Bounded copying is subsequential: Implications for metathesis and reduplication*

Jane Chandlee
Linguistics and Cognitive Science
University of Delaware
Newark, DE
janemc@udel.edu

Jeffrey Heinz
Linguistics and Cognitive Science
University of Delaware
Newark, DE
heinz@udel.edu

Abstract

This paper first defines the conditions under which copying and deletion processes are subsequential: specifically this is the case when the process is bounded in the right ways. Then, if we analyze metathesis as the composition of copying and deletion, it can be shown that the set of attested metathesis patterns fall into the subsequential or reverse subsequential classes. The implications of bounded copying are extended to partial reduplication, which is also shown to be either subsequential or reverse subsequential.

1 Introduction

This paper presents a computational analysis of copying and deletion in metathesis and partial reduplication and establishes the necessary conditions for such patterns to be subsequential. More specifically, it is shown that such patterns fall into the subsequential or reverse subsequential classes if the copying (for both cases) and deletion (for the case of metathesis only) are bounded in the right ways.

The classification of natural language patterns by the Chomsky Hierarchy (Chomsky, 1956) is one means of distinguishing the complexity of the patterns found in various linguistic domains. Syntactic patterns, for example, may be context-free (e.g. English nested embedding, (Chomsky, 1956)) or context-sensitive (e.g. Swiss German crossing dependencies (Schieber, 1985)), while phonological patterns (i.e. patterns that can be described with rewrite rules of the form A $\Rightarrow$ B / C ___ D, where A, B, C, and D are regular expressions) have been shown by Johnson (1972) and Kaplan and Kay (1994) to be regular.

The regular class of patterns, however, is in fact too large to correspond exactly to phonology (Heinz, 2007; Heinz, 2009; Heinz, 2010). Rather, it seems that phonological patterns fit into a subclass of the regular patterns. Since the *subsequential class* is a proper subset of the regular class (Oncina et al., 1993; Mohri, 1997), it is therefore a useful candidate, especially because of its attractive computational properties (Mohri, 1997). Restricting the class of phonological patterns in this way has implications for learning, since the subsequential but not the regular class is identifiable in the limit from positive data (Oncina et al., 1993).

Using the formalism of finite state transducers (FSTs), we will show that metathesis and partial reduplication patterns can be described with subsequential FSTs. Subsequential FSTs are deterministic weighted transducers in which the weights are strings and multiplication is concatenation.

The analysis defines generally the conditions necessary for metathesis and partial reduplication to be subsequential. Representative examples of the empirical phenomena that can be so classified are shown in (1). (1-a) is an example of local metathesis, (1-b) is an example of metathesis around an intervening segment, and (1-c) is an example of partial reduplication.

*We thank the anonymous reviewers for useful questions and suggestions.This research is supported by grant #1035577 from the National Science Foundation.

Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON2012), pages 42–51,
Montréal, Canada, June 7, 2012. ©2012 Association for Computational Linguistics

(1) a. Rotuman:
hosa ⇒ hoas 'flower'
b. Cuzco Quechua:
yuraq ⇒ ruyaq 'white'
c. Tagalog:
sulat ⇒ susulat 'will write'

This kind of analysis sheds light on the nature of the relation itself independent of the particular theory used to account for it. It does not matter whether the metathesized or reduplicated forms exemplified here are derived via a series of SPE-style rules or with ranked constraints in OT. The mapping (e.g. <hosa, hoas> for Rotuman) remains the same in either case. Additionally, the analysis has implications for any model beyond the phonological domain that uses finite-state methodology, including (but not limited to) artificial intelligence (Russell and Norvig, 2009), bioinformatics (Durbin et al., 1998), natural language processing (Jurafsky and Martin, 2008), and robotics (Belta et al., 2007; Tanner et al., 2012).

The structure of the paper is as follows. Section two provides the formal definitions necessary for the analysis. Section three presents an analysis of subsequential copying, and section four presents an analysis of subsequential deletion. Section five turns to the analysis of metathesis as the composition of copying and deletion and proves the conditions for metathesis to be subsequential. Section six extends this analysis to partial reduplication. Section seven discusses the implications of the distinctions drawn by the computational analysis for both typology and learning. Section eight concludes.

2 Preliminaries

If Σ is a fixed finite set of symbols (an *alphabet*), then Σ^* is the set of all finite length strings formed over this alphabet, and $\Sigma^{\leq k}$ is the set of all strings of length less than or equal to *k*. A language is a subset of Σ^*. ϵ is the empty string. The length of a string *s* is $|s|$; thus $|\epsilon| = 0$. The *prefixes* of a string *s*, written $Pr(s)$, are $\{u \in \Sigma^* : \exists v \in \Sigma^* \text{ such that } s = uv\}$. The *suffixes* of a string *s*, written $Suf(s)$, are $\{u \in \Sigma^* : \exists v \in \Sigma^* \text{ such that } s = vu\}$. $Suf_n(s)$ is a suffix of *s* of length *n*. The nonempty, proper prefixes of a string *s* is written $Pr_{\text{prop}}(s)$.

