

# LINGUISTIK INTERNATIONAL

HERAUSGEGEBEN VON

HEINRICH WEBER  
SUSANNE BECKMANN  
ABRAHAM P. TEN CATE  
WILFRIED KÜRSCHNER  
KAZIMIERZ SROKA  
INGO WARNKE  
LEW ZYBATOW

BAND 26



PETER LANG

Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien

MARKÉTA ZIKOVÁ / MOJMÍR DOČEKAL  
(EDS.)

## SLAVIC LANGUAGES IN FORMAL GRAMMAR

PROCEEDINGS OF FDSL 8.5, BRNO 2010



PETER LANG

Internationaler Verlag der Wissenschaften

**Bibliographic Information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.d-nb.de>.

Cover design: Uli Weber

ISSN 1436-6150  
ISBN 978-3-631-63609-1

© Peter Lang GmbH  
Internationaler Verlag der Wissenschaften  
Frankfurt am Main 2012  
All rights reserved.

All parts of this publication are protected by copyright. Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution. This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

[www.peterlang.de](http://www.peterlang.de)

**Content**

Preface

Mojmír Dočekal Negation and Durative Adverbials.....	1
Dominika Dziubała-Szrejbrovska <i>Pięć kobiet czy pięciu mężczyzn? – Towards the Solution of a Numeral Puzzle in Slavic Languages.....</i>	17
Hana Gruet-Skrabalova What Kind of Element Is <i>že</i> in Czech? .....	33
Petr Jäger, Vladimír Petkevič, Alexandr Rosen & Hana Skoumalová Towards a Treebank for All Tastes.....	49
Tomáš Jelínek Automatic Rule-Based Correction of Stochastic Syntactic Annotation of Czech.....	65
Slavica Kochovska Clitic Co-occurrence Restrictions in Macedonian Wh-Questions.....	81
Pavel Kosek Word Order of Conditional Auxiliary Clitics in the Czech Language of the Baroque Period.....	101
Vesselina Laskova How <i>kogato</i> Differs from <i>kato</i> : Analysis of Bulgarian Adverbial Clauses...	119
Evgenia Markovskaya Derivational Account of Gender in Deverbal Nominals in Russian.....	135
Roksolana Mykhaylyk Scrambling to Middlefield.....	149
Vladimír Petkevič Syntactic Functions of Infinitive in Contemporary Czech.....	165
Tobias Scheer Yers and Epenthetic Vowels in Polish.....	179

Barbora Štindlová, Svatava Škodová, Alexandr Rosen & Jirka Hana Annotating Foreign Learners' Czech.....	205
Beata Trawiński AND-Type versus WITH-Type Conjunctions: Towards a Corpus-Based Study.....	221
Ewa Willim Concord in Polish Coordinate NPs as Agree.....	233
Erin Zaroukian Homophony and Russian Copular Constructions.....	255
<b>Workshop on Slavic Prefixes and Prepositions</b>	
Petr Biskup Slavic Prefixes and Adjectival Participles.....	271
Lucie Medová Anticausatives Are Derived Unergatives.....	291
Bartosz Wiland Prefix Stacking, Syncretism, and the Syntactic Hierarchy.....	307
Markéta Ziková Lexical Prefixes and Templatic Domains: Prefix Lengthening in Czech.....	325
Rok Žaucer The Syntax of Perdurative-prefixed Verbs and the VP-internal/ VP-external Prefix Distinction.....	339

## Preface

This volume comprises papers from the third half time conference *Formal Description of Slavic Languages 8.5* which took place on November 25–27 2010, at Masaryk University, Brno, in the Czech Republic.

The program of the conference, which was completed on the basis of anonymous abstracts, consisted of talks focused on topics ranging from phonology, syntax and semantics to computational linguistics. Nevertheless, all the talks described various Slavic languages from a formal point of view. Out of 29 papers presented at the conference, 21 then appear in a revised and edited version in the present volume.

It was our pleasure to include additional 5 papers from the Workshop on Slavic Prefixes and Prepositions – a very pleasant and thought provoking part of the conference, in our opinion – into a separate section of this volume.

We would like to thank our colleagues who reviewed either abstracts for the conference or papers for this book. And we happily acknowledge the financial support from the Grant Agency of the Czech Republic (GAČR 405/09/0677).

The editors

Markéta Ziková  
Mojmír Dočekal

## References

- Novák, Václav & Zdeněk Žabokrtský. 2007. Feature Engineering in Maximum Spanning Tree Dependency Parser. In *Proceedings of the 10th International Conference, TSD*. Berlin, Heidelberg: Springer-Verlag, LNCS 4629. 92–98.
- Petkevič, Vladimír. 2006. Automatické rozpoznání infinitivu. Případová studie jako příspěvek k automatické disambiguaci českých textů. In F. Čermák & R. Blatná (eds.), *Studie z korpusové lingvistiky I. Korpusová lingvistika: Stav a modelové přístupy*. Praha: NLN. 226–253.
- Šmilauer, Vladimír. 1966. *Novočeská skladba*. Praha: SPN.

## Yers and Epenthetic Vowels in Polish

Tobias Scheer (University of Nice, CNRS 6039)

## 1. Introduction

Polish features two competing patterns of root-final cluster vocalisation in Gpl: monomorphemic -CC# may (*wiadr-o* – *wiader* ‘pail Nsg, Gpl’) or may not vocalize (*cyfr-a* – *cyfr* ‘number Nsg, Gpl’). However, (monomorphemic) clusters always vocalize in presence of a C-initial (or yer-initial) suffix, even those that do not in Gpl: *wiader-k-o* ‘id., dim.’, *cyfer-k-a* ‘id., dim.’. The pattern is also lexically variable: some roots have both vocalized and non-vocalized forms in free variation (*wydr-a* – *wydr/wyder* ‘otter Nsg, Gpl’). Finally, there is cross-speaker variation as well: some speakers may accept a vocalized or unvocalized version of a root in Gpl, while others may not (e.g. *kurw-a* – *kurew/?kurw* ‘whore Nsg, Gpl’).

It is shown below that the pattern cannot be analysed with the regular instrument that is put to use for (Slavic) vowel-zero alternations, i.e. lexically present vowels that are made inaudible by phonological computation (the yers). Most of the literature does not talk about the pattern at all: this is the case for instance of two of the three books that have founded the generative analysis of Polish, Gussmann (1980) and Rubach (1984). By contrast, the third book in this category, Laskowski (1975:29ff), offers a very careful survey of the extremely intricate empirical situation and provides rich material (see also Bajerowa 1953). Laskowski (1975) is couched in linear SPE, and he considers all vowel-zero alternations the result of epenthesis (rather than of deletion, see section 3.1 below). His conclusion is that all kinds of lexical items need to be diacritically marked as an exception to all kinds of rules.<sup>1</sup> Szpyra (1995:97) reaches the same obvious conclusion, but formulates the need for lexical marking in theory-neutral terms: ‘the logical conclusion is that the presence versus absence of yers is largely unpredictable und must therefore be marked in the lexical representation of the relevant items.’

Another interpretation of the pattern that is found in the literature gives up on its phonological character altogether: Gussmann’s (2007) argues that vowel-zero alternations of the kind described are instances of allomorphy, i.e. managed outside of the phonology (section 4.1). Cyran’s (2005) analysis of the (non-)vocalization of word-final clusters, although not explicitly (because forms

<sup>1</sup> The exception feature that Laskowski (1975:25ff) uses is ‘[α RM*f* i]’ where *RM*f** means ‘morphological rule’ and *i* is a variable that refers to the number of the rule.

with C/yer-final suffixes are not considered), also results in an allomorphic solution (section 4.2).

Finally, Bethin (1992:146ff) argues for a scenario whereby vowel-zero alternations are based on regular yers, except in loanwords where they are of epenthetic origin. It is argued below that this approach is on the right track, but needs to be refined: there is no difference between loanwords and native vocabulary (extension of the scope of Bethin's epenthetic analysis), and all vowel-zero alternations in loans do not originate in epenthesis (restriction of the scope of Bethin's epenthetic analysis). Also, the difference between *cyfr-a* – *cyfr* and *wiadr-o* – *wiader* is certainly unpredictable and hence encoded in the lexical recording of each item – but not by way of Laskowski's lexical diacritics, which are also used by Bethin (whose solution is discussed in greater detail in section 5.1 and 5.2). Instead of placing diacritics into phonological representations that alter the course of the phonological computation, I argue that the lexical opposition is achieved by contrasting properties of the (autosegmental) representation itself. That is, there are three (and only three) distinct structures: 1) stable vowels (i.e. which do not alternate), 2) alternating vowels that appear in clusters in Gpl (*wiadr-o* – *wiader*) and 3) alternating vowels that do not appear in clusters in Gpl, but surface before C/yer-initial suffixes (*cyfr-a* – *cyfr* – *cyfer-k-a*). In the analysis below, this triple lexical contrast is expressed in the vocabulary of strict CV phonology (Lowenstamm 1996, Scheer 2004, Cyran 2010).

