MaxEnt is as transparent as OT and HG

□ Tesar (2014) extends the distinction between transparency and opacity from rule-based to arbitrary categorical phonological grammars through the notion of output-drivenness. Anttila and Magri (2018) extend entailments such as the one used to define output-drivenness from categorical to probabilistic phonological grammars. Building on these two developments, this paper investigates transparency/opacity for probabilistic MaxEnt grammars. The main result is that the same condition on the faithfulness constraints which is responsible for the transparency of categorical OT and HG also suffices to ensure the transparency of MaxEnt. This result is surprising in light of Anttila and Magri's finding that ME breaks many of the entailments predicted by OT and HG.

□ Let $\langle \mathsf{x}, \mathbb{A}, \mathbb{B} \rangle = \mathsf{y}$ denote the SPE derivation which starts with the UR x and applies rule $\mathbb{A}$ followed by rule $\mathbb{B}$ to yield the SR y. Hence $\langle \mathsf{x}, \mathbb{A}, \mathbb{B} \rangle = \langle \langle \mathsf{x}, \mathbb{A} \rangle, \mathbb{B} \rangle$. Recall that rule $\mathbb{B}$ **counter-feeds** rule $\mathbb{A}$ in this derivation iff $\mathbb{A}$ applies vacuously to x (1a) but not to the result of applying $\mathbb{B}$ to x (1b).

(1) a. $\langle \mathsf{x}, \mathbb{A} \rangle = \mathsf{x}$
b. $\langle \mathsf{x}, \mathbb{B}, \mathbb{A} \rangle \neq \langle \mathsf{x}, \mathbb{B} \rangle$

My first contribution is that counter-feeding entails **chain shifts**. More precisely, consider the form $\mathsf{z} = \langle \mathsf{x}, \mathbb{B} \rangle$. Suppose that $\langle \mathsf{z}, \mathbb{A}, \mathbb{B} \rangle$ is not a **Duke of York** (DOY) derivation. This assumption is mild because DOYs are rare. Under this mild assumption, if $\mathsf{y} = \langle \mathsf{x}, \mathbb{A}, \mathbb{B} \rangle$ is a counter-feeding derivation, the SPE phonology corresponding to the ordered rules $\langle \mathbb{A}, \mathbb{B} \rangle$ yields the chain shift $\mathsf{x} \rightarrow \mathsf{y} \not\rightarrow \mathsf{y}$.

□ Here is the proof. Recall that the derivation $\langle \mathsf{z}, \mathbb{A}, \mathbb{B} \rangle$ would be a DOY if $\mathbb{A}$ modifies z (2a) but $\mathbb{B}$ subsequently gets back the initial form z (2b). The claim follows by applying the SPE phonology $\langle \mathbb{A}, \mathbb{B} \rangle$ to the form $\mathsf{y} = \langle \mathsf{x}, \mathbb{A}, \mathbb{B} \rangle$ as in (3), obtaining a form different from y, whereby the chain shift. In step (3a), I have used the definition of the form y as the result of applying $\langle \mathbb{A}, \mathbb{B} \rangle$ to x. In step (3b), I have used the counter-feeding condition (1a), whereby $\mathbb{A}$ applies vacuously to x. Since $\mathsf{z} = \langle \mathsf{x}, \mathbb{B} \rangle$, the counter-feeding condition (1b) can be restated as $\langle \mathsf{z}, \mathbb{A} \rangle \neq \mathsf{z}$. In other words, the DOY condition (2a) holds. The other DOY condition (2b) must therefore fail. Since $\mathsf{z} = \langle \mathsf{x}, \mathbb{B} \rangle$, failure of (2b) means $\langle \langle \mathsf{x}, \mathbb{B} \rangle, \mathbb{A}, \mathbb{B} \rangle \neq \langle \mathsf{x}, \mathbb{B} \rangle$, whereby (3c). In step (3d), I have used again the counter-feeding condition (1a). Finally in step (3e), I have used again the definition $\mathsf{y} = \langle \mathsf{x}, \mathbb{A}, \mathbb{B} \rangle$.

