Vowel Dispersion and Kazakh Labial Harmony
Adam McCollum
University of California, San Diego

Abstract
This paper presents novel data on labial harmony in Kazakh (Turkic) that is typologically anomalous under Kaun’s analysis of labial harmony. I argue herein that the data receives a principled explanation if perceptual weakness, which is argued to motivate labial harmony, is redefined in terms of systemic relations couched in Dispersion theory rather than feature combinations. Using this reconception of perceptual weakness, I develop an analysis of the Kazakh pattern in a probabilistic Harmonic Grammar, using scalar and conditional constraints that directly reference perceptual distances calculated over the Kazakh vowel space. The dispersionary analysis presented herein is shown to fit the Kazakh data better than a feature-based analysis. Furthermore, the dispersionary analysis is shown to make the correct predictions for four languages with harmony patterns drastically different from the Kazakh pattern.

1 Introduction

Cross-linguistically, labial, or rounding harmony is subject to a number of restrictions (Steriade 1981; Cole & Trigo 1988; Kaun 1995; Walker 2001; van der Hulst & Moskal 2013). First, labial harmony is parasitic on another harmony or feature agreement pattern. Second, the class of harmonic vowels is restricted by height and/or palatality to a subset of round vowels. Kaun (1995, 2004) contends that this subset of round vowels is definable in terms of perceptual salience, and that labial harmony operates on perceptually weak vowels, extending the temporal span of less salient contrasts. Thus, vowels whose [RD] feature is only weakly realized are more likely to trigger harmony. In addition, Kaun notes that languages asymmetrically produce strings of height-uniform round vowels. In this way, harmony not only is initiated by more similar vowels, it also preferentially produces strings of similar, often identical, vowels. Both of these issues relate back to the observation that more similar segments are more likely to interact in harmony (Ultan 1973:53; Rose & Walker 2004: §2.2; Wayment 2009).

In this article I present new data from Kazakh, a Turkic language of Central Asia. Crucially, contemporary Kazakh exhibits trigger preferences that do not conform to Kaun’s typology. To account for the Kazakh pattern I redefine perceptual similarity in dispersionary (Liljencrants & Lindblom 1972; Lindblom 1986), rather than featural terms. This reconception of perceptual similarity is elegantly formalized via constraints that directly reference perceptual distances between harmonic pairs. Additionally, the Kazakh pattern is gradient, and to account for this gradience the analysis is couched in a Maximum Entropy Harmonic Grammar (Smolensky 1986; Goldwater & Johnson 2003) with scalar and conditional constraints. The analysis developed herein is further shown to make the correct predictions for four languages with harmony pattern decidedly
different from Kazakh. This paper thus redefines the typology of labial harmony in dispersionary terms, offering a more empirically accurate and testable means for evaluating perceptual similarity in labial harmony.

The paper is organized as follows. In §2 I describe vowel harmony in contemporary Kazakh, noting in particular trigger strength differences, as well as asymmetrical application within- and across-morpheme boundaries. In §3 I discuss previous analyses of labial harmony, in particular Kaun’s typological study (1995, 2004). In §4 perceptual similarity is operationalized along dispersion-theoretic lines, and in §5 these distances are used to construct a formal analysis of the pattern in a probabilistic grammar. §6 outlines the various predictions made by the present proposal, and relates this analysis to other constraint-based dispersion-theoretic work (e.g. Flemming 2002; Padgett 2004), and Kaun’s feature-based analysis of harmony. In §7 I offer concluding remarks and potential avenues for further investigation.

2 The vowel harmony pattern

In this section I describe labial harmony in contemporary colloquial Kazakh, focusing on the various restrictions on its application. I first provide background information about the language and its vowel inventory in §2.1. Then in §2.2 I describe the application of palatal and labial harmony in colloquial Kazakh.

2.1 Background and vowel inventory

Kazakh is a Turkic language of the Kipchak branch (Kirchner 1998), with over ten million speakers in Central Asia, primarily in the Republic of Kazakhstan. Muhamedowa (2015:xix-xx) and Kara (2002:4) report three dialects of Kazakh, but note few differences between them (Grenoble 2003:150). Most studies report palatal, or backness harmony in Kazakh, a property shared with the majority of Turkic languages (Menges 1947; Korn 1969; cf. Vajda 1994). In addition, Kazakh employs labial harmony, which is the focus of this paper. As affixes are concatenated via suffixation, these two harmony processes extend rightward from root to suffixes.

Writers have disagreed on the number of vowel phonemes in the language, with as few as five, and as many as eleven vowels reported (Menges 1947; Dzhunisbekov 1972; Kirchner 1998; Kara 2002; Yessenbayev et al. 2012; Sharipbay 2013; McCollum 2015; Muhamedowa 2015). Writers agree though on a common set of eight vowels, /ə ɔ ʊ ɐ ɪ ʏ/. Three additional vowels might be phonemic in the modern language, /i æ u/. As these three vowels are lexically rare, and phonologically peripheral to the general harmony system (Dzhunisbekov 1972:39-41; Bowman & Lokshin 2014:2; Kirchner 1998:322), they will be excluded from further discussion.

2.2 Vowel harmony

In this section I describe the palatal and labial harmony patterns in Kazakh, as attested during fieldwork in southeastern Kazakhstan with eleven native speakers, (mean age 33.5 yrs, range 19-46 yrs). Data was collected via semi-formal, conversationally-based elicitation in the target language. In §2.2.1 I show that palatal harmony is typically word-delimited while in §2.2.2 I demonstrate that labial harmony is subject to morphological as well as inventory-related restrictions.
on both triggers and targets. These three classes of restrictions will interact significantly in the analysis presented in §4.

2.2.1 Palatal harmony
Kazakh words usually contain vowels from only one palatal set. As palatality is determined by the root, (i.e. the word-initial syllable), and labiality is at least partially dependent on the root, height (high versus non-high) is the only aspect of non-initial vowels that does not alternate. In (1-2) below, both root-internal and suffix instantiations of palatal harmony are presented under the assumption that non-initial root vowels are, like suffixes, targets for harmony (Harrison & Kaun 2000; Kabak & Weber 2013; cf. Clements & Sezer 1982).

In (1) all possible harmonic combinations of vowel height are exemplified within roots. Words from multiple syntactic classes are shown below, specifically nouns (1a,c,d,f,g), adjectives (1e,h), and a verb (1b). Harmony among front vowels is shown in (1a-d), and harmony among back vowels is exemplified in (1e-h).

(1) Root-internal harmony
a. tizim ‘list’          e. quzuʃ ‘red’
 b. turke- ‘register (V.)’ f. quran ‘hawk’
c. esik ‘door’           g. qazu ‘horse sausage’
d. ʒebe ‘arrow’          h. qara ‘black’

In (2) all possible harmonic combinations of vowel height between roots and suffixes are shown. In (2a-d) the locative suffix is realized with a non-high vowel that agrees with the root in palatality. In (2h-j) the high vowel of the accusative suffix also agrees with the root vowel in palatalit. Furthermore, iterative palatal harmony is exemplified in (2e-g). Generally, palatal harmony is coextensive with the word.

(2) Suffix harmony
a. birz-de ‘1P-LOC’       h. birz-di ‘1P-ACC’
b. kez-de ‘time-LOC’       i. kez-di ‘time-ACC’
c. quz-da ‘girl-LOC’      j. quz-du ‘girl-ACC’
d. qaz-da ‘goose-LOC’     k. qaz-du ‘goose-ACC’
e. tizim-de-gi-ler-i-м-iz-de ‘list-LOC-REL-PL-POSS-1-PL-LOC’
f. turke-l-di-ŋ-iz-der ‘register-PASS-PST-2-FORM-PL’
g. quran-dar-u-м-uet-da ‘hawk-PL-POSS-1-PL-LOC’

In summary, palatal harmony in Kazakh is iterative (2e-g), not sensitive to height restrictions, and generally word-bounded. Furthermore, all syntactic classes undergo harmony (1-2). To reiterate, I will assume throughout, following Harrison & Kaun (2000), that all non-initial vowels are targeted by harmony, both root-internal and suffix. Palatal harmony is not subject to the kinds of restrictions shown for labial harmony, as will be described below. It is more pervasive synchronically and diachronically (Johanson 1998a:32-33; Róna-Tas 1998:72-73; Erdal 1998:138-140).

1 Exceptions to this generalization include compounds, unassimilated foreign loans, and words containing invariant morphemes, like the instrumental-comitative suffix /Men/.
2.2.2 Labial harmony

Labial harmony is robustly reported in Menges (1947:59-64) and Korn (1969:101-102), but more recent writers have noted increasing restrictions on this particular harmony pattern (Dzhunisbekov 1972; Kirchner 1998; Kara 2002; McCollum 2015) in contrast to the palatal pattern observed above. Concretely, labial harmony in contemporary Kazakh is confined to a smaller domain of application (Balakaev 1962; Dzhunisbekov 1972; Vajda 1994; Kirchner 1998; McCollum 2015), and harmony asymmetrically occurs more often within roots than across morpheme boundaries (Vajda 1994; McCollum 2015). These restrictions are demonstrated in (3-5) below.

Within roots several generalizations emerge from the data in (3). First, high vowels are far more likely to undergo harmony than are non-high vowels (3a-d). Second, harmony is generally triggered by the front vowels, [ʏ ø], as well as the back vowel [ʊ], but not by [ɔ] (3d). The frequency data reported below was determined impressionistically after initial data collection. These judgements were then confirmed by comparing acoustic properties of non-initial vowels to those of underlying (i.e. initial) vowels (Zsiga 1997:234-235).

(3) Root-internal harmony

<table>
<thead>
<tr>
<th>Attested variants</th>
<th>Frequency of harmony</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ʒʏzɡ ~ ʒyzk</td>
<td>(10/11)</td>
<td>'ring'</td>
</tr>
<tr>
<td>b. kɔsɔk ~ kɔstɔk</td>
<td>(10/11)</td>
<td>'desert carrot'</td>
</tr>
<tr>
<td>c. qɔłɔn ~ qɔłun</td>
<td>(16/17)</td>
<td>'colt'</td>
</tr>
<tr>
<td>d. qɔzʊ ~ qozʊ</td>
<td>(1/10)</td>
<td>'lamb'</td>
</tr>
<tr>
<td>e. tɔlɔk ~ tylek</td>
<td>(1/11)</td>
<td>'graduate'</td>
</tr>
<tr>
<td>f. tɔbe</td>
<td>(0/11)</td>
<td>'hill'</td>
</tr>
<tr>
<td>g. qɔłaq</td>
<td>(0/10)</td>
<td>'ear'</td>
</tr>
<tr>
<td>h. bɔłat</td>
<td>(0/10)</td>
<td>'steel'</td>
</tr>
</tbody>
</table>

The inertness of [ɔ] as a trigger for harmony is evident in Figure 1 below. Figure 1 plots the normalized F1-F2 of the root-internal high vowels in [qazuz] 'horse sausage' and [qozut] 'lamb.' If we assume that the acoustic properties (e.g. F1-F2) of vowels derived from harmony approximate those of underlying vowels (e.g. Zsiga 1997:234-235), then in [qazuz] labial harmony should produce vowels more similar to [ʊ] than [u]. However, in almost every case F1-F2 of high back vowels after [ə] and [ɛ] more closely resemble initial [u]. Note that the second syllable vowel after [ɛ] approximates F1-F2 of initial [u] only once, but approximates initial [u] in most instances. Also, while lip rounding depresses both F2 and F3 (Ladefoged 2001:41,46), in Kazakh F3 is not a reliable predictor of rounding (McCollum 2015).

2 Also confer Abuov (1994:41, 44), who reports labial harmony identical to that reported in Menges’ and Korn’s data. Abuov explicitly focuses on literary Kazakh, whereas Menges and Korn do not identify the register described. Curiously, Poppe (1965:183) reports no labial harmony in Kazakh.
Figure 1: Plot of non-initial vowel F1-F2 with mean [ɯ] and [ʊ] in initial. Solid ellipses show 1 SD around mean formant values for initial [ɯ] and [ʊ] (N=365), and + and ⋄ represent tokens of the second syllable vowels in [qɔzu] and [qazu] (N=24).