The three variations mentioned (within Gpl, cross-roots and cross-speaker) are then a consequence of alternative or hesitating lexicalisation of the three lexical representations. As Bethin's, this solution is thus purely phonological and instantiates a currently developed idea in minimalist syntax: variation reduces to variation in the lexicon (the so-called Chomsky-Borer Conjecture: Biberauer 2008; Baker 2008; Roberts & Holmberg 2010).

The gist of the analysis is that an important piece of the standard Slavic yer-based account of vowel-zero alternations needs to be abandoned: it is not true that all vowels which alternate with zero are underlyingly yers – Bethin (1992:153) says that '[v]owel-zero alternations in Polish are not attributable to a unique underlying representation'. Some are yers (in my analysis, but not in Bethin's, those that vocalize in Gpl: *wiadr-o* – *wiader*), while others are not (those that do not vocalize: *cyfr-a* – *cyfr*). The latter are epenthetic vowels, i.e. lexically absent and inserted in order to repair an ill-formed structure (three consonants in a row in surface description, two empty nuclei in a row in the analysis below).<sup>2</sup>

<sup>2</sup> The present chapter is a development of a section in Scheer (2010), where the analysis was first introduced. A more complete version of the chapter, including heteromorphemic clusters and an analysis of unvocalized -CC-C-V forms (section 2.3), appears as Scheer (forth b).

For lack of space, a number of important questions cannot be addressed in the chapter. One is the study of heteromorphemic clusters, i.e. where the first consonant belongs to the root and the second to a suffix (e.g. *sluż-b-a* – *slużb* – *slużeb-n-y* 'service Nsg, Gpl, adj.'). Mono- and heteromorphemic clusters are systematically treated on a par in the literature (including by Bethin, an exception being Cyran 2005), although they have quite different properties. Another issue is the comparison of the Gpl with the other zero case marker, Nsg (Asg).

## 2. To vocalize or not to vocalize...

### 2.1 Pattern A and B (in monomorphemic clusters)

Table (1) below illustrates a set of alternations that I will call pattern A where (monomorphemic) root-final CCs do not vocalize in Gpl, but where a vowel does break up the cluster in presence of a C/yer-initial suffix. The leftmost column shows that there is no evidence to the end that the sonority slope of the cluster plays a role: TR, RT and RR clusters may follow the pattern at hand (T is shorthand for obstruents, R for sonorants). Also, (more or less recent) loanwords and native vocabulary are both represented and do not show any different behaviour.

The table contains a number of items (in italics) where the cluster does not vocalize before C/yer-initial suffixes. Section 2.3 below collects and discusses words of this type. Finally, items in brackets identify words that natives will probably never had heard or used, and whose status is thus ad-hoc.

- (1) pattern A: -CC# does not vocalize in Gpl, but vocalizes before C/yer-initial suffixes

TR#	Tr	CC-V# Nsg	CC# Gpl	CeC-C C/yer-initial suffix	gloss
		cyfr-a	cyfr	cyfer-k-a	number
		zebr-a	zebr	zeber-k-a	zebra
		tundr-a	tundr	tunder-k-a	tundra
		bistr-o	bistr	(bister-k-o)	bistro
		Tatr-y	Tatr	tater-nik	toponym
		algebra	algebr		algebra
		siostr-a	siostr		sister
		sutr-a	sutr	(suter-k-a)	sutra
		Piotr-a	Piotr	<i>Piotr-k-a (!)</i>	first name
		ikr-a	ikr	(ikier-k-a)	fish eggs
		podagr-a	podagr	(podagier-k-a)	gout

	Tn	blizn-a	blizn	(blizen-k-a) (blizn-k-a (!))	scar
		tętn-o	tętn	tęten-k-o	pulse
		piętn-o	piętn	pięten-k-o	stamp
	Tm	wydm-a	wydm	wydem-k-a	dune
		taśm-a	taśm	tasiem-k-a	ribbon, tape
		jarzm-o	jarzm	jarzem-k-o	yoke
		pism-o	pism	pisem-k-o	document
				pisem-n-y	
		drachm-a	drachm	drachem-k-a	drachma
	Tv	pletw-a	pletw	pletew-k-a	fin (fish)
		warstw-a	warstw	warstew-k-a	layer
		sakw-a	sakw	sakiew-k-a	pannier
	RT#	rT	farb-a	farb	farb-k-a (!)
			kart-a	kart	kart-k-a (!)
			musztard-a	musztard	musztard-k-a (!)
		rv	larw-a	larw	larw-k-a (!)
			barw-a	barw	barw-n-y (!)
			ścierw-o	ścierw	ścierw-nik (!)
			bulw-a	bulw	bulw-k-a (!)
			salw-a	salw	salw-k-a (!)
		IT	walk-a	walk	walecz-n-y
			palt-o	palt	palet-k-o
			Kielc-e	Kielc	Kielc-k-i
		NT	słońc-e	słońc	słonecz-n-y
					słonecz-k-o
			legend-a	legend	legend-k-a (!)
			rumb-a	rumb	rumb-k-a (!)
	TT#	cht	placht-a	placht	plachet-k-a
			szlacht-a	szlacht	szlachet-n-y
					szlachet-k-a (arch)
	sT#	sp	wysp-a	wysp	wysep-k-a
	RR#	rm	form-a	form	forem-n-y
					forem-k-a
			firm-a	firm	firem-k-a
		m	urn-a	urn	uren-k-a
			tawern-a	tawern	taweren-k-a
					tawern-k-a (!)
	lm	palm-a	palm	palem-k-a	palm

I call pattern B the set of alternations where (monomorphemic) root-final CCs vocalize in Gpl (as well as in presence of a C/yer-initial suffix). Illustration is provided under (2) below.<sup>3</sup>

(2) pattern B: CC# vocalize in Gpl and also before C/yer-initial suffixes

		CøC-V	CeC#	CeC-C	gloss
		Nsg	Gpl	C/yer-initial suffix	
TR#	Tr	żebr-o	zeber	zeber-k-o	rib
		srebr-o	sreber	sreber-k-o	silver
				srebr-n-y (!)	
		piętr-o	pięter	pięter-k-o	floor
		futr-o	futer	futer-k-o	fur
		lustr-o	luster	luster-k-o	mirror
		jądr-o	jąder	jąder-k-o	nucleus
				jądr-n-y (!)	
		wiadr-o	wiader	wiader-k-o	pail
		iskr-a	iskier	iskier-k-a	sparkle
		chuchr-o	chucher	chucher-k-o	weakling
	TI	ciepl-o	ciepeł	ciepeł-k-o	warmth
				ciepl-n-y (!)	
		szabl-a	szabel	szabel-k-a	sword
		krzesł-o	krzesel	krzesel-k-o	chair
		kukł-a	kukieł	kukieł-k-a	puppet
		jagł-a	jagieł	(jagieł-k-i)	millet
		igł-a	igieł	igieł-k-a	needle
		cegł-a	cegieł	cegieł-k-a	brick
	Tn	krosn-o	krosien	krosien-k-o	loom
		wiosn-a	wiosen	wiosen-k-a	spring
		okn-o	okien	okien-k-o	window
		sukn-ia	sukien	sukien-k-a	dress
		sukn-o	sukien	(sukien-k-o)	cloth
		bagn-o	bagien	bagien-k-o	mud
	Tv	listw-a	listew	listew-k-a	lath
		mątw-a	mątew	(mątew-k-a)	cuttlefish
	RR#	durni-a	dureń	peret-k-a	fool
		perł-a	peret	peret-k-a	perl

While the list of pattern B items under (2) aims at exhaustivity (e.g. Laskowski 1975:29ff; Bethin 1992:146ff; Cyran 2003:176ff,188, 2005; Gussmann

<sup>3</sup> All Polish data in this chapter have been controlled and enriched by Gienek Cyran, to whom I am indebted.

2007:230ff), pattern A items under (1) are but a (representative) selection of what can be found in the Polish lexicon. Numerically, then, non-vocalisation in Gpl (pattern A) appears to be more common than vocalisation (pattern B).

Finally, table (2) shows that unlike for pattern A the sonority slope of the cluster is a relevant conditioning factor for pattern B: if any type of cluster can be of the non-vocalising type (pattern A), RT clusters appear to be unable to vocalize (pattern B). The same may be stated the other way round: only TR clusters seem to be able to vocalize in Gpl. This generalisation was made by Bethin (1992:149) for loanwords: "if the vowel does appear, it is more likely to appear within final sequences of rising sonority. Sequences which form optimal syllable codas of falling sonority [i.e. word-final RT and RR clusters in her examples] [...] generally do not acquire the alternation". Cyran (2005) (also Cyran 2003:176ff, 2010:160ff) makes the same observation.

