An IBSP Description of Sanskrit /n/-Retroflexion

Anonymous NAACL submission

Abstract

Graf and Mayer (2018) analyze the process of Sanskrit /n/-retroflexion (nati) from a subregular perspective. They show that nati, which might be the most complex phenomenon in segmental phonology, belongs to the class of input-output tier-based strictly local languages (IO-TSL). However, the generative capacity and linguistic relevance of IO-TSL is still largely unclear compared to other recent classes like the interval-based strictly piecewise languages (IBSP; Graf, 2017, 2018). This paper argues that nati is also IBSP, albeit at the cost of a much more convoluted description.

1 Introduction

Research in computational phonology has determined that all phonological patterns fit in the class of finite-state languages (Kaplan and Kay, 1994). The study of subregular phonology explores how characterizations of phonological phenomena can be further restricted by identifying suitable subclasses of the regular languages. This route of study enables us to formally classify the bounds on complexity of phonological computations, which provides new insights for typology and learnability (see Heinz 2018 and references therein).

One phenomenon that has proven to be particularly complex is /n/-retroflexion in Sanskrit, also known as nati. The nasal /n/ undergoes retroflexion whenever it appears immediately before a sonorant and there is a retroflex somewhere to its left. While this interaction of local and non-local factors is already unusual, the true complexity of the process comes from various blocking effects. It has been known since Graf (2010) that nati — when viewed as a phonotactic constraint on surface forms — is star-free. Recently, an alternative bound has been established in the form of input-output tier-based strictly local languages (IO-TSL; Graf and Mayer, 2018).

IO-TSL is an extension of the empirically well-supported class TSL (Heinz et al., 2011). Whereas the subclasses I-TSL and O-TSL of IO-TSL enjoy independent empirical support (De Santo and Graf, 2019; Mayer and Major, 2018), IO-TSL seems to be needed for no other phonological phenomena besides nati. In addition, the formal properties of IO-TSL are not well-understood. It isn’t even known whether IO-TSL is a subclass of the star-free languages. By contrast, the class of interval-based strictly piecewise languages (IBSP; Graf, 2017, 2018) is properly star-free, handles a wide range of phonotactic phenomena, and has even been applied to syntax (Shafiei and Graf, 2019). For all these reasons, an IBSP description of nati would be preferable to the current IO-TSL description.

In this paper, I argue that nati can be given an IBSP description, but the resulting grammar is much more convoluted than the IO-TSL analysis. While the basic cases of nati are very natural from an IBSP perspective, the interactions of blocking effects muddy this clear picture. The structure of the paper is as follows: IBSP is formally defined in Section 2, adapting the more general format proposed in (Graf, 2018). Section 3 then walks the reader through the nati analysis, starting from the simplest case and refining the IBSP grammar with each new complication. Section 4 reflects on the status of the analysis and what limitations of IBSP make nati so difficult to account for.

2 Preliminaries

Graf (2017) first defined the class of interval-based strictly piecewise (IBSP) string languages as an extension of the strictly piecewise (SP) languages (Rogers et al., 2010). IBSP enriches SP with locality domains, and the checking of SP-dependencies is limited to these locality domains.
IBSP properly subsumes SP, but also the classes SL and TSL, all three of which play a major role in subregular phonology. Graf (2018) further generalizes the format of locality domains to account for phenomena that had previously been analyzed in terms of I-TSL. Only this more general version can handle nati.

Intuitively, an IBSP intervals involves definitions of (i) the left and right domain edge, (ii) a finite number \( k \) of open slots, and (iii) the fillers that can occur between open slots. Fillers and domain edges are defined through \( k \)-intervals, also called \( k \)-vals. The IBSP grammar also supplies a list of forbidden \( k \)-grams. A string is well-formed iff there is no way to instantiate the \( k \)-val in such a manner that the configuration of open slots matches a forbidden \( k \)-gram.

While IBSP is originally defined in terms of first-order logic (Graf, 2017), I adopt the newer definition of Shafiei and Graf (2019) as it also subsumes the generalized intervals of Graf (2018).