The third generalization to take from the examples above is that labial consonants do not trigger or spread harmony, which is demonstrated in examples like [tizim] ‘list’ in (1a), [tizim-de-gt-ler-\textunderscore m-iz-de] ‘list-LOC-REL-PL-1.POSS-PL-LOC’ in (1e), and [tobe] ‘hill’ in (2d).

Compared to (3), labial harmony is even more restricted in (4). In no case does harmony assimilate a vowel across a suffix boundary with 100% frequency, regardless of root vowel quality, suffix height, or morphosyntactic category. Specifically, the high vowel of the past tense suffix in (4a-d) and the non-high vowel of the conditional suffix in (4e-h) typically surface as unrounded despite the roundness of the initial vowel. Longer words are shown in (4i-k), where the same generalizations hold: outside of roots, harmony is infrequent, even after [ɣ], as in (4i-j).

(4) Suffix harmony

<table>
<thead>
<tr>
<th>Attested variants</th>
<th>Freq. of harmony</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. kyl-dv ~ kyl-di</td>
<td>(7/22)</td>
<td>‘laugh-PST.3’</td>
</tr>
<tr>
<td>b. øl-dv ~ øl-di</td>
<td>(1/20)</td>
<td>‘die-PST.3’</td>
</tr>
<tr>
<td>c. qor-du ~ qor-du</td>
<td>(1/18)</td>
<td>‘construct-PST.3’</td>
</tr>
<tr>
<td>d. qps-to ~ qps-tu</td>
<td>(1/20)</td>
<td>‘add-PST.3’</td>
</tr>
<tr>
<td>e. kyl-se</td>
<td>(0/15)</td>
<td>‘laugh-COND’</td>
</tr>
<tr>
<td>f. øl-se</td>
<td>(0/14)</td>
<td>‘die-COND’</td>
</tr>
<tr>
<td>g. qor-sa</td>
<td>(0/13)</td>
<td>‘construct-COND’</td>
</tr>
<tr>
<td>h. qps-sa</td>
<td>(0/13)</td>
<td>‘add-COND’</td>
</tr>
<tr>
<td>i. ʒyžk-tv ~ ʒyžk-ti</td>
<td>(1/9)</td>
<td>‘ring-ACC’</td>
</tr>
<tr>
<td>j. ʒyžk-ter-di</td>
<td>(0/23)</td>
<td>‘ring-PL-ACC’</td>
</tr>
</tbody>
</table>

Based on the frequency data in (3) and (4), along with that in Figure 1, it is clear that [ə] does not generally trigger harmony while [ɣ], [ʊ], and [ø] do to varying degrees. Two additional facts
are significant. First, both palatal and labial harmony are strictly local, affecting all segments within the domain of harmony. More concretely, each consonant is phonetically realized according to the backness and roundness of the word-initial vowel (Dzhunisbekov 1980, 1991:83-94; McCollum 2015:342-343). For instance, McCollum (2015:342-343) shows that after round vowels spectral energy of the fricative [s] is significantly lowered, an acoustic correlate of lip rounding (see Ní Chiosáin & Padgett 2001:125-126 for Turkish). Second, rounding of suffixes in (4) does occasionally occur. The rounding of suffixes is construed as coarticulatory due to its infrequency and variability. When rounding obtains, it almost always targets the first suffix only. In connected speech there are occasional instances of rounding on both first and second suffix vowels if both are high. In contrast, the rounding of root-internal high vowels after [γ o ø], as in (3), is regarded as phonological, being almost exceptionless.

Across morpheme boundaries, harmony is more frequently attested in a class of suffixes like the converbial and passive-reflexive suffixes, which are composed of a single underlying consonant. When these morphemes are concatenated to consonant-final roots, epenthesis repairs the illegal consonant cluster. The realization of the converbial suffix is shown below in (5). The converbial suffix consists of a voiceless bilabial stop, as in (5a-b), but a preceding epenthetic vowel occurs after consonant-final roots, shown in (5c-h). The epenthetic vowel is always high, and agrees with the root in palatality, and optionally in rounding (5e-g).

(5) Converbial suffix

<table>
<thead>
<tr>
<th>Attested variants</th>
<th>Frequency of harmony</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. qara-p</td>
<td></td>
<td>‘look-CVB’</td>
</tr>
<tr>
<td>b. soyle-p</td>
<td></td>
<td>‘speak-CVB’</td>
</tr>
<tr>
<td>c. qat-up</td>
<td></td>
<td>‘harden-CVB’</td>
</tr>
<tr>
<td>d. kes-ip</td>
<td></td>
<td>‘cut-CVB’</td>
</tr>
<tr>
<td>e. kyl-yp ~ kyl-ip</td>
<td>(26/31)</td>
<td>‘laugh-CVB’</td>
</tr>
<tr>
<td>f. qor-op ~ qor-up</td>
<td>(13/18)</td>
<td>‘construct-CVB’</td>
</tr>
<tr>
<td>g. øl-yp ~ øl-ip</td>
<td>(10/18)</td>
<td>‘die-CVB’</td>
</tr>
<tr>
<td>h. qps-op ~ qps-up</td>
<td>(1/15)</td>
<td>‘add-CVB’</td>
</tr>
</tbody>
</table>

The frequency of attestation is recorded for the optional cases of harmony in (5e-h). After [γ], [ø], and [ø], the epenthetic vowel is rounded 84%, 72%, 54%, and 7% of the time, respectively.

In summary, labial harmony in colloquial Kazakh applies to high vowels within roots, except after [a]. In suffixes harmony is no longer categorical, but applies with varying frequency in second (and rarely third) syllable suffix vowels. In addition, root vowels exert varying degrees of assimilatory force, with [γ] triggering harmony more often than [ø], [ø], or [a]. From the root-internal and suffix data presented above the following hierarchy of trigger strength emerges: γ > ø > ø > a. Broadly speaking, high vowels are better triggers than non-high vowels, and front vowels are better triggers than back vowels. I will demonstrate in §4-5 that the trigger strength hierarchy corresponds to differences in perceptual distance between the triggering vowel and its unrounded harmonic counterpart. Thus, we will eventually collapse the distinctions in height and backness just noted to distinctions in perceptual distance, where triggers that are more similar to their harmonic counterpart serve as better triggers for harmony. In the following section I relate these findings to the literature on labial harmony, particularly asymmetrical trigger strength relations previously reported in Kaun (1995, 2004).
3 Background literature

In §3.1 I survey previous literature on labial harmony, noting its parasitic nature and the role of phonetic grounding presented in previous work. In §3.2 I detail Kaun’s typological observations, discussing their relevance for the Kazakh data presented above.

3.1 Labial harmony

Both formal (e.g. Krämer 2003:120-133; Nevins 2010:30-31; Ko 2012:ch. 5) and functional analyses (Kaun 1995, 2004; Walker 2001) of labial harmony have been provided in the literature. Most analyses have, interestingly, converged on two separate generalizations. First, previous work has often noted that labial harmony is parasitic, in that its application depends on the agreement of some other feature(s), in many cases, vowel height (Steriade 1981; Mester 1988; Cole & Trigo 1988; van der Hulst & Smith 1988; Cole & Kisseberth 1994: §5; Finley 2008b:§9; Jurgec 2011:ch. 8). For instance, in Hixkaryana (Derbyshire 1979) labial harmony applies only when trigger and target are high. In Galab (Sasse 1974), prefixes undergo harmony only when trigger and target are non-high. The requirement that trigger and target agree in height is often noted for Yawelmani (Newman 1944; Kuroda 1967; Archangeli 1984; Mester 1988). Additionally, in the vast majority of cases, labial harmony operates in languages with an additional harmony pattern (Ultan 1973:55; van der Hulst & van der Weijer 1995:534). Even in languages with another harmony pattern these height restrictions are well attested. For instance, in Kachin Khakass (Korn 1969), palatal harmony is pervasive but labial harmony applies only when trigger and target are high. Furthermore, in Tafi (Bobuafor 2013), regressive ATR harmony affects all prefixes regardless of height, but progressive labial harmony targets and is triggered by non-high vowels only. Thus, even in languages where labial harmony co-exists with another harmony pattern additional conditions are often necessary for its application. Van der Hulst & Moskal (2012:48-29) suggest that labial harmony simply cannot exist without another type of feature agreement, vowel harmony or otherwise (cf. Li 1996:§5.6).

Second, phonetic grounding has often been used in formal analyses to account for trigger and target asymmetries in labial harmony. Kaun (1995:1) argues that “bad vowels spread”, suggesting a strong relationship between a vowel’s perceptual weakness and its propensity to trigger harmony. Kaun argues that a vowel’s perceptual weakness depends on the degree of lip rounding necessary for its articulation and the amount of acoustic modulation accomplished thereby. As for capturing target asymmetries, Kirchner’s (1993:8) co-occurrence constraint *[+round][−high], adopted in Kaun (1995:§5.5), is argued to capture the typological infrequency of round vowels lower in the vowel space, as well as the diminished degree of lip rounding possible for lower vowels. In Kaun’s analysis, this portrays non-high vowels as poor targets for harmony. Thus, a variety of analyses have connected the typological patterns found for labial harmony with the phonetic properties of round vowels. It should be noted though that while many analyses of labial harmony have implemented functional constraints, others have used formal mechanisms without significant reference to functional grounding (Cole & Trigo 1988; Nevins 2010; Ko 2012; Moskal 2012).

In the above cases the additional restriction on harmony is putatively featural, but Vaux (1993) suggests that labial harmony may also be constrained by systemic factors. Analyzing labialization of non-high vowels in Turkic, Vaux contends that in languages with three distinct vowel heights harmony produces non-high round vowels, but does not in systems with only two phonological height distinctions. In short, the harmonic pattern is, in part, derivable from the shape and size of a given language inventory. The use of systemic considerations will play a crucial role in the analysis presented in §4.
3.2 Typological generalizations

As noted above, Kaun argues that perceptually weaker vowels are better triggers, and conversely, perceptually salient vowels are poorer harmony triggers. She contends that harmony serves to extend the temporal span of a weak contrast to improve chances of reliable discrimination. Following Suomi (1983), she links vowel harmony with perception, as well as articulation, using several cross-linguistic phonetic studies of lip rounding (i.e. Terbeek 1977; Linker 1982; Boyce 1990) to inform her functional analysis of labial harmony. Specifically, she proposes the following five generalizations to account for the harmony patterns reported among the world’s languages.

(8) Kaun’s typological generalizations (2004:92)
1. Non-high vowels are better triggers than high vowels
2. Front vowels are better triggers and targets than back vowels
3. High vowels are better targets than non-high vowels
4. Cross-height harmony is dispreferred
5. Short vowels are better triggers than long vowels

I will discuss the above generalizations in order, suggesting possible re-interpretations where relevant. I will also relate each of these generalizations back to the Kazakh data presented in §2.2.

Addressing Kaun’s first two generalizations, she reports that in Linker’s (1982) cross-linguistic study of lip rounding high vowels display more lip rounding than non-high vowels. One thing to note is that the difference in lip rounding between high and non-high vowels varied significantly between languages. For instance, the difference between [y] and [ø] was minimal in Swedish and French, but more robust in in Finnish. Among the back vowels the difference between [u] and [o] was larger in Finnish and French, but smaller in Swedish. While in the five languages studied high vowels exhibited more lip rounding, it is also evident that the particular lip rounding gestures differed significantly between languages. As for perception, Kaun cites two findings from Terbeek (1977). First, Terbeek finds that high vowels were perceived as more rounded than non-high vowels. Second, he reports that back vowels were judged to be more rounded than front vowels, relating to Kaun’s second generalization above. Kaun uses these findings to argue that where the articulatory lip rounding gesture is smaller in magnitude, its acoustic consequences will be more subtle, leading to greater perceptual difficulty in discrimination. In sum, because non-high and front vowels are produced with less lip rounding, they are perceptually less salient than high and back vowels, making them better triggers of labial harmony.