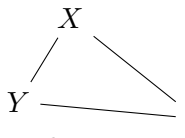
(2) a. $\langle \mathsf{z}, \mathbb{A} \rangle \neq \mathsf{z}$
b. $\langle \mathsf{z}, \mathbb{A}, \mathbb{B} \rangle = \mathsf{z}$

(3) $\langle \mathsf{y}, \mathbb{A}, \mathbb{B} \rangle \overset{(a)}{=} \langle \langle \mathsf{x}, \mathbb{A}, \mathbb{B} \rangle, \mathbb{A}, \mathbb{B} \rangle \overset{(b)}{=} \langle \langle \mathsf{x}, \mathbb{B} \rangle, \mathbb{A}, \mathbb{B} \rangle \overset{(c)}{\neq} \langle \mathsf{x}, \mathbb{B} \rangle \overset{(d)}{=} \langle \langle \mathsf{x}, \mathbb{A} \rangle, \mathbb{B} \rangle \overset{(e)}{=} \mathsf{y}$

□ A **categorical** grammar G (construed as a mapping from URs to SRs) is called **idempotent** provided it yields no chain shifts. Equivalently, it satisfies the implication (4): whenever G maps a UR x to a SR y, it also maps y (construed as a UR) to itself. Since counter-feeding entails chain-shifts as shown above, non-idempotency generalizes counter-feeding opacity beyond rule-based phonology.

(4) $G(\mathsf{x}) = \mathsf{y} \Longrightarrow G(\mathsf{y}) = \mathsf{y}$

(5) $F(\mathsf{x}, \mathsf{z}) \leq F(\mathsf{x}, \mathsf{y}) + F(\mathsf{y}, \mathsf{z})$



Magri (2018) shows that OT and HG grammars are idempotent provided each faithfulness constraint F satisfies the **triangle inequality** (TI) in (5): the violations $F(\mathsf{x}, \mathsf{z})$ assigned to the mapping of a UR x to a SR z are never larger than the violations $F(\mathsf{x}, \mathsf{y})$ assigned to the mapping of the UR x to some other SR y plus the violations $F(\mathsf{y}, \mathsf{z})$ assigned to the mapping of y (construed as a UR) to the SR z. The intuition is that F measures the phonological distance between URs and SRs and it thus complies with an axiom on distances which says that any side XZ of a triangle is shorter than the sum of the other two sides XY and YZ. The well-known resistance of OT and HG to chain shifts is then explained by the fact that most faithfulness constraints in the literature indeed satisfy the TI.

□ This paper extends these results to **MaxEnt** (ME). A ME grammar G^{ME} maps a UR x to a probability distribution $G^{\mathrm{ME}}(\cdot \,|\, \mathsf{x})$ over the candidate set of x. Following Anttila and Magri (2018),

(6) $G^{\mathrm{ME}}(\mathsf{y} \,|\, \mathsf{x}) \leq G^{\mathrm{ME}}(\mathsf{y} \,|\, \mathsf{y})$

I generalize the idempotency entailment (4) to this probabilistic setting as the condition (6) that the probability $G^{\mathrm{ME}}(\mathsf{y} \,|\, \mathsf{y})$ assigned to the consequent faithful mapping of y to itself is not smaller than the probability $G^{\mathrm{ME}}(\mathsf{y} \,|\, \mathsf{x})$ of the antecedent mapping of the UR x to this SR y. Condition (4) is indeed a special case of (6) for probability distributions which are categorical, namely concentrated on a single candidate. My second contribution is that the faithfulness TI (5) also suffices to ensure that ME satisfies (6) and is thus idempotent (whenever the URs x and y share the same candidate set).

