Computing Vowel Harmony:  
The Generative Capacity of Search and Copy

Search and Copy (SC, [Nevins, 2010; Mailhot and Reiss, 2007; Samuels, 2009a,b]) is a procedural model of vowel harmony (VH) in which underspecified vowels trigger searches for targets that provide them with features. We analyze the weak generative capacity of SC. When search proceeds in a single direction, SC can model a phonological process only if it can be represented as a tier-based input strictly local function (TISL, [Chandlee, 2014; Hao and Andersson, 2014]). The composition of multiple VH rules is however not TISL. When search proceeds in both directions, SC can model non-finite-state functions.

Background: VH is traditionally thought of as a process where features spread from a fully-specified vowel (the donor) to underspecified vowels (recipients). For example, Turkish has suffixes with vowels underspecified for backness. These vowels agree in backness with the root, e.g. the vowel /A/ in the suffixes /-lAr/ and /-A/, where A is the feature bundle [$\alpha$BACK, $-$HIGH, $-$ROUND].

(1)  

<table>
<thead>
<tr>
<th></th>
<th>'house'</th>
<th>'horse'</th>
<th>'address'</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>/ev/ $\rightsquigarrow$ [ev]</td>
<td>/at/ $\rightsquigarrow$ [at]</td>
<td>/adres/ $\rightsquigarrow$ [adres]</td>
</tr>
<tr>
<td>PL</td>
<td>/ev-lAr/ $\rightsquigarrow$ [ev-ler]</td>
<td>/at-lAr/ $\rightsquigarrow$ [atlar]</td>
<td>/adres-lAr/ $\rightsquigarrow$ [adres-ler]</td>
</tr>
</tbody>
</table>

In rule-based phonology, harmony can be implemented with an iterative output-based rule like (2a) or a simultaneous input-based rule like (2b). Both formulations are donor-based: the backness feature is spread from the donor to the recipient.

(2) a. A $\rightarrow$ [$\alpha$BACK] / [$\alpha$BACK]C* _ b. A $\rightarrow$ [$\alpha$BACK] / [$\alpha$BACK]{$\{C,A\}$}* _

SC replaces the notion of iterative and simultaneous rules with a search operation. For an underlying form like /adres-lAr-A/2, SC computes the surface form [adres-ler-e] 'at the addresses' in (1). According to SC, each of the two As initiates a search for a vowel with the feature [$\alpha$BACK]. A direction parameter specific to this process indicates that the search should proceed leftward. The search terminates when the first vowel with the desired features is found, and the value [$\alpha$BACK] carried by this vowel is copied to the A initiating the search.

Tier-Based Input Strictly Local Functions: Let $\Sigma$ be an alphabet, and let $\Gamma \subseteq \Sigma$ be a tier. A function $f : \Sigma^* \rightarrow \Sigma^*$ is $k$-input strictly local on tier $\Gamma$ ($k$-TISL), where $k > 0$, if $f$ is computed by a sub-sequential finite-state transducer (SFST) whose states encode the $k - 1$ most recent symbols in the input stream, ignoring symbols not in $\Gamma$. To see that SC models TISL functions, let us consider the computation of /adres-lAr-A/2 $\rightsquigarrow$ [adres-ler-e] 'at the addresses' in (1). According to SC, each of the two As searches for an appropriate donor vowel, and both searches end at the /e/. In fact, since this /e/ is the last vowel with an [$\alpha$BACK] feature specification, every search triggered by an A terminates at this e. Therefore, SC can be simulated using an SFST $T$ as follows. Each state represents the identity of the most recent symbol with an [$\alpha$BACK] specification. Every A encountered by $T$ is assigned the value of [$\alpha$BACK] indicated by its current state, and all other symbols are copied to the output stream faithfully. The function computed by $T$ is then 2-TISL on tier $\Gamma$, where $\Gamma$ is the set of all [$\pm$BACK] vowels. This corresponds to the set of fully specified vowels because only fully specified vowels have [$\pm$BACK] in Turkish.
Computing Vowel Harmony: The Generative Capacity of Search and Copy

Bidirectional Search and Copy In Woleaian, the verbal theme vowel /-A-/ is underspecified for \([\pm\text{LOW}]\). It surfaces as [c] if the closest vowels on both sides are \([-\text{LOW}]\), and it surfaces as [a] otherwise. (3) illustrates this using data from [Howard 1972].

\[1\text{SG} /\text{ülüm-A-ji}/ \rightsquigarrow [\text{ülümej}] \quad 3\text{SG} /\text{ülüm-A-la}/ \rightsquigarrow [\text{ülümal}]\]

[Nevins 2010] analyzes Woleaian as a bidirectional search initiated by the A. When the A is equidistant from two appropriate donors, the copying step is successful only if both donors contribute the same feature value to the recipient. In the 3SG from in (3), the copying step is said to fail, and the \([\pm\text{HIGH}]\) receives a default valuation of \([-\text{HIGH}]\). In general, this predicts that in a verb of the form /V_1C^-A-C^*V_2/., the A harmonizes with both \(V_1\) or \(V_2\). Such a system is not subsequential, but is Weakly Deterministic (Gainor et al. 2012; Heinz and Lai 2013). This is represented as a leftward search that copies the \([+\text{HIGH}]\) feature if it is found, followed by a rightward search that copies the \([-\text{HIGH}]\) feature if it is found.

Multiple harmonies: We have shown that individual spreading processes modeled by unidirectional SC are TISL. However, since the TISL functions are not closed under composition (Chandlee et al. 2018), two harmony processes applied sequentially may result in a non-TISL mapping. For example, in addition to the backness harmony process shown in (1), Turkish high vowels harmonize in roundness with previous vowels. The genitive suffix is /-In/, where /I/ is an underspecified vowel bearing the features \([\text{BACK}, -\text{HIGH}, \text{ROUND}]\). In the singular, /-In/ gets its backness and roundness features from the root: sap-in, son-un; but in the plural, it gets its backness feature from the plural suffix: sap-lar-in, son-lar-in, *son-lar-un. Thus, any monolithic TISL implementation of the two harmony systems would require A to be on the tier, which contradicts the fact that backness harmony requires A not to be on the tier.

\[
\begin{array}{c|c|c|c|c}
\text{Tier} & \text{SG} /\text{sap}/ & \rightsquigarrow & [\text{sap}] & \text{PL} /\text{sap-lAr}/ & \rightsquigarrow & [\text{sap-lAr}] \\
\hline
\text{SG} /\text{son}/ & \rightsquigarrow & \text{PL} /\text{son-lAr}/ & \rightsquigarrow & [\text{son-lAr}] \\
\hline
\text{GEN} /\text{son-In}/ & \rightsquigarrow & [\text{son-In}] & \text{PL, GEN} /\text{son-lAr-In}/ & \rightsquigarrow & [\text{son-lAr-In}] \\
\end{array}
\]

Conclusion: We have shown that unidirectional SC is TISL and that bidirectional SC is not finite-state. This result improves Gainor et al. (2012) sub-sequential bound on unidirectional harmony systems, since Nevins (2010) analyzes all such systems using unidirectional SC. Because the composition of multiple VH patterns is not TISL, this suggests the use of multiple TISL functions or multiple tiers (Aksënova 2018).