As for the third generalization above, Kaun contends that while vowel trigger preferences depend on intrinsic perceptual weakness, target preferences derive from perceptual salience. Since high rounded vowels involve more lip rounding (Linker 1982), and high vowels are perceived as more rounded (Terbeek 1977), Kaun argues that high vowels better cue the roundness of the initial vowel. This reasoning is used to explain why labial harmony in Turkish affects high vowels only.

Fourth, Kaun motivates the dispreference for cross-height labial harmony (Steriade 1981:5) in the phonetic implementation of the phonological feature [round]. She argues that rounding harmony involves the multiple linkage of one particular [round] feature, and that because lip gestures differ according to vowel height, the phonetic modulation of the lips to accommodate these differences militates against cross-height harmony (1995: §5.6). The constraint formulated to ban cross-height harmony is GESTURAL UNIFORMITY [RD]. However, this constraint is inadequate. The typological tendency is not only to strings of same-height round vowels, but rather to strings of identical vowels. Only eastern Mongolian dialects allow same-height where trigger and target
disagree in backness. In these languages, some root-internal vowels underwent umlaut, which resulted in the exception to this cross-linguistic tendency (Svantesson 1985:§3.1). In all other instances, same-height harmony produces sequences of identical vowels. For this reason Kaun’s analysis, as well as formal analyses that require height identity (Mester 1988; van der Hulst & Moskal 2013) miss an important generalization—labial harmony preferentially produces sequences of identical vowels.

One might object that in cases like Turkish there is no clear preference for sequences of identical vowels in labial harmony since round vowels of both heights trigger harmony. In a study on regressive labial harmony in loanwords Kaun (1999:97) finds that the Turkish high vowels are better triggers for regressive labial harmony than the non-high vowels. More concretely, in loans with an initial consonant cluster, a high vowel is epenthesized cluster-medially, and that epenthetic vowel optionally undergoes regressive labial harmony (Yavaş 1980; Clements & Sezer 1982:247). Kaun found that for all speakers, high vowels obligatorily triggered harmony, but non-high vowels obligatorily triggered harmony in only two of nine speakers. Thus, Kaun’s findings could be construed as evidence for a sequential identity preference in Turkish that is inactive in the native lexicon.

Lastly, the length asymmetry noted in (8) above was not included in Kaun (1995), but was added after Li (1996) reported that in Baiyina Orochen and Evenki, two Tungusic languages, short vowels trigger harmony but long vowels do not. In a general sense, Kaun’s generalization may relate to phonetic studies by Bennett (1968) and Ainsworth (1972), which found that duration was an important factor in vowel discrimination, particularly in more crowded portions of the vowel space. Note, though, that Kaun (2004) assumes that only contrastive length may affect harmony.

In summary, labial harmony patterns are typically parasitic, depending on the agreement of some condition(s) orthogonal to labialization. These conditions are often interpreted as phonetic, functional factors. Specifically, perceptual weakness is argued to undergird the trigger strength asymmetries, while perceptual salience motivates the target preferences noted in Kaun (1995, 2004). In the Kazakh data presented in §2.2.2, high triggers are, contra Kaun’s typology, preferential triggers for harmony. One might argue that this stems from the preference for high targets in the language in tandem with the preference for same-height harmony. In Section 4 below I present acoustic data that further suggest that Kazakh does not conform to Kaun’s generalizations. Crucially, I argue in §4-6 that a systemic, Dispersion theoretic (Liljencrants & Lindblom 1972; Lindblom 1986) conception of weakness, where perceptual weakness is defined in terms of contrasts rather than abstract feature bundles, provides a superior analysis for Kazakh. Moreover, in §7 I use the analysis of Kazakh presented below to re-cast the larger typology of labial harmony along dispersionary lines.

4 Operationalizing perceptual similarity

In this section I develop a Dispersion theoretic analysis of labial harmony in contemporary Kazakh, suggesting that weakness should be defined in dispersionary terms without necessary reference to abstract features. In §4.1 I introduce Dispersion Theory (Liljencrants & Lindblom 1972; Lindblom 1986), and its more recent instantiations (Schwartz et al. 1997; Flemming 2002, 2004, 2008a,b; Padgett 2004). In §4.2 I compute perceptual distances for the Kazakh inventory. The distances calculated in this section are crucially referenced in the analysis presented in §5.
4.1 Dispersion theory

Dispersion theory (henceforth DT; Liljencrants & Lindblom 1972; Lindblom 1986) attempts to model the fitness of an inventory rather than that of an individual segment. Under the assumption that vowels should be sufficiently dispersed in the acoustic space, DT models the acoustic space without explicit reference to features, but rather to the local positions of each phonemic vowel in the vowel space. Following Liljencrants & Lindblom (1972), most work in DT has sought to explain the particular shapes that vowel inventories often take (Lindblom 1975, 1986; Schwartz et al. 1997; Flemming 2002). For example, the frequency of the three-vowel system containing /i a u/ is explained by the large distances between the three vowels in the system within DT. In Flemming (2002) this dispersionary motivation is formalized with a set of Optimality theoretic (Prince & Smolensky 2004) constraints that enforce minimum distance in opposition to other constraints favoring contrast maximization. The interaction of these conflicting constraints in Flemming (2002) define the shape of the vowel space for a given language. Whereas previous work used dispersion heuristically to define inventories, Flemming’s work (2002; 2004; 2008a,b) assumes that speakers synchronically access prototype-like representations rather than abstract features. Figure 2 below shows Flemming’s representational system. Perceptual distances are calculated in spatial rather than featural terms by counting grid spaces between two given vowels.

Figure 2: Flemming’s vowel space (2002:30)

<table>
<thead>
<tr>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
</tr>
<tr>
<td>y</td>
</tr>
<tr>
<td>i</td>
</tr>
<tr>
<td>u</td>
</tr>
<tr>
<td>i</td>
</tr>
<tr>
<td>y</td>
</tr>
<tr>
<td>o</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>ø</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>œ</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

Similarly, Padgett (2004), uses an ERB scale to define the Russian vowel space, with SPACE constraints militating against category prototypes (assumed to be represented by mean formant values) insufficiently dispersed in the acoustic space. These two approaches, however, differ significantly in the amount of detail they reference. Flemming abstracts from language-specific detail onto a grid-like vowel space while Padgett directly references mean formant values. Using Dispersion theory, perceptual weakness under Kaun’s larger analysis of weakness-driven harmony can be redefined to accommodate language-specific phonetic detail, as in Padgett (2004). In §4.2 I describe vowel data collected and the calculation of perceptual distances that inform the treatment of Kazakh labial harmony developed in §4.3.

4.2 Determining perceptual distances

In this section I apply the methodologies used in previous DT work to the Kazakh vowel inventory to compute perceptual distances, focusing on the eight vowels that regularly participate in harmony. The perceptual distances calculated below are referenced in the formal analysis presented in §5.

4.2.1 Kazakh vowel inventory

Data for the present study was collected during fieldwork in southeastern Kazakhstan. Data from eleven speakers is presented below. Of the eleven, eight were female. Mean age was 33.5 years,
with a range of 19-46 years. Most speakers were from southeastern Kazakhstan. More detailed information regarding elicitation protocol and speaker metadata is presented in McCollum (2015:331-333). Root vowel data (N=2,490) was elicited to assess the relative positioning of each vowel in the vowel space. Mean F1-F2 for the eight phonemic vowels under study is shown below in Figure 3.

Within each set of putative front and back vowels, the high and mid vowels do not differ in height, in accordance with Johanson (1998b:94) and Kara (2002:9), which report a lowering of the historical high vowels in Kazakh.\(^3\)

**Figure 3: Plot of Mean Root Vowel F1-F2 (z) with 1 standard deviation ellipses**

Crucially, the data in Figure 3 suggest that the likelihood a vowel will trigger harmony is not, contra Kaun, dependent on its height and backness. Despite the fact that that /ɨ/ and /ø/, and /ʊ/ and /ɔ/ are realized with almost identical vocalic gestures, respectively (Dzhunisbekov 1980:21), they trigger different patterns of harmony. If high vowels should be poorer triggers, it is surprising that [ɨ] is the best trigger and [ɔ] is the poorest. If front vowels are better triggers than back vowels, the tendency for harmony after these two vowels is accounted for, but the fact that harmony is more likely after [ø] than [ʊ] is unexplained. Moreover, since [ɨ] and [ø] are of the same approximate height, as are [ʊ] and [ɔ], the infrequency of post-initial [ʊ] and [ɔ] is surprising. Target preference seems to have little to do with vowel height in Kazakh. Thus, for both triggers and targets, labial harmony in Kazakh is typologically anomalous under Kaun’s analysis.

It is clear that vowel height, defined in terms of F1, is insufficient to differentiate these vowels, but vowel length does distinguish the historical high from non-high vowels. High vowels are significantly shorter (Dzhunisbekov 1972:75; Kirchner 1998:319; Kara 2002:9; Washington 2015). Table 2 presents duration data (N=32) for eight vowels produced in analogous environments by Speaker 11.

\(^3\) Moreover, vowel height shifts are not unattested in Turkic. The high vowels have lowered and the mid vowels have raised in contemporary Tatar (Poppe 1968; Comrie 1997).
Table 2: Mean duration (in ms., with standard deviations) of Kazakh vowels in initial syllables

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Mean Duration (SD)</th>
<th>Vowel</th>
<th>Mean Duration (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʏ</td>
<td>22.5 (8.0)</td>
<td>ʊ</td>
<td>43.3 (14.5)</td>
</tr>
<tr>
<td>ɪ</td>
<td>47.7 (10.9)</td>
<td>ɯ</td>
<td>30.1 (12.5)</td>
</tr>
<tr>
<td>ø</td>
<td>90.1 (18.5)</td>
<td>ɔ</td>
<td>78.4 (27.1)</td>
</tr>
<tr>
<td>e</td>
<td>69.7 (23.0)</td>
<td>a</td>
<td>62.4 (8.0)</td>
</tr>
</tbody>
</table>

When duration was averaged over the two putative heights, mean duration was 35.9 ms for the historically high vowels, and 75.2 ms for the historically non-high vowels. While it is possible to construct an analysis that uses contrastive length instead of height, I show in §6 though that this analytical move still fails to provide the empirical coverage provided by the dispersionary analysis developed below.

In addition to duration, the mid (long) vowels /e ø ɔ/ are distinguished from the high (short) vowels by diphthongization. The mid vowels have developed onglides that agree in labiality. Thus, /e/ is realized as [je], while /ø/ and /ɔ/ are realized as [wø], and [wɔ], respectively (Dzhunisbekov 1980:21; Vajda 1994:619-624). The diphthongization of these phonemes is most evident word-initially.

In summary, the Kazakh harmony pattern does not conform to a number of Kaun’s generalizations. As the mid and high vowels do not differ in height, and moreover are composed of the same vocalic gestures, Kaun’s analysis cannot easily explain the trigger and target asymmetries found in the language.

4.2.2 Quantifying dispersion in the Kazakh vowel space

Converting raw acoustic data into perceptual distances is integral to the analysis presented herein. In this section I replicate the general procedures outlined in Schwartz et al. (1997) and Padgett & Tabain (2005) to calculate perceptual distances between harmonic counterparts. Raw formant values in Hertz are first converted to ERB to more accurately reflect human perceptual abilities. Second, the data was normalized (Labonov 1971) to account for interspeaker physiological differences. Third, normalized ERB values were warped to more heavily weight F1 over F2 to reflect increased perceptual sensitivity to differences in F1 (Flanagan 1955; Lindblom 1975). The perceptual distances calculated in this section will be used in the analysis presented in §5. Following Padgett & Tabain (2005:23), F3 is excluded from analysis. By-vowel mean F1-F2 with standard deviations are reported in Table 3 below.

