Proceedings of the Society for Computation in Linguistics

Volume 1 Article 26

2018

T-orders across categorical and probabilistic constraint-based phonology

Arto Tapani Anttila
Stanford University, anttila@stanford.edu

Giorgio Magri
CNRS, magrigrg@gmail.com

Recommended Citation

Tapani Anttila, Arto and Magri, Giorgio (2018) "T-orders across categorical and probabilistic constraint-based phonology," *Proceedings of the Society for Computation in Linguistics*: Vol. 1 , Article 26.
DOI: https://doi.org/10.7275/R5VH5M18
Available at: https://scholarworks.umass.edu/scil/vol1/iss1/26

T-orders across categorical and probabilistic constraint-based phonology

Arto Anttila (Stanford University) and Giorgio Magri (CNRS)

□ *Implicational universals* (Greenberg 1966) state that, if a language has property P, it also has property $\widehat{P}$. For instance, if a language allows a complex margin cluster, it also allows simpler clusters (e.g., CCCV→CCV→CV). Another much-discussed example comes from dialects of English: if t/d deletes before vowels (cost us ∼ cos' us), it also deletes before consonants (cost me ∼ cos' me). More formally, consider a typology $\mathfrak{T}$ of phonological grammars (construed as mappings from underlying to surface forms). The entailment $(\mathsf{x}, \mathsf{y}) \xrightarrow{\mathfrak{T}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ holds between two underlying/surface form pairs (x, y) and $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ provided each grammar in $\mathfrak{T}$ which maps x to y also maps $\widehat{\mathsf{x}}$ to $\widehat{\mathsf{y}}$. This relation $\xrightarrow{\mathfrak{T}}$ is a partial order on mappings, called the *T-order* induced by $\mathfrak{T}$ (Anttila and Andrus 2006).

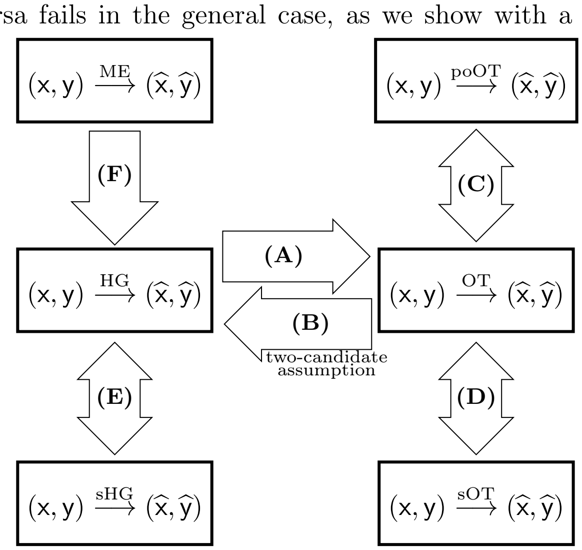
□ Implicational universals can also be statistical: if t/d deletion is variable in both pre-vocalic and pre-consonantal position, deletion is more frequent before consonants than vowels across all known dialects (Guy 1991, Kiparsky 1993, Coetzee 2004). We thus extend the notion of T-order to a typology $\mathfrak{T}$ of *probabilistic* phonological grammars (construed as functions from underlying forms to probability distributions over surface forms) as follows: the entailment $(\mathsf{x}, \mathsf{y}) \xrightarrow{\mathfrak{T}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ holds provided each grammar in $\mathfrak{T}$ assigns a probability to $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ which is at least as large as the probability it assigns to (x, y). The categorical definition is a special case of the probabilistic one as categorical grammars can be construed as probabilistic grammars assigning 0/1 probabilities.

□ T-orders are important because they impose strict limits on possible phonological patterns, including statistical patterns. In other words, they "measure" the amount of typological structure. They derive directly from the phonological theory, they need not be learned, nor can they be subverted by learning (Becker, Ketrez, and Nevins 2011). They have also been used to model categorical and gradient phonotactic judgments (Anttila 2008). This paper develops the formal theory of T-orders in constraint-based phonology. We consider T-orders in the two categorical frameworks of OT ($\xrightarrow{\text{OT}}$) and HG ($\xrightarrow{\text{HG}}$) and the four probabilistic frameworks of stochastic OT ($\xrightarrow{\text{sOT}}$), stochastic (or noisy) HG ($\xrightarrow{\text{sHG}}$), partial order OT ($\xrightarrow{\text{poOT}}$), and Max Ent ($\xrightarrow{\text{ME}}$). We address two questions:

Q1: *what are the relationships among T-orders in these six frameworks?* This question is crucial because one of the main virtues of T-orders is indeed that they allow for cross-framework comparisons of typological structure, even bridging across categorical and probabilistic frameworks.

Q2: *for each of these six frameworks, how can we express the T-order $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ in terms of just the constraint violation profiles of (x, y) and $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ and their loser candidates?* These constraint conditions would be important to help us understand what it really means that a T-order holds. Furthermore, they would allow us to compute T-orders without having to compute the entire typology, which is difficult in the categorical case (the typology can be large) and impossible in the probabilistic case (the typology of probability distributions is infinite!).