In terms of the analysis to be developed below, this means that the distribution of alternating vowels in root-final clusters is not entirely arbitrary, i.e. lexical: only the rising sonority slope TR (and maybe RR: table (2) mentions two cases) allows for the acquisition of an alternating vowel. Whether a TR cluster acquires/possesses a yer, though, is entirely arbitrary (i.e. lexical). The question regarding the sonority-based condition is not further pursued below. Cyran (2005) (also Cyran 2003:176ff, 2010:160ff) offers an analysis in terms of his CSL (Complexity Scales and Licensing) model: (word-final) TRs are more difficult to license (by the final empty nucleus) than word-final RTs (this is also what the above quote from Bethin implies). Therefore 'easy' word-final clusters, i.e. RTs, are safe, while breakdown, i.e. vocalisation, is lurking for more fragile TRs.

## 2.2 Roots with free variation (and other types of variation)

A number of roots (with monomorphemic clusters) has both vocalized and unvocalized forms in free variation in Gpl, as shown under (3) below (e.g. Lasowski 1975:40; Bethin 1992:125; Gussmann 2007:230; Cyran 2005, 2010:170f).

- (3) roots with free variation in Gpl  
but vocalization is the ONLY option with C/yer-initial suffixes

	CøC-V	CC# / CeC#	CeC-C	gloss
	Nsg	Gpl	C/yer-initial suffix	
TR# Tr	wydr-a	wydr/wyder	wyder-k-a	otter
	biodr-o	bioder/biódr	bioder-k-o	hip
Tv	brzytw-a	brzytw/brzytew	brzytew-k-a	razor
	pochw-a	pochew/pochw	pochew-k-a	vagina/sheath
	poszw-a	poszew/poszw	poszew-k-a	duvet cover
	bitw-a	bitw/bitew	bitew-n-y	battle

	kotw-a	kotew/kotw		anchor
	tratw-a	tratew/tratw	tratew-k-a	raft
Tm	pasm-o	pasm/pasem	pasem-k-o	wisp
	karcz-m-a	karczem/karcz-m	karczem-n-y	inn
	ciżm-y	ciżm/ciżem	ciżem-k-a	shoes (arch.)
RT# IT	kalk-a	kalk/kalek	kalecz-k-a	carbon paper
	rT	kurw-a	kurew-k-a	whore
		torb-a	toreb-k-a	bag
RR# rn	sarn-a	sarn/saren	saren-k-a	roe deer
	żarn-a	żarn/żaren	żaren-k-a	quern
	ziarn-o	ziarn/ziaren	ziaren-k-o	grain
			ziarn-k-o (!)	
	mn gumn-o	gumien/gumn	gumien-k-o	barnyard
	wn grzyw-n-a	grzywien/grzyw-n	grzywien-k-a	fine
	łn wełn-a	wetel/wełn	(wełen-k-a)	wool

The trend to only allow for vocalized forms in TR clusters is also visible here, but less so than with non-variable vocalisation. And as before, vocalisation before C/yer-initial suffixes is systematic, i.e. the only option.

Diachronically, Polish is visibly following a movement from a stage where no clusters were vocalized at all in Gpl to a situation where more and more roots implement vocalized forms. One indicator is the numerical situation (non-vocalisation is the standard, only a limited number of roots is able to vocalize), another is the fact that for many roots the vocalized variant is substandard and stigmatized. This may be seen when looking at the recommendation of normative dictionaries like Szober's (1969, "dictionary of correct Polish"), who warns for instance against Gpl *cyfer* (Nsg *cyfr-a*). Finally, the interpretation of the sonority restrictions on vocalisation that was mentioned in the previous section also supports this diachronic scenario: vocalisation exists in Gpl because "difficult" clusters, i.e. word-final TRs, break down successively along a lexical diffusion perspective (while "easy" clusters are safe).

Vocalized forms are also substandard and/or dialectal in Nsg/Asg, which is the other zero case marker in Polish (apart from Gpl) (but which cannot be discussed in this chapter for lack of space). For example, *wiatr*, *Piotr*, *metr*, *filtr*, *bóbr* 'wind, Peter, meter, filter, beaver Nsg' are commonly encountered as *wiater*, *Pioter*, *meter*, *filter*, and this identifies uneducated speakers, or speakers of non-standard varieties. It is true, however, that the movement can also go in the other direction: vocalized *swetr* 'jumper Nsg' is standard, but unvocalized *swetr* is commonly heard.

In sum, the core of words has predictable (non-)vocalisation and does not show any variation: roots belong either to pattern A or B. A fair amount of words, though, show variation, either free or socially relevant, and the exact set

of words that belong to the three categories (A, B or variable) is a matter of inter-speaker variation.

### 2.3 Cases of non-vocalisation before C/yer-initial suffixes

Tables (4) and (7) below show cases of non-vocalisation of (monomorphemic) root-final CCs before C/yer-initial suffixes (see namely Laskowski 1975:39 and Bethin 1992:148). Their number is relatively small, but they need to be accounted for. This section proposes a *pretheoretical* analysis, arguing that these items are lexicalized, that is stored independently from the root (there is no derivational activity between the root and the items in question).

Table (4) provides those items whose Nsg ends in -CC-V, i.e. the category of words that this chapter is restricted to (because the only zero marker whose influence is considered is Gpl).

#### (4) CC# that do not vocalize before a C/yer-initial suffix (Nsg: -CC-V)

	Nsg	Gpl	C/yer-initial suffix	gloss	
A	tr	Piotr-a	Piotr	Piotr-k-a	first name
	rb	farb-a	farb	farb-k-a	paint
	rt	kart-a	kart	kart-k-a	card
	rd	musztard-a	musztard	musztard-k-a	mustard
	rw	larw-a	larw	larw-k-a	grub
		barw-a	barw	barw-n-y	colour
		ścierw-o	ścierw	ścierw-nik	corpse
	lw	bulw-a	bulw	bulw-k-a	root tuber
		salw-a	salw	salw-k-a	salvo
	rn	tawern-a	tawern	tawern-k-a	tavern
				<i>tawern-k-a (!)</i>	
	nd	legend-a	legend	legend-k-a	legend
	mb	rumb-a	rumb	rumb-k-a	rumba
B	TR	srebr-o	sreber	srebr-n-y	silver
				but: sreber-k-o	
		ciepl-o	ciepeł	ciepl-n-y	warmth
		światł-o	świateł	świetl-n-y	light
				białoskrzydł-n-y	white-winged
	jądr-o	jąder	jędr-n-y	firm, robust	

It may be seen that both patterns A and B deliver -CC-C-V forms, i.e. where the root-final cluster does not break up even before a C/yer-initial suffix. It also appears that the same root may derive vocalising as well as non-vocalising items: *srebr-n-y* vs. *sreber-k-o* (the same goes for *wietrz-n-y* vs. *wiater-ek* from table (7)). Hence whatever the lexical properties of the root, the vocalisation before

C/yer-initial suffixes does not depend on them. Rather, we are facing lexical idiosyncrasy of the derived words: either they are independent lexical recordings (and hence there is no synchronic derivation based on the root), or there are two separate roots (root allomorphy), one deriving vocalized, the other unvocalized items.

It is shown in Scheer (forth b) that unvocalized -CC-C-V items are the result of root-allomorphy, rather than of distinct lexical recordings of whole words (this piece could not be included in the present chapter for lack of space). There is a lexical item *srebr<sub>1</sub>-* that derives *srebr-o*, *sreber* and *sreber-k-o*, and there is another lexical item *srebr<sub>2</sub>-* that is used in *srebr-n-y*. *Srebr<sub>1</sub>-* and *srebr<sub>2</sub>-* are lexically distinct, and the difference lies in the relationship between the b and the r: while the two consonants form a branching onset in *srebr<sub>2</sub>-* and therefore make the r trapped upon the concatenation of -n-y, they are separated by a floating piece of meody (i.e. a yer) in *srebr<sub>1</sub>-*.

Note that the cluster -brn- of *srebr-n-y*, or CRC clusters more generally speaking, are well-formed and perfectly unspectacular in Polish: sonorants that occur between consonants (or in #RC and CR# position) are called trapped (as opposed to syllabic, see Scheer 2008, 2009). Examples are *trwać* 'to last', *brzmieć* 'to sound', *klnę* 'I curse', *plwocina* 'sputum', *krnąbrny* 'unruly', *brnąć* 'to wade' etc.

In any event, *srebr-n-y* is *not* synchronically derived from *srebr-o*. Of course it was derived from *srebr-o* at some point in the history of Polish, i.e. when the adjective *srebr-n-y* was first created – but the output of this primitive derivation was then stored in the lexicon and from that point in time on, derivationally speaking, had got nothing to do with the root *srebr-o* anymore.

