Interface Constraints and Frequency in Russian Compound Stress

Maria Gouskova and Kevin Roon
New York University

This paper has two goals. The first is to describe the patterns of secondary stress assignment in Russian compounds. Russian lexical stress is famously complex, and secondary stress in compounds reveals previously unnoticed properties of the system. An understanding of compound stress may resolve some debates in the analyses of Russian stress. Our second goal is to contribute to the study of how frequency interacts with phonological markedness. There is an oft-noted correlation between high frequency and relative phonological unmarkedness (Martin 2007, Zipf 1949, and others). Russian presents a correlation of a different variety: phonological markedness signals morphological complexity. Specifically, secondary stress, which is an anomalous feature for Russian words, is more likely to occur on low-frequency words, and we argue that its placement encodes morphological complexity. Low frequency requires a more robust indication of morphological complexity. We analyze the interaction between frequency and morphological complexity in Russian compound stress in terms of constraint indexation in Optimality Theory (Prince and Smolensky 2004). An analysis of Russian requires that indexation be available for morphological interface constraints, not just for faithfulness constraints (see also Pater 2008, Flack 2007, Gouskova 2007).

The paper is organized as follows. Section 1 overviews the morphology of Russian compounds. §2.1 provides a bit of background on Russian primary stress. §2.2 lays out the patterns of secondary stress.

*For valuable feedback, we would like to thank Tuuli Adams, Adam Buchwald, Lisa Davidson, Amanda Dye, Alec Marantz, John Singler, Jason Shaw, the NYU Ph-group, two anonymous reviewers, and the audience of FASL 17.
in compounds that we found in our study. The phonological analysis is presented in §3. Finally, §4 concludes.

1 Morphology of compounds

1.1 Kinds of Russian compounds

Russian has three types of compounds: coordinating, truncated, and subordinating (Townsend 1975: 201-207, Molinsky 1973). Coordinating compounds consist of (at least) two whole words, with each bearing its own inflection: [gús-i-lébed-i] ‘geese (nom. pl.)’ and swans (nom. pl.).\(^1\) Stress in these appears on each constituent. Truncated compounds consist of at least two bases truncated from the right, typically to one closed syllable. A single inflection for the whole compound appears on the rightmost stem: [kol-xóz] ‘collective farm’ (from [kolektívnoje] ‘collective’ and [xozlajstvo] ‘farm’), the company name [vněj-pròm-tèx-obmén] (from [vnějnij] ‘external,’ [promíflennij] ‘industrial,’ [texnítjeskij] ‘technical,’ and [obmén] ‘exchange’). Stress in these appears on each stem, but in older, frequent compounds such as [kolxóz], there is only one stress, on the rightmost stem. In subordinating compounds, which are our primary focus, stems are combined with a theme vowel (orthographic -e- or -o-, similar to Greek (Nespór and Ralli 1996)): [oborôn-o-sposóbnost] ‘defense capability’ (from [oborón-a] ‘defense’ and [sposóbnost] ‘capability’). The morphological head is the rightmost stem, which also bears the inflection for the whole compound. The rightmost stem is always stressed, which we attribute to a requirement for morphological heads to be stressed (Revithiádou 1999:28). This requirement is never violated in compounds. The presence of secondary stress on the first stem depends on complex conditions (discussed in §2.2). The generalizations concerning stress and frequency seem to apply both to truncated and subordinating compounds, so our analysis accounts for both.

\(^1\)We use a fairly broad transcription: stress is transcribed throughout (primary as an acute [á], and secondary as a grave [à]), but we do not systematically mark vowel reduction, devoicing, or palatalization.
1.2 Morphological and prosodic structure

Our assumptions about the morphological and prosodic structure of Russian compounds are as follows. Morphosyntactically, coordinat-
ing compounds consist of separate syntactic words, and prosodically, they are concatenations of prosodic words (ω) into a phonological phrase (Φ), as in [(gusı)x{lèbedi}x]Φ. In subordinating and truncated compounds, on the other hand, the stems are combined into a single syntactic word, which contains at least two stems and a linker morpheme. Phonologically, therefore, these compounds constitute single, non-recursive prosodic words, though some of them may have more than one foot: the name of the film studio {(mòs)F_t (film)F_t}ω (cf. [moskvá] ‘Moscow’ and [film] ‘film’) and {go.lo(vo.lóm)F_tka}ω ‘puzzle’ (cf. [golová] ‘head’ and [lomát] ‘to break’).

