Phonology **33** (2016). Supplementary materials

Computationally, tone is different

Adam Jardine
University of Delaware

Supplementary materials

Mathematical definitions and proof

1 Notation

Basic knowledge of set theory is assumed. An alphabet is a finite set of symbols; if Σ is an alphabet, let Σ^* denote the set of all finite strings, including the empty string λ, over Σ. Let $|w|$ denote the length of string w. If w and u are strings, let wu denote their concatenation. If w is a string and X is a set of strings, then let wX denote the set of strings resulting from concatenating w to each string in X. The PREFIXES of a string $w \in \Sigma^*$ are $Pr(w) = \{u \in \Sigma^* \mid \exists v \in \Sigma^* \text{ such that } w = uv\}$. The prefixes of a set of strings $L \subseteq \Sigma^*$ are $Pr_{set}(L) = w \in \cap_{x \in L} Pr(x)$.

The LONGEST COMMON PREFIX (lcp) of a set of strings is the longest prefix shared by all strings in the set: $lcp(L) = w$ such that $w \in Pr_{set}(L)$ and $\forall w' \in Pr_{set}(L)$, $|w'| \leq |w|$. For example, $lcp(\{aaa, aab\}) = aa$, because aa is the longest prefix shared by both aaa and aab.

If Σ and Δ are alphabets, a RELATION is some subset of $\Sigma^* \times \Delta^*$. A relation R is a MAP (or FUNCTION) iff for all $w \in \Sigma^*$, $(w, v), (w, v') \in R$ implies $v = v'$.

The TAILS of x in given a relation R, denoted $T_R(x)$, are $T_R(x) = \{(y, v) \mid t(xy) = uv, u = lcp(t(x\Sigma^*))\}$. If R is a map, it is a SUBSEQUENTIAL map iff its sets of tails are finite; that is, the set $\cup_{w \in \Sigma^*}\{T_R(w)\}$ is of finite cardinality.

2 Subsequential finite-state transducers

A finite-state transducer (FST) is a sextuple $(q_i, F, Q, \Sigma, \Delta, \delta)$, where Q is the finite set of states, $q_i \in Q$ is the initial state, $F \subseteq Q$ is the set of final states and $\delta \subseteq Q \times \Sigma^* \times \Delta^* \times Q$ is the transition function. The recursive extension of the transition function δ^* is defined as:

- $\delta \subseteq \delta^*$
- $(q, \lambda, \lambda, q) \in \delta^*$ for all $q \in Q$
- $(q, x, y, r) \in \delta^*$ and $(r, a, b, s) \in \delta$ implies $(q, xa, yb, s) \in \delta^*$

The relation that a FST describes is defined as $R(t) = \{(x, y) \in \Sigma^* \times \Delta^* \mid \exists q_f \in F$ such that $(q_i, x, y, q_f) \in \delta^*\}$.

A FST is DETERMINISTIC iff $\forall q \in Q$ and for all $\sigma \in \Sigma$, (q, σ, v, r), $(q, \sigma, v', r') \in \delta$ implies $v = v'$ and $r = r'$. SUBSEQUENTIAL FSTs (SFSTs) are deterministic FSTs with an added output function $\omega : Q \to \Delta^*$, which specifies for each state an output string to be written when the machine ends on that state. Thus, a SFST is a septuple $(q_i, F, Q, \Sigma, \Delta, \delta, \omega)$. The relation that a SFST describes is defined as $R(t) = \{(x, yz) \in \Sigma^* \times \Delta^*$: $\exists q_f \in F$ such that $(q_i, x, y, q_f) \in \delta^*$ and $\omega(q_f) = z\}$.

Theorem 1 states an important relationship between SFSTs and subsequential maps.

THEOREM 1 (Mohri 1997)
A relation R is a subsequential map iff it is recognised by a SFST for which each state in the machine corresponds to a set of tails in R.

3 Proof that UTP is not subsequential

The proof is exactly Heinz & Lai (2013)'s proof for the non-subsequentiality of sour grapes.

Proof. Let *UTP* be the map discussed in the main text. The following shows that for all distinct $n, m \in \mathbb{N}$, $T_{UTP}(\text{HØ}^m) \neq T_{UTP}(\text{HØ}^n)$. As $\mathbb{N}$ is infinite, this means there must be infinitely many states in the canonical SFST for *UTP*, which by Theorem 1 means there is no SFST describing it.

If $x = \text{HØ}$, $lcp(UTP(x\Sigma^*)) = \text{H}$, because $UTP(x\Sigma^*)$ includes both HLL (which is equal to $UTP(\text{HØØ})$) and HHH (which is equal to $UTP(\text{HØH})$) and thus there is no shared prefix of $UTP(x\Sigma^*)$ longer than H. Thus for all $n \neq 2$, $(\text{Ø},\text{L}^n) \notin T_{UTP}(\text{HØ})$, i.e. (Ø,LL) is the only possible tail with Ø as the first member of the tuple.

If $x = \text{HØØ}$, $lcp(UTP(x\Sigma^*)) = \text{H}$, because $UTP(x\Sigma^*)$ includes both HLLL and HHHH. Thus for all $n \neq 3$, $(\text{Ø},\text{L}^n) \notin T_{UTP}(\text{HØ})$, i.e. (Ø,LLL) is the only possible tail with Ø as the first member of the tuple.

We can see then that for any distinct $n \in \mathbb{N}$, $(\text{Ø},\text{L}^k) \in T_{UTP}(\text{HØ}^n)$ only if $k = n + 1$. For $m \in \mathbb{N}$, $m \neq n$, $(\text{Ø},\text{L}^j) \in T_{UTP}(\text{HØ}^m)$ only if $j = m + 1$. Thus for all distinct n and m, $k \neq j$, and so $T_{UTP}(\text{HØ}^n) \neq T_{UTP}(\text{HØ}^m)$.