

Annual Workshop on Formal Approaches to Slavic Linguistics

The Toronto Meeting **2006**

edited by
Richard Compton
Magdalena Goledzinowska
Ulyana Savchenko

Michigan Slavic Publications
Ann Arbor 2007

Michigan Slavic Publications is a non-profit organization associated with the Department of Slavic Languages and Literatures of the University of Michigan. Its goal is to publish titles which substantially aid the study and teaching of Slavic and East European languages and cultures. The present volume, based on a conference held at the University of Toronto, Toronto, Canada, in May 2006, continues a series of conference proceedings devoted to formal approaches to Slavic linguistics.

Michigan Slavic Materials, 52

Series editor
Jindřich Toman
ptydepe@umich.edu

collection © Michigan Slavic Publications 2007
individual contributions © authors

Library of Congress Cataloging-in-Publication Data

Workshop on Formal Approaches to Slavic Linguistics (15th : 2006 : Toronto, Ont.)

Formal approaches to Slavic linguistics. The Toronto meeting 2006 / editors, Richard Compton, Magdalena Goledzinowska, Ulyana Savchenko. -- 1st ed.

v. cm. -- (Michigan Slavic materials ; vol. 52)

Revised and edited versions of papers presented at the fifteenth annual meeting of Formal Approaches to Slavic Linguistics, held at University of Toronto, Toronto, Canada, May 12-14, 2006.

Includes bibliographical references.

Contents: Clitic reduplication in Bulgarian : towards a unified account / Olga Arnaudova and Iliana Krapova -- A derivational approach to microvariation in Slavic binding / John Frederick Bailyn -- On the clausal and NP structure of Serbo-Croatian / Zeljko Boskovic -- Why clitics cannot climb out of CP : a discourse approach / Jakub Dotlacil -- Double object construction in Croatian : arguments against appl / Martina Gracanin-Yuksef -- Possessive reflexives in Russian / Anna Grashchenkova and Pavel Grashchenkov -- Parasitic gaps in Russian / Natalia Ivlieva -- (Non) D-linking at the Macedonian left periphery / Slavica Kochovska -- An anti-intervention effect in Czech splits : an argument for late / MergeIvona Kucerova -- On the structure and the distribution of negative concord items in Slavic / Masakazu Kuno -- Polish nickname formation : the case of allomorph selection / Anna Lubowicz, Nathan Go, Nancy Huang, and Sara McDonald -- Last-conjunct agreement in Slovenian / Franc Marusic, Andrew Nevins and Amanda Saksida -- On the grammaticalization of the 'have'-perfect in Slavic / Krzysztof Migdalski -- Micro-variation in clitic-doubling in the Balkan Slavic dialectal continuum / Olga Miseska Tomic -- Split phrases in colloquial Russian : a corpus study / Asya Pereltsvaig -- The syllable is not a valid constituent : evidence from two Serbo-Croatian language games / Olivier Rizzolo -- Multiple WH relatives in Slavic / Catherine Rudin -- The functional structure of imperative phrase markers : evidence from adult and child Slovenian imperatives / Dominik Rus -- The syntactic properties and diachronic development of postnominal adjectives in Polish / Pawel Rutkowski -- On the status of word-initial clusters in Slavic and elsewhere / Tobias Scheer -- Jer vowels in Russian prepositions / Olga Steriopolo -- Another look at multiple WH-questions in Czech / Anne Sturgeon -- Nothing wrong with finite T : non-agreeing accusative impersonal sentences / Luka Szucsich -- Possessives within and beyond NPs / Helen Trugman -- Subjunctive complements of modal verbs in Bulgarian and Macedonian / Valja Werkmann.

ISBN 978-0-930042-97-4 (pbk. : alk. paper) 1. Slavic languages--Grammar--Congresses. I. Compton, Richard. II. Goledzinowska, Magdalena. III. Savchenko, Ulyana. IV. Title. V. Title: Toronto meeting 2006.

PG59.W67 2006

491.8'045--dc22

2007014906

Michigan Slavic Publications
Department of Slavic Languages and Literatures
University of Michigan
3040 MLB, Ann Arbor, MI 48105-1275
michsp@umich.edu
www.lsa.umich.edu/slavic

Preface

Many individuals and organizations contributed to the success of *FASL* 15. Aside from the stimulating presentations and a great audience, we also had fun! We would like to acknowledge the generosity of everyone who participated in coordinating the meeting and the proceedings.

The conference was organized by Magdalena Goledzinowska and Diane Massam from the Department of Linguistics and Christina Kramer from the Department of Slavic Languages and Literatures. Ulyana Savchenko shared in the administrative responsibilities and Richard Compton designed and maintained the website at www.chass.utoronto.ca/fasl15/.

The following organizations provided generous financial support for the conference: the Social Sciences and Humanities Research Council of Canada, the Faculty of Arts and Science at the University of Toronto, the Slavic and East European Language Resource Center at the University of North Carolina at Chapel Hill and Duke University, St. Vladimir Institute in Toronto, the Radovan I. Matanić Bookstore in Toronto, Erudit Russian Books in Toronto, the Department of Linguistics and the Department of Slavic Languages and Literatures at the University of Toronto.

We are greatly indebted to the late Professor Ed Burstynsky from the Department of Linguistics at the University of Toronto, without whose support we could not have enjoyed the feast that was the lunch at the St. Vladimir Institute. A great linguist, teacher and member of the Ukrainian community in Toronto, Ed is missed by countless numbers of colleagues, students and friends.

FASL 15 could not have happened without the time and energy of our student volunteers. A big thank you goes to Michael Barrie, Sandhya Chari, Sarah Clarke, Benjamin Flight, Chiara Frigeni, Amanda Greber, Catherine Macdonald, Vladislav Malik, Vanessa Shokeir and Tanya Slavin.