If L is a language then the prefixes of L are $Pr(L) = \bigcup_{s \in L} Pr(s)$ and the nonempty proper prefixes of L are $Pr_{\text{prop}}(L) = \bigcup_{s \in L} Pr_{\text{prop}}(s)$. A language L is finite iff there exists some k such that $L \subseteq \Sigma^{\leq k}$. For any $w \in \Sigma^*$, the *good tails* of w in L is $T_L(w) = \{v \in \Sigma^* | wv \in L\}$. Two prefixes u_1 and u_2 are Nerode equivalent with respect to some language L iff they share the same good tails: $u_1 \sim_L u_2 \Leftrightarrow T_L(u_1) = T_L(u_2)$. In fact, a language L is *regular* iff the partition induced over Σ^* by $\sim_L$ has finite cardinality. Note that every finite language is regular. If L_1 and L_2 are languages then $L_1 L_2 = \{uv \mid u \in L_1 \text{ and } v \in L_2\}$.

Definition 1. *(Oncina et al., 1993) A* subsequential finite state transducer (SFST) *is a six-tuple (Q, Σ, Δ, q_0, δ, σ), where Q is a finite set of states, Σ is the input alphabet, Δ is the output alphabet, $q_0 \in Q$ is the initial state, $\delta \subset (Q \times \Sigma \times \Delta^* \times Q)$ is the transition function, and σ: $Q \Rightarrow \Delta^*$ is a partial function that assigns strings to the states in Q. The edges, E, of the SFST are a finite subset of $(Q \times \Sigma^* \times \Delta^* \times Q)$. SFSTs are deterministic, meaning they are subject to the condition $(q,a,u,r),(q,a,v,s) \in E \Rightarrow (u=v \wedge r=s)$.*

Definition 2. *(Oncina et al., 1993) A* path *in an SFST τ is a sequence of edges in τ, $\pi = (q_0, x_1, y_1, q_1)(q_1, x_2, y_2, q_2) \cdots (q_{n-1}, x_n, y_n, q_n)$. Π_τ is the set of all possible paths over τ. A path π can also be expressed as (q_0, x, y, q_n) where $x = x_1 x_2 ... x_n$ and $y = y_1 y_2 \cdots y_n$. The transduction τ realizes is the function t: $\Sigma^* \Rightarrow \Delta^*$ such that $\exists (q_0,x,y,q) \in \Pi_\tau$ and $t(x) = y\sigma(q)$.*

A relation describable with an SFST is a subsequential relation. If *f* and *g* are relations, $\circ$ denotes the composition, where $(g \circ f)(x) = g(f(x))$. Subsequential relations are closed under composition:

Theorem 1 ((Mohri, 1997) Theorem 1)**.** *Let f : $\Sigma^* \Rightarrow \Delta^*$ and g : $\Delta^* \Rightarrow \Omega^*$ be subsequential functions, then $g \circ f$ is subsequential.*

Let R be a relation. The reverse relation $R^r = \{<x^r, y^r>: <x,y> \in R\}$. A relation is reverse subsequential if its reverse relation is subsequential.

3 Subsequential copying

This paper will ultimately prove the conditions under which metathesis and partial reduplication, two processes which can be analyzed as involving copying, are subsequential relations. To do this it is first necessary to define a copy relation in general. A copy can be place either before or after the original—these two processes can be distinguished as pre-pivot or post-pivot copying (where the pivot is an intervening string from the set *U*).

Definition 3. *Let L,U,X,R be languages.*

1. *The rule* $\emptyset \Rightarrow X\,/\,LXU\_\_\;R$ *is a* post-pivot *copy relation.*

2. *The rule* $\emptyset \Rightarrow X\,/\,L\_\_\;UXR$ *is a* pre-pivot *copy relation.*

Kaplan and Kay (1994) show that if L, X, U, and R are regular languages, then the copy relations above are regular relations. One goal of this paper is to identify the conditions on L, X, U, and R which make the above relations subsequential.

Further distinctions can be drawn among regular copy relations based on which of the surrounding contexts of the copy and original are of bounded length.