**Definition 2.1** (k-val). A segmented \( k \)-interval \( (k \geq 0) \) over alphabet \( \Sigma \), or simply segmented k-val, is a tuple \( \langle L, R, F_i \rangle_{0 \leq i \leq k} \) such that:

- \( L, R \subseteq \Sigma \cup \{\varepsilon\} \) specify the left edge and right edge, respectively, and
- \( F_i \subseteq \Sigma \) specifies the \( i \)-th filler slot.

**Definition 2.2** (IBSP-k). Let \( \Sigma \) be some fixed alphabet and \( \times, \times \notin \Sigma \) two distinguished symbols. An IBSP-k grammar over \( \Sigma \cup \{\times, \times\} \) is a pair \( G := \langle i, S \rangle \), where \( i \) is a segmented \( k \)-val over \( \Sigma \cup \{\times, \times\} \) and \( S \subseteq (\Sigma \cup \{\times, \times\})^k \) is a set of forbidden \( k \)-grams. A string \( s \in \Sigma^* \) is generated by \( G \) iff there is no \( k \)-gram \( u_1 \ldots u_k \in S \) such that \( \times^{k} s \times^{k} \) is a member of the language

\[
(\Sigma \cup \{\times, \times\})^* \times L \times F_0^* \times \{u_1\} \times F_1^* \times \{u_2\} \times \ldots \times F_{k-1}^* \times \{u_k\} \times F_k^* \times R \times (\Sigma \cup \{\times, \times\})^*
\]

The language \( L(G) \) is the set of all \( s \in \Sigma^* \) that are generated by \( G \). A stringset \( L \) is IBSP-k iff \( L = L(G) \) for some IBSP-k grammar \( G \).

The reader may skip ahead to (1) and (2) for a depiction of a concrete IBSP interval and its application to an illicit string.

In IBSP, all possible instantiations of a locality domain must be evaluated. If at least one of them yields a match for an illicit \( k \)-gram, the whole string is discarded. By default, fillers allow each open slot to be arbitrarily far away from the next one. However, adjacency of the \( i \)-th and \( i + 1 \)-th open slot can be enforced by stipulating \( F_{i+1} = \emptyset \).

Mixing such empty fillers with normal fillers allows IBSP to capture phonotactic constraints in which local and non-local dependencies interact. As will see next, this isn’t needed for simplified version of nati, but will be crucial once the full range of facts is considered (Sec. 3.3 and subsequent sections).

### 3 Data and Analysis

**Nati** is a left-to-right long-distance assimilation process with a single trigger, a single target, and several conditions for blocking. While nati is usually described as a process — i.e. a mapping from underlying forms to surface forms — I treat it as a phonotactic phenomenon. That is to say, nati is reanalyzed as a constraint on the distribution of \([n]\) in surface forms, making it a matter of string languages rather than string transductions. This is in line with the previous work done by Graf and Mayer (2018), which will henceforth be referred to as GM.

The discussion starts with the simplest cases of nati and continually refines the IBSP description as new data is considered. The final version is presented in Sec. 3.5.

Several notational conventions will be adopted for the remainder of this paper: Sanskrit examples have their triggers and targets bolded, while active blockers are underlined. All the examples are taken from GM and Ryan (2017). IBSP interval diagrams are represented in a pictorial fashion: domain edges are large, green rectangles, fillers are vertically offset boxes in red, and open slots are blue squares.

3.1 Long-distance assimilation

**Nati** starts out with the basic constraint that a nasal target /\([n]\)/ becomes \([n]\) when preceded arbitrarily far to the left by a non-lateral retroflex continuant in \(\{\overline{\theta}, \overline{\lambda}, \overline{\lambda}^\theta, \lambda, \overline{\lambda}\}\). GM formalize this as the constraint “no \([n]\) may appear in the context \(R \ldots \).”, where \( R \) is one of the triggers listed in the preceding sentence.

GM’s constraint is easily expressed in terms of IBSP. Our grammar consists of a single forbidden unigram, which is \([n]\). The interval spans from \( R \) to the right word edge \$. Fillers may contain any-
thing except a word edge, which captures that nati cannot apply across word boundaries.

(1) IBSP interval (Version 1)

$$\text{R} \quad \text{n} \quad \text{S}$$

For the sake of succinctness, the interval above lists the forbidden unigram directly in the open slot. While this is non-standard, I believe it makes the analysis easier to follow once the complexity of the intervals starts to increase.

