door leenwoorden geïntroduceerd is een voortzetting vormt van wat onafhankelijk is opgemerkt in de Servo-Kroatische inheemse morfologie: het onderscheid tussen wat we noemen de paradigmatische en de niet-paradigmatische patronen. De derivationele patronen geïntroduceerd door leenwoorden worden praktisch toegevoegd aan de 'harde', niet-paradigmatische laag van het lexicon, die door lage productiviteit en weinig semantische transparantie gekenmerkt is.

Terwijl we in dit proefschrift tegen de veronderstelling hebben gepleit dat leenwoorden een gebied zijn waarin grammatica een kans krijgt om 'zichzelf uit te drukken', hebben we de casus aangetoond dat, als leenwoordintegratiepatronen de nodige conceptuele autonomie wordt verleend, het bestuderen daarvan tot nieuwe inzichten over de realiteit van de inheemse patronen kan leiden. Met andere woorden, in plaats van het geven van een universeel antwoord op de vraag naar de relatie tussen leenwoordintegratiepatronen en grammatica, concluderen we dat deze relatie van nogal uiteenlopende aard kan zijn, en dat het rekening houden met deze relatie onderdeel moet worden van de methodologie van descriptief adequate linguïstische modellen, vooral wanneer dergelijke modellen toegepast worden op talen die in contact zijn met andere talen. Wat echter universeel het geval is, is dat leenwoordintegratie het creëren van een lexicaal item is op basis van een oppervlaktevorm onder de controle van Lexicaal Conservatisme en dat tweetaligen contact-specifieke kennis hebben van de correspondenties tussen de structuren van beide talen, de intertalige mappings.
De belangrijkste consequentie van ons model die, naar wij hopen, moeilijk te negeren zal zijn, is dat elk patroon dat kan worden geanalyseerd als een geval van fonologische integratie hand in hand gaat met verschillende andere integratieprocessen, die allemaal bij het vormen van een nieuw lexicaal item horen. Een dergelijk patroon is typisch een gevolg van conventionalisering binnen de tweetalige gemeenschap, en geen spontane reactie van een grammatica op een nieuwe input. Kortom, ons model heeft niet alleen consequenties voor de manier waarop leenwoordverschijnselen moeten worden verklaard, maar ook voor hoe ze moeten worden beschreven.

Omdat ons model betrekkelijk autonome onderzoeksgebieden tracht te verbinden, dwingt het toepassen ervan onderzoekers om over de oorspronkelijke grenzen van deelgebieden heen te stappen en nieuwe relaties aan te gaan. Terwijl we hopen beide soorten gerelateerdheid die ons model definiëren op de agenda van leenwoordonderzoek gezet te hebben, blijft er nog steeds veel over dat toekomstig onderzoek zal moeten verbinden. We hopen dat leenwoordonderzoek, eenmaal bevrijd van de grammaticale bias, op zoek zal gaan naar de precedenten van leenwoordpatronen binnen het inheemse domein. In dit streven, waar we slechts sporadisch aanzijn begonnen, is er veel te leren van lexicon-interne lexicalisatieprocessen (zoals de lexicalisatie van afkortingen) en taal-interne mappings (zoals interdialectale mappings). Tot slot hopen we dat het toekomstige onderzoek consequenties zal hebben voor taalcontactsuitaties die expliciet buiten beschouwing zijn gelaten in onze discussie (bijvoorbeeld situaties waarin geen sprake is van een duidelijke grens tussen codes of geen replicatie van lexicaal materiaal).
Sažetak na srpskohrvatskom

Imigracijska služba leksikona: Prolegomena za teoriju integracije tuđica

Jezici imaju univerzalni potencijal da preuzimaju elemente iz drugih jezika i inkorporiraju ih u svoju strukturu. Cilj lingvističkog bavljenja tuđicama jeste pronalaženje odgovora na pitanja koja se javljaju kada jezični kontakt dovede do preuzimanja reči ili elemenata sličnih rečima. Cilj ove disertacije je stvaranje prostora za konceptualno razdvajanje gramatike i integracije tuđica, odnosno sprečavanje da se njihova povezanost ili identičnost podrazumeva, kao što je slučaj u mnogim modernim pristupima tuđicama. Nakon što se ovo konceptualno razdvajanje završi, proučavanje odnosa između gramatike i integracije tuđica može (i treba) postati delom novog istraživačkog programa.