Another interesting fact about unvocalized items before C/yer-initial suffixes is that the kind of variation described in the previous section is absent. For instance, *srebr-n-y* is the only possibility for this item: nobody says or can say \**sreber-n-y*. I could not come by any item parallel to the *wydr-a* – *wydr/wyder* pattern, i.e. where a -CC-C-V form would show free (or inter-speaker) variation with a -CeC-C-V form. For any given root-suffix combination, either one or the other are found – never both. This effect is predicted by the lexicalisation-based analysis: a lexical item (such as *srebr-n-*) that is unsubjected to derivational activity cannot vary – except in case there were two distinct "primitive" derivations, which have produced two different lexical entries.<sup>4</sup>

<sup>4</sup> The absence of variation concerns only -CCs that 1) are monomorphemic and 2) based on a vowel-final Nsg -CC-V. I have come across one item with free variation in a heteromorphemic cluster: *myd-el-niczka/myd-l-niczka* 'soap dish' (from *myd-l-o* 'soap', data from Laskowski 1975:39). Other items based on identical constructions/derivations have only a form with a stable vowel: *masel-niczka* 'butter dish' (from *mas-l-o* 'butter'), *piekiel-nic-a* 'witch' (from *piek-l-o* 'hell'). Unfortunately heteromorphemic clusters cannot be discussed in this chapter for lack of space. There is also one item I am aware of where a Nsg form in



For the sake of completeness, table (7) below gathers those items with unvoiced root-final -CCs before C/yer-initial suffixes whose Nsg is distinct from -CC-V: 1) -CC# (*wiatr* – *wietrz-n-y*), 2) -CC-ek# (*Piotr* – *Piotr-ek*), 3) -CC-C-V# (*jabl-k-o*). Before considering these data, note that Nsg -CC# clusters most often vocalize in presence of -ek (whose vowel is a yer), at least in native vocabulary. This is shown under (6).

## (5) Nsg -CC# → Nsg -CeC-ek

Nsg	Gpl	C/yer-initial suffix	gloss
wiatr		wiater-ek	wind
bóbr		bober-ek	beaver
filtr		filter-ek	filter
trefl		trefel-ek	club

## (6) Nsg -CC# → Nsg -CeC-ek

Nsg	Gpl	C/yer-initial suffix	gloss
wiatr		wiater-ek	wind
bóbr		bober-ek	beaver
filtr		filter-ek	filter
trefl		trefel-ek	club

## (7) CC# that do not vocalize before a C/yer-initial suffix (Nsg -CC# or -CC-eC# or -CC-C-V)

Nsg	Gpl	C/yer-initial suffix	gloss
a. no specificity			
wiatr		wietrz-n-y	wind
		but: wiater-ek	
pieśń		piosn-k-a	song
		piosen-k-a	
–		Jędr-ek	first name
Piotr		Piotr-ek	first name
alarm		alarm-ek <sup>5</sup>	alarm
uniform		uniform-ek	uniform
film		film-ek	film

-CC#, *pieśń* 'song', derives forms with free variation before a C/yer-initial cluster: *pi-osn-k-a/piosen-k-a* 'id., dim.'. The root-final cluster is most probably monomorphemic synchronically speaking, although it has a heteromorphemic origin (compare CS \*pěti > Cz pět, Po pieć 'to sing').

<sup>5</sup> *Alarm-ek*, *uniform-ek* and *film-ek* are quoted by Bethin (1992:148), but the native that I have consulted has a strong preference for forms in -ik: *alarm-ik*, *uniform-ik* and *film-ik*.

Park	parcz-ek	park	
b. s+C clusters			
myśl	myśl-nik	thought	
c. derived from adjectives			
mędr-ek	mędr-k-a	know-it-all	
	mędr-szy	wiser (compar.)	
podl-ec	podl-c-a	rascal	
d. isolated words			
	jabl-ek	jabl-k-o	apple
	bedł-ek	bedł-k-a	agaric (mushroom)
		Siedl-ce	city in Mazowia
		nabiodr-ek	cuisse
Sewr	sewr-ski	city in France	

The categorisation under (7) follows Laskowski (1975:39), who tries to identify specific patterns that lead to non-vocalisation before C/yer-initial suffixes. For instance, he singles out s+C as a vocalisation inhibitor, and this may sound plausible given the standard suspicion that s+C are solidary and some kind of contour segment (affricate) (e.g. Selkirk 1982:346ff; Carr 1993:212). This generalisation, however, does not fare well since we have already come across the pattern A item *wysp-a* – *wysp* – *wysep-k-a*. Finally, whether deadjectival derivation has any bearing on vocalisation remains to be seen: Laskowski provides only two items.

All in all, what table (7) shows is again lexical idiosyncrasy: which words do not vocalize before a C/yer-initial suffix cannot be predicted from any of their properties.

### 3. The regular yer-analysis fails: the difference between A- and B-roots must be lexical

#### 3.1 Insertion disqualified in the standard analysis

Pattern B is regular in the realm of Slavic vowel-zero alternations: vowels that alternate with zero are underlying yers, which vocalize in presence of a following yer, and otherwise remain mute: *wiadro* is /wiadEro/, and the yer E appears on the surface in Gpl /wiadEr-O/ → *wiader* (Gpl is a yer itself) as well as before a yer-initial suffix /wiadEr-Ek-o/ → *wiader-k-o*.

This analysis of Slavic vowel-zero alternations is the insight encoded in the Lower rule that was introduced by Lightner (1965) and adapted to various linear and autosegmental frameworks in the following decades (Gussmann 1980;

Rubach 1984, 1986; Gussmann & Kaye 1993; Scheer 2005, see the overviews in Cyran 2005 and Scheer 2011). For the time being it does not matter which implementation of Lower is used: the only thing that matters is that *all* versions share the basic rule according to which vowels that alternate with zero are underlying yers.

In other words, the standard analysis is based on deletion, rather than on insertion. The two perspectives were debated at length in the literature: are alternating vowels underlyingly absent and inserted, or present and deleted? Insertion-based analyses have been proposed by, among others, Laskowski (1975), Czaykowska-Higgins (1988) and Piotrowski (1992). They are convincingly refuted by Gussmann (1980:26ff), Rubach (1984:28f, 1993:134ff) and Szypra (1992a:280ff, 1995:94ff).

Among the arguments in favor of deletion, the following are decisive. In languages such as Russian where more than one vowel alternates with zero, it cannot be predicted which vowel will appear in which morpheme. That is, the presence of an alternating *e* in *d'én' – dn'-á* 'day Nsg, Gsg', against alternating *o* in *són – sn-á* 'dream Nsg, Gsg', is a lexical property of the root. An insertion-based analysis would not know which vowel to epenthesize into which root.<sup>6</sup> The second reason is that there is no context for insertion. The motor for insertion is held to be the avoidance of heavy clusters (or unsyllabifiable/extrasyllabic consonants): the Gpl of Russian *lask-a* 'weasel Nsg' and *bobr-á* 'beaver fur Nsg' is *lások* and *bob'ór*, respectively; in the insertion perspective, Gpl forms undergo epenthesis in order to avoid final *-sk#* and *-br#* clusters. This cannot be the reason, though, since Russian happily tolerates these clusters in *lask-a – lask* 'caress Nsg, Gpl' and *bobr-á – bóbr* 'beaver Gsg, Nsg'. The same situation is found in other Slavic languages. Cyran (2005) makes the argument based on Polish data: compare *swetr-a – sweter* 'jumper Gsg, Nsg', *lalk-a – lalek* 'doll Nsg, Gpl', *durni-a – dureń* 'fool Gsg, Nsg' (where clusters vocalize) with *wiatr-u – wiatr* 'wind Gsg, Nsg', *walk-a – walek* 'fight Nsg, Gpl', *cierni-a – cierní* 'thorn Gsg, Nsg' (where the same clusters do not vocalize).

### 3.2 Pattern A misbehaves: Lower predicts that it cannot exist

While pattern B is a regular instantiation of Lower, pattern A cannot be analysed with the standard tool. As was mentioned, a critical ingredient of the theory is that *all* vowels that alternate with zero are underlying yers. Since there is a vowel-zero alternation *cyfr-a – cyfer-k-a* 'number Nsg, dim.', there must be a yer separating the two root-final consonants: the root identifies as /cyfEr/. We also

<sup>6</sup> It is also not the case that the quality of the alternating vowel may be predicted from the palatal vs. non-palatal character of the preceding consonant. In *bobr-á – bob'ór* 'beaver fur Nsg, Gpl' for example, an *o*, not an *e*, appears after a palatalized labial. This issue is further discussed in Scheer (forth a), also in the light of e-o alternations.

know that the Gpl is a yer, since it provokes the vocalisation of the preceding yer in pattern B: *wiadro* /wiadEr-o/ – *wiader* /wiadEr-O/. Since the inflection marker is of course the same in both patterns, we are expecting a vocalized yer in Gpl /cyfEr-O/ – but alas the Gpl is unvocalized: *cyfr*.

In short, the standard theory predicts that pattern A cannot exist: once we are sure that the root has a yer, this yer must surface before the Gpl-yer. A logical way out of this dilemma would be to assume an allomorphy for Gpl: the case marker would be a yer when attached to B-roots, but literally nothing when combining with A-roots. Hence Gpl /cyfEr<sub>A</sub>-ø/ surfaces as *cyfr* following the regular application of Lower because the trigger for vocalisation is absent. On the other hand, the suffixal yer in /wiadEr<sub>B</sub>-O/ produces the regular vocalisation of the root yer since this time the Gpl marker is a yer.

