MaxEnt fails at reasoning by transitivity

Arto Anttila (Stanford University) and Giorgio Magri (CNRS, University of Paris 8, SFL)

OT, HG, and Stochastic or Noisy HG (SHG) support reasoning by transitivity: if a constraint A dominates/outweights B, and B dominates/outweights C, then A dominates/outweights C. We present a general result about implicational universals of Maximum Entropy (ME) typologies and derive from it the new observation that ME does not support reasoning by transitivity.

Universals. Following Anttila & Magri (2017), we say that a ME or SHG typology satisfies the implicational universal $(\mathsf{x},\mathsf{y}) \to (\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ provided that any grammar in the typology, corresponding to any non-negative weights, realizes the consequent underlying form $\widehat{\mathsf{x}}$ as the consequent winner $\widehat{\mathsf{y}}$ with a probability that is at least as large as the probability with which the same grammar realizes the antecedent underlying form x as the antecedent winner y. For instance, the implication $(/\mathsf{cost\#us}/, [\mathsf{cos.us}]) \to (/\mathsf{cost\#me}/, [\mathsf{cos.me}])$ is a universal of the typology of English dialects because t-deletion is always at least as probable before consonants as before vowels.

Main result. Given an implication $(\mathsf{x},\mathsf{y}) \to (\widehat{\mathsf{x}},\widehat{\mathsf{y}})$, we say that a constraint C is *even* provided it does not distinguish between the antecedent and consequent winners y and $\widehat{\mathsf{y}}$ because it assigns them the same number of violations $C(\mathsf{x},\mathsf{y}) = C(\widehat{\mathsf{x}},\widehat{\mathsf{y}})$. Furthermore, we denote by z and $\widehat{\mathsf{z}}$ the generic antecedent and consequent losers, namely any candidates of the antecedent and consequent underlying forms x and $\widehat{\mathsf{x}}$ different from the antecedent and consequent winners y and $\widehat{\mathsf{y}}$. The main result of this paper is stated in the box below. It provides a necessary constraint condition for the implication $(\mathsf{x},\mathsf{y}) \to (\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ to qualify as a ME universal. This constraint condition captures the intuition that the antecedent losers z are at least as good as the consequent losers $\widehat{\mathsf{z}}$, whereby the share of the probability mass that the antecedent losers z leave for the antecedent winner y cannot be larger than the share that the consequent losers $\widehat{\mathsf{z}}$ leave for the consequent winner $\widehat{\mathsf{y}}$.

> *Suppose that an implication $(\mathsf{x},\mathsf{y}) \to (\widehat{\mathsf{x}},\widehat{\mathsf{y}})$ is a ME universal. Then, for every consequent loser $\widehat{\mathsf{z}}$ and for every constraint $\widehat{C}_{even}$ that is even between the antecedent and consequent winners y and $\widehat{\mathsf{y}}$, there ought to exist some antecedent loser z that "is at least as good" in the sense that:* ***(A)*** *z violates the even constraint $\widehat{C}_{even}$ at most as much as $\widehat{\mathsf{z}}$: $\widehat{C}_{even}(\mathsf{x},\mathsf{z}) \leq \widehat{C}_{even}(\widehat{\mathsf{x}},\widehat{\mathsf{z}})$;* ***(B)*** *$\mathsf{z}$ satisfies every other even constraint that is satisfied by $\widehat{\mathsf{z}}$.*

An example. To understand the implications of this result, we focus on the two underlying forms /kyg/ and /tag/. They are listed by *Gen* with the 128 candidate surface realizations obtained by changing underlying stop voicing, aspiration, and place (between velar and coronal) as well as vowel rounding in all possible ways. The constraint set consists of three families of constraints. Constraints for place: *Dorsal (penalizes non-coronal place) and IdentPlace. Constraints for vowel rounding: *RoFro (penalizes front rounded vowels) and IdentRound. Laryngeal constraints: AspOns (requires voiceless stops to be aspirated in onsets); *VoicedCoda (penalizes voiced codas); IdentVOT (penalizes corresponding segments with different VOTs; Flemming 2001). The latter broad faithfulness constraint that penalizes discrepancies in both voicing and aspiration simultaneously is crucial for our specific example. Yet, our general point on transitivity in ME is independent of any assumptions about the phonetics/phonology interface.

The OT and HG typologies corresponding to these candidates and constraints satisfy the implicational universal $(/\mathsf{kyg}/, [\mathsf{k^hyg}]) \to (/\mathsf{tag}/, [\mathsf{t^hag}])$: every OT or HG grammar that contains the antecedent mapping, also contains the consequent mapping. Since categorical HG and probabilistic SHG share the same implicational universals (Anttila & Magri 2017), this universal also holds in SHG. Finally, we show that this universal holds in ME as well. No matter how we choose the non-negative constraint weights, every SHG or ME grammar assigns at least as much probability to the consequent mapping as to the antecedent mapping. This generalization makes good intuitive sense given the constraints employed. Faithfulness-wise, the antecedent and consequent mappings are on equal standing (both violate IdentVOT once to satisfy AspOns). Markedness-wise, the antecedent winner $[\mathsf{k^hyg}]$ is worse than the consequent winner $[\mathsf{t^hag}]$ because of the marked front rounded vowel (penalized by *RoFro) and the additional marked velar (penalized by *Dorsal).

