

JAYE PADGETT

THE EMERGENCE OF CONTRASTIVE PALATALIZATION IN RUSSIAN

Abstract. The well-known contrast in Russian between palatalized and non-palatalized consonants originated roughly one thousand years ago. At that time consonants were allophonically palatalized before front vowels, as in *danʲI* 'tribute'. When the 'jer' (high, lax) vowels disappeared in certain positions, the palatalization formerly triggered by the front jer remained, leading to a palatalization contrast across most consonant types, e.g., *danʲ* 'tribute' vs. *dan* 'given' (< *danʃ*). At the same time or soon thereafter, a rule is said to have been established by which /i/ surfaced as [j] after non-palatalized consonants, e.g., *ot imʲenʲi* 'on behalf of' (< *otʃ imʲenʲi*). This paper analyzes these two sound changes within a version of Dispersion Theory (DT, Flemming 1995a) elaborated by Ní Chiosáin & Padgett (2001) and Padgett (1997, to appear). DT differs from other current models of phonology in its fundamentally *systemic* orientation: constraints evaluate not only isolated forms as is usual, but *sets* of forms in contrast. References to these systems of contrast is key to the statement of constraints governing the perceptual distinctiveness of contrasts on the one hand, and constraints directly penalizing merger (neutralization) on the other. The analysis of the Russian facts here illustrates how this theory works, and provides an explanation for the otherwise mysterious allophonic /i/ → [j] rule, and for the historical emergence of this rule as a consequence of the loss of the jers.

Keywords: Russian, Old Russian, Old East Slavic, palatalization, velarization, jer, contrast, perceptual distinctiveness, merger, neutralization.

0. INTRODUCTION¹

Russian consonants famously contrast in secondary palatalization, as in *nos* 'nose' vs. *nosʲ* 'he carried', and *vʲes* 'weight' vs. *vʲesʲ* 'entire (masc.sg.)'. The beginnings of this palatalization contrast can be traced back about a millennium. At that time consonants in Old Russian, or Old East Slavic, developed allophonic secondary palatalization before front vowels, as in *danʲI* 'tribute'. A series of diverse sound changes ensued that resulted in this secondary palatalization becoming phonemic. To use Jakobson's (1929) term, palatalization was *phonologized*. This paper investigates the beginnings of this process in Russian, the famous loss of the jers. 'Jer' is the traditional term for the Late Common Slavic high, lax vowels [ɪ,ʊ]. In certain positions, including word-finally, these vowels were lost. But consonants that had been palatalized due to a front jer remained so, e.g., *danʲ* (< *danʲɪ*). A contrast was thus established between palatalized consonants and non-palatalized consonants: compare *dan* 'given' (< *danʃ*). Of particular interest here, as soon as this occurred a new rule was established in Russian, which holds today as well: under the usual interpretation, /i/ backs to [j] after non-palatalized consonants.

Therefore former *otʊ imʲenʲi* ‘on behalf of’, for example, became *ot imʲenʲi* after [ʊ] dropped. Building on Padgett (2001), I argue that what occurred is in fact something different: before [i] non-palatalized consonants became *velarized*: *otʊ imʲenʲi* > *otʲ imʲenʲi*.

These facts are analyzed from the perspective of Dispersion Theory (henceforth DT, Flemming 1995a, Ní Chiosáin & Padgett 2001, Padgett 1997, to appear). DT, with its singularly *systemic* approach to phonology, possesses important advantages over other current theories in accounting for sound changes. Chief among these is DT’s explicit appeal to constraints regulating the perceptual distinctiveness of contrast on the one hand, and constraints directly penalizing neutralization on the other. This paper largely motivates the first of these, while Padgett (to appear) argues for both. These functional notions are not new to historical phonology – they are familiar especially from the work of Martinet (1952, 1955, 1964). But their explanatory potential has not been adequately explored, I believe, in part because they have never been made explicit enough. Related to this are serious methodological differences between generative phonology and most work in historical phonology. A strength of generative phonology is formal rigor, or at least a degree of predictive explicitness. A strength of traditional work in historical phonology, beginning with Jakobson (1929), is the emphasis on the role of the phonological system as a whole in shaping sound changes, and vice versa. Yet these strengths rarely come together. DT is unusual among formal theories in attempting to bridge this gap. I will show that DT’s appeal to the perceptual distinctiveness of contrast explains why consonants became velarized before [i] after the loss of the jers – that is, explains the ‘/i/ → [ɨ] rule’ that holds today of Russian.

The remainder of the paper is laid out as follows: following a presentation of the synchronic Russian facts involving palatalization in §1, §2 lays out in detail the dispersion theory model. §3 shows how this model explains the synchronic Russian facts, and unifies them with similar facts from other languages. §4 approaches the same facts historically, making the connection to the loss of the jers. §5 is the conclusion.

1. PALATALIZATION IN CONTEMPORARY STANDARD RUSSIAN

There are five vowel phonemes in Contemporary Standard Russian (CSR): /a,e,i,o,u/. The consonantal phonemes are given below. Most consonants are ‘paired’ (a slavistic term) for the palatalization contrast, e.g., /p/ vs. /pʲ/ and so on. Nine are traditionally viewed as unpaired: the velars, the post-alveolars, and /ts/ and /j/.² It will become clear from the following discussion why they did not pair up when palatalization was phonologized in Old East Slavic.

p	pʲ	t	tʲ		k
b	bʲ	d	dʲ		g
f	fʲ	s	sʲ	ʃ	ʃʲ
v	vʲ	z	zʲ	ʒ	ʒʲ
		ts		tʃ	
m	mʲ	n	nʲ		
		l	lʲ		
		r	rʲ		
				j	

Figure 1. Consonants of contemporary standard Russian.

Within morphological words, palatalization is contrastive before back vowels, word-finally, and to some extent pre-consonantly ((1a-c) respectively).

- | | |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (1) | <p>a. <i>mat</i> ‘foul language’ <i>mʲat</i> ‘crumpled (past part.)’
 <i>vol</i> ‘ox’ <i>vʲol</i> ‘he led’
 <i>suda</i> ‘court of law (gen.sg.)’ <i>sʲuda</i> ‘here, this way’</p> <p>b. <i>mat</i> ‘foul language’ <i>matʲ</i> ‘mother’
 <i>krof</i> ‘shelter’ <i>krofʲ</i> ‘blood’
 <i>ugol</i> ‘corner’ <i>ugolʲ</i> ‘(char)coal’</p> <p>c. <i>polka</i> ‘shelf’ <i>polʲka</i> ‘polka’
 <i>vʲetka</i> ‘branch’ <i>fʲetʲka</i> (name)
 <i>gorʲka</i> ‘hill’ <i>gorʲko</i> ‘bitterly’</p> |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Matters are more complicated before front vowels. Before /e/, historically native words in CSR are palatalized, as in (2a). However, historical loanwords in CSR can feature non-palatalized consonants before /e/, (2b). Palatalization is therefore contrastive before /e/ to a limited extent. This affects only roots, though; at morpheme boundaries consonants are invariably palatalized before /e/, as in *brat* ‘brother’ vs. *bratʲe* ‘(prep.sg.)’, *tent* ‘tent’ vs. *tentʲe* ‘(prep.sg.)’.

- | | |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (2) | <p>a. <i>sʲestʲ</i> ‘to sit down’ b. <i>tent</i> ‘tent’
 <i>pʲetʲ</i> ‘to sing’ <i>tennis</i> ‘tennis’
 <i>vʲetʲer</i> ‘wind’ <i>kep</i> ‘cap’</p> |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Palatalization is contrastive before /i/ regardless of the morphology. According to a well-known rule, however, /i/ is retracted to high, central, unrounded [ɨ] after non-palatalized consonants, as shown below (Trubetzkoy 1969, Avanesov & Sidorov 1945, Halle 1959, Hamilton 1980, Farina 1991).

(3)	<i>bʲit</i>	‘beaten’	<i>b#</i>	‘way of life’
	<i>tʲikatʲ</i>	‘to tick’	<i>tʲikatʲ</i>	‘to address in familiar form’
	<i>sʲito</i>	‘sieve’	<i>s#to</i>	‘sated (neut.sg.)’

Before leaving this general description of the palatalization contrast, one important point should be highlighted. Though the Russian contrast is often characterized as involving ‘plain’ vs. palatalized consonants, in reality the non-palatalized consonants are often velarized, as noted by Trubetzkoy (1969), Reformatzkii (1958), Fant (1960), Öhman (1966), Purcell (1979), and Evans-Romaine (1998), among others. (As we will see, the details involving the realization of palatalization and velarization are somewhat intricate.) This fact establishes a parallel between Russian and other languages exhibiting a contrast in secondary palatalization, such as Irish and Marshallese: non-palatalized consonants are velarized in these languages too. Hence the contrast would be better characterized as one involving consonantal backness or ‘tonality’, and not simply palatalization. Velarization has often been overlooked in phonological work on Russian, and yet it is important for understanding Russian phonology. For example, Padgett (2001) demonstrates based on a phonetic study that the /i/ seen above is more appropriately characterized as /i/ with velarization of the preceding consonant. That is, *b#* ‘way of life’ is actually *bʲit*, and so on. This too follows a cross-linguistic pattern: in Russian, Irish, and Marshallese, velarization of non-palatalized consonants is especially salient before front vowels. This turns out to be just one means by which the potential contrast between *bʲit* and *bit* is avoided in languages having contrastive palatalization. (For instance, some languages simply neutralize the contrast in this environment.) As Ní Chiosáin & Padgett (2001) note, it is before front vowels, and especially [i], that a contrast between truly plain and palatalized consonants would be perceptually most disfavored. In a discussion of Irish, they show that DT provides an explanation for a ‘shift’ in the realization of the palatalization contrast, from plain vs. palatalized to velarized vs. palatalized (or even plain), in front-vowel environments. These ideas will be extended to Russian in what follows.