□ We give a **complete answer** to **Q1**, summarized in the diagram below. The HG typology is usually at least as large as the OT typology. Thus, a HG T-order $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{HG}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ entails the OT T-order $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{OT}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ (arrow A). The vice versa fails in the general case, as we show with a simple example. Yet, our **first result** is that the vice versa *surprisingly* does hold when both x and $\widehat{\mathsf{x}}$ have only two candidates each (arrow B). This result is significant because applications of T-orders often limit themselves to two candidates. For example, analyses of free variation often have two viable contenders, e.g., the presence vs. absence of t/d in English, front vs. back vowel harmony in Finnish (Ringen and Heinämäki 1999), and voiced vs. devoiced geminates in Japanese (Pater 2009, Coetzee and Kawahara 2013). Indeed, Pater shows that optional devoicing in Japanese loanwords (dog:u ∼ dok:u 'dog') requires constraint ganging and thus constitutes evidence against OT, for HG. Surprisingly, our result says that the OT and



Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 201-202.
Salt Lake City, Utah, January 4-7, 2018

HG T-orders predicted by his analysis are provably identical given that his analysis satisfies the two-candidate assumption. In other words, our result provides further evidence for Pater's claim that the HG typological structure is often in practice only mildly richer than the OT structure.

□ Our **second result** is that the T-orders in stochastic OT (Boersma 1998) and stochastic HG (Boersma and Pater to app.) are identical to the T-orders in categorical OT and HG (arrows E-D). Despite the stochastic typologies being infinite (they contain an infinite number of probability distributions) while the categorical typologies being finite, they provide exactly the same typological structure. Here is the intuition behind our result for OT (same reasoning holds for HG). A ranking vector assigns a probability to each ranking. The probability assigned by stochastic OT to a mapping is the sum of the probabilities of the rankings consistent with that mapping. Thus, the probability assigned by stochastic OT to $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ is at least as large as that assigned to (x, y) for every ranking vector (as required by $\xrightarrow{\text{sOT}}$) iff every ranking consistent with (x, y) is also consistent with $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ (as required by $\xrightarrow{\text{OT}}$). The result for stochastic OT extends to partial order OT (Anttila 1998; Anttila and Cho 1998), as the latter is a special case of stochastic OT (arrow C).

□ Our **third result** is that a T-order $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{ME}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ in MaxEnt entails the T-orders $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{HG}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ and $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{sHG}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ in categorical and stochastic HG (arrow F). The vice versa is shown to fail even when there is a unique constraint and both x and $\widehat{\mathsf{x}}$ have a unique loser candidate! Thus, although MaxEnt and stochastic HG both look like "innocuous" probabilistic variants of HG (they both depend on weighted averages of constraint violations), the typological structure imposed by MaxEnt (as quantified through T-orders) is substantially weaker than that imposed by stochastic HG, which indeed was shown above to coincide with the typological structure of categorical HG.

□ Turning to **Q2**, because of the equivalences established above, we only need constraint conditions for T-orders in (categorical) HG and OT and in MaxEnt. To this end, we call *antecedent difference vector* a vector with a component for each constraint C and that component is defined as the difference $C(\mathsf{x}, \mathsf{z}) - C(\mathsf{x}, \mathsf{y})$, where (x, y) is the antecedent of the T-order $(\mathsf{x}, \mathsf{y}) \rightarrow (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ considered, (x, z) is one of its loser candidates, and $C(\mathsf{x}, \mathsf{z})$ and $C(\mathsf{x}, \mathsf{y})$ are the number of violations assigned to them by that constraint C. The *consequent difference vectors* are defined analogously, as pitting the winner mapping $(\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ in the consequent of the T-order against one of its loser mappings $(\widehat{\mathsf{x}}, \widehat{\mathsf{z}})$.

□ When there are only two constraints, the antecedent difference vectors can be plotted as the points in the plane of Fig. 1. The dark-gray region is their corresponding *cone*. The points which are larger (component-wise) than some point in this cone yield the light-gray region. Our **fourth result** is that the HG T-order $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{HG}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ holds (for any number of constraints) iff each consequent difference vector lives in this light-gray region, namely it is larger than some vector in the cone generated by the antecedent difference vectors. We derive this result from the *Farkas-Minkowski theorem* of convex geometry (Boyd and Vandenberghe 2004). While the definition of the HG T-order $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{HG}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ is *expensive* to check directly (because it involves *universal* quantification over weights), the latter geometric characterization is *easy* to check (because it admits an algebraic formulation which involves *existential* quantification over weights). We thus release the first provably efficient algorithm (coded in `Python`) to establish HG T-orders. The geometric characterization and the corresponding algorithm extend from HG to OT through the HG-to-OT-portability result in Magri (2013).

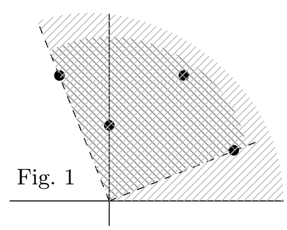


Fig. 1

□ The dark-gray region in Fig. 2 is the *polyhedron* generated by the antecedent difference vectors (represented by the black dots). The points which are larger than some point in this polyhedron yield the light-gray region. Our **fifth result** is that a necessary condition for the ME T-order $(\mathsf{x}, \mathsf{y}) \xrightarrow{\text{ME}} (\widehat{\mathsf{x}}, \widehat{\mathsf{y}})$ (for any number of constraints) is that each consequent difference vector lives in this light gray region, namely it is larger than some vector in the polyhedron generated by the antecedent difference vectors. This condition is also sufficient when both x and $\widehat{\mathsf{x}}$ have at most three candidates. For a larger number of candidates, we derive a more involved sufficient condition which is nonetheless stronger than the necessary condition. The task of closing the gap between necessary and sufficient conditions is left for future research. Nonetheless, we will show that the formal conditions obtained suffice to completely characterize ME T-orders in a number of realistic test cases taken from the literature.

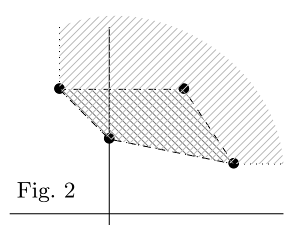


Fig. 2