Representing phonology in the formal comparison of phonological representations

Workshop on The Role of Representation in Computational Phonology

13 January 2026 – King's College (University of Cambridge) – OCP26

Nick Danis, Washington University in St. Louis
nsdanis@wustl.edu

https://www.nickdanis.com/

Why formally compare representations?

(non-exhaustive)

1. Compare some input representation to some target/output representation of some process to learn about the computational nature of that process (e.g. Jardine 2017)
2. Compare some potential lexical representation with a target surface representation to determine "what type of information needs to be directly stored in long-term memory" (Nelson 2024)
3. Compare proposal A from the literature to proposal B from the literature to determine differences in predictions/complexity/content/?? (e.g. Strother-Garcia 2019, Oakden 2020, Jardine, Danis & Iacoponi 2021)

Meta-theoretical comparisons

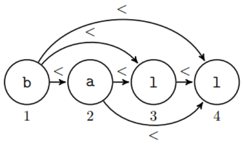
General enterprise

- Define phonological structures in finite model theory (Libkin 2004)
- Provide logical transductions, or translations, from one theory to another (Courcelle 1994)

For phonologists

1. What can computational phonologists learn about or conclude from transductions defined over representations across theoretical proposals?
2. What can practitioners of such proposals conclude from the same transductions?
3. ***Under what circumstances can two representations be considered notationally equivalent?***

Figure 3.2: A visual representation of $\mathcal{M}^{<}_{ball}$



$$\mathcal{M}^{<}_{ball} \stackrel{\text{def}}{=} \langle \mathcal{D}; \{R_<, R_a, R_b, R_l\}; \varnothing \rangle$$

$$\mathcal{D} \stackrel{\text{def}}{=} \{1, 2, 3, 4\}$$

$$R_< \stackrel{\text{def}}{=} \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,3\rangle, \langle 2,4\rangle, \langle 3,4\rangle\}$$

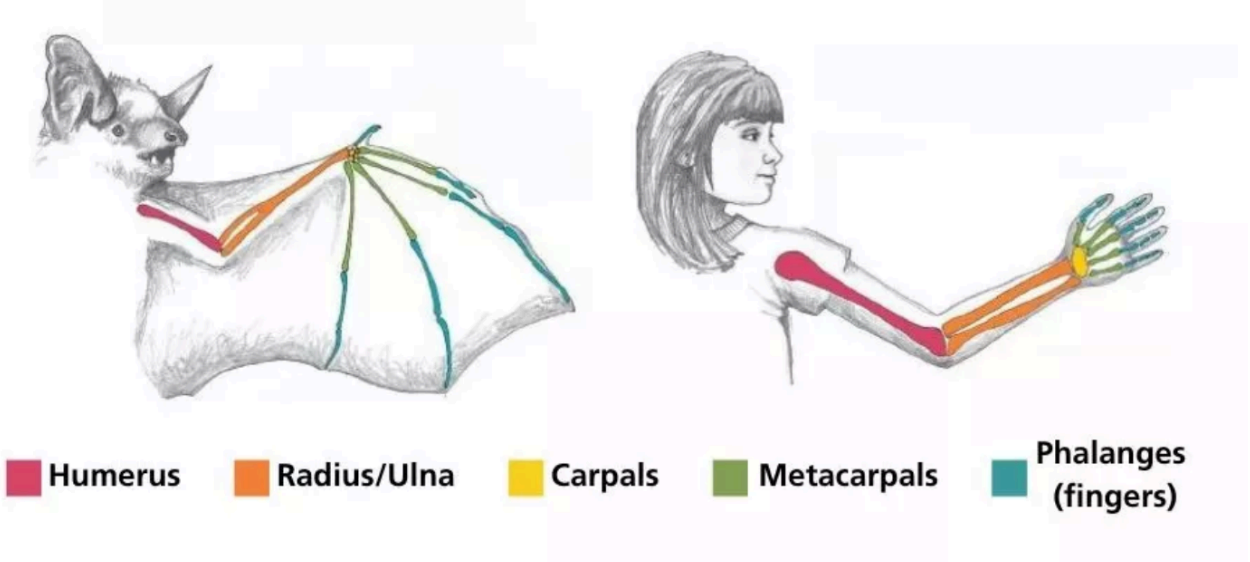
$$R_a \stackrel{\text{def}}{=} \{2\}$$

$$R_b \stackrel{\text{def}}{=} \{1\}$$

$$R_l \stackrel{\text{def}}{=} \{3, 4\}$$

from Strother-Garcia 2019: 19

An analogy



from https://www.nps.gov/subjects/bats/how-bats-fly.htm

Framework and preliminaries

Definitions

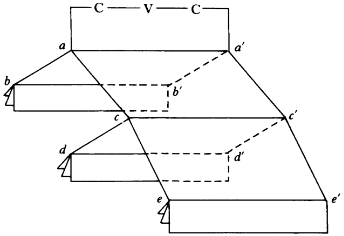
- I use **proposal** to mean some phonological theory or model as defined and proposed in the primary literature, and **theory** to mean its specific model-theoretic implementation (Libkin 2004, Strother-Garcia 2019)
- A **segmental theory** or proposal is one that focuses only on the structure of segments[1], as opposed to syllabic, metrical, or other types of possible phonological representations
  - **linear representational theory** here means those theories where there is only a single total linear ordering relation (or function) defined between segments in a theory (e.g. strings or feature matrices)
  - **nonlinear representational theory** then are those with more than a single ordering relation (or function) defined (e.g. autosegmental representations), or one using general precedence

[1] or segment-like things, as not every proposal intends to actually define what a *segment* is

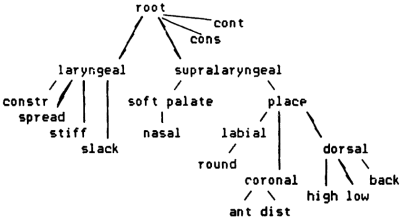
(Some) Segmental Proposals

$$\begin{bmatrix} +\text{voc} \\ -\text{cons} \\ +\text{high} \\ -\text{back} \\ -\text{ant} \\ -\text{cor} \\ +\text{cont} \\ -\text{nasal} \\ -\text{strid} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{voc} \\ -\text{cons} \\ +\text{high} \\ -\text{back} \\ -\text{ant} \\ -\text{cor} \\ +\text{cont} \\ -\text{nasal} \\ -\text{strid} \end{bmatrix}$$

