Evidence for Classifying Metathesis Patterns as Subsequential

Jane Chandlee, Angeliki Athanasopoulou, and Jeffrey Heinz
University of Delaware

1. Introduction

This paper presents a computational analysis of metathesis patterns that distinguishes three categories of metathesis that differ in their computational complexity. These categories are local metathesis, bounded long distance metathesis, and unbounded long distance metathesis. Using the formalism of finite state automata, it is established that the first two categories are *subsequential*, while the third category is not (in fact, it is not even *regular*). These terms will be discussed in more detail below, but the overall distinction is one of complexity: the subsequential class is more restrictive than the regular class. Assigning a pattern to the subsequential class then identifies it as less complex than a pattern that is non-regular. Furthermore, the patterns identified as subsequential are robustly attested in the world's languages, whereas the non-regular patterns are much less common and in fact only attested diachronically. Thus this result suggests an upper bound for how complex a synchronic phonological pattern can be.

The outline of this paper is as follows. Section 2 presents the Chomsky Hierarchy of language patterns and discusses recent findings and hypotheses for where to classifying phonological patterns on the hierarchy. Section 3 defines subsequential FSTs and demonstrates how they can be used to describe phonological patterns. Section 4 presents the computational analysis of metathesis patterns. Section 5 discusses the typological implications of the analysis, as well as the implications for learning. Section 6 indicates directions for future work and concludes.

2. The Chomsky Hierarchy and the Subregular Hypothesis

Chomsky (1956) presents a means to classify languages and language patterns based on their degree of complexity or how expressive they are. This hierarchy is shown in Figure 1.

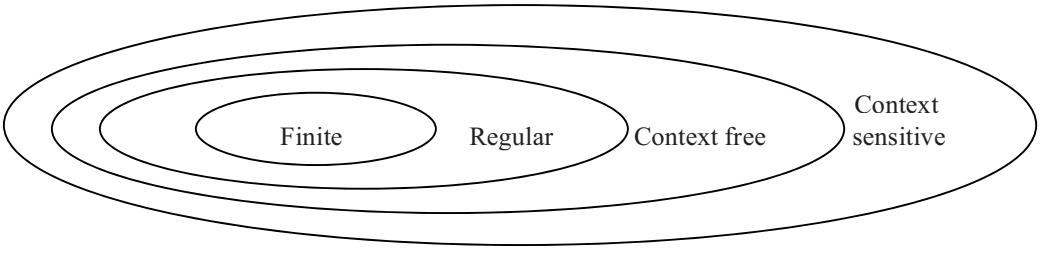

Figure 1. The Chomsky Hierarchy

One use of the hierarchy has been to identify a difference in the complexity of patterns from different domains of linguistic knowledge (Heinz and Idsardi 2011). For example, we know that at least some

* Special thanks to members of the Phonetics and Phonology Lab at the University of Delaware for valuable discussion and feedback on this work.

 Proceedings of the 29th West Coast Conference on Formal Linguistics, ed. Jaehoon Choi et al., 303-309. Somerville, MA: Cascadilla Proceedings Project.

syntactic patterns are context free or context sensitive (such as English Nested Embedding (Chomsky 1956) and Swiss German Crossing Dependencies (Shieber 1985), respectively). Phonological patterns, on the other hand, appear to fit into the regular class: Johnson (1972) and Kaplan and Kay (1994) have shown that rewrite rules of the form A → B / C __ D, where A, B, C, and D are regular expressions, describe regular relations. It is precisely rules of this type that we use to describe phonological patterns (in an SPE style analysis).

However, in one sense the regular class is too large to directly coincide with the class of phonological patterns: there are regular patterns that do not resemble phonological patterns. For example, a pattern in which well-formed words have an even number of sibilants can be described by a regular expression, but no such pattern is attested in the world's languages. Thus it may be more accurate to consider phonological patterns as belonging to a *subclass* of the regular class, in other words to say that phonology is *subregular* (Heinz 2007, 2009, 2010). Such a hypothesis would be supported if it could be shown that phonological patterns fit into a well-defined subclass of regular relations.

One candidate for this subclass is the class of subsequential relations, which are a proper subclass of the regular relations (Mohri 1997). It has been shown that common phonological processes, such as epenthesis, deletion, substitution, local assimilation and dissimilation (Koirala 2010), and vowel harmony (Gainor et al. 2011) fall into the subsequential class of relations. The analysis that will be presented in Section 5 will show that local and bounded long distance metathesis patterns also belong to this class.

