Optimizing over subsequences generates context-sensitive languages

Andrew Lamont
University of Massachusetts Amherst
alamont@linguist.umass.edu

Abstract

Phonological generalizations are finite-state. While Optimality Theory is a popular framework for modeling phonology, it is known to generate non-finite-state mappings and languages. This paper demonstrates that Optimality Theory is capable of generating non-context-free languages, contributing to the characterization of its generative capacity. This is achieved with minimal modification to the theory as it is standardly employed.

1 Introduction

Phonological generalizations are finite-state (Johnson, 1972; Kaplan and Kay, 1994; see Heinz, 2018, for a recent overview); that is, input-output mappings can be modeled using finite-state transducers and phonotactic well-formedness can be modeled using finite-state acceptors.

Optimality Theory (OT; Prince and Smolensky, 1993/2004) is a framework that is commonly used to model phonology. While some restricted variants of OT are finite-state (Frank and Satta, 1998; Eisner, 2000, 2002; Riggle, 2004; see Hulden, 2017, for a recent overview), standard OT, as it is employed by practicing phonologists, is known to generate non-finite-state mappings and languages (Eisner, 1997; Frank and Satta, 1998).

OT is a special instance of Harmonic Grammar (Legendre et al., 1990), which can model arbitrary computations (Smolensky, 1992). While the exact generative capacity of OT has not yet been characterized, it has recently been shown to produce non-context-free mappings (Lamont, 2019a,b). This paper contributes to the literature on OT by demonstrating its capacity to generate non-context-free languages using constraints defined over subsequences. Subsequences are finite literals composed of ordered symbols that are not necessarily adjacent.[1] They contrast with

---

[1] Trivially, all strings of length 1 are subsequences. In this paper, *subsequences* refer to subsequences of length ≥ 2.

substrings, whose constituent elements are contiguous. Figure 1 illustrates the subsequences of length 2 in the string *example*. Of these twenty-one subsequences, six are also substrings of length 2: *e. . . x, x. . . a, a. . . m, m. . . p, p. . . l, l. . . e.*

In the literature on phonotactics as formal languages, subsequences have been used to model non-local phenomena (Heinz, 2007, 2010, 2014; Rogers et al., 2010; Graf, 2017). For example, if a language disallows words from surfacing with more than one lateral consonant, it can be modeled as banning the subsequence *l. . . l*. Languages defined by banning a finite set of subsequences belong to the Strictly Piecewise languages, which are properly contained within the class of regular languages.

Strictly Piecewise languages impose inviolable constraints on subsequences: A string belongs to a language if and only if it does not contain any banned subsequences. In OT, all constraints are violable, and violations are minimized whenever possible. Consequently, in addition to modeling non-local restrictions, constraints on subsequences are often used to minimize the distance between two objects (McCarthy and Prince, 1993; Hyde, 2012, 2016). For example, Figure 2 illustrates a string of syllables, represented as σ, that belong to a prosodic word, whose edges are marked with square brackets. Two syllables are parsed into a foot, indicated by parentheses. The number of $) \ldots \sigma \ldots]$ subsequences indicates how far the foot is from the right edge of the prosodic word, calculated over intervening syllables.

When only one foot is parsed, aligning it to the right edge of the prosodic word eliminates $) \ldots \sigma \ldots]$ subsequences: $[\sigma\sigma\sigma\sigma(\sigma\sigma)]$. However, they are unavoidable when multiple feet are parsed. In these cases, the pressure to minimize $) \ldots \sigma \ldots]$ subsequences determines the prosodification. For example, Table 1 illustrates four parses of a seven syllable string that contain three disyllabic feet and one monosyllabic foot. The position of the monosyllabic foot affects the total number of $) \ldots \sigma \ldots]$ subsequences.
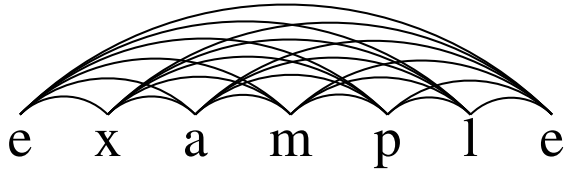
528

Figure 1: The string *example* contains twenty-one subsequences of length 2: *e. . . x, e. . . a, e. . . m, e. . . p, e. . . l, e. . . e, x. . . a, x. . . m, x. . . p, x. . . l, x. . . e, a. . . m, a. . . p, a. . . l, a. . . e, m. . . p, m. . . l, m. . . e, p. . . l, p. . . e, l. . . e.*
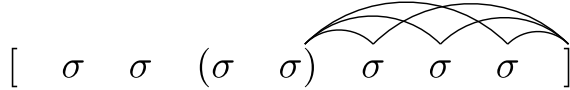

Figure 2: Minimizing the number of $)\ldots\sigma\ldots]$ subsequences minimizes the distance between the right edge of the foot and the right edge of the prosodic word.