This solution based on Gpl-allomorphy is outlandish and can be safely disregarded. It supposes roots that select a particular affix, which means that the head is selected by its complement. Allomorphy, however, works the other way round: heads (affixes) select complements (roots), if anything.

### 3.3 The difference between A- and B-roots must be lexical

It thus needs to be understood what the difference is between A- and B-roots. Obviously, this difference is *lexical*: membership in either class is arbitrary (with the reserve discussed regarding the sonority slope of the root-final cluster), there are doublets, and there is substantial dialectal, register-related and inter-speaker variation. The distinction thus bears all signs of variable lexical representations.

The puzzle, then, is to identify the lexical property that opposes vocalizing and non-vocalizing roots in a system where the Gpl marker is invariable. We know that there is no solution in the frame of the standard Lower-based analysis, but of course the basic insight of Lower cannot be called into question given the massive amount of evidence that it is able to account for. What we need, then, is a scenario where the computation is uniform (Lower) for both A- and B-roots, but based on distinct lexical representations. This is an invitation to reconsider the Slavic mantra according to which *all* alternating vowels are underlying yers.

## 4. The easy way out: root allomorphy (Gussmann 2007)

### 4.1 Gussmann (2007): Standard GP

Instead of trying to isolate the lexical specificity of A-roots, the Gpl puzzle may also be taken to be the witness of irregularity, which leaves no hope for a management under the roof of the Lower rule. The alternative, then, is root allomorphy. This is the option favoured by Gussmann (2007), whose general project

is to reduce computational activity to a strict minimum: the labour is outsourced to morphology and an increased number of lexical recordings (allomorphy) on the one hand, and to the phonology-phonetics mapping on the other (see Scheer 2010). In this context, the situation in the Gpl leads Gussmann to doubt that vowel-zero alternations are managed by phonology at all, despite their regularity elsewhere (in Polish and Slavic).

Gussmann (2007:230) hints at a non-phonological reason for the failure of A-roots to vocalize in word-final position: 'when the nouns become the input to further derivations, the floating vowel normally appears in them'. In other words, Gpl *cyfr* remains unvocalized because it is underived, while *cyfer-k-a* is vocalized because the root was subject to derivational activity.

Gussmann (2007:233) concludes that there are two distinct lexical recordings for every A-root: one where the root-final cluster is separated by a nucleus (which contains a floating piece of melody), and one where the root-final cluster is a branching onset. This is shown under (8) below.

(8) root allomorphy: a non-phonological solution

a. allomorph 1: yer present

O N O N O N  
| | | |  
c y f e r

b. allomorph 2: yer absent

O N O N  
| | ^  
c y f r

Gussmann's analysis is couched in (Standard) Government Phonology where word-final consonants are onsets of an empty nucleus (Kaye 1990). He also adopts Scheer's (2004:§76, 2005) representation of yers as floating pieces of melody: stable vowels are lexically associated as under (9) (*bies* - *bies-a* 'devil Nsg, Gsg'), while alternating vowels (yers) are lexically floating pieces of melody as under (9) (*pies* - *ps-a* 'dog Nsg, Gsg').

(9) stable vs. alternating vowels

a. lexically associated vowel:  
stable

O N O N  
| | |  
b<sup>j</sup> e s

b. floating piece of melody: alternating  
vowel (yer)

O N O N  
| | |  
p<sup>j</sup> e s

On this analysis, the association of floating pieces of melody (yers) to their syllabic constituent, i.e. their pronunciation, depends on the status of the following nucleus. In case this nucleus is filled (i.e. associated to some melody), it can govern its lefthand neighbour (as under (10)); if on the other hand the following nucleus is unpronounced itself (either because it is empty as under (10) or because it is itself governed as under (10)), it cannot dispense government. Gov-

ernment thus acts as an association-inhibitor: floating pieces of melody associate iff their nucleus is ungoverned. Association is blocked in case it is governed.

(10) government-based Lower

a. *wiadr-o*: yer governed

Gvt  
┌───┐  
O N O N O N  
| | | | |  
w<sup>j</sup> a d e r o

b. Gpl *wiader*: yer ungoverned

Gvt  
┌───┐  
O N O N O N  
| | | | |  
w<sup>j</sup> a d e r

c. yer-initial suffix: yer ungoverned

Gvt Gvt  
┌───┐ ┌───┐  
O N O N O N O N  
| | | | | | |  
w<sup>j</sup> a d e r e k o

The fact that Lower describes a lateral relation between two nuclei  $ON_1ON_2$  whereby the status of  $N_2$  (presence/absence of a yer) determines the vocalisation of  $N_1$  is obvious (this is explained at greater length in Scheer 2005, 2011 and Scheer & Ziková 2010). The analysis in Government Phonology merely gives a name to this lateral relation: government.

In this environment, Gussmann's allomorphy-based analysis works like this: underived forms of A-roots select for the allomorph (8) where the root-final cluster is a branching onset. Therefore, in Gpl the root does not contain any yer that could surface, and the result is /cyfr-/ → *cyfr*. In derived forms on the other hand, the allomorph (8) is selected, and yer-initial suffixes trigger regular yer-vocalisation along the lines of (10).

The variation that is related to the vocalisation of root-final clusters, then, is the result of distinct lexical representations: doublets (such as *wydra* - *wydr/wyder* 'otter Nsg, Gpl') are due to the competition of two independent lexical representations, one along the lines of A-roots with the allomorphy described, the other along B-roots where no allomorphy occurs and the only lexical form is the one under (8), i.e. the one that bears a yer.

4.2 Cyran (2005): CVCV, and only epenthesis

Let us now consider Cyran's (2005) analysis, which goes down the same road as Gussmann's, albeit not explicitly since forms with C/yer-final suffixes are not

examined. Unlike Gussmann, Cyran evolves in CVCV (Lowenstamm 1996; Scheer 2004; Cyran 2010) where no branching constituents are left. His model allows for lateral relations among adjacent consonants in both directions. Following the classical take of Standard GP, he holds that in a cluster it is always the obstruent that dominates the sonorant. Hence TR clusters contract a progressive relationship  $T \rightarrow R$ , while RT (and RR, TT) identify as  $R \leftarrow T$ . The empty nucleus enclosed in such a domain is circumscribed and may remain empty without being governed. That is,  $T\emptyset_1R\emptyset_2\#$  and  $R\emptyset_1T\emptyset_2\#$  are well-formed (in Polish where final empty nuclei can license both RT and TR clusters) despite the presence of two empty nuclei in a row because  $\emptyset_1$  is taken care of by the consonantal domain (while  $\emptyset_2$  is licensed by virtue of being domain-final).

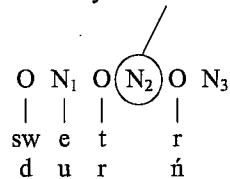
Cyran (2005) only considers the (non-)vocalization of word-final CC# clusters, i.e. before the two zero case markers Nsg (Asg) and Gpl. He is not concerned with the situation of the same clusters in presence of C/yer-initial suffixes. On the basis of this empirical situation, he argues for a purely epenthetic analysis of vowel-zero alternations: all alternating vowels that appear in word-final clusters in Polish are epenthetic. Their insertion repairs an ill-formed structure that accommodates two empty nuclei in a row: the leftmost receives a vowel (note that in his system final empty nuclei are unable to govern). This is shown under (11)a below.

(11) vowel-zero alternations according to Cyran (2005)

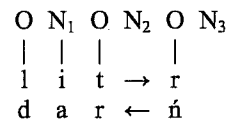
a. pattern B: CC# broken up

b. pattern A: stable CC#

lexically marked as unlockable



result: *sweter* 'jumper Nsg',  
*dureń* 'fool Nsg'



result: *litr* 'liter Nsg', *darn* 'sod Nsg'

The reason why epenthesis occurs under (11), but not under (11), is the lexical marking of N<sub>2</sub> under (11) as 'unlockable'. Interconsonantal relations are contracted whenever they can; their presence under (11) is thus the regular case. As was mentioned, their existence makes the structure well-formed since they 'lock' one of the two empty nuclei in a row. The presence of an alternating vowel under (11) shows that in these roots there is no consonantal domain. Since the difference between *sweter* and *litr* is an idiosyncratic property of each root, it

must be marked in the lexicon. Something thus prevents the consonants under (11) to establish a domain, and Cyran holds N<sub>2</sub> responsible: this nucleus is lexically marked as unlockable. Therefore there are two phonologically active empty nuclei in a row under (11), and the leftmost is subject to epenthesis (the rightmost is licensed because it is domain-final). In sum, then, items with unlockable N<sub>2</sub>'s belong to pattern B, while roots with regular N<sub>2</sub>'s instantiate pattern A.

Like in Gussmann & Kaye (1993), one of the two obstacles for epenthetic analyses that was mentioned in section 3.1 is overcome, but the other is not addressed. The context for insertion, which is not predictable from the surface, is lexically specified, here in terms of unlockable empty nuclei. That the analysis is not exportable to (Slavic) languages where more than one vowel alternates (like Russian) is not discussed.