We are also grateful for the musical performance by F-Zero, an interdepartmental band of linguists and language enthusiasts. It is thanks

Contents

Preface	v
<i>Olga Arnaudova and Iliana Krapova</i> Clitic Reduplication in Bulgarian: Towards a Unified Account	1
<i>John Frederick Bailyn</i> A Derivational Approach to Microvariation in Slavic Binding	25
<i>Željko Bošković</i> On the Clausal and NP Structure of Serbo-Croatian.....	42
<i>Jakub Dotlačil</i> Why Clitics Cannot Climb out of CP: A Discourse Approach	76
<i>Martina Gračanin-Yuksek</i> Double Object Construction in Croatian: Arguments Against Appl ⁰	94
<i>Anna Graščenkova and Pavel Graščenkov</i> Possessive Reflexives in Russian.....	113
<i>Natalia Ivlieva</i> Parasitic Gaps in Russian.....	132
<i>Slavica Kochovska</i> (Non) D-Linking at the Macedonian Left Periphery.....	141
<i>Ivona Kučerová</i> An Anti-Intervention Effect in Czech Splits: An Argument for Late Merge	161
<i>Masakazu Kuno</i> On the Structure and the Distribution of Negative Concord Items in Slavic	180

<i>Anna Lubowicz, Nathan Go, Nancy Huang, and Sara McDonald</i>	
Polish Nickname Formation: The Case of Allomorph Selection	196
<i>Franc Marušič, Andrew Nevins and Amanda Saksida</i>	
Last-Conjunct Agreement in Slovenian	210
<i>Krzysztof Migdalski</i>	
On the Grammaticalization of the 'have'-perfect in Slavic	228
<i>Asya Pereltsvaig</i>	
Split Phrases in Colloquial Russian: A Corpus Study.....	245
<i>Olivier Rizzolo</i>	
The Syllable Is Not a Valid Constituent: Evidence from Two Serbo-Croatian Language Games.....	264
<i>Catherine Rudin</i>	
Multiple Wh Relatives in Slavic	282
<i>Dominik Rus</i>	
The Functional Structure of Imperative Phrase Markers: Evidence from Adult and Child Slovenian Imperatives.....	307
<i>Paweł Rutkowski</i>	
The Syntactic Properties and Diachronic Development of Postnominal Adjectives in Polish	326
<i>Tobias Scheer</i>	
On the Status of Word-Initial Clusters in Slavic (And Elsewhere)	346
<i>Olga Steriopolo</i>	
Jer Vowels in Russian Prepositions.....	365
<i>Anne Sturgeon</i>	
Another Look at Multiple Wh-Questions in Czech	386

<i>Luka Szucsich</i>	
Nothing Wrong with Finite T: Non-Agreeing Accusative Impersonal Sentences	401
<i>Olga Mišeska Tomić</i>	
Micro-variation in Clitic-Doubling in the Balkan Slavic Dialectal Continuum	420
<i>Helen Trugman</i>	
Possessives within and beyond NPs.....	437
<i>Valja Werkmann</i>	
Subjunctive Complements of Modal Verbs in Bulgarian and Macedonian	458

FASL 15, 346-364
Michigan Slavic Publications
2007

On the Status of Word-Initial Clusters in Slavic (and Elsewhere)*

Tobias Scheer
Université de Nice, CNRS 6039

1 Introduction

It is generally admitted that gaps in word-initial consonant clusters are systematic, not accidental. An oft-quoted illustration is the pair *blick* - *lbick*. Both items are not actual English words, however the former could be one because its initial cluster is well formed (sonority increases), while the latter could not for its initial cluster violates sonority sequencing. That sonority sequencing is a property of English grammar is shown by the different attitude that natives adopt in regard of the two words at hand: *lbick* is not a possible word for any speaker, while *blick* could enter the language at any time if it acquired a meaning.

The conclusion, then, is that the set of existing initial clusters in a language qualifies as a natural class and is defined by grammar. In the English case and in many other languages, the natural class in question may be described by the statement "within initial clusters, sonority must increase" (s+C clusters are well-known exceptions; they do not need to be further discussed here). On this count, non-occurring clusters such as #lb are absent since they violate grammar (systematic gap), not because of some lexical accident (accidental gap).

In this paper, I aim at showing that this line of reasoning holds true

* This article is a piece of a larger project regarding the phonological representation of extra-phonological information (Direct Interface, cf. below). The part which is not specific to Slavic has appeared in Scheer (1999; 2004a: §83, 381, 390); the part related to Slavic has been orally presented (e.g. Scheer 2000b, Krísto and Scheer 2005), and appears in Scheer (forth). Thanks to Markéta Ziková who helped preparing the version of the corpus that is used here (version 5,2), and which has been cleared against the Czech and Slovak National Corpora.

only for TR-only languages like English, i.e. where sonority increases in all word-initial clusters.¹ In anything-goes languages where #RT clusters occur on top of #TR sequences, however, the absence of some particular initial cluster is *always* accidental. For instance, #rt occurs in Polish (e.g., *rteć* "quicksilver"), but #rp does not. Unlike in the case of the non-occurring English #lb, Polish grammar, I argue, does not outlaw #rp, which is just as well-formed as #rt. Another obvious reflection of the difference between TR-only and anything-goes languages is the fact that the former instantiate all obstruent-liquid combinations (with the pervasive exception of #tl, #dl), whereas the latter may operate an arbitrary choice among #RT clusters (cf. the Polish example) while still showing all possible obstruent-liquid combinations (but typically including #tl, #dl, as e.g. Czech and Polish).

A consequence of this view is a strictly binary typology. An examination of the surface suggests that anything-goes languages may be more or less close to the TR-only bottom line: Classical Greek has just a few non-TR clusters, Slavic languages show quite a number of them, while Moroccan Arabic instantiates all logically possible sequences of Ts and Rs. Counter to this impression of gradience, a consequence of what I propose is that *all* anything-goes languages share the same grammar as far as word-initial clusters are concerned: *any* sequence is well-formed, and those which do not occur may enter the language tomorrow. Thus #rp is absent in both English and Polish, but for different reasons: it is a systematic gap in the former, but a mere accidental gap in the latter. Hence there are only two types of languages phonologically speaking: those which impose a restriction on initial clusters (TR-only), and those which do not (anything-goes).

The following arguments in favour of this perspective are brought to bear. In Slavic anything-goes languages (which will be examined in detail), new words (loans, acronyms) with non-occurring initial clusters may freely enter the language. Also, it is not true that occurring clusters (or non-occurring clusters for that matter) constitute a natural class: looking at them from all possible angles, there is no principle that allows to characterise all and only those sequences which are (non-)existing. By contrast, the hypothesis crediting lexical accident is supported by a striking diachronic generalization: *all* modern Slavic #RT clusters have

¹ Here and henceforth, T represents any obstruent, R any sonorant. Unless otherwise specified, #RT is shorthand for all non-#TR clusters, i.e. #RT, #RR and #TT.

come into being through the loss of a yer. The two consonants of a Common Slavic (CS) sequence #RyerT, however, were of course not subject to any co-occurrence restriction. Therefore their reunion through the loss of the yer creates a randomly structured sequence, both as far as its members and as gaps are concerned: #rp does not exist in any Slavic language simply because CS happened not to feature any lexical item beginning with #r-yer-p.

