Output-drivenness and partial phonological features

Giorgio Magri

Abstract. Tesar (2013) develops the notion of *output-drivenness*, provides guarantees that OT grammars satisfy it, and demonstrates its learnability implications. This squib discusses the extension of Tesar's theory to a representational framework with partial phonological features. It considers a hierarchy of notions of output-drivenness of increasing strength which can be defined within this extended framework. It determines the strongest notion of output-drivenness which holds in the case of partial features. And it shows that the learnability implications discussed by Tesar carry over to a framework with partial features only if feature undefinedness is effectively treated by identity faithfulness constraints as an additional feature value.

Keywords: Optimality Theory; partial features; underspecification; learnability; inconsistency detection; output-drivenness.

1 Introduction

Tesar's (2013) notion of *output-drivenness* formalizes the intuition that any discrepancy between an underlying and a surface (or *output*) form is *driven* exclusively by the goal of making the surface form fit the phonotactics. Non-output-drivenness unifies various opaque phonological phenomena such as *chain shifts* (Łubowicz 2011) and *derived environment effects* (or *saltations*; White 2014). Tesar's theory of output-drivenness within *Optimality Theory* (OT; Prince and Smolensky 2004) has two parts. First, Tesar develops general sufficient constraint conditions for the output-drivenness of OT grammars. Second, Tesar distills the learnability implications of output-drivenness for the classical *inconsistency detection* approach (Merchant 2008) to learning underlying forms in OT.

Tesar's theory is developed within a feature-based representational framework which assumes all phonological features to be *total*, namely defined for every segment. Yet, a number of phonetic and phonological arguments have been put forward in favor of *partial* features. To start, if manner and place features describe the degree and the place of an oral constriction, how can they be defined for laryngeal segments [h ɦ ʔ] which involve no oral constriction (McCarthy 1988)? As another example, given that a feature such as [DELAYED RELEASE] is simply irrelevant for sonorants, how could its value be defined for sonorants (Hall 2007)? Besides theoretical perspicuity, empirical arguments have also been provided in favor of partial features. For instance, allowing dorsal features such as [HIGH], [LOW], [FRONT], and [BACK] to be undefined for a labial segment such as [p] captures the fact that the position of the tongue in the production of a medial [p] does not seem to have a specific articulatory target but simply executes the most convenient transition between the positions required for the two segments flanking [p] (Keating 1988). Finally, various approaches to phonological *underspecification* (Steriade 1995) have explored the assumption that certain segments lack a value for certain features at certain levels of phonological representation, either universally (*radical* underspecification: Kiparsky 1982; Archangeli 1984) or language-specifically (*contrastive* or *archiphonemic* underspecification: Mester and Ito 1989; Inkelas 1995; Dresher 2009).

This squib thus tackles the problem of extending Tesar's theory of output-drivenness to a representational framework which allows for partial phonological features.[1] Section 2 extends

Tesar's notion of output-drivenness to a representational framework with partial features. A hierarchy of notions of output-drivenness are considered, ordered by their strength. Section 3 (together with a final appendix) pinpoints the strongest notion of output-drivenness which allows Tesar's guarantees for OT output-drivenness to be extended from total to partial phonological features. Two approaches to identity faithfulness constraints relative to partial features are compared, which differ in whether disparities in feature definedness (the feature is defined for only one of two corresponding segments) are or not penalized just as disparities in feature value (the feature assigns different values to two corresponding segments). Section 4 shows that the learnability implications uncovered by Tesar extend from total to partial phonological features only when phonological identity penalizes disparities in feature definedness as well. Section 5 concludes that choices pertaining to the proper definition of featural identity have non-trivial implications for learnability when Tesar's framework is extended to partial features.

2 Output-drivenness with partial features

This section extends Tesar's notion of output-drivenness to a representational framework with partial features, leading to a hierarchy of notions of output-drivenness.

2.1 Output-drivenness

Consider two different underlying forms $\mathbf{a}$ and $\mathbf{b}$ which share a surface form $\mathbf{x}$ among their candidates,[2] so that both $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$ count as *candidate pairs*. Suppose that the underlying form $\mathbf{b}$ is more similar to the surface form $\mathbf{x}$ than the underlying form $\mathbf{a}$ is. In other words, the candidate pair $(\mathbf{a}, \mathbf{x})$ has less *internal similarity* than the candidate pair $(\mathbf{b}, \mathbf{x})$. Tesar captures this assumption through the condition

(1) $(\mathbf{a}, \mathbf{x}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{x})$

where $\leq_{\text{sim}}$ is a *similarity order* properly defined among candidate pairs (or, more precisely, among candidate pairs which share the surface form). Suppose that a grammar G maps the less similar underlying form $\mathbf{a}$ to the surface form $\mathbf{x}$, namely $G(\mathbf{a}) = \mathbf{x}$. Intuitively, this means that $\mathbf{x}$ is phonotactically licit and that $\mathbf{x}$ is not too dissimilar from $\mathbf{a}$. Since the phonotactic status of $\mathbf{x}$ does not depend on the underlying form and furthermore $\mathbf{x}$ is even more similar to $\mathbf{b}$ than it is to $\mathbf{a}$, the grammar G should map also the more similar underlying form $\mathbf{b}$ to that same surface form $\mathbf{x}$, namely $G(\mathbf{b}) = \mathbf{x}$. Tesar calls any grammar which abides by this logics *output-driven*.

Definition 1 (Output-drivenness) *A grammar G is* output-driven *relative to a similarity order $\leq_{sim}$ provided the following condition holds*

(2) *If $G(\mathbf{a}) = \mathbf{x}$ and $(\mathbf{a}, \mathbf{x}) \leq_{sim} (\mathbf{b}, \mathbf{x})$, then $G(\mathbf{b}) = \mathbf{x}$*

for any underlying/surface form candidate pairs $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$ sharing the surface form $\mathbf{x}$.

2.2 Representational framework

The notion of output-drivenness is predicated on a similarity order $\leq_{\text{sim}}$. How should it be defined? In order to tackle this question conveniently, I make three restrictive assumptions on the representational framework. First, I assume that underlying and surface forms are strings of segments drawn from a finite segment set Σ (e.g., Σ is some set of segments from the IPA table), without any additional structure. Second, I assume that underlying and surface forms in a candidate pair $(\mathbf{a}, \mathbf{x})$ are strings of the same length, namely $\mathbf{a} = a_1 \cdots a_\ell$ and $\mathbf{x} = x_1 \cdots x_\ell$. An underlying segment a_i and a surface segment x_j *correspond* to each other (in the sense of McCarthy and Prince 1993) provided they occupy the same position in the two corresponding strings $\mathbf{a}$ and $\mathbf{x}$, namely $i = j$. Thus, each underlying (surface) segment has one and only one corresponding surface (underlying) segment, so that epenthesis, deletion, coalescence, and breaking are not allowed (see Tesar 2013, chapter 2 for the extension to a framework which allows for deletion and epenthesis, although not for breaking and coalescence). Finally, I assume that any pair of strings of the same length counts as a candidate pair, so that idiosyncratic restrictions on candidacy are not allowed (see Tesar 2013, section 3.3.1 for the extension to a framework which relaxes this candidacy *completeness* assumption).

2.3 Similarity order in the case of total features

Assume that the segments in the segment set Σ are distinguished through certain phonological features collected together in the feature set Φ. A generic phonological feature φ is a function which takes a segment x and returns a certain feature value $\varphi(x)$. Features can take two values (in the case of *binary* features) or more than two values (in the case of *multi-valued* features). For instance, [VOICE] is a binary feature which only takes values $+$ and $-$ while [PLACE] can be construed as a multi-valued feature which takes the three values L for labial, C for coronal, and D for dorsal (see for instance de Lacy 2006, section 2.3.2.1.1). Tesar assumes that the features in the feature set Φ are all *total* (relative to Σ), namely defined for each segment in the segment set Σ. Under this assumption, he provides a definition of the similarity order which can be adapted as follows to the representational framework adopted in this paper. The superscript "total" in the notation "$\leq_{\text{sim}}^{\text{total}}$" makes explicit the assumption that every feature φ is total.[3]

Definition 2 (Similarity order with total features) *Consider two candidate pairs* $(\mathbf{a}, \mathbf{x})$ *and* $(\mathbf{b}, \mathbf{x})$ *which share the surface string* $\mathbf{x} = x_1 \cdots x_\ell$ *so that the two underlying strings* $\mathbf{a}$ *and* $\mathbf{b}$ *have the same length* ℓ*, namely have the shape* $\mathbf{a} = a_1 \cdots a_\ell$ *and* $\mathbf{b} = b_1 \cdots b_\ell$*. The relation* $(\mathbf{a}, \mathbf{x}) \leq_{sim}^{total} (\mathbf{b}, \mathbf{x})$ *holds provided the following disjunction holds*

$$
\text{(3)} \quad \begin{aligned} \textit{Either:} &\quad \varphi(b_i) = \varphi(x_i) \\ \textit{Or:} &\quad \varphi(b_i) = \varphi(a_i) \end{aligned}
$$

for every $i = 1, \ldots, \ell$ *and for every feature* φ *in the feature set* Φ*.*

The first disjunct of (3) says that the segment b_i of the more similar underlying string $\mathbf{b}$ patterns alike the corresponding segment x_i of the surface string $\mathbf{x}$ relative to the feature φ. Suppose that this disjunct fails because the feature φ assigns different values to the two segments b_i and x_i. The second disjunct must then hold. It says that the segment a_i of the less similar underlying string $\mathbf{a}$ patterns alike the segment b_i of the more similar underlying string $\mathbf{b}$ relative to the feature φ. This entails in particular that also a_i disagrees with x_i relative to the feature φ. The condition $(\mathbf{a}, \mathbf{x}) \leq_{\text{sim}}^{\text{total}} (\mathbf{b}, \mathbf{x})$ thus indeed formally captures the intuition that the underlying string $\mathbf{a}$ is at most as similar to the surface string $\mathbf{x}$ as the underling string $\mathbf{b}$ is.