Now recall that forms with C/yer-initial suffixes are not in the empirical scope of Cyran (2005). What would be their analysis in terms of his apparatus? The solution is necessarily based on some kind of allomorphy: the regular pattern B under (11) where N<sub>2</sub> vocalizes before C/yer-initial suffixes (*wiadr-o* – *wiader* – *wiader-k-o*) is without problems (like for the regular yer-based analysis), but pattern A cannot be derived by purely phonological means. In order to avoid vocalization in Gpl *cyfr*, there must be a consonantal domain as under (11), which however will prevent vocalization in \**cyfr-k-a*. Like in Gussmann's analysis, A-items with C/yer-initial suffixes must thus be derived from a lexical form with an unlockable N<sub>2</sub> as under (11) (*cyfer-k-a*), while Gpl necessarily instantiates a form with a locked N<sub>2</sub> as under (11).

## 5. A purely phonological solution for the A-B variation

5.1 Bethin's (1992) epenthesis is on the right track, but its scope is both too narrow and too wide

The point made in this chapter is that there is no need to recur to an extra-phonological solution for the Polish pattern A-B variation. A purely phonological management is possible if the invitation that was mentioned at the end of section 3 is followed: it is not true that all vowels which alternate with zero are underlying yers. Note that this statement is entirely independent of any particular implementation of Lower: it will work with linear as much as with autosegmental systems, and within the latter does not make any selection either. Hence, the analysis below is done in strict CV, but the same point could be made in any other framework that can express a representational difference between the presence and the absence of yers.

Recall that Bethin has already made this point in her 1992 book: "[v]owel-zero alternations in Polish are not attributable to a unique underlying representa-

tion", Bethin 1992:153). This section takes a closer look at how exactly her dual system works, i.e. which cases of vowel-zero alternations are exactly the result of yer vocalisation, and which ones are due to epenthesis. Also, the motivation of epenthesis is examined.

The major line of division that Bethin draws is between native vocabulary and loanwords. She motivates epenthesis on the grounds of vowel-zero alternations in prepositions where Lower plays no role (*w domu* 'in the house', *z doktorem* 'with the doctor' vs. *we wtorek* 'on Tuesday', *ze stolem* 'with the table') and then argues that since grammar allows for both yers and epenthesis, there is ambiguity for speakers when they come across a vowel-zero alternation: "[t]his ambiguity is most likely to arise in cases where the evidence for a [-cons] node [her representation of yers] in the conditioning environment is questionable, as for example, might be the case in the inflectional paradigm for the nom sg and gen pl desinences. More information is needed to study how Polish speakers interpret the vowel-zero alternation" (Bethin 1992:146).

She leaves it at that for alternations that occur in inflectional paradigms, and turns to the study of loanwords. The vowel-zero alternations that are found here, she argues, are the result of epenthesis. While examining the data, though, she notes that native vocabulary and loans behave alike: "[m]any borrowings do not exhibit the vowel-zero alternation at all" (Bethin 1992:148), and on the same page "[s]ince there are so many native nouns that do not exhibit the alternation as expected, or exhibit it partly [...], the absence of the vowel-zero alternation in borrowed forms is not particularly significant."

This is certainly true: we have already seen in section 2.1 that loanwords and native vocabulary are equally represented in pattern A and B, produce items that are subject to Gpl variation and derive forms that refuse to vocalize before a C/yer-initial suffix. Hence there is no reason to believe that loans and native items are any different regarding the workings of vowel-zero alternations. Bethin's distribution of the two managements whereby epenthesis occurs in loans and yers are found elsewhere, then, cannot be correct. If it is true that there is an epenthetic management besides yers, the question thus arises how we know which particular vowel-zero alternation is driven by which mechanism. It is argued below 1) that there is epenthesis, 2) that its motivation is not the one that Bethin uses (rather than being motivated by alternations in prepositions, it is enforced by the configuration created by C/yer-initial suffixes), and that 3) the distribution of the two mechanisms is along the opposition between pattern A (epenthesis) and pattern B (yers).

On this view, the scope of epenthesis in Bethin (1992) is too narrow: not only loans are concerned. But it is also too wide: not all vowel-zero alternations that occur in loans are the result of epenthesis. Loans such as *perl-a* – *perel* – *perel-k-a* 'perl Nsg, Gsg, adj.' also occur in pattern B, which faithfully instantiate regular yer-based alternations.

Another difference between the analysis that is developed below and Bethin's is the motivation for epenthesis, and this is due to the different perspectives that are induced by different representational environments: while Bethin follows the line of attack of all epenthesis-based analyses whereby a structure needs to be repaired because of a problem encountered with a consonant (which is unsyllabifiable or extrasyllabic), the need for repair in the analysis below comes from the illegal situation of a nucleus (which cannot remain empty in presence of another internal empty nucleus to its right).

As we will see, a side-effect of the nucleus-focussed and hence government-based analysis is the correct prediction of exactly the nucleus that is subject to epenthesis: in  $C\emptyset_1C\emptyset_2CV$ , V governs  $\emptyset_2$  but  $\emptyset_1$  remains orphan. It is thus  $\emptyset_1$ , not  $\emptyset_2$ , that is subject to epenthesis: /*cyfɔr-ɔk-a*/ → *cyferka* (not \**cyfreka*). Bethin (1992:152) wonders why in the  $C_1C_2C_3$  cluster that is in need of repair epenthesis only ever occurs between  $C_1$  and  $C_2$ , rather than between  $C_2$  and  $C_3$ . Her answer is cyclicity: *cyferka* is cyclically derived from [[[cyfr]k]a], and epenthesis breaks up fr because this is the first cluster that is encountered by the derivation on the innermost cycle. This cannot be the reason, though, since Gpl *cyfr* then should also be subject to epenthesis: it is made only of the innermost cycle. The basic equation is precisely that the *same* item, [cyfr], behaves differently according to whether it occurs alone (result: *cyfr*) or is followed by a C/yer-initial suffix (result: *cyfer-k-a*). Cyclic derivation is unable to discriminate between the two options because there is no look-ahead: when the innermost cycle of [[[cyfr]k]a] is computed, the computation sees nothing else than [cyfr] – it does not know that there will be a C/yer-initial suffix on a later cycle.

In sum, I argue that Bethin (1992) was on the right track for the solution of the puzzle by proposing two distinct mechanisms for vowel-zero alternations (yers and epenthesis), but that the scope of both was ill-defined. Also, the shift from a consonant- to a nucleus-focussed view, i.e. to the lateral government-based interpretation of Lower, allows for a better understanding of the precise locus of epenthesis.

Finally, another article to be mentioned in this context is Worth (1968), which examines vowel-zero alternations in Russian in a structuralist environment. The author identifies exactly what we have seen in Polish: pattern A vs. pattern B according to whether -CCs vocalize in word-final position, but systematic vocalisation before C/yer-initial suffixes. The solution of the Polish puzzle that is presented below follows Worth's analysis: vowels that appear before C/yer-initial suffixes in pattern A roots are epenthetic, while they are due to yer vocalisation in pattern B roots.

### 5.2 The key to the puzzle: internal vs. final

Before we proceed with the analysis, a pre-theoretical ingredient is missing. Recall from section 4 that Gussmann ascribes the contrasting behaviour of Gpl *cyfr* 'number' (no vocalisation) and *cyfer-k-a* 'id., dim.' (vocalisation) to the fact that the latter is derived, while the former is not. This is the non-phonological way of looking at things. But there is also a phonological way to capture the contrast. Regularity (or uniformity) is encountered when the root-final CC is word-internal: vocalisation is regular (pattern A *cyfer-k-a* like pattern B *wiader-k-o*, except for the lexicalized cases discussed in section 2.3) and unsubjected to variation even for those items that show variation in Gpl (*wydra* produces both Gpl *wydr* and Gpl *wyder*, but only *wider-k-a*). By contrast, irregularity and variation are observed when the cluster is word-final: here vocalisation is a matter of lexical marking (Gpl *cyfr* vs. Gpl *wiader*), and this is also where the locus of variation is found (*wydra* - *wydr* / *wyder*).

Again Bethin (1992:152) has understood that this is the critical contrast: regarding *forem-n-y* (*form-a* - *form*), she writes that "the sonorant is no longer at the end of the word, and a vowel is epenthesized to facilitate syllabification. Although word medial adjunction seems to be characteristic of a few items in Polish such as *piosnka* 'song', *srebrny* 'silver', *cieplny* 'thermal', the usual adjustment of borrowings into Polish phonology seems to call for vowel epenthesis." In Bethin's analysis, unsyllabifiable root-final consonants such as the *m* in Gpl *form* or the *r* in Nsg *filtr* are saved by being directly adjoined to a higher prosodic constituent (the phonological word):  $[[[for]_{\sigma}m]_{\omega}]$  and  $[[[filt]_{\sigma}r]_{\omega}]$  are well-formed since all consonants are integrated into prosodic structure. Lexical marking then discriminates between items that favour this solution (which is the regular way to go) and those where adjunction to the phonological word is blocked: this is the case of Nsg *cyngiel* 'trigger' (< German *Züngel*), where the word-final consonant remains unsyllabifiable and therefore can only be saved by epenthesis (Bethin 1992:150, see Cyran 2005 for an overview of this analytic strand).