Now we add the constraint OCP that penalizes surface words that contain two stops that are too similar because they agree in both place and laryngeal specifications (Berkley 2000; McCarthy 1994, Frisch *et al.* 2004; Coetzee and Pater 2008). Obviously, the addition of this OCP constraint does not compromise the universal $(/\mathsf{kyg}/, [\mathsf{k^hyg}]) \to (/\mathsf{tag}/, [\mathsf{t^hag}])$ in categorical OT and HG. The

same holds for probabilistic SHG, again because HG and SHG share the same universals. The situation is dramatically different in ME. As soon as the OCP constraint is added, this universal (/kyg/, [kʰyg]) → (/tag/, [tʰag]) fails in ME: we can construct non-negative constraint weights such that the ME probability of the antecedent mapping is larger than the ME probability of the consequent mapping. We now use the result boxed above to make sense of this puzzling observation.

Reasoning by transitivity holds in OT. We only focus on the constraints and the candidates described in ERC notation (Prince 2002) in Fig. 1. We represent each candidate as a sequence of two squares, with the convention that the first (second) square is filled if AspOns (*VoiCoda, respectively) is violated. In order for the antecedent winner [kʰyg] □■ to beat the antecedent loser [kyg] ■■ in OT, AspOns must outrank IdentVOT. In order for the antecedent winner [kʰyg] □■ to also beat the antecedent loser [kʰyk] □□, Ident must in turn outrank *VoiCoda. By transitivity, AspOns must outrank *VoiCoda. The latter ranking AspOns ≫ *VoiCoda ensures that the consequent winner [tʰag] □■ beats the consequent loser [tak] ■□. Without OCP, this ranking AspOns ≫ *VoiCoda does not need to be inferred by transitivity because it is required in order for the antecedent winner [kʰyg] □■ to beat the antecedent loser [kyk] ■□. With OCP instead, the latter antecedent loser provides no evidence for this ranking AspOns ≫ *VoiCoda, that must therefore be inferred by transitivity. We conclude that, when OCP is added, the implication (/kyg/, [kʰyg]) → (/tag/, [tʰag]) is an OT universal that holds *by transitivity*.

☞[kʰyg] □■	AspOns	Ident	*VoiCo	OCP
[kyg] ■■	W	L		
[kʰyk] □□		W	L	
[kyk] ■□	W		L	W

☞[tʰag] □■	AspOns	Ident	*VoiCo	OCP
[tak] ■□	W		L	

Figure 1

Reasoning by transitivity holds in HG. Analogous considerations hold in HG. Figure 2 replaces ERCs with actual numbers of violations. In order for the antecedent winner [kʰyg] □■ to beat the antecedent loser [kyg] ■■ in HG, the weight of AspOns must be larger than that of Ident. In order for the antecedent winner [kʰyg] □■ to also beat the antecedent loser [kʰyk] □□, the weight of Ident must in turn be larger than that of *VoiCoda. By transitivity, the weight of AspOns must be larger than the weight of *VoiCoda. The latter weighting $w_{\text{AspOns}} > w_{*\text{VoiCoda}}$ ensures that the consequent winner [tʰag] □■ beats the consequent loser [tak] ■□. Without OCP, this weighting $w_{\text{AspOns}} > w_{*\text{VoiCoda}}$ does not need to be inferred by transitivity because it is required in order for the antecedent winner [kʰyg] □■ to beat the antecedent loser [kyk] ■□. With OCP instead, the latter antecedent loser provides no evidence for this weighting $w_{\text{AspOns}} > w_{*\text{VoiCoda}}$, that must therefore be inferred by transitivity. We conclude that, when OCP is added, the implication (/kyg/, [kʰyg]) → (/tag/, [tʰag]) is an HG universal that holds *by transitivity*.

/kyg/ ■■	AspOns	Ident	*VoiCo	OCP
☞[kʰyg] □■		1	1	
[kyg] ■■	1		1	
[kʰyk] □□		2		
[kyk] ■□	1	1		1

/tag/ ■■	AspOns	Ident	*VoiCo	OCP
☞[tʰag] □■		1	1	
[tak] ■□	1	1		

Figure 2

Reasoning by transitivity fails in ME. The four constraints in Fig. 1-2 are all even between antecedent and consequent mappings (/kyg/, [kʰyg]) and (/tag/, [tʰag]): both mappings satisfy AspOns and OCP and both violate Ident and *VoiCoda once. Let us focus on the consequent loser [tak] ■□. Is there an antecedent loser that does as well in the sense that it satisfies conditions (A) and (B) boxed above? The antecedent losers [kyg] ■■ and [kyk] ■□ flout condition (B): they violate *VoiCoda and OCP respectively while the consequent loser [tak] ■□ satisfies them both. The antecedent loser [kʰyk] □□ satisfies both *VoiCoda and OCP as required by condition (B), but flouts condition (A) because it violates Ident more (twice) than the consequent loser [tak] ■□ (only once). The boxed result thus explains why the universal (/kyg/, [kʰyg]) → (/tag/, [tʰag]) fails in ME when OCP is added. Without OCP, the consequent loser [tak] ■□ and the antecedent loser [kyk] ■□ share the same violation profile, whereby the latter satisfies both boxed conditions (A) and (B).

Conclusion. As suggested by the abstract notation □□, these considerations extend beyond the example considered. Whenever we have the ERC pattern in Fig. 1, the resulting universal holds in OT because $M_1 \gg F$ and $F \gg M_2$ entail $M_1 \gg M_2$ by transitivity. Yet, that universal always fails in ME. In conclusion, while OT, HG, and SHG support reasoning by transitivity, ME does not.