2.4 Graphical representation of the similarity order

The disjunction (3) is the heart of Tesar's definition 2 of the similarity order. Towards the goal of extending this definition to the case of partial features, it is useful to pause to introduce a graphical representation of this disjunction. To start, suppose that the feature φ which figures in this disjunction is total and binary, namely assigns one of the two values $+$ or $-$ to any segment. In this case, Tesar's disjunction can be stated as the condition that the two pairs of feature values $(\varphi(a_i), \varphi(x_i))$ and $(\varphi(b_i), \varphi(x_i))$ be connected by an arrow in (4). Since φ is binary, there are four pairs of feature values to consider. These four pairs are sorted into two groups separated by a vertical line, because we only compare pairs of feature values which share the second value. The straight arrows represent the first disjunct of (3). The loop arrows represent the second disjunct and thus ensure that the corresponding similarity order is reflexive.

(4)
$$\begin{array}{c|c} \circlearrowright & \circlearrowright \\ (+,+) & (-,-) \\ \uparrow & \uparrow \\ (-,+) & (+,-) \\ \circlearrowleft & \circlearrowleft \end{array}$$

Next, consider the case where the feature φ is again total but multi-valued. For concreteness, assume that φ is the feature [PLACE] and that it takes the three values L, C, and D. In this case, Tesar's disjunction (3) can be stated as the condition that the two pairs of feature values $(\varphi(a_i), \varphi(x_i))$ and $(\varphi(b_i), \varphi(x_i))$ be connected by an arrow in (5). Since the feature considered is three-valued, there are nine pairs of feature values to consider. These nine pairs are sorted into three groups separated by a vertical line, because we only compare pairs of feature values which share the second value. Again as above, the straight arrows represent the first disjunct of (3) and the loop arrows represent the second disjunct.

(5)
$$\begin{array}{ccc|ccc|ccc} & \circlearrowright & & & \circlearrowright & & & \circlearrowright & \\ & (\textsc{l}, \textsc{l}) & & & (\textsc{c}, \textsc{c}) & & & (\textsc{d}, \textsc{d}) & \\ \nearrow & & \nwarrow & \nearrow & & \nwarrow & \nearrow & & \nwarrow \\ (\textsc{c}, \textsc{l}) & & (\textsc{d}, \textsc{l}) & (\textsc{l}, \textsc{c}) & & (\textsc{d}, \textsc{c}) & (\textsc{l}, \textsc{d}) & & (\textsc{c}, \textsc{d}) \\ \circlearrowleft & & \circlearrowleft & \circlearrowleft & & \circlearrowleft & \circlearrowleft & & \circlearrowleft \end{array}$$

In conclusion, Tesar's definition 2 of the similarity order consists of two steps. One step defines a partial order among pairs of feature values, such as those represented in (4) or (5). The other step then "lifts" that partial order from pairs of feature values to pairs of strings by requiring the ordering to hold for the values assigned by each feature to each segment.

2.5 Similarity orders in the case of partial features

Tesar's definition 2 of the similarity order $\leq^{\text{total}}_{\text{sim}}$ assumes that each feature in the feature set Φ is total, namely defined for each segment in the segment set Σ. This assumption guarantees that it makes sense to consider and compare the values $\varphi(a_i)$, $\varphi(b_i)$, and $\varphi(x_i)$ assigned by any feature φ to any of the segments a_i, b_i, x_i involved. Consider now a slightly more general representational framework, which allows for a feature φ to be *partial*, namely undefined for some segments in the segment set Σ. This assumption requires some tampering with the disjunction

(3) at the heart of Tesar's original definition 2, as a certain feature φ could be undefined for one or more of the three segments a_i, b_i, x_i. For concreteness, assume that φ is binary. For the sake of visualization, let me represent the fact that φ is undefined for a certain segment through the condition that φ assigns it the "dummy" value 0 (see Hayes 2009, section 4.8). Three different possibilities are then plotted in (6).

(6) a.

$$\begin{tikzcd}
& (+,+) \arrow[loop above] & & & (-,-) \arrow[loop above] & & & (0,0) \arrow[loop above] & \\
(-,+) \arrow[loop below] \arrow[ur] & & (0,+) \arrow[loop below] & (+,-) \arrow[loop below] \arrow[ur] & & (0,-) \arrow[loop below] & (+,0) \arrow[loop below] & & (-,0) \arrow[loop below]
\end{tikzcd}$$

b.

$$\begin{tikzcd}
& (+,+) \arrow[loop above] & & & (-,-) \arrow[loop above] & & & (0,0) \arrow[loop above] & \\
(-,+) \arrow[loop below] \arrow[ur] & & (0,+) \arrow[loop below] \arrow[ul] & (+,-) \arrow[loop below] \arrow[ur] & & (0,-) \arrow[loop below] \arrow[ul] & (+,0) \arrow[loop below] & & (-,0) \arrow[loop below]
\end{tikzcd}$$

c.

$$\begin{tikzcd}
& (+,+) \arrow[loop above] & & & (-,-) \arrow[loop above] & & & (0,0) \arrow[loop above] & \\
(-,+) \arrow[loop below] \arrow[ur] & & (0,+) \arrow[loop below] \arrow[ul] & (+,-) \arrow[loop below] \arrow[ur] & & (0,-) \arrow[loop below] \arrow[ul] & (+,0) \arrow[loop below] \arrow[ur] & & (-,0) \arrow[loop below] \arrow[ul]
\end{tikzcd}$$

Again, the loop arrows on each pair are needed in order for the resulting similarity order to be reflexive. The option (6a) ignores pairs which contain the dummy value "0" (apart for the loop arrows). In other words, it assumes that the dummy value cannot contribute to any feature disparity. This option represents the minimal extension of Tesar's diagram (4). The option (6c) represents instead the maximal extension. It effectively assumes that the dummy value "0" can lead to disparities just as a third plain feature value, as shown by comparison with Tesar's diagram (5) for the three-valued feature [PLACE]. In other words, it treats a binary partial feature analogously to a total three-valued feature. The intermediate option (6b) instead encodes the fact that "0" is not just a plain third feature value, by avoiding any straight arrows in the rightmost block.

The three disjunctions (7a), (7b), and (7c) in the definition below correspond to the three diagrams (6a), (6b), and (6c) when the feature φ is binary. In the sense that one of these three disjunctions holds provided the pairs of feature values $(\varphi(a_i), \varphi(x_i))$ and $(\varphi(b_i), \varphi(x_i))$

are connected by an arrow in the corresponding diagram. Again, the straight arrows represent the first disjunct and the loop arrows represent the second disjunct.[4] The three similarity orders thus obtained for a framework with partial features are distinguished by the superscripts "sparse", "medium", and "dense", which intuitively reflect the relative number of pairs of values connected by straight arrows in the three diagrams (6). If the feature φ is total, the three disjunctions (7) are equivalent to Tesar's original disjunction (3). If all the features in the feature set Φ are total, the two definitions 2 and 3 thus define the same similarity order.

Definition 3 (Similarity order with partial features) *Consider two candidate pairs* $(\mathbf{a}, \mathbf{x})$ *and* $(\mathbf{b}, \mathbf{x})$ *which share the surface string* $\mathbf{x} = x_1 \cdots x_\ell$ *so that the two underlying strings* $\mathbf{a}$ *and* $\mathbf{b}$ *have the same length* ℓ*, namely have the shape* $\mathbf{a} = a_1 \cdots a_\ell$ *and* $\mathbf{b} = b_1 \cdots b_\ell$*. The relation* $(\mathbf{a}, \mathbf{x}) \leq_{sim}^{sparse} (\mathbf{b}, \mathbf{x})$*,* $(\mathbf{a}, \mathbf{x}) \leq_{sim}^{med} (\mathbf{b}, \mathbf{x})$*, or* $(\mathbf{a}, \mathbf{x}) \leq_{sim}^{dense} (\mathbf{b}, \mathbf{x})$ *holds if and only if the corresponding disjunction (7a), (7b), or (7c) holds*

(7) a. Either: feature φ *is defined for* a_i*,* b_i *and* x_i *and furthermore* $\varphi(b_i) = \varphi(x_i)$*.*

Or: feature φ *is undefined for both* b_i *and* a_i *or else it is defined for both and furthermore* $\varphi(b_i) = \varphi(a_i)$*.*

b. Either: feature φ *is defined for* b_i *and* x_i *and furthermore* $\varphi(b_i) = \varphi(x_i)$*.*

Or: feature φ *is undefined for both* b_i *and* a_i *or else it is defined for both and furthermore* $\varphi(b_i) = \varphi(a_i)$*.*

c. Either: feature φ *is undefined for both* b_i *and* x_i *or else it is defined for both and furthermore* $\varphi(b_i) = \varphi(x_i)$*.*

Or: feature φ *is undefined for both* b_i *and* a_i *or else it is defined for both and furthermore* $\varphi(b_i) = \varphi(a_i)$*.*

for every $i = 1, \ldots, \ell$ *and for every feature* φ *in the feature set* Φ*.*

2.6 A hierarchy of notions of output-drivenness

The three diagrams (6a), (6b), and (6c) contain an increasing number of arrows. As a result, the three corresponding similarity orders $\leq_{\text{sim}}^{\text{sparse}}$, $\leq_{\text{sim}}^{\text{mid}}$, and $\leq_{\text{sim}}^{\text{dense}}$ hold among an increasing number of candidate pairs. In other words, the implications in (8) hold for any two candidate pairs $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$ which share the surface form $\mathbf{x}$.