**Footnotes:**
4 F3 is not incorporated into the model developed herein for multiple reasons. First, F3 was found not to significantly improve vowel discrimination in a discriminant analysis. Second, F3 was highly variable within speakers, and although F3 was typically lower for rounded vowels, it was determined to be an insignificant parameter in the model developed (McCollum 2015; see also Padgett & Tabain 2005:23). Thirdly, the model performed poorer using a perceptual second formant affected by F3 (see Schwartz et al. 1997:263 and de Boer 2004:448 for deriving F2').
5 If one were to compute perceptual distances, Δ, based on raw formant values, one would obtain the following Euclidean distances (in Hz, with no warping of F2): t-ʏ=253.2, u-ø=302.7, a-ɔ=408.5, e-ø=819.1.
Raw formant values were then transformed into ERB (Equivalent Rectangular Bandwidth) using the equation provided in Reetz & Jongman (2008:245-246). Converting the data into ERB scales the data in a more perceptually plausible way. Additionally, Schwartz and colleagues warp the vowel space with a parameter, $\lambda$, such that the ratio of $F_2$-to-$F_1$ ranges is $0.75-0.5:1$ (see also de Boer 2000; Padgett 2004; Padgett & Tabain 2005). As in Lindblom (1975), this weights $F_1$ over higher formant contrasts, as $F_1$ has been found to be a more robust acoustic cue (Flanagan 1955). Furthermore, the $F_1$ dimension appears to be obligatorily exploited in vowel systems, whereas in vertical vowel systems (e.g. Kabardian, Marshallese), $F_2$ is allophonically determined. For this reason it is necessary to warp the vowel space to reflect the salience of $F_1$ over $F_2$. ERB-transformed $F_1$-$F_2$ are presented in Table 4.\(^{6}\)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Mean $F_1$ (SD)</th>
<th>Mean $F_2$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>13.03 (1.16)</td>
<td>17.50 (1.23)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>10.17 (0.80)</td>
<td>15.32 (1.27)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>10.68 (1.60)</td>
<td>17.90 (0.94)</td>
</tr>
<tr>
<td>$\omega$</td>
<td>10.29 (1.33)</td>
<td>15.92 (1.23)</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>8.89 (0.80)</td>
<td>21.97 (1.09)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>8.97 (0.80)</td>
<td>18.21 (1.67)</td>
</tr>
<tr>
<td>$\imath$</td>
<td>9.11 (1.27)</td>
<td>20.51 (1.04)</td>
</tr>
<tr>
<td>$\upsilon$</td>
<td>8.65 (1.08)</td>
<td>19.30 (1.36)</td>
</tr>
</tbody>
</table>

After conversion to ERB, data was normalized (Labonov 1971) to reduce inter-speak anatomical and physiological differences, allowing for across-speaker comparisons. Results from this transformation are shown in Table 5.

To achieve the 0.625:1 midpoint in the range of $F_2$:F1 ratios proposed by previous writers, $\lambda$ was set to 0.723.\(^{7}\) In Table 5 the perceptual distance between $[\imath]$ and $[\upsilon]$ is 0.365, much smaller

---

\(^{6}\) If perceptual distances are computed on the basis of $F_1$-$F_2$ in ERB, the following values emerge: (with no warping of $F_2$) \(1-\upsilon=1.29\), \(\imath-\alpha=2.02\), \(\alpha-\sigma=3.60\), \(e-\phi=3.76\).

\(^{7}\) The distances, if computed without adjusting the weight of $F_2$, would be \(1-\upsilon=0.447\), \(\imath-\alpha=0.777\), \(e-\phi=1.138\), \(\alpha-\sigma=1.521\). In principle, the target ratio could have fallen anywhere in the range proposed by
than any other distances reported. Moreover, the relative distances between all four pairs correspond to the trigger strength hierarchy described in §2.2.2, namely, \( \gamma > \sigma > \varnothing > \sigma \).

Table 5: Normalized ERB Mean root vowel F1-F2 with standard deviations

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Mean F1 (SD)</th>
<th>Mean F2 (SD)</th>
<th>Weighted Euclidean Distance ((\lambda(F2)=0.723))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>1.339 (0.452)</td>
<td>-0.169 (0.298)</td>
<td>1.451</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>-0.032 (0.354)</td>
<td>-0.828 (0.340)</td>
<td>1.451</td>
</tr>
<tr>
<td>(\mu)</td>
<td>0.117 (0.657)</td>
<td>0.103 (0.272)</td>
<td>0.568</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>-0.007 (0.587)</td>
<td>-0.664 (0.353)</td>
<td>0.568</td>
</tr>
<tr>
<td>(e)</td>
<td>-0.637 (0.375)</td>
<td>1.196 (0.265)</td>
<td>0.823</td>
</tr>
<tr>
<td>(\varnothing)</td>
<td>-0.595 (0.399)</td>
<td>0.059 (0.482)</td>
<td>0.823</td>
</tr>
<tr>
<td>(i)</td>
<td>-0.542 (0.532)</td>
<td>0.745 (0.227)</td>
<td>0.365</td>
</tr>
<tr>
<td>(y)</td>
<td>-0.788 (0.459)</td>
<td>0.372 (0.388)</td>
<td>0.365</td>
</tr>
</tbody>
</table>

The trigger strength hierarchy after conversion to ERB, normalization, and warping differs from the hierarchy obtained from raw formant values, which is \( \gamma > \sigma > \sigma > \varnothing \). This difference in the ranking of \([\alpha]\) and \([\varnothing]\) derives from the independently motivated transformations discussed above. These transformations highlight the difference between the acoustic and perceptual properties of vowels, and this difference underlies the change in relative rankings for these two vowels.

In Figure 4 below the vowel space is depicted along with the weighted Euclidean distances between each harmonic pair. In Table 6 weighted Euclidean distances are shown between each vowel in the vowel space.

Figure 4: Perceptual distances between means of labial harmonic counterparts

Schwartz and colleagues. Although the distances would be affected slightly, the overall relations would not change.
The three transformations applied to the data, ERB conversion, normalization, and asymmetric weighting of F1 over F2 result in the distances shown in Table 5. Raw formant values were converted to ERB to translate acoustic data into a more reasonable approximation of human perceptual abilities. In this section I use these distances to motivate a set of constraints for the analysis of harmony in Kazakh.

5 Analysis

In this section I use the perceptual distances determined in §4.2.2 above to develop an analysis of the data presented in (3-5), repeated in (10-12) below, in Maximum Entropy Harmonic Grammar (MaxEnt, Smolensky 1986; Legendre et al. 1990; Goldwater & Johnson 2003).

5.1 Maximum Entropy Harmonic Grammar

Harmonic Grammar (henceforth, HG; Legendre et al. 1990; Pater 2009) is similar to canonical Optimality Theory (Prince & Smolensky 2004) in many ways. However, HG uses weighted rather than strictly-ranked constraints to derive output predictions (Pater 2009). Notably, HG allows ganging effects where multiple constraints can cumulatively outrank a higher-weighted constraint. When modeling categorical data, the output candidate that receives the most harmonic score, determined by a summed product of constraint violations multiplied by their respective weights, is deemed the winner.

To analyze variable data like that presented in §2.2.2 a probabilistic version of HG is necessary. In addition to handling varying frequencies of harmony, the probabilistic grammar used herein allows us to more precisely describe the relationship between trigger perceptual distance and the rate of harmony. The particular version I used throughout this section is Maximum Entropy HG (MaxEnt; Goldwater & Johnson 2003; Hayes & Wilson 2008; McPherson & Hayes 2016). This implementation of HG assigns a harmony score, \( H \). The harmony of each candidate is assessed using the following equation (Hayes & Wilson 2008:383):

\[
H(x) = \sum_{i=1}^{N} w_i \cdot C_i(x)
\]

where

---

8 These are approximations only. Without actual perceptual data these findings must suffice, but future research on vowel perception is needed for a fuller understanding of the interaction between vowel dispersion and harmony.

9 Noisy HG (Boersma & Pater 2008) would also presumably account for the data well. This paper does not compare these two grammatical. For the interested reader, see Jesney (2007); Potts et al. (2010), and McPherson & Hayes (2016).
\(w_i\) is the weight of the \(i^{th}\) constraint, \\
\(C_i(x)\) is the number of times that \(x\) violates the \(i^{th}\) constraint, and \\
\(\sum_{i=1}^{N}\) denotes summation over all constraints \((C_1 \ldots C_N)\)

Thus, the harmony score, \(H(x)\), for each output candidate is the summed total of constraint violations multiplied by their respective weights. All constraints weights are assumed to be negative herein (cf. Kimper 2011).

From the harmony score, \(H(x)\), the probability of output candidate \(x\), \(P_x\), is calculated by taking the exponent of the harmony score normalized by the summed score of the exponents of all competing output candidate forms:

\[
(11) \, P_x = \frac{e^{H(x)}}{\sum_{j=1}^{M} e^{H(y)}}
\]

In the end, harmony scores with more negative values result in lower probabilities because the more negative \(H(x)\) is, the smaller the numerator will be in (11) above. Thus, the ideal candidate would have a harmony score of zero, incurring no violations of faithfulness or markedness constraints (cf. Kimper 2011). The relation between \(H(x)\) and \(P_x\) is contingent upon the set of output competitors under examination, as in (11). Thus, a harmony score of, for instance -5, may very well represent a high probability outcome if other outputs have harmony scores significantly lower than -5. On the other hand, a score of -5 may indicate an extremely low probability candidate if other candidates’ harmony scores are well above -5.

The goal of MaxEnt HG is to match predicted probability with attested frequency. As the Kazakh data is decidedly gradient, using this type of model provides a nice analytic fit to the data under study, which is reviewed below in §5.2.

5.2 A review of the data

The data presented in §2.2.2 is briefly reviewed below. Labial harmony overwhelmingly applies root-internally when the target is high (or alternatively, short, as in 12a-c) and when the trigger is \([\text{y} \, \text{o} \, \text{o}]\). In contrast, \([\text{z}]\) does not typically initiate harmony (12d). Furthermore, non-high (or alternatively, long) vowels /e a/ also do not typically undergo harmony (12e-h).

\[
(12) \quad \text{Root-internal harmony}
\]

<table>
<thead>
<tr>
<th>Attested variants</th>
<th>Frequency of harmony</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ʒyzyk ~ ʒyzik</td>
<td>(10/11)</td>
<td>‘ring’</td>
</tr>
<tr>
<td>b. kɔskyk ~ kɔskik</td>
<td>(10/11)</td>
<td>‘desert carrot’</td>
</tr>
<tr>
<td>c. qɔłun ~ qɔłun</td>
<td>(16/17)</td>
<td>‘colt’</td>
</tr>
<tr>
<td>d. qɔzø ~ qɔzøu</td>
<td>(1/10)</td>
<td>‘lamb’</td>
</tr>
<tr>
<td>e. tylɔk ~ tylek</td>
<td>(1/11)</td>
<td>‘graduate’</td>
</tr>
<tr>
<td>f. tɔbe</td>
<td>(0/11)</td>
<td>‘hill’</td>
</tr>
<tr>
<td>g. qołaq</td>
<td>(0/10)</td>
<td>‘ear’</td>
</tr>
<tr>
<td>h. bɔɬat</td>
<td>(0/10)</td>
<td>‘steel’</td>
</tr>
</tbody>
</table>
Regardless of trigger or target vowel quality, harmony does not typically obtain in suffixes, as in (13). When harmony does obtain across morpheme boundaries, it is more likely in high (short) vowels (13a-d, e-h).