Nesporno je da tuđice u jezik uvode bilingvali (u najširem smislu reči), i to ponajpre kao rezultat preključivanja kodova — dakle, preuzimajući reč iz leksikona jezika izvornika. Nedvojbeno je, takođe, da na kraju procesa pozajmljivanja tuđica ima reprezentaciju u leksikonu jezika primaoca. Nije, međutim, lako utvrditi kada preuzeta reč prestaje biti rezultat preključivanja kodova i postaje integrirana tuđica, tj. otvoreno je pitanje od kojeg trenutka reč ima reprezentaciju u leksikonu jezika primaoca. Distinkcija između preključivanja kodova i pozajmljivanja jedna je od najkontroverznijih tema u savremenoj literaturi o preključivanju kodova (kao što je objašnjeno u prvom poglavlju) i svaki pokušaj primene ove distinkcije na koherentan način vodi postavljanju daljih važnih pitanja: Za koga data reč mora postati integrirana? (Za pojedinačnog govornika? Za celu zajednicu?). Je li u pitanju nepovratan proces ili se reč može preuzeti, zaboraviti, pa zatim ponovno preuzeti? (I, ponovo, od strane koga?). U literaturi je ovim pitanjima posvećeno mnogo pažnje i ispravan odgovor je, čini se: sve gorenavedeno. Drugim rečima, tuđice se ne uklapaju u jednostavnu teorijsku shemu, već prisiljavaju istraživače da ostanu svesni različitih dimenzija u kojima one postoje i imaju značaj: pojedinačne i grupne, sincrhorne i dijahrone itd.

Mnogi su faktori ovde isprepletani i različite su perspektive moguće. Stoga ne iznenađuje činjenica da se zanimanje za tuđice javilo u nekolikim područjima lingvistike, prelazeći granice disciplina, programa i ideologija. S jedne strane su tuđice tradicionalno tema proučavanja jezičnog kontakta, koje se smatra delom sociolingvistike. U ovoj oblasti, u kojoj dominiraju
pristupi bazirani na uzusu (eng. *usage-based approaches*), tuđice se tretiraju kao deo kontinuuma povezanih fenomena koji uz ostalo čine i preključivanje kodova, prelazak na drugi jezik, mešanje jezika i kreolizacija. S druge strane, tuđice su objekat opsežnih istraživanja unutar fonologije, kojom dominiraju tzv. formalni pristupi. Kako pokazuje diskusija u drugom poglavlju, tuđice zapravo nikada nisu jasno pozicionirane unutar ove tradicije, pa i u današnjem diskursu u velikoj meri ostaje implicitno gde se u arhitekturi jezičke sposobnosti nalaze procesi vezani za tuđice. Drugim rečima, za formalne su fonologe tuđice pre svega zagonetka koju treba rešiti poznatim sredstvima i uz što manje uvođenja mehanizama koji bi se odnosili samo na tuđice.

Cilj modela predloženog u ovoj disertaciji jeste da osnaži oblast formalnih istraživanja integracije tuđica (a) primenom uvida iz sociolinguističkih istraživanja na formalne modele te (b) uključivanjem morfološke i morfosintaktičke integracije u sliku kojom danas dominira fonološka integracija. Tokom koncipiranja novog modela, naše je istraživanje bilo vođeno nekolikim pitanjima klasičnim u modernoj lingvistici: a) Je li proces kojim se bavimo sinhronijski ili dijahronijski? Ako je sinhronijski, pripada li proces sinhronoj gramatici ili nekom drugom modulu? b) Je li dati proces refleks internalizovanog znanja govornika ili nastaje u zajednici? Ako je proces refleks internalizovanog znanja govornika, kako je taj proces mentalno predstavljen? Kao rezultat primene ovog metoda predložen je novi model koji omogućava uvide u sumeđu (engl. *interface*) među jezicima. Ovaj model formalizuje fonološku, morfološku i morfosintaktičku integraciju tuđica. Pored toga ove vrste integracije su iz perspektive našeg modela isprepletani aspekti istog šireg procesa: leksikalizacije. Leksikalizacija je ovde shvaćena kao formiranje nove leksičke stavke na temelju površinske forme iz jezika izvornika. Predloženi model u tom smislu karakterizira određena teličnost, baš kao i danas preovlađujuće adaptacijske modele (razmatrane u drugom poglavlju). Ali za razliku od adaptacijskih modela, u kojima procesi koji se odnose na tuđice imaju rezultat koji se ne razlikuje od domaćih reči, krajnja tačka u našem intergacijskom modelu jeste nastanak leksičke stavke koju karakterizira potpuna funkcionalnost (i koja u principu može biti distinktivno strana).