(8) $(\mathbf{a},\mathbf{x}) \leq^{\text{sparse}}_{\text{sim}} (\mathbf{b},\mathbf{x}) \quad\Longrightarrow\quad (\mathbf{a},\mathbf{x}) \leq^{\text{med}}_{\text{sim}} (\mathbf{b},\mathbf{x}) \quad\Longrightarrow\quad (\mathbf{a},\mathbf{x}) \leq^{\text{dense}}_{\text{sim}} (\mathbf{b},\mathbf{x})$

According to definition 1, a grammar G is output-driven relative to a similarity order $\leq_{\text{sim}}$ provided it satisfies the following implication for any two candidate pairs $(\mathbf{a},\mathbf{x})$ and $(\mathbf{b},\mathbf{x})$: if G maps $\mathbf{a}$ to $\mathbf{x}$ and $(\mathbf{a},\mathbf{x}) \leq_{\text{sim}} (\mathbf{b},\mathbf{x})$, then G also maps $\mathbf{b}$ to $\mathbf{x}$. Suppose that the similarity order $\leq_{\text{sim}}$ is so sparse that it only holds of identical pairs, namely $(\mathbf{a},\mathbf{x}) \leq_{\text{sim}} (\mathbf{b},\mathbf{x})$ if and only if $\mathbf{a}$ and $\mathbf{b}$ coincide. Obviously, any grammar is output-driven relative to this similarity order. In other words, this sparsest similarity order yields the weakest (namely trivial) notion of output-drivenness. In general, the sparser the similarity order $\leq_{\text{sim}}$ (namely the fewer candidate pairs are in a similarity relation), the weaker the corresponding notion of output-drivenness. The implications (8) among the three similarity orders $\leq^{\text{sparse}}_{\text{sim}}$, $\leq^{\text{mid}}_{\text{sim}}$, and $\leq^{\text{dense}}_{\text{sim}}$ thus yield the reverse implications (9) among the three corresponding notions of output-drivenness.

(9) G is output-driven relative to $\leq^{\text{dense}}_{\text{sim}}$ $\Longrightarrow$
G is output-driven relative to $\leq^{\text{med}}_{\text{sim}}$ $\Longrightarrow$
G is output-driven relative to $\leq^{\text{sparse}}_{\text{sim}}$

In conclusion, (9) defines a hierarchy of notions of output-drivenness for a representational framework which allows for partial features. The rest of the paper investigates OT guarantees for and learnability implications of these various notions of output-drivenness.

3 Establishing output-drivenness for OT grammars

This section pinpoints the strongest notion of output-drivenness in the hierarchy (9) which allows Tesar's guarantees for OT output-drivenness to be extended from total to partial phonological features. The discussion has subtle implications for the definition of featural identity.

3.1 Featural identity in the case of total features

Conditions on the output-drivenness of OT grammars take the form of conditions on the constraint set used to define them. Since I am assuming that underlying (surface) segments have one and only one surface (underlying) correspondent, the only relevant faithfulness constraints are featural identity constraints (McCarthy and Prince 1993). To start, assume that the features in the feature set Φ are all total relative to the segment set Σ. The definition of the *identity* faithfulness constraint IDENT_φ corresponding to a total feature φ is recalled in (10).

(10) a. For any underlying and surface segments a and x:

$$\textsc{Ident}_{\varphi}(a,x)=\begin{cases}1 & \text{if } \varphi(a)\neq\varphi(x)\\ 0 & \text{otherwise}\end{cases}$$

b. For any underlying and surface strings $\mathbf{a}=a_1\cdots a_\ell$ and $\mathbf{x}=x_1\cdots x_\ell$:

$$\textsc{Ident}_{\varphi}(\mathbf{a},\mathbf{x})=\sum_{i=1}^{\ell}\textsc{Ident}_{\varphi}(a_i,x_i)$$

Clause (10a) defines the faithfulness constraint $\textsc{Ident}_{\varphi}$ for a pair (a,x) of underlying/surface segments: it is only violated if the feature φ assigns different values $\varphi(a)$ and $\varphi(x)$ to the two segments a and x. Clause (10b) extends the definition from segments to strings, by summing over pairs of corresponding segments. Clause (10b) only considers pairs of underlying and surface strings of the same length, as required for them to form a candidate pair.

3.2 Establishing OT output-drivenness in the case of total features

For which constraint sets does it happen that the OT grammars corresponding to all constraint rankings all qualify as output-driven? Under the current assumption that the features in the feature set Φ are all total, we focus on output-drivenness relative to the similarity order $\leq_{\text{sim}}^{\text{total}}$ provided by Tesar's definition 2 (the other similarity orders provided by definition 3 collapse with $\leq_{\text{sim}}^{\text{total}}$ when the features are all total). Tesar's answer to this question is recalled in theorem 1, adapted to the restrictive representational framework considered here (see Tesar 2013, chapters 2 and 3 for the extension to a representational framework which allows for epenthesis and deletion, although not for breaking and coalescence).

Theorem 1 (Output-drivenness with total features) *Assume that the faithfulness constraint set consists of the identity faithfulness constraints (10) relative to a set Φ of features which are all total (either binary or multi-valued). The OT grammar corresponding to any ranking of the constraint set is output-driven relative to the similarity order $\leq_{sim}^{total}$ provided by definition 2.*

Tesar's theorem 1 makes no assumptions on the markedness constraints. Only the faithfulness constraints matter for output-drivenness. If in particular the faithfulness constraints are all of the identity-type (10), output-drivenness holds.

3.3 Featural identity in the case of partial features: two options

The definition (10) of the identity faithfulness constraint $\textsc{Ident}_\varphi$ assumes the future φ to be total. How should this definition be adapted when the feature φ is partial? Two options are readily available. According to the *stronger* definition (11a), the identity faithfulness constraint penalizes corresponding segments which differ either in feature definedness (the feature is defined for only one of the two segments) or in feature value (the feature is defined for both segments but assigns them different values). According to the *weaker* definition (11b), the identity faithfulness constraint only penalizes pairs of corresponding segments which differ in feature value, but it assigns no violations when the feature is defined for only one of the two segments. Both definitions are trivially extended from pairs of corresponding segments to pairs of corresponding strings (of the same length) through the same clause as (10b).

(11) For any underlying and surface segments a and x:

a. $$\textsc{Ident}_\varphi^{\text{strong}}(a,x) = \begin{cases} 1 & \text{if } \varphi \text{ is defined for only one of } a \text{ and } x \\ & \text{or it is defined for both but } \varphi(a) \neq \varphi(x) \\ 0 & \text{otherwise} \end{cases}$$

b. $$\textsc{Ident}_\varphi^{\text{weak}}(a,x) = \begin{cases} 1 & \text{if } \varphi \text{ is defined for both } a \text{ and } x \text{ but } \varphi(a) \neq \varphi(x) \\ 0 & \text{otherwise} \end{cases}$$

The labels "strong" and "weak" capture the fact that $\textsc{Ident}_\varphi^{\text{strong}}$ is stronger (or more stringent) than $\textsc{Ident}_\varphi^{\text{weak}}$ (whenever the latter assigns a violation, the former does as well).

Which of these two definitions turns out to be preferable? As recalled in section 1, partial features arise in the theory of phonological underspecification or have been motivated by phonetic considerations. The latter phonetic literature is oblivious to the issue of the proper definition of identity constraints. Turning to underspecification theory, Harrison and Kaun (2001) entertain the hypothesis that Hungarian has an underlying vowel /E/ underspecified for quality. They consider the identity constraint $\textsc{Ident}_{[\textsc{quality}]}$ and assume that it is weakly defined: it does not penalize candidates which put the underlying underspecified vowel /E/ in correspondence with a fully specified surface vowel such as [e] or [æ]. As another example, Colina (2013) assumes that some dialects of Galician have a voiced velar obstruent G underspecified for continuancy. She considers the identity constraint $\textsc{Ident}_{[\textsc{cont}]}$ and assumes that it is weakly defined, namely that "outputs underspecified for continuancy do not violate $\textsc{Ident}_{[\textsc{cont}]}$ due to the fact that there is no continuancy specification that could correspond to the input" (p. 90). Yet, this endorsement of the weak definition (11b) of feature identity seems not supported by substantial arguments: indeed, the examples just mentioned turn out to be consistent with the strong construal (11a) as well.[5] This squib will show that learnability considerations instead do bear on the choice between weak and strong featural identity, favoring the latter.

3.4 Establishing OT output-drivenness in the case of partial features

Tesar's output-drivenness theorem 1 assumes that the faithfulness constraint set consists of identity faithfulness constraints all relative to total features. How does the theorem extend to partial features? The answer to this question is provided by the following theorem 2. Tesar's output-drivenness theorem 1 extends to a representational framework with partial features when the strong notion (11a) of featural identity is adopted, no matter which of the three options $\leq_{\text{sim}}^{\text{sparse}}$ or $\leq_{\text{sim}}^{\text{mid}}$ or $\leq_{\text{sim}}^{\text{dense}}$ is used to extend Tesar's original similarity order $\leq_{\text{sim}}^{\text{total}}$. The situation is different when the weak notion (11b) of feature identity is adopted instead: the extension only holds relative to $\leq_{\text{sim}}^{\text{sparse}}$ or $\leq_{\text{sim}}^{\text{mid}}$, but it does not hold relative to the strongest notion of output-drivenness relative to the similarity order $\leq_{\text{sim}}^{\text{dense}}$.