It is not possible to replicate these effects with inviolable constraints on subsequences. Formally, this is because some strings that are not in the language are themselves subsequences of strings that are in the language. For example, to ban the string $*[(\sigma)(\sigma\sigma)]$, one must ban one of its subsequences. However, because $*[(\sigma)(\sigma\sigma)]$ is itself a subsequence of $[(\sigma\sigma)(\sigma\sigma)]$, the latter contains every subsequence of the former, and $*[(\sigma)(\sigma\sigma)]$ cannot be banned without incorrectly banning $[(\sigma\sigma)(\sigma\sigma)]$. This example illustrates that violable constraints on subsequences in OT generate languages more expressive than Strictly Piecewise languages. Along similar lines, Koser and Jardine (2020) demonstrate that violable constraints on substrings in OT are more expressive than inviolable constraints.

Optimization and violability contribute much more expressivity than these results suggest. Eisner (1997, 2000) demonstrated that OT can generate context-free languages with subsequence constraints, and this paper pushes his result into non-context-free languages. The main result of this paper has wide-reaching consequences for phonologists, as many standard constraint families are defined over subsequences. Examples include alignment constraints (McCarthy and Prince, 1993; Hyde, 2012, 2016), conjoined constraints (Smolensky, 1993, 2006; Alderete, 1997), co-occurrence constraints (Suzuki, 1998; Pulleyblank, 2002), SHARE constraints (McCarthy, 2010; Mullin, 2011), and the family of surface corre-

Parse	Total
$[(\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	12
$[(\sigma\sigma)(\sigma)(\sigma\sigma)(\sigma\sigma)]$	11
$[(\sigma\sigma)(\sigma\sigma)(\sigma)(\sigma\sigma)]$	10
$[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	9

Table 1: Right-aligning a monosyllabic foot in odd-parity syllable strings minimizes the number of $)\ldots\sigma\ldots]$ subsequences.

spondence constraints (Walker, 2000; Hansson, 2001, 2010; Rose and Walker, 2004; Bennett, 2013, 2015).

Eisner's result, that OT with subsequence constraints generates context-free languages, is presented in section 2. It further argues that restricting the set of subsequence constraints available to OT does not limit its generative capacity. This paper's contribution, that OT generates context-sensitive languages with constraints on subsequences, is presented in section 3. The result is illustrated with case studies on prosodic parsing and non-local dissimilation, and a proof of the general case is provided.

2 Violable Subsequence Constraints Can Divide Strings into Halves

Eisner (1997, 2000) demonstrated that with violable constraints on subsequences, mappings in Optimality Theory can target the centers of strings. In his example, the Midpoint Pathology, a feature in the input shifts so as to surface in the center of the output. Taking stress as that feature, the Midpoint Pathology is defined in (1) as an input-output mapping, where σ represents a syllable, and $\acute{\sigma}$ represents a stressed syllable. The output language is homomorphic to the archetypal non-finite-state language $a^n b^n$, with the allowance of an additional a or b.

(1) $F_{\text{midpoint}} : \sigma^i \acute{\sigma} \sigma^j \mapsto \sigma^k \acute{\sigma} \sigma^l$ where
$i + j = k + l$ and $|k - l| \le 1$

In OT, a set of candidates is generated from an input string, and evaluated by a ranked set of constraints. Constraints are functions that map candidates onto a number of violation marks, assigning as many violations as there are specific structures in the candidate or specific changes made to the input to generate the candidate. The candidate that is lexicographically minimal in its

/ σ́σσσσσσ /	ALIGN(σ, $\acute{\sigma}$, σ)	IDENT(stress)
a. $\acute{\sigma}\sigma\sigma\sigma\sigma\sigma$	$0 + 0 + 1 + 2 + 3 + 4 + 5 = 15$	
b. $\sigma\acute{\sigma}\sigma\sigma\sigma\sigma$	$0 + 0 + 0 + 1 + 2 + 3 + 4 = 10$	2
c. $\sigma\sigma\acute{\sigma}\sigma\sigma\sigma$	$1 + 0 + 0 + 0 + 1 + 2 + 3 = 7$	2
→ d. $\sigma\sigma\sigma\acute{\sigma}\sigma\sigma$	$2 + 1 + 0 + 0 + 0 + 1 + 2 = 6$	2
e. $\sigma\sigma\sigma\sigma\acute{\sigma}\sigma$	$3 + 2 + 1 + 0 + 0 + 0 + 1 = 7$	2
f. $\sigma\sigma\sigma\sigma\sigma\acute{\sigma}\sigma$	$4 + 3 + 2 + 1 + 0 + 0 + 0 = 10$	2
g. $\sigma\sigma\sigma\sigma\sigma\sigma\acute{\sigma}$	$5 + 4 + 3 + 2 + 1 + 0 + 0 = 15$	2