Finally, I show that a particular phonological theory, so-called CVCV where the peculiarities of the word-initial site follow from the presence or absence of an initial CV unit, precisely predicts that there are only two possible grammars for the beginning of the word. By the same token, a prediction is made to the effect that there can be only one "extrasyllabic" consonant at most. While this appears to be true, regular extrasyllabic analyses of #RT clusters do not impose any restriction on the number of extrasyllabic consonants that could occur in a row.

2 #RT Clusters across Languages

Cross-linguistically, #TR-only and anything-goes systems exhaust the record of languages that admit initial clusters: in some languages no clusters occur at all; in others, only #TRs are found (English, Italian etc.), while in a third group clusters of any sonority profile occur (e.g. modern occidental Semitic, Berber, Slavic). What natural language does not provide for is the fourth logical possibility, i.e. a situation where only non-#TRs are found word-initially: there are no #RT-only languages. Therefore #TR and #RT clusters entertain an implicational relationship: if a language possesses #RTs, it will also feature #TRs; the reverse, however, is not true. All theories need to account for this distribution.

At first sight, thus, we seem to face a binary typology: a cluster-admitting language is either #TR-only or anything-goes. However, let us have a closer look at the domestic typology of the latter group. Languages following this pattern are not very frequent, and their distribution seems to follow fairly robust genetic/ geographic patterns (this of course does not exclude their presence elsewhere). Clements (1990: 288ss) for example provides a list of anything-goes languages, and these typically include (or occur in) modern occidental Semitic (e.g. Moroccan Arabic), Berber, Slavic, Greek, Caucasian languages and Salish (native American Northwest, e.g. Bagemihl 1991).²

² Greenberg (1978: 258) records 12 anything-goes languages, but his data are not

Among these languages, the number and nature of #RT, #RR and #TT clusters show important variation: in some cases only a few sonority-offending sequences are present, while in others many such groups exist. Still in others, all logically possible combinations of two consonants indeed occur word-initially. Moroccan Arabic is on this end of the spectrum (see Scheer 2004a: §383-385 for further detail).

On the other end, Classical Greek is known for word-initial #pt, #kt (as well as aspirated versions thereof) and #mn, but lacks all other sonority-offending groups. One way to approach this pattern would be to acknowledge that the language allows for #TT and #RR (and of course #TR), but not for #RT. This does not explain, however, why it is far from instantiating all logically possible #TTs and #RRs: even on this assumption, the number of existing clusters is only a very small subset of the number of logically possible clusters.³

In the following section, the Slavic family is examined in detail with respect to this variation.

3 #RT Clusters in Slavic

3.1 *The Corpus: Constitution and Organisation*

Over the past years (a first version has appeared in Scheer 2000a), I have attempted to establish a data base which provides an exhaustive record of all words that begin with a sonorant-obstruent cluster in 13 Slavic languages: Czech, Slovak, Polish, Upper Sorbian, Lower Sorbian, Kashubian (West), Bulgarian, Macedonian, Bosno-Serbo-Croatian, Slovenian (South), Russian, Ukrainian, Belarusian (East). For the time being, the inquiry does not include #TT and #RR sequences. This is only because a corpus containing all non-TR clusters represents a workload that would have exceeded my resources. I have therefore privileged the clusters that are most offending for sonority sequencing. Hopefully at some point the list of Slavic #RR- and #TT-words will also be available.

Unfortunately, the corpus is too large to be included in the present article. It is available at www.unice.fr/dsl/tobweb/classes.htm#sldata,

reliable: among these are French, Persian and Icelandic.

³ I am not familiar with the distribution of all anything-goes languages quoted, but a reasonable assumption is that there is none which restricts offending initial clusters to either #RT, #TT or #RR (or any two-member combination thereof) and within this putative group exhausts all logical possibilities. A reviewer points out Georgian (Caucasus) as a putative counterexample. I leave the question open for further study.

where its internal organization and the methodology that has been used for its construction are also explained in detail. Words are arranged according to the Common Slavic root that they represent. All in all, this produces a set of 41 CS roots.

3.2 Broad Results: More or Less #RT-Friendly Slavic Languages

(1) distribution of #RT clusters over Slavic languages

		West						South				East		
		Cz	Sk	USo	LSo	Po	Ka	Bu	Ma	BSC	Sn	Bru	Uk	R
jT	jd	+												
	jh	+												
	js	+												
rT	rb									+				
	r̄ts	+				+								
	r̄tj	+												
	rk, řk	+												
	rd, r̄d̄z, r̄d̄z̄	+				+				+				+
	rz	+								+				
	r̄s	+		+	+	+						+	+	
	rt	+								+		+	+	
	rv, řv	+								+		+	+	
lT	lb	+				+								+
	lg, lh	+				+	+					+	+	
	l̄s	+	+			+								
	lz	+		+		+								
	lk	+	+			+								
	lp	+	+											
	ls, l̄ç	+	+			+								+
	lj	+											+	
lv	+					+						+	+	
mT	md	+	+			+								+
	mg, mh	+				+	+							+
	m̄s	+				+								+
	mz	+	+	+		+								+
	mx	+				+								+
	m̄j	+	+	+		+	+							+
	mk	+				+	+							+
	m̄j̄													+
	ms, m̄ç	+	+			+								+
	mt	+												+
Total: 31	28	8	4	1	20	4			5			12	16	

Table (1) above shows the broadest information provided by the corpus: for every #RT recorded, it indicates whether ("+") or not (blank) it occurs in the 13 individual Slavic languages (no indication is given of the

number of words/ roots that incarnate a particular cluster).⁴

The result is clearly scalar: every Slavic language seems to make its own selection among #RT clusters, whose number may range from "zero" (Bulgarian, Macedonian, Slovenian, Belarusian) over "almost none" (Sorbian 1,4, Kashubian 4), "some" (Slovak 8) and "quite some" (Ukrainian 12, Russian 16) to "a whole lot" (Polish 20, Czech 28). No language, however, attests the full set of logically possible #RT clusters. Even the most permissive systems are far from that: out of 126 logically possible #RT sequences (6 sonorants, 21 obstruents), Polish selects 20, i.e. about 16%. Czech attests 28 combinations out of 108 logically possible clusters (6 sonorants, 18 obstruents), which amounts to 26%.