Definition 4. *Let* L, X, U, *and* R *be regular languages.*

1. *A pre-pivot or post-pivot copy relation is* I-bounded *iff U is a finite language (I is for 'intervening').*

2. *A pre-pivot copy relation is* L-bounded *iff L is a finite language (L is for 'left-context').*

3. *A post-pivot copy relation is* R-bounded *iff R is a finite language (R is for 'right-context').*

4. *A copy relation is* T-bounded *iff X is a finite language (T is for 'target').*

Theorem 2 below states that a pre-pivot, T-bounded, I-bounded, and R-bounded copy relation is subsequential. To understand the idea behind the proof, consider the abstract pre-pivot relation schematized in the following SFST for a particular $l \in L$, $u \in U$, $x \in X$, and $r \in R$.[1] Note that each transition in the figure represents a series of transitions and states (depending on the lengths of the strings involved).

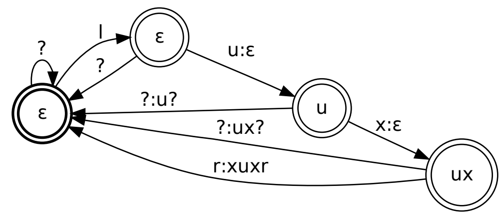

Figure 1: An SFST schematizing a pre-pivot T-bounded, I-bounded, and R-bounded copy relation.

The transitions for which the output is the empty string can be thought of as the machine withholding the output until it verifies that it has found the context for copying. Thus, out of the state labelled 'ux' the output on 'r' is 'xuxr', which is the copy followed by the segments for which there was no output (i.e. the segments in *u* and *x* that were being 'held'). This mechanism of holding is why the bounds on the lengths of the strings are necessary. If there were no upper bound on the length of the words in U, X, or R, the machine would have to hold a potentially infinite number of strings, which would in turn require infinitely many states. Without these bounds, no SFST can be constructed.

Theorem 2. *A regular pre-pivot copy relation that is T-bounded, I-bounded, and R-bounded is subsequential.*

Proof. Let C be a pre-pivot, T-bounded, I-bounded, and R-bounded regular copy relation. Then there exists a regular language L and finite languages U, X, and R such that C is described by the rewrite rule $\emptyset \Rightarrow$ X / L__ UXR.

An SFST is constructed for C as follows. The states Q are the set of good tails of L and the non-empty proper prefixes of UXR. Formally, let $\pi_L = \{T_L(w) \mid w \in Pr(L)\}$. Then

$$Q = (\pi_L \cup Pr_{\text{prop}}(UXR))$$

Since L is a regular language, there are finitely many elements of π_L. Since U, X, and R are finite lan-

[1] Following (Beesley and Karttunen, 2003), in this and all other FSTs in the paper, '?' represents any symbol or string except for those for which other transitions out of that state are defined. The states of the machine are labeled with the string mapped to them by the σ function.

guages, $Pr_{\text{prop}}(UXR)$ is also finite. Therefore Q is finite.

The initial state $q_0 = T_L(\epsilon)$.

The sigma function is defined as follows. $\forall q \in Q$,

$$\sigma(q) = \begin{cases} \epsilon & \text{iff } q \in \pi_L \\ q & \text{otherwise} \end{cases}$$

The transition function is defined in two parts. First, for all $s \in Pr(L)$ and $a \in \Sigma$:

$(T_L(s), a, a, T_L(sa)) \in E$ iff $s, sa \in Pr(L)$ and $s \notin L$
$(T_L(s), a, \epsilon, a) \in E$ iff $s \in L$ and $a \in Pr(UXR)$
$(T_L(s), a, a, T_L(\epsilon)) \in E$ otherwise

Second, for all s in the nonempty proper prefixes of UXR and $a \in \Sigma$:

$(s, a, \epsilon, sa) \in E$ iff $s, sa \in Pr_{\text{prop}}(UXR)$
$(s, a, xuxra, T_L(\epsilon)) \in E$ iff $(\exists x \in X)$ $(\exists u \in U)(\exists r \in Pr_{\text{prop}}(R))[s = uxr, ra \in R]$
$(s, a, \sigma(s)a, T_L(\epsilon)) \in E$ otherwise

It follows directly from this construction that the SFST recognizes the specified copy relation. □

Theorem 3 below states that a post-pivot copy relation need only be T- and R-bounded to be subsequential. The idea behind the proof is demonstrated by the abstract post-pivot copy relation schematized in the following SFST for a particular *x*, *u*, and *r*.

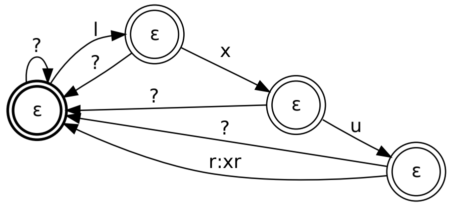

Figure 2: An SFST for a post-pivot T-bounded and R-bounded copy relation.