(1) Morphological and prosodic structures for Russian truncated and subordinating compounds

<table>
<thead>
<tr>
<th>a. Truncated compounds</th>
<th>b. Subordinating compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;word&quot;</td>
<td>&quot;word&quot;</td>
</tr>
<tr>
<td>&quot;stem&quot;</td>
<td>&quot;stem&quot;</td>
</tr>
<tr>
<td>&quot;stem&quot;</td>
<td>&quot;stem&quot;</td>
</tr>
<tr>
<td>Ø_{linker}</td>
<td>themeV</td>
</tr>
<tr>
<td>&quot;stem&quot;</td>
<td>Ø_{linker}</td>
</tr>
<tr>
<td></td>
<td>ω</td>
</tr>
<tr>
<td>(F_t)</td>
<td>σ*</td>
</tr>
<tr>
<td></td>
<td>Ft+</td>
</tr>
<tr>
<td></td>
<td>σ*</td>
</tr>
</tbody>
</table>

We assume that the theme vowel -e/-o- forms a morphologi-
cal constituent with the left-hand stem. Phonologically, this vowel is clearly syllabified with the last consonant of the left-hand stem: root-final consonants retain a voicing contrast in left-hand compound stems (/golov-o-lom-k-a/→[golavo.lom.ka], not */golovo.o...| ‘puzzle’). Since Russian has devoicing at the ends of prosodic words, the consonant is not prosodic-word-final. The theme vowel is also not
prosodic-word-final based on reduction patterns. The morphological affiliation of the vowel is harder to determine: we are not aware of morphosyntactic evidence that points either way in Russian (though Krott et al. 2001 find that the left-hand stem has a greater effect than the right-hand stem on the choice of linking element in Dutch compounds). Some work on Greek linking vowels makes the same assumption, though others assume that the vowel is epenthetic and not morphological (see Ralli (2003) for an overview). The epenthetic analysis does not seem appropriate for Russian, since the theme vowel sometimes appears in hiatus contexts (see (2)).

With this background on the morphology of Russian compounds, we now move on to the phonology of stress.

2 Secondary stress in Russian compounds

2.1 Background on main stress placement

Russian stress is lexically contrastive, and its position cannot be predicted from the phonological shape of the word. It is also strongly culminative: in single-root words, there is only one main stress, regardless of the number of syllables: e.g., [vǐ-kristal-iz-ova-tj-sja] ‘to crystallize.’ Compounds present the only robust context for secondary stress. Main stress in compounds always falls on the last syllable of the right stem in pretonic position, but to [a] in pretonic position. The reduction pattern in pretonic position would indicate that the vowel is footed into an iamb with the following stressed syllable (Crosswhite 1999): if there were a prosodic word boundary separating the two syllables, we would expect the vowel to reduce to schwa. Alternatively, the pretonic vowel could have different quality due to tone spreading from the stressed syllable (Bethin 2006), but there is still no evidence that there is a strong prosodic boundary between the theme vowel and the following stem. In true word-final positions (e.g., oborôn[a] góroda ‘defense of the city’), vowel reduction does not seem to depend on where the stress falls in the following word, but this is something that should be investigated further.

The other context for secondary stress is certain foreign prefixes (súper-, psévdo-, óper-). We analyze these as lexical exceptions to the “one-stress-per-word” generalization (see §3): these are lexically accented prefixes whose accents cannot be deleted even if this means that the word ends up with two stresses. Alternatively, one could posit that these are stems (or roots) in their own right, as Peperkamp (1997) does for Italian. We would like to avoid this route, since
stem, and its position is determined by the accentual properties of the root and affixes (see Roon 2006). To understand how secondary stress is assigned, we have to present some background on main stress placement, since the lexical subclass of the first compound stem determines to some extent whether it will bear secondary stress.

As reported by Zaliznjak (1977), a majority (≈92%) of nominal stems in Russian have fixed stress on some syllable of the stem throughout the inflectional paradigm (Pattern A, [tetrád-l-Ø]~[tetrád-i] ‘notebook nom. sg.~gen/dat/loc. sg.’). About 6% of stems have stress on the inflectional suffix, and if there isn’t an overt suffix, on the last syllable of the stem (Pattern B, [fert-á]~[fert-Ø] ‘feature’ nom. sg.~gen. pl.). The remainder (about 2%) of the stems have mobile stress, which alternates between inflection stress and either initial (Pattern C, [kólokol]~[kolokol-á] ‘bell nom sg.~nom pl.’) or stem-final stress (Pattern D, [kolbas-á]~[kolbás-i] ‘sausage’ nom. sg.~nom. pl.).