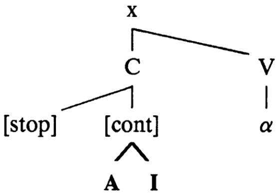
Chomsky & Halle 1968: 336



Clements 1985: 229



Sagey 1986: 2



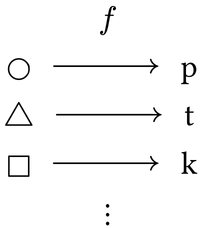
van de Weijer 1996: 77

Segmental Theories

In addition to their model-theoretic implementation, as a tool to aid in comparing segmental theories, we assume an **atomic segment mapping** for each:

Definition 1: Atomic segment mapping.

For each segmental proposal represented as a finite model, we also define a bijective mapping from the full range of possible segmental structures in that theory to atomic symbols (e.g. IPA characters that capture the intended linguistic meaning of said structures).



Note that there is some analytical freedom in how structures are mapped to IPA characters; this has consequences later.

Logical Transductions

- Illustrative transduction example from Nelson 2024: 106: a translation between strings (over *t, d, n, a*) and feature bundles (with [voi], [son], [syl])
- **output structure** defined in terms of **input structure**

$$\phi_{\texttt{voi}}(x) \stackrel{\text{def}}{=} d(x) \vee n(x) \vee a(x) \quad (4.13)$$

$$\phi_{\texttt{son}}(x) \stackrel{\text{def}}{=} n(x) \vee a(x) \quad (4.14)$$

$$\phi_{\texttt{syl}}(x) \stackrel{\text{def}}{=} a(x) \quad (4.15)$$

Rule (4.13) can be read: *some position x is labeled [voi] in the output if it is labeled d, or n, or a in the input*

- output structure is then an order of positions, each with multiple labels (unary relations) representing phonological features

Logical Transductions

- When the input is a featural representation, there is the following translation back into a string:

$$\phi_t(x) \overset{\text{def}}{=} \neg\mathtt{voi}(x) \wedge \neg\mathtt{son}(x) \wedge \neg\mathtt{syl}(x) \tag{4.16}$$

$$\phi_d(x) \overset{\text{def}}{=} \mathtt{voi}(x) \wedge \neg\mathtt{son}(x) \wedge \neg\mathtt{syl}(x) \tag{4.17}$$

$$\phi_n(x) \overset{\text{def}}{=} \mathtt{voi}(x) \wedge \mathtt{son}(x) \wedge \neg\mathtt{syl}(x) \tag{4.18}$$

$$\phi_a(x) \overset{\text{def}}{=} \mathtt{voi}(x) \wedge \mathtt{son}(x) \wedge \mathtt{syl}(x) \tag{4.19}$$

- The two theories are therefore **bi-interpretable** (formal definition forthcoming)
- Additionally, and crucially, not just structures but any rule/constraint/logical sentence given in terms of one theory is also translatable into the logic of the other theory; they are equally expressive

Interpretation Domains

- Miller 2001 provides a framework for comparing different syntactic formalisms also using model theory to explicitly define a *strong generative capacity* for syntax
- The framework of *interpretation domains* and *interpretation functions* allows for formal meta-theoretical comparisons
  - "That is, **we must be able to specify which are the intended interpretations of a structural description in a given formalism**. For this purpose we introduce the notion of *Interpretation Domain*. Interpretation Domains are set up to provide explicit characterizations of linguistically significant properties of sentences, independently of specific formalisms." (Miller 2001: 9, emph. mine)
- Example: one Interpretation Domain is predicted constituent structure of a natural language sentence. Formalism A is a phrase structure tree, formalism B is a dependency parse. An **Interpretation Function** is defined that takes this formalism and outputs a set-theoretic representation of the constituency structure of each; these can now be compared directly.

Three Interpretation Domains for Phonology

Contrast Preservation
Do the theories predict the same set of phonological contrasts?

Natural Class Preservation
Do the theories predict the same sets of natural classes?

Feature Class Preservation
Do the theories predict the same groups of features behave similarly?

- Each interpretation domain codifies some *intended linguistic property* of some potential theory
- Some interpretation domains might not be relevant for a specific comparison—this is expected and exactly why they must be defined and enumerated
- Each domain relates to logical transductions in different ways; some as a consequence, others as seemingly orthogonal
- **(Non-)equivalence across these domains must be considered when claiming notational equivalence across theories**

What do transductions tell us?

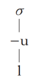
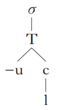
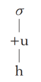
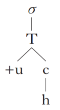
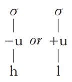
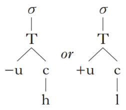
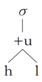
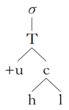
Contrast Preservation

- Discussed in detail in Oakden 2020: contrast preservation is when "no contrasts present in one model are lost in the process of translation into the other." (Oakden 2020: 263)
- The definition of **bi-interpretable** used in Oakden is stronger than that used in previous work (e.g. Strother-Garcia 2019), so a transduction that meets this criteria will be called here **strongly bi-interpretable**.

Definition 2: Strongly bi-interpretable.

Given two theories T_1 and T_2, an interpretation F from T_1 to T_2, and G from T_2 to T_1, the theories T_1 and T_2 are strongly bi-interpretable iff the mapping of $F \circ G$ "produces the same mapping as (i.e. is isomorphic to) the identity map that maps every bundled structure to itself", and likewise for $G \circ F$ (Oakden 2020: 281).

Oakden 2020

tone	Yip (1989)	Bao (1990)
low level L		
high level H		
mid level M		
high falling HM		

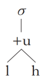
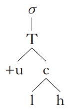
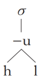
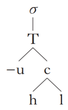
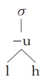
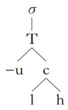
tone	Yip (1989)	Bao (1990)
high rising MH		
low falling ML		
low rising LM		

Table I

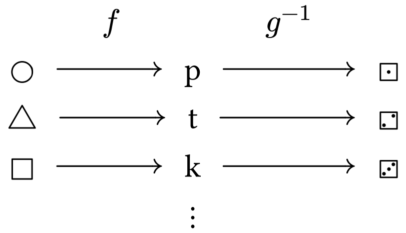
Level and contour tonal contrasts in Yip (1989) and Bao (1990).