Since the machine finds the original *x* before it has to produce the copy, the segments in *u* do not have to be held. The bounding is only necessary for *x* itself, and for the right context of the copy *r*. Thus when the original precedes the copy, the bounding on *u* is no longer a necessary condition for subsequentiality.

Theorem 3. *A regular post-pivot copy relation that is T-bounded and R-bounded is subsequential.*

The proof of the Theorem 3 (omitted) is similar to the one for Theorem 2 but slightly more complicated by the fact that U can be any regular language.

The reverse of the relation in the proof of Theorem 3 would be a pre-pivot copy relation that is T-bounded and L-bounded. This would reverse the pattern in Figure 2, except the left context and not the right context would be bounded. Such a pattern is not subsequential, but it is reverse subsequential.

Corollary 1. *A regular pre-pivot copy relation that is T-bounded and L-bounded is reverse subsequential.*

4 Subsequential deletion

As with copying, the deletion relations relevant for metathesis come in two flavors, depending on whether the deleted string precedes or follows the one that remains.

Definition 5. *Let L,U,X,R be languages.*

1. *The rule X ⇒ ∅ / LXU__ R is a* post-pivot *deletion relation.*
2. *The rule X ⇒ ∅ / L__ UXR is a* pre-pivot *deletion relation.*

Kaplan and Kay (1994) show that if L, X, U, and R are regular languages, then the deletion relations above are regular relations. Another goal of this paper is to identify the conditions on L, X, U, and R which make the relations above subsequential.

We can thus provide parallel definitions for T-bounded, I-bounded, and R-bounded deletion relations.

Definition 6. *Let L,X,U,R be regular languages.*

1. *A pre-pivot or post-pivot deletion relation is* I-bounded *iff U is a finite language.*
2. *A pre-pivot deletion relation is* L-bounded *iff L is a finite language.*
3. *A post-pivot deletion relation is* R-bounded *iff R is a finite language.*
4. *A deletion relation is* T-bounded *iff X is a finite language.*

Figure 3 schematizes a pre-pivot regular deletion relation that is T-bounded, I-bounded, and R-bounded.

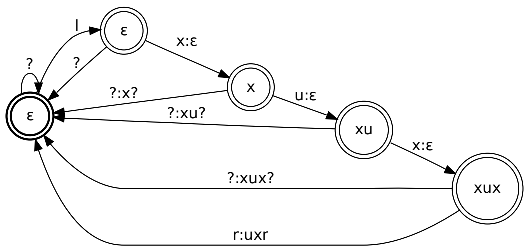

Figure 3: An SFST for a pre-pivot T-bounded, I-bounded, and R-bounded deletion relation.

As with the copying, due to the need for ϵ-transitions, the possibility of constructing this machine depends on the bounding of the length of the deleted string, the intervening string, and the string that makes up the right context.

Theorem 4. *A regular pre-pivot deletion relation that is T-bounded, I-bounded, and R-bounded is subsequential.*

Proof. Let D be a pre-pivot, T-bounded, I-bounded, and R-bounded regular deletion relation. Then there exists a regular language L and finite languages X, R, and U such that D is defined by the rewrite rule X ⇒ ∅ / L__ UXR.

An SFST is constructed for D as follows. The states Q are the set of good tails of L and the nonempty proper prefixes of XUXR. Formally, let $\pi_L = \{T_L(w) \mid w \in Pr(L)\}$. Then

$$Q = (\pi_L \cup Pr_{\text{prop}}(XUXR))$$

Since L is regular, π_L is finite. Since U, X, and R are finite languages, $Pr_{\text{prop}}(XUXR)$ is also finite. Therefore Q is finite.

The initial state $q_0 = T_L(\epsilon)$. The sigma function is defined as in Theorem 2.

Also as in Theorem 2, the transition function is defined in two parts. The first part, where all $s \in Pr(L)$ and $a \in \Sigma$ are considered, is the same as in Theorem 2 except for one case below.

$$(T_L(s), a, \epsilon, a) \in E \quad \text{iff} \quad s \in L \text{ and } a \in Pr(XUXR)$$

The second part, where all s in the nonempty proper prefixes of XUXR and $a \in \Sigma$ are considered, is constructed as follows.