Theorem 2 (Output-drivenness with partial features) *Assume that the faithfulness constraint set consists of the identity faithfulness constraints relative to a set Φ of possibly partial phonological features.* ***[A]*** *If the strong definition (11a) of featural identity is adopted, the OT grammar corresponding to any ranking of the constraint set is output-driven relative to any of the similarity orders $\leq_{sim}^{sparse}$, $\leq_{sim}^{mid}$ or $\leq_{sim}^{dense}$ provided by definition 3.* ***[B]*** *If the weak definition (11b) of featural identity is adopted, the OT grammar corresponding to any ranking of the constraint set is output-driven relative to either similarity order $\leq_{sim}^{sparse}$ or $\leq_{sim}^{mid}$ but not relative to $\leq_{sim}^{dense}$.*

The proof of theorem 2A is straightforward. In fact, given a feature φ which is partial and n-valued, consider the corresponding feature $\widehat{\varphi}$ which coincides with φ when the latter is defined and otherwise assigns the "dummy value" 0 ("dummy" means that 0 does not figure among the n values taken by the original feature φ). Obviously, $\widehat{\varphi}$ is a total and $(n+1)$-valued feature. Furthermore, the strong identity constraint $\textsc{Ident}_{\varphi}^{\text{strong}}$ defined as in (11a) out of the partial and n-valued feature φ is identical to the faithfulness constraint $\textsc{Ident}_{\widehat{\varphi}}$ defined as in (10) out of the total and $(n+1)$-valued feature $\widehat{\varphi}$. Let $\widehat{\Phi}$ be the feature set obtained from the original feature set Φ by replacing each partial feature φ with the corresponding total feature $\widehat{\varphi}$. Obviously, the similarity order $\leq_{\text{sim}}^{\text{dense}}$ constructed by definition 3 out of the original partial feature set Φ coincides with the similarity order $\leq_{\text{sim}}^{\text{total}}$ constructed by definition 2 out of the derived total feature set $\widehat{\Phi}$. Tesar's theorem 1 ensures that, if identity constraints are defined in terms of strong feature identity (11a), the resulting OT grammars are all output-driven relative to $\leq_{\text{sim}}^{\text{dense}}$. Output-drivenness relative to $\leq_{\text{sim}}^{\text{sparse}}$ and $\leq_{\text{sim}}^{\text{mid}}$ finally follows from the hierarchy (9).

Turning to theorem 2B, the final appendix provides a proof that output-drivenness holds relative to $\leq_{\text{sim}}^{\text{sparse}}$ and $\leq_{\text{sim}}^{\text{mid}}$ in the case of weak featural identity. This proof is a simple extension of Tesar's proof of the original theorem 1, which is in turn greatly simplified taking advantage of the simplified representational framework assumed here. The rest of this section provides a counterexample that output-drivenness instead fails relative to the similarity order $\leq_{\text{sim}}^{\text{dense}}$ in the case of weak feature identity. The segment set Σ consists of the four segments *ð, z, b* and *f*. The feature set Φ consists of the single feature $\varphi = [\textsc{strident}]$. This feature is assumed to be only defined for coronals (see for instance Hayes 2009), namely only for *ð* (which has value $-$) and *z* (which has value $+$), not for *b* and *f*. The constraint set contains only one faithfulness constraint, namely the weak identity faithfulness constraint $\textsc{Ident}^{\text{weak}}_{[\textsc{strident}]}$ defined as in (11b). The constraint set furthermore contains two markedness constraints: *NonSibFric, which encodes a preference for sibilants among fricatives (and thus punishes [ð] and [f]); and *Labial, which encodes the markedness of labial place compared to coronal place (and thus punishes [b] and [f]). The two tableaux in (12) show a ranking which maps /ð/ to [b], because the two non-sibilant fricatives [ð] and [f] are ruled out by *NonSibFric and the sibilant fricative [z] is ruled out by $\textsc{Ident}^{\text{weak}}_{[\textsc{strid}]}$. The underlying form /f/ is instead mapped to [z], because the feature [strident] is undefined for /f/ and the mapping to the strident [z] thus does not violate $\textsc{Ident}^{\text{weak}}_{[\textsc{strid}]}$ by virtue of the weak construal of featural identity.

(12) a.

/ð/	IDENT$^{\text{weak}}_{[\text{STRID}]}$	*NONSIBFRIC	*LABIAL
[ð]		*!	
[z]	*!		
☞ [b]			*
[f]		*!	*

b.

/f/	IDENT$^{\text{WEAK}}_{[\text{STRID}]}$	*NONSIBFRIC	*LABIAL
[ð]		*!	
☞ [z]			
[b]			*!
[f]		*!	*

Let [b] play the role of the surface string $\mathbf{x}$ and let /ð/ and /f/ play the roles of the two underlying strings $\mathbf{a}$ and $\mathbf{b}$ respectively, as stated in (13). The similarity order $\leq^{\text{dense}}_{\text{sim}}$ is dense enough to hold between the two candidate pairs $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$ thus defined, as stated in (13). In fact, this similarity order $\leq^{\text{dense}}_{\text{sim}}$ is defined in terms of the disjunction (7c). The second disjunct of this disjunction fails, because the feature $\varphi = [\text{STRIDENT}]$ is defined for /ð/ (which plays the role of the unique segment of $\mathbf{a}$) but is undefined for /f/ (which plays the role of the unique segment of $\mathbf{b}$). Yet, the first disjunct does hold, because the feature $\varphi = [\text{STRIDENT}]$ is undefined for both /f/ (which plays the role of the unique segment of $\mathbf{b}$) and [b] (which plays the role of the unique segment of $\mathbf{x}$).

(13) $(\mathbf{a}, \mathbf{x}) \leq^{\text{dense}}_{\text{sim}} (\mathbf{b}, \mathbf{x})$, where $(\mathbf{a}, \mathbf{x}) =$ (/ð/, [b]) and $(\mathbf{b}, \mathbf{x}) =$ (/f/, [b]).

The OT grammar described in (12) is thus not output-driven relative to $\leq^{\text{dense}}_{\text{sim}}$, since it maps $\mathbf{a} =$ /ð/ but not $\mathbf{b} =$ /f/ to $\mathbf{x} =$ [b], despite the fact that $\mathbf{b} =$ /f/ is more similar to $\mathbf{x} =$ /b/ than $\mathbf{a} =$ /ð/ is, when similarity is measured through $\leq^{\text{dense}}_{\text{sim}}$.[6] The problem illustrated here falls under Tesar's (chapter 4) pattern of *distinction only at lesser similarity*: for the lesser similarity input $\mathbf{a} =$ /ð/, the faithfulness constraint IDENT$^{\text{weak}}_{[\text{STRID}]}$ distinguishes between the candidates (/ð/, [b]) and (/ð/, [z]); while for the greater similarity input $\mathbf{b} =$ /f/, the constraint does not distinguish between (/f/, [b]) and (/f/, [z]), because /f/ is undefined for stridency.

The following section explores the learnability implications of this conclusion that partial features threaten output-drivenness according to weak featural identity. This result is also relevant in its own right, as it contributes to an OT literature which accounts for opaque (and thus non-output-driven) child patterns (such as chain shifts) through the assumption that the child temporarily entertains underlying representations which are underspecified for certain features (see Dinnsen and Barlow 1998 and references therein).

4 Learnability implications of output-drivenness

This section shows that, within the hierarchy (9) of notions of output-drivenness, only output-drivenness relative to $\leq_{\text{sim}}^{\text{dense}}$ is strong enough for the learnability implications uncovered by Tesar to extend from total to partial phonological features.

4.1 Inconsistency detection

Tesar (2013, chapters 7 and 8) focuses on the implications of output-drivenness for the following classical formulation of the language learning problem (for a more detailed formulation of this problem, see Merchant 2008, section 1.1; Tesar 2013, chapter 6.2; and Magri 2015, section 2). Adhering to the generative perspective, a language learner is granted full knowledge of the typological space. The typology is defined in OT terms, namely through all possible rankings of a given constraint set. The learner's training data consist of a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ of surface forms sampled from the target language. The learner's task is to infer simultaneously a lexicon of corresponding underlying forms $\{\mathbf{a}_1, \mathbf{a}_2, \ldots\}$ and a constraint ranking such that the OT grammar corresponding to that ranking maps each of the inferred underlying forms $\mathbf{a}_1, \mathbf{a}_2, \ldots$ to the corresponding given surface forms $\mathbf{x}_1, \mathbf{x}_2, \ldots$.[7]

Knowledge of the target ranking would help the learner to reverse-engineer the target lexicon of underlying forms. On the other hand, knowledge of the target lexicon of underlying forms would allow the learner to easily infer the target ranking from the underlying/surface form mappings. The challenge raised by the learning problem is that both the lexicon of underlying forms and the ranking are unknown and thus need to be learned simultaneously. A natural strategy to cope with this challenge is to maintain partial lexical and grammatical/ranking information and to increment them iteratively by "boosting" one type of partial information with the other. Ignoring for the moment issues of algorithmic efficiency, this approach can be implemented through the scheme (14), explained below.

(14)
$$\begin{tikzcd} L_{\mathbf{x}} \arrow[r, bend left, "\text{ERI step}"] & R \arrow[l, bend left, "\text{ELI step}"] \end{tikzcd}$$

The learner represents its current ranking information through a set R of currently *admis-*

sible constraint rankings, which is initialized to the set of all possible rankings. The learner represents its current lexical information through a set $L_{\mathbf{x}}$ of currently *admissible* underlying forms for each training surface form $\mathbf{x}$, which is initialized to the set of all possible underlying forms for $\mathbf{x}$. The learner then enriches the current ranking and lexical information by iterating the following two steps. One step extracts ranking information (*ERI step*): the learner tries to zoom closer to the target constraint ranking, by eliminating from the set R of currently admissible rankings any rankings which are inconsistent with the current lexical information. The other step extracts lexical information (*ELI step*): the learner tries to zoom closer to the target underlying forms, by eliminating from the set $L_{\mathbf{x}}$ of currently admissible underlying forms for a training surface form $\mathbf{x}$ any underlying forms which are inconsistent with the current ranking information. The learner iterates the two ERI and ELI steps until, hopefully, the set R of admissible rankings has been reduced to just one ranking (or to just a few rankings, which all capture the target grammar). And the set $L_{\mathbf{x}}$ of admissible underlying forms for each training surface form $\mathbf{x}$ has been reduced to just one underlying form (or just a few underlying forms which only differ for features which are not contrastive). Because both the ERI and the ELI steps prune inconsistent options, the resulting learning scheme (14) is called *inconsistency detection* (Kager 1999; Tesar et al. 2003; Merchant 2008; Tesar 2006, 2013).