In analyses of Russian stress (Melvold 1990, Idsardi 1992, Halle 1996, Halle and Vergnaud 1987, Halle 1973, Alderete 1999), three positions compete for default status: initial, post-stem, and final or desinence. There is no consensus in the literature as to the default (see Crosswhite et al. 2003)—all analyses have to appeal to lexical exceptions, suggesting that no one generalization can be made over the entire system. Regardless of what default is posited, every analysis treats stems with fixed stress on the 2nd or 3rd syllable as underlingly accented, so we will take this to be the strongest generalization emerging from the literature. We will also assume that all Pattern A stems have underlying stress, and that Patterns B, C, and D do not.

there is no evidence that these morphemes have root status—for one thing, they cannot head words of their own. Positing that they are stems based on stress alone amounts to circularity. Our analysis does not explain, however, why prefixes but not suffixes can bear secondary stress.

4In subsequent examples, we indicate the stress patterns of stems with subscript letters A-D.
2.2 Secondary stress

Existing descriptions of secondary stress in Russian compounds rely on the intuitions of individual native speakers, and since the patterns are variable and involve some optionality, the works do not always transcribe secondary stress consistently (Avanesov 1964, Yoo 1992, Kuznetsova 2006). We investigated them more systematically in a production study. Three native Russian speakers from Moscow read a list of 144 compounds. Each speaker read the list twice. The words were placed in the frame napíšano ______ píšť ráz ‘X is written five times’, chosen to avoid potential stress clash effects on the left-hand side. The words were transcribed for the presence of secondary stress by both authors, who consulted in cases of disagreement. The generalizations we extracted from the data are summarized below.

Normally, two requirements must be met for secondary stress to appear. First, the left-hand stem must have fixed stress (Pattern A, as described in §2.1). Second, secondary stress must be at least two unstressed syllables to the left of the primary stress. As shown in (2a), secondary stress does not surface if the syllables are too close to each other.

(2) Patterns of secondary stress in Russian compounds

a. No secondary stress: one syllable would separate stresses
kanat-o-xódets ‘tightrope walker’ kanát ‘tightrope’
ver-o-lómstvo ‘treachery’ věr ‘faith’

b. Secondary stress: two syllables separate stresses
věr-o-isprobědání je ‘denomination’ věr-a ‘faith’
oborón-o-sposôbnost ‘defense capability’ oborón-a ‘defense’
bômb-o-ubějîFFE ‘bomb shelter’ bômb-a ‘bomb’

Secondary stress normally does not appear on Pattern B and Pattern C stems even if there is enough room for two unstressed syllables to separate stresses:

(3) Pattern B and C stems do not have secondary stress

golov-o-krajûnîje ‘vertigo’ cf. golov-a ‘head’
ogn-e-tuťatel ‘fire extinguisher’ cf. ogón ‘fire’
korabl-e-krajûnîje ‘shipwreck’ cf. koráb ‘ship’
There are exceptions to the rhythmic generalization (as noted also by Yoo 1992, Avanesov 1964). In low-frequency words, secondary stress may appear even when there is only one syllable separating the stresses (see (4)). Crucially, in many of these low-frequency words, secondary stress surfaces in a position that does not correspond to an underlying accent. For example, none of the analyses of Russian stress assume that [jestestv-ó] ‘nature’ is underlingingly stressed on the second syllable.

(4) Low-frequency stems get secondary stress

jestěstv-o-vědenije ‘natural science’ cf. jestestv-ó B ‘nature’
kukurůz-o-vǒd ‘maize grower’ cf. kukurůz-a A ‘maize, corn’

Moreover, secondary stress may even surface in a syllable adjacent to primary stress: in compounds with vowelless (yer) stems, the theme vowel bears secondary stress (see (5)). Compounds with “linen” and “ice” tend to be infrequent words, so it is impossible to tell a priori whether these compounds have secondary stress because of low frequency or for another reason, for example because they contain relatively marked consonant clusters. We are currently investigating this question in a follow-up study.

(5) Yer stems get stress

líd-ò-bůr ‘ice breaker’ cf. líd B ‘ice’
lín-ò-zavód ‘linen factory’ cf. lín B ‘linen’

The effect of frequency on secondary stress realization is shown in Figure 1 for the 150 compounds we recorded. Each stimulus compound was classified according to its frequency in the Russian-language search engine Yandex (http://yandex.ru). High frequency words were the 35 most frequent stimuli, low frequency words were the 35 least frequent, and the rest were classified as middle frequency.

5Yandex includes inflected forms of each compound in the total number of hits, whereas Google treats case forms such as golovolomk-a and golovolomk-i as different words.
As shown in the graph, the patterns of secondary stress realization are more or less the same in the high and middle frequency compounds, but they are reversed in the low frequency compounds.