Table 2: The Midpoint Pathology (Eisner, 1997, 2000): Stress shifts onto the middle syllable in odd-parity words, and onto either of two middle syllables in even-parity words. Violations of ALIGN(σ, $\acute{\sigma}$, σ) are split up by syllable.

concatenated violations is returned as output. In the Midpoint Pathology, stress shifts to minimize the violations of the alignment constraint (McCarthy and Prince, 1993; McCarthy, 2003) defined in (2), which assigns a candidate as many violations as $\acute{\sigma} \ldots \sigma \ldots \sigma$ and $\sigma \ldots \sigma \ldots \acute{\sigma}$ subsequences it contains.

(2)   ALIGN(σ, $\acute{\sigma}$, σ): For every syllable σ, if there is a stressed syllable $\acute{\sigma}$, assign one violation mark for every syllable that intervenes between σ and $\acute{\sigma}$.

The tableau in Table 2 illustrates stress shifting onto the middle syllable of a seven syllable string. The candidate set contains the input string / σ́σσσσσσ / (2a) and the six candidates derived from it by shifting the stress to another syllable (2b-g). The middle column shows the number of violations ALIGN(σ, $\acute{\sigma}$, σ) assigns each candidate; for clarity, the violations incurred by each syllable are shown separately. Violations of ALIGN(σ, $\acute{\sigma}$, σ) decrease as stress approaches the center syllable. For completeness, the rightmost column shows the number of violations assigned by the constraint IDENT(stress), which penalizes syllables whose stress value in the input was changed. The ordering between constraints indicates that ALIGN(σ, $\acute{\sigma}$, σ) is ranked above IDENT(stress). The candidate with medial stress (2d) is returned as output because its violation vector $(6, 2)$ is lexicographically minimal. If IDENT(stress) were ranked above ALIGN(σ, $\acute{\sigma}$, σ), candidate (2a) would be returned as output.

As this example demonstrates, shifting stress onto the medial syllable minimizes the violations of ALIGN(σ, $\acute{\sigma}$, σ); see section 3 for a proof that this holds for any length input. Hyde (3; 2008; 2012; 2016) argues that the Midpoint Pathology

/σ́σσσσσσ/	*$\acute{\sigma} \ldots \sigma \ldots \sigma$	ID(stress)
a. $\acute{\sigma}\sigma\sigma\sigma\sigma\sigma$	15	
b. $\sigma\acute{\sigma}\sigma\sigma\sigma\sigma$	10	2
c. $\sigma\sigma\acute{\sigma}\sigma\sigma\sigma$	6	2
d. $\sigma\sigma\sigma\acute{\sigma}\sigma\sigma$	3	2
e. $\sigma\sigma\sigma\sigma\acute{\sigma}\sigma$	1	2
→ f. $\sigma\sigma\sigma\sigma\sigma\acute{\sigma}\sigma$	0	2
→ g. $\sigma\sigma\sigma\sigma\sigma\sigma\acute{\sigma}$	0	2

Table 3: Asymmetric alignment does not motivate the Midpoint Pathology.

is an artifact of the symmetrical nature of ALIGN(σ, $\acute{\sigma}$, σ). In particular, if the constraint penalized only $\acute{\sigma} \ldots \sigma \ldots \sigma$ or $\sigma \ldots \sigma \ldots \acute{\sigma}$ subsequences and not both, then stress would be drawn to one edge rather than the center. The tableau in Table 3 illustrates this. Here, only $\acute{\sigma} \ldots \sigma \ldots \sigma$ subsequences are penalized, and stress is drawn to the right edge, surfacing on either of the last two syllables (3f-g).

While asymmetrical constraints avoid the Midpoint Pathology specifically, they motivate other mappings that target the centers of strings. For example, the constraint ALLFEET-RIGHT (Hyde, 2008, 2012, 2016) penalizes $) \ldots \sigma$ subsequences that occur within a prosodic word. By restricting its application to a given prosodic word, ALLFEET-RIGHT is similar to but distinct from penalizing $) \ldots \sigma \ldots]$ subsequences. As discussed in section 1, this constraint pulls monosyllabic feet to the right edge of prosodic words, and as the tableau in Table 4 illustrates, it also balances the size of multiple prosodic words. All the candidates in this tableau are parsed into two prosodic words and are exhaustively footed. To save space, constraints that enforce these conditions are omitted. The