Contrast Preservation

While Oakden 2020 uses *contrast preserving* to mean exactly those translations that are strongly bi-interpretable, the definition of contrast preservation offered here is one that is not dependent on the existence of a defined transduction.

Definition 3: Contrast Preservation (transduction neutral).

If f is the atomic segment mapping for some theory T_1 and g is the atomic segment mapping for some theory T_2, then T_1 and T_2 are contrast preserving iff $f \circ g^{-1}$ is a bijection (where g^{-1} is the inverse mapping, i.e. from atomic symbols to structures)



Contrast Preservation

Theorem 1: Contrast preservation follows from strong bi-interpretability.

If two segmental theories T_1 and T_2 are strongly bi-interpretable, then there exist atomic segment mappings for T_1 and T_2 that are contrast preserving.

Proof: If F is a translation from $T_1 \to T_2$, and G vice versa, then by definition if T_1 and T_2 are strongly bi-interpretable, $F \circ G$ and $G \circ F$ are isomorphic to the identity map. The identity map is a bijection, so $F \circ G$ and $G \circ F$ are bijective. If $F \circ G$ is bijective, then F and G are each bijective. Map every structure M in T_1 to some arbitrary unique symbol; this is its atomic segment mapping f. Associate that same symbol to $F(M)$; this is the atomic segment mapping g for T_2. $f \circ g^{-1}$ must then itself be bijective and therefore T_1 and T_2 are contrast preserving, by definition.

Contrast Preservation

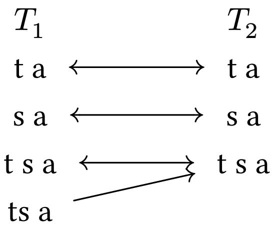
Splitting the definition of contrast preservation from that of (strong) bi-interpretability allows us to ask:

Can there be a situation where T_1 and T_2 are contrast preserving (as defined here), but are **not** strongly bi-interpretable under MSO?

- In a purely mathematical sense, probably?: two theories with the equal cardinality of structures are contrast preserving in a trivial sense
- In the extension of those representations ever proposed for phonology, probably not?
  - Nelson 2022 provides translations for strings and representations in articulatory phonology, which are fairly far apart on the representational spectrum

Contrast non-preservation and pattern complexity

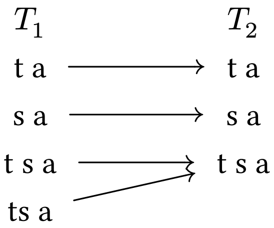
Inspired by conference website, consider the contour segment (T_1) vs. cluster (T_2) analyses of affricates such that some true affricate *ts* is in the atomic segment mapping codomain of T_1 but not T_2

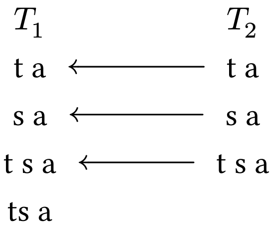


Spaces added between atomic elements for clarity.

T_1 has a contrast between clusters and true affricates, while T_2 does not, thus the two are **not contrast preserving** and further **are not strongly bi-interpretable**.

functions





surjection (but not injection): every element in the codomain (T_2) is mapped to from some element in the domain (T_1)

injection (but not surjection): no two elements in the domain (T_2) map to the same element in the codomain (T_1)

a **bijection** is both an injection and a surjection

Contrast non-preservation and pattern complexity

Theorem 2: Strongly bi-interpretable linear theories preserve complexity classes(?).

If a pattern is of certain complexity class C_k when modeled under some **linear** representational theory T_1, and T_1 is strongly bi-interpretable with T_2, then the pattern modeled under T_2 is of the complexity class C_j, where potentially $j = k$[2]

- Sketch of proof: this is essentially defining an isomorphism for formal languages of strings between two alphabets Σ and Σ'; such an operation there preserves the complexity class. Assuming only linear represtational theories means there is never a situation where a non-local process becomes local (or vice versa).
- Non-linear representations can definitely lower the complexity of a pattern: Jardine 2016: certain tonal patterns are non-local when computed over string representations, but local when computed over autosegmental representations.

[2]The k of the necessary constraint might change based on the exact configuration of the structure, but with finite models in a linear theory the k remains bounded.

What transductions *don't* tell us

Natural Class Preservation

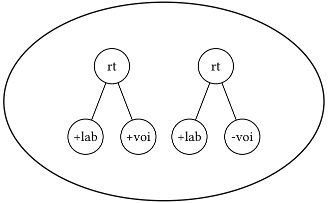
> In view of this, if a theory of language failed to provide a mechanism for making distinctions between more or less natural classes of segments, this failure would be sufficient reason for rejecting the theory as being incapable of attaining the level of explanatory adequacy. (Chomsky & Halle 1968: 335)

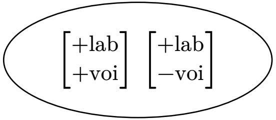
> In Logical Phonology […], rules refer to natural classes by definition: a statement that cannot be formulated in terms of natural classes is not a rule. (Volenec & Reiss 2020: 28)

- Two theories are **natural class preserving** if they predict the same sets of natural classes across their possible structures.
- In the respective grammatical systems for most proposed representations, the possible natural classes directly influence the rules or constraints over that structure.