3. Subsequential Finite State Transducers

Regular relations can be defined as those relations describable with the formalism of a finite state transducer (FST) (for details see Beesley and Karttunen 2003). *Subsequential* relations are defined as those describable with *subsequential* FSTs. Subsequential FSTs are a special kind of FST. Before defining them formally, we convey the main ideas behind these machines through an example and informal discussion.

Consider a relation called 'voice assimilation' that describes the common process of word-final obstruent voicing assimilation. This relation would include the input-output pairs <kætz, kæts>, <kagz, kagz>, etc. Figure 2 contains a fragment of the subsequential FST that describes this relation (just the fragment relevant to the alternation of <kætz, kæts> is shown for clarity).

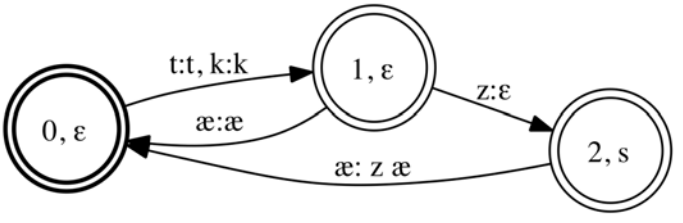

Figure 2. Fragment of subsequential FST for the relation 'voice assimilation'

The states of the machine are numbered and labeled with an assigned string. This string is appended to the final output of the transducer and is part of the machine's definition (it is given by a function called the sigma function, see below). The machine essentially reads in the input string /kætz/ and looks for the trigger for the z/s alternation: that trigger is a word-final voiceless obstruent-z cluster. Thus anytime it finds a voiceless obstruent it moves into state 1; if it does not then find a [z] it returns to state 0 and continues with the next segment of the string. In most cases the machine just outputs the same segment it read in, but if, while in state 1, it does find a [z], it 'holds onto' that [z] and moves into state 2. This is represented by the z:ε transition, which means 'input a z and output the empty string'. So in state 2, the current output is [kæt]. If the end of the input string is reached, as it is in this example, then the string mapped to state 2 by the sigma function, [s], is appended to the output: [kæt+s] to give the correct result of [kæts]. In this way the machine achieves the desired alternation.

Three characteristics distinguish subsequential FSTs from FSTs in general. First, they are deterministic on the input, which means that at any given state, every possible readable input transitions the machine to at most one other state. Second, their definition includes a function that maps states of the machine to strings. These are the strings that are appended to the output if the end of the input string is reached in that state. Third, all states are final.

Formally, a *subsequential FST* is a six-tuple $\tau = (Q, X, Y, q_0, E, \sigma)$, where Q is a finite set of states, X is the input alphabet, Y is the output alphabet, $q_0 \in Q$ is the initial state, $E \subset (Q \times X \times Y^* \times Q)$ is the transition function, and $\sigma: Q \Rightarrow Y^*$ is a partial function that assigns output strings to the states in Q (definition from Oncina et al. 1993).

A *path* in τ is a sequence of transitions $\pi = (q_0, x_1, y_1, q_1)(q_1, x_2, y_2, q_2)\ldots(q_{n-1}, x_n, y_n, q_n)$ where $q_1 \in Q$, $x_i \in X$, $y_i \in Y^*$, $1 \leq i \leq n$. Because τ is deterministic, the path can be written as $\pi = (q_0, x_1, x_2 \ldots x_n, y_1, y_2 \ldots y_n, q_n)$. Let Π_τ be the set of all possible paths over τ. The function realized by τ is the partial function $t: X^* \Rightarrow Y^*$ defined as $t(x) = y\sigma(q)$ iff $\sigma(q)$ is defined and $(q_0, x, y, q) \in \Pi_\tau$ (definition from Oncina et al. 1993).

Subsequential FSTs permit at most one output for every input, though Mohri (1997) has generalized them to p-subsequential machines that permit at most p outputs per input.

4. Analysis of Metathesis

Before turning to the analysis of metathesis, it is necessary to first clarify what types of patterns are being considered as metathesis, since there is some variation in how the term is used in the literature. Metathesis is typically defined as an alternation in which two segments switch positions, as in the Rotuman example in (1) (from Churchward 1940).

(1) hula → hual ‘moon’

In (1), the incomplete form of the noun is derived from the complete form via a process of word-final CV metathesis. Another class of patterns that has been described involves a single segment moving leftward or rightward from its original position. A diachronic example from Spanish is in (2) (from Lipski 1990):

(2) *costra > crosta ‘crust’

Some authors refer to the pattern exemplified in (2) as ‘displacement’ rather than metathesis, since only one segment moves. However, it is possible to analyze the pattern in (1) the same way, with only the vowel moving rightward. In that sense, the movement is the same as in (2) – it just takes place over a shorter distance.