$(s, a, \epsilon, sa) \in E$ iff $s, sa \in Pr_{\text{prop}}(XUXR)$
$(s, a, uxra, T_L(\epsilon)) \in E$ iff$(\exists x, x' \in X)$
$(\exists u \in U)(\exists r \in Pr_{\text{prop}}(R))[s = xux'r, ra \in R]$
$(s, a, \sigma(s)a, T_L(\epsilon)) \in E$ otherwise

It follows directly from this construction that the SFST recognizes the deletion relation D. □

As for a post-pivot deletion relation, the properties of T-bounding and R-bounding are sufficient for subsequentiality since the intervening set U occurs before the deletion. As with Theorem 3, the proof of Theorem 5 is omitted.

Theorem 5. *A regular post-pivot deletion relation that is T-bounded and R-bounded is subsequential.*

Lastly, if *u* is not bounded in a pre-pivot deletion relation, but *l* and *x* are, the relation is reverse subsequential.

Corollary 2. *A regular pre-pivot deletion relation that is T-bounded and L-bounded is reverse subsequential.*

5 Metathesis as the composition of copying and deletion

Metathesis has traditionally been viewed as an operation of transposition, in which segments switch positions. Under another view, metathesis can be considered as the result of two separate processes, a copy process followed by deletion of the original segment the copy was made from (Blevins and Garrett, 1998; Blevins and Garrett, 2004). Take, for example, the metathesis process in Najdi Arabic (Abboud, 1979) in which a word with a CaCCat template surfaces as CCaCat:

(2) /naʕɟat/ ⇒ [nʕaɟat] ‘ewe’

An independent process of deletion (CaCaC ⇒ CCaC) is also observed in Arabic dialects. So the change in (2) could be achieved via the two processes in (3). The result after both processes corresponds to metathesis (4).

(3) a. Copy: **CV$_1$CC** ⇒ CV$_1$CV$_1$C
b. Delete: CV$_1$CV$_1$C ⇒ **CCV$_1$C**

(4) Metathesis: $\mathbf{CV_1CC} \Rightarrow \mathbf{CCV_1C}$

This analysis provides a way of classifying attested patterns according to the type of copying and deletion involved. Cross-linguistic surveys (Blevins and Garrett, 1998; Blevins and Garrett, 2004; Hume, 2000; Buckley, 2011; Chandlee et al., to appear) reveal that in a large number of metathesis patterns there is some bound on the length of the string that intervenes between the copied segment and its original. A classic example is found in the Rotuman language, in which the incomplete form of a word is derived from the complete form via word-final consonant-vowel metathesis (Churchward, 1940). The general rule for the example in (5-a) is in (5-b).

(5) a. hosa ⇒ hoas 'flower'
b. CV ⇒ VC / V __ #

If we decompose this metathesis into its component copy and deletion operations, the copy portion would be as in (6):[2]

(6) $V_1CV_2\# \Rightarrow V_1V_2CV_2\#$

Applying Definitions 3 and 4 to this example, we first classify the Rotuman pattern as pre-pivot copying. Since the length of the string between the original segment and the copy (u = C) is bounded by 1, the copying is I-bounded. The copying is also T-bounded, since a single vowel is copied ($x = V_2$). And it is R-bounded, since the original vowel is word-final ($r = \epsilon$). An FST for this pattern is shown in Figure 4. Note that when the right context is the empty string, the copying is achieved via the σ function rather than a transition.

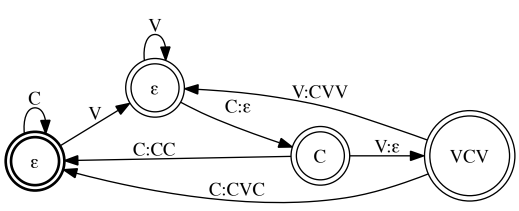

Figure 4: An SFST for the copy process of Rotuman CV-metathesis.

[2]This analysis assumes that it is the vowel that metathesizes. The same surface form would obtain if the consonant metathesized ($C_1V \Rightarrow C_1VC_1 \Rightarrow VC_1$). For evidence that the vowel is indeed the segment involved in CV metathesis, see (Heinz, 2005).

Another example of I-bounded copying is in an optional metathesis process in Cuzco Quechua, in which sonorants metathesize across an intervening vowel (Davidson, 1977):

(7) yuraq ⇒ ruyaq 'white'

Under a copy+deletion analysis of metathesis, this pattern would involve two copy processes followed by two deletions, one for the 'r' and one for the 'y'. Schematizing just the process for 'r', we can see in (8) that the length of the intervening string is again bounded by 1 (the 'y' is removed for clarity).[3] The copied string x (the liquid) is also bound by 1.