(13) Suffix harmony

<table>
<thead>
<tr>
<th>Attested variants</th>
<th>Freq. of harmony</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. kvl-dv ~ kvl-di</td>
<td>(7/22)</td>
<td>‘laugh-PST.3’</td>
</tr>
<tr>
<td>b. øl-dv ~ øl-di</td>
<td>(1/20)</td>
<td>‘die-PST.3’</td>
</tr>
<tr>
<td>c. qor-du ~ qor-duu</td>
<td>(1/18)</td>
<td>‘construct-PST.3’</td>
</tr>
<tr>
<td>d. qɔs-tu ~ qɔs-tuu</td>
<td>(1/20)</td>
<td>‘add-PST.3’</td>
</tr>
<tr>
<td>e. kvl-se</td>
<td>(0/15)</td>
<td>‘laugh-COND’</td>
</tr>
<tr>
<td>f. øl-se</td>
<td>(0/14)</td>
<td>‘die-COND’</td>
</tr>
<tr>
<td>g. qor-sa</td>
<td>(0/13)</td>
<td>‘construct-COND’</td>
</tr>
<tr>
<td>h. qɔs-sa</td>
<td>(0/13)</td>
<td>‘add-COND’</td>
</tr>
<tr>
<td>i. ʒyzyk-ty ~ ʒyzyk-ti</td>
<td>(1/9)</td>
<td>‘ring-ACC’</td>
</tr>
<tr>
<td>j. ʒyzyk-ter-di</td>
<td>(0/23)</td>
<td>‘ring-PL-ACC’</td>
</tr>
</tbody>
</table>

However, in the converbial suffix, shown in (14), harmony applies more consistently than in the examples above. The potential reasons for this asymmetry are discussed in §2.2.2.

(14) Converbial suffix

<table>
<thead>
<tr>
<th>Attested variants</th>
<th>Frequency of harmony</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. qorra-p</td>
<td></td>
<td>‘look-CVB’</td>
</tr>
<tr>
<td>b. soļle-p</td>
<td></td>
<td>‘speak-CVB’</td>
</tr>
<tr>
<td>c. qat-up</td>
<td></td>
<td>‘harden-CVB’</td>
</tr>
<tr>
<td>d. kes-ip</td>
<td></td>
<td>‘cut-CVB’</td>
</tr>
<tr>
<td>e. kvl-yp ~ kvl-ip</td>
<td>(26/31)</td>
<td>‘laugh-CVB’</td>
</tr>
<tr>
<td>f. qor-op ~ qor-up</td>
<td>(13/18)</td>
<td>‘construct-CVB’</td>
</tr>
<tr>
<td>g. øl-yp ~ øl-ip</td>
<td>(10/18)</td>
<td>‘die-CVB’</td>
</tr>
<tr>
<td>h. qɔs-op ~ qɔs-up</td>
<td>(1/15)</td>
<td>‘add-CVB’</td>
</tr>
</tbody>
</table>

The crux of the analysis presented herein lies in connecting the perceptual distances from §4.2.2 with the above data. With the perceptual distances calculated above in mind, the front vowel [ʏ] is least dispersed from its counterpart, making it the most perceptually weak, and accordingly, most aided by harmony. For that same lack of dispersion, [ʏ] is most likely to surface from the underlying form of its counterpart, [ɪ], because it involves the least perceptible adjustment to an input vowel to satisfy harmony. The opposite holds for the back vowel, [ɔ]. Because [ɔ] is most dispersed from its counterpart it is least aided by harmony, and because this difference in perceptual distance is why it fails to surface as a result of harmony. These generalizations are encoded in the constraint set defined below.
5.3 Defining the constraint set

In this section, I introduce a set of constraints to capture the labial harmony data in MaxEnt HG. First, a harmony-driving constraint is necessary, and such a constraint should make reference to perceptual similarity. I employ a scalar alignment constraint here, but compare consider a conditional alignment constraint at final model comparison. Note that I do not analyze palatal harmony here, which would employ a separate set of constraints.

5.3.1 Harmony-driver

In harmonic grammars, the weight of a given constraint may be scaled according to some relevant scale to assign asymmetric penalties to certain structures. In (15) a general alignment constraint is introduced.

(15) General Alignment

\[ \text{ALIGN-R(RD, WD)} \] align the roundness of the initial vowel to the right edge of the word.

In HG, all constraints are assigned weights, \( w_c \). Scalar constraints assign a scaling factor to \( w_c \), that augments the weight of violations in relation to some factor, like sonority, sequential distance, or in this case, perceptual distance. The scaling mechanism for \( \text{ALIGN-R(RD, WD)} \) is defined in (16).

(16) Definition: Scalar \( \Delta \text{[ALIGN-R(RD, WD)]} \)

given a constraint weight, \( w_A \), and

a scale, \( \Delta = \{0, 0.001 \ldots n\} \) corresponding to perceptual distance between round-unrounded harmonic counterparts,

For every vowel to which the roundness of the initial vowel with perceptual distance \( \delta \in \Delta \) is not aligned, assign a weighted violation of \( w_A \cdot \delta \).

The scale employed is a linear scale relating perceptual distance to the probability that harmony will obtain after a given trigger vowel (cf. Kimper 2011; Zymet 2014, in prep.). Figure 5 presents root perceptual distance and pooled frequency of application, showing a strongly linear relationship between the perceptual distance between a trigger vowel and its harmonic counterpart and the application of harmony. In Figure 5 it is apparent that perceptual distance is negatively correlated with rate of application, as argued throughout §2.2.2.
Using scalar constraints forces a redefinition of the equation in (13) to include the possibility of a scale for each constraint, as shown in (17).

\[
H(x) = \sum_{i=1}^{N} w_i \cdot C_i(x) \cdot S_i(x)
\]

where

- \( w_i \) is the weight of the \( i^{th} \) constraint,
- \( C_i(x) \) is the number of times that \( x \) violates the \( i^{th} \) constraint,
- \( S_i(x) \) is the scale that \( C_i(x) \) is scaled by (herein, \( \Delta \) for perceptual distance), and
- \( \sum_{i=1}^{N} \) denotes summation over all constraints \( (C_1 \ldots C_N) \)

Scaling provides a way to restrict the relations between triggers in the vowel inventory. Rather than fitting the model to each segment individually, which offers no restriction on possible relations, scaling formally delimits the possible types of relations.

To determine the optimal scaling various ratios (alternatively, slopes) of \( w_{ALIGN}^{[1]} : w_{ALIGN}^{[3]} \), ranging from 1:1 to 1,000:1 were applied to the data and the scale that best fit the data, 2.3:1, was selected. The optimal model scaled the weight of the alignment constraint thusly.

\[
\begin{align*}
Y & : 1 \cdot w_{ALIGN} \\
\phi & : 0.894 \cdot w_{ALIGN} \\
\phi & : 0.762 \cdot w_{ALIGN} \\
\text{o} & : 0.435 \cdot w_{ALIGN}
\end{align*}
\]
In this model the penalty for violating ALIGN-R after trigger (root) \( \gamma \) is 2.3 times greater than after \( \beta \). Using the perceptual distance between a trigger vowel and its counterpart to scale the weighted alignment constraint produces probabilistic predictions from the interaction of constraint violation, scaled weights, and the larger probabilistic grammar, as in McPherson & Hayes (2016). The conditional alignment constraint is more internally complex, relying on internal arguments to define the nature and scope of application. Internally, the conditional constraint establishes a threshold to categorically distinguish between triggers and non-triggers. In a scaled alignment constraint, however, no such distinction is made. For these reasons, in addition to differences in empirical coverage, a scalar alignment constraint was used for the analysis, although comparisons between models with each type of harmony-driver is shown in §5.4.

5.3.2 Markedness constraints
 Notice that the domain of harmony in the above constraint is set to the word, although harmony is typically coextensive with the root. In longer words, like \( \text{ʒy} \text{zyk-\text{tι}} \) ‘ring-ACC’ harmony does not extend to the right edge of the word. If, however, harmony is set to align root roundness to the right edge of the root, (ALIGN-R(RD,RT)), this predicts no harmony across morpheme boundaries, which is inaccurate.

In order to distinguish root-internal from suffix harmony I use the following markedness constraint banning the extension of a labial gesture across a morpheme boundary, CRISPEDGE[RD,MORPH] (Itō & Mester 1999; Walker 2011).

(19) CRISPEDGE[RD,MORPH]- assign a violation to every instance of [RD] that is associated with more than one morpheme.

Two kinds of evidence undergird the proposed constraint. First, there is some evidence that indicates articulatory gestures are more tightly coordinated within morphemes than across morphological boundaries. Cho (2001) finds that palatalization of \([t]\) and \([n]\) across morpheme boundaries in Korean is less consistent than tautomorphemic palatalization. There is phonological evidence, as well, that suggests the importance of morphological boundaries for assimilation. In a number of languages, harmony processes are known to be sensitive to these boundaries. In Yeyi, a Bantu language of southern Africa, regressive assimilation of \([i]\) to \([u]\) consistently obtains within roots, but is reportedly more variable across a root boundary (Seidel 2008:47-48). Second, more categorical blocking by morphology has been noted in a variety of languages. Hansson (2010:326-327) notes that consonant harmony is frequently limited to morphological roots. For vowel harmony, Archangeli & Pulleyblank (2007:365) report that Ngbaka ATR harmony is similarly root-delimited. Along these same lines, Pulleyblank (2002) analyzes height harmony in C’Lela using root- and word-delimited faithfulness constraints.

One additional markedness constraint is necessary for the analysis, *[RD]*, which penalizes both epenthetic and underlying vowels that surface as round. In this way, the relative frequency of harmony in the converbial affix is differentiated from that in the past tense morpheme because the past tense morpheme has an underlying high vowel, but the converb does not.

(20) *[RD]*- assign a violation to every round vowel

Epenthetic vowels often pattern differently from underlying vowels in harmony systems. Hahn (1991:49-51) reports that labial harmony in closely related Uyghur applies only to epenthetic vowels, leaving underlying vowels unaffected. This same asymmetry is reported in Marash Armenian (Vaux 1998:169), where underlying vowels undergo palatal harmony while epenthetic
vowels additionally undergo labial harmony. Uffmann (2006) shows that in Sranan and Samoan, languages without general harmony, epenthetic vowels typically harmonize in backness with a preceding vowel rather than surfacing with some default specification.¹⁰

5.3.3 Faithfulness constraints
In addition to the above markedness constraints, two IDENT constraints (McCarthy & Prince 1995) are necessary. A general IDENT-IO[RD] constraint bans changes to input [RD] specifications.

(21) IDENT-IO[RD]- assign a violation to every input-output pair that disagree for [RD].

I assume throughout that harmonic vowels are fully specified underlyingly, bearing the feature [-RD] (cf. Inkelas 1995), both based on evidence from language games in Turkic (Harrison & Kaun 2000), and the improved performance of the models developed below when non-initial vowels were underlyingly specified for [-RD]. Lip rounding, leading the hearer to construct underlying representations with [-RD] rather than [+RD]. Harrison & Kaun (2000) argue that lexicon optimization should operate over regularly alternating segments, in addition to non-alternating segments (see also Krämer 2003:§6.1.2-3). Also, when the models shown below were compared to models that used inputs with non-initial labialization, the underlyingly unrounded model performed much better. When [+RD] inputs were used, Ident-IO[RD] received a weight of 0, while *[RD] was given more priority. In canonical OT, *RD would obligatorily outrank Ident-IO[RD], but this is not necessary in the MaxEnt model used herein.¹¹ If speakers optimize their lexical representations based on surface patterns, then it seems reasonable that Kazakh speakers utilize unrounded inputs because doing so optimizes the generalizations to be extracted from labial harmony patterns in the language. This second reason hinges upon the constraint set, as a different constraint set would result in differing degrees of model fit.

In addition to the general faithfulness constraint above, I employ a conditional faithfulness constraint that militates against any perceptually large modifications to an input [RD] feature.¹² As with the alignment constraint formulated above, this faithfulness constraint directly references the perceptual distances calculated in §4.2.¹³

---

¹⁰ It is also possible to construe harmony as affecting the converb more than the past tense marker because harmony spans one consonant in the converb but two in the past tense morpheme. Xiangru & Hahn (1989:272-273) report that in Ili Turki labial harmony is blocked by clusters. Krämer (2001) and Mahanta (2008) find the same phenomenon in Yucatec Maya and Assamese, respectively, suggesting that harmony may not span a mora. While it appears that coda consonants are moraic in Kazakh, harmony does span multiple consonants within roots, like [bvrkvt] ‘golden.eagle.’ Choosing either way to differentiate frequency of harmony in CVB and PST.3 respectively does not affect the analysis, though.