Natural Class Preservation

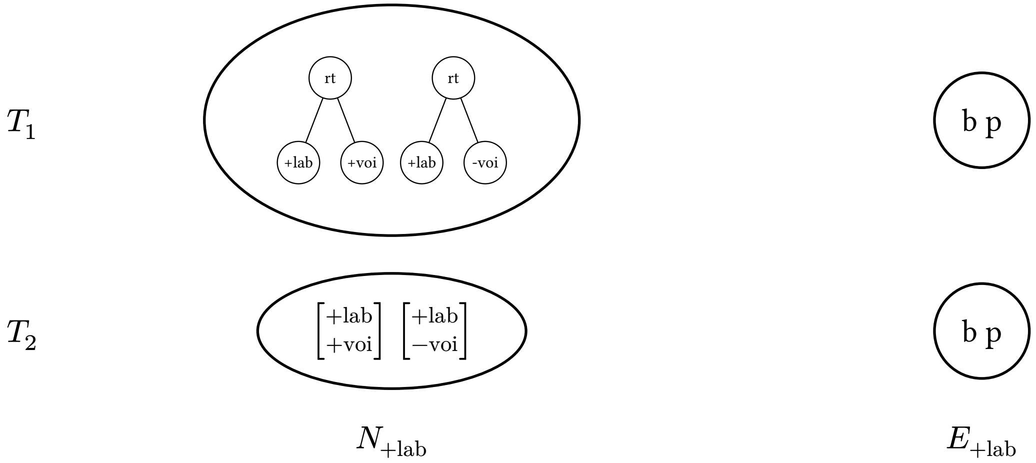
- Given some theory T, a **natural class** in that theory are all the segmental structures sharing some piece of connected substructure.
- For example, in a theory where a node can have the label **[+labial]**, then the natural class for **[+labial]** is the set of all segmental structures containing a node labeled **[+labial]** and none that do not.
- Because this is a set of *structures*, it cannot be compared directly against some other theory that builds structure differently:





Natural Class Extensions

If N is a natural class of structures in T as defined above, and f is the atomic segment mapping for T, then the **natural class extension** of N is the set $E = \{f(x) : x \in N\}$



Natural Class Extensions

Definition 4: Natural Class Preservation.

Two theories T_1 and T_2 are natural class preserving iff the set of all natural class extensions of T_1 exactly equals the set of all natural class extensions of T_2

Two theories: unified and v-features

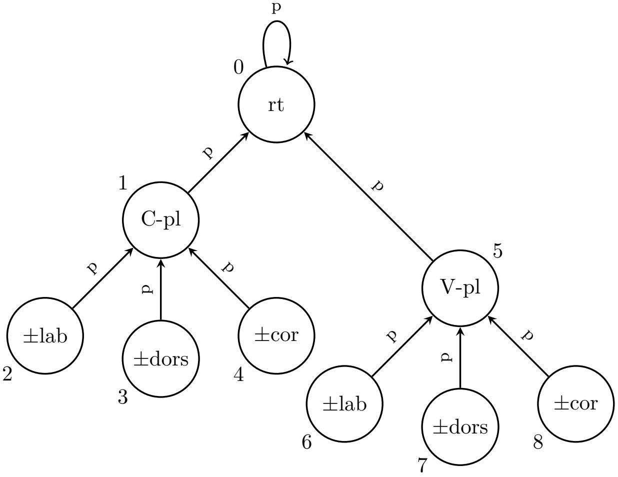


Figure 1: **unified** theory uses same unary labels for features distinguishing consonants and vowels, but includes V-place node

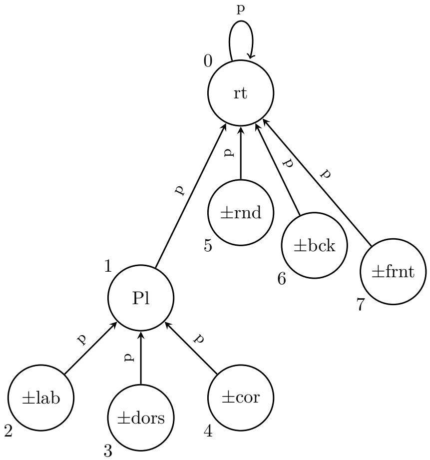


Figure 2: **v-features** theory uses unique labels for features distinguishing vowels, no separate V-place node