□ Here is the proof. Using the analytical expression of ME, the idempotency inequality (6) is made explicit in (7). Here, $H_{\mathbf{w}}(\mathrm{UR}, \mathrm{SR})$ is the HG **harmony** of the mapping $(\mathrm{UR}, \mathrm{SR})$ relative to the **weight vector w**. The sums in the denominators run over all candidates z in the shared candidate set which are different from y. I multiply the numerator and the denominator of the left hand side of (7) by $e^{H_{\mathbf{w}}(\mathsf{y},\mathsf{y})}$. I also multiply the numerator and the denominator of the right hand side by $e^{H_{\mathbf{w}}(\mathsf{x},\mathsf{y})}$. After some simplifications, I can thus rewrite (7) as in (8) (with the inequality reversed). Since a faithfulness constraint F assigns zero violations to the faithful mapping (y, y), I can add $F(\mathsf{y}, \mathsf{y})$ to the left hand side of the TI (5), obtaining (9a). Furthermore, since a markedness constraint M is oblivious to the URs, the identity (9b) holds. Since the harmony $H_{\mathbf{w}}(\cdot, \cdot)$ is the weighted sum of the constraint violations multiplied by -1, the two inequalities (9) yield the harmony inequality $H_{\mathbf{w}}(\mathsf{x}, \mathsf{z}) + H_{\mathbf{w}}(\mathsf{y}, \mathsf{y}) \geq H_{\mathbf{w}}(\mathsf{y}, \mathsf{z}) + H_{\mathbf{w}}(\mathsf{x}, \mathsf{y})$ (the inequality is reversed because of -1), whereby (8) (because the exponential is increasing).

$$(7)\quad \frac{e^{H_{\mathbf{w}}(\mathsf{x},\mathsf{y})}}{e^{H_{\mathbf{w}}(\mathsf{x},\mathsf{y})} + \sum_{\mathsf{z}\neq\mathsf{y}} e^{H_{\mathbf{w}}(\mathsf{x},\mathsf{z})}} \leq \frac{e^{H_{\mathbf{w}}(\mathsf{y},\mathsf{y})}}{e^{H_{\mathbf{w}}(\mathsf{y},\mathsf{y})} + \sum_{\mathsf{z}\neq\mathsf{y}} e^{H_{\mathbf{w}}(\mathsf{y},\mathsf{z})}}$$

$$(8)\quad \sum_{\mathsf{z}\neq\mathsf{y}} e^{H_{\mathbf{w}}(\mathsf{x},\mathsf{z})+H_{\mathbf{w}}(\mathsf{y},\mathsf{y})} \geq \sum_{\mathsf{z}\neq\mathsf{y}} e^{H_{\mathbf{w}}(\mathsf{x},\mathsf{y})+H_{\mathbf{w}}(\mathsf{y},\mathsf{z})}$$

$$(9)\quad \begin{array}{ll} \text{a.} & F(\mathsf{x},\mathsf{z}) + F(\mathsf{y},\mathsf{y}) \leq F(\mathsf{x},\mathsf{y}) + F(\mathsf{y},\mathsf{z}) \\ \text{b.} & M(\mathsf{x},\mathsf{z}) + M(\mathsf{y},\mathsf{y}) = M(\mathsf{x},\mathsf{y}) + M(\mathsf{y},\mathsf{z}) \end{array}$$

□ I now turn to counter-bleeding opacity. Recall that rule $\mathbb{B}$ **counter-bleeds** rule $\mathbb{A}$ in the derivation $\langle \mathsf{x}, \mathbb{A}, \mathbb{B}\rangle$ provided $\mathbb{A}$ applies non-vacuously to x (10a) but vacuously to the result of applying $\mathbb{B}$ to x (10b). My third contribution is that counter-bleeding entails **saltations**. More precisely, suppose that $\mathbb{A}$ does not bleed $\mathbb{B}$ in the derivation $\langle \mathsf{x}, \mathbb{A}, \mathbb{B}\rangle$. This assumption is mild because the derivation $\langle \mathsf{x}, \mathbb{B}, \mathbb{A}\rangle$ would otherwise be a case of **mutual bleeding**, which is uncommon (Bakovic 2011). Under this mild assumption, if $\mathsf{z} = \langle \mathsf{x}, \mathbb{A}, \mathbb{B}\rangle$ is a counter-bleeding derivation, the SPE phonology corresponding to the ordered rules $\langle \mathbb{A}, \mathbb{B}\rangle$ yields the saltation $\mathsf{x} \curvearrowright \mathsf{y} \curvearrowright \mathsf{z}$: the form $\mathsf{y} = \langle \mathsf{x}, \mathbb{B}\rangle$ is mapped to itself but x is mapped to z, despite the fact x is closer to y than to z, as y is obtained from x through rule $\mathbb{B}$ only while z is obtained from x through both non-vacuous rules $\mathbb{A}$ and $\mathbb{B}$. The proof is omitted for space, but it is similar in spirit to the proof of my first contribution.

