The computational cost of generalizations: An example from micromorphology

Sedigheh Moradi, Alëna Aksënova and Thomas Graf

Overview Morphotactics has been argued to be limited to the formal class of tier-based strictly local languages (Aksënova et al., 2016). We claim that the level of the complexity of a pattern largely depends on the way it is morphologically analyzed. Using an example from adjectival inflection in Noon (NIGER-CONGO), we show that the complexity of this pattern falls in two different classes in the subregular hierarchy if viewed from different perspectives. In particular, the traditional segmentation of Noon affixes (Soukka 2000) yields a 3-TSL grammar, while the same pattern is 3-SSTSL under the perspective of micromorphology (Stump 2017). Both grammars require a locality window of 3 segments; however, the micromorphology-based analysis shows an increase in formal complexity, although it reduces the grammar size by defining complex affixes in terms of simpler ones.

Subregular Hierarchy Subdivisions of the regular class of formal languages into weaker classes yield a hierarchy of subregular languages (Rogers and Pullum 2011) that can efficiently capture the complexity of different parts of language (Heinz 2010; Graf and Heinz 2016). This paper focuses on 2 subclasses within this hierarchy: TSL and SS-TSL.

A *tier-based strictly local* (TSL) grammar captures dependencies by blocking or allowing substrings of a certain length. However, instead of evaluating the whole string, it evaluates the *tier* on which only selected elements (tier alphabet) are projected (Heinz and Rawal 2011). TSL grammars capture long-distance dependencies, but the interaction of local and long-distance processes requires a more complex subregular class. *Structure sensitive tier-based strictly local* (SS-TSL) languages are a proper extension of the TSL class. They project items on the tier while taking local context into account (De Santo 2018). In other words, SS-TSL languages allow for the encoding of long-distance dependencies that interact with some local restrictions.

Data Adjectives exhibit a complex pattern of inflection in Noon noun-class system (Soukka 2000; Stump 2017). The inflectional affixes studied here are the attributive prefix and the definite suffix. Indefinite adjectives agree with the noun in number via the attributive prefix. Definite adjectives, apart from acquiring the attributive prefix, express definiteness via the definite suffix.

Noun Class	*Indefinite*	*Definite*
C1	wi-yak	wi-yak-wum
C2	fi-yak	fi-yak-fum

Soukka (2000) observes that each attributive prefix consists of a class marker and a prefixal formative *-i-*, and each definite suffix consists of a class marker and a suffixal formative *-um*. Stump (2017) uses this example to characterize the micromorphology hypothesis, according to which *an affix can itself be morphologically complex*, cf. table below.

Noun Class	*Indefinite*	*Definite*
C1	w-i-yak	w-i-yak-w-um
C2	f-i-yak	f-i-yak-f-um

Analysis *Under the micromorphological perspective*, the class markers and the affixal formatives do not form a single unit. Therefore, they all need to be treated by the grammar separately. Following Aksënova et al. (2016), we assume the length of the stem in a word to be potentially unbounded to accommodate the compounding processes, and deploy the marker # in order to indicate the edges of the stem. We use the following notation: CM – class markers, PF – prefixal formative, SF – suffixal formative, RT – stem. Then, a class marker can be used in the following cases:

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 367-368.
New York City, New York, January 3-6, 2019

1. Indefinite adjectives: CM-PF-#-RT-#
2. Definite adjectives: CM-PF-#-RT-#-CM-SF

In indefinite adjectives, the only position where CM can appear is immediately before PF. However, in definite adjectives, not only is the position of the initial CM bound in the same way as in the previous case, but the second CM is also restricted to the position immediately preceding SF. Moreover, the two instances of CMs must be identical, i.e., they must agree across a potentially unbounded number of morphemes within a word. A TSL grammar cannot capture this pattern because not every PF or SF must be projected onto the tier, but only those that immediately follow CMs. This rule can be encoded by an SS-TSL grammar.

The positive SS-TSL grammar for the Noon pattern is G_{SSTSL} = (⋊-CM-PF, CM-PF-#, PF-#-#, #-#-CM, #-CM-SF, CM-SF-⋉, #-#-⋉), and it operates over the tier T = {CM, PF$^{\text{after } cm}$, SF$^{\text{after } cm}$, #}.

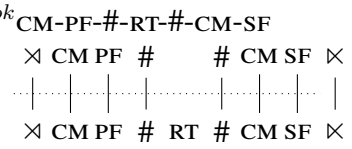

Fig. 1: SS-TSL

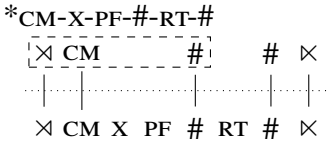

Fig. 2: SS-TSL [cont.]

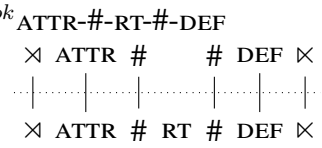

Fig. 3: TSL

The sequence CM-PF-#-RT-#-CM-SF (Fig. 1) is well-formed, because both CMs are immediately followed by PF and SF, and the resulting tier satisfies the defined grammar G_{SSTSL}. However, the sequence CM-X-PF-#-RT-# (Fig. 2), where X represents any intervening affix, is correctly ruled out: PF cannot be projected on the tier, therefore making it ill-formed.

Under the traditional perspective, the CM-PF and CM-SF sequences are single morphemes ATTR (attributive prefix) and DEF (definite suffix), respectively.

3. Indefinite adjectives: ATTR-#-RT-#
4. Definite adjectives: ATTR-#-RT-#-DEF

This pattern is easily accountable in a TSL manner, because none of the elements require structural sensitivity: viewing those two affixes as an atomic unit implicitly solves the locality challenge. The corresponding 3-local positive TSL grammar G_{TSL} = (⋊-ATTR-#, ATTR-#-#, #-#-⋉, #-#-DEF, #-DEF-⋉) operates over the tier T = {ATTR, DEF, #}, see Fig. 3.

Conclusion We discussed a morphotactic pattern that varies in its computational complexity based on how it is analyzed. If viewed from the traditional perspective, the pattern of adjectival agreement in Noon falls into the subregular class of tier-based strictly local languages. However, analyzed micromorphologically, the same pattern is computationally more complex, and needs a structure sensitive tier-based strictly local grammar in order to be captured. We are not discriminating one morphotactic approach over the other one. Instead, in this paper, we show that the encoding of the formalism largely affects its computational complexity: It is not always the case that simplifying the way basic elements are represented reduces the overall complexity of the resulting system.

Aksënova, A., **Graf**, T. and **Moradi**, S. (2016) Morphotactics as tier-based strictly local dependencies. *SIGMORPHON* 14. • **De Santo**, A. (2018) Extending TSL to account for interactions of local and non-local constraints. *SCiL* 2018. • **Graf**, T. and **Heinz**, J. (2016) Tier-based strict locality in phonology and syntax. Ms. • **Heinz**, J. (2010) Learning long-distance phonotactics. *LI* 41. • **Heinz**, J., **Rawal**, C. and **Tanner**, H. G. (2011) Tier-based strictly local constraints for phonology. *ACL 49*. • **Rogers**, J. and **Pullum**, G. (2011) Aural pattern recognition experiments and the subregular hierarchy. *JLLI* 20. • **Soukka**, M. (2000) *A descripive grammar of Noon: A Cangin language of Senegal.* • **Stump**, G. (2017) Rule conflation in an inferential realizational theory of morphotactics. *Acta Linguistica Academica* 64.