Figure 1: Effect of frequency on secondary stress realization

3 Analysis

We assume that by default, compounds have two prominences—one for each root-based stem. Rhythmic and faithfulness constraints may override this default, so not all compounds will surface with secondary stress. For low-frequency words, however, the requirement for each stem to have a prominence is ranked higher, so it overrides the rhythmic and faithfulness constraints.

This pattern arises through the interaction of the following constraints.\(^6\) In (6), we define an interface constraint $\text{STEM} \rightarrow \text{PROM}$, which requires each morphological stem to contain at least one segment that projects a phonological prominence (cf. Alderete’s 1999

\(^6\)We assume a bracketed grid representation for stress (Hayes 1995). We also assume that Headedness “a PrWd dominates a Foot” is not violated, so each word has to have at least one stress.
POST-STEM-PROM, also Revithiadou 1999). Since stems are often nested inside each other, the constraint must apply at the level of the maximal projection for each stem. This constraint conflicts with markedness constraints on rhythm (see (7)). These include a modified anti-clash constraint *STRONGCLASH (following Nespor and Vogel 1989; cf. *FtFt of Kager 1994) and some constraints whose interaction favors words with just one prominence, including ENDRULE-L. ENDRULE-R is never violated in Russian, since the main stress is always the rightmost and usually the only stress in the word.

(6) Morphology-phonology interface constraint

STEM→PROM (St→Pr): “For each stem, some segment affiliated with the stem projects a prominence on the grid.”

[One instantiation indexed to low-frequency words, the other applies to all]

(7) Prosodic markedness constraints

a. *STRONGCLASH (*S-CLASH): “assign a violation mark for every pair of adjacent columns of strong beats”

\[
\begin{array}{cccc}
\text{word-level beat} & \text{foot-level beat} & \text{syllable-level beat} \\
\text{x} & \text{x} & \text{x} \\
\text{x} & \text{x} & \text{x} & \text{x}
\end{array}
\]

b. ENDRULE-L (ER-L): “A word-level prominence is not preceded by another prominence at the word level.” (after Prince 1983; see also McCarthy 2003)

Finally, faithfulness is also active in the pattern. We adopt Alderete’s (1999) accentual faithfulness, defined informally as follows.

(8) Faithfulness to accent

MAX(Accent) “No deletion of accent”

DEP(Accent) “No insertion of accent”

NOFLOP(Accent) “No movement of accent”
We start with the phonology of normal compounds. These compounds are by default required to have two stresses—unlike non-compounds, which can only surface with one stress even if more than one morpheme is accented underlingly. This is because \textsc{EndRule-L} dominates MAX, requiring that the main stress be the only stress in non-compounds. \textsc{Stem→PROM} in turn dominates \textsc{EndRule-L}, and so two stresses surface in compounds:\footnote{We use comparative tableaux (Prince 2000). Users not familiar with this format should ignore “W” and “L.”}

Tableau 1: Compounds project two prominences, whereas non-compounds project one

<table>
<thead>
<tr>
<th>Compound</th>
<th>\textsc{St→Pr}</th>
<th>ER-L</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>/vi-, kristál, -iz, -ova, -t, -s(\text{v})/</td>
<td>a. vikristalizovat(\text{i})s(\text{v})</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>b. vikristalizovat(\text{i})s(\text{v})</td>
<td>*!W</td>
<td>L</td>
</tr>
<tr>
<td>/rabót, -o-, saposob-, sost(\text{i})/</td>
<td>c. rabót-o-saposobnost(\text{i})</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. rabot-o-saposobnost(\text{i})</td>
<td>*!W</td>
<td>L</td>
</tr>
</tbody>
</table>

\textsc{St→Pr} will be violated when the underlying position of the stress on the left-hand stem is too close to the main stress. This would violate *\textsc{Strong-Clash}, so stress must be deleted in such words:

Tableau 2: Compounds normally do not have stress clashes; underlying accent is deleted to avoid clash

<table>
<thead>
<tr>
<th>Compound</th>
<th>*\textsc{S-Clash}</th>
<th>\textsc{St→Pr}</th>
<th>ER-L</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>/rabót, -o-, dat-, el(\text{v})/</td>
<td>a. rabót-o-dátel(\text{i})</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. rabót-o-dátel(\text{i})</td>
<td>*!W</td>
<td>L</td>
<td>*W</td>
</tr>
</tbody>
</table>