2. DISPERSION THEORY

2.1. Idealization and systemic phonology

The analysis here is cast within a version of Dispersion Theory (Flemming 1995a), a theory that translates functional insights of Adaptive Dispersion Theory (Lindblom 1986, 1990) into Optimality Theory (Prince & Smolensky 1993). Extended to historical analysis, Dispersion Theory might also be viewed as an attempt to make more precise ideas of Martinet (1952, 1955, 1964) in particular. (For other work within Dispersion Theory see Ní Chiosáin & Padgett 2001, Padgett 1997, to appear, and Minkova & Stockwell this volume.) The next few subsections lay out and motivate the theory, while laying the groundwork for the discussion of palatalization.

The key idea of DT is that wellformedness must be evaluated not simply over isolated forms, but also with respect to the larger system of contrasts into which those forms enter. The first thing to make clear, therefore, is what precisely the object of evaluation is. What is ‘the system’, and how do we evaluate a form’s role within it? In essence, the idea is that we evaluate not single forms, as is usual, but languages (Flemming 1999). This sounds daunting at first, but it can be made very manageable once we idealize the situation sufficiently (Ní Chiosáin & Padgett 2001). This only makes explicit the kind of idealizing that phonologists do anyway.

Suppose we were analyzing /l/ velarization in American English. The facts, simplifying a bit, are that ‘clear’ /l/ occurs in onsets and ‘dark’ or velarized /l/ in rhymes. A typical analysis might consider forms such as *leaf* and *feel*. The former is [lif] and not *[tɪf] (for many speakers), the latter [fiɫ], not *[fil]. Once an analysis has been constructed that generates [lif] and [fiɫ] but rules out *[tɪf] and *[fil], we might well consider ourselves finished. We would not be expected to consider also *leek* and *keel*, to take just two other words at random having the shape, lateral-[i]-consonant, since no one imagines that the non-lateral consonants in these particular words have anything to do with /l/ velarization. If we agree that vowel quality is irrelevant (again possibly simplifying the reality), then we need not consider, say, *lake* and *kale* either. If position in the syllable – within the onset vs. within the rhyme – is truly all that matters, then there is no pressing reason to explicitly consider any other forms at all. This is reasonable because the four forms derived are *representative* of everything that matters. (Obviously this argument does not hinge on whether it is four forms, or six, or some other relatively small number.) In *practice*, therefore, we have idealized severely, entertaining a possible world in which the only words that could exist are [lif], [fil], [tɪf], and [fiɫ]. In *principle* there are no such limitations. For instance, should we discover that vowel quality does matter, then we would be obliged to consider more forms in our idealization. The point of idealization is to make analysis possible, and it is unavoidable. We could not pass every possible word of English through our constraint tableaux (and no one would want us to). We can often make do with an extremely small number of forms.

This reasoning holds regardless of the theoretical framework employed. Translating the hypothetical scenario above into Optimality Theoretic terms, we might say that the inputs and candidates considered must be chosen from these four forms and no others. This amounts to a kind of “tactical constraint on richness of the base and on GEN” (see Prince & Smolensky 1993 on these two notions). It is ‘tactical’ because in fact richness of the base and GEN hold as usual, and again, should we decide that other properties of a form are relevant to the analysis, we would have to address those forms as well. In most work within OT, even though the number of forms considered is in fact quite small, the idea of an idealization is not explicitly addressed. The proposal here, following Ní Chiosáin & Padgett (2001), and Padgett (to appear), is to make the working idealization clear *up front*. This is necessary because of the assumption that candidates are languages. Without some clear limitations on what a candidate ‘language’ could be, it would be impossible to reliably evaluate them or understand what the relevant competitors might be.

Any idealization assumed always depends on the phenomenon to be analyzed, since what matters for /l/ velarization differs from what matters for, say, vowel harmony. In order to analyze the facts of Old Russian, I will rely on the idealization shown below. In what follows, an input, or a candidate output, will be understood to be any subset of the forms implied by (4).

$$(4) \quad \text{Word} = m^{(x)}V_1l^{(y)}(V_2), \text{ where } V_1 = \{i, \ddot{i}\}; V_2 \in \{I, \sigma\}; x, y \in \{j, \gamma\}$$

That is, a ‘language’ must consist of a set of words having the shape $mVl(V)$, where /m/ and /l/ can each be plain, palatalized, or velarized, and the vowels can be as shown. This idealization allows in all 54 possible words, including $mil\sigma$, $m^{\gamma}i\ddot{i}\sigma$, $m^jil^{\gamma}I$, $m\ddot{i}I$, $m\ddot{i}\ddot{i}$, $m^{\gamma}\ddot{i}^j$, and so on. Of course, no candidate *need* include very many of these words, so long as it is clear what is ruled in and what is ruled out. We will see how this works below.

2.2. *Perceptual distinctiveness*

There are two reasons for taking the objects of evaluation to be ‘languages’ and not simply forms in isolation. The first, of central importance to this paper, involves DT’s appeal to perceptual distinctiveness of contrasts. A constraining assumption of DT is that markedness constraints are grounded in independently motivated properties of the human mind and physiology. (This line of reasoning extends in particular ways the general notion of ‘grounding’ of constraints, as in Archangeli & Pulleyblank 1994.) The two important sources of grounding explored by DT – but not necessarily the only ones around³ – are the familiar competing notions of articulatory effort, on the one hand, and perceptual distinctiveness of contrasts, on the other. The latter, obviously, is inherently comparative: *two or more* forms are distinct from each other to some degree. We must therefore evaluate sets of forms – our ‘languages’ – and not forms in isolation.

For arguments that perceptual distinctiveness is crucial to an accurate conception of markedness, the reader is referred to the references on DT given earlier, and to phonetic works cited in those references, such as Lindblom (1986). Ní Chiosáin & Padgett (2001) and Padgett (2001) discuss one that is relevant here and has already been mentioned: there is a tendency, in languages having contrastive palatalization, for non-palatalized consonants to be velarized. To a first approximation, for example, Russian has contrasts such as /b^j/ vs. /b^γ/, with *no* plain /b/, in spite of appearances in Figure 1. (I begin with this simple characterization, and show later how it must be refined.) From the perspective of any markedness theory that ranks all segments along a single scale, this presents a problem: since plain consonants must be ‘less marked’ than either palatalized or velarized ones, it should not be possible for a language to have only the latter two kinds. This is easy to see in Optimality Theoretic terms, as in (5). Assume first (and uncontroversially) that markedness constraints prohibiting consonants with a secondary articulation universally outrank constraints against their plain counterparts, as shown. If all of

These constraints are relativized to the auditory dimension of contrast in question, here consonantal F2 (C-F2), and they correspond roughly to Flemming's (1995a) 'Minimal Distance' constraints.⁴ The ranking seen in (6c) is universal, reflecting the need (all things equal) to maximize the perceptual spacing of contrasts. The number of Space constraints can differ across auditory dimensions. Since a three-way contrast is the upper limit for consonantal backness, we can assume that $\text{SPACE}_{\text{C-F2}} \geq 1/3$ is in GEN, in OT terms. (That is, no candidates will be generated that would violate this constraint.) This leaves $\text{SPACE}_{\text{C-F2}} \geq 1/2$ and $\text{SPACE}_{\text{C-F2}} \geq 1$ to be ranked in constraint tableaux.

Generative phonology has long assumed that it is the inventory of phonological features that provides our theory of possible contrasts. This is the essence of distinctive feature theory. For example, to explain why a three-way contrast for consonantal backness is the largest possible, that theory stipulates that there are only three feature values available for this dimension, [front], [back] (or [+/-back]), and unspecified. (The precise choice of features and values is not the issue here.) DT departs from this view in claiming that upper limits on contrast follow directly from output constraints on perceptual distinctiveness, as seen here. Placing $\text{SPACE}_{\text{C-F2}} \geq 1/3$ in GEN does the work of the distinctive feature theory stipulation just mentioned. Since SPACE constraints guarantee that contrasts will not be overgenerated, we are free to enlarge the inventory of phonological features to include distinctions that are never contrastive, if they are important for stating phonological generalizations. Recent work argues that they are, including Steriade (1994, 2000), Flemming (1995b, 2001), Kirchner (1997), Boersma (1998), Zhang (2000), Ní Chiosáin & Padgett (2001), and Padgett (2002).