$/\sigma\sigma\sigma\sigma\sigma\sigma\sigma/$	AllFt-R
a. $[(\sigma)]\,[(\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+12=12$
b. $[(\sigma)]\,[(\sigma\sigma)(\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+11=11$
c. $[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)(\sigma\sigma)]$	$0+10=10$
d. $[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+9=9$
e. $[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+6=6$
f. $[(\sigma)(\sigma\sigma)]\,[(\sigma)(\sigma\sigma)(\sigma\sigma)]$	$2+6=8$
g. $[(\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma)(\sigma\sigma)]$	$2+5=7$
h. $[(\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$2+4=6$
i. $[(\sigma\sigma)(\sigma)]\,[(\sigma)(\sigma\sigma)(\sigma\sigma)]$	$1+6=7$
j. $[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma)(\sigma\sigma)]$	$1+5=6$
k. $[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$1+4=5$
$\rightarrow$ l. $[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]$	$2+2=4$

Table 4: AllFeet-Right prefers that when two prosodic words are parsed, they are balanced in size. Violations are separated by prosodic word.

violations of AllFeet-Right decrease as the difference in size between the two prosodic words decreases, and candidate (4l) is returned as output. In practice, prosodic words typically reflect morphosyntactic structure. However, because the constraints that enforce such correspondences are violable (such as the family of Match constraints; Selkirk, 2011), mappings like those illustrated in Table 4 are predicted to be possible.

(3) AllFeet-Right: For every foot in a prosodic word, assign one violation for every syllable it precedes within the same prosodic word.

The balanced prosodic word mapping is defined in (4). Like the Midpoint Pathology, its application depends on identifying the center syllable, and its output language is homomorphic to $a^n b^n$ with an extra a or b. Thus, even though asymmetric alignment cannot target the center of a prosodic word, it can target the center of a string parsed into two prosodic words.

(4) $F_{\text{balanced}} : \sigma^i \mapsto [(\sigma\sigma)^j(\sigma)^k][(\sigma\sigma)^l(\sigma)^m]$
    where $i = j + k + l + m$, $k \leq 1$, $m \leq 1$,
    and $j = l$

The mappings in this section depend on the same underlying mechanism: divide a string of syllables into two parts, and minimize the subsequences of at least length 2 that occupy each part. This reduces the difference in size between the two parts to at most one syllable. With the Midpoint Pathology, the two parts are defined by syllables
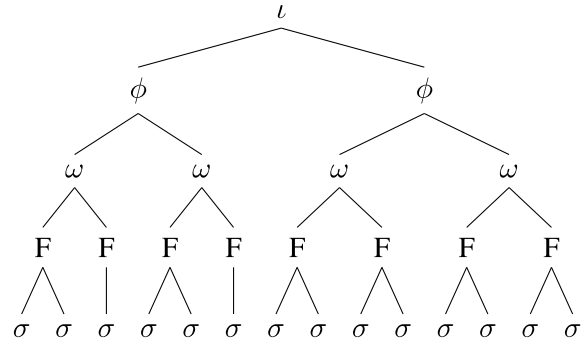

Figure 3: Hierarchical prosodic structure: syllables (σ) are parsed into feet (F), which are parsed into prosodic words (ω), which are parsed into phonological phrases (ϕ), which are parsed into an intonational phrase (ι).

that precede the stressed syllable and syllables that follow the stressed syllable. With balanced prosodic words, the two prosodic words define the parts. This mechanism is independent of the subsequences themselves, provided that they are at least of length 2. Thus, restricting which subsequences a constraint can penalize can only block specific mappings. No restrictions on the set of subsequences prevents them from dividing strings in half. This is proved formally at the end of section 3. The mappings in this section only divided strings into two equal parts, generating context-free languages. The next section presents mappings that divide strings into three or more equal parts, generating non-context-free languages.

3 Violable Subsequence Constraints Can Divide Strings into Arbitrarily Many Equally Sized Parts

As the previous section demonstrated, subsequence constraints can be used to divide strings into two equal parts, generating context-free languages. Parsing strings into more than two equal parts generates non-context-free languages. With balanced prosodic words, this follows from parsing a string of syllables into more than two prosodic words. This can be motivated by hierarchical prosodic structure (Nespor and Vogel, 1986; Selkirk, 1984), rather than stipulated arbitrarily. Figure 3 illustrates a standard five-level prosodic hierarchy, where prosodic words are dominated by phonological phrases, which are dominated by an intonational phrase. By requiring these top two levels to dominate exactly two daughter nodes, the string of syllables is parsed