The following analysis does not distinguish patterns based on how many segments move: both of the pattern types exemplified in (1) and (2) are classified as metathesis. A distinction will be made, however, based on how *far* the segment(s) move. Specifically, the relevant distinction is between a local (or adjacent) pattern and two types of long distance (non-local) patterns.

4.1. Local metathesis

A local metathesis pattern is one in which the segments involved are adjacent, as in the Rotuman example above in (1). Additional examples are shown in (3) (Churchward 1940).

(3) a. tiko → tiok ‘flesh’
b. hosa → hoas ‘flower’

The Rotuman pattern can be described with the rule $C_1V_2 \rightarrow V_2C_1 / V_1$ __ #: the adjacent segments C_1V_2 appear in the opposite order when in the context of being word-final and following a vowel. More generally, a local metathesis pattern can be described with a rule of the form ab → ba / C __ D, where C and D are regular expressions that define the contextual trigger for the metathesis. Such a

pattern can also be described with a subsequential FST. An FST for the Rotuman pattern is shown in Figure 3. For readability, this machine abstracts away from the actual segment inventory of Rotuman and uses generic C(onsonant) and V(owel) labels on the transitions. Without this abstraction away from the inventory, the machine would be much larger, requiring separate states and transitions for all possible CV combinations.

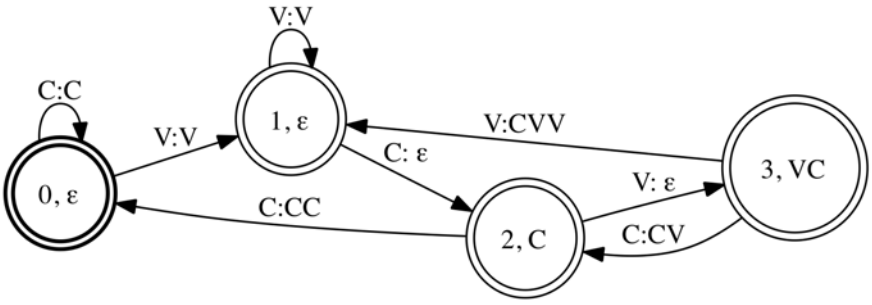

Figure 3. Subsequential FST for the Rotuman metathesis pattern

To see how the FST works, consider the derivation of *tiok* from *tiko*. Again the machine reads in the input string one segment at a time; as it reads in the input it is looking for the context of metathesis (a word-final VCV sequence). To start, it takes the transition marked C:C out of state 0, and this corresponds to reading in the initial [t] and just outputting that [t]. It then proceeds to state 1 via the V:V transition, again just reading in the [i] and outputting [i]. When it next reads in another consonant, it proceeds to state 2 via the C:ε transition, which again means that nothing is outputted and the C segment is being 'held'. The reason the machine holds onto the segment is that it has seen a portion of the context for metathesis: it has seen a VC. It now has to confirm whether the rest of the context is present before deciding what to do with the held consonant. It then moves to state 3 via the V:ε transition – it again has to hold onto the vowel until it determines that the VCV sequence is word-final. Since the end of the input string has been reached, the sequence is in fact word-final and the machine returns the held CV ('ko') in the opposite order via the sigma function, which has mapped the string VC to state 3. If the metathesis context had not been found, say if the input turned out to be just *tik*, then the held segment ('k') would have just been returned by the sigma function when the string ended in state 2. Thus a local metathesis pattern can be described by a subsequential FST and is therefore subsequential.

4.2. Long distance metathesis: bounded

In long distance (LD) metathesis patterns the segment(s) involved in the movement move over an intervening string of one or more segments. An example (from Davidson 1977) is found in Cuzco Quechua, in which a liquid and glide segment exchange positions over an intervening vowel:

(4) yuraq → ruyaq 'white'

Such a pattern can be described by a rule such as aBc → cBa / C __ D, where B represents a string of intervening segments. In the case of Cuzco Quechua, B just represents the single intervening vowel. More generally, as long as the number of segments in B is bounded (i.e. there exists some *k* such that the length of B is less than or equal to *k*), then B represents a finite number of strings and a subsequential machine with finitely many states can be constructed to accommodate all of those strings. Such a pattern is called a bounded long distance metathesis pattern. An abstract version of the machine that can describe this class of patterns is in Figure 4.