Table 1 lists some data points that are relevant for this base case. The form of the instrumental singular suffix /-ena/ alternates based on whether the root it attaches to contains a trigger for nati. For the sake of exposition, I also include an illicit nonce variation (indicated by the gloss “N/A”).

| Form      | Gloss      | Nati? | Licit? |
|-----------|------------|-------|--------|
| kâm-ena   | ‘by human’ | ×     | ✓      |
| manu-äja  | ‘by desire’| ✓     | ✓      |
| manu-ëna  | N/A        | ✓     | ×      |

Table 1: Forms showing basic nati; Ryan (2017), p. 305

The reader may wonder why an analogous nonce form kâm-ëna isn’t included in Tab. 1. In this nonce form, [n] would undergo nati without a suitable trigger, which should be illicit. However, this presupposes a view of nati as a process. From the perspective of phonotactics, it is not obvious that this nonce form is actually illicit because [n] can occur independently of nati. The phonotactics of nati only concern the distribution of [n], not [n], so only the former need to be considered here.

Let us now see how the locality domain in (1) captures the well-formedness of the first two forms in Tab. 1 while also ruling out the illicit nonce form. First, kâm-ëna is well-formed because it lacks a retroflex, so there is no suitable left edge for the interval in (1). Hence the locality domain cannot be established at all, so there are no open slot configurations to check against the list of forbidden unigrams. As a result, the string is well-formed.

The second example is manu-ëna, which does allow for numerous instantiations of the interval. In all of them, the interval spans from [s] to the right word edge, and the only difference is what segments make up the fillers and which one ends up in the open slot. But since manu-ëna does not contain any [n], the open slot never matches the forbidden unigram. Consequently, this string is also deemed well-formed. In contrast to the first example, where well-formedness followed from the inability to instantiate any locality domain, this example allows for many distinct instantiations but none of them yield a forbidden configuration of open slots.

This leaves us with the illicit manu-ëna. It works exactly like the second case, except that now there is an instantiation that results in a match with the forbidden unigram n. This particular instantiation is depicted below.

(2) IBSP interval: manu-ëna

$$\text{manu} \quad \text{j-e} \quad \text{n} \quad \text{a} \quad \text{S}$$

So far, IBSP has not done anything that could not be accomplished by simpler means, e.g. an SP grammar. But as we start adding on conditions and exceptions, IBSP intervals will quickly become indispensible.

### 3.2 Unconditional blocking by intervening coronals

We now turn to the first of the nati-blocking effects: /n/-retroflexion can be blocked if a coronal segment appears between trigger and target. The set of relevant coronals includes retroflexes but excludes the glide [j] as the latter is both a sonorant and a coronal — see Ryan (2017) for further discussion. Table 2 lists a particular example of coronal blocking, an illicit nonce form, and a nonce form that illustrates what the surface form would look like if coronals weren’t blockers.

| Form       | Gloss        | Nati? | Blocking? | Licit? |
|------------|--------------|-------|-----------|--------|
| vaq-anamam | no gloss     | x     | ✓         | ✓      |
| vaq-anamam | N/A          | ✓     | x         | x      |

Table 2: Forms showing blocking by intervening coronals; Hansson (2001), p. 227

In GM, the forbidden context for [n] is updated to $R\overline{C}...$, where $C$ matches every segment that is not a coronal, including [j]. To represent this in IBSP, we modify the first filler in (1) so that it may not contain any coronals either. If a string contains a coronal, it must go in the open slot or the second filler. Either way, no subsequent [n] can appear in
the open slot, and consequently the string will be
deemed well-formed.

(3) **IBSP interval (Version 2)**

At the same time, strings without coronals will still
be judged illicit. This is illustrated below for the
nonsense form *vaqm-ananām.*

(4) **IBSP interval: vaqm-ananām**

Note that [n] itself is a coronal blocker, so any
subsequent [n] in a word loses its eligibility as
a target for *nati.* The only exception to this is
geminate /nn/ sequences where both /n/ become
retroflexed. However, this could also be treated
as a separate process of progressive local assimila-
tion. I put this issue aside for now, but it will be
revisited in Sec. 4.