531

$/\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma/$	ALLFEET-RIGHT
a. $[(\sigma)]\,[(\sigma)]\,[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+0+25=25$
b. $[(\sigma)]\,[(\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+0+20=20$
c. $[(\sigma)]\,[(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+1+16=17$
d. $[(\sigma)]\,[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+2+12=14$
e. $[(\sigma)]\,[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+4+9=13$
f. $[(\sigma)]\,[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+6+6=12$
g. $[(\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+0+16=16$
h. $[(\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+1+12=13$
i. $[(\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+2+9=11$
j. $[(\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+4+6=10$
k. $[(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+1+1+9=11$
l. $[(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+1+2+6=9$
m. $[(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+1+4+4=9$
n. $[(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+2+2+4=8$
o. $[(\sigma\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+0+12=12$
p. $[(\sigma\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+1+9=10$
q. $[(\sigma\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+0+2+6=8$
r. $[(\sigma\sigma)]\,[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+0+4+4=8$
s. $[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma\sigma)]$	$0+1+1+6=8$
t. $[(\sigma\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$0+1+2+4=7$
u. $[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)(\sigma)]$	$1+1+1+4=7$
$\rightarrow$ v. $[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]\,[(\sigma\sigma)(\sigma\sigma)]$	$1+1+2+2=6$

Table 5: ALLFEET-RIGHT prefers that when four prosodic words are parsed, they are balanced in size. Violations are split up by prosodic word.

into four prosodic words. Note that because this hierarchy is finite, it can be represented by a finite-state grammar (Yu, 2019).

As expected, ALLFEET-RIGHT has two effects on strings parsed into four prosodic words. First, in odd-parity prosodic words, the monosyllabic foot appears at the right edge. Second, no two prosodic words differ in size by more than one syllable. The tableau in Table 5 illustrates these effects with a fourteen syllable string. A large number of candidates are omitted to reduce the size of this tableau. These include candidates with more or fewer than four prosodic words, and candidates where monosyllabic feet do not surface at the right edge of their prosodic word. It can be verified that those candidates incur more violations of ALLFEET-RIGHT, as in Tables 1 and 4. The other omitted candidates are identical to the presented candidates, except with their prosodic words in another order. For example, the candidate chosen as output (5v) represents a set containing six possible parses, which incur the same number of violations of ALLFEET-RIGHT.

This quartering mapping is defined in (5). The language it generates is homomorphic to the context-sensitive stringset $a^n b^n c^n d^n$ with allowances for an extra a, b, c, or d.

(5) $F_{\text{quarter}} : \sigma^i \mapsto$
$[(\sigma\sigma)^j(\sigma)^k][(\sigma\sigma)^l(\sigma)^m][(\sigma\sigma)^n(\sigma)^o][(\sigma\sigma)^p(\sigma)^q]$
where $i = j + k + l + m + n + o + p + q$,
$k \leq 1, m \leq 1, o \leq 1, q \leq 1$, and $j = l = n = p$

As noted in the previous section, ALLFEET-RIGHT is distinct from a constraint that penalizes $)\ldots\sigma$ subsequences. In particular, because it is restricted to subsequences within prosodic words, it undercounts when multiple prosodic words are parsed, as Figure 4 illustrates. The next case study demonstrates that this property is irrelevant to the generative capacity of OT with subsequence constraints.

To that end, consider the constraint *X...X defined in (6). *X...X is a non-local variant of *GEMINATE that penalizes subsequences of length 2 whose constituent segments are identical. It has not been proposed as a serious phonological
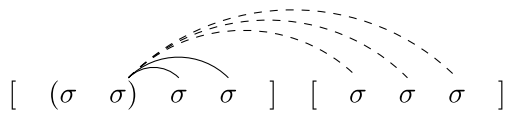
Figure 4: ALLFEET-RIGHT only penalizes the two $)\ldots\sigma$ subsequences indicated with solid lines, and not the three indicated with dashed lines.

/llllllll/	*X...X	IDENT(lateral)
a. llllllll	$28 + 0 = 28$	
b. lllllllr	$21 + 0 = 21$	1
c. llllllrr	$15 + 1 = 16$	2
d. lllllrrr	$10 + 3 = 13$	3
→ e. llllrrrr	$6 + 6 = 12$	4

Table 6: *X. . . X prefers that strings that contain both laterals and rhotics have equal numbers of them. Violations are separated into laterals and rhotics.

constraint, but rather as supporting evidence for this paper's result. Unlike ALLFEET-RIGHT, there are no restrictions on which subsequences it evaluates.

(6) *X. . . X: Assign one violation for every subsequence $\alpha\ldots\beta$ where $\alpha = \beta$.

The tableau in Table 6 illustrates the effect of this constraint on non-local liquid dissimilation. In this example, the class of liquid consonants comprises only alveolar laterals and rhotics, and the constraint IDENT(lateral) penalizes changing one into the other. The input contains eight laterals (6a). The candidates shown are derived by changing underlying laterals into rhotics (6b-e). As in Table 5, permutations of these strings incur equal numbers of violations. This tableau demonstrates that as the difference in the number of laterals and rhotics decreases, so does the number of violations of *X. . . X. Any string with four laterals and four rhotics is returned as output (6e).