Two theories: unified and v-features

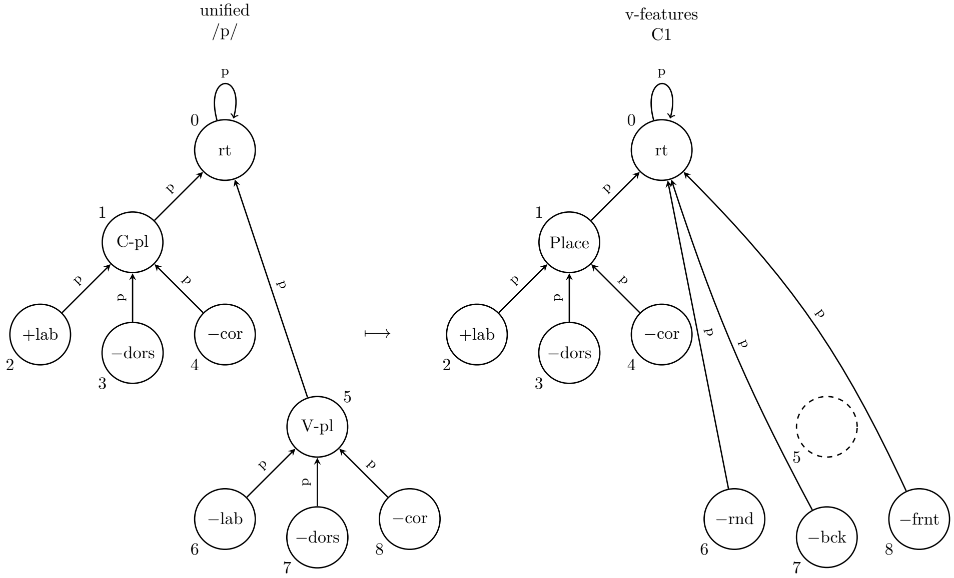


Figure 3: Output of transduction from unified to v-features

Two theories: unified and v-features

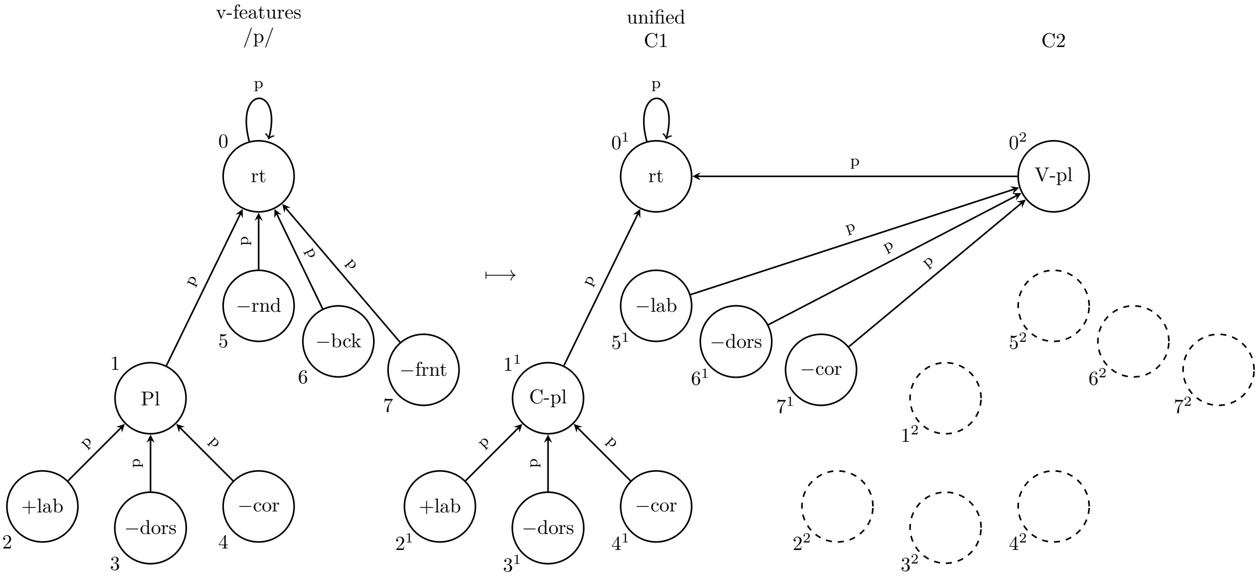


Figure 4: Outout of transduction from v-features to unified

Two theories: unified and v-features

- **unified** and **v-features** are **contrast preserving**
  - each with 6 fully-specified binary features
- **unified** and **v-features** are **strongly QF-bi-interpretable**
  - full transduction rules in appendix
- yet **unified** and **v-features** are *not* natural class preserving
  - **unified** contains a natural class for each value of each place feature whose natural class extensions are not part of the natural class extensions of **v-features** (this is intentional and expected, see Danis 2025 for further discussion)
- do we want to call them notationally equivalent?

Feature Class Preservation

The enterprise of feature geometry involves a cross-linguistic investigation of which features seem to behave/change together in some process, and working this into the representations themselves, e.g. if all all place features tend to assimilate together, then there must be some constituent in the segmental structure that includes all place features. (Clements 1985, Sagey 1986, Mester 1986, McCarthy 1988, among many others)

- If two theories group the same (or related) features into the same constituency structure, they are **feature class preserving**.
- **Problem**: constituents of features cannot be compared directly if their labels (names) do not match (e.g. [labial] vs [round]).
- **Solution**: Utilize the logical transduction to compare constituency in terms of relevant node indices such that constituents in both theories can be compared directly in a set theoretic way.

Feature Class Preservation

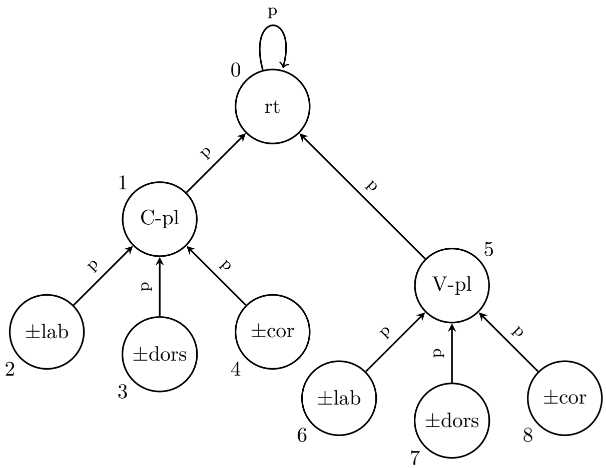


Figure 5: **unified**

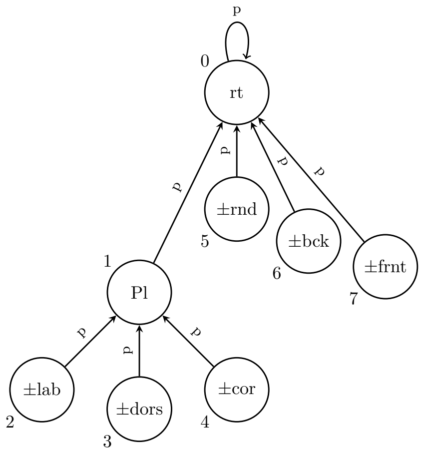


Figure 6: **v-features**

the sets of terminal nodes cannot be compared directly as their labels differ

Feature Class Preservation

The property of feature class preservation is defined as an MSO sentence over one theory while defining relations for the other theory based on the rules of the transduction.

1. define a relation D that captures the domination relation for the model in theory A in the logic of theory A

$$xDy = \varphi(x, y)$$

2. use the transduction to define D' in the logic of B *but still evaluated over model A*

$$xD'y = \varphi'(x, y)$$

3. define two MSO sentences defining constituency, one using R and one using R', and compare the resulting sets

Feature Class Preservation

Defining Constituency

- define R and R' as the transitive closure of D and D' (which is in general MSO definable) to get the general dominance relation
- assume the following helper predicate:

$$\text{terminal}(x) = \neg\exists y[xDy]$$

- the definition of constituent in theory A:

$$\text{Constit}_{\text{A}}(X) = \exists y \forall x[X(x) \leftrightarrow yRx \wedge \text{terminal}(x)]$$

a constituent is the set of positions X for some node y such that position x is contained in X iff y generally dominates x and x is a terminal node