¹² It tentatively seems possible to extend this reasoning to palatal harmony in Kazakh, as the most distinct palatal harmonic pair [a]-[e], is also the pair most likely to block harmony. It might be possible to extend the present work to encompass palatal harmony as well, but I leave that for future work.

¹³ It is also possible to scale the faithfulness constraint by perceptual distance. Pooled rate of application by target type was not easily fit by a linear model, though. Thus, for this paper it seems reasonable to confine the analysis to only one scaled alignment constraint.
(22)  Conditional Faithfulness

\[
\text{IDENT-IO}[\text{RD}/\Delta > x] - \text{assign a violation for every input-output [RD] pair whose perceptual distance, } \Delta, \text{ exceeds } x.
\]

This constraint is in the spirit of Steriade’s (2001) P-Map proposal in several ways (see also Zuraw 2007 and Hayes & White 2015). Generally, Steriade contends that speakers have access to the relative perceptibility of a set of given contrasts. Second, she argues that speakers often choose the least perceptible alternation to avoid undue deviation from some form. If features are completely eschewed, it is conceivable that to avoid violating (22) a set of vowels could re-pair, like \([\alpha]-[\circ]\) instead of the actual \([\alpha]-[\circ]\), or \([i]-[\sigma]\) instead of the attested \([e]-[\sigma]\). I will assume that this re-pairing is precluded by both the presence of featural representations, as well as the same modification to lexicon optimization proposed by Harrison & Kaun (2000) above. The essence of their argument is that speakers have analogical access to the, in this case, vowels that alternate with one another, and they can use this access to both create underlying forms, and ostensibly, maintain harmonic pairings. Likewise, this constraint bans perceptually large deviations from input forms, although this constraint formulates faithfulness in terms of inputs and outputs, as opposed to *MAP constraints, which enforce output-output relations. Since \([\sigma]\) and \([\mathtt{s}]\) do not surface non-initially the distance which is permissible to adjust a vowel’s acoustic realization must be smaller than 0.823, the distance between \([e]\) and \([\sigma]\). Because \([\circ]\) surfaces via harmony the threshold must also be greater than 0.568, the distance between \([\circ]\) and \([\mathtt{u}]\) (see Table 5 and Figure 4 above). Therefore, the following version of this constraint defines the actual set of undergoers in contemporary Kazakh.

(23)  \text{IDENT-IO}[\text{RD}/\Delta > 0.7] - \text{assign a violation for every input-output [RD] pair whose perceptual distance, } \Delta, \text{ exceeds } 0.7.

The exact perceptual distance used in (23) is somewhat arbitrary, serving to differentiate only between alternations that are regularly attested and those that are at best rarely attested. Thus, shifting underlying /e/ to \([\sigma]\) (0.823) or /a/ to \([\mathtt{s}]\) (1.451) incurs a violation of this constraint, but changing \([i]\) to \([\vee]\) (0.365) or \([u]\) to \([\circ]\) (0.568) does not because these two alternations do not exceed the threshold established by the constraint, 0.7. As initial-syllable (root) vowel alternations are unattested in the language I will assume a highly-ranked positional faithfulness constraint (Beckman 1997) to preclude the possibility of unrounding initial-syllable vowels to avoid harmony in the tableaux below.

In total, the analysis uses the following five constraints. First, a scalar alignment constraint, ALIGN-R([RD,WD]) drives harmony rightward throughout the word. Second, CRISPEDGE([RD,MORPH]) limits the scope of labialization to one morpheme. Third,*RD bans round vowel generally. Fourth, IDENT-IO[RD] bans input-output mismatches among vowels, and finally, the conditional faithfulness constraint, IDENT-IO ([RD]/\Delta > 0.7) prohibits input-output mismatches for the feature [RD] that exceed the perceptual threshold, 0.7. As noted above, a positional faithfulness constraint is assumed, but was not included actual model calculations.

5.4  Determining the weights

Goodness of fit was determined by the log likelihood of a given model. The log likelihood is the summed product of log probabilities, \(\log(P_x)\), and the attested frequencies of the candidates under examination, which results in a negative number with a large absolute value. Constraint weights
were assigned in order to maximize the model’s log likelihood. As the search space for MaxEnt models is free of local maxima (Della Pietra et al. 1997), a variety of optimization algorithms converge on one approximate solution. Accordingly, I used Excel’s Solver add-in (Fylstra et al. 1998) to find the maximum (least negative) log likelihood for a set of input-output forms and their frequency, in conjunction with the constraint set defined above and the violations incurred by each output form.\footnote{I did not use Wilson & George’s (2009) MaxEnt Grammar Tool because constraint violations can only be input as integers in MaxEnt Grammar Tool, which limits the precision of the calculations made herein that rely on non-integer values.}

One pitfall in machine learning is overfitting the data. Essentially, a model may learn to describe the data well but may perform poorly when asked to predict outputs from unseen data. To counter this, it is common to separate the dataset into training and test sets. This reduces the chances of producing a descriptive model rather than an inferential model that is able to predict, in this case, the probability of labial harmony for the larger population. Accordingly, I randomly assigned 25% of the dataset to test, and the remaining 75% was used for training. After fitting the model to the training data, the model was then cross-validated on the unseen test data.

The optimal weights assigned to the training data are shown below in (24).

\begin{equation}
\text{(24) Constraint weights}
\begin{align*}
\text{ALIGN-R(RD,WD)} & : -7.663 \\
\text{IDENT-IO[V/Δ>0.7]} & : -5.94 \\
\text{CRISPEDGE(RD,MORPH)} & : -4.239 \\
\text{IDENT-IO[V]} & : -2.854 \\
*\text{RD} & : -1.486 \\
\end{align*}
\end{equation}

Given the constraint weights in (24), harmony scores for [qʊɫɯn] and [qʊɫʊn] in (25) are -8.318 and -5.826, respectively.

\begin{equation}
\text{(25)}
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\text{/qʊɫɯn/} & \text{ALIGN-R(RD,WD)} & \text{IDENT-IO[V/Δ>0.7]} & \text{CRISPEDGE} & \text{IDENT-IO[V]} & \ast\text{RD} & \text{Harmony} & \text{Predicted probability} & \text{Observed frequency} \\
\hline
\text{Weights} & -7.663 & -5.94 & -4.239 & -2.854 & -1.486 & \text{ } & \text{ } & \text{ } \\
\hline
\text{qʊɫɯn} & -7.663 & *0.894 = -6.85 & -1.468 & \text{ } & -6.85 + \text{ } & -8.318 \rightarrow 0.076 & 0.059 \\
\hline
\text{qʊɫʊn} & -2.854 & \text{ } & \text{-2.972} & \text{ } & \text{-2.972} + \text{-2.854} & -5.826 \rightarrow 0.924 \rightarrow .941 \\
\hline
\end{array}
\end{equation}

Note that the penalty for violating ALIGN-R for [qʊɫɯn] is scaled by 0.894, the scaling factor associated with the perceptual distance between [ʊ] and [ɯ]. The summed number of violations multiplied by the relevant weights equals -8.318 and -5.826 respectively. The exponent of these

\footnote{For more information see the supplemental materials for McPherson & Hayes (2016), found at: http://www.dartmouth.edu/~mcpherson/papers-and-handouts/Harmony_supplements.html}
two harmony scores results in the probabilities: \( P_{q0\mu n} = 0.059 \) and \( P_{q0\nu n} = 0.94 \). These probabilities may be compared to output frequencies, and the difference between attested frequency and predicted probability is the error for each input-output mapping, which is approximately 0.017 in (25) above.\(^{16}\)

In (26) below, output \([οlɪp] \) and \([οlʏp] \) from input \(/οl-p/ \) ‘die-CVB’ are considered. In (26) the disharmonic candidate, \([οlɪp] \), violates the scalar ALIGN-R constraint and *RD while the harmonic candidate, \([οlʏp] \), violates CRISPEDGE while also violating *RD twice. As above, the model fits the data nicely, with only .028 error.

\begin{align*}
(26)
\end{align*}

<table>
<thead>
<tr>
<th>(/οl-p/)</th>
<th>ALIGN-R(RD,WD)</th>
<th>IDENT-IO[RD/Δ&gt;0.7]</th>
<th>CRISP EDGE</th>
<th>IDENT-IO[RD]</th>
<th>*RD</th>
<th>Harmony</th>
<th>Predicted probability</th>
<th>Observed frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted</td>
<td>-7.663</td>
<td>-5.94</td>
<td>-4.239</td>
<td>-2.854</td>
<td>1.486</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>οlɪp</td>
<td>-7.663 *</td>
<td>0.762 = -5.839</td>
<td>-4.239</td>
<td>1.486</td>
<td>1.486 + 8.732 = 7.211</td>
<td>0.472</td>
<td>0.444</td>
<td></td>
</tr>
<tr>
<td>οlʏp</td>
<td>-4.239</td>
<td></td>
<td>-1.486 * 2 violations = -2.972</td>
<td>-4.839 + 2.972 = 1.867</td>
<td>0.528</td>
<td>0.556</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall, the model performed well on both training and test data, averaging 0.0356 and 0.0359 error, respectively. The summed log-likelihood of the model on training and test was -115.705. In Table 6, the log likelihoods of the conditional alignment model compared to various implementations of a scaled model are shown for comparison. Note the range of log likelihoods produced by the various scaling relations tested. The optimal model accounted for the vast majority of variance in the attested data, \( r^2 = .82 \). In Table 6 below log likelihood and Akaike’s Information Criterion (AIC) for each model is presented. AIC is a common statistical measure derived from the log likelihood of a model that penalizes models with more parameters (all of the models compared below used four constraints). Smaller AIC values indicate better model performance. As is evident, a number of scalings (1.5-5:1) produce similar log likelihoods and AIC. In the simple models compared below, models with an AIC of less than approximately 2 do not involve a significant difference in model fit (Burnham & Anderson 2002: chs.3-4).\(^{17}\)

---

\(^{16}\) During the computation one significant idealization was made- only two outputs candidates were evaluated, following Hayes & McPherson (2016). MaxEnt models assign a probability even to harmonically bounded outputs (Jäger & Rosenbach 2006). In a fuller analysis, forms like \([qɯɫɯn] \) and \([qʊɫn] \) would each be given a certain, albeit small, probability of occurrence although these ways to avoid harmony are unattested.

\(^{17}\) Models were additionally tested in which a perceptual second formant, F2’, based on the algorithm in Schwartz et al. (1997:263, but models using F2’ performed poorer than the models used above (optimal scaling at 2.6:1, \( r^2 = .62 \), log likelihood= -120.53, AIC= 249.06).
Table 6: Summed log likelihoods of some of the models tested

<table>
<thead>
<tr>
<th>Model (# of parameters)</th>
<th>Scale</th>
<th>log likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional (6)</td>
<td>N/A</td>
<td>-119.883</td>
<td>251.766</td>
</tr>
<tr>
<td>Scaled (5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinal</td>
<td>-121.663</td>
<td>253.326</td>
<td></td>
</tr>
<tr>
<td>1:1</td>
<td>-149.073</td>
<td>308.146</td>
<td></td>
</tr>
<tr>
<td>1.5:1</td>
<td>-115.737</td>
<td>241.474</td>
<td></td>
</tr>
<tr>
<td>2:1</td>
<td>-115.723</td>
<td>241.446</td>
<td></td>
</tr>
<tr>
<td>2.2:1</td>
<td>-115.723</td>
<td>241.446</td>
<td></td>
</tr>
<tr>
<td>2.3:1</td>
<td>-115.705</td>
<td>241.41</td>
<td></td>
</tr>
<tr>
<td>2.4:1</td>
<td>-115.713</td>
<td>241.426</td>
<td></td>
</tr>
<tr>
<td>3:1</td>
<td>-115.716</td>
<td>241.432</td>
<td></td>
</tr>
<tr>
<td>4:1</td>
<td>-115.732</td>
<td>241.464</td>
<td></td>
</tr>
<tr>
<td>5:1</td>
<td>-115.997</td>
<td>241.994</td>
<td></td>
</tr>
<tr>
<td>6:1</td>
<td>-116.309</td>
<td>242.618</td>
<td></td>
</tr>
<tr>
<td>8:1</td>
<td>-116.829</td>
<td>243.658</td>
<td></td>
</tr>
<tr>
<td>25:1</td>
<td>-118.217</td>
<td>246.434</td>
<td></td>
</tr>
</tbody>
</table>

The scaled models shown in Table 6 above were also compared to a model with ordinal scaling (as in e.g. McPherson & Hayes 2016) and a model using a conditional alignment constraint. The model with ordinal scaling assigned a weight to the alignment constraint based on ranked rather than calculated perceptual distance. For instance, violations to ALIGN-R after [y] were multiplied by 4 while violations to ALIGN-R after [ø] were multiplied by 2, as these numbers reflect the relative rankings of perceptual distance in the vowel space. The ordinal model performed relatively well (log likelihood -121.663, AIC 253.326), but as well as models that referenced actual distances rather than rankings.