This mapping is defined in (7). Its output language is context-free, and homomorphic to permutations of $a^n b^n$ with the usual allowance of an additional a or b.

(7) $F_{\text{liquid}} : \{l,r\}^i \mapsto \{l,r\}^i$ where the difference between the number of laterals and rhotics is not greater than 1.

To generate a context-sensitive stringset with this constraint, one just has to consider a third

/ppppppppp/	*X...X	IDENT(place)
a. ppppppppp	$36 + 0 + 0 = 36$	
b. pppppppt	$28 + 0 + 0 = 28$	1
c. ppppppptt	$21 + 1 + 0 = 22$	2
d. pppppptttt	$15 + 3 + 0 = 18$	3
e. pppptttt	$10 + 6 + 0 = 16$	4
f. ppppppptk	$21 + 0 + 0 = 21$	2
g. ppppptkk	$15 + 0 + 1 = 16$	3
h. pppptkkk	$10 + 0 + 3 = 13$	4
i. pppptkkkk	$6 + 0 + 6 = 12$	5
j. ppppttkk	$10 + 1 + 1 = 12$	4
k. ppppttkkk	$6 + 1 + 3 = 10$	5
→ l. ppptttkkk	$3 + 3 + 3 = 9$	6

Table 7: *X. . . X prefers that strings that contain labials, coronals, and dorsals have equal numbers of stops at each place of articulation. Violations are separated into labials, coronals, and dorsals.

segment type. The tableau in Table 7 on the next page illustrates this case with dissimilation targeting major place. Here, the segment inventory comprises voiceless stops specified as labial, coronal, or dorsal. The input is a string of nine labial stops (7a), and candidates derived from it by changing the place features are shown (7b-l). The constraint IDENT(place) penalizes changes made to major place features. Unsurprisingly, as the differences between the numbers of each stop decrease, so do the violations of *X. . . X. The output is any string with three labial stops, three coronal stops, and three dorsal stops (7l).

The mapping is defined in (8); it generates a language homomorphic to permutations of $a^n b^n c^n$, with one additional a, b, or c.

(8) $F_{\text{place}} : \{p,t,k\}^i \mapsto \{p,t,k\}^i$ where the difference between any two sets of stops defined by place is not greater than 1.

*X. . . X has the same effect as constraints like ALLFEET-RIGHT: it divides the string into a fixed number of parts, and requires those parts be as similar to each other in size as possible by penalizing their subsequences. *X. . . X differs only in that it does not require that the subparts form contiguous substrings.

The mappings in this section demonstrated constraints over subsequences dividing strings into a fixed number of groups of equal size. In all cases, if the groups were not exactly equal, they could differ by at most one element. This generalizes to strings of all lengths and all

533

numbers of groups: minimizing the difference between group sizes minimizes the number of violating subsequences. Before presenting a proof of this result, it is necessary to establish that subsequences of length 2 grow quadratically in a string's length, in particular, a string of length n has $\frac{n^2-n}{2}$ subsequences of length 2. As a base case, a string of length 2 has 1 such subsequence: $\frac{2^2-2}{2} = 1$. Inductively, assume that a string of length n has $\frac{n^2-n}{2}$ subsequences of length 2. Adding one segment adds n subsequences of length 2, and it can be verified that $\frac{n^2-n}{2} + n = \frac{(n+1)^2 - (n+1)}{2}$. The proof of the main result of this paper follows.

Proof. For a string s composed of k disjoint sets, let $s_1, s_2, \ldots, s_k$ denote the cardinality of each set, and define the constraint function $M : s \to \mathbb{N}$ as

$$M(s) = \sum_{i=1}^{k} \frac{s_i^2 - s_i}{2}$$

Let u be a string composed of k disjoint sets such that one set has two more members than another:

$$u_i \geq u_j + 2$$

By way of contradiction, assume that the composition u minimizes the function M.

Consider an alternate composition v, identical to u except that one element has been moved from the larger set to the smaller set:

$$v_i = u_i - 1$$
$$v_j = u_j + 1$$

By assumption, we have $M(u) < M(v)$, from which we derive a contradiction:

$$\frac{u_i^2 - u_i}{2} + \frac{u_j^2 - u_j}{2} < \frac{v_i^2 - v_i}{2} + \frac{v_j^2 - v_j}{2}$$

$$u_i^2 - u_i + u_j^2 - u_j < v_i^2 - v_i + v_j^2 - v_j$$

$$u_i^2 - u_i + u_j^2 - u_j < (u_i - 1)^2 - (u_i - 1)$$

$$+ (u_j + 1)^2 - (u_j + 1)$$

$$u_i^2 - u_i + u_j^2 - u_j < u_i^2 - 3u_i + u_j^2 + u_j + 2$$

$$2u_i < 2u_j + 2$$

$$u_i < u_j + 1$$

This contradicts the fact that u_i is at least as great as $u_j + 2$. Therefore, no composition of a string whose component sets differ in cardinality by two or more minimizes M. $\square$

This proves that violable constraints in Optimality Theory that penalize subsequences of length 2 can divide any length input into a fixed number of equally sized parts, generating context-sensitive stringsets.