The wording of SPACE presupposes a notion 'potential minimal pair'. Following Padgett (to appear), I suggest this be made precise by means of the simple extension of correspondence theory (McCarthy & Prince 1995) depicted in (7). Suppose we index the segments of a word from left to right, in a way identical for all words. (For instance, we can begin with the number 1 and continue until we run out of segments, as shown.) Then we can define a potential minimal pair as two words, all but one of whose corresponding segments are identical. (This definition ignores pairs of words that have a different number of segments, for the sake of simplicity.) (7a-b) are therefore both potential minimal pairs, though only (7a) *passes* the minimal pair test for English. To 'pass the minimal pair test', roughly speaking, is to be sufficiently distinct such that a difference in meaning could be supported. It is this notion of distinctness that SPACE constraints attempt to make more precise.

$$(7) \quad \begin{array}{cc} \text{a. } & b_1 \ \alpha_2 \ t_3 \\ & | \ | \ | \\ & b_1 \ \alpha_2 \ k_3 \end{array} \quad \begin{array}{cc} \text{b. } & b_1 \ \alpha_2 \ t_3 \\ & | \ | \ | \\ & b_1 \ \alpha_2 \ t_3^h \end{array}$$

We can now illustrate how SPACE constraints work. Suppose that for the sake of this illustration we consider only 'words' of the form $b^{(j/x)}V$, in which /b/ is plain, palatalized, or velarized, and V is some vowel. Under this idealization there are only three words a candidate 'language' could have. All three are shown in the input in Tableau 1. (Placing all possible words in the input is equivalent to adhering to

richness of the base, modulo the idealization. See §4.) Candidate 1a is fully faithful to this input, but fares worst of all on SPACE. SPACE constraints consider separately each possible pairing of words within a candidate language. In 1a there are three – [b^jV] vs. [bV], [bV] vs. [b^yV], and [b^jV] vs. [b^yV]. Of these, the first two violate SPACE_{≥1/2} (see (6a)), hence the two violations counted for this constraint. Every pairing within this candidate violates SPACE_{≥1}, since this constraint requires that every segment have the entire perceptual space to itself. Candidate 1b has only one pair of words, but this pair also violates both SPACE_{≥1/2} and SPACE_{≥1}. (Note that the words within each candidate are arranged so as to suggest their perceptual similarity.) 1c violates only SPACE_{≥1}, and so is the best pairing of words possible. 1d has the same violations as 1c, because the pair [dV] vs. [b^jV] (or [b^yV]) is *not* a minimal pair, and so passes SPACE vacuously. (Refer again to (6b).) Finally, 1e passes all SPACE constraints vacuously. The best way to satisfy SPACE, therefore, is to attempt no contrast.

Tableau 1. How SPACE_{C-F2} evaluates candidate languages

b ^j V bV b ^y V	SPACE _{≥1/2}	SPACE _{≥1}
a. b ^j V bV b ^y V	**	***
b. b ^j V bV	*	*
c. b ^j V b ^y V		*
d. b ^j V b ^y V dV		*
e. b ^j V		

Before going on to address the markedness paradox discussed above (refer again to (5)), we must deal with issues of contrast and faithfulness.

2.3. Contrast and neutralization avoidance

In standard OT, contrast is maintained by faithfulness constraints. A typical faithfulness constraint, following the correspondence theoretic formulation of McCarthy & Prince (1995), is given in (8). The approach to DT assumed here and in Padgett (to appear), as opposed to earlier ones, likewise employs faithfulness constraints.

- (8) IDENT(PAL): Let S_i and S_o be corresponding consonants of the input and output. Then S_i is [αback] iff S_o is [αback].

Padgett (to appear) argues in addition for a new faithfulness constraint, shown in (9). This constraint is analogous to the constraint UNIFORMITY of McCarthy &

Prince (1995), but it applies over *words* rather than segments. Though conventional faithfulness constraints preserve contrast, they do so only indirectly, by mitigating against changes in potentially contrasting words. Flemming (1995a) and Padgett (to appear) argue that phonology must appeal more *directly* to neutralization avoidance, that is, the desire to maintain contrasts in the first place. *MERGE does this, doing similar work to Flemming's (1995a) 'Maintain Contrast' constraints, though recasting the notion in terms of faithfulness. This is the second reason, besides the appeal to perceptual distinctiveness, for DT's systemic approach to phonology, in which inputs and candidates are sets of forms ('languages'): obviously, the formulation of *MERGE requires this view.

- (9) *MERGE: No output word has multiple correspondents in the input.

The following tableau shows how the idea works. From here on subscripts refer to entire words and not to single segments as is usual in correspondence theory. Candidate 2a is fully faithful to the input, but candidate 2b has merged input /b^yV/ and /bV/, as the subscripts show, and so violates *MERGE. It also violates IDENT(PAL). Candidate 2c does *not* violate *MERGE, but it does violate another faithfulness constraint involving continuancy.

Tableau 2. How *MERGE evaluates candidate languages

b ^j V ₁ bV ₂ b ^y V ₃	*MERGE	IDENT(PAL)	IDENT(CONT)
a. b ^j V ₁ bV ₂ b ^y V ₃			
b. b ^j V ₁ bV _{2,3}	*	*	
c. b ^j V ₁ bV ₂ β ^y V ₃			*

It should be emphasized that 2b involves a *merger* of words (by virtue of some feature change), and not any *deletion* of words, nor deletion of segments. This is indicated by the subscript notation. What *does* change in 2b is the consonantal [back] specification of input /b^yV/. It should be clear that any time *MERGE is violated, some conventional faithfulness constraint must be violated as well, since neutralization entails some change in feature value, segment make-up, segment order, or the like. There is overlap, therefore, between *MERGE and other faithfulness constraints. In spite of this overlap, the two notions are crucially distinct. It is possible to violate conventional faithfulness without violating *MERGE, as 2c shows. This candidate imagines a shift of [b^y] to bilabial [β^y] in a language that previously lacked [β^y]. In general, conventional faithfulness constraints are required in order to explain the tendency to preserve identity even when merger is not at stake. On the other hand, *MERGE explains facts that conventional faithfulness constraints cannot, as Padgett (to appear) shows. *MERGE does not play a crucial role here, in contrast, though it comes up in §4.

3. EXPLAINING THE SYNCHRONIC PATTERN

With the basic ideas of the theory in place, this section shows how the intricate distribution of distinctive palatalization and velarization of Russian is derived. Consider again the input language shown in Tableau 3. (The content of this section follows Ní Chiosáin & Padgett 2001 in most respects.) Given the ranking $SPACE_{\geq 1/2} \gg IDENT(PAL) \gg SPACE_{\geq 1}$, the optimal candidate must be either 3c or d, one with the most dispersed contrast. This is the kind of candidate that conventional markedness assumptions cannot derive, as seen in (5). Candidates 3c and d differ only in the fate of underlying /bV/: it merges with /b^yV/ in 3c and with /b^jV/ in 3d. Both outcomes seem plausible in principle, and they can be distinguished in a particular case by factoring IDENT(PAL) into two constraints, one preserving input [+back] values, the other input [+back] values.⁵

Tableau 3. Dispersed contrast

$b^jV_1 bV_2 b^yV_3$	$SPACE_{\geq 1/2}$	IDENT(PAL)	$SPACE_{\geq 1}$	*C ^j	*C ^y
a. $b^jV_1 bV_2 b^yV_3$	*!*		***	*	*
b. $b^jV_1 bV_{2,3}$	*!	*	*	*	
c. $b^jV_1 b^yV_{2,3}$		*	*	*	*
d. $b^jV_{1,2} b^yV_3$		*	*	*	*
e. $b^jV_{1,2,3}$		**!		*	
f. $bV_{1,2,3}$		**!			
g. $b^yV_{1,2,3}$		**!			*

If both SPACE constraints outrank faithfulness, then a ‘language’ with no contrast in consonantal backness will be favored, as in Tableau 4. Under this scenario, articulatory markedness constraints make the choice, and the plain [b] is preferred.

Two aspects of the account here are worth stressing. First is the bidimensional approach to markedness: outputs favored by perceptually based constraints do not necessarily subsume those favored by articulatory constraints, or vice versa. What is a real problem for conventional markedness as in (5) receives a straightforward resolution here. Second is DT’s systemic approach to perceptual distinctiveness: SPACE constraints evaluate not [b^jV], [bV], or [b^yV] in isolation, but the perceptual distance between *pairs* of such output forms. This is also necessary for an adequate explanation of the facts, and distinguishes DT even from other approaches to phonology that appeal to perceptual distinctiveness. Compare for example Steriade’s (2001) proposal to establish hierarchies of faithfulness constraints (or in other work, markedness constraints) according to perceptual distinctiveness in a given *syntagmatic* context. Word-final obstruent devoicing occurs, for instance, rather than

prevocalic devoicing, according to this account, because IDENT(VOICE)/__V is distinguished from, and universally outranks, IDENT(VOICE)/__#. This follows from the fact that the contrast between voiced and voiceless obstruents is perceptually more distinct in the former context. It is true that perceptual distance depends on syntagmatic context – see also below. And this sort of account works excellently for binary contrasts such as that of voicing, where the question is simply to contrast or not in a particular context. However, for contrast dimensions that allow three or more contrasting degrees, it is not enough to sanction contrast or not: we must be allowed to regulate *how much* contrast, in the sense of perceptual distance. This is a matter of *paradigmatic* context, that is, the system of contrasts into which a form enters, which context-dependent faithfulness (or markedness) does not address. Therefore it shares with unidimensional markedness the problem exemplified by (5) of failing to explain dispersed contrasts as in Tableau 3c-d. Why does Russian have [l̥] and [l̥ʲ] but no [l̥]? SPACE constraints handle both paradigmatic and syntagmatic context in a unified way, as we will see.

