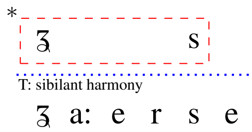
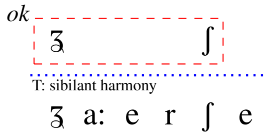
A Game of Tiers: Exploring the Formal Properties of TSL Languages

Keywords: phonology, computational linguistics, subregular hypothesis, formal language theory

Overview Following the attempt to use formal language theory to describe the complexity of linguistics processes, it has been recently suggested that unbounded dependencies in phonotactics, morphology, and even syntax can all be captured by the class of Tier-based Strictly Local languages (TSL). Here I present patterns that have been reported to be problematic for this approach, and discuss how we can account for them by small changes to TSL formal mechanisms. Exploring the resulting new classes, I also show how solutions adopted in the literature to capture problematic phonotactic dependencies give us more expressive power than what is desirable.

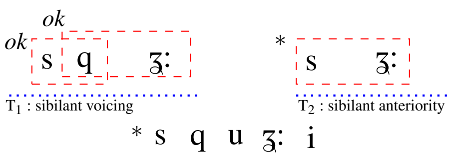
TSL Phonology Many dependencies in phonology can be captured by *local constraints* that only make distinctions on the basis of contiguous subsequences of segments up to some length *k*. For example, a (*k*=2) local dependency requiring /s/ to surface as [z] when followed by [l] can be captured by a grammar that only contains the sequence *zl* (or, alternatively, the negative constraint *sl). Dependencies that are not locally bounded do not fit this pattern. Heinz *et al.* (2011) argue that long-distance dependencies are *tier-based strictly local*: a tier is defined as the projection of a subset of the segments of the input string, and the grammar constraints act only over that subset. For instance, the example below (from AARI, an Omotic language of south Ethiopia) shows how to enforce long-distance sibilant harmony (in anteriority) by projecting a tier *T* only containing sibilants from the input, and ban contiguous ʒs and sʒ on *T*.

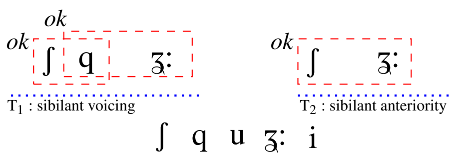




Limitations of TSL (1) McMullin (2016) discusses several languages that struggle to be captured by the way TSL currently combines a projection mechanism and strictly local dependencies. For example, he presents a case of long-distance Sibilant Harmony in IMDLAWN TSHLHIYT, where sibilants agree both in voicing and anteriority, unless there is an intervening voiceless obstruents somewhere in the string. In that case, agreement in voicing (but not in anteriority) is blocked. This pattern cannot be described by a single TSL grammar: to block agreement in voicing, we would need to project a tier of sibilants *and* obstruents. Once such tier has been built though, we cannot stop obstruents from also blocking the agreement in anteriority. In the literature, this problem has been resolved via the conjunction of distinct TSL grammars. However, it can be shown via formal proof that this operation is not well-defined for TSL: it is unclear what expressive power we get by combining the constraints of multiple TSL grammars.

Multiple-Tier SL Even leaving complex patterns aside, intuitively we want our system to generate strings that conform to all the constraints of the natural language we are describing. To account for this, and cover patterns like the ones in McMullin (2016), I introduce *Multiple Tier Strictly Local (MTSL)* languages, a generalization of TSL that allows us to enforce multiple Tier-based constraints at the same time, while providing a better formal characterization of the conjunction operation. Informally, MTSL can be defined as TSL languages projecting over multiple tiers and enforcing a – potentially different – set of constraints for each tier. Then, a string is well-formed for the language iff every substring on each projected tier is well-formed.

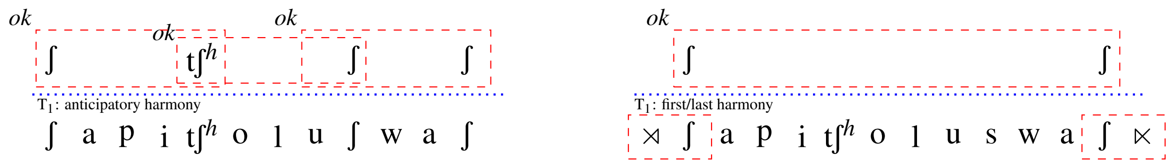




The above example from IMDLAWN TSHLHIYT shows how – by projecting a separate tier for each harmony type – the MTSL grammar correctly rules out the ill-formed **[squʒːi]** (on the left), where voicing harmony is blocked on T_1 but anteriority is violated on T_2, while ruling in the well-formed **[ʃquʒ:i]** (on the right).

Limitations of TSL (2) The MTSL class gives us a way to cope with phonotactic processes that require local constraints over different sets of segment. However, there are patterns for which the problem lies in not projecting enough information onto a tier. Baek (2016) reports cases of unbounded stress in EASTERN CHEREMIS and DONGOLESE NUMBIAN, that seem to need tier-projection of more than just substring segments. In these languages, the primary stress falls on the right-most non-final heavy syllable or - if there is none - on the initial syllable. Baek shows that no TSL grammar is able to generate well-formed strings like, for instance, **Ĺ**LLLH, while also ruling out ill-formed strings like **Ĺ**LLLHH. In order to account for these patterns, she proposes a variant of TSL that can project structural features (e.g. initial/final syllable).

Structure-Sensitive TSL Instead of projecting bundles of features encoding structural information, I show that this kind of expressivity can be accomplished by increasing the locality window of the projection mechanism. For example, we can allow the grammar to put an element on the tier iff it is in an initial/final position in the input string. While the TSL projection mechanism is based on each segment's individual (also, 1-local) properties, this new class of grammars – *Structure-Sensitive TSL (SS-TLS)* – relies on structural relationships between segments in the input string to choose what to project. This adjustment captures Baek's unbounded stress patters, and other processes reported as problematic for TSL (e.g. culminativity, cf. Heinz (2014)).



In the example above, a TSL grammar for anticipatory harmony in SAMALA (on the left) is extended to SS-TSL (Pseudo-SAMALA first-last harmony, on the right). The grammar wants to project sibilants on a tier, but can do it iff the sibilant segment is in an initial/final position in the input string. The explicitly defined local sensitivity of SS-TSL then, shows us that by including structural information on a tier – as suggested by Baek (2016) – we allow for the generation of undesired patterns. By providing a formal understanding of the structural-projection idea, I explore this increase in generative power and the relationship between the TSL, MTSL and SS-TSL classes. In the future, it would be interesting to see whether some combination of these classes can cover all desired patterns while avoiding the unnatural ones.

Conclusion The last decade has provided support for the subregular hypothesis as a strong computational theory of language complexity. Since a growing body of literature is exploring the TSL boundary from a typological – as well as experimental – point of view, here I discuss the generative power of classes obtained by extending TSL via minor modifications to its projection mechanism. These new classes offer different ways to account for controversial patterns in the literature, and a careful understanding of their formal relationship to TSL will help the search for the right subregular class for phonotactics.

Baek, Hyunah. 2016. *Unbounded Stress: TSL with structural features*. Ms., Stony Brook University. • **Heinz, Jeffrey**. 2014. Culminativity times harmony equals unbounded stress. In *Word Stress: Theoretical and Typological Issues*, Chap 8. • **Heinz, Jeffrey**, Chetan Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints in phonology. In *Proceedings of ACL 49th*, 58–64. • **McMullin, Kevin**. 2016. *Tier-based locality in long-distance phonotactics: learnability and typology*. Doctoral dissertation, University of British Columbia.