Budući da su za integracijske procese važni i sinhroni i dijahroni, i individualni i grupni aspekti, naš model, predstavljen u četvrтом, petom i šestom poglavlju, sadrži dve različite perspektive koje nam omogućuju da modeliramo i objasnimo različite aspekte ponašanja tuđica, ne gubeći iz vida one aspekte koji se u datom trenutku nalaze van fokusa. Više dijahronijska komponenta našeg modela fokusira se na načine na koje se obeležja ulaze
Sažetak na srpskohrvatskom

forme iz jezika izvornika čuvaju u procesu integracije u leksikon (koji deli cela jezična zajednica). Deo tog procesa su na primer stvaranje paradigme, dodeljivanje morfosintaktičkih obeležja itd. Druga, više dijahronijska komponenta modela prikladnija je za analizu pravilnosti koje su deo znanja dvojezičnih govornika. Te pravilnosti, koje nazivamo *međujezičnim preslikavanjima*, nastaju u okviru dvojezične zajednice i sadrže upute za pretvaranje elemenata jezika izvornika u odgovarajuće elemente jezika primaoca.


U drugom i trećem poglavlju razmatra se postojeća literatura o fonološkoj integraciji tudića. Primećujemo najpre da zapravo postoje dve oblasti koje se bave fonologijom tudića, koje i terminološki razdvajamo. Postoji, s jedne strane, literatura koja analizira kako tudiće prolaze kroz niz transformacija, te, s druge, literatura koja se usredotočuje na strukturu postkontaktnog leksikona, a posebno na formalizaciju iznimnih čuvanja markiranih struktura u tudićama. Prvu oblast zovemo istraživanjem pozajmljivananja i razmatramo je u drugom poglavlju, dok drugu oblast nazivamo istraživanjem pozajmljenosti i razmatramo je u trećem poglavlju. Diskusija u drugom poglavlju ističe da su postojeći pristupi iz oblasti pozajmljivanja presudni za način na koji danas konceptualiziramo tudiće. Ipak, diskurs u ovoj oblasti je donekle ograničen činjenicom da su ciljevi bavljenja tudićama unapred postavljeni. Ti ciljevi se razlikuju za dve podoblasti: s jedne strane, pristupi bazirani na ulaznom materijalu velikim se delom posvećuju raspri oko najčešće/standarnog ulaznog materijala za fonološku adaptaciju tudića, dok, s druge strane, pristupi bazirani na Teoriji optimalnosti pokušavaju pozicionirati integraciju tudića unutar postojećih modela gramatike i to

U četvrtom poglavlju, u kojem su predstavljene osnovne sastojine našeg modela, te petom poglavlju, u kojem se raspravlja o morfološkoj i morfosintaktičkoj integraciji, razmatraju se i najvažnije konceptualne i terminološke inovacije koje model donosi. Ova dva poglavlja objednajena su dijahronijskom perspektivom, koju nazivamo integracijskom putanjom: integracija se posmatra kao leksikalizacije pojedinačne ulazne reči. U četvrtom poglavlju najpre utvrđujemo početak i kraj integracijske putanje: ona počinje površinskiom formom iz jezika izvornika (umetnutom kao rezultat preključivanja kodova), a završava se formiranjem kompletne leksičke stavke u leksikonu jezika primaoca. Dok je krajnja tačka različito okarakterisana u različitim jezicima, početni oblik ima nekoliko ključnih karakteristika, bez obzira na jezik izvornik. One se ogledaju u činjenici da početna forma nema morfološka i morfosintaktička obeležja, činjenici da je u pitanju površinski oblik, te činjenici da početna forma ima specifično značenje. Da bi se takva forma konvertirala u leksičku stavku, mora se dogoditi temeljita promena reprezentacije. Pokazali smo da je taj proces pod kontrolom leksičke tendencije koja takođe igra ulogu u čuvanju površinskih obeležja unutar leksikona — leksički kozervatizma (Steriade 1997). Na kraju integracijske putanje, ukoliko se ukaže potreba, fonološka obeležja koja je leksički kozervatizam uspešno zaštitio prelaze pod zaštitu odgovarajućeg ograničenja
VERNOSTI (koje može, ali ne mora biti indeksirano na tuđice). U petom poglavlju, u kojem se razmatra morfosintaktička integracija, pokazujemo da je integracija uvek i morfološka. Tuđice, naime, u pravilu ne ulaze u fonologiju jezika: one uvek i pre svega postaju deo leksikona jezika primaoca i „fonološka“ integracija je uvek vođena integracijom u leksikon. Takođe smo utvrdili da cilj istraživanja tuđica ne bi trebao biti predviđanje rezultata jednostavnog mapiranja između oblika reči u jeziku izvorniku i oblika odgovarajuće reči u jeziku primaocu. Cilj našeg modela jeste da ograniči prostor mogućih interakcija u složenom procesu u kojem su isprepletani gramatika, leksikon i znanje dvojezičnih govornika, i u kojem se određuju fonološke, morfološke, sintaktičke i semantičke karakteristike nove leksičke stavke. U nizu kratkih studija slučaja pokazujemo kako taj proces kontrolise ista tendencija: leksički konzervatizam, koji nameće očuvanje fonoloških karakteristika izvornog površinskog oblika u novoformiranoj paradigmi.