Tableau 4. Articulatory simplicity

$b^jV_1 bV_2 b^yV_3$	SPACE $\geq 1/2$	SPACE ≥ 1	IDENT(PAL)	*C ^j	*C ^y
a. $b^jV_1 bV_2 b^yV_3$	*!*	***		*	*
b. $b^jV_1 bV_{2,3}$	*!	*	*	*	
c. $b^jV_1 b^yV_{2,3}$		*!	*	*	*
d. $b^jV_{1,2} b^yV_3$		*!	*	*	*
e. $b^jV_{1,2,3}$			**	*!	
f. $bV_{1,2,3}$			**		
g. $b^yV_{1,2,3}$			**		*!

For completeness, let us consider the other candidates shown in these tableaux. Candidate 4a will be favored if IDENT(PAL) is undominated. As noted, this is an attested pattern as well, at least for [l]. If it turns out not to be possible for other consonants, such as obstruents, this would imply that SPACE constraints must be further broken down according to consonant type, with SPACE $\geq 1/2$ being in GEN (inviolable) for obstruents. Candidate 4b, or the analogous candidate having only [bV] and [b^yV], represents the possibility of a contrast that is maintained but not maximally dispersed. It is possible to output the former if *C^y is undominated, and the latter if *C^j is undominated. If this is a problem, it is a problem for the standard theory as well. Candidates 4e and g, ‘languages’ having exclusively palatalized or velarized segments respectively, should presumably never win. And they will not, given the articulatory markedness hierarchy: plain segments harmonically bound their complex counterparts, as can be seen in Tableau 4.

The discussion above considers differences among consonants while abstracting away from the vocalic environment. In reality consonants occur in a range of environments having a significant effect on the perceptual distinctiveness of palatalization contrasts. Compare for instance the contrasts [b^lu] vs. [bu] and [b^li] vs. [bi]. It is clear that the latter contrast is perceptually much less distinct, because an off-glide [j] is acoustically very similar to [i]. Conversely for velarized consonants, a contrast such as [b^ɣi] vs. [bi] is much more salient than one between [b^ɣu] and [bu]. (Here ‘[b^ɣu]’ should be understood as an attempt to increase the velar constriction in comparison to ‘[bu]’). Since [u] is already velarized, the effect of this is necessarily slight.) To get an idea of the possibilities, consider Figure 2, an attempt to convey the perceptual difference between various consonant-vowel syllable pairs. The figure takes each Russian vowel as a context for a contrast in consonantal backness on a preceding consonant. For each vowel, the perceptual distinctiveness of a palatalized versus plain, a plain versus velarized, and a palatalized versus velarized, consonant, are compared. For each CV sequence, first and second formant values were measured at consonantal release, where cues to consonantal backness predominate. Following Ménard et al. (2002) these values were converted from Hertz to Bark, a better measure of perceptual differences, and ‘backness’ for each consonant was taken to be the value in Bark of F2-F1. Once these values were found, the difference in backness (measured in this way) between each pair of syllables was taken. The lighter bars indicate data based on the author’s own attempt at pronunciations of these sequences, and the darker ones those of a native speaker of Russian.⁶ (Many of these sequences are not possible in Russian; the native speaker could only be asked to produce those that are.)

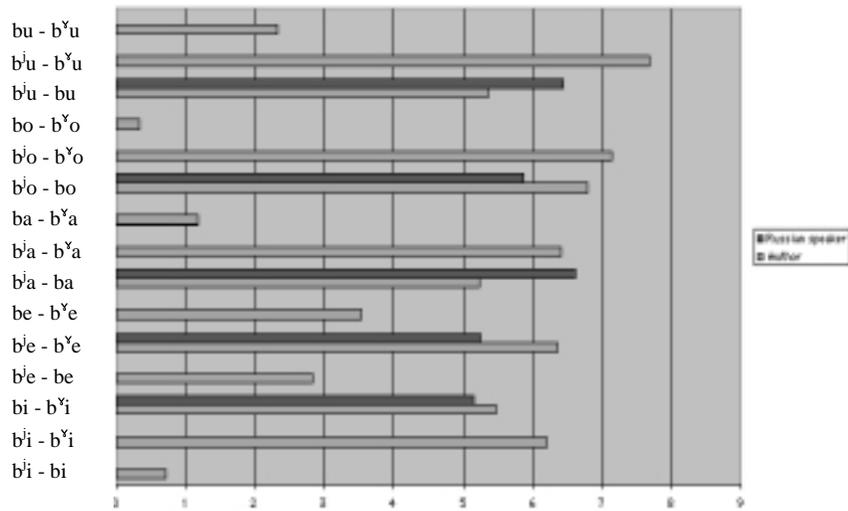


Figure 2. Difference in backness between selected CV sequences at consonantal release (where ‘backness’ = F2-F1 in Bark).

Since this chart is based on quite a limited amount of data, we should consider only the grossest differences among contrasts suggested by it. In addition, it should be kept in mind that we are focusing solely on ‘backness’ as indicated by *formant transitions* at release. There are other correlates of ‘palatalization’ that must be borne in mind in the larger picture, such as affrication in the case of coronals; see Padgett (2001). To avoid this complication, here we consider only labials.

With these caveats in mind, the five worst contrasts based on this diagram, differing by 3 Bark or less, are [bu] versus [b^yu], [bo] versus [b^yo], [ba] versus [b^ya], [be] versus [b^ye], and [bi] versus [b^ji].⁷ These are just the contrasts in which the only secondary articulation brought to bear is of the same backness as the following vowel. On the other hand, of the rest, all but perhaps [be] versus [b^ye] represent differences approximating or exceeding one-half of the full backness range (about 11 Bark for the author). I will therefore assume that these ‘good’ contrasts pass the constraint $SPACE_{C-F2} \geq 1/2$ seen above, while the ‘bad’ ones, and [be] versus [b^ye], do not.

Consider now forms having the shape $b^{(i/y)}V$, where V is [i] or [u] (a total of six possible words), shown in the tableau below. This tableau repeats the ranking of Tableau 3, which achieves dispersed contrast. The faithful candidate 5a violates $SPACE \geq 1/2$ twice, for the poor contrasts [b^ji] vs. [bi] and [bu] vs. [b^yu]. 5b is just as bad by this constraint, since it preserves *only* the poor contrasts. 5c-e pass this constraint, but 5e neutralizes to an unnecessary degree. 5c-d differ only in overall articulatory complexity. 5c is optimal by this final criterion. Here and throughout I assume that where merger occurs, a consonant merges with its closest neighbor, e.g., [b^j] or [b^y] with plain [b], and [b] with the perceptually closer of [b^j] and [b^y] given the vowel context. (That is, /bi/ merges with [b^ji], while [bu] merges with [b^yu].)⁸

Tableau 5. Context-dependent dispersed contrast

$b^i i_1$ bi_2 $b^j u_4$	$b^y i_3$ bu_5 $b^y u_6$	$SPACE \geq 1/2$	IDENT(PAL)	$SPACE \geq 1$	*C ^j	*C ^y
a. $b^i i_1$ bi_2 $b^j u_4$	$b^y i_3$ bu_5 $b^y u_6$	*!*		*****	**	**
b. $b^i i_1$ $bi_{2,3}$ $bu_{4,5}$ $b^y u_6$		*!*	**	**	*	*
c. $bi_{1,2}$ $b^j u_4$	$b^y i_3$ $bu_{5,6}$		**	**	*	*
d. $b^i i_{1,2}$ $b^j u_4$	$b^y i_3$ $b^y u_{5,6}$		**	**	***!	***!
e. $bi_{1,2,3}$ $bu_{4,5,6}$			***!*			

The overall pattern that emerges, based on these constraint rankings, is one in which a well-dispersed palatalized vs. plain contrast is maintained before back vowels, while a *plain* vs. *velarized* one holds before front vowels. Two aspects of these results should be emphasized, since they highlight distinctive properties of DT. First, the nature of the contrast ‘shifts’, depending on vocalic environment, in just such a way as to maximize the perceptual distinctiveness of the contrast. Second, once the needs of perceptual distinctiveness (and faithfulness) are met, as in 5c-d, it is articulatory simplicity that decides the rest. In other words, a palatalized versus velarized contrast is avoided as articulatory ‘overkill’ given the perceptual needs. As Ní Chiosáin & Padgett (2001) point out, this is precisely the realization of the palatalization contrast before high vowels in Irish, as shown in (10a). As foreshadowed in §1, I extend this claim to Russian as well, as in (10b).

(10) Contrast shift in Irish (a) and Russian (b)

a.	<i>fʷu:</i>	‘worth’	<i>fu:ə</i>	‘hate’
	<i>bi:</i>	‘be (imp.)’	<i>b^ɣi:</i>	‘yellow’
b.	<i>b^hust</i>	‘bust’	<i>busts^ɣi</i>	‘soccer cleats’
	<i>bit</i>	‘beaten’	<i>b^ɣit</i>	‘way of life’

Within (10b), it is only the facts before /i/ that require discussion; the palatalized vs. plain contrast before /u/ is uncontroversial. Padgett (2001) demonstrates that Russian ‘ɨ’ is best characterized as velarization of the preceding consonant before /i/: the second formant of this ‘vowel’ is quite low at the release of the preceding consonant, but is virtually identical to that of [i] at its end. Hence the transcription *b^ɣit* instead of *b^hi* (and incidentally *busts^ɣi* for *busts^h*). (Many phonetic descriptions of Russian note the ‘diphthongized’ pronunciation of ‘ɨ’, which when stressed is really not at all the high, central, unrounded [ɨ] it is often said to be.) At the same time, before [i] the palatal off-glide of ‘palatalized’ consonants is only weakly present or not present at all, a fact noted in phonetic descriptions (e.g., Jones & Ward 1969, Zubkova 1974). (Again it is only this off-glide that is under discussion; palatalized coronal obstruent stops, for example, are typically affricated to some degree, and this provides another cue to the presence of palatalization.)