(8) uraq ⇒ ruraq

L-bounded copying is exemplified in a diachronic metathesis pattern found in a South Italian dialect of Greek (Rohlfs, 1950):

(9) Classical South Italian Greek
gambros ⇒ grambo[4] 'son-in-law'

In this pattern, a non-initial liquid surfaces in the initial onset cluster. The original position of the liquid varies, which means there is no bound on the length of the intervening string. However, the location of the copy is always the initial cluster, which means the left context of the copy *is* bounded by the length of the maximum onset.

The copying in metathesis is only the first step - the second is deletion of the original. The deletion rule must be complementary to the copying, in the sense that if the copying is pre-pivot, the deletion will be post-pivot (10), and vice versa (11).

(10) a. Copy: $\emptyset \Rightarrow$ x / v__ uxw
b. Delete: x $\Rightarrow \emptyset$ / vxu __ w

(11) a. Copy: $\emptyset \Rightarrow$ x / vxu __ w
b. Delete: x $\Rightarrow \emptyset$ / v__ uxw

As stated in section 2, if two relations are subsequential, then the composition of these relations will also be subsequential. Thus, whether or not the metathesis is subsequential depends on its component copy and deletion processes. This result is es-

[3]This assumes the two copy-deletions occur simultaneously. If one were to precede the other, then the bound would be 2.

[4]The [s] is deleted in an unrelated process.

tablished in the following theorem.

Theorem 6. *If a copy relation* f *is either (1) post-pivot, T-bounded, I-bounded and R-bounded or (2) pre-pivot, T-bounded, and R-bounded, and a deletion relation* g *is either (1) post-pivot, T-bounded, I-bounded and R-bounded or (2) pre-pivot, T-bounded, and R-bounded, then the metathesis relation* $g \circ f$ *is subsequential.*

Proof. By Theorems 2 and 3, f is subsequential. By Theorems 4 and 5, g is subsequential. By Theorem 1, $g \circ f$ is subsequential. □

6 Partial Reduplication

Unlike metathesis, reduplication is analyzed by phonologists of all stripes as involving copying. Traditionally, two categories of reduplication have been described: full (or total) and partial. Full reduplication involves copying the entire string and affixing it to the original. A classic example is found in Indonesian (Sneddon, 1996), to express the plural:

(12) a. buku 'book'
b. buku-buku 'books'

In contrast, partial reduplication involves copying a designated portion of the string and affixing it as either a prefix, suffix, or infix. These options can be schematized as in (13)(Riggle, 2003), in which local means the reduplicant attaches adjacent to the material it copies and nonlocal means the reduplicant copies a non-adjacent portion of the string:[5]

(13) a. local prefixation: CV-CVZ
b. nonlocal prefixation: CV-ZCV
c. local suffixation: ZCV-CV
d. nonlocal suffixation: CVZ-CV
e. infixation: C_1VC_1Z

An example of local prefixation is found in Tagalog, in which the future of a verb is derived from the stem with a CV reduplicative prefix (Blake, 1917):

(14) sulat ⇒ susulat 'will write'

An example of the infixation shown in (13-e) can be found in Pima, in which the plural is derived by infixing a copy of the initial consonant after the first vowel (Riggle, 2006):

(15) sipuk ⇒ sispuk 'cardinals'

Since under the current analysis of metathesis as copy+delete, the only difference between a metathesis pattern and partial reduplication is that the latter does not involve the second process of deletion, we should predict that partial reduplication patterns will also be subsequential if the copying is bounded as per the definitions given above. We can clearly see that in the local patterns (13-a) and (13-c), the original and the copy are adjacent and therefore (vacuously) they are I-bounded. For the infixation, the copy likewise appears at a fixed distance from the original. In all three cases, the amount of material to be copied fits a template and is thus bounded by the length of the template (i.e. they are T-bounded). Therefore by Theorem 2, these partial reduplication patterns are subsequential. The SFST for local prefixation (13-a) is shown in Figure 5.

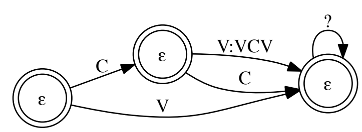

Figure 5: An SFST for CV-CVZ partial reduplication.

The SFST for local suffixation (13-c) is shown in Figure 6.

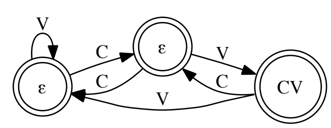

Figure 6: An SFST for ZCV-CV partial reduplication.

And the SFST for infixation (13-e) is shown in Figure 7.

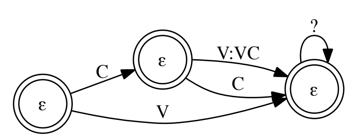

Figure 7: An SFST for C_1VC_1Z partial reduplication.