U šestom se poglavlju razmatra „sirhrona“ komponenta predloženog modela. U ovom poglavlju pokazujemo kako su međujezična preslikavanja, korespondencije između elemenata jezika izvornika i odgovarajućih elemenata jezika primaoca, deo jezične realnosti baš koliko i gramatike dvaju jezika. Ova druga komponenta dovrašava uklanjanje iz našeg modela onoga što smo nazvali *gramatizmom* — tendencije da se promene u tuđicama sagledavaju kao posledica gramatike. Iako međujezična preslikavanja na prvi pogled mogu izgledati vrlo slična gramatički, pokazujemo kako su ona konvencionalizovana, znatno jednostavnija od preslikavanja koja su kodirana pomoću gramatika, te opremljena samo za konverziju onih ulaznih struktura koje se javljaju u stvarnom kontaktu. Tvrđimo, na kraju, da su komponenta usredotočena na leksički kozervatizam i komponenta usredotočena na međujezična preslikavanja jednako ključne za uspostavu integralnog modela integracije tuđica. Iako se dve perspektive ne mogu koristiti istovremeno — leksički kozervatizam i međujezične korespondencije se ne mogu istovremeno promatrati u akciji — ustanovili smo da kada se međujezične korespondencije menjaju, ta promena uvek biva u smeru boljeg zadovoljavanja leksičkog kozervatizma. Drugim rečima, leksički kozervatizam može se samo unaprediti.

U sedmom poglavlju, u kojem je predstavljen privremeni sažetak našeg modela, demonstriramo kako model omogućuje davanje eksplicitnog opisa stadijuma jezičnog kontakta koji većina pristupâ fonologiji tuđica rutinski podrazumeva: onog stadijuma u kojem postoje međujezična preslikavanja. Mi taj stadijum, gde veliki broj reči podržava međujezična preslikavanja, tako da mnoge druge reči postaju pozajmljive ili potencijalno-već-pozajmljene, nazivamo *stanjem isprepletanosti jezika*. Ovo poglavlje takođe najavljuje četiri poglavlja koja donose studije slučaja u kojima je pokazano kako model...
predložen u ovoj disertaciji interaguje s većim kolekcijama podataka. Ova poglavlja takođe doprinose onome što se može opisati kao tipologija odnosa između integracije tuđica i gramatike.

Rasprava o konsonantskoj geminaciji u tuđicama u osmom poglavlju donosi katalog srodnih fenomena koji se javlja u različitim jezicima i dele isti okidač - Leksički konzervatizam. Nalazi ovog poglavlja ilustruju konstelaciju koja se, kako samo tvrdili tokom uspostavljanja našeg modela, mora uzeti u obzir: tuđice podležu fonološkoj modifikaciji koja nije uzrokovana fonologijom jezika primaoca i usled koje se fonologija jezika primaoca ne menja. Drugim rečima, osmo poglavlje predstavlja empirijski argument za konceptualnu autonomiju obrazaca koji se javljaju u integraciji tuđica.

Tri preostale studije slučaja fokusiraju se na podatke iz istog jezika, srpskohrvatskog, i otkrivaju dosta složene odnose između integracije reči i gramatike jezika primaoca. Deveto poglavlje, koje se bavi nepostojanim a, pokazuje kako su ograničenja leksičkog konzervatizma čiji se efekat vidi u tuđicama samo vrh ledenog brega širih ograničenja na leksikalizaciju koja diktira Leksički konzervatizam. Pokazuje se da leksički konzervatizam odlučuje i o tome na koji način domaće reči mogu biti reanalizirane. Nalazimo, naime, obrasce koje je moguće reprezentovati i koji su dopušteni gramatikom, ali koji se ipak javljaju samo u rezultatima _wug_-eksperimenata, jer ne postoji niti jedna leksikalizacijska putanja kojom bi takvi obrasci nastali.