Consider now words with the vowels [e,o] instead. The predictions are the same but for one point: according to Figure 2, we concluded that contrasts such as [be] vs. [b^ɣe] fall short of satisfying $SPACE \geq 1/2$, unlike [bi] vs. [b^ɣi]. Given this, more articulatory complexity is required, in the form of palatalization of [b^he], in order to meet the perceptual distinctiveness requirement, as shown below. (The less harmonic candidates from the previous tableau have been omitted here, but would be treated analogously.) The reason for this difference between the high vowels and the mid vowels is well understood: the range of F2 (or F2-F1) values achievable shrinks as vowels lower, becoming a smaller fraction of the full F2 range. It is this fact, for example, that explains why languages frequently do not contrast front and back low vowels.

Tableau 6. Before mid vowels

	SPACE \geq 1/2	IDENT(PAL)	SPACE \geq 1	*C ^J	*C ^Y
b ^j e ₁ be ₂ b ^y e ₃ b ^j o ₄ bo ₅ b ^y o ₆					
a. b ^j e ₁ be ₂ b ^y e ₃ b ^j o ₄ bo ₅ b ^y o ₆	*! **		*****	**	**
b. be _{1,2} b ^y e ₃ b ^j o ₄ bo _{5,6}	*!	**	**	*	*
c. b ^j e _{1,2} b ^y e ₃ b ^j o ₄ bo _{5,6}		**	**	**	*

Again this predicts the facts in the case of Russian. ‘Palatalized’ consonants before [e] are indeed clearly palatalized, as shown in (11a). As noted in §1, there are many historical loans having non-palatalized consonants before [e]. These consonants are in fact always velarized, (11b). Moreover, sequences of non-palatalized consonant plus [e] are permitted across morphological word boundaries for a limited number of native Russian words related to the deictic pronoun *eto* ‘this/that’, as shown in (11c) (as well as for historically borrowed words). Here again the consonants are velarized.

- (11) a. *s^jestʲ* ‘to sit down’ b. *t^yent* ‘tent’
 p^jetʲ ‘to sing’ *t^yennis* ‘tennis’
 v^jetʲer ‘wind’ *k^yep* ‘cap’
- c. *v + etom* → *v^yetom* ‘in this/that’
 k + etomu → *k^yetomu* ‘toward this/that’

The facts are different this time in the case of Irish (Ní Chiosáin & Padgett 2001): ‘plain’ consonants are velarized before [e] as in Russian, but ‘palatalized’ consonants have little or no palatal off-glide before [e]. Perhaps this difference is related to a difference between the languages in the pronunciation of mid vowels: Russian [e] and [o] are low-mid, in fact best transcribed [ɛ] and [ɔ] in most environments, while the Irish vowels are [e] and [o]. (In fact, the pattern discussed here for Irish holds only of long, tense vowels.) Given the connection between vowel height and F2 noted above, [bɛ] and [b^yɛ] must be less distinct than are [be] and [b^ye] (or equivalently, [be] and [b^je] are less distinct than are [bɛ] and [b^jɛ]), providing a possible motivation for the required off-glide of Russian in [b^jɛ].

Finally, Tableau 7 shows what happens before the vowel [a]. According to the account, palatalized consonants should contrast with plain ones in this context, since this contrast satisfies SPACE \geq 1/2. This prediction seems correct for both Irish and Russian. That is, [ba] differs from [b^yai] and [b^yae] in lacking any clear velarization. Here, it should be noted, and in other contexts such as preconsonantly or word-

finally, the facts involving velarization are least clear. (See the references from §1.) At least for some speakers, consonants even here might be velarized, perhaps weakly. However, Purcell's (1979) phonetic study found that F2 values were much less variable for palatalized consonants due to a following vowel than for non-palatalized ones, a finding that seems consistent with the view that velarization is weaker or absent.⁹

Tableau 7. Before the low vowel

$b^j a_1$	ba_2	$b^y a_3$	SPACE \geq 1/2	IDENT(PAL)	SPACE \geq 1	*C ^j	*C ^y
a. $b^j a_1$	ba_2	$b^y a_3$	*!		***	*	*
b. $b^j a_1$	$ba_{2,3}$			*	*	*	
c.	$ba_{1,2}$	$b^y a_3$	*!	*	*		*
d. $b^j a_1$		$b^y a_{2,3}$		*	*	*	*!

4. HISTORICAL ANALYSIS

The previous section shows how the complex distribution of palatalized, velarized, and plain consonants in Russian and Irish can be explained given the tenets of Dispersion Theory. We now turn to the historical facts to be analyzed. The goal of the rest of the paper is to show how the famous /i/ → [i̯] allophonic rule, understood here as velarization of a consonant before [i], emerged as a natural consequence of the loss of the jers. Though a connection between the loss of the jers and this rule (and distinctive palatalization) have long been recognized, the precise reason for the connection has never been made clear, so far as I know. §4.1 provides the facts and analysis of Old Russian before the changes. §§4.2 and 4.3 deal with the changes of interest, the loss of the jers and the introduction of the /i/ → [i̯] rule.

4.1. Old Russian at the beginning

Common Slavic is generally considered to have begun its disintegration into major Slavic dialects around the sixth century A.D. (Of course, there is always some arbitrariness in such dates. See, however, Shevelov 1965, Carlton 1991.) What resulted are three major dialects, South Slavic, West Slavic, and East Slavic. Old East Slavic, also known as Old Russian, is the parent language of Russian, Belorussian, and Ukrainian. It is well attested in documents dating from the tenth century. The change to be analyzed here, the loss of the jers, began about the time of historical attestation, and was certainly over by the thirteenth century. I will use the term 'Old Russian' here, since I explicitly consider this history in light of Contemporary Standard Russian, without reference to Belorussian or Ukrainian. The discussion of Old Russian here is largely based on Sobolevskii (1907/1962),

Jakobson (1929), Chernykh (1962), Borkovskii and Kuznetsov (1963), Filin (1972), Kiparsky (1979), and Ivanov (1990).

The phoneme inventory of Old Russian, at about the tenth century, is shown below. /I,Ů/ are the Slavic ‘jers’, vowels usually considered to have been short (and probably lax) counterparts of /i,u/. /ě/ denotes a vowel that was probably either a diphthong /ie/ or simply higher than /e/. (In modern dialects it is realized as /e/, /ie/, and /i/.)

i	ĭ	u	p	t	k
I		Ů	b	d	g
ě			s	s ^j	ʃ ^j x
e		o	v	z	z ^j
	a			ts ^j	tʃ ^j
			m	n	n ^j
			l	l ^j	
			r	r ^j	j

Figure 3. Old Russian phonemes.

As can be seen, while Old Russian had a much richer vowel system than does CSR, its consonantal inventory was smaller. In particular, it did not have the pervasive palatalization contrast of CSR. Though the sounds /ts^j,ʃ^j,z^j,tʃ^j/ were palatalized, they had no non-palatalized counterparts of the same manner and place. This is because they were derived earlier in Common Slavic by a series of palatalizing mutations affecting the velars when adjacent to front vocoids. (/tʃ^j/ is still palatalized today in CSR, but the other three sounds lost their palatalization later.) Five other palatalized phonemes at this stage were possibly paired: /s,z,n,l,r/ vs. /s^j,z^j,n^j,l^j,r^j/. /s^j,z^j/ were derived along with /ts^j/ by velar mutations (from /x,g/ respectively). The palatalized coronal sonorants were derived by a historical merger of /n,l,r/ + /j/. Jakobson (1929) assumes that palatalization was not contrastive even for these sounds, and Lunt (1956) argues this point explicitly. This conclusion rests on one’s treatment of the vocalic system. The paired treatment assumes the vowel phonemes shown above, while Jakobson’s and Lunt’s posits the extra vowel phonemes /æ/ and /y/ (/ü/). Even under the former view, compared to the distribution of the palatalization contrast in CSR, that of these Old Russian sounds was at best very limited. Within roots, /s^j,z^j/ contrasted with /s,z/ only before /ě/; /n^j,l^j,r^j/ contrasted with /n,l,r/ only before /u/. Across a morpheme boundary, these sounds contrasted with their plain counterparts before most front vowels and before /a,u/. There was no palatalization contrast before any other vowels; nor was there a contrast word-finally or before other consonants: Old Russian syllables were open.

Recall the idealization adopted in §2.1 for this analysis: a ‘language’ must consist of a set of words having the shape $mV_1l(V_2)$, where [m] and [l] can each be plain, palatalized, or velarized, V_1 can be either of [i,ĭ], and V_2 can be either of [I,Ů]. Of the 54 words made possible by this idealization, only four were actually possible words of Old Russian:

$$(12) \quad m^i\dot{i}I \quad m^i\dot{i}I^j \quad m^i\dot{i}I^j\dot{I} \quad m^i\dot{i}I^j\dot{I}^j$$

In Old Russian, consonants before front vowels were allophonically palatalized. Palatalized consonants did not occur at all before [ɨ,ʊ], the back vowels considered in the idealization. There is no evidence that the language had velarization. (However, it should be kept in mind that evidence for such conclusions is at best very indirect; this is true even of the evidence for allophonic palatalization, as we will see.) In addition, syllables were open, as noted (with very limited exceptions, see Lunt 1956). These generalizations together account for (12).