The second type of model compared above, the conditional alignment model, motivated harmony over the data by reference to perceptual distance but without scaled violation. A conditional constraint internally references perceptual similarity such that harmony is motivated only for the subset of the data fulfilling the condition imposed constraint-internally. This parallels Kaun’s use of conditional alignment constraints, like ALIGN-R([RD]/[l-BK]), which motivates harmony only when the trigger is [-back]. In the same way, it is possible for a harmony-driving constraint to index the perceptual similarity of a root-vowel contrast, like the following:

(27) Conditional Alignment

ALIGN-R(RD/Δ[m,n]<X,WD)-

align the roundness of root vowel, m, to the right edge of the word if the perceptual distance between round root vowel, m, and its harmonic counterpart, n, is less than x.

The value of Δ was set to 1, but could’ve been set to any value greater between 0.823 and 1.451 to accomplish the same effect. The conditional model performed better than the ordinally-scaled model, (log likelihood -119.883, AIC 251.776). The conditional model included two alignment constraints, a general constraint in addition to the conditional constraint. The general alignment constraint was necessary to compel harmony after [ø], as harmony is attested, albeit infrequently,
after [ə]. The conditional model thus used one more parameter than the scaled models. This accounts for the slight increase in AIC for the conditional model.

In summary, the analysis presented above accounts for gradient and categorical data in Kazakh labial harmony by using a probabilistic Harmonic Grammar. The perceptual distances calculated in §4 are used to develop a phonetically-informed set of constraints that interact with more abstract constraints referencing underlying phonological structure as well as morphological constituency. The interplay of these surface and structural factors provide an empirically accurate model of the variable harmony pattern in colloquial Kazakh. The analysis provides a principled account for why the trigger strength hierarchy -υ > ʊ > ø > ɔ exists in the language. The definition of weakness presented in Kaun (1995) is reinterpreted in dispersionary terms. The analysis also extends current work (e.g. Hsu & Jesney 2015 a,b) on scalar constraints in Harmonic Grammar.

6 Comparing Kaun’s analysis

In this Section I compare the proposed analysis with Kaun’s analysis. Kaun’s (1995, 2004) typology of labial harmony provides a compelling formulation for the general motivations for harmony. However, I show below in a MaxEnt grammar Kaun’s constraint set does not fit the data as the dispersionary model.

Kaun’s typology of labial harmony systems predicts feature-class behavior in harmony. More specifically, she assumes that a set of universal features are crucial for explaining the patterns attested in her survey, specifically the features relevant for Kazakh, [high], and [back]. Specifically, Kaun (2004) uses three types of constraints to model the languages under study: alignment constraints, faithfulness constraints, and markedness constraints. Her markedness constraints are of two varieties: featural and sequential co-occurrence restrictions. She uses two featural markedness constraints, *RoRo, which bans a combination of the features [+round] and [-back], and *RoLo, which militates against the co-occurrence of [+round] and [-high]. Additionally, GESTURAL UNIFORMITY [RD], prohibits sequences of round vowels that disagree in height. As for faithfulness constraints, Kaun (2004) assumes a highly ranked initial-syllable faithfulness constraint (Beckman 1997), but the only faithfulness constraint implemented is DEP(LINK), which bans additional linkages to a given autosegment. Her alignment constraints include a number of conditional variants, like ALIGN-R([RD]/[-BK]) and ALIGN-R([RD]/[-HI]), to capture the typological tendencies for front vowels and non-high vowels to be better harmony triggers.

For Kaun’s analysis, the following set of contrastive features for the Kazakh vowel inventory is necessary. However, recall from §4 that the putative mid and high vowels do not really differ in F1, making the use of the feature [high] less phonetically motivated. Using the feature [long] would maintain the phonetic motivation, but is more generally problematic as a feature. Both possible choices result in the same outcome, making the choice between the two trivial.

(28) Feature specifications for Kaun’s analysis

<table>
<thead>
<tr>
<th></th>
<th>back</th>
<th>+back</th>
</tr>
</thead>
<tbody>
<tr>
<td>-round</td>
<td>round</td>
<td>-round</td>
</tr>
<tr>
<td>+high</td>
<td>ɪ</td>
<td>ʏ</td>
</tr>
<tr>
<td>-high</td>
<td>ë</td>
<td>ø</td>
</tr>
</tbody>
</table>

Kaun’s model, with its conditional alignment constraints, ALIGN-R([RD]/[-BK]), and ALIGN-R([RD]/[-HI]), used alongside an unscaled alignment constraint, performs far poorer than the
dispersionary model presented in §4. Kaun’s model produces a log likelihood of -211.09, which is worse than all of the models compared in Table 6. Her model produces a mean error of .244 and .274 on training and test runs, respectively. Moreover, the correlation between predicted and attested forms is decidedly lower for Kaun’s analysis ($r^2=.29$) compared to the dispersionary analysis ($r^2=.82$). Additionally, Kaun’s analysis requires three harmony-driving constraints, increasing the total number of model parameters to 7. This increase in model complexity is penalized by the model selection criteria used herein. As a result of both model complexity and poorer fit, the AIC of Kaun’s model is 436.18, which is far greater than the 241.41 associated with the optimal model. Despite the complexity of Kaun’s model, her constraint set cannot account for the gradient empirical data presented above as well as the dispersionary constraint set.

In addition to the empirical findings, there is no meaningful theoretical explanation in Kaun’s analysis why [v] is most likely to both trigger harmony and surface via harmony, while [ɔ] is least likely to trigger harmony and surface due to harmony. While it might be argued that the combined force of multiple alignment constraints and the relevant markedness constraint (*ROLO) account for the asymmetric activity of these two vowels, her analysis offers no principled reason as to why this is so. For instance, Kaun’s model cannot distinguish between the actual pair [a]-[ɔ] and a hypothetical pair [v]-[ɔ], although the second pair would be more similar than the first. The influence of vowel height on the perceptibility of the contrast is ignored. In contrast, the dispersion theoretic conception of harmony, by taking into account the acoustic relations between each harmonic pair, elegantly unifies both aspects of [v] and [ɔ]’s participation in the harmony system. The front vowel [v] is least dispersed from its counterpart, making it the most perceptually weak, and accordingly, most aided by harmony. For that same lack of dispersion, [v] is most likely to surface from the underlying form of its counterpart, /ɪ/, because it involves the least perceptible adjustment to an input vowel to satisfy harmony. The opposite holds for the back vowel, [ɔ]. Because [ɔ] is most dispersed from its counterpart it is least aided by harmony, and this difference in perceptual distance is also why it fails to surface via harmony. The system deems it too costly to turn underlying /a/ to [ɔ], and thus prefers faithfulness to harmony in this context.

Therefore, the analysis proposed in §5 is superior to Kaun’s model for Kazakh. In addition to empirical coverage, the dispersionary framework for labial harmony offers a more principled explanation for both the trigger and target asymmetries noted in contemporary Kazakh. In the following section I extend the dispersionary analysis presented above to three labial harmonic languages outside the Turkic language family in order to demonstrate that the analysis presented above has relevance for languages with significantly different harmony patterns.

7 Testing the model

In order to test the general validity of the model proposed above, I used phonetic data from four non-Turkic languages with labial harmony: three dialects of Mongolian, Khalkha, Baarin and Šiliingol, as well as Solon, a Tungusic language. Crucially, all languages show preference for low vowels to participate in harmony rather than high vowels, which aligns more with Kaun’s predictions than the Kazakh data. Below I demonstrate that in each of these four languages dispersion makes the correct predictions, despite the fact that the harmony patterns are quite distinct from the Kazakh pattern.

7.1 Khalkha Mongolian

Labial harmony in Khalkha is triggered by non-high vowels and targets non-high vowels only (Svantesson 1985:318-320; Kaun 1995:48-53), as is demonstrated in (29). In (29a-d), non-high
suffix vowels surface as unrounded, even after high round vowels, shown in (29c,d). In (29e-h), though, non-high round vowels trigger labialization of following non-high suffix vowels.

(29) Khalkha (Svantesson 1985)
   a. jav-laː ‘go-PST’
   b. deːl-eːr ‘coat-INST’
   c. toːlai-gaːr ‘hare-INST’
   d. uz-leː ‘jump-INST’
   e. ɔɾ-ɨː ‘enter-PST’
   f. ɔɾ-ɔː ‘enter-PFV’
   g. og-loː ‘give-PST’
   h. tor-ɔː ‘be born-PFV’
   i. ɔɾ-ʊːl-aː ‘enter-CAUS-PFV’
   j. tor-uːl-eːd ‘be born-CAUS-PFV’
   k. xɔt-iːx ‘town-REFL.GEN’
   l. tomr-iːxoː ‘iron-REFL.GEN’

The high unrounded vowel, [i], is transparent to harmony, shown in (29k,l), but the high round vowels, [u] and [ʊ], block harmony in Khalkha, as in (29i,j). As the high round vowels do not trigger harmony (29c,d) nor do they surface via harmony from underlying /i/, the perceptual distance between [u] and [i] should be greater than the perceptual distance between [e] and [o] and between [a] and [a].

Svantesson (1985:290-291) recorded three repetitions of each long phonemic vowel in initial position from one Khalkha speaker. Using the procedures detailed in §4.2.2, I calculated perceptual distances between each relevant vowel pair from F1-F2 measurements presented in Svantesson (1985).

Mean normalized F1-F2 and corresponding perceptual distances for each harmonic pair are shown in Table 7. Since non-high vowels trigger and surface via harmony, the distances between [a]-[ɔ] and [e]-[o] are expected to be less than the distances between [i]-[ʊ] and [i]-[u].

Table 7: Mean F1-F2 (ERB) and corresponding perceptual distances in Khalkha

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>F2</th>
<th>Harmonic Pairing</th>
<th>Perceptual Distance ($\lambda(F2)=0.473$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>13.604</td>
<td>17.827</td>
<td>a-ɔ</td>
<td>2.196</td>
</tr>
<tr>
<td>ɔ</td>
<td>11.69</td>
<td>15.549</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>9.919</td>
<td>18.429</td>
<td>e-ɔ</td>
<td>1.135</td>
</tr>
<tr>
<td>ɔ</td>
<td>9.964</td>
<td>16.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>10.467</td>
<td>20.008</td>
<td>i-ʊ</td>
<td>2.597</td>
</tr>
<tr>
<td>ʊ</td>
<td>10.031</td>
<td>14.591</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>8.39</td>
<td>20.909</td>
<td>i-u</td>
<td>3.361</td>
</tr>
<tr>
<td>u</td>
<td>8.232</td>
<td>13.804</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The perceptual distances between the non-high vowels is less than the distance between the two high vowels. Note that only attested (i.e. ATR and [high] harmonic) pairings were considered.
7.2 Inner Mongolian dialects

In dialects of Mongolian spoken in China, as in Khalkha, non-high vowels trigger harmony on following non-high vowels. However, in these dialects, roughly corresponding to the Shuluun Höh data presented in Kaun (1995:53-58), the vowel inventory is notably larger than the Khalkha inventory. The front vowels /æ/ and /œ/ are phonemic in the language, emerging diachronically from umlauted back vowels. /ø/ is also reported for these varieties of Mongolian, but this vowel was not found during Svantesson’s data collection (1985:290).