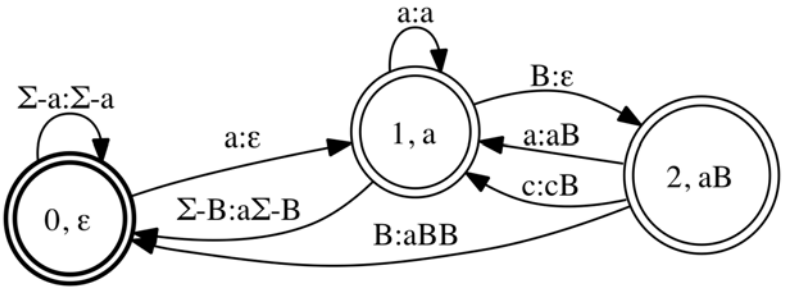

Figure 4. Subsequential FST to describe the class of bounded long distance metathesis patterns

In the machine in Figure 4, 'B' actually represents a series of transitions and states – however many would be required to accommodate all of the strings bounded by length *k*. But other than that the machine works the same as the one for the local metathesis cases in Figure 3; it reads in the segments and holds the ones that potentially form the context for metathesis. In this case the metathesis is accomplished by the transition marked 'c:cB' – the 'c' segment is returned in front of the intervening string.

4.3. Long distance metathesis – unbounded

Compare the Cuzco Quechua bounded long distance pattern to one such as in South Italian Greek, exemplified in (5) (Rohlfs 1924):

(5)	Classic Greek	South Italian Greek	
	kopros	kropo	'dung'
	gambros	grambo	'son-in-law'

The data in (5) show a diachronic process of liquid metathesis that is also common to several of the Romance languages. Again such a pattern could be described with a rule such as aBc → cBa / C __ D, but in this case the description of the pattern does not specify or restrict how many segments the liquid crosses over. It always moves to the initial onset, but how many segments it crosses over to get to that position depends on the length of the word (i.e. in the above examples it crosses over two and three segments, respectively, but in a hypothetical word like *gambobros* it would have to cross over five segments). This means that the length of the intervening string B also depends on the length of the word, which in turn means that with no principled upper bound on the length of B, there is no *k* such that the length of the intervening string B is less than or equal to *k*. Thus, such a pattern is an *unbounded* long distance metathesis pattern.

Unbounded LD metathesis patterns cannot be described by a subsequential FST. Looking again at the machine in Figure 4 for the bounded LD cases – we noted that 'B' in the machine actually represents the set of states and transitions needed to read in any of the finite number of strings bounded by *k*. In particular, the 'B:ε' transition from state 1 to state 2 represents all the segments that the machine has to hold onto as it checks for the metathesis triggering context. Without that upper bound of *k*, the set of strings represented by 'B' is no longer finite – it is infinite, and therefore the machine cannot be constructed (i.e. it cannot have finitely many states). Thus, unbounded LD metathesis patterns are not subsequential, and in fact, they are not even regular, since no FST can be constructed to describe them.

5. Further implications

In the previous section, it was shown that two categories of metathesis – local and bounded long distance – are subsequential, while a third category – unbounded long distance – is not. Interestingly, this distinction is reflected in the typology of metathesis patterns: the subsequential patterns are also the ones attested synchronically, while the patterns that are not subsequential are only attested

diachronically. More precisely, there appear to be no synchronic cases of unbounded long distance metathesis (Harris and Halle 2005, Buckley 2011). A summary of these findings is presented in Table 1.

Metathesis pattern	Regular	Subsequential	Attested
Local	√	√	synchronic
Bounded LD	√	√	synchronic
Unbounded LD	X	X	diachronic

Table 1. Typology of metathesis patterns

Since subsequential relations are a proper subclass of regular relations, as noted above, then the result that local and bounded LD metathesis patterns are subsequential necessarily means they are regular as well. Unbounded LD metathesis, however, is neither subsequential nor even regular.

These distinctions raise two significant questions. One is why are the synchronic patterns subsequential? The hypothesis we are supporting, along with those cited above who have likewise classified other phonological patterns as subsequential, is that subsequentiality is a psychologically real boundary for what makes a possible pattern a phonological one. The second question is, if phonological patterns are restricted to the subsequential class, why do non-subsequential/non-regular patterns appear even diachronically? One answer that has been argued for by Blevins and Garrett (1998, 2004) is that certain phonetic cues can extend over multiple segments – this anticipatory coarticulation is then misperceived as the segmental source originating in a different position. The segment is then eventually reanalyzed as being in this position, and the result is what appears diachronically to be movement or metathesis. The types of cues susceptible to such bleeding over segments include rhoticity and aspiration – the very types of sounds that appear in unbounded LD metathesis patterns.