This situation strongly contrasts with the one found in TR-only languages where all possible obstruent-liquid clusters exist (at least all stop-liquid clusters, except of course the notorious #tl, #dl).⁵

On the bottom line, Slavic languages may thus be reasonably divided into three groups: one where #RT clusters are common (Czech, Polish, Russian, Ukrainian, Slovak), one where they are absent (Bulgarian, Macedonian, Slovenian, Belarusian), and one where #RT clusters are so rare that their synchronic status may be called into question (Upper Sorbian, Lower Sorbian, Kashubian).

The remaining language, Bosno-Serbo-Croatian, is to be counted out altogether since the sonorant of all #RT clusters (only #rT occurs) is syllabic, hence has vocalic function and is not part of any cluster at all. Its syllabicity can be told from various tests, e.g. the fact that it may bear stress (as in *rvati se* "to tussle") and even constitute the only syllabic peak of the word (*rt* "cape (in the sea)"). Also, it is guaranteed that the initial sonorant is not syllabic but trapped in the five permissive languages, (see Scheer 2004a: §240, in press a for further discussion of this difference): stress is initial in Czech, penultimate in Polish, but the R of #RT clusters is never stressed in either language (were the R syllabic, it would receive stress e.g. in Cz *rtut'* and Po *rtęć* "quicksilver" where the vowel is tonic;

⁴ It does not seem to make sense to distinguish #Rd and #R̄d or, for that matter, s, z and ś, ź as well as the corresponding affricates, t̄ [w] and l, g and h. All these pairs appear together in the same line. Even though there is reason to believe that Cz ř is not a sonorant, I continue mentioning it (together with r).

⁵ The situation for nasals is different: usually #TN does not occur (e.g. English, French), and even if it does (like in German), the variety of clusters found is vanishingly small (#kn, but not #pn, #km etc.). Perhaps it is more appropriate to talk about TL-only languages anyway (where L is a liquid).

more tests leading to the same diagnostic, also for the Eastern languages, are provided in the references quoted).

3.3 Slavic #RTs Never Form a Natural Class

Let us now look at the qualitative aspect of Slavic #RT clusters. Were they controlled by grammar, both the occurring and the non-occurring #RTs, in a given language, should form a natural class. All attempts at parsing the #RT-set of any of the 9 #RT-displaying languages under (1) into a natural class are vain. Whatever the criterion or the feature or combination of features used (sonority, nasality, place etc.), the #RT-set of all languages will resist exhaustive assignment: some clusters that according to the natural class should exist are absent, and some that are outlawed do occur.

In order to see this, let us follow the strongest case strategy. Polish is by far the best studied Slavic language as far as phonology is concerned, and this is especially true for initial clusters. Starting with Kuryłowicz (1952), a traditional topic in Polish phonology has been to find the guiding principle which is able to tell occurring from non-occurring initial clusters. The exhaustive inventory of Polish initial clusters on which all analytic work is based has been established by Sawicka (1974) (see also Rowicka 1999: 309ss and Scheer 2004a: §§375, 622). Relevant analytic literature includes Rubach and Booij (1990), Gussmann (1991), Cyran and Gussmann (1998, 1999) and Rowicka (1999).

Kuryłowicz' (1952) double onset analysis sets the frame for later attempts at solving the problem that are couched in Government Phonology: he contends that Polish respects sonority sequencing just like other languages, but unlike these allows words to begin with two onsets in a row, both of which may branch. Implicit in this line of thought of course is the existence of an intervening empty nucleus. However, as Cyran and Gussmann (1998:129) point out, this approach still overgenerates a lot: "While it [Kuryłowicz' proposal] succeeds remarkably well in covering the existing forms by reducing the heavy consonant groups to simple one- or two-member sequences, it does so at the expense of predicting a massive number of forms which do not and cannot exist. [...]. It is easy to think of numerous cases where the mirror-image situation [of existing #CC clusters] is not possible: although we find [kr, pr, gn, tn] [...], no reversing of elements is possible *[rk, rp, ng, nt]." Cyran and Gussmann (1998,1999) then provide a government-based analysis that improves on Kuryłowicz' results, but admit that even then Polish clusters resist a proper characterisation in terms of natural classes: "in fact [r] can only be

followed by some obstruents and never by sonorants, while [n] cannot be followed by anything. Likewise [m] can be followed but not preceded by a sonorant. [...]. Regularities of this sort fail to result from the licensing mechanism called PG. [...] These complex issues are not fully understood at present" Cyran & Gussmann (1998: 135).

We are thus left with an anarchic picture that no principle seems to be able to explain. It is certainly possible to argue that our current understanding of possible guiding principles is too poor, and that progress in phonological theory one day will shed light on the mystery. This perspective, however, is not really promising because rather than about theory we are talking about basic descriptive categories (sonority, place, manner etc.), which are not likely to evolve. Also, the strongest case for a systematic gap analysis is certainly Polish: linguists have invested more time into the study of Polish initial clusters than into the study of initial clusters of any other Slavic language. If the result is negative here, it is unlikely to be positive elsewhere.

Let us therefore think of a solution which dispenses with the idea that things should be explained at all: the clusters and gaps are not enforced by grammar; rather, they are the result of lexical accident. The following section provides strong diachronic support for this alternative.

3.4 All Slavic #RTs Are Produced by the Loss of a Yer

Contrasting with the synchronically anarchic situation, there is an obvious and absolutely exceptionless diachronic generalisation that is brought to light by the corpus: *all #RT-words in all Slavic languages have been produced by the loss of an intervening yer*. That is, all modern #RTs come from a CS #R-yer-T sequence. This hard distributional fact can hardly be accidental. All theories need to account for it: there is a causal relation between the loss of yers and the particular #RTs that occur.

This causal relation is immediately obvious when considering the fact that as in any other language, there was no co-occurrence restriction between C_1 and C_2 of a Common Slavic # C_1 -V- C_2 sequence. In case V happened to be a yer, thus, a # C_1 - C_2 cluster was "mechanically" created as the yer was lost. In absence of any reaction against the new cluster, we therefore do not expect any co-occurrence restriction between the two consonants in the resulting language. And, of course, no particular distribution of C_1 and C_2 either: the choice of C_1 and C_2 was arbitrary in CS, and so it is in the new clusters. In short, thus, Czech initial #RTs for example are just CS #RyerTs minus the yers.