4.2 Speeding-up the ERI step

At the ERI step, the learner eliminates from the current set R of admissible rankings any ranking which is inconsistent with the current lexical information. A ranking is inconsistent with the current lexical information provided there exists some training surface form $\mathbf{x}$ such that every corresponding admissible underlying form in $L_{\mathbf{x}}$ fails at being mapped to $\mathbf{x}$ by that ranking. The ERI step can thus be made explicit as in (15).

(15) **ERI step:** eliminate from R any ranking whose corresponding OT grammar maps every underlying form in $L_{\mathbf{x}}$ into a surface form different from $\mathbf{x}$.

Tesar shows that this formulation (15) of the ERI step can be hugely simplified when the OT typology explored by the learner consists of grammars which are all output-driven relative to a similarity order $\leq_{\text{sim}}$. In fact, assume that the lexicon $L_{\mathbf{x}}$ of currently admissible underlying forms for each training surface form $\mathbf{x}$ admits a most similar underlying form when similarity is measured through the similarity order $\leq_{\text{sim}}$, as stated in (16).

(16) There exists a *most similar admissible underlying form*, namely there exists an underlying form $\mathbf{b}$ which belongs to $L_{\mathbf{x}}$ and satisfies the similarity inequality $(\mathbf{a}, \mathbf{x}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{x})$ for every other underlying form $\mathbf{a}$ in $L_{\mathbf{x}}$.

The ERI step (15) can then be equivalently reformulated as in (17). In fact, suppose that a ranking is eliminated by the original ERI step (15). This means that the corresponding grammar fails on every underlying form in $L_{\mathbf{x}}$, namely maps it to a surface form different from $\mathbf{x}$. Hence, that grammar fails in particular on the underlying form $\mathbf{b}$, as $\mathbf{b}$ belongs to $L_{\mathbf{x}}$. The ranking considered is thus also eliminated by the reformulated ERI step (17). Vice versa, assume by contradiction that there is some ranking which is eliminated by the reformulated ERI step (17) but not by the original ERI step (15). Since that ranking is not eliminated by the original ERI step (15), there exists at least one admissible underlying form $\mathbf{a}$ in $L_{\mathbf{x}}$ which is mapped to $\mathbf{x}$ by the corresponding OT grammar. Since that grammar is $\leq_{\text{sim}}$-output-driven (because of the assumption that all grammars in the typology are $\leq_{\text{sim}}$-output-driven) and since $(\mathbf{a}, \mathbf{x}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{x})$ (because of the assumption (16) that $\mathbf{b}$ is the most similar admissible underlying form), that grammar must also map $\mathbf{b}$ to $\mathbf{x}$. The ranking considered could thus not have been eliminated by the reformulated ERI step (17).

(17) **ERI step (reformulation based on output-drivenness):** eliminate from R any ranking whose corresponding OT grammar maps $\mathbf{b}$ to a surface form different from $\mathbf{x}$, where $\mathbf{b}$ is the most similar underlying form in $L_{\mathbf{x}}$, which exists by (16).

The original ERI step (15) requires the learner to look at all the admissible underlying forms in $L_{\mathbf{x}}$. The reformulation (17) achieves the same net result by looking only at one, namely only at the most similar one. This simplification is substantial. In fact, there are no implementations of the original ERI step (15) which are efficient.[8] The reformulation (17) can instead be executed efficiently.[9] This is indeed the most spectacular learnability implication of Tesar's notion of output-drivenness (see Magri 2015 for discussion).

4.3 Existence of the most similar admissible underlying form in the case of total features

In order to benefit from this spectacular speed-up, we need to establish the assumption (16) that the set $L_{\mathbf{x}}$ of currently admissible underlying forms for the training surface form $\mathbf{x}$ contains a most similar underlying form. This set $L_{\mathbf{x}}$ is initialized to the entire set of underlying forms for $\mathbf{x}$ and then iteratively pruned at the ELI step. Within the restrictive representational framework adopted in this paper, the surface form $\mathbf{x}$ is a segment string $\mathbf{x} = x_1 \cdots x_\ell$ of some length ℓ. The set $L_{\mathbf{x}}$ is thus initialized to the set of all underlying strings $\mathbf{a} = a_1 \cdots a_\ell$ of the same length ℓ (because candidate pairs consist of strings of the same length). The ELI step then works segment by segment, feature by feature: at each iteration, the ELI step tries to set the value of one of the features in Φ for one of the ℓ underlying segments, and eliminates from $L_{\mathbf{x}}$ all the underlying strings which have a different value for that feature and that segment.

With this little background on the current set $L_{\mathbf{x}}$ of admissible underlying forms, we can now take a closer look at the crucial assumption (16) that $L_{\mathbf{x}}$ admits a most similar underlying form. Assume for now that the features in the feature set Φ are all total and thus consider Tesar's original similarity order $\leq_{\text{sim}}^{\text{total}}$ provided by definition 2. Consider the underlying form $\mathbf{b} = b_1 \cdots b_\ell$ defined segment by segment and feature by feature as in (18).[10]

(18) a. If a feature φ has not yet been set for the ith underlying segment, then b_i has the same value for that feature as the ith surface segment x_i.

b. If a feature φ has already been set for the ith underlying segment to a certain value, then b_i has that value for that feature.

The form $\mathbf{b}$ defined in (18) validates the crucial assumption (16) because it is the admissible underlying form most similar to $\mathbf{x}$. In fact, condition (18b) says that the form $\mathbf{b}$ respects all the feature values which have been set so far and thus ensures that $\mathbf{b}$ belongs to $L_{\mathbf{x}}$ and thus counts as admissible. Furthermore, the underlying form $\mathbf{b}$ is most similar to $\mathbf{x}$ when similarity is measured relative to $\leq_{\text{sim}}^{\text{total}}$. In fact, consider any other underlying form $\mathbf{a} = a_1 \cdots a_\ell$ in the set $L_{\mathbf{x}}$. If a feature φ has not yet been set for the ith segment, clause (18b) guarantees that the two segments b_i and x_i have the same value for feature φ. This validates the first disjunct of the disjunctive definition (3) of the similarity order $\leq_{\text{sim}}^{\text{total}}$. If a feature φ has instead been set to a certain value for the ith segment, then the segment b_i has that value by clause (18b) and the segment a_i shares that same value (because underlying forms whose ith segment has a different value have been pruned from $L_{\mathbf{x}}$), validating the second disjunct of the definition (3) of $\leq_{\text{sim}}^{\text{total}}$.

4.4 Existence of the most similar admissible underlying form in the case of partial features

Let me now consider the case where the feature set Φ instead contains partial features. As noted in the preceding section, in this case we have at our disposal the hierarchy of similarity orders $\leq_{\text{sim}}^{\text{sparse}}$, $\leq_{\text{sim}}^{\text{med}}$, and $\leq_{\text{sim}}^{\text{dense}}$ provided by definition 3. The same reasoning as above shows that the form $\mathbf{b}$ defined analogously to (18) satisfies the crucial assumption (16) when similarity is measured relative to $\leq_{\text{sim}}^{\text{dense}}$. This is not surprising, as it was noted above that the similarity order $\leq_{\text{sim}}^{\text{dense}}$ represented in (6c) effectively treats the dummy value "0" which marks undefinedness as a plain feature value (e.g., it treats a partial binary feature as a total ternary feature). The situation is very different for the other two similarity orders $\leq_{\text{sim}}^{\text{sparse}}$ and $\leq_{\text{sim}}^{\text{med}}$: these similarity orders are too sparse to validate the crucial assumption (16) of a most similar admissible underlying form. To illustrate, let's focus on the latter similarity order $\leq_{\text{sim}}^{\text{med}}$ and consider the following minimal counterexample. Consider the feature $\varphi = [\text{STRIDENT}]$ and assume it is only defined for coronals (see for instance Hayes 2009). Consider a surface form $\mathbf{x} = x_1 \cdots x_\ell$. whose ith segment x_i is not coronal and is thus not assigned any value by the feature φ. Suppose that the feature φ has not been valued yet for the ith segment. Thus, $L_{\mathbf{x}}$ contains in particular: an underlying form $\mathbf{a}' = a'_1 \cdots a'_\ell$ whose ith segment a'_i is not coronal and thus undefined for stridency; an underlying form $\mathbf{a}'' = a''_1 \cdots a''_\ell$ whose ith segment a''_i is a strident coronal; and an underlying form $\mathbf{a}''' = a'''_1 \cdots a'''_\ell$, whose ith segment a'''_i is a non-strident coronal. We thus need in particular to compare the three candidate pairs $(\mathbf{a}', \mathbf{x})$, $(\mathbf{a}'', \mathbf{x})$, $(\mathbf{a}''', \mathbf{x})$ relative to the similarity order $\leq_{\text{sim}}^{\text{med}}$. This requires in particular checking whether any two of the pairs of feature values $(\varphi(a'_i), \varphi(x_i))$, $(\varphi(a''_i), \varphi(x_i))$, and $(\varphi(a'''_i), \varphi(x_i))$ are connected through a straight arrow in the diagram (6b). Since the feature φ is undefined for the surface segment x_i, the rightmost block of this diagram applies. But that block has no straight arrows. In other words, $\leq_{\text{sim}}^{\text{mid}}$ is too sparse and establishes no similarity relations among $(\mathbf{a}', \mathbf{x})$, $(\mathbf{a}'', \mathbf{x})$, and $(\mathbf{a}''', \mathbf{x})$. The crucial assumption (16) of a most similar admissible underlying form thus fails when similarity is measured relative to $\leq_{\text{sim}}^{\text{mid}}$. And the speed-up promised by output-drivenness is thus lost. Analogous considerations hold for the similarity order $\leq_{\text{sim}}^{\text{sparse}}$.