### 3.3 Mandatory adjacency to sonorant

In order for /n/ to undergo *nati,* it must also be
immediately followed by a vowel, a glide, [m], or
[n] itself. More succinctly, the following segment
must be a non-liquid sonorant (Whitney, 1889).

For example, in the form *bghman,* *nati* does not
apply as [n] occurs at the very end of the word
without any subsequent sonorant. Similarly, *nati*
does not apply in *caṇ-a-n-ti,* in this case because
[t] is not a sonorant. Sanskrit has some nasals be-
sides [m] and [n] that are non-liquid sonorants, but
as those cannot follow [n] for independent reasons
(Emeneau, 1946) they do not matter for the pur-
poses of this paper.

| Form     | Gloss               | *Nati?* | Sonorant? | *Licit?*
|----------|---------------------|---------|-----------|---------|
| *caṇ-a-n-ti* | ‘wander (3Pl)’   | ×       | ×         | ✓       |
| *bghman*   | ‘brahman’          | ×       | ×         | ✓       |
| *bghmana*  | N/A                | ✓       | ✓         | ✓       |

Table 3: Forms showing mandatory adjacency to son-
orant; (Hansson, 2001), p. 229 and (Ryan, 2017), p. 318

GM represent the new illicit context for [n] as
*R*C...*S,* where *S* is a suitable sonorant. We will
use the same definition of *S* to add a second open
slot to the interval in (3). The list of illicit uni-
grams is now expanded to illicit bigrams. It is no
longer just [n] that is forbidden, but rather any bi-
gram of the form *nS.* Keep in mind that coronal
blocking is still active, though.

(5) **IBSP interval (Version 3)**

This is the only interval that could possibly cause
the IBSP grammar to reject the string as the first
open slot is filled by n. However, as the second
open slot is not a sonorant, the open slot configura-
tion does not match any of the forbidden bigrams.

The well-formedness of *bghdana* follows for the
very same reason: there is no way of instantiat-
ing the locality domain so that the two open slots
would contain [n] and a sonorant, respectively.

At the same time, *bghdana* is correctly ruled
out as illicit.

(7) **IBSP interval: bghmana**

### 3.4 Conditional blocking by preceding velar
and labial plosives

Coronal consonants are not the only blockers of
*nati:* velar and labial plosives also block the pro-
cess, but only if 1) the plosive immediately pre-
cedes the target nasal, and 2) a left root bound-
ary (✓) occurs somewhere between the trigger and
the plosive. Blocking is contingent on both condi-
tions being met, as is exemplified by the data in
Tab. 4. In *paṛ-ṭ/niṇ-ṭa-ti,* *nati* still occurs across
a left root boundary due to the absence of a plosive
immediately before /n/. In √\textipa{aga-nh}, \textipa{nati} can target an \(n\) after an immediately preceding velar plosive /g/ because the left root boundary does not occur between the triggering retroflex and the plosive. Only in \(\textipa{(ab)i-pqa-√gh\text{\textipa{n}-an-ti)}\) does \textipa{nati} fail as there is both a plosive and a root boundary, both of which occur in the relevant positions.

The filler before the third open slot is set to \textipa{none} so that it can only be filled by whatever segment immediately precedes /n/. The fillers surrounding the first open slot are more complex. The ban against coronals is carried over from coronal blocking, but in addition these filters may not contain a root boundary either. As a result, a root boundary that occurs somewhere between the triggering retroflex and a suitable plosive must go into the first open slot. The conjunction of all these factors ensures that if a string contains a suitable root boundary and plosive, they will always occur in the first two open slots.

In the next step, we expand the list of forbidden bigrams of the form \textipa{nS} to forbidden 4-grams of the form \textipa{φnS}. Here \textipa{φ} represents a large number of bigrams. As \textipa{nati} is only blocked whenever the first open slot is a root boundary and the second open slot is a plosive, /n/\textipa{S} is illicit if

1. the first open slot is not a root boundary, or
2. the second open slot is not a plosive, or
3. both 1 and 2 hold.