The main reason consonants before front vowels are thought to have been palatalized is the phonologization of palatalization that followed: it seems likely that for phonemic palatalization to have arisen, consonants must have been allophonically palatalized. One wonders whether consonants before [i] had any more appreciable an off-glide in Old Russian than they do now. For the sake of discussion I will follow others in assuming they did. I take allophonic palatalization before front vowels to be forced by a constraint PAL(ATALIZE). (Allophonic palatalization is plausibly another contrast dispersion effect, as Flemming 1995a argues, but I do not pursue this point here.) In order for this constraint to have any effect, it must dominate both IDENT(PAL) and *C^j, as shown below.

Tableau 8. Allophonic palatalization

milɪ	PAL	IDENT(PAL)	*C ^j
a. milɪ	*!*		
b. m ^j ilʲɪ		**	**

In OT, the tenet of richness of the base (ROTB; Prince & Smolensky 1993:191-6) holds that every possible linguistic form is a licit input. It is therefore the job of the output constraint hierarchy alone to winnow down all of the input possibilities to those that conform to the requirements of a given language. Since we are operating within an idealization in which only 54 forms are linguistically possible, richness of the base in our terms implies that *all* of these 54 forms together in principle make up the input. But it is clear from just the input considered above why PAL » IDENT(PAL), *C^j.

Since palatalization was not contrastive, we can also infer *C^j » IDENT(PAL), as shown below.

Tableau 9. No contrastive palatalization

m ^j ɨlʲʊ	*C ^j	IDENT(PAL)
a. m ^j ɨlʲʊ	*!*	
b. mɨlʲʊ		**

Assuming no velarization, we can also see that $*C^y$ must outrank IDENT(PAL) (which mitigates against *any* changes in consonantal backness):

Tableau 10. No velarization

$m^y i l^y \emptyset$	$*C^y$	IDENT(PAL)
a. $m^y i l^y \emptyset$	$*!*$	
b. b. $m i l \emptyset$		**

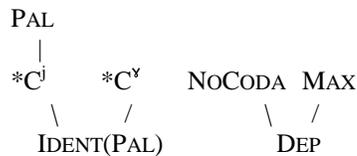
Because Old Russian banned closed syllables, the constraint banning them, NOCODA, must have dominated either DEP, a constraint that prohibits insertion of segments, or MAX, one that prohibits deletion of segments. For the sake of discussion I assume it was DEP, and that the vowel inserted was [ɔ]:

Tableau 11. No closed syllables

$m i l$	NOCODA	MAX	DEP
a. $m i l$	$*!$		
b. $m i$		$*!$	
c. c. $m i l \emptyset$			*

To sum up so far, the constraints and rankings shown below are sufficient to select the four forms of (12), assuming all 54 possible forms as input. It would obviously be difficult to show this in one constraint tableau, but the conclusion should be clear based on what we have seen. Since palatalization was not yet phonemic, the relevance of the earlier dispersion theory discussion is not yet apparent, but this will soon change.

(13) Old Russian constraints and rankings



4.2. *The loss of the jers*

Old Russian phonemic palatalization can be said (as in Jakobson 1929, Lunt 1956, Filin 1972, and Kiparsky 1979) to have originated in earnest with a sound change soon to occur: the loss of the jers. Before this occurred, the front jer [ɪ] had triggered allophonic palatalization, as all front vowels did, on the preceding consonant. When the jers disappeared (only in so-called ‘weak’ positions, see below), this palatalization was retained, and minimal pairs such as those shown in (14) were created for the first time. This sound change was the first source of a palatalization contrast across *all* of the ‘paired’ consonants of CSR, and in non-prevocalic environments.

- | | | | | |
|------|---------------------------------------------------|-----------|-----------------------------------------------------|-------------------|
| (14) | <i>dan</i> σ > <i>dan</i> | ‘given’ | <i>klad</i> σ > <i>klad</i> | ‘buried treasure’ |
| | <i>dan</i> ^ɪ > <i>dan</i> ^ʲ | ‘tribute’ | <i>klad</i> ^ɪ > <i>klad</i> ^ʲ | ‘load’ |

The loss of the jers and its consequences illustrates a common pattern of phonologization in sound change: an allophonic feature becomes phonemic when its conditioning environment is lost. Compare, for example, the well-attested change across languages of *cvn* > *c̄vn* > *c̄v̄*, leading to distinctive vowel nasalization.

All jers [ɪ, ʊ] disappeared from Russian by one of two means: either they were deleted, or they merged with [e, o]. Which fate befell a jer depended on whether it occupied a ‘strong’ or a ‘weak’ position. The generalization, known as Havlik’s Law, is that every odd-numbered jer deleted, counting leftwards from the end of the word; if a non-*jer* vowel intervened, the count would begin again. Some examples are given below (from Bethin 1998 and Kiparsky 1979).

- | | | | | |
|------|---------------------|---|--------------------|------------------|
| (15) | <i>fɪvɪtsʲɪ</i> | > | <i>fʲvʲetsʲ</i> | ‘tailor’ |
| | <i>fɪvɪtsʲa</i> | > | <i>fʲevʲtsʲa</i> | ‘tailor (gen.)’ |
| | <i>rʊpʊtʊ</i> | > | <i>rʊpʊt</i> | ‘murmur’ |
| | <i>rʊpʊtu</i> | > | <i>rʊpʊtu</i> | ‘murmur (dat.)’ |
| | <i>otʊxodʊmʲikʲ</i> | > | <i>otʊxodnʲikʲ</i> | ‘hermits (acc.)’ |

The alternating pattern strongly suggests the influence of some metrical organization. Bethin (1998) and Zec (to appear) both assume that jers deleted when they occupied the weak syllable of a trochee. A complicating factor is that ‘weak’ positions did not behave identically. The first jers to delete, judging by the historical records, were those in word-final position. This implies a hierarchy of weak (or strong) positions, of which word-final weak syllables were the ‘weakest of the weak’. As it happens, in the idealization assumed for our analysis, jers occur only word-finally. Our interest lies not in the details of the causes of jer deletion, but in the result that palatalization was phonologized, a result that held wherever jers deleted. For this purpose, word-final jers are fully representative.

Given the special status of the word-final jers, it will be useful to posit a constraint *JER]_{WD}, though this should ultimately be understood as an interaction of

more basic constraints. Before the deletion of jers this constraint must have been dominated by NOCODA, since it was assumed above that /mⁱl/ → [mⁱl^σ] (see Tableau 11). Once the jers were lost, the opposite ranking held, as shown below. Given richness of the base, inputs like /mⁱlⁱI/ remain possible and must now be ruled out by the grammar. Assuming that such inputs surfaced as shown, *JER]_{WD} also dominated MAX. Here and below I assume that the constraint PALATALIZE remains in force, that is, continues to dominate *C^j and IDENT(PAL). So consonants before front vowels will always be palatalized.

Tableau 12. No word-final jers: *JER]_{WD} » NOCODA, MAX

m ⁱ l ⁱ I	*JER] _{WD}	NOCODA	MAX
a. m ⁱ l ⁱ I	*!		
b. b. m ⁱ l ⁱ		*	*

Of more interest to us is the retention of the formerly allophonic palatalization, and the consequences of this. This is what led to the phonologization of palatalization, that is, rendered it phonemic. Had palatalization instead dropped along with the jers, then the contrast formerly maintained by those jers would have been lost as well: [mⁱlⁱ] would have merged with [mⁱl], and so on. This merger did not occur. Let us assume that this is due to IDENT(PAL), though see below on the possible relevance of *MERGE. This means that the former constraint ranking *C^j » IDENT(PAL) no longer held, as shown.

Tableau 13. Retention of palatalization: IDENT(PAL) » *C^j

m ⁱ l ⁱ	IDENT(PAL)	*C ^j
a. a. m ⁱ l ⁱ		**
b. m ⁱ l	*!	*

We have not yet brought to bear the results of the earlier DT analysis of palatalization. Consider once again the perceptual distinctiveness constraint, SPACE_{C-F2} ≥ 1/2. In the simplest case, where vowel context is irrelevant (see (6a) in §2.2), this constraint has the effect of forcing consonants to be velarized when in contrast with palatalized consonants. This is precisely the case here, where we are dealing with non-prevocalic consonants. In order for this velarization to occur, it must also be the case that SPACE outranks *C^y. Assuming this to be true, word-final consonants would have become velarized when palatalization became phonemic, rather than remain plain. In order to see this, we must return to considering sets of forms rather than forms in isolation. In Tableau 14, the input consists of the four

forms seen earlier (in (12)), just those that were possible outputs at the previous stage. Again, by richness of the base, these remain possible *inputs*. Word-final jers delete, and palatalization is retained, for reasons already given. Candidates 14b-c differ only in the new presence of word-final velarization: SPACE requires it.