This perspective explains the absence of co-occurrence restrictions on modern #RTs and hence the failure of attempts at characterising them as a natural class. And it directly captures the causal relation between the existing #RTs and the loss of yers.

There is thus good reason to believe that modern #RTs are lexical accident, i.e. the regular Common Slavic lexical accident catered down to modern languages. Missing #RTs are therefore accidental gaps: there happened not to be any CS word with an intervening yer.

3.5 Predictions and Suggestions

This analysis makes two predictions and suggests that the diachronic scenario for non-Slavic anything-goes languages could have been identical or similar.

The first prediction concerns the creation of lexical items: since grammar does not rule out any #RT cluster, new words may enter the language with any logically possible sequence. Hence #lb does not occur in either English or Ukrainian; *lbick*, however, is a possible word in the latter, but not in the former. The English TR-only grammar rules it out, while the Ukrainian anything-goes grammar admits any cluster. This contrast is confirmed by the harshly negative attitude of English (or French, or German etc.) natives, as opposed to the judgement of Czech, Polish etc. natives ("it sounds strange, but I could imagine a word of that shape"), which is comparable to the English situation of *blick*.

Possible sources for lexical creation are loans and acronyms. The former supposes a donor language with #RT-words, which is not easy to come by even in a globalised world: Slavic languages are unlikely to borrow from modern occidental Semitic, Berber or Salish. One obvious case, however, are Caucasian languages which feature #RT-words and have been in intimate contact with Russian. Looking at Russian loans of Caucasian origin, the prediction is borne out:⁶ the words *Mcyri* "poem by Lermontov, and the corresponding character", *Mckheta* "town in Georgia", *Mtacminda* "mountain in Tbilisi", *rkaciteli* "popular brand of wine", *Rza* "personal name" all bear an #RT cluster that does not occur in native Russian items ([*mts,mt,rk,rz*]). They are nonetheless borrowed without any modification of the initial cluster, receive regular inflection

⁶ On the basis of responses to a query published on Linguist List in February 2001 (volume 12-358), the following data have been verified with several native speakers. Judgments are uniform.

etc. Trying to make an English, German, Italian etc. native accept this kind of borrowing fails miserably.

The second prediction concerns non-occurring #RTs and is of diachronic nature: the reason why a given #RT does not exist in a modern Slavic language is diachronic accident: either there was no CS basis, i.e. there happened not to be any CS word with this particular #R-yer-T sequence, or the existing CS basis did not make it to the modern language (or was modified). That is, the reason why there is no #nT in any Slavic language must be the absence of a CS #n-yer-T basis. Note that on phonological grounds this absence, compared to the frequent occurrence of #mT, is unexpected: m is certainly marked with respect to n; hence if anything, #mT should be missing. The diachronic prediction, again, appears to be correct: etymological dictionaries (e.g. Havlová et al. 1989-2006:557s, Holub and Kopečný 1952: 241, Machek 1957: 321) do not have a single CS #n-yer-T-V root on record.⁷

The overall diachronic scenario is thus as follows: Common Slavic was a regular TR-only language until the loss of yers "blindly" created offending #RT sequences. Individual Slavic dialects had different responses to this new situation: either they maintained the TR-only grammar, or they switched to an anything-goes grammar. Languages in the former case either refused yer loss that would have created #RT clusters (hence yers were vocalised even in weak position), or modified the resulting #RTs (all kinds of strategies are attested, cf. the corpus: metathesis, epenthesis, loss of C₁ or C₂). On the other hand, languages in the latter situation did not show any reaction. This split sets apart Belarusian, Slovenian, Bulgarian and Macedonian (following the former pattern) from Czech, Polish, BSC, Russian and Ukrainian (following the latter pattern). Within the latter group, the varying number of different #RTs is the result of further diachronic evolution: individual roots may or may not have survived, may or may not have been modified etc.

Finally, the diachronic scenario together with the claim that there are only two grammars regarding initial clusters suggests that #RT clusters in non-Slavic anything-goes languages could also have arisen through the

⁷ The sequence #n-yer-T must indeed be followed by a vowel since otherwise the yer will be regularly vocalised: in roots of the shape #n-yer-T-C it occurs in so-called strong position. Dictionaries actually offer one single item of this kind, CS **nštv* "trough" (< IE **nigw* "washing", e.g. gr *víceiv* "to wash"), which indeed shows regular yer vocalisation in all modern reflexes: BSC *načve*, Cz *necky*, Po *niecka*, Old Ru *načvy*.

loss of intervening vowels. On this assumption, the erratic occurrence of particular #RTs in Slavic is due to the fact that in Common Slavic 1) only two vowels out of eleven (the yers) have fallen out and 2) any vowel can be the first vowel of a word. By contrast in a language like Classical Arabic which has produced the modern varieties where all logically possible #RTs occur, the prediction is that either more vowels have been lost in #C₁-V-C₂ sequences, or the distribution of the first vowel of a word is not random. The latter scenario happens to be true: Semitic has non-concatenative morphology, which means that the three consonants of a stem carry its lexical meaning, while the two intervening vowels are actual morphemes: in Cl. Arabic katab-a "write 3m sg act pf", the first (short) [a] is the marker for active voice and thus will be present in all verbs (the second [a] is a class marker). Now diachronically, all Cl. Arabic short vowels have become schwa in the dialects at hand, while all long vowels have become short. Schwa, however, alternates with zero according to the pattern [kɪb] (representing /kɪb-i/ from Cl. Arabic katab-a "write 3m sg act pf") vs. [kɪb-u] (representing /kɪb-u/ from Cl. Arabic katab-uu "write 3m pl act pf"). Hence the 3m sg act pf of all verbs in modern occidental varieties will "mechanically" produce #C₁C₂ from underlying /C₁iC₂ibC₃-i/. Therefore all logically possible #RTs indeed do occur in Moroccan Arabic: unlike in Common Slavic where only two eleventh of the lexicon were concerned (on the assumption that all vowels had equal distribution), the intervening vowel is lost in all verbs.⁸

4 CVCV: The Initial CV Imposes the Binary Typology at Hand

4.1 CVCV and the Representation of Extra-Phonological Information

This section shows that a particular phonological theory, so-called CVCV, offers only space for two different grammars regarding the beginning of the word, i.e. exactly the picture established by the preceding discussion.