5 Conclusion

Suppose that a certain underlying form $\mathbf{a}$ is mapped to a certain surface form $\mathbf{x}$. Suppose that another underlying form $\mathbf{b}$ is at least as similar to $\mathbf{x}$ as $\mathbf{a}$ is. Output-drivenness then requires that the underlying form $\mathbf{b}$ be mapped to that surface form $\mathbf{x}$ as well. Output-drivenness is thus predicated on a notion of similarity which is formalized through a similarity order $\leq_{\text{sim}}$ defined among candidate pairs $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$ which share the surface form $\mathbf{x}$. Tesar (2013) develops a theory of output-drivenness under the assumption that all phonological features are total. Under this assumption, he manages to define the similarity order in such a way that it is not too strong, so that the corresponding notion of output-drivenness can be guaranteed for large typologies of OT grammars. And at the same time it is not too weak, so that the corresponding notion of output-drivenness has substantial learnability implications, for instance by validating the crucial condition (16) that lexicons of admissible underlying forms defined by feature-based inconsistency detection admit a unique most similar underlying form.

This squib has tackled the problem of extending output-drivenness to a framework which allows for partial features. Within this framework, the intuitive notion of similarity which underlies output-drivenness can be formalized through a hierarchy of similarity orders $\leq_{\text{sim}}^{\text{sparse}}$, $\leq_{\text{sim}}^{\text{mid}}$, and $\leq_{\text{sim}}^{\text{dense}}$, defined in section 2. They yield a hierarchy of corresponding notions of output-drivenness, ordered by their relative strength. Section 3 has established that output-drivenness relative to the similarity orders $\leq_{\text{sim}}^{\text{sparse}}$ and $\leq_{\text{sim}}^{\text{mid}}$ holds irrespectively of the details of the proper definition of featural identity. That is instead not the case for the stronger notion of output-drivenness relative to the similarity order $\leq_{\text{sim}}^{\text{dense}}$: it requires the strong definition (11a) of feature identity which also penalizes disparities in feature definedness, while the weak definition (11b) of feature identity does not suffice, because it only penalizes disparities in feature value. The difference between these two approaches to feature identity seems not to have been discussed in the phonological literature. Yet, it is consequential for phonological learnability: output-drivenness relative to the similarity orders $\leq_{\text{sim}}^{\text{sparse}}$ and $\leq_{\text{sim}}^{\text{mid}}$ is too weak to support the extension of Tesar's learnability implications from total to partial features, as shown in section 4. The stronger notion of output-drivenness relative to $\leq_{\text{sim}}^{\text{dense}}$ is needed instead, as that is the only similarity order which guarantees the crucial condition (16). In conclusion, the extension of output-drivenness from total to partial phonological features has subtle implications for the notion of featural identity: the strong construal of feature identity is needed, because the weak construal leads to a mismatch between what is needed (the stronger notion of output-drivenness relative to $\leq_{\text{sim}}^{\text{dense}}$ required for learnability) and what can be afforded (the weaker notion of output-drivenness relative to $\leq_{\text{sim}}^{\text{sparse}}$ or $\leq_{\text{sim}}^{\text{mid}}$ which is required for guarantees on OT output-drivenness).[11]

A Proof of theorem 2

This appendix provides a proof of theorem 2B, which extends theorem 1 to partial phonological features. For convenience, the proof is broken up into the three steps, corresponding to lemmas 1-3. The proof is a straightforward extension of the analysis developed in Tesar (2013, chapter 3), adapted to the restrictive phonological framework considered here. Throughout this appendix, Σ is a finite set of segments and Φ is a finite set of segmental features, which can be either binary or multi-valued, total or partial. A candidate pair is any pair of strings of segments of the same length.

A.1 The faithfulness output-drivenness condition

Lemma 1 guarantees output-drivenness of the OT grammar corresponding to any ranking of a constraint set whose faithfulness constraints all satisfy the *faithfulness output-drivenness condition* (FODC) stated in (19). No assumptions are made on the markedness constraints or on the similarity order. The proof is taken from Tesar (2013, section 3.2).

Lemma 1 *Consider an arbitrary similarity order $\leq_{sim}$ among candidate pairs. Assume that every faithfulness constraint F in the constraint set satisfies condition (19) for any candidate pairs $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$ such that $(\mathbf{a}, \mathbf{x}) \leq_{sim} (\mathbf{b}, \mathbf{x})$ and any string $\mathbf{y}$ of the same length as $\mathbf{x}$.*

(19) a. *If $F(\mathbf{b}, \mathbf{y}) < F(\mathbf{b}, \mathbf{x})$, then $F(\mathbf{a}, \mathbf{y}) < F(\mathbf{a}, \mathbf{x})$.*

b. *If $F(\mathbf{a}, \mathbf{x}) < F(\mathbf{a}, \mathbf{y})$, then $F(\mathbf{b}, \mathbf{x}) < F(\mathbf{b}, \mathbf{y})$.*

Then, the OT grammar corresponding to any ranking of the constraint set is output-driven relative to that similarity order $\leq_{sim}$.

Proof. Assume that the OT grammar corresponding to a certain ranking maps the less similar underlying string $\mathbf{a}$ to the surface string $\mathbf{x}$. Let me show that it then also maps the more similar underlying string $\mathbf{b}$ to that same surface string $\mathbf{x}$. This means that I have to show that the candidate pair $(\mathbf{b}, \mathbf{x})$ beats any other candidate pair $(\mathbf{b}, \mathbf{y})$ according to that ranking. The assumption that $\mathbf{a}$ is mapped to $\mathbf{x}$ entails in particular that the candidate pair $(\mathbf{a}, \mathbf{x})$ beats the candidate pair $(\mathbf{a}, \mathbf{y})$. This means in turn that one constraint C which prefers the winner pair $(\mathbf{a}, \mathbf{x})$ to the loser pair $(\mathbf{a}, \mathbf{y})$ is ranked above every constraint $C_1, C_2, \ldots$ which instead prefers $(\mathbf{a}, \mathbf{y})$ to $(\mathbf{a}, \mathbf{x})$, as represented in (20).

(20)
$$\begin{array}{cl} C & \text{a constraint which prefers } (\mathbf{a}, \mathbf{x}) \text{ to } (\mathbf{a}, \mathbf{y}) \\ | & \\ C_1, C_2, \ldots & \text{all the constraints which prefer } (\mathbf{a}, \mathbf{y}) \text{ to } (\mathbf{a}, \mathbf{x}) \end{array}$$

If the constraint C top ranked in (20) is a markedness constraint, then it does not care whether the underlying form is $\mathbf{a}$ or $\mathbf{b}$. The fact that it prefers $(\mathbf{a}, \mathbf{x})$ to $(\mathbf{a}, \mathbf{y})$ thus entails that it also prefers $(\mathbf{b}, \mathbf{x})$ to $(\mathbf{b}, \mathbf{y})$. If instead C is a faithfulness constraint, this entailment is guaranteed by the FODC (19b). The ranking conditions (20) can thus be updated as in (21).

(21)
$$\begin{array}{cl} C & \text{a constraint which } \boxed{\text{prefers } (\mathbf{b}, \mathbf{x}) \text{ to } (\mathbf{b}, \mathbf{y})} \\ | & \\ C_1, C_2, \ldots & \text{all the constraints which prefer } (\mathbf{a}, \mathbf{y}) \text{ to } (\mathbf{a}, \mathbf{x}) \end{array}$$

Consider a constraint which incorrectly prefers $(\mathbf{b}, \mathbf{y})$ to $(\mathbf{b}, \mathbf{x})$. If it is a markedness constraint, then again it also prefers $(\mathbf{a}, \mathbf{y})$ to $(\mathbf{a}, \mathbf{x})$, namely it is one of the constraints $C_1, C_2, \ldots$ ranked at the bottom of (21). If instead it is a faithfulness constraint, that same conclusion is guaranteed by the FODC (19a). The ranking conditions (21) can thus be updated as in (22).

(22)
$$\begin{array}{cl} C & \text{a constraint which prefers } (\mathbf{b}, \mathbf{x}) \text{ to } (\mathbf{b}, \mathbf{y}) \\ | & \\ C_1, C_2, \ldots & \text{all the constraints which } \boxed{\text{prefer } (\mathbf{b}, \mathbf{y}) \text{ to } (\mathbf{b}, \mathbf{x})} \end{array}$$

The ranking conditions (22) say that $(\mathbf{b}, \mathbf{x})$ wins over $(\mathbf{b}, \mathbf{y})$, as a constraint C which prefers $(\mathbf{b}, \mathbf{x})$ to $(\mathbf{b}, \mathbf{y})$ is ranked above every constraint $C_1, C_2, \ldots$ which instead prefers $(\mathbf{b}, \mathbf{y})$ to $(\mathbf{b}, \mathbf{x})$. Since this conclusion holds for any candidate $\mathbf{y}$, the OT grammar considered maps the more similar underlying string $\mathbf{b}$ to the surface string $\mathbf{x}$, as required by output-drivenness. $\square$

A.2 Simplifying the faithfulness output-drivenness condition

Lemma 1 states the FODC (19) for an arbitrary faithfulness constraint F. Lemma 2 specializes this condition to identity faithfulness constraints. This lemma and the following lemma 3 hold irrespectively of the choice between the strong construal (11a) and the weak construal (11b) of the identity constraint $\textsc{Ident}_\varphi$ relative to a partial feature φ. Yet, since the case of the strong construal is already covered by theorem 2A, I focus here on the case of the weak construal. The superscript "weak" introduced in subsection 3.3 is suppressed to simplify the notation.