As for the nonlocal patterns, first consider the case of suffixation, (13-d). The string represented by Z

[5]These schemas assume CV reduplication, but the analysis would proceed the same for any reduplicant template, CC, CVC, etc.

is not bounded, and therefore this pattern is not I-bounded. It is, however, R-bounded, since the copy is always affixed to the end of the word. The right context is the empty string, which is (vacuously) bounded. Thus, by Theorem 3 this pattern is also subsequential; the FST is presented in Figure 8.

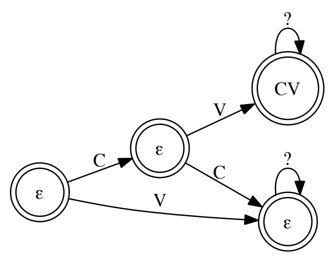

Figure 8: An SFST for a CVZ-CV pattern.

As for the last pattern, nonlocal prefixation, we have the opposite situation: the machine does have to hold a potentially infinite number of segments while it searches for the original, which means the pattern is not subsequential. However, this partial reduplication pattern is reverse subsequential by Corollary 1: reversing the string CV-ZCV gives VCZ-VC, which is R-bounded and identical to the nonlocal suffixation that was already argued to be subsequential.

To summarize, the attested partial reduplication patterns all appear to be subsequential or reverse subsequential.

This leaves us with full reduplication. Full reduplication is non-regular—although the position of the copy is fixed, the amount of material that is copied is not: full reduplication is not T-bounded. Again, with no principled upper limit on the length of words, a machine that copies an entire string cannot be finite state, much less subsequential. This distinction separates full and partial reduplication in terms of computational complexity—the implication being that these processes may be better viewed as distinct phenomena rather than subclasses of a single process.

7 Discussion

The analyses of subsequential copying and deletion presented above have revealed the conditions under which metathesis and partial reduplication patterns are subsequential. Under a copy+deletion analysis of metathesis, all metathesis patterns are T-bounded, since only one segment is copied (and then deleted) at a time. Partial reduplication is also T-bounded, assuming what is copied fits a certain template. Thus the T-bounded requirement on subsequential copying excludes only full reduplication from the subsequential class.

The analysis of metathesis relied on generalized representations of such patterns that appear to be typologically justified. A large number of attested metathesis patterns are considered 'local', which amounts to being I-bounded by 1. Metathesis patterns described as 'long distance' are also either I-bounded or else L-bounded - patterns such as Romance liquid movement (Vennemann, 1988) and Romani aspiration displacement (Matras, 2002) are striking in that they all affect the initial onset of the word.[6] A type of logically possible pattern that appears to be unattested is one in which no context of the copying - left, right, or intervening - is bounded. Such a pattern would not be subsequential, and in fact could not be described with any FST (i.e. it is non-regular). In this way the restriction of metathesis to the subsequential class finds support in the typological evidence. Also (apparently) unattested is an R-bounded metathesis pattern - one which targets the end of the word - though this is readily found in the typology of partial reduplication. It remains for future work whether such a metathesis pattern does exist, and if not, whether further distinctions need to be drawn to account for why R-bounded copying only appears as partial reduplication.

Narrowing the computational bound of possible phonological patterns from the regular class to the subsequential class also has implications for learning. It is known that the class of regular languages is not identifiable in the limit from positive data (Gold, 1967), but the subsequential class is: Oncina et al. (1993) have shown this class to be learnable by the OSTIA algorithm. Although (without modification) OSTIA does not do so well in practice on real data sets (Gildea and Jurafsky, 1996), future work may reveal algorithms that fare better if the hypothesis space can be restricted even further (i.e. to a sub-

[6]The L-bounded patterns also appear to be restricted to the diachronic domain. See (Chandlee et al., to appear)

class of the subsequential relations).

8 Conclusion

This paper has argued for an analysis of metathesis as the composition of copying and deletion. Such an analysis provides a computational link between metathesis and partial reduplication, which extends into a classification of these patterns as subsequential based on the bounded nature of the copying and (in the case of metathesis) deletion. The typology of attested patterns aligns well with the classifications proposed here, suggesting a tighter computational bound on phonological patterns than the one established by Johnson (1972) and Kaplan and Kay (1994).

Thus we can add metathesis and partial reduplication to the phonological processes that have previously been shown to be subsequential - see (Koirala, 2010) for substitution, insertion, and deletion, and Gainor et al. (to appear) for vowel harmony. It will be interesting to see to what extent morphological processes, including templatic morphology, and prosodic circumscription, also fit into this class.

References

P. F. Abboud. 1979. The verb in Northern Najdi Arabic. *Bulletin of the School of Oriental and African Studies*, 42:467-499.