4 Conclusion

Optimality Theory is known to generate non-finite state mappings and languages (Eisner, 1997; Frank and Satta, 1998), even with constraints defined over strings (Riggle, 2004; Gerdemann and Hulden, 2012; Heinz and Lai, 2013; Hao, 2019; Lamont, 2019a,b). This paper contributes to this literature by demonstrating that constraints over subsequences generate context-sensitive languages under optimization. This result has wide-ranging impacts for the field of phonology, as a number of commonly employed constraint types are defined over subsequences.

Acknowledgments

References

John Alderete. 1997. Dissimilation as local conjunction. In Kiyomi Kusumoto, editor, *Proceedings of NELS 27*, pages 17–32. Graduate Linguistics Student Association, Amherst, MA.

Wm. G. Bennett. 2013. *Dissimilation, Consonant Harmony, and Surface Correspondence*. Ph.D. thesis, Rutgers, The State University of New Jersey.

Wm. G. Bennett. 2015. *The Phonology of Consonants*. Cambridge University Press, Cambridge.

Jason Eisner. 1997. What constraints should OT allow? Paper presented at LSA 71. Available at `http://roa.rutgers.edu/article/view/215`

Jason Eisner. 2000. Directional constraint evaluation in Optimality Theory. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 257–263. **DOI:** `https://doi.org/10.3115/990820.990858`

Jason Eisner. 2002. Comprehension and compilation in Optimality Theory. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 56–63. **DOI:** `https://doi.org/10.3115/1073083.1073095`

Robert Frank and Giorgia Satta. 1998. Optimality Theory and the generative complexity of constraint violability. *Computational Linguistics*, 24(2):307–315.

Dale Gerdemann and Mans Hulden. 2012. Practical finite state Optimality Theory. In *Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing*, pages 10–19.

Thomas Graf. 2017. The power of locality domains in phonology. *Phonology*, 34(2): 385–405. **DOI:** `https://doi.org/10.1017/S0952675717000197`

Gunnar Ólafur Hansson. 2001. *Theoretical and Typological Issues in Consonant Harmony*. Ph.D. thesis, University of California, Berkeley.

Gunnar Ólafur Hansson. 2010. *Consonant Harmony: Long-Distance Interactions in Phonology*. University of California Press, Berkeley, CA.

Yiding Hao. 2019. Finite-state Optimality Theory: Non-rationality of Harmonic Serialism. *Journal of Language Modeling*, 7(2):49–99. **DOI:** `https://doi.org/10.15398/jlm.v7i2.210`

Jeffrey Heinz. 2007. *Inductive Learning of Phonotactic Patterns*. Ph.D. thesis, University of California Los Angeles.

Jeffrey Heinz. 2010. Learning long-distance phonotactics. *Linguistic Inquiry*, 41:623–661. **DOI:** `https://doi.org/10.1162/LING_a_00015`

Jeffrey Heinz. 2014. Culminativity times harmony equals unbounded stress. In Harry van der Hulst, editor, *Word Stress: Theoretical and Typological Issues*, pages 255–275. Cambridge University Press, Cambridge. **DOI:** `https://doi.org/10.1017/CBO9781139600408.012`

Jeffrey Heinz. 2018. The computational nature of phonological generalizations. In Larry M. Hyman and Frans Plank, editors, *Phonological Typology*, pages 126–195. De Gruyter Mouton, Berlin. **DOI:** `https://doi.org/10.1515/9783110451931-005`

Jeffrey Heinz and Regine Lai. 2013. Vowel harmony and subsequentiality. In *Proceedings of the 13th Meeting on the Mathematics of Language*, pages 52–63.

Mans Hulden. 2017. Formal and computational verification of phonological analyses. *Phonology*, 34(2):407–435. **DOI:** `https://doi.org/10.1017/S0952675717000203`

Brett Hyde. 2008. Alignment continued: Distance-sensitivity, order-sensitivity, and the midpoint pathology. Unpublished manuscript, Washington University. Available at `http://roa.rutgers.edu/article/view/1028`.