An underlying stress could in principle be realized somewhere other than its underlying location, but this option is ruled out by an undominated No\textsc{FloP(Accent)}. Thus far, we’ve accounted for Pattern A stems, which we assume have underlying stress. For roots that lack underlying stress, \textsc{St→Pr}
cannot be satisfied by inserting stress. This suggests that Dep(Acc) dominates ST→Pr:

Tableau 3: Stress cannot be inserted on underlyingly unaccented stems

<table>
<thead>
<tr>
<th>/golov_G, -o-, kruž-, čenij/</th>
<th>Dep</th>
<th>ST→Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. golov-o-kruženíje</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. golőv-o-kruženíje</td>
<td>*!W</td>
<td>L</td>
</tr>
</tbody>
</table>

We now turn to low-frequency compounds, which satisfy ST→Pr for each stem even if it means inserting stress and violating rhythm. ST→Pr is doubly instantiated in the hierarchy, and the higher-ranked indexed Stem→PromL applies to low-frequency compounds. This constraint is ranked above Dep(Accent), so an accent must be inserted even if one isn’t present underlyingly:

Tableau 4: Low-frequency stems have prominence, even if it must be inserted

<table>
<thead>
<tr>
<th>/jestežtv_B, -o-, ispitanije_L</th>
<th>ST→PR_L</th>
<th>Dep</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. jestežtv-o-ispitanije</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. jestežtv-o-ispitanije</td>
<td>*!W</td>
<td>L</td>
</tr>
</tbody>
</table>

Under this analysis, even stems containing vowelless roots should have stress, which is placed on the only available syllable: the one with the theme vowel as its nucleus (recall from §1 that we take the theme vowel to be part of the first stem). This placement of accent violates both Dep(Accent) and *Strong-Clash:

Tableau 5: Even the theme vowel may be stressed in low-frequency compounds

<table>
<thead>
<tr>
<th>/pď_B, -o-, bur_LEX</th>
<th>ST→PR_L</th>
<th>Dep</th>
<th>*S-Clash</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. pď-o-bůr</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. Pd-o-bůr</td>
<td>*!W</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>
Forms such as [lʲdębûr] and [jestěstvoispitáñije] present essential evidence for our argument that this pattern is driven by a morphophonological interface constraint rather than by indexed faithfulness. Since these forms lack stress underlingly, their stress patterns cannot be due to the promotion of \textsc{max(ACC)} to the top of the hierarchy. These forms violate faithfulness in order to satisfy the interface constraint.

Although we have been talking about this pattern in terms of indexation to frequency, we believe this is a proxy for a more abstract distinction. The grammar provides two different instantiations of the constraint in the hierarchy, but whether the relevant property is low frequency or formal register may be determined outside the grammar proper. It may even be that the indexation is quite arbitrary. This would explain forms such as [zêmîl-e-délets] ‘farmer’ (from [zeml]-áč ‘earth’ and del- ‘to do, make’), which unexpectedly surface with secondary stress in violation of both \textsc{dep(ACC)} and \textsc{*strong-clash}. These pattern with low-frequency compounds—an option made available by generic indexation. The prediction of this analysis is that accent can be inserted on such stems, but it will not be deleted on Pattern A stems.

Finally, our analysis has nothing to say about the location of inserted secondary stress. Why, for example, is [jestěstv-o-ispitáñije] stressed on the second syllable and not on the first? There are many possible explanations for this, which we cannot treat fully here, but we mention a few. One possibility is that the same principles are at work here as elsewhere in the language: in the genitive plural and in derived affixed forms (e.g., [jestěstvenno] ‘naturally’), the stress in this stem is on the last syllable, just as in the compound. Another possibility is that stress placement is determined by some related output form, which serves as a transderivational correspondence base for the compound (Benua 1997). This seems initially plausible for some forms, but even a cursory look at the left-hand stems suggests that the choice of base is not a simple matter. It is also possible that some of the mobile stress stems (Patterns B, C, and D) actually have underlying stress, which the grammar treats differently from Pattern A stress.
4 Conclusion

Russian compound stress is sensitive to two factors. First, a left constituent will surface with secondary stress if it is underlyingly accented and secondary stress does not create a stress clash. Second, low-frequency compounds are more likely to surface with secondary stress than higher-frequency compounds. We have accounted for this by proposing a morpho-phonological constraint requiring each morphological stem to project a prominence on the metrical grid. This constraint is indexed to low-frequency compounds. Its ranking above rhythmic and faithfulness constraints requires low-frequency compounds to have secondary stress even if they are underlyingly unaccented or if there is a stress clash. Secondary stress thus encodes morphological complexity in Russian compounds.

References


Pater, Joe. 2008. Morpheme-Specific Phonology: Constraint Indexa-