Hence \textipa{φ} corresponds to any combination of segments that matches one of the three conditions above. If the first two open slots in an instantiated interval do not match \textipa{φ}, \textipa{nati} won’t be enforced, capturing the described blocking effect. This is illustrated below for \(\textipa{(ab)i-pqa-√gh\text{\textipa{n}-an-ti)}\).

\[
\begin{array}{c}
\textbf{(9) IBSP interval: } \textipa{(ab)i-pqa-√gh\text{\textipa{n-an-ti)}} \\
\end{array}
\]

Any configuration where the first two open slots are not √ and a plosive will match \textipa{φ}, triggering a \textipa{nati} violation if the remaining two open slots are filled by /n/ and a sonorant. As a concrete example, consider the nonce form \textipa{pqa-√min-a-ti}.

\[
\begin{array}{c}
\textbf{(10) IBSP interval: } \textipa{pqa-√min-a-ti} \\
\end{array}
\]
The reader is urged to verify for themselves that the remaining forms in Tab. 4 are handled correctly by this grammar.

One additional wrinkle is that the introduction of new open slots has created an “escape hatch” for coronals. In previous versions, a coronal had to go into the first or second open slot, or the third filler. But these are now the third and fourth open slot and the fifth filler. While coronals are still banned in the first and second filler, they could go into the first or second open slot. But since \( \phi \) currently matches coronals, too, we no longer capture coronal blocking. Fortunately, the fix is easy. We further restrict the shape of \( \phi \) so that it does not match any open slot configuration with a coronal. Overall, this leaves the following patterns for \( \phi \):

\[
\begin{array}{c|c|c|c}
1 & 2 & 3 & 4 \\
\hline
\checkmark & -P \land -C & -C & -R \\
\checkmark & -P \land -C & -C & -R \\
\checkmark & -P \land -C & -C & -R \\
\checkmark & -P \land -C & -C & -R \\
\end{array}
\]

Given a list of suitable list of segments for Sanskrit, \( \phi \) can be compiled out into a list of bigrams. These bigrams are then prefixed with every possible instantiation of \( nS \) to arrive the list of forbidden 4-grams.

3.5 Conditional blocking by following retroflex

Even though the grammar in (8) is already fairly complicated, it still does handle the last layer of \( \text{nati} \): if a retroflex appears arbitrarily far to the right of the target \( /n/ \), \( /n/-\text{retroflexion} \) may be blocked. Blocking only occurs when both of the following two conditions are met: 1) a left root boundary intervenes between the trigger and the target, and 2) there is no coronal between the target \( /n/ \) and blocking retroflex. Condition 2) is particularly peculiar. Essentially, the appearance of a coronal consonant between \( /n/ \) and its following retroflex blocks the blocking of \( \text{nati} \) by said retroflex, so that the process applies as usual.

| Form   | Gloss       | Nati? | Licit? |
|--------|-------------|-------|--------|
| \( \text{pt}-\checkmark /\text{nasi}-\text{tum} \) | 'to vanish (inf.)' | \X \ | \checkmark |
| \( \text{pt}-\checkmark /\text{nasi}-\checkmark \) | 'leader' | \checkmark | \checkmark |
| \( \text{pt}-\checkmark /\text{a}-\text{t} \) | 'unite (2s)' | \checkmark | \checkmark |

Table 5: Ryan (2017), p. 325

The form \( \text{pt}-\checkmark /\text{nasi}-\text{tum} \) in Tab. 5 shows the following retroflex action as a blocker when a left root boundary intervenes between \( /n/ \) and \( /s/ \). On the other hand, the retroflex is not a blocker in \( \text{pt}-\checkmark /\text{nasi}-\checkmark \) due to the coronal intervening between \( /n/ \) and \( /a/ \). Finally, \( \text{pt}-\checkmark /\text{a}-\text{t} \) is a case where the retroflex does not block in the absence of an intervening root boundary.

We can follow the same approach as in section 3.4 to handle this complication. That is to say, we include yet another two conditional slots following the target nasal, and its mandatory adjacent sonorant. As the interval now gets exceedingly long, graphical depictions have to be broken up again across multiple lines.

(11) IBSP interval (Version 5, Final)

This time, open slot 3 tracks the presence of a coronal, and open slot 4 indicates whether a retroflex is present. Once again we have to forbid these segments in the neighboring fillers to ensure that if such a segment is present, it must go into one of these open slots.