If it is understood that whether a word is a loan or not plays no role, Bethin's analysis thus opposes pattern A and pattern B by the idiosyncratic ability of a root to have its last consonant adjoined to the phonological word (pattern A: yes, pattern B: no). Crucially, though, this opposition is only workable word-finally since adjunction to the phonological word is a form of extraprosodicity (Rubach & Booij 1990), and the Peripherality Condition restricts extra-Xity (extrametricality, extrasyllabicity, extraprosodicity) to word edges (e.g. Hayes 1995:57f; Clements 1990:290). This is what Bethin means when she says that adjunction to the phonological word is impossible word-medially. She then runs into 'a few items' (*piosnka* etc.) where (in violation of the Peripherality Condition) Rubach & Booij (1990) (also Rubach 1997) argue for an extrasyllabic analysis of the middle consonant of the CRC cluster. Her way out is again to call on the native

vs. loan contrast (word-internal adjunction to the phonological word is possible in words belonging to the former, but not in the latter set), but this does not work: Bethin (1992:148) herself provides a list of loans that refuse to vocalize before C/yer-initial suffixes (e.g. *barw-a* - *barw* - *barw-n-y* 'colour Nsg, Gpl, adj.' (< German *Farbe*)).

I believe that as before, Bethin's general direction was correct: the contrast between the word-internal and word-final locus is responsible for the distribution of variability (final) and stability (internal). The representational environment in which she evolved, though, did not allow her to bring home this intuition.

In Government Phonology (Standard and strict CV alike), the internal-final contrast translates into the difference between internal and final empty nuclei: the nucleus that decides on the vocalisation of the putative yer in the preceding cluster under (10) is word- (or domain-) final, but word- (or domain-) internal under (10). It is well known that the right edge of words allows for more clustering than what can be found word-internally (e.g. Broselow 2003). This and other specific properties of the right edge have been translated into Government Phonology as a difference in the lateral actorship of final empty nuclei (FEN), as opposed to internal empty nuclei. In short, FEN can do more than their internal peers, i.e. they may be able to license and govern where internal empty nuclei are unable to dispense lateral forces (e.g. Charette 1990, 1992; Scheer 2004:§524; Cyran 2010).

### 5.3 A government-based analysis that predicts the locus of epenthesis

With this in mind we can finally introduce the purely phonological analysis of the A-B variation that the reader is waiting for: the basic insight is the lexical contrast that appears under (12) below.

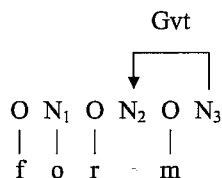
#### (12) lexical contrast between A- and B-roots

a. A-root: yer absent	b. B-root: yer present
O N O N O N	O N O N O N
f o r m	w <sup>j</sup> a d e r

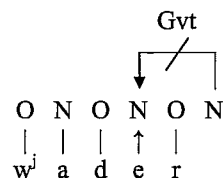
The Gpl marker is literally nothing (this is the regular analysis in government-based Lower, see Scheer 2005, 2011). Hence in Gpl the representations under (12) are interpreted as such. This is shown under (13) below.

## (13) derivation of Gpl

a. A-root: form  
FEN governs empty nucleus



b. B-root: *wiader*  
FEN is unable to govern contentful nucleus



FEN are able to govern empty nuclei (as under (13)), but not nuclei that have a lexical content (i.e. a floating piece of melody, as under (13)).<sup>7</sup> Therefore the yer of B-roots surfaces: its nucleus is ungoverned ((13)). Under (13), however, nothing can surface in A-roots since they lack floating pieces of melody (yers). This is also the reason why the FEN is able to govern the preceding nucleus: it is empty.

In Nsg forms, the vowel in the final nucleus will always govern the preceding nucleus, irrespectively of whether it is empty or not. The result are non-vocalized forms with both A- and B-roots (*form-a*, *wiadr-o*).

C/yer-initial suffixes identify as under (14) below: they begin with a floating piece of melody (the yer), which after concatenation ends up in the root-final empty nucleus.

## (14) lexical representation of C/yer-initial suffixes

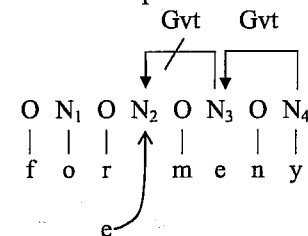
O	N	
e	n	adjectival <i>-n-y</i>
e	k	diminutive <i>-ek</i> , <i>-k-a</i> , <i>-k-o</i>
		etc.

The derivation of forms with C/yer-initial suffixes, then, is as under (15).

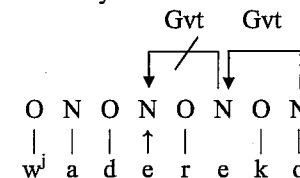
<sup>7</sup> Were FEN able to also govern contentful nuclei, /p<sup>h</sup>Esø/ would not surface as *pies*, but as \**ps* (Scheer 2004:§541).

## (15) derivation of forms with C/yer-initial suffixes

a. A-roots: epenthesis



b. B-roots: yer vocalisation



The result is the same on the surface (presence of a vowel) with both A- and B-roots, but the vowel that appears has different origins: it is the regular yer that is lexically present in B-roots, but results from epenthesis in A-roots. In the latter, indeed, the concatenation of the suffix and the application of government produces a configuration that features two unpronounced empty nuclei in a row (N<sub>2</sub> and N<sub>3</sub>). Such a structure is ill-formed. N<sub>3</sub> is governed and therefore cannot surface; but being itself unpronounced it cannot govern N<sub>2</sub>, which therefore remains orphan. In this situation, the structure is repaired by an epenthesis that fills in the orphan nucleus N<sub>2</sub>, i.e. the one that is not governed (N<sub>3</sub> has no demands since it is governed).<sup>8</sup>

Note that N<sub>3</sub>, the root-final nucleus, is word-internal under (15), but word-final under (13) when the Gpl is derived. It is empty in both cases, and this emptiness makes it unable to govern in internal position: this is why epenthesis is triggered. It is a good governor, though, when occurring in word- (or domain-) final position as under (13): as was mentioned in the previous section, this is how Government Phonology expresses the opposition between the internal and the final locus: FEN "can do more" than their internal peers.

On this analysis, the variation observed is strictly lexical: doublets are produced when speakers have lexicalized both A- and B-forms for the same root. The evolution being in the sense from A- to B-roots, high style is conservative, and B-forms may be socially stigmatized.

## 6. Conclusion

Following Bethin (1992), I have argued that Polish Gpl alternations compel us to abandon a fundamental ingredient of the regular analysis of Slavic vowel-zero

<sup>8</sup> A ground rule in Government Phonology is indeed that empty nuclei need to be governed: ungoverned empty nuclei make a representation ill-formed. The detail is a little more complicated (and also depends on the brand of GP that one subscribes to), but does not matter here. Relevant literature includes Kaye (1992); Gussmann & Kaye (1993); Cyran (2010:116ff) provides a survey.

alternations: it is not the case that all alternating vowels are underlying yers. Some are, but others are not, and only analysis will tell who is who. In the Polish case, the critical diagnostic is the behaviour of (monomorphemic) stem-final clusters in Gpl: vowels that appear before C/yer-initial suffixes in pattern A roots are epenthetic (*cyfer-k-a*, cluster unvocalized in Gpl: *cyfr-a – cyfr*) while they represent vocalized yers in pattern B roots (*wiader-k-o*, cluster vocalized in Gpl: *wiadr-o – wiader*).

If Polish has alternating vowels that are the result of epenthesis, other Slavic languages may have epenthetic vowels as well. Czech for instance is not among them: in this language vocalisation in Gpl is absolutely regular (*form-a – forem – ne-forem-n-ý* 'form Nsg, Gpl, adj.' and so on). In Polish terms, Czech is a language where all roots are of the B-type; and hence where all vowel-zero alternations represent yer vocalisation. Diachronically speaking, then, it may be the case that Polish is on the way to become like Czech (the movement is from A- to B-roots).

Finally, the analysis of Polish makes a prediction regarding Slavic languages where more than one vowel alternates with zero (e.g. of the Eastern family): in case they feature the Polish pattern and thus have epenthetic vowels, there must be a way to predict which vowel (e or o in Russian for example) will be inserted. Either it is always the same vowel, i.e. e or o, or the quality of the vowel must be able to be predicted from the consonantal environment. Russian happens to instantiate the Polish pattern (Worth 1968), and the prediction may thus be tested: those alternating vowels which appear in presence of yer-initial suffixes, but not in Gpl (e.g. *igr-á – igr – igór-k-a* 'game Nsg, Gpl, dim. '), must not be able to sustain the lexical contrast between e and o.

## References

Items followed by the mention WEB are available at [www.unice.fr/scheer](http://www.unice.fr/scheer).