Lemma 2 *Consider the weakly defined identity faithfulness constraint* $\textsc{Ident}_\varphi$ *relative to a feature* φ *possibly partial relative to the segment set* Σ. *The FODC (19) specializes to this specific case* $F = \textsc{Ident}_\varphi$ *as follows:*

(23) a. *If* $\sum_{i\in I} \textsc{Ident}_\varphi(b_i, y_i) < \sum_{i\in I} \textsc{Ident}_\varphi(b_i, x_i)$, *then* $\sum_{i\in I} \textsc{Ident}_\varphi(a_i, y_i) < \sum_{i\in I} \textsc{Ident}_\varphi(a_i, x_i)$.

b. *If* $\sum_{i\in I} \textsc{Ident}_\varphi(a_i, x_i) < \sum_{i\in I} \textsc{Ident}_\varphi(a_i, y_i)$, *then* $\sum_{i\in I} \textsc{Ident}_\varphi(b_i, x_i) < \sum_{i\in I} \textsc{Ident}_\varphi(b_i, y_i)$.

where the sums run over the set I *of those indices* $i = 1, \ldots, \ell$ *where the two surface strings* $\mathbf{x} = x_1 \cdots x_\ell$ *and* $\mathbf{y} = y_1 \cdots y_\ell$ *disagree relative to the feature* φ*:*

$$\text{(24)} \quad I = \left\{ i \in \{1, \ldots, \ell\} \;\middle|\; \begin{array}{l} \varphi \textit{ is defined for only one of the two segments } x_i \textit{ and } y_i \\ \textit{or it is defined for both but assigns them different values} \end{array} \right\}$$

Proof. Focus on the first FODC (19a); an analogous reasoning holds for the second FODC (19b). By (10)/(11), the identity faithfulness constraint $\textsc{Ident}_\varphi$ is defined for a candidate pair of strings by summing over pairs of corresponding segments. The first FODC (19a) can thus be made explicit as in (25) when the faithfulness constraint F is $\textsc{Ident}_\varphi$.

$$\text{(25)} \quad \text{If } \underbrace{\sum_{i=1}^{\ell} \textsc{Ident}_\varphi(b_i, y_i)}_{F(\mathbf{b},\mathbf{y})} < \underbrace{\sum_{i=1}^{\ell} \textsc{Ident}_\varphi(b_i, x_i)}_{F(\mathbf{b},\mathbf{x})}, \text{ then } \underbrace{\sum_{i=1}^{\ell} \textsc{Ident}_\varphi(a_i, y_i)}_{F(\mathbf{a},\mathbf{y})} < \underbrace{\sum_{i=1}^{\ell} \textsc{Ident}_\varphi(a_i, x_i)}_{F(\mathbf{a},\mathbf{x})}.$$

In step (26a) and in (27), the sum over $\{1, \ldots, \ell\}$ has been split into two sums over I and over its complement. In step (26b), y_i has been replaced with x_i in the second sum which runs over the complement of I, because x_i and y_i agree relative to the feature φ for every index $i \notin I$ and the faithfulness constraint $\textsc{Ident}_\varphi$ thus cannot distinguish between them.

$$(26) \quad \sum_{i=1}^{\ell} \textsc{Ident}_{\varphi}(b_i, y_i) \overset{(a)}{=} \sum_{i \in I} \textsc{Ident}_{\varphi}(b_i, y_i) + \sum_{i \notin I} \textsc{Ident}_{\varphi}(b_i, y_i)$$
$$\overset{(b)}{=} \sum_{i \in I} \textsc{Ident}_{\varphi}(b_i, y_i) + \sum_{i \notin I} \textsc{Ident}_{\varphi}(b_i, x_i)$$

$$(27) \quad \sum_{i=1}^{\ell} \textsc{Ident}_{\varphi}(b_i, x_i) = \sum_{i \in I} \textsc{Ident}_{\varphi}(b_i, x_i) + \sum_{i \notin I} \textsc{Ident}_{\varphi}(b_i, x_i)$$

Because of (26) and (27), the inequality in the antecedent of (25) is equivalent to the inequality in the antecedent of (23a), as the quantity $\sum_{i \notin I} \textsc{Ident}_{\varphi}(b_i, x_i)$ appears on both sides and can thus be ignored. An analogous reasoning shows that that the inequality in the consequent of (25) is equivalent to the inequality in the consequent of (23a). □

A.3 Verifying the faithfulness output-drivenness condition

The following lemma 3 guarantees that identity faithfulness constraints satisfy the FODC (23) when the similarity order is properly defined, thus completing the proof of theorem 2.

Lemma 3 *The weakly defined identity faithfulness constraint* $\textsc{Ident}_{\varphi}$ *relative to any (possibly partial) feature* φ *in* Φ *satisfies the FODC (23) relative to the similarity orders* $\leq_{sim}^{sparse}$ *or* $\leq_{sim}^{mid}$ *provided by definition 3.*

Proof. Focus on the first FODC (23a); an analogous reasoning holds for the second FODC (23b). Furthermore, focus on the similarity order $\leq_{\text{sim}}^{\text{mid}}$; the claim then obviously extends to the sparser similarity order $\leq_{\text{sim}}^{\text{sparse}}$. The similarity order $\leq_{\text{sim}}^{\text{mid}}$ is defined in terms of the disjunction (7b), repeated in (28).

(28) a. Either: feature φ is defined for both b_i and x_i and furthermore $\varphi(b_i) = \varphi(x_i)$.

b. Or: feature φ is undefined for both b_i and a_i or it is defined for both and $\varphi(b_i) = \varphi(a_i)$.

Each of the two disjuncts (28a) and (28b) individually entails each of the two implications (29). That is obvious for the second disjunct (28b), as it says that the two segments a_i and b_i do not differ relative to the feature φ and thus cannot be distinguished by the faithfulness constraint $\textsc{Ident}_{\varphi}$. Furthermore, the first disjunct (28a) entails the implication (29a) because it ensures that its consequent is true, namely that $\textsc{Ident}_{\varphi}(b_i, y_i) = 1$. In fact, feature φ is defined for segment y_i, because otherwise the antecedent $\textsc{Ident}_{\varphi}(a_i, y_i) = 1$ would be false according to the weak construal (11b) of feature identity. Feature φ is also defined for segment

b_i, as ensured by the disjunct (28a). Finally, the two segments b_i and y_i disagree relative to the feature φ, because the disjunct (28a) says that $\varphi(b_i) = \varphi(x_i)$ and the hypothesis that $i \in I$ says that $\varphi(x_i) \neq \varphi(y_i)$. Finally, the first disjunct (28a) also entails the implication (29b) because it ensures that its antecedent is false, namely that $\text{IDENT}_\varphi(b_i, x_i) = 0$. In fact, the disjunct (28a) says in particular that $\varphi(b_i) = \varphi(x_i)$, so that the segment pair (b_i, x_i) does not violate IDENT_φ.

(29) For every $i \in I$:

a. if $\text{IDENT}_\varphi(a_i, y_i) = 1$, then also $\text{IDENT}_\varphi(b_i, y_i) = 1$;

b. if $\text{IDENT}_\varphi(b_i, x_i) = 1$, then also $\text{IDENT}_\varphi(a_i, x_i) = 1$.

The chain of inequalities in (30) finally shows that the faithfulness constraint IDENT_φ indeed satisfies the first FODC (23a). Steps (30a) and (30c) are guaranteed by the implications (29a) and (29b), respectively. Step (30b) is guaranteed by the antecedent of the FODC (23a).

(30) $$\sum_{i \in I} \text{IDENT}_\varphi(a_i, y_i) \overset{(a)}{\leq} \sum_{i \in I} \text{IDENT}_\varphi(b_i, y_i) \overset{(b)}{<} \sum_{i \in I} \text{IDENT}_\varphi(b_i, x_i) \overset{(c)}{\leq} \sum_{i \in I} \text{IDENT}_\varphi(a_i, x_i)$$

The logical structure of this proof can be highlighted as follows. The FODC (23a) for the identity faithfulness constraint IDENT_φ only looks at those indices $i \in I$ where the two candidates $\mathbf{x}$ and $\mathbf{y}$ differ for the feature φ, yielding the two reverse implications (29a) and (29b). The lemma trivially follows from these two implications, as shown in (30). □

Notes

[1]Another dichotomy which has figured prominently in the literature on feature systems is the one between *univalent/privative* features and *binary/multivalued* features. Although crucial within derivational frameworks, this distinction seems to play only a marginal role within OT. For instance, Wetzels and Mascaró (2001, p. 237) address the issue of whether the feature [VOICE] is privative or binary and conclude that "it is [...] much harder to argue against privativity in OT than it is in derivational phonology. This is due to the fact that with the IDENT and AGREE constraint families it is as easy to refer to the absence of a privative feature as it is to refer to the unmarked value of a binary feature." Since Tesar's theory of output-drivenness is casted within OT, the extension of the theory to privative features is a less pressing issue than its extension to partial features.

[2]Boldfaced letters from the beginning of the alphabet $(\mathbf{a}, \mathbf{b}, \ldots)$ and from the end of the alphabet $(\ldots \mathbf{x}, \mathbf{y}, \mathbf{z})$ are used to denote underlying and surface forms, respectively.

[3]The relation $\leq_{\text{sim}}^{\text{total}}$ is obviously reflexive and transitive. In order for it to also be antisymmetric (and thus qualify as a partial order among candidate pairs), the feature set Φ must be rich enough: for any two different segments in Σ, there has got to exist a feature in Φ which assigns them a different value.

[4]As noted above, each of the nodes in the diagrams (6) must come with a loop arrow in order to ensure reflexivity of the resulting similarity relation. Hence, the three disjunctions in (7) all share the same second disjunct corresponding to the loop arrows. The three disjunctions only differ for their first disjunct, corresponding to the straight arrows.

[5]Indeed, nothing would in change in tableaux E and F of Harrison and Kaun (2001, pp. 221-222) if $\text{IDENT}^{\text{weak}}_{[\text{QUALITY}]}$ where replaced with $\text{IDENT}^{\text{strong}}_{[\text{QUALITY}]}$. Analogously, nothing would change in tableau (18) in Colina (2013, p. 92) if $\text{IDENT}^{\text{weak}}_{[\text{CONT}]}$ where replaced with $\text{IDENT}^{\text{strong}}_{[\text{CONT}]}$. Thanks to an anonymous reviewer for advice on this point.