We then expand the list for forbidden 4-grams to forbidden 6-grams. The 4-gram pattern \( \phi nS \) is expanded to \( \phi nS \phi' \). Just like \( \phi \) describes the illicit segments for 1 and 2, \( \phi' \) handles open slots 3 and 4 in (11). However, \( \phi' \) cannot be described independently of \( \phi \) as the relevance of slots 3 and 4 for blocking depends on the presence of a root boundary in open slot 1. Hence the options for \( \phi \) and \( \phi' \) have to be specified in conjunction:

\[
\begin{array}{c|c|c|c|c}
1 & 2 & 3 & 4 & 5 \\
\hline
\checkmark & -P \land -C & -C & -R & -\checkmark \\
\checkmark & -P \land -C & -C & -R & -\checkmark \\
\checkmark & -P \land -C & -C & -R & -\checkmark \\
\checkmark & -P \land -C & -C & -R & -\checkmark \\
\end{array}
\]

The interval in (11), with the list of forbidden 6-grams above, is the final version of the IBSP grammar for \( \text{nati} \) (although other potential variants are...
discussed in Sec. 4). This is a good point to reevaluate some of the earlier data points. For example, we can model some examples that illustrate conditional blocking of intervening velar/labial plosives like so:

(12) **IBSP interval: pqa\(\sqrt{\text{mi\-n\-a\-t}}\)**

\[
\text{p} \quad [s, C, \sqrt{\text{a}}] \quad \text{m} \quad [s, C, \sqrt{i}]
\]

The instantiated locality domain looks quite similar to its previous iteration in (10). The main difference is that rather than having [t] and [i] in the filler following the nS sequence, those segments are pushed into the open slots that check for the presence of an anti-blocking coronal and/or blocking retroflex. The configuration of conditional slots matches \(\sqrt{1}, \neg P \land \neg C, \neg C, \neg R\), which is one that enforces *nati*. Consequently, the presence of an [n] in the open slot where it is forbidden causes the string to be rejected. If [n] had undergone *nati* as required, the string would not have been deemed illicit by the grammar.

The string \(\text{pqa}\sqrt{\text{g}^{'\text{n\-a\-t}}\text{ti}}\), on the other hand, is still well-formed. Even when [n] appears in the open slot, this does not yield an illicit configuration of open slots due to the presence of a root boundary in open slot 1 and a plosive in open slot 2.

(13) **IBSP interval: \((ab)\sqrt{i}\text{pqa}\sqrt{\text{g}^{'\text{n\-a\-t}}\text{ti}}\)**

\[
(ab)^{\text{pi}} [s, C, \sqrt{\text{a}}] [s, C, \sqrt{i}]
\]

1. Furnish an open slot for every type of segment that can potentially matter for the dependency.
2. If an open slot needs to track the presence of some segment of type \(X\), do not allow the surrounding fillers to contain \(X\).
3. Whatever implicational relations hold between the relevant segments are compiled out into a list of forbidden \(k\)-grams.

While each step is conceptually simple, the sheer number of open slots and potential combinations of segments make an IBSP analysis of *nati* a daunting task.

In fact, the analysis presented here still involves major simplifications. As mentioned in Sec. 3.2, geminate /\(n^{'}/\) becomes geminate /\(n^{'}/\) under *nati*. This is not captured by the current grammar, but corresponding modifications could be made. If geminate /\(n^{'}/\) is modeled as underlying /\(nn^{'}/\), then the list of forbidden 6-grams can be modified to also block /\(n^{'}/\). Then /\(n^{'}/\) would be the only possible surface form. If, on the other hand, /\(n^{'}/\) is a single symbol, then the 6-grams must be modified such that /\(n^{'}/\) is forbidden even if the following segment is not a sonorant (since the geminate, metaphorically speaking, acts as its own sonorant). Needless to say, the resulting list of forbidden 6-grams obfuscates the relevant dependencies even more.

Another problem is that as the size of the \(k\)-val grows, shorter strings are automatically considered well-formed. An interval with 6 open slots cannot be instantiated in a string that only consists of 5 symbols. Graf (2017) allow strings to be padded out by additional edge markers to enforce the required minimal length. But this means that the list of 6-grams also needs to be extended to handle cases where some open slots contain word edges. At this point, inspecting the grammar for correctness is no longer humanly possible.