- Bajerowa, Irena. 1953. Z historii polskich grup spółgłoskowych. *Biuletyn Polskiego Towarzystwa Językoznawczego* 12, 193–203.
- Baker, Mark C. 2008. The macroparameter in a microparametric world. In T. Biberauer (ed.), *The Limits of Syntactic Variation*. Amsterdam: Benjamins. 351–373.
- Bethin, Christina. 1992. *Polish Syllables: the Role of Prosody in Phonology and Morphology*. Columbus, OH: Slavica.
- Biberauer, Theresa. 2008. Introduction. In T. Biberauer (ed.), *The Limits of Syntactic Variation*. Amsterdam: Benjamins. 1–72.
- Broselow, Ellen. 2003. Marginal phonology: phonotactics at the edge. *The Linguistic Review* 20, 159–193.
- Carr, Philip. 1993. *Phonology*. London: Macmillan. WEB.
- Charette, Monik. 1990. Licence to govern. *Phonology* 7, 233–253. WEB.
- Charette, Monik. 1992. Mongolian and Polish meet Government Licensing. *SOAS Working Papers in Linguistics and Phonetics* 2, 275–291.
- Clements, George. 1990. The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckmann (eds.), *Papers in Laboratory Phonology I*. Cambridge: CUP. 283–333.
- Cyran, Eugeniusz. 2003. *Complexity Scales and Licensing Strength in Phonology*. Lublin: KUL.
- Cyran, Eugeniusz. 2005. Sound patterns of Polish: phonotactic paradoxes at the right edge of words. *Studies in Polish Linguistics* 2, 61–89.
- Cyran, Eugeniusz. 2010. *Complexity scales and licensing in phonology*. Berlin: Mouton de Gruyter.
- Czaykowska-Higgins, Ewa. 1988. *Investigations into Polish morphology and phonology*. Cambridge, MA: The MIT Press.
- Gussmann, Edmund. 1980. *Studies in Abstract Phonology*. Cambridge, MA: The MIT Press.
- Gussmann, Edmund. 2007. *The Phonology of Polish*. Oxford: OUP.
- Gussmann, Edmund & Jonathan Kaye. 1993. Polish notes from a Dubrovnik Café: I. The yers. *SOAS Working Papers in Linguistics and Phonetics* 3, 427–462.
- Hayes, Bruce. 1995. *Metrical Stress Theory. Principles and Case Studies*. Chicago, London: University of Chicago Press.
- Kaye, Jonathan. 1990. 'Coda' licensing. *Phonology* 7, 301–330. WEB.
- Kaye, Jonathan. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics* 2, 293–313. WEB.
- Laskowski, Roman. 1975. *Studia nad morfonologią współczesnego języka polskiego*. Wrocław, Warszawa, Kraków, Gdańsk: Wydawnictwo Polskiej Akademii Nauk.
- Lightner, Theodore. 1965. *Segmental Phonology of Contemporary Standard Russian*. Doctoral dissertation, MIT.
- Lowenstamm, Jean. 1996. CV as the only syllable type. In J. Durand & B. Laks (eds.), *Current trends in Phonology. Models and Methods*. Salford, Manchester: ESRI, 419–441. WEB.
- Piotrowski, Marek. 1992. Polish yers in non-linear phonology. In U. Dressler, H. Luschützky, O. Pfeiffer & J. Rennison (eds.), *Phonologica 1988*. Cambridge: CUP. 215–227. WEB.
- Roberts, Ian & Anders Holmberg. 2010. 'Introduction: Parameters in minimalist theory. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: CUP. 1–57.
- Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: The Structure of Polish*. Dordrecht: Foris.
- Rubach, Jerzy. 1986. Abstract vowels in three dimensional phonology: the yers. *The Linguistic Review* 5, 247–280. WEB.
- Rubach, Jerzy. 1993. *The Lexical Phonology of Slovak*. Oxford: Clarendon.
- Rubach, Jerzy. 1997. Extrasyllabic consonants in Polish: Derivational Optimality Theory. In I. Roca (ed.), *Derivations and Constraints in Phonology*. Oxford: Clarendon. 551–581.
- Rubach, Jerzy & Geert Booij. 1990. Edge of constituent effects in Polish. *Natural Language and Linguistic Theory* 8, 427–463.
- Scheer, Tobias. 2004. *A Lateral Theory of Phonology. Vol.1: What is CVCV, and why should it be?* Berlin: Mouton de Gruyter.
- Scheer, Tobias. 2005. Slavic Vowel-Zero Alternations and Government Phonology: Two Approaches, One Solution. In S. Franks, F. Gladney & M. Tasseva-Kurkatchieva (eds.),



- Formal Approaches to Slavic Linguistics 13. The South Carolina Meeting.* Ann Arbor: Michigan Slavic Publications. 300–311. WEB.
- Scheer, Tobias. 2008. Syllabic and Trapped Consonants in (Western) Slavic: the Same but yet Different. In G. Zybatow, L. Szucsich, U. Junghanns & R. Meyer (eds.), *Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003*. Frankfurt am Main: P. Lang. 149–167. WEB.
- Scheer, Tobias. 2009. Syllabic and trapped consonants in the light of branching onsets and licensing scales. In G. Zybatow, U. Junghanns, De. Lenertová & P. Biskup (eds.), *Studies in Formal Slavic Phonology, Morphology, Syntax, Semantics and Information Structure*. Frankfurt am Main: P. Lang. 411–426. WEB.
- Scheer, Tobias. 2010. Review of Gussmann (2007) *The Phonology of Polish*. *Studies in Polish Linguistics* 5, 109–158.
- Scheer, Tobias. 2011. Slavic Yers. In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (eds.), *The Blackwell Companion to Phonology*. New York: Wiley-Blackwell.
- Scheer, Tobias. forth a. Why Russian vowel-zero alternations are not different, and why Lower is correct. *Journal of Language and Verbal Behaviour*.
- Scheer, Tobias. forth b. Variation is in the lexicon: yer-based and epenthetic vowel-zero alternations in Polish. In E. Cyran, B. Szymanek & H. Kardela (eds.), *In memoriam Edmund Gussmann*. Berlin: de Gruyter.
- Scheer, Tobias & Markéta Ziková. 2010. The Havlík Pattern and Directional Lower. In D. Zec & W. Browne (eds.), *Formal Approaches to Slavic Linguistics. The Second Cornell Meeting 2009*. Ann Arbor: Michigan Slavic Publications. 471–486.
- Selkirk, Elisabeth. 1982. The syllable. In H. van der Hulst & N. Smith (eds.), *The Structure of Phonological Representations, Part II*. Dordrecht: Foris. 337–384.
- Szober, Stanisław. 1969. *Słownik Poprawnej Polszczyzny*. 7th edition Warsaw: Państwowy Instytut Wydawniczy.
- Szpyra, Jolanta. 1992. Ghost segments in nonlinear phonology: Polish yers. *Language* 68, 277–312. WEB.
- Szpyra, Jolanta. 1995. *Three Tiers in Polish and English Phonology*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Worth, Dean. 1968. Vowel-zero alternations in Russian derivation. *International Journal of Slavic Linguistics and Poetics* 11, 110–123.

## Annotating Foreign Learners' Czech

Barbora Štindlová & Svatava Škodová (Technical University, Liberec)  
Alexandr Rosen & Jirka Hana (Charles University, Prague)

### 1. Introduction

One of the challenges of contemporary corpus linguistics is the compilation and annotation of corpora consisting of texts produced by non-native speakers. In addition to morphosyntactic tagging and lemmatisation, such texts can be annotated by information relevant to the specific nonstandard use. Cases of deviant language use can be corrected and identified by a tag specifying the type of the error. Because of the properties of Czech, namely rich inflection, derivation, agreement, and a largely information-structure-driven constituent order, it is not straightforward to design an annotation scheme satisfying all requirements on the description of errors produced by non-native learners. Our proposal aims at an optimal solution that is still realistic given the annotation costs and the demands of the corpus users.

After an overview of issues related to learner corpora in section 2 and a brief introduction to the project of a learner corpus of Czech in section 3 we present the issues of annotation in section 4 and the concept of our annotation scheme in section 5, followed by a description of the annotation process in section 6.

### 2. Learner corpus

A learner corpus, also called interlanguage or L2 corpus, is a computerised textual database of language as produced by foreign/second language (L2) learners (Leech 1998:xiv). It is a very powerful resource in the research of second language acquisition (SLA) and foreign language teaching (FLT). It serves as a repository of authentic data about a specific variety of natural language (Granger 2003), namely the *learner language*, in the context of SLA and FLT (Selinker 1972) often called *interlanguage* (IL).

Learner corpora allow to compare non-native and native speakers' language, or to compare interlanguage varieties. They can be studied on the background of national corpora, which helps to track various deviations from standard usage in the language of non-native speakers, such as frequency patterns – cases of overuse or underuse – or *foreign-soundingness* as compared with the language of native speakers. Recent studies have focused primarily on the frequency of use of separate language elements (e.g. Ringbom 1998), collocations and prefabs (e.g.