$$(10)\quad \begin{array}{ll} \text{a.} & \langle \mathsf{x}, \mathbb{A}\rangle \neq \mathsf{x} \\ \text{b.} & \langle \mathsf{x}, \mathbb{B}, \mathbb{A}\rangle = \langle \mathsf{x}, \mathbb{B}\rangle \end{array}$$

□ Consider two mappings (x, z) and (y, z) which share the same SR z. Suppose that the UR y is more similar to the shared SR z than the UR x is. In other words, that the mapping (y, z) has more "internal similarity" than the other mapping (x, z). This assumption is formalized as $(\mathsf{x}, \mathsf{z}) \leq_{\mathrm{sim}} (\mathsf{y}, \mathsf{z})$, where $\leq_{\mathrm{sim}}$ is a **similarity order**, namely an ordering of mappings based on their internal similarity. Tesar (2014) calls **output-driven** a categorical grammar G which satisfies the implication (11): whenever G maps the less similar UR x to z, it also maps the more similar UR y to z. In other words, an output-driven grammar yields no saltations $\mathsf{x} \curvearrowright \mathsf{y} \curvearrowright \mathsf{z}$. Furthermore, output-drivenness entails idempotency, as (4) is a special case of (11) where the UR y more similar to z is the most similar one, namely y itself. Hence, an output-driven grammar also yields no chain shifts. Since opacity means counter-feeding or counter-bleeding which in turn mean chain shifts or saltations, output-drivenness generalizes transparency beyond rule-based phonology.

$$(11)\quad G(\mathsf{x}) = \mathsf{z} \Longrightarrow G(\mathsf{y}) = \mathsf{z}$$

□ In order to make progress on the theory of output-drivenness, we need to define the similarity order $\leq_{\mathrm{sim}}$. Magri (2018b) shows that Tesar's (2014) concrete definition of $\leq_{\mathrm{sim}}$ can be axiomatized as follows: $(\mathsf{x}, \mathsf{z}) \leq_{\mathrm{sim}} (\mathsf{y}, \mathsf{z})$ holds provided each faithfulness constraint F satisfies (12). This inequality says that the less similar mapping (x, z) "makes up" not only for every faithfulness violation of the more similar mapping (y, z) but also for every discrepancy between the two forms x and y which play the role of the two URs. With this definition of $\leq_{\mathrm{sim}}$, Magri (2018) shows that OT and HG grammars are output-driven provided again each faithfulness constraint F satisfies the TI (5).

$$(12)\quad F(\mathsf{x}, \mathsf{z}) \geq F(\mathsf{y}, \mathsf{z}) + F(\mathsf{x}, \mathsf{y})$$

□ Again following Anttila and Magri (2018), I generalize the categorical output-drivenness entailment (11) to the probabilistic ME setting through the condition (13) that the probability $G^{\mathrm{ME}}(\mathsf{z} \,|\, \mathsf{y})$ of the consequent more similar mapping (y, z) is not smaller than the probability $G^{\mathrm{ME}}(\mathsf{z} \,|\, \mathsf{x})$ of the antecedent less similar mapping (x, z). My fourth contribution is that the faithfulness TI (5) also suffices to ensure that ME satisfies (13) and is thus output-driven. The proof is omitted for space, but similar to the proof of my second contribution. In conclusion, OT, HG, and ME are transparent under the same condition: that faithfulness constraints measure phonological distance in compliance with the metric TI (5).

$$(13)\quad G^{\mathrm{ME}}(\mathsf{z} \,|\, \mathsf{x}) \leq G^{\mathrm{ME}}(\mathsf{z} \,|\, \mathsf{y})$$