K.R. Beesley and L. Karttunen. 2003. *Finite State Morphology.* Stanford: CSLI Publications.

C. Belta, A. Bicchi, M. Egerstedt, E. Frazzoli, E. Klavins and G. J. Pappas. 2007 Symbolic planning and control of robot motion. *IEEE Robotics and Automation Magazine*, 14(1): 61–71.

F. R. Blake. 1917. Reduplication in Tagalog. *The American Journal of Philology*, 38: 425-431.

J. Blevins and A. Garrett. 1998. The origins of consonant-vowel metathesis. *Language*, 74(3):508-556.

J. Blevins and A. Garrett. 2004. The evolution of metathesis. In B. Hayes, R. Kirchner, and D. Steriade (eds.) *Phonetically Based Phonology*. Cambridge: Cambridge UP, 117-156.

E. Buckley. 2011. Metathesis. In M. van Oostendorp, C.J. Ewen, E. Hume, and K. Rice (eds.) *The Blackwell Companion to Phonology*, Volume 3. Wiley-Blackwell.

J. Chandlee, A. Athanasopoulou, and J. Heinz. to appear Evidence for classifying metathesis patterns as subsequential. *Proceedings of the 29th West Coast Conference on Formal Linguistics* Somerville: Cascadilla.

N. Chomsky. 1956. Three models for the description of language. *IRE Transactions on Information Theory 113124.* IT-2.

C. M. Churchward. 1940. *Rotuman grammar and dictionary*. Sydney: Methodist Church of Australasia, Department of Overseas Missions.

J. O. Davidson, Jr. 1977. A Contrastive Study of the Grammatical Structures of Aymara and Cuzco Quechua Ph.D. dissertation. University of California, Berkeley.

R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison 1998 *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* Cambridge University Press.

B. Gainor, R. Lai, and J. Heinz. to appear Computational characterizations of vowel harmony patterns and pathologies. *Proceedings of the 29th West Coast Conference on Formal Linguistics.* Somerville: Cascadilla.

D. Gildea and D. Jurafsky. 1996. Learning bias and phonological-rule induction. *Computational Linguistics*22, 497-530.

E.M. Gold. 1967. Language identification in the limit. *Infomation and Control*10, 447-474.

J. Heinz. 2005. Optional partial metathesis in Kwara'ae. *Proceedings of AFLA 12, UCLA Working Papers*, 91-102.

J. Heinz. 2007. The inductive learning of phonotactic patterns. Doctoral dissertation, University of California, Los Angeles.

J. Heinz. 2009. On the role of locality in learning stress patterns. *Phonology* 26: 303-351.

J. Heinz. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41, 623-661.

E. Hume. 2000. *Metathesis Website*. www.ling.ohio-state.edu/ ehume/metathesis.

C. D. Johnson. 1972. *Formal Aspects of Phonological Description* The Hague: Mouton.

D. Jurafsky and J.H. Martin 2008 *Speech and Language Processing, 2nd edition.* Pearson Prentice Hall.

R. Kaplan and M. Kay. 1994. Regular models of phonological rules systems. *Computational Linguistics* 20: 331-378.

C. Koirala. 2012. Strictly Local Relations. Ms. University of Delaware.

Y. Matras. 2002. *Romani: a linguistic introduction*. Cambridge: Cambridge UP.

M. Mohri. 1997. Finite-state transducers in language and speech processing. *Computational Linguistics*, 23: 269-311.

J. Oncina, P. Garcia, and E. Vidal. 1993. Learning subsequential transducers for pattern recognition interpretation tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5): 448-457.

J. Riggle. 2003. Nonlocal reduplication. *Proceedings of the 34th annual meeting of the North Eastern Linguistic Society*.

J. Riggle. 2006. Infixing reduplication in Pima and its theoretical consequences. *Natural Language and Linguistic Theory*, 24: 857-891.

G. Rohlfs. 1950. *Historische grammatik der unteritalienischen Gräzität*, München: Verlag der Bayerischen Akademie der Wissenchaften.

S. Russell and P. Norvig. 2009 *Artificial Intelligence: a modern approach.* Upper Saddle River, NJ: Prentice Hall

S. Schieber. 1985 Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8: 333-343.

J. N. Sneddon. 1996. *Indonesian: a comprehensive grammar*, Routledge.

H. G. Tanner, C. Rawal, J. Fu, J. L. Piovesan, and C. T. Abdallah. 2012 Finite abstractions for hybrid systems with stable continuous dynamics. *Discrete Event Dynamic Systems* ,22(1):83-99.

T. Vennemann. 1988. *Preference laws for syllable structure and the explanation of sound change.* Berlin: Mouton de Gruyter.