Tableau 14. Non-prevocalic velarization: $SPACE_{C-F_2 \geq I/2} \gg *C^y$

	$m^j i l^j r_1$ $m^j i l^j \sigma_3$	$m^j i l^j r_2$ $m^j i l^j \sigma_4$	SPACE	$*C^y$
a.	$m^j i l^j_1$ $m^j i l^j_3$	$m^j i l^j_2$ $m^j i l^j_4$	*!*	
b.	$m^j i l^j_1$ $m^j i l^y_3$	$m^j i l^j_2$ $m^j i l^y_4$		**

There is strong reason to believe this is right. At a later period, a new sound change affected Russian, in which [e] backed to [o], e.g., [lʲet] > [lʲot] ‘flight’. This occurred only to [e] that was preceded by a palatalized consonant and followed by a non-palatalized one.¹⁰ Why should the change occur in the context C^j_C , but not in the context $C^j_C^j$? Following Andersen (1978), I assume that the following consonant must in fact have been velarized, and so it constituted a backing context.

The next section shows how this systemic approach to the facts, and especially the appeal to SPACE, makes possible a new explanation for the well-known /i/ → [ɨ] allophonic rule of Russian.

4.3. The reanalysis of [ɨ]

We saw in §1 that CSR maintains a palatalization contrast before the phoneme /i/, and that non-palatalized consonants are velarized in this context. The latter requirement is pervasive and regular in Russian. Whenever a non-palatalized consonant precedes /i/ within a phonological phrase, that consonant is velarized. This can be seen, for example, when a non-palatalized consonant precedes an /i/-initial word (16a), or an /i/-initial suffix (16b). As we saw in §3, the degree of perceptible velarization depends on context, so that before the vowel /a/ (also shown) these consonants would not be transcribed as velarized.

- (16) a. $k^y ivanu$ ‘to Ivan’ $k anato l^j iju$ ‘to Anatoly’
 $brat^y ivana$ ‘Ivan’s brother’ $brat anato l^j ija$ ‘Anatoly’s brother’
- $v^y ital^j iju$ ‘to Italy’ $v arm^j en^j iju$ ‘to Armenia’
 $nad^y ital^j ijej$ ‘above Italy’ $nad arm^j en^j ijej$ ‘above Armenia’

- b. *konʲi* ‘kitty (in a game, pl.)’ *kona* (gen.sg.)
 konʲi ‘horse (pl.)’ *konʲa* (gen.sg.)

As should be clear from earlier discussion, this alternation is usually understood differently in the literature on Russian. Specifically, /i/ is said to become [ɨ] after non-palatalized consonants. (See Padgett 2001 for extensive discussion.) That is, the alternation is taken to involve not the consonant but the following vowel: [ivan] ‘Ivan’ vs. [k ɨvanu] ‘to Ivan’, etc.

There was certainly a time in the history of Russian, or Slavic in any case, when the ‘ɨ’ transcribed by linguists (more commonly transcribed ‘y’) did indeed exist as a vowel. (17) illustrates the history of the contrast between this vowel and /i/ (see Shevelov 1965). ‘ɨ’ descends from Common Slavic /u:/, and the contrast illustrated in (17a) existed long before Slavic developed palatalization (whether contrastive or allophonic). This vowel is said to have lost its roundness, and perhaps fronted somewhat, at a later stage (17b). (At some point also the originally quantitative oppositions were reanalyzed as involving vowel quality.) This change was part of a chain shift that also included a shift of /au/ to /u/ (see Padgett to appear). At a later stage still, (17c), allophonic palatalization before front vowels arose. Most interesting is the shift from (17c) to (17d). It is widely held that when palatalization phonologized, with the loss of the jers, the opposition shown here was reanalyzed as involving not the vowels, but the consonants. Where once the distinction rested on the vowel backness, with palatalization a mere redundancy before /i/, now palatalization became the distinctive feature, and it was the vowel [ɨ] whose backness became redundant. This is indicated by the change in underlying representations. It is widely assumed also that the reanalysis had no phonetic consequences, but we return to this question below. The evidence for the reanalysis is clear: from this time on it was no longer consonants that alternated according to the vowel (front or back), but the reverse. For example, forms like [stol] leveled their stems so that all ended in a non-palatalized consonant. This caused /i/ of the nominative plural to back to [ɨ], as shown in (17e). Other nouns that retained palatalized stems have [i] for both nominative and accusative plural. (Compare the forms in (16b) again.) This historical reanalysis is the origin of the well-known allophonic rule of CSR assumed by most researchers.

- (17) a. *stolu:* ‘table (acc.pl.)’ *stoli:* (nom.pl.)
 b. *stolʲ* *stoli*
 c. *stolʲ* /stolʲ/ *stolʲi* /stoli/
 d. *stolʲ* /stoli/ *stolʲi* /stolʲi/
 e. *stolʲ* /stoli/ *stolʲ* /stoli/

Indeed, the loss of the jers essentially coincided with, or triggered, this alternation, where none had existed before. As back jers disappeared from word-final position, for example, scribes began using the symbol for ‘ɨ’ in place of that for

/i/ in a following word, as shown below. This is the origin of the alternations seen in (16a).

- (18) *ot* *im'i* > *ot* *im'en'i* 'on behalf of'
v *istoba* > *v* *zba* 'in the hut'

There is reason to believe that when the jers fell, 'i' was already [i] preceded by a velarized consonant, as the analysis above suggested for CSR. First, scholars have long entertained the possibility that 'i' had a diphthongized pronunciation well before the loss of the jers, in the Late Common Slavic period. Some transcriptions of 'i' in Slavic words borrowed into neighboring languages have 'ui' or 'oi' for this vowel (see Shevelov 1965). This is certainly suggestive of something resembling [ɥi]. But such transcriptions were not the rule, and digraphs sometimes represent not diphthongs but something perceived as intermediate between the symbols employed. We are on firmer ground inferring a pronunciation like [ɥi] for Old Russian after the loss of the jers. Sobolevskii (1907/1962:42-3) notes that in musical texts, a prolonged vowel was indicated by means of repeating the vowel letter for each note. In the case of 'i', however, the symbol for 'i' was prolonged by the *front* vowel symbol, e.g., '*edinom iisl'no*' for '*edinom isl'no*' (see also Kiparsky 1979:95). This is how 'i' is sung today, and it is far more consistent with the view that 'mi' is [m^ɥi] than it is with the view that it is [mɨ]. (Again, see Padgett 2001 for extensive discussion.) These musical transcriptions date from the twelfth century, at just the time of, or soon after, the loss of the jers.

I assume here the conservative view that 'i' was [i] before the loss of the jers, and was reanalyzed upon the loss of the jers as [i] with the preceding consonant velarized. (If 'i' was already pronounced something like [ɥi], then the reanalysis of [Cɥi] as [C^ɥi] would be even more straightforward.) How should we understand the reanalysis in terms of the DT account? Consider the contrast [Cⁱ] vs. [Cɨ] from the point of view of SPACE: how different are the consonantal F2 values of these two sequences? Abstracting away from effects of a consonant's major place of articulation, the comparison is between the F2 values of [j] and [ɨ]. This is because the release of [C^j] is palatalized, while that of [Cɨ] is coarticulated with the following [ɨ]. But this difference falls well below that necessary to satisfy SPACE_C. $F_2 \geq 1/2$, representing an F2 differential on a par with [ɤe] vs. [e] (see Figure 2) or worse. On the other hand, the contrast in consonantal F2 in the case of [Cⁱ] vs. [C^ɥi] satisfies this constraint easily, as we have seen.

Before we can continue, however, we must reconsider our definition of 'potential minimal pair' presupposed by the formulation of SPACE. The current formulation of SPACE in (6b) relies on a definition of minimal pair given in §2.2: two words, all but one of whose corresponding segments are identical. According to this definition, [mⁱ] and [mi] are a minimal pair, as are [m^ɨ] and [mɨ]. But [mⁱ] and [mɨ] are not. They differ not only in their consonants ([mⁱ] vs. [m]) but in their vowels ([i] vs. [ɨ]). In the same way, pairs such as [mi] vs. [m^u] are not a minimal pair, and they were

assumed to vacuously satisfy $\text{SPACE}_{\text{C-F}_2} \geq 1/2$ in the analyses above. This segment-based understanding of ‘minimal pair’ follows the traditional intuition, but it has undesirable consequences. The problem illustrated here is very general: any time two words differ in more than one segment, they will vacuously satisfy SPACE. This is a problem, because it is possible for a pair to differ in more than one segment, while *none* of those differences is perceptually good enough. This is precisely the case, I suggest, with [m^hi] vs. [m^hɨ]: not only is the perceptual distance at consonantal release small, but so is that of the vowels. Compare (6a) to (19), where the F2 spacing of high vowel systems is considered. In any system contrasting [i], [ɨ], and [u], clearly no vowel can occupy one-half or more of the total F2 range.

(19)	...i.... ...ɨ.... ...u....	Each segment gets 1/3 of the perceptual space
i.....u.....	Each segment gets 1/2 of the perceptual space
ɨ.....	Each segment gets 1/1 of the perceptual space

In order to address these facts, first, let us generalize $\text{SPACE}_{\text{C-F}_2} \geq 1/2$ to $\text{SPACE}_{\text{F}_2} \geq 1/2$, which applies over F2 generally, whether of consonantal release or of vowels. Suppose also that we follow Lindblom (1992) in taking the unit of comparison to be not segments, but CV demisyllables. (VC demisyllables should qualify also, but I ignore that fact here.) That is, let a minimal pair be any two words, all but one of whose corresponding *demisyllables* are identical. Then the effect of $\text{SPACE}_{\text{F}_2} \geq 1/2$ is to require minimal pairs (so defined) to differ by at least one-half of the total F2 range. Like Lindblom, I assume that demisyllables are compared at two points – at consonantal F2 and at the vowel target F2. To satisfy SPACE, demisyllables must differ sufficiently in *at least one* of these two places.¹¹

The tableau below repeats Tableau 14, but now compares the previously winning candidate 14b – here 15b – with another candidate in which [ɨ] has been altered to [i] and the preceding consonant velarized. Candidate 15c has more violations of *C^y (as well as *i, not shown) than any other candidate. But given the ranking $\text{SPACE} \geq 1/2 \gg *C^y$ already motivated, the reanalysis follows. With the revised formulation of SPACE, candidate 15b now has two SPACE violations, one for each pair differing only in [m^hi] vs. [m^hɨ]; 18c has none. (15a has four, because not only does [m^hi] vs. [m^hɨ] violate SPACE, but so does [l^h] vs. [l], recall. This implies that coda consonants, or perhaps the [VI] sequences, constitute ‘demisyllables’.)