[6]The similarity orders $\leq_{\text{sim}}^{\text{mid}}$ and $\leq_{\text{sim}}^{\text{sparse}}$ are instead sparse enough not to hold between these two candidate pairs $(\mathbf{a}, \mathbf{x})$ and $(\mathbf{b}, \mathbf{x})$. In fact, $\leq_{\text{sim}}^{\text{mid}}$ is defined in terms of the disjunction (7b). The second disjunct of this disjunction again fails. And the first disjunct fails as well, because it requires in particular that the feature $\varphi = [\text{STRIDENT}]$ be defined for both /f/ (which plays the role of the unique segment of $\mathbf{b}$) and [b] (which plays the role of the unique segment of $\mathbf{x}$), which is not the case. The fact that the OT grammar described in (12) maps $\mathbf{a} =$ /ð/ but not $\mathbf{b} =$ /f/ to $\mathbf{x} =$ [b] is not an obstacle to its output-drivenness relative to $\leq_{\text{sim}}^{\text{sparse}}$ or $\leq_{\text{sim}}^{\text{mid}}$, since $\mathbf{b} =$ /f/ is not more similar to $\mathbf{x} =$ /b/ than $\mathbf{a} =$ /ð/ is, when similarity is measured through the more demanding standards of $\leq_{\text{sim}}^{\text{mid}}$ or $\leq_{\text{sim}}^{\text{sparse}}$.

[7]Strictly speaking, the learning problem thus formulated is trivial: it admits the trivial solution where the underlying forms are all identical to the given surface forms and the OT grammar corresponds to a ranking with faithfulness constraints at the top. Indeed, more precise formulations of this learning problem considered in the literature rule out this trivial solution through additional conditions. First, the problem formulation is usually refined with some restrictiveness condition which rules out the choice of a grammar with faithfulness at the top (Tesar 2013,

chapter 9). Second, the problem formulation is usually refined with conditions on the lexicon of underlying forms which rule out the choice of fully faithful underlying forms. For instance, the learner is typically assumed to be trained not on a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ of unadorned surface forms but on a paradigm of morphologically decomposed surface forms annotated with the corresponding meanings. For instance, a learner of Dutch would be trained on surface forms such as $\mathbf{x}_1$ =[pɑt+∅] and $\mathbf{x}_2$ =[pɑd+ən], split up into stem and suffix and annotated with the corresponding meanings TOADSING and TOADPL. These meanings say that the two surface strings $\mathbf{x}_1, \mathbf{x}_2$ correspond to two underlying strings which share the underlying stem. Since the two surface stems differ in voicing of their final consonant, the final consonant of the shared underlying stem must be non-faithful relative to one of the two surface stems. To simplify the presentation and avoid orthogonal issues, this section focuses on the bare formulation of the learning problem, without additional conditions on the grammars or the underlying forms.

[8] Let me clarify this statement. So far, I have assumed the learner to represent its current ranking information in terms of a set R of currently admissible rankings which is initialized to the set of all possible rankings and then pruned iteratively at the ERI step. This is of course unfeasible: there are just too many rankings! A more compact representation of the current ranking information is needed. A natural idea is to represent a possibly large set R of constraint rankings through a small set of ranking conditions. Prince (2002) develops the formalism of *elementary ranking conditions* (ERCs) and Merchant (2008) suggests using that specific format to compactly represent the learner's ranking information. Thus, the learner starts from an empty set of ERCs, corresponding to the set of all constraint rankings. And the ERI step (15) updates the current set of ERCs by enriching it with the additional set of ERCs corresponding to the set of rankings which map at least one underlying form in $L_{\mathbf{x}}$ to the surface form $\mathbf{x}$. Merchant (2008) develops a sophisticated technique to tackle the difficult problem of computing the latter set of ERCs. Unfortunately, his technique relies on the operation of *fusional closure*. There are no results (I am aware of) on the efficient computation of the fusional closure, although it is unlikely that it can be computed efficiently in the general case. In conclusion, the price to pay for compressing the data structure from sets of constraint rankings to sets of ERCs is that the original formulation of the ERI step (15) becomes computationally demanding.

[9]Let me clarify this statement. Tesar shows that the most $\leq_{\text{sim}}^{\text{total}}$-similar admissible underlying form $\mathbf{b}$ in $L_{\mathbf{x}}$ can be computed quickly (without enumerating all forms in $L_{\mathbf{x}}$) through the prescription in (18) below. Suppose that the learner represents its current ranking information in terms of sets of ERCs, as explained in footnote 8. The reformulated ERI step (17) can be described in terms of ERCs as follows: enrich the current set of ERCs with the additional set of ERCs which are satisfied by all and only the rankings which map the most similar underlying form $\mathbf{b}$ to the surface form $\mathbf{x}$. The latter set of ERCs involves a single underlying form $\mathbf{b}$ and can therefore be computed efficiently.

[10]There is an obvious problem which looms out of the definition (18): if there are dependencies between the features already set and those still unset, there might exist no form $\mathbf{b}$ which satisfies this definition. This is a general problem which afflicts this approach, which is orthogonal to the issue raised by partial features.

[11]An anonymous reviewer points out that partial features raise no threat to alternative approaches to the problem of learning underlying forms, such as those by Jarosz (2006) and Riggle (2006). Yet, those approaches and Tesar's approach have very different goals. Tesar's theory of output-drivenness is explicitly construed as an attempt at using assumptions on constraints and representations to improve learning speed and algorithmic efficiency. Indeed, the major learnability implication of output-drivenness is a substantial speed-up of one of the subroutines of the inconsistency detection approach, as recalled in section 4. The mentioned alternative approaches make no representational assumptions but ignore issues of algorithmic efficiency. For instance, Riggle (p. 348-349) explicitly acknoweldges that "sorting, updating, and storing a potentially huge number of ⟨grammar, input-set⟩ pairs in a realistic on-line fashion that doesn't require unreasonable amounts of memory [...] is one of the biggest problems in scaling up from toy grammars to real grammars."

References

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology and morphology. Doctoral Dissertation, MIT.

Beckman, Mary E. 1999. *Positional faithfulness: An Oprimality Theretic treatment of phono-*

logical asymmetries. New York: Garland Publishing (Outstanding dissertations in Linguistics).

Colina, Sonia. 2013. Galician gelada: in defense of underspecification in Optimality Theory. *Lingua* 133:84–100.

Dinnsen, Daniel A., and Jessica A. Barlow. 1998. On the characterization of a chain shift in normal and delayed phonological acquisition. *Journal of Child Language* 25:61–94.

Dresher, B Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge University Press.

Hall, T. A. 2007. Segmental features. In *The cambridge handbook of phonology*, ed. Paul de Lacy, 311–334. Cambridge: Cambridge University Press.

Harrison, K. David, and Abigail Kaun. 2001. Patterns, pervasive patterns, and feature specification. In *Distinctive feature theory*, ed. T. Alan Hall, 211–236. Mouton de Gruyter.

Hayes, Bruce. 2009. *Introductory phonology*. Wiley-Blackwell.

Inkelas, Sharon. 1995. The consequences of optimization for underspecification. In *Proceedings of the Northeastern Linguistic Society 25*, ed. J. Beckman, 287–302. GLSA.

Jarosz, Gaja. 2006. Richness of the base and probabilistic unsupervised learning in Optimality Theory. In *Proceedings of the 8th meeting of the ACL special interest group on computational phonology at HTL-NAACL 2006*, 50–59. Association for Computational Linguistics.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 5:275–292.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In *Linguistics in the morning calm*, ed. I-S Yang, 3–91. Hanshin.

de Lacy, Paul. 2006. *Markedness. reduction and preservation in phonology*. Cambridge University Press.

Łubowicz, Anna. 2011. Chain shifts. In *Companion to phonology*, ed. Marc van Oostendorp, Beth Ewen, Colin Hume, and Keren Rice, 1717–1735. Wiley-Blackwell.

Magri, Giorgio. 2015. Review of Tesar's (2014) 'Output-driven phonology: Theory and learning'. *Phonology* 31:525–556.

McCarthy, John J. 1988. Feature geometry and dependency: a review. *Phonetica* 45:84–108.

McCarthy, John J., and Alan Prince. 1993. Generalized alignment. In *Yearbook of morphology 1993*, ed. Geert Booij and Jaap van Marl, 79–153. Dordrecht: Kluwer.

Merchant, Nazarré. 2008. Discovering underlying forms: contrast pairs and ranking. Doctoral Dissertation, Rutgers University.

Mester, Armin, and Junko Ito. 1989. Feature predictability and underspecification: palatal prosody in Japanese mimetics. *Language* 65:258–293.

Prince, Alan. 2002. Entailed ranking arguments. Ms., Rutgers University, New Brunswick, NJ. Rutgers Optimality Archive, ROA 500. Available at http://www.roa.rutgers.edu.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell. As Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. Also available as ROA 537 version.

Riggle, Jason. 2006. Using entropy to learn OT grammars from surface forms alone. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, ed. Donald Baumer, David Montero, and Michael Scanlon, 346–353. Cascadilla Proceedings Project.

Steriade, Donca. 1995. Underspecification and markedness. In *Handbook of phonological theory*, ed. John Goldsmith, 114–174. Oxford: Blackwell.

Steriade, Donca. 2001. Directional asymmetries in place assimilation. In *The role of speech perception in phonology*, ed. Hume, Elizabeth V. and Johnson, Keith, 219–249. Academic Press.

Tesar, Bruce. 2006. Faithful contrastive features in learning. *Cognitive Science* 30:863–903.

Tesar, Bruce. 2013. *Output-driven phonology: Theory and learning*. Cambridge Studies in Linguistics.

Tesar, Bruce, John Alderete, Graham Horwood, Nazarré Merchant, Koichi Nishitani, and Alan Prince. 2003. Surgery in language learning. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, ed. Gina Garding and Mimu Tsujimura, 477–490. Somerville, MA, USA: Cascadilla Proceedings Project.

Wetzels, W. Leo, and Joan Mascaró. 2001. The typology of voicing and devoicing. *Language* 77:207–244.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130.1:96–115.