The analysis presented above has implication for learning as well. As stated before, these findings on metathesis support the hypothesis that phonology is not only regular, but subsequential. This establishes a tighter computational bound on the set of possible phonological patterns than that posited by Johnson (1972) and Kaplan and Kay (1994). From the learning perspective, this means that the hypothesis space can be restricted to subsequential patterns. Indeed, the class of subsequential relations is identifiable in the limit from positive data, and existing algorithms can provably identify subsequential patterns (Oncina et al. 1993).

6. Conclusions and future work

Though algorithms such as OSTIA (Oncina et al. 1993) exist that can learn subsequential patterns from positive data, the amount of data required for them to learn phonological patterns does not appear to be present in dictionaries (Gildea and Jurafsky 1996). However, as Gildea and Jurafsky's study indicates, algorithms may do much better if the hypothesis space for phonological patterns can be further restricted beyond just the subsequential class. In fact, just as the regular class is too large because regular patterns exist that are not phonological, subsequential patterns also exist that are not phonological (i.e. subsequentiality is a necessary, but not a sufficient criterion for a phonological pattern). Thus phonology may actually correspond to a *subclass* of the subsequential class. More work is needed to identify what that class might be.

To conclude, it has been established that local and bounded long distance metathesis patterns belong to the subsequential class of relations. A third type of metathesis, unbounded long distance metathesis, does not belong to this class, and in fact does not even belong to the regular class. Nonetheless, this non-regular metathesis type appears to be restricted to the diachronic domain, and thus the analysis presented here provides evidence for limiting synchronic phonology and the learning hypothesis space to the subsequential class.

References

Beesley, Kenneth and Kartunnen, Lauri. (2003). *Finite State Morphology*. Stanford, CA: CSLI Publications.

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

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

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

Chomsky, Noam. (1956). Three models for the description of language. *IRE Transactions on Info Theory* 113124. IT-2.

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

Davidson, Joseph Orville, Jr. (1977). A contrastive study of the grammatical structures of Aymara and Cuzco Kechua. Doctoral dissertation, UC Berkeley.

Gainor, Brian, Lai, Regine, and Heinz, Jeffrey. (2011). Computational characterizations of vowel harmony patterns and pathologies. Talk given at WCCFL 29, University of Arizona, Tucson, AZ.

Gildea, Daniel and Jurafsky, Daniel. (1996). Learning bias and phonological-rule induction. *Computational Linguistics 22* (4), 497-530.

Harris, James and Halle, Morris. (2005). Unexpected plural inflections in Spanish: Reduplication and metathesis. *Linguistic Inquiry 36* (2), 195-222.

Heinz, Jeffrey. (2007). The inductive learning of phonotactic patterns. Doctoral dissertation, UCLA.

Heinz, Jeffrey. (2009). On the role of locality in learning stress patterns. *Phonology 26,* 303-351.

Heinz, Jeffrey. (2010). Learning long-distance phonotactics. *Linguistic Inquiry 41,* 623-661.

Heinz, Jeffrey, and Idsardi, William. (2011). Sentence and word complexity. *Science, 333,* 295-297.

Johnson, C. Douglas. (1972). *Formal Aspects of Phonological Description*. The Hague: Mouton.

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

Koirala, Cesar. (2010). Strictly local relations. Ms.

Lipski, John M. (1990). Metathesis as template matching: A case study from Spanish. *Folia Linguistica Historica 11*, 89-104.

Mohri, Mehryar. (1997). FSTs in language and speech processing. *Computational Linguistics 23,* 269-311.

Oncina, José, García, Pedro, and Vidal, Enrique. (1993). Learning subsequential transducers for pattern recognition interpretation tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence 15*(5), 448-457.

Rohlfs, G. (1924). *Griechen und Romanen in Unteritalien: En Beitrag zur Geschichte der unteritalienischen Grazitat.* Biblioteca dell Archivum Romanium, Series 2,7. Geneve: Leo S. Olschki.

Shieber, Stuart. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy 8*, 333-343.

Proceedings of the 29th West Coast Conference on Formal Linguistics

edited by Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman

Cascadilla Proceedings Project Somerville, MA 2012

Copyright information

ISBN 978-1-57473-451-5 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Chandlee, Jane, Angeliki Athanasopoulou, and Jeffrey Heinz. 2012. Evidence for Classifying Metathesis Patterns as Subsequential. In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, ed. Jaehoon Choi et al., 303-309. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2715.