CVCV is an offspring of Government Phonology (Kaye et al. et al. 1990, Harris 1994) that has been introduced by Lowenstamm (1996).⁹ The

⁸ Amimi and Bohas (1996) discuss the Arabic pattern in greater detail. The Greek situation (why is there only a small subset of possible #TTs, why are there no #RT clusters at all?) has a diachronic solution as well, which is exposed in Seigneur-Froli (2003, 2006, in press).

⁹ Relevant references are, among others, Scheer (2004a), Scheer and Szigetvári (2005), Cyran (2003), Szigetvári (1999).

central idea of this theory is the lateralisation of structure and causality: instead of the familiar syllabic arborescence, lateral relations among constituents are responsible for the effects observed. CVCV takes this line of reasoning to its logical end: it holds that syllabic constituency boils down to a strict sequence of non-branching Onsets and non-branching Nuclei in all languages. There are no Codas and no Rhymes, and the minimal syllabic unit that may be manipulated is an Onset followed by a Nucleus.

In traditional approaches, syllabic arborescence assures the function of binding together different constituents, thereby identifying their grouping into higher units. In CVCV, this function is shifted onto lateral relations that are assumed to hold between constituents: Government and Licensing. Effects that are usually attributed to the fact that a given segment belongs to this or that syllabic constituent are claimed to stem from the configuration regarding Government and Licensing that it is involved in. Space restrictions preclude further discussion of the properties of the theory, and the reader must be referred to the references mentioned in note 9.

In CVCV, the beginning of the word is represented by an empty Onset followed by an empty Nucleus, the initial CV. That is, Lowenstamm (1999) proposes a truly phonological identity for what traditionally has only a diacritic existence, i.e. "#". This take follows the principle of Direct Interface according to which extra-phonological information must not come down to the phonology as a diacritic (such as SPE-type boundary symbols, brackets in Lexical Phonology or the Prosodic Hierarchy). Rather, only objects that are needed by the domestic phonology in absence of any issue related to the interface may represent morpho-syntactic information. Hence a CV unit (but neither of the diacritics mentioned) qualifies as a representative of morpho-syntactic information in the phonology.¹⁰

The initial CV thus represents morpho-syntactic information in the phonology. It is therefore absent from the lexical recording of words. Like in all other theories (SPE, Prosodic Phonology, OT etc.), the representation to which phonological computation applies is pieced together from the lexical entries of the morphemes and objects that represent morpho-syntactic information. The latter (#s, the Prosodic Hierarchy, brackets etc. in the theories mentioned, the initial CV here) are thus distributed by morpho-syntax, typically through some kind of

¹⁰ See Scheer (2005, in press b, forth) for further detail regarding Direct Interface.

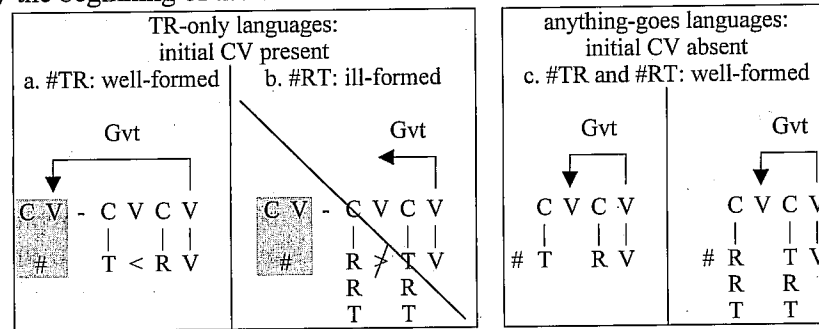
Translator's Office (the Readjustment Component in SPE, mapping rules in Prosodic Phonology, ALIGN and WRAP in OT).

Therefore, the distribution of the initial CV is not automatic: languages may or may not ship off a phonological exponent of the morpho-syntactic information "beginning of the word". Crucially, though, there are only two possible ways of handling this device: the initial CV may or may not be sent down to the phonology. There is no third possibility. Note that this contrasts with other interface theories where the number of different diacritics that may come down to the phonology is unlimited in principle (# and the like, various prosodic constituents).

4.2 The Beginning of the Word in CVCV

Given these premises, (2) below shows the representation of word-initial clusters in languages where the initial CV (grey-shaded) is distributed (which will turn out to be the TR-only type), and in those where the beginning of the word is left unmarked by any extra-phonological information (absence of the initial CV, these languages will turn out to be the anything-goes type).

(2) the beginning of the word in CVCV



The ground rule in Government Phonology is that empty Nuclei do not come for free: they need to be governed in order to exist. However, only phonetically expressed Nuclei can govern, and every governor can only govern one target. Therefore, two empty Nuclei in a row are ill-formed. In the version of CVCV that I use (others include Szigetvári 1999, Cyran 2003 and Rowicka 1999), all lateral relations are head-final. Also, the representation of TR clusters, i.e. traditional branching Onsets, is as under (2a): being more complex (i.e. made of more phonological primes), sonorants may establish a relation called Infrasegmental

Government over obstruents (represented as "<" under (2a), as ">" under (2b)). This relation has the effect of satisfying the Empty Category Principle (ECP) of the enclosed Nucleus.¹¹

With that in mind consider (2a) and (2b): both representations present two empty Nuclei, i.e. the one separating the initial cluster and the one from the initial CV. Under (2a), the former is taken care of by the relation between the two consonants, while the ECP of the latter is satisfied through Government from the first vowel of the word. The representation is thus well-formed. Under (2b), however, no relation between consonants can be established: either it would have to be left-headed (RT), or equal complexity does not allow for it (RR, TT). Therefore the intervening empty Nucleus still requires to be silenced, and the first vowel of the word would need to simultaneously govern two empty Nuclei. Since this is impossible, the structure is ill-formed.

On the bottom line, thus, the difference in complexity between obstruents and sonorants, together with the pressure that the initial CV puts on the structure by loading it with an extra empty Nucleus, causes the well-formedness of #TRs and the ill-formedness of #RTs. Or, in other words, the presence of the empty CV enforces the restriction of initial clusters to #TRs.

On the other hand, its absence under (2c) has the effect of leaving the structure with only one empty Nucleus that requires care no matter what the initial cluster. Therefore, the sonority slope of the sequence is irrelevant: whatever it is, the first vowel of the word will be able to govern the intervening empty Nucleus. The representation will thus be well-formed in any event: (2c) describes anything-goes languages.