Constituents in unified

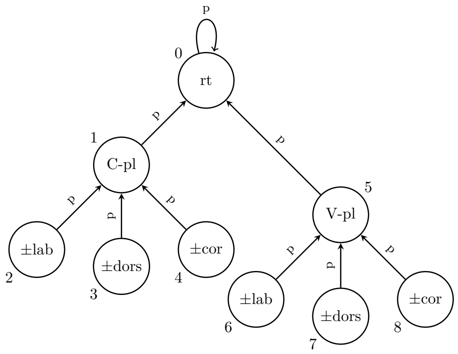
$$xDy := \text{parent}(y) = x \land x \neq y$$

$$\text{Constit}_{\text{unified}}(X) := \exists y \forall x [X(x) \leftrightarrow yRx \land \text{terminal}(x)]$$

for **unified**,

$$\begin{aligned} R = \{&(0,1),(0,2),(0,3),(0,4),\\ &(0,5),(0,6),(0,7),(0,8)\\ &(5,6),(5,7),(5,8),\\ &(1,2),(1,3),(1,4)\} \end{aligned}$$

so the following sets satisfy $\text{Constit}_{\text{unified}}(X)$:
$\{2,3,4,6,7,8\}, \{2,3,4\}, \{6,7,8\}$



unified

Constituents in v-features

Define D'

Current definition for D: $xDy := \text{parent}(y) = x \wedge x \neq y$

Relevant transduction rules:

$$\text{parent}(x^1) := \begin{cases} (\text{parent}(x))^1 \Leftrightarrow \neg\ \text{vowelFeature}(\text{parent}(x)) \\ (\text{parent}(\text{parent}(x)))^1 \Leftrightarrow \text{vowelFeature}(\text{parent}(x)) \end{cases}$$

$$\text{vowelFeature} = +\text{round}(x) \vee +\text{front}(x) \vee +\text{back}(x) \vee -\text{round}(x) \vee -\text{front}(x) \vee -\text{back}(x)$$

The function is defined with cases; these become conjuncts:

$$\begin{aligned} xD'y := {} & (\text{parent}(y))^1 = x \leftrightarrow \neg\ \text{vowelFeature}(\text{parent}(y)) \wedge \\ & (\text{parent}(\text{parent}(y)) = x \leftrightarrow \text{vowelFeature}(\text{parent}(y)) \wedge \\ & x \neq y \end{aligned}$$

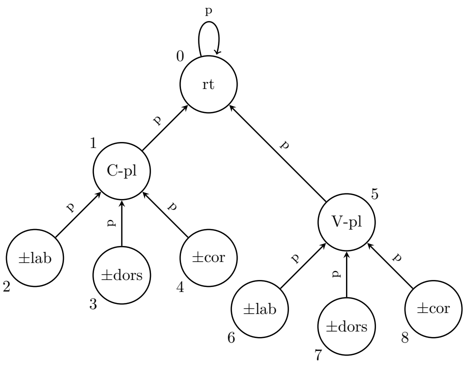
Feature Class Preservation

$$xD'y := (\text{parent}(y))^1 = x \leftrightarrow \neg\ \text{vowelFeature}(\text{parent}(y)) \wedge$$
$$(\text{parent}(\text{parent}(y)) = x \leftrightarrow \text{vowelFeature}(\text{parent}(y)) \wedge$$
$$x \neq y$$

for **v-features** but defined over **unified**,

$$R' = \{(0,1),(0,2),(0,3),(0,4),$$
$$(0,6),(0,7),(0,8),$$
$$(1,2),(1,3),(1,4)\}$$

so the following sets satisfy $\text{Constit}_{\text{v-feat}}(X)$:
$\{2,3,4,6,7,8\}, \{2,3,4\}$



unified

Feature Class Preservation

$$\text{constituents of unified} = \{\{2,3,4,6,7,8\},\{2,3,4\},\{6,7,8\}\}$$

$$\text{constituents of v-features} = \{\{2,3,4,6,7,8\},\{2,3,4\}\}$$

- the nodes $\{6,7,8\}$ form a constituent in **unified** due to the V-place node, but no such constituent exists in **v-features**
- the constituent dominated by the root node, $\{2,3,4,6,7,8\}$, is present in both sets even though the labels of these nodes differ across theories
- by using the transduction rules for translating the definition of the parent function for **v-features** into the logic of **unified** means we can determine the constituents using the same node indices

On notational equivalence

Notational equivalence

At what point can we call two theories *notationally equivalent*?

Notational equivalence

A QF transduction is extremely restricted in the degree to which the output can differ from the input because QF is a weak logical language limited to local operations. QF-bi-interpretability can therefore be considered an indication of notational equivalence. (Strother-Garcia 2019: 39)

…we can conclude that separated and bundled representations are bi-interpretable in a strict model theoretic sense. Within the framework adopted here, the models do not differ in any non-trivial way in terms of their structure. Condition (1b)[3] for notational equivalence is thus satisfied. (Oakden 2020: 286)

[3] "Two models represent the same set of abstract properties, differing only superficially."

Notational equivalence

- In their intended grammatical frameworks, the changes in natural class definitions have real and tangible effects for both the naturalness (or even possibility) of capturing a process and for the predicted typology of languages.
- No card-carrying phonologist before 2010 would call strings and feature bundles notationally equivalent—because the expectation then is that the representations would be used in the grammatical frameworks they were intended for.
- The appeal of Miller 2001's approach is that we can now safely and comfortable say that two representations or formalisms are *equivalent in some domain* and potentially not equivalent in others
- Specific equivalence under some transduction should be referred to as such, especially after determining exactly what similarities in complexity or expressivity must follow from a transduction. (Oakden 2020 has excellent discussion on this.)