Brett Hyde. 2012. Alignment constraints. *Natural Language and Linguistic Theory*, 30:789–836. **DOI:** `https://doi.org/10.1007/s11049-012-9167-3`

Brett Hyde. 2016. *Layering and Directionality*. Equinox, Sheffield.

C. Douglas Johnson. 1972. *Formal Aspects of Phonological Description*. Mouton, The Hague. **DOI:** `https://doi.org/10.1515/9783110876000`

Ronald Kaplan and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics*, 20:331–378.

Nate Koser and Adam Jardine. 2020. The complexity of optimizing over strictly local constraints. In *Proceedings of the 43rd Annual Penn Linguistics Conference*, pages 125–134.

Andrew Lamont. 2019a. Majority rule in harmonic serialism. In *Supplemental Proceedings of the 2018 Annual Meeting on Phonology*, Washington, D.C., Linguistic Society of America. **DOI:** `https://doi.org/10.3765/amp.v7i0.4546`

Andrew Lamont. 2019b. Precededence is pathological: The problem of alphabetical sorting. In *Proceedings of the 36th West Coast Conference on Formal Linguistics*, pages 243–249, Somerville, MA. Cascadilla Proceedings Project.

Géraldine Legendre, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, page 884–891, Hillsdale, NJ. Erlbaum.

John J. McCarthy. 2003. OT constraints are categorical. *Phonology*, 20(1):75–138. **DOI:** `https://doi.org/10.1017/S0952675703004470`

John J. McCarthy. 2010. Autosegmental spreading in Optimality Theory. In John A. Goldsmith, Elizabeth Hume, and W. Leo Wetzels, editors, *Tones and Features*, pages 195–222. Walter de Gruyter, Berlin.

John J. McCarthy and Alan Prince. 1993. Generalized alignment. In Geert Booij and Jaap van Marle, editors, *Yearbook of Morphology*, pages 79–153. Kluwer, Dordrecht. **DOI:** `https://doi.org/10.1007/978-94-017-3712-8_4`

Kevin Mullin. 2011. Strength in harmony systems: Trigger and directional asymmetries. Unpublished manuscript, University of Massachusetts Amherst. Available at `https://www.academia.edu/393775/Strength_in_Harmony_Systems_Trigger_and_Directional_Asymmetries`

Marina Nespor and Irene Vogel. 1986. *Prosodic Phonology*. Foris Publications, Dordrecht.

Alan Prince and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing, Malden, MA.

Douglas Pulleyblank. 2002. Harmony drivers: No disagreement allowed. In *Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistics Society*, pages 249–267. **DOI:** `https://doi.org/10.3765/bls.v28i1.3841`

Jason Riggle. 2004. *Generation, Recognition, and Learning in Finite State Optimality Theory*. Ph.D. thesis, University of California Los Angeles.

James Rogers, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome, and Sean Wibel. 2010. On languages piecewise testable in the strict sense. *The Mathematics of Language*, 10/11:255–265. **DOI:** `https://doi.org/10.1007/978-3-642-14322-9_19`

Sharon Rose and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language*, 80(3):475–531. **DOI:** `https://doi.org/10.1353/lan.2004.0144`

Elisabeth Selkirk. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press, Cambridge.

Elisabeth Selkirk. 2011. The syntax-phonology interface. In John A. Goldsmith, Jason Riggle, and Alan C. L. Yu, editors, *The Handbook of Phonological Theory*, 2nd edition, pages 435–484. Blackwell Publishing, Oxford. **DOI:** `https://doi.org/10.1002/9781444343069.ch14`

Paul Smolensky. 1992. Harmonic Grammars for formal languages. In *Advances in Neural Information Processing Systems 5*, pages 847–854, Morgan Kaufmann Publishers Inc., San Francisco.

Paul Smolensky. 1993. Harmony, markedness, and phonological activity. Paper presented at Rutgers Optimality Workshop 1. Available at `http://roa.rutgers.edu/article/view/88`.

Paul Smolensky. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature-domain markedness. In Paul Smolensky and Géraldine Legendre, editors, *The Harmonic Mind: From Neural*

Completeness to Optimality-Theoretic Grammar, volume II, pages 27–160. MIT Press, Cambridge, MA.

Keiichiro Suzuki. 1998. *A Typological Investigation of Dissimilation*. Ph.D. thesis, University of Arizona.

Rachel Walker. 2000. Long-distance consonantal identity effects. In *Proceedings of WCCFL 19*, pages 532–545. Somerville, MA. Cascadilla Press.

Kristine M. Yu. 2019. Parsing with Minimalist Grammars and prosodic trees. In Robert C. Berwick and Edward P. Stabler, editors, *Minimalist Parsing*, pages 69–109. Oxford University Press, Oxford. **DOI:** `https://doi.org/10.1093/oso/9780198795087.003.0004`