4.3 Benefits: Binary Typology, *#RT-Only, No Multiple Extrasyllabicity

It follows that CVCV and the initial CV make a prediction to the effect that there can be only two different grammars in natural language regarding the left edge of the word: either the initial CV is present and the TR-restriction is enforced, or it is absent and any cluster is well-formed. This is precisely what the empirical discussion of sections 2 and 3 demands: a theory that does not care for the apparent surface gradience of more or less permissive languages (#TR plus a little, some, quite some, a

¹¹ The two ways of satisfying the ECP that have been mentioned are enough for the present purpose. More detail regarding the phonological ECP and the functioning of branching Onsets is available in Scheer (2004a:§14).

lot or all #RTs). As far as grammar is concerned, there are only two types of languages, TR-only and anything-goes.

Moreover, if a language has #RTs, it must also possess #TRs. This is so since the presence of the former implies the absence of the initial CV, which in turn makes any cluster well-formed. By contrast, the presence of #TRs in a language does not allow to predict whether #RTs also occur: they may (if the initial CV is absent) or may not (if it is present). And in any event, #RT-only languages are not an option: if #RTs exist, #TRs must also occur. Now recall from the beginning of section 2 that this is precisely the cross-linguistic typology that theory is called to derive. The absence of #RT-only languages and the implicational relationship between #RTs and #TRs thus come for free with the theory of the beginning of the word that has been laid out.

Finally, let us look at another approach to #RTs that, like the analysis presented, does not try to force existing and non-existing items into natural classes. In generative phonology, the traditional analysis of clusters that violate sonority sequencing is along the lines of extrasyllabicity. On this count, a syllabification algorithm operates over a lexically unsyllabified string and leaves the sonority-offenders unparsed. For example, Polish *rtęć* "quicksilver" will end up as <r>teć where angled brackets identify the unparsed, i.e. extrasyllabic element. Phonology may then operate, and at some point of the derivation, the extrasyllabic consonant is reintegrated into the syllabic or prosodic arborescence (otherwise it would not make it to the surface). The exact location to which extrasyllabic consonants are adjoined differs across analyses: sometimes they end up in truly syllabic constituents (the Onset in the case quoted, e.g. Hall 1992: 122ss), sometimes in higher prosodic constituents (the prosodic word, e.g. Rubach and Booij 1990, Rubach 1997, in violation of the Strict Layer Hypothesis). In the former case, sonority sequencing is said to hold only at the "deep" level where syllabification operates (while an Onset can contain anything on the surface), while the latter does not need any specific statement since no co-occurrence restrictions are defined for prosodic constituents anyway.

The problem with all versions of extrasyllabicity is that they allow for random strings of extrasyllabic consonants: since prosodic constituents and Onsets (at the surface level) can contain anything, words with three, five or 17 extrasyllabic consonants in a row, e.g. /rgtflpen/, should be well-formed. Obviously natural language does not produce this kind of monsters. I show in Scheer (2004a: §373,b) that even systems which have

an extremely liberal reputation such as Salish languages and Polish do not produce extrasyllabic clusters.

Unlike the extrasyllabic approach, the theory outlined above predicts that there could be no more than one extrasyllabic consonant in a row. A (left-edge) extrasyllabic consonant is a consonant that occurs at the left edge of a word and whose Nucleus is governed (C₁ under 2c). Hence the adjunction of any further consonant supposes the presence of an additional empty Nucleus. Two or more empty Nuclei in a row as e.g. in /rətøken/, however, are ill-formed since the first vowel of the word will be unable to govern all of them. Therefore one sonority-offending consonant at most can be accommodated on the count of CVCV.

5 Conclusion

The foregoing pages have applied the theory of Direct Interface to the beginning of the word: rather than by diacritics such as "#" or elements of the Prosodic Hierarchy (e.g. phonological words), the representation of morpho-syntactic information in the phonology can only be effected by truly phonological material. Unlike the former (#s neither demand nor invalidate initial clusters), the latter have a direct impact: the initial CV burdens the phonology with an additional empty Nucleus, which makes non-#TR clusters illegal.

Therefore, a prediction is made to the effect that natural language can only produce two different grammars regarding the beginning of the word: TR-only and anything-goes. All surface gradience suggesting that some languages are more permissive than others reduces to either pattern. Eventually missing clusters are accidental, not systematic; they may be happily filled in by new words.

References

Items followed by "WEB" are available at www.unice.fr/dsl/tobias.htm.

- Amimi, Abdeljabber, and Bohas, Georges. 1996. Les formes nominales [CeCC] et [CCeC] en arabe marocain ou la persistance des schèmes. *Revue canadienne de linguistique* 41: 1-28.
- Bagemihl, Bruce. 1991. Syllable structure in Bella Coola. *Linguistic Inquiry* 22: 589-646.

- Clements, George. 1990. The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I*, eds. John Kingston and Mary Beckmann, Cambridge: Cambridge University Press, 283-333.
- Cyran, Eugeniusz, and Gussmann, Edmund. 1998. Polish consonantal sequences: a phonological testing ground. In *Structure and Interpretation*, ed. Eugeniusz Cyran, Lublin: Pase. WEB, 127-138.
- Cyran, Eugeniusz, and Gussmann, Edmund. 1999. Consonant clusters and governing relations: Polish initial consonant sequences. In *The Syllable, Views and Facts*, eds. Harry van der Hulst and Nancy Ritter, Berlin: Mouton de Gruyter, 219-248.
- Cyran, Eugeniusz. 2003. *Complexity Scales and Licensing Strength in Phonology*. Lublin: KUL.
- Greenberg, Joseph. 1978. Some generalizations concerning initial and final clusters. In *Universals of Human Language*, ed. Joseph Greenberg, Stanford, CA: Stanford University Press, 243-279.
- Gussmann, Edmund. 1991. Polish syllable structure: A hypothesis and its problems. In *Words are Physicians for an Ailing Mind*, eds. Maciej Grochowski and Daniel Weiss, München: Sagner, 207-213.
- Hall, Tracy. 1992. *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Niemeyer.
- Harris, John. 1994. *English sound structure*. Oxford: Blackwell.
- Havlová, Eva, Erhart, Adolf, and Janyšková, Ilona (eds). 1989-2006. *Etymologický slovník jazyka staroslověnského. 1st - 13th booklet A-sice*. Praha: Nakladatelství Československé Akademie Věd/ Akademie Věd České Republiky.
- Holub, Josef, and Kopečný, František. 1952. *Etymologický slovník jazyka českého*. Praha: Státní nakladatelství učebnic.
- Kaye, Jonathan, Lowenstamm, Jean, and Vergnaud, Jean-Roger. 1990. Constituent structure and government in phonology. *Phonology* 7: 193-231. WEB.
- Kristo, László, and Scheer, Tobias. 2005. The beginning of the word in Slavic. Paper presented at Formal Description of Slavic Languages 6, Potsdam. WEB.
- Kuryłowicz, Jerzy. 1952. Uwagi o polskich grupach spółgłoskowych [Remarks on Polish consonantal groups]. *Biuletyn Polskiego Towarzystwa Językoznawczego* 11: 54-69.
- Lowenstamm, Jean. 1996. CV as the only syllable type. In *Current Trends in Phonology. Models and Methods*, eds. Jacques Durand and Bernard Laks, 419-441. Salford, Manchester: ESRI. WEB.
- Lowenstamm, Jean. 1999. The beginning of the word. In *Phonologica 1996*, eds. John Rennison and Klaus Kühnhammer, La Hague: Holland Academic Graphics. WEB, 153-166.
- Machek, Václav. 1957. *Etymologický slovník jazyka českého a slovenského*. Praha: Nakladatelství Československé Akademie Věd.