The ‘backing of /i/ to [ɨ]’ shown in (18), reinterpreted as velarization of the previous consonant, is thus seen as a consequence of the emerging palatalization contrast.

Tableau 15. Reanalysis of *ĭ*

	$m^jil^jI_1$ $m^jil^jO_3$	$m^jil^jI_2$ $m^jil^jO_4$	SPACE	*C ^y
a.	$m^jil^j_1$ $m^jil^j_3$	$m^jil^j_2$ $m^jil^j_4$	*!***	
b.	$m^jil^j_1$ $m^jil^y_3$	$m^jil^j_2$ $m^jil^y_4$	*!*	**
c.	$m^jil^j_1$ $m^jil^y_3$	$m^yil^j_2$ $m^yil^y_4$		****

There is one final matter to address. Tableau 16 shows all of the constraints relevant to the distribution of backness (F2) contrasts, and several of the key candidates we have seen. The ranking *JER]_{WD} » NOCODA, MAX still ensures that word-final jers cannot surface, though this is not shown here. As can be seen, the constraint hierarchy does indeed prefer 16c to 16a-b. (This tableau also shows that SPACE dominates IDENT(PAL).) However, suppose we reconsider the question why palatalization did not simply disappear along with word-final jers. Candidate 16d represents this option. The word-final laterals of this candidate *vacuously* satisfy SPACE, because they are identical. This candidate therefore ties with the desired output on this constraint. (The desired, but losing candidate, is indicated with a frowning face.) It also ties on IDENT(PAL). Since it has fewer articulatory markedness violations, 16d emerges as optimal.

Tableau 16. Contrast versus neutralization

	$m^jil^jI_1$ $m^jil^jO_3$	$m^jil^jI_2$ $m^jil^jO_4$	SPACE	IDENT(PAL)	*C ^j	*C ^y
a.	$m^jil^j_1$ $m^jil^j_3$	$m^jil^j_2$ $m^jil^j_4$	*!***		****	
b.	$m^jil^j_1$ $m^jil^y_3$	$m^jil^j_2$ $m^jil^y_4$	*!*	**	****	**
☹ c.	$m^jil^j_1$ $m^jil^y_3$	$m^yil^j_2$ $m^yil^y_4$		****	**!***	**!***
d.	$m^jil_{1,3}$	$m^yil_{2,4}$		****	*	*

However, there are two straightforward ways to rule out 19d. One involves a plausible change in how IDENT(PAL) violations are counted. Suppose we decompose

this constraint into separate ones, relativized to the underlying value of [back] that must be preserved. Only 19d involves a loss of underlying palatalization, so that a constraint such as IDENT(-BK) would rule it out.¹² The other solution is to invoke *MERGE (see §2.3). As can be seen, 19c and 19d differ in another substantive way: though both involve changes in underlying feature values, only 19d involves neutralization of contrast. This is illustrated below, for just the two candidates of interest.

Tableau 17. Contrast versus neutralization

	$m^i i^j T_1$ $m^i i^j T_2$ $m^i i^j \sigma_3$ $m^i i^j \sigma_4$	*MERGE	IDENT(PAL)	*C ^j	*C ^y
☞ a.	$m^i i^j_1$ $m^y i^j_2$ $m^i i^y_3$ $m^y i^y_4$		****	***!***	***!***
b.	$m^i i^j_{1,3}$ $m^y i^j_{2,4}$	*!*	****	*	*

5. CONCLUSION

Optimality Theory, more than other theories of grammar, opens the door to functional accounts of sound change, and of phonology. This is because the theory resolves an apparent contradiction between the universality of functional constraints and the language-particularity of grammars: in OT, grammars are constructed out of *universal* constraints, yet these are violable and ranked in a *language-particular* way.

Dispersion Theory, in turn, differs from other functional theories in its ‘systemic’ orientation. First, it relies crucially on constraints that govern the perceptual distinctiveness of phonemic contrasts, the SPACE constraints. These compare *pairs* of forms. Second, it makes use of a very direct notion of neutralization avoidance, in the form of *MERGE. This refers to a *set* of input forms that may or may not merge in the input-output mapping. The main goal of this paper was to motivate the first of these, by demonstrating the importance of the perceptual distinctiveness of contrast to the well-known variation in Russian between [i] and ‘i’. It has long been understood that this variation is intimately related to the palatalization contrast. But the basis of this relationship has remained in important respects unclear. Why should /i/ maintain an allophone ‘i’ after non-palatalized consonants? Why do such allophonic rules – or *any* allophonic rules – exist at all? Here, extending to a new area ideas of Ní Chiosáin & Padgett (2001) and Padgett (2001), I offer a DT account of the historical rise of [C^yi] in Russian, one that derives [C^yi] from independent functional considerations, and in doing so better motivates the connection to palatalization. The account also makes clear what Russian has in common with Irish, Marshallese, and other languages having contrastive palatalization. No account of Russian can afford to ignore the larger cross-linguistic pattern.

There is certainly a good deal of work still to be done. Perhaps the most obvious need is for research on the best formulation of SPACE constraints, and on the best way of grounding them in phonetic fact. Still, I hope it is clear that a formal approach to these problems is both conceivable and worth exploring.

University of California, Santa Cruz

6. NOTES

¹ I am grateful to Paul Boersma, Dylan Herrick, Eric Holt, Donka Minkova, Nathan Sanders, Jennifer Smith, Caro Struijke, and two anonymous reviewers for feedback that improved this paper greatly. I would also like to thank the participants in my winter quarter UCSC seminar, where some of this work was first aired.

² One might wonder whether /ts/ vs. /tʃ/ should count as paired, and similarly for /ʃ/ vs. /ʃʲ/, or even /ʃ/, /ʒ/ vs. /ʃʲ/, /ʒʲ/, respectively. The term ‘paired’ traditionally involves more than surface contrast, including also various alternations between palatalized and non-palatalized consonants, a topic for another paper. We are on firmer ground in stating that velars are in fact paired in CSR, though with a limited distribution (see Padgett to appear and references therein).

³ The question of possible sources of grounding is an empirical one. Some have argued, for example, that processing limitations shape phonology (see Frisch 1996 for just one example).

⁴ Flemming’s original MINDIST formulation of these notions has some disadvantages. See Padgett 1997 and Boersma 1998.

⁵ There is another way of distinguishing 3c, d that is worth exploring. Suppose that a merged word like bʲV_{1,2} were to count not as one form but two for the purposes of markedness constraints. Then 3c, d would differ in markedness violations, one having two violations of C^j and one of C^y, the other the reverse.

⁶ Thanks to Aia Vladimirsky for agreeing to make recordings. For both this speaker and the author, three tokens of each CV sequence were recorded. These were digitized and analyzed using Praat software (Boersma & Weenik, available from <http://www.fon.hum.uva.nl/praat/>) using the burg algorithm.

⁷ The fact that [bu] versus [b^yu] is ‘more distinct’ than [bo] versus [b^yo] in the author’s tokens suggests either aberrant formant measurements, inconsistent pronunciations by the author, or both.

⁸ This implies that faithfulness would be sensitive to the same scales of perceptual similarity that SPACE constraints are, certainly a plausible idea. It is not pursued here, but see Padgett 2001b.

⁹ Some of the conclusions drawn in this section depend on the fact that non-palatalized consonants are velarized, and not *labio-velarized*, that is, not C^w. The latter means a lower F2 value than for a merely velarized consonant, and so better potential contrast with C^j. Were non-palatalized consonants labio-velarized, then the contrast [be] versus [b^we] might pass SPACE_{≥1/2}, unlike [be] versus [b^ye]. Similarly, [ba] versus [b^wa] might be good enough, unlike [ba] versus [b^ya].

¹⁰ [o] also replaced [e] word-finally after a palatalized consonant, but most researchers agree that these instances were not phonological, but involved cases of morphological analogy.

¹¹ This does not actually solve the problem of vacuous application noted above, but simply ‘promotes’ it from the segment to the demisyllable. Now, if a potential minimal pair differs in two or more demisyllables, they will therefore pass SPACE, even if neither difference is perceptually adequate. I leave this matter to later work.

¹² Yet another plausible way of reinterpreting IDENT(PAL) violations would be to consider in more detail the question of consonantal F2. Take the input /mʲilʊ/, which surfaces as [mʲilʲ] in 19c but as [mʲil] in 19d. Since /l/ of /mʲilʊ/ is coarticulated with a following [ʊ], it should have an F2 value very similar to that of [lʲ], and more different from that of plain word-final [l]. If IDENT were sensitive to such distinctions, then 19c could win.