Natural class and learnability

- Gildea & Jurafsky 1996 test the OSTIA (*onward subsequential transducer inference algorithm*, Oncina, Garcia & Vidal 1993) against synthetic but **naturalistic** input-output pairs in an effort to learn the English flapping rule t → dx / V́ r* _ V
- Even though the OSTIA algorithm can provably to learn any subsequential function in the limit from positive data, it **fails** to learn the correct pattern from naturalistic data
- The learning task is **successful** after the implementation of three learning biases Gildea & Jurafsky 1996 implement, one of which is the idea that “[p]honologically similar segments behave similarly.” (Gildea & Jurafsky 1996: 508)

Natural class information as a learning aid

Initial attempt with **unmodified** OSTIA algorithm:

- relatively high error rate (did not learn the pattern exactly)
- obscene number of states (just look at it)

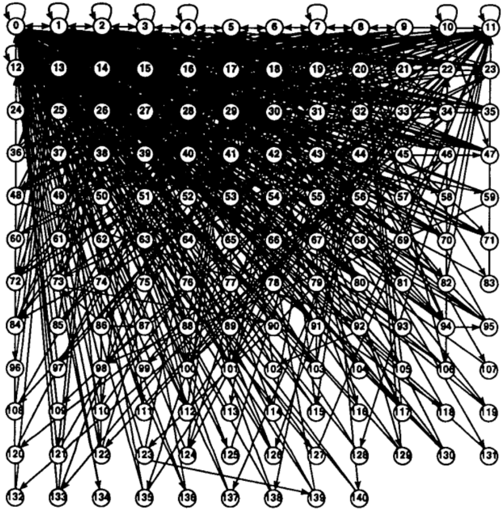


Figure 7
First attempt of OSTIA to learn flapping. Transducer induced on 25,000 samples.

from Gildea & Jurafsky 1996: 507

Natural class information as a learning aid

Adding **Faithfulness**[4] bias:

- much lower error rate (0.06% down from 4.46%)
- down to 3 states
- cannot generalize: if a particular segment was not in the right position in the training data, it is excluded from the rule context

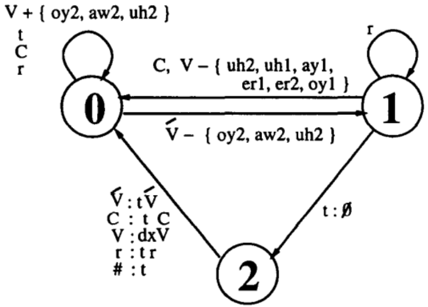


Figure 12
Flapping transducer induced with alignment, trained on 25,000 samples.

from Gildea & Jurafsky 1996: 507

[4] state transitions should have identical inputs and outputs as much as possible

Natural class information as a learning aid

Adding **Community** (natural class information) bias:

- essentially correct
- generalizes process to all appropriate vowels

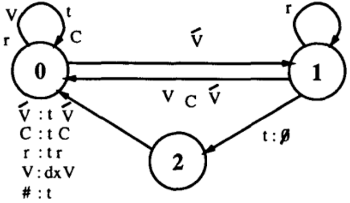


Figure 14
Flapping transducer induced from 50,000 samples.

from Gildea & Jurafsky 1996: 507

"The intuition that OSTIA is missing, then, is the idea that phonological constraints are sensitive to phonological features that pick out certain equivalence classes of segments. Since the beginning of generative grammar, and based on Jakobson's early insistence on the importance of binary oppositions (Jakobson 1968; Jakobson, Fant, and Halle 1952), phonological features, and not the segment, have generally formed the vocabulary over which linguistic rules are formed. **Giving such knowledge to OSTIA would allow it to hypothesize that if every vowel it has seen has acted a certain way, that the rest of them might act similarly.**" (Gildea & Jurafsky 1996: 514, emph. mine)

Natural class information as a learning aid

- For algorithms to learn from more naturalistic data than what is currently required, they (at least) must be able to generalize over classes of segments
- The case of Gildea & Jurafsky 1996 is just one example of this; other examples across various learning paradigms include:
  - the algorithm in Gouskova & Gallagher 2020, which crucially assumes that segments participating in long-distance interactions form a natural class
  - the Output Driven Learner of Tesar 2013, which crucially uses feature information to reduce the search space of appropriate underlying forms (among for many other things)
  - for the MaxEnt learner (Hayes & Wilson 2008), "it is the natural classes determined by the features, rather than the features themselves, that determine the content of a constraint." (p. 391)

Summary

Summary

- Logical transductions between model-theoretic implementations of phonological representations offers a rigorous and precise way to track differences in expressivity between theories
- However, even translations defined with weak logic, like quantifier-free first-order logic, allow for relevant linguistic differences between the representations, such as predicted natural classes or feature class information
- The research program here puts forth a way to marry the purely logical approach of comparing representations with one that tracks relevant linguistic properties, without abandoning the precision and rigor of finite model theory
- While the exact consqeunces of these linguistic properties, such as natural class preservation, is understudied in terms of a representation's expressive power, it most definitely matters for learning over such structures, especially with naturalistic data

thank you!

and thanks to Adam Jardine, Jeff Heinz, Scott Nelson, and many others who have engaged with this work in the past, either as an audience or directly.

References

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, Massachusetts: The MIT Press.

Clements, G. N. 1985. The geometry of phonological features. *Phonology* 2(1). 225–252. https://doi.org/10.1017/S0952675700000440.

Courcelle, Bruno. 1994. Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science* 126(1). 53–75. https://doi.org/10.1016/0304-3975(94)90268-2.

Danis, Nick. 2025. Logical transductions are not sufficient for notational equivalence. *Proceedings of the Annual Meetings on Phonology*. https://doi.org/10.7275/amphonology.3022.

Gildea, Daniel & Daniel Jurafsky. 1996. Learning Bias and Phonological-Rule Induction. (Ed.) Julia Hirschberg. *Computational Linguistics* 22(4). 497–530. https://aclanthology.org/J96-4003/. (12 January, 2026).

References 2

Gouskova, Maria & Gillian Gallagher. 2020. Inducing nonlocal constraints from baseline phonotactics. *Natural Language & Linguistic Theory* 38(1). 77–116. https://doi.org/10.1007/s11049-019-09446-x.

Hayes, Bruce & Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39(3). 379–440. https://doi.org/10.1162/ling.2008.39.3.379.

Jardine, Adam. 2016. *Locality and non-linear representations in tonal phonology*. phdthesis.

Jardine, Adam. 2017. On the Logical Complexity of Autosegmental Representations. In Makoto Kanazawa, Philippe de Groote & Mehrnoosh Sadrzadeh (eds.), *Proceedings of the 15th Meeting on the Mathematics of Language*, 22–35. London, UK: Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-3403.

Jardine, Adam, Nick Danis & Luca Iacoponi. 2021. A Formal Investigation of Q-Theory in Comparison to Autosegmental Representations. *Linguistic Inquiry* 52(2). 333–358. https://doi.org/10.1162/ling_a_00376.