- Rowicka, Grażyna. 1999. On ghost vowels. A strict CV approach. Doctoral dissertation, University of Leiden.
- Rubach, Jerzy, and Booij, Geert. 1990. Edge of constituent effects in Polish. *Natural Language and Linguistic Theory* 8: 427-463.
- Rubach, Jerzy. 1997. Extrasyllabic Consonants in Polish: Derivational Optimality Theory. In *Derivations and Constraints in Phonology*, ed. Iggy Roca, 551-581. Oxford: Clarendon.
- Sawicka, Irena. 1974. *Struktura grup spółgłoskowych w językach słowiańskich*. Wrocław, Warszawa, Kraków, Gdańsk: Wydawnictwo Polskiej Akademii Nauk.
- Scheer, Tobias. 1999. A theory of consonantal interaction. *Folia Linguistica* 32: 201-237. WEB.
- Scheer, Tobias. 2000a. De la Localité, de la Morphologie et de la Phonologie en Phonologie. Habilitation thesis, University of Nice. WEB.
- Scheer, Tobias. 2000b. Why Moroccan Arabic tolerates anything word-initially, but Slavic does not. Paper presented at the 8th Manchester Phonology Meeting, Manchester. WEB.
- Scheer, Tobias. 2004a. *A Lateral Theory of Phonology. Vol.1: What is CVCV, and why should it be?* Berlin: Mouton de Gruyter.
- Scheer, Tobias. 2004b. A better solution for extrasyllabicity than extrasyllabicity. Paper presented at GLOW 27, Thessaloniki. WEB.
- Scheer, Tobias. 2005. We need a translator's office, but the buffer has to go: Direct Interface. Paper presented at the 36th Poznań Linguistic Meeting, Poznań. WEB.
- Scheer, Tobias. In press a. Syllabic and trapped consonants in (Western) Slavic: Different but still the same. In *Investigations into Formal Slavic Linguistics*, eds. Gerhild Zybatow and Luka Szucsich. Frankfurt am Main: Lang.
- Scheer, Tobias. In press b. Why the Prosodic Hierarchy is a diacritic and why the Interface must be Direct. In *Sounds of Silence*, eds. Jutta Hartmann, Veronika Hegedus and Henk van Riemsdijk. Amsterdam: Elsevier. WEB.
- Scheer, Tobias. forth. *A Lateral Theory of Phonology. Vol.2: On Locality, Morphology and Phonology in Phonology*. Berlin: Mouton de Gruyter.
- Scheer, Tobias, and Szigetvári, Péter. 2005. Unified representations for the syllable and stress. *Phonology* 22: 37-75.
- Seigneur-Froli, Delphine. 2003. Diachronic consonant lenition and exotic word-initial clusters in Greek: A unified account. In *Studies in Greek Linguistics. Proceedings of the 23rd annual meeting of the department of linguistics*, eds. M. Stavrou-Sifaki and A. Fliatouras, Thessaloniki: University of Thessaloniki, 345-357. WEB.
- Seigneur-Froli, Delphine. 2006. Le statut phonologique du début de mot grec. Doctoral dissertation, University of Nice.
- Seigneur-Froli, Delphine. In press. Greek pterio: neither extrasyllabic nor magic. In *Phonologica 2002*, eds. Friedrich Neubarth, Markus Pöchtrager and John Rennison. Berlin: Mouton de Gruyter.

Szigetvári, Péter. 1999. VC Phonology: a theory of consonant lenition and phonotactics. Doctoral dissertation, Eötvös Loránd University Budapest.

Tobias Scheer
 Laboratoire "Bases, Corpus, Langage" UMR 6039
 Faculté des Lettres
 98, Boulevard E. Herriot
 F-06204 Nice
 FRANCE
 scheer@unice.fr

FASL 15, 365-385

Michigan Slavic Publications
 2007

Jer Vowels in Russian Prepositions*

Olga Steriopolo
 University of British Columbia

The goal of this paper is to describe and analyze *jer* vowels in Russian monosyllabic prepositions *v/vo* 'in', *s/so* 'with/from', and *k/ko* 'to'. I show that patterns of *jer* realization in Russian prepositions are different from those in Russian prefixes. I present empirical generalizations on the behavior of prepositional *jers* based on collected data and propose an OT analysis.

Monosyllabic prepositions with *jer* vowels are especially interesting because they alternate between a syllable of their own and a single consonant, thus adding to the problem of the canonical syllable and extra-syllabicity in Russian.

This paper is organized as follows. Section 1 presents the background of the study of *jer* vowels. Section 2 compares the behavior of prepositional *jers* with the behavior of prefixal *jers*. Section 3 lays out assumptions about syllabification and prosodic parsing in Russian. Section 4 illustrates empirical generalizations on *jer* vowels in prepositions. Section 5 proposes an OT analysis of the data. Finally, section 6 presents the conclusions of the study.

1 Background

All Slavic languages have one or two vowels that alternate with zero. Such vowels are called *jer* (or *yer*) vowels. Russian has two *jer* vowels:

* Many thanks to Gunnar Hansson, Patricia Shaw, and Joe Stemberger for their helpful comments and suggestions. Special thanks to the audience at the FASL 15 and two anonymous reviewers for their valuable feedback. This work was supported in part by SSHRC grant 410-2006-2166, awarded to Martina Wiltschko and by SSHRC grant 410-2004-0710, awarded to Gunnar Ólafur Hansson.