In the other direction, the interval may still be too large. For instance, coronals cannot go into the first or second filler, leaving only the first open slot as an option for a coronal somewhere to the left of /\(n^{'}/\). If a string contains two coronals, neither one of which is adjacent to /\(n^{'}/\), the interval cannot be instantiated at all. In this case this is unproblematic since coronals would block *nati* anyways, so either way the string is deemed well-formed. But the situation is reversed with coronals after

4 Discussion and conceptual remarks

The IBSP analysis developed over the course of Sec. 3 is admittedly convoluted. At a high level of abstraction, the strategy employed here boils down to a few simple tricks:
\(/n/\), which undo blocking of \textit{nati} by a retroflex. If a string contains two coronals between \(/n/\) and such a retroflex, the interval won’t be instantiated and the string will incorrectly be treated as well-formed. Again one could fix this by adding more open slots and modifying the list of forbidden \(k\)-grams. But the resulting grammar would be utterly unintelligible.

For all these reasons, IBSP does not provide an insightful or elegant perspective of \textit{nati}, in particular compared to GM’s IO-TSL treatment. Nonetheless it is a useful observation that \textit{nati} can be given an IBSP description. The discrepancy we find between IBSP and IO-TSL touches on a larger issue for subregular phonology: to what extent should succinctness and elegance of description be a criterion in the classification of empirical phenomena? If formalism \(X\) strictly speaking generates the right string pattern, but a more powerful class provides a more natural perspective than \(X\), which one of the two is closer to cognitive reality?

5 Conclusion

I have argued that a process as complex as \textit{nati}, which can be viewed as an interaction between local and non-local dependencies with intervening material that provides blocking effects, can be modeled in IBSP. Since IBSP enjoys independent empirical support, this result makes \textit{nati} look like less of an outlier in the phonological landscape. However, the proposed grammar is fairly complicated and lacks linguistic naturalness. Future work could revisit these findings along two dimensions. On a formal level, it might be possible to extend IBSP grammars with mechanisms that allow for more succinct descriptions without increasing generative capacity. From a linguistic perspective, one might try to reassess the empirical status of \textit{nati} with respect to which of its components are most natural under an IBSP-analysis. If these aspects turn out to be on empirically shaky ground, this might provide indirect evidence for IBSP as a model of natural language phonotactics.

\[\text{References}\]

Aniello De Santo and Thomas Graf. 2019. Structure sensitive tier projection: Applications and formal properties. In Formal Grammar, pages 35–50, Berlin, Heidelberg. Springer.

M.B. Emeneau. 1946. The nasal phonemes of sanskrit. Language, page 22(2):86–93.

Thomas Graf. 2010. Comparing incomparable frameworks: A model theoretic approach to phonology. University of Pennsylvania Working Papers in Linguistics, 16(2):Article 10.

Thomas Graf. 2017. The power of locality domains in phonology. Phonology, 34:385–405.

Thomas Graf. 2018. Locality domains and phonological c-command over strings. In Proceedings of "NELS" 2017.

Thomas Graf and Connor Mayer. 2018. Sanskrit n-retroflexion is input-output tier-based strictly local. In Proceedings of "SIGMORPHON" 2018, pages 151–160.

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

Jeffrey Heinz. 2018. The computational nature of phonological generalizations. Phonological Typology, Phonetics and Phonology, pages 126–195.

Jeffrey Heinz, Chetan Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints in phonology. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 58–64.

Ronald M. Kaplan and Martin Kay. 1994. Regular models of phonological rule systems. Computational Linguistics, 20(3):331–378.

Connor Mayer and Travis Major. 2018. A challenge for tier-based strict locality from uyghur backness harmony. volume 10950 of Lecture Notes in Computer Science, pages 62–83. Formal Grammar, Springer, Berlin, Heidelberg.

James Rogers, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Vischer, David Wellcome, and Sean Wibel. 2010. On languages piecewise testable in the strict sense. 6149:255–265.

Kevin Ryan. 2017. Attenuated spreading in sanskrit retroflex harmony. Linguistic Inquiry, 48 (2):299–340.

Nazila Shafiei and Thomas Graf. 2019. The subregular complexity of syntactic islands. Ms., Stony Brook University.

William Dwight Whitney. 1889. Sanskrit Grammar. Oxford University Press, London.