One interesting complication to labial harmony in Inner Mongolian dialects is the prohibition of front rounded vowels in suffixes, exemplified in (30) below. In (30a,b) [œ] triggers labial harmony on the instrumental suffix, which always surfaces as a back vowel. However, in (30c,d), harmony does not obtain when the suffix is underlying [-back], as in the comitative suffix. In these cases, disharmony occurs.

(30) Inner Mongolian (Svantesson 1985)
   a. нөөгөр ‘dog-INST’
   b. мөржөр ‘horse-INST’
   c. өд-тә ‘star-COM’
   d. өбс-тә ‘grass-COM’

Svantesson (1985:290-291) presents F1-F3 for tokens produced by one speaker of two Mongolian dialects, Baarin and Šiliingol, spoken in China. As with the Khalkha data above, the same procedures from §4.2.2 were used to calculate perceptual distances. Like in Khalkha, the distances between the non-high round vowels and their harmonic counterparts should be less than the distances between the high vowels and their counterparts. Additionally, the distance between the front round vowel, [œ], and its counterpart, [æ], as well as the distance between [æ] and [ɔ], are expected to be less than the distances between the high vowels. This last prediction, in particular, is relevant because these dialects permit [ɔ] to follow [œ], which is typologically unusual because languages with labial harmony either have only back round vowels, or exhibit palatal harmony. I predict that this unusual sequence is permissible, at least in part, because the distance between these two vowels is relatively small. Mean F1-F2 and perceptual distances are shown in Table 8 below.
Table 8: Mean F1-F2 (ERB) and perceptual distances in Baarin and Šiliingol

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>F2</th>
<th>Harmonic Pairing</th>
<th>Perceptual Distance (λ(F2)=0.496)</th>
<th>F1</th>
<th>F2</th>
<th>Harmonic Pairing</th>
<th>Perceptual Distance (λ(F2)=0.441)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>14.464</td>
<td>17.827</td>
<td>a-ɔ</td>
<td>3.414</td>
<td>13.834</td>
<td>17.329</td>
<td>a-ɔ</td>
<td>2.46</td>
</tr>
<tr>
<td>o</td>
<td>10.208</td>
<td>15.002</td>
<td>o-ə</td>
<td>4.328</td>
<td>10.186</td>
<td>21.455</td>
<td>o-ə</td>
<td>3.023</td>
</tr>
</tbody>
</table>

As with Khalkha, the high vowel pairs are marked by greater perceptual distances, while the non-high vowels are more proximate to one another in the vowel space. The [æ]-[œ] pairing is the least dispersed in both dialects, followed by [ə]-[o], and [a]-[ɔ]. The relative distance between the two high vowel pairings differs by dialect, but this is not significant, as both pairs are the most dispersed in each dialect.

As suggested above, the relationship between [œ] and [ɔ] is particularly interesting. Cross-height harmony is common, but cross-backness harmony is rare, either because the inventory has no front round vowels, or because palatal harmony operates in the language. A few Tungusic languages like Sibe and Sanjiazi Manchu (Li 1996:§5.5-5.6) are known to exhibit similar cross-backness harmony patterns. It seems possible that in both cases, cross-height and cross-backness, that the distance between the initial and subsequent round vowels matters for harmony.

7.3 Solon

In addition to the Mongolian dialects above, Svantesson (1985:296) provides acoustic data from one speaker of Solon, a Tungusic language also spoken in Inner Mongolia. The labial harmony pattern in Solon, which is demonstrated below in (31), is almost identical to that of Khalkha Mongolian. As above, non-high vowels trigger and undergo harmony, while the high vowels, [u] and [ʊ], block harmony.

(31) Solon (Li 1996:104)
   a. olda 'quilt'
   b. ulda 'meat'
   c. əʃxɔ 'fish'
   d. toʃo 'cloud'

Based on the harmony pattern, the non-high vowel pairs are predicted to be less dispersed than the high vowel pairs. When the phonetic data presented by Svantesson was used to calculate
perceptual distances, this prediction was borne out. Table 9 below shows mean F1-F2 and perceptual distances for Solon vowel pairings.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>F2</th>
<th>Harmonic Pairing</th>
<th>Perceptual Distance (λ(F2)=0.298)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>12.329</td>
<td>18.261</td>
<td>a-ɔ</td>
<td>0.904</td>
</tr>
<tr>
<td>ɔ</td>
<td>11.69</td>
<td>16.114</td>
<td>a-ɔ</td>
<td>0.904</td>
</tr>
<tr>
<td>ə</td>
<td>10.403</td>
<td>16.314</td>
<td>a-ɔ</td>
<td>0.904</td>
</tr>
<tr>
<td>o</td>
<td>10.317</td>
<td>14.853</td>
<td>e-o</td>
<td>0.443</td>
</tr>
<tr>
<td>ɪ</td>
<td>10.208</td>
<td>21.468</td>
<td>e-o</td>
<td>0.443</td>
</tr>
<tr>
<td>ʊ</td>
<td>10.295</td>
<td>14.249</td>
<td>e-ʊ</td>
<td>2.151</td>
</tr>
<tr>
<td>i</td>
<td>8.044</td>
<td>22.349</td>
<td>e-ʊ</td>
<td>2.151</td>
</tr>
<tr>
<td>u</td>
<td>8.364</td>
<td>13.353</td>
<td>e-ʊ</td>
<td>2.151</td>
</tr>
</tbody>
</table>

As in the above-discussed languages, the harmonic pairings in Solon confirm the predictions made by the model. The non-high vowels are markedly less dispersed than the high vowel pairs.

In this section I have presented phonetic data from four different languages, all of which show a strikingly different pattern from Kazakh. In all four languages non-high vowels are better triggers for harmony than high vowels, and relatedly, non-high vowels surface via harmony while the high vowels do not. In each case the perceptual distances between harmonic pairings made correct global predictions as to the nature of labial harmony in each language. Thus, the data presented above offers evidence that the dispersionary analysis has cross-linguistic relevance for redefining perceptual weakness in labial harmony.

8 Discussion

In §8.1 I compare this model to other Dispersion Theoretic work in OT. In §8.2 I briefly discuss teleology and dispersion.

8.1 Comparisons with other DT models in Optimality Theory

Two types of Dispersion theoretic analyses have been developed within Optimality Theory—continuous and discretized. Padgett (2004) uses continuous SPACE constraints to analyze vowel reduction in Russian. Furthermore, Padgett (2003) contends that non-contrastive phonetic detail is necessary for analysis, and theoretically, this level of detail provides a restrictive framework for analysis. In contrast, Flemming’s model of Dispersion theory (2002) abstracts phonetic detail into discrete scales (cf. Flemming 2006).

Flemming’s model proposes three conflicting goals: one, to maximize the number of contrasts in the language; two, to maximize distance between those contrasts, and three, to minimize effort. Crucial to Flemming’s analysis is the model of the vowel space presented in Figure 3 above, shown again in Figure 8, which is used to calculate distances between contrastive vowels.
To calculate the distance between a contrastive pair is extremely simple, count the gridlines between each member of the pair for a given acoustic dimension (e.g. F1 or F2). Thus, [e] and [ø] differ in F2 by only one, as do [ɪ] and [ʏ]. Problematically, both the [ɑ]-[ɔ] and [ɯ]-[ʊ] contrasts involve both an F1 and F2 difference of one grid space. In these terms, then, the [ɑ]-[ɔ] contrast is justifiably less likely to initiate harmony than the front vowels, but the propensity of [ʊ] to trigger harmony is unaccounted for. In short, Flemming’s model can only predict that front vowels are more perceptually similar than back vowels, and has nothing more to differentiate these four harmonic pairs. This coarse-grained representation is thus insufficient for capturing the dispersionary generalizations presented herein.

Flemming (2006) argues that incorporating contextual phonetic detail like gestural overlap, and among other things, language-specific coarticulatory effects is desirable for analysis, which renders the original abstraction to a discretized model of the vowel space superfluous. If representations are truly enriched by language-specific detail, there is no need to map language-specific vowels to a universal grid for analysis. This lends credence to a language-specific systemic approach, removing the inconsistency of simultaneously enriched and impoverished representations. Also unlike Flemming, the present work follows Padgett (2004) both in using acoustic detail, and couching dispersion within Optimality Theory’s conception of input-output faithfulness.

8.2 Teleology and dispersion

While this work diverges from Flemming in some significant ways, it fundamentally agrees that perceptual distance is active in the synchronic grammar. Lindblom’s (1986) formulation of DT is couched in evolutionary (i.e. emergent) terms, but Flemming’s use of dispersion theory assumes speaker access to phonetic detail that diverges in some ways from Lindblom’s original proposal. If viewed along the hyper-hypoarticulation continuum (Lindblom 1990; Lindblom et al. 1995), then a more dispersed vowel space is one means of clearer communication. In fact, dispersed vowel spaces are often reported for infant-directed speech (Ratner 1984; Kuhl et al. 1997).

Flemming’s position, which is the position taken here, namely that speakers access dispersion synchronically, has been contested by various authors (de Boer 2000; Wedel 2006; Boersma & Haman 2008; Hall 2011; Vaux & Samuels 2015). Both Hall (2011) and emergentist works (de Boer 2000; Wedel 2006; Boersma & Haman 2008; Vaux & Samuels 2015) object to the teleological underpinnings of Flemming’s MiNDist and Padgett’s Space constraints, instead contending for non-phonological factors in the development of vowel spacing. Interestingly, Kaun’s analysis of

---

18 Flemming’s vowel space is not always uniform. Compare the grid in Flemming (2006:4) with the one above, which leaves [ui] out and further distinguishes [ɔ] from [ɑ] via an additional height distinction. The need to modify the putative universal vowel space reflects the coarse granularity of his model, and moreover, the ability to modify the model is also problematic for a restrictive theory of dispersion.
labial harmony is similarly teleological, assuming that speakers initiate phonological harmony in order to aid weak vowel contrasts. It is perhaps unsurprising then that these Optimality theoretic interpretations of DT are compatible with Kaun’s weakness-driven harmony analysis.

It is important to note that even if the teleological underpinnings of the present analysis are discarded, the analysis is in no way undermined, because the analysis does not depend on synchronic access to the details of the vowel space, but uses this as indicative without a necessarily causal relationship between dispersion and harmonic activity. If, however, vowel dispersion was a consequence and not the cause the asymmetries in labial harmony outlined in §3.2, it is unclear what would drive these asymmetries. For this reason it seems most plausible to maintain that speakers have synchronic access to dispersion in addition to featural representations.

9 Conclusion

In this paper I have presented new data from colloquial Kazakh showing a variable labial harmony pattern. I have argued, in agreement with Kaun (1995, 2004), that labial harmony is motivated by perceptual weakness, but that perceptual weakness must be defined in dispersionary rather than featural terms. The central claim of this paper is that in labial harmony perceptual weakness is linked to a weighted Euclidean distance between a round vowel and its unrounded counterpart. I presented a novel framework through which to articulate Kaun’s general observation, recasting the various predictions made in systemic rather than atomistic vowel-intrinsic terms. I demonstrated that dispersionary constraints outperform Kaun’s featural constraints in MaxEnt grammar. To further evaluate the predictions made by the model, I examined four other languages, finding that in each case the dispersion-based predictions are borne out in each language’s labial harmony pattern. By using phonetic data in addition to impressionistic judgments this paper quantifies both the rate of harmony and its dispersionary motivation for both categorical and gradient harmony in Kazakh. The possibility to use this unified approach to phonetic and phonological assimilation offers a testable means by which to both evaluate the claims made herein and to further refine our understanding of vowel harmony.

References


Vowel Dispersion and Kazakh Labial Harmony


Hayes, Bruce, Robert Kirchner, and Donca Steriade, (eds.) (2004). Phonetically based phonology. Cambridge University Press.


Vowel Dispersion and Kazakh Labial Harmony


Adam G. McCollum
University of California, San Diego
agmccoll@ucsd.edu