References 3

Libkin, Leonid. 2004. *Elements of Finite Model Theory*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-07003-1.

McCarthy, John. 1988. Feature geometry and dependency: A review. *Phonetica* 84–108. https://doi.org/10.1159/000261820.

Mester, Armin. 1986. *Studies in Tier Structure*. phdthesis.

Miller, Philip H. 2001. *Strong Generative Capacity: The Semantics of Linguistic Formalism* (Lecture Notes). Center for the Study of Language, Information. https://press.uchicago.edu/ucp/books/book/distributed/S/bo3612630.html. (4 May, 2024).

Nelson, Scott. 2022. A Model Theoretic Perspective on Phonological Feature Systems. In Allyson Ettinger, Tim Hunter & Brandon Prickett (eds.), *Proceedings of the Society for Computation in Linguistics 2022*, 1–10. online: Association for Computational Linguistics. https://aclanthology.org/2022.scil-1.1/. (18 December, 2025).

References 4

Nelson, Scott. 2024. *The Computational Structure of Phonological and Phonetic Knowledge.* phdthesis.

Oakden, Chris. 2020. Notational equivalence in tonal geometry. *Phonology* 37(2). 257–296. https://doi.org/10.1017/S0952675720000123.

Oncina, J., P. Garcia & E. Vidal. 1993. Learning subsequential transducers for pattern recognition interpretation tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(5). 448–458. https://doi.org/10.1109/34.211465.

Sagey, Elizabeth. 1986. *The Representation of Features and Relations in Non-Linear Phonology.* phdthesis.

Strother-Garcia, Kristina. 2019. *Using model theory in phonology: a novel characterization of syllable structure and syllabification.* Thesis. https://udspace.udel.edu/handle/19716/25084. (24 October, 2022).

References 5

Tesar, Bruce. 2013. *Output-Driven Phonology*. Cambridge University Press.

Volenec, Veno & Charles Reiss. 2020. Formal Generative Phonology. *Radical: A Journal oh Phonology* 2. 1–148. https://radical.cnrs.fr/volenec-reiss-formal-generative-phonology/. (26 September, 2022).

Weijer, Jeroen van de. 1996. *Segmental Structures and Complex Segments*. Tübingen: Max Niemeyer Verlag.

Appendix

$$\mathtt{rt}(x^1) := \mathtt{rt}(x) \tag{1}$$
$$\mathtt{Place}(x^1) := \mathtt{C\text{-}place}(x) \tag{2}$$
$$+\mathtt{lab}(x^1) := +\mathtt{lab}(x) \wedge \mathtt{C\text{-}place}(\mathit{parent}(x)) \tag{3}$$
$$+\mathtt{cor}(x^1) := +\mathtt{cor}(x) \wedge \mathtt{C\text{-}place}(\mathit{parent}(x)) \tag{4}$$
$$+\mathtt{dors}(x^1) := +\mathtt{dors}(x) \wedge \mathtt{C\text{-}place}(\mathit{parent}(x)) \tag{5}$$
$$-\mathtt{lab}(x^1) := -\mathtt{lab}(x) \wedge \mathtt{C\text{-}place}(\mathit{parent}(x)) \tag{6}$$
$$-\mathtt{cor}(x^1) := -\mathtt{cor}(x) \wedge \mathtt{C\text{-}place}(\mathit{parent}(x)) \tag{7}$$
$$-\mathtt{dors}(x^1) := -\mathtt{dors}(x) \wedge \mathtt{C\text{-}place}(\mathit{parent}(x)) \tag{8}$$
$$+\mathtt{round}(x^1) := +\mathtt{lab}(x) \wedge \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{9}$$
$$+\mathtt{front}(x^1) := +\mathtt{cor}(x) \wedge \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{10}$$
$$+\mathtt{back}(x^1) := +\mathtt{dors}(x) \wedge \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{11}$$
$$-\mathtt{round}(x^1) := -\mathtt{lab}(x) \wedge \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{12}$$
$$-\mathtt{front}(x^1) := -\mathtt{cor}(x) \wedge \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{13}$$
$$-\mathtt{back}(x^1) := -\mathtt{dors}(x) \wedge \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{14}$$
$$\mathit{parent}(x^1) := (\mathit{parent}(x))^1 \Leftrightarrow \neg\mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{15}$$
$$\mathit{parent}(x^1) := (\mathit{parent}(\mathit{parent}(x)))^1 \Leftrightarrow \mathtt{V\text{-}place}(\mathit{parent}(x)) \tag{16}$$

Figure 7: Transduction rules from unified to v-features

$$\texttt{rt}(x^1) := \texttt{rt}(x) \tag{17}$$

$$+\texttt{lab}(x^1) := +\texttt{round}(x) \vee +\texttt{lab}(x) \tag{18}$$

$$+\texttt{cor}(x^1) := +\texttt{front}(x) \vee +\texttt{cor}(x) \tag{19}$$

$$+\texttt{dors}(x^1) := +\texttt{back}(x) \vee +\texttt{dors}(x) \tag{20}$$

$$-\texttt{lab}(x^1) := -\texttt{round}(x) \vee -\texttt{lab}(x) \tag{21}$$

$$-\texttt{cor}(x^1) := -\texttt{front}(x) \vee -\texttt{cor}(x) \tag{22}$$

$$-\texttt{dors}(x^1) := -\texttt{back}(x) \vee -\texttt{dors}(x) \tag{23}$$

$$\texttt{C-place}(x^1) := \texttt{Place}(x) \tag{24}$$

$$\texttt{V-place}(x^2) := \texttt{rt}(x) \tag{25}$$

$$\textit{parent}(x^1) := (\textit{parent}(x))^1 \Leftrightarrow \neg\text{vowelFeature}(x) \tag{26}$$

$$\textit{parent}(x^1) := (\textit{parent}(x))^2 \Leftrightarrow \text{vowelFeature}(x) \tag{27}$$

$$\textit{parent}(x^2) := x^1 \Leftrightarrow \texttt{rt}(x) \tag{28}$$

Figure 8: Transduction rules from unified to v-features