Analiza glagolskih tuđica u desetom poglavlju pokazuje da, iako je opravdano konceptualno razdvajanje upotrebe sufiksa u integraciji tuđica i produktivnosti toga sufiksa u jeziku primaocu, upotreba sufiksa u procesu pozajmljivanja može uticati na njegovu morfosintaksu u izvedenicama unutar jezika primaoca. Ovo poglavlje donosi i najdetaljniji opis aktivnog međujezičnog preslikavanja u ovoj disertaciji.

Jedanaesto poglavlje pokazuje kako je predviđena nefleksibilnost i niska produktivnost derivacijske morfologije koju uvode tuđice zapravo nastavak tendencije koja je nezavisno uočena u domaćoj srpskohrvatskoj morfologiji: razlike između onoga što nazivamo paradigmatskim i neparadigmatskim obrascima. Derivacijski obrasci uvedeni tuđicama praktički se pridružuju „tvrdom” neparadigmatskom sloju leksikona, koji karakterizira niska produktivnost i malo semantičke transparentnosti.

Iako smo u ovoj disertaciji iznosili argumente protiv pretpostavke da su tuđice područje u kojem gramatika dobija priliku „da se izrazi”, studije slučaja su pokazale da, ako obrasci integracije tuđica dobiju potrebnu konceptualnu autonomiju, njihovo proučavanje može dovesti do daljnje...
rasvetljavanja domaćih obrazaca. Drugim rečima, umesto da damo univerzalizirani odgovor na pitanje odnosa između integracije tuđica i gramatike, možemo zaključiti da je ovaj odnos može biti različite prirode, a da uzimanje toga odnosa u obzir treba uvrstiti u zadatke deskriptivno adekvatnih jezičnih modela, posebno kada se takvi modeli primenjuju na jezike koji su u kontaktu s drugim jezikima. Ostaje, međutim, univerzalno tačno da integracija tuđica jeste stvaranje leksičke stavke na temelju površinskog oblika i pod kontrolom leksičkog kozervatizma, te da bilingvali imaju znanja o korespondencijama između struktura dvaju jezika — međujezična preslikavanja, specifična za dati kontakt.

Ovde predstavljeni model zamišljen je kao pokušaj da se oblast fonološke integracije tuđica učini relevantnom za susedne oblasti — istraživanje preključivanja kodova i morfološke integracije — ali i za opšta jezična pitanja, kao što su arhitektura gramatike i leksikona. Najvažnija posledica našeg modela, koju će, nadamo se, biti teško ignorirati, jeste da svaki uzorak koji se može analizirati kao slučaj fonološke integracije ide ruku pod ruku s različitim drugim integracijskim procesima, a sve kao sastavni deo formiranja nove leksičke stavke. Takav obrazac je obično rezultat konvencionalizacije unutar bilingvalne zajednice, a ne spontana reakcija gramatike na strani ulazni materijal. Ukratko, naš model ne samo da ima konsekvence za način na koji treba objasniti integracijske obrasce, već i za to kako ih treba opisati.

Marko Simonović was born in 1983 in Zaječar, Yugoslavia. While still at high school he attended courses in biology and linguistics at Petnica Research Centre. In 2002 he enrolls in the BA programme in Dutch language and literature at Belgrade University, and in 2003 in the BA programme in General linguistics at the same university. He obtains his bachelor’s degrees in 2006 and 2007, respectively. In 2007, he receives a Huygens scholarship for the international research MA program in linguistics of Utrecht University. In 2009 he obtains his master’s degree (cum laude) from the University of Utrecht with a thesis on loanword phonology. He has had the opportunity to write his own PhD proposal and conduct his PhD research at UiL OTS, Utrecht University. During his PhD he spent one semester at Collegio dei Fiamminghi in Bologna. Marko Simonović has participated in dozens of linguistic conferences, organised three workshops and published articles in various edited volumes, as well as in two international journals.

His other academic interests include sociolinguistics (especially normativity and emancipation), science and technology studies and new materialisms (especially Karen Barad’s agential realism). Marko Simonović is also an online columnist for Peščanik and Dutch review.