

# The computational cost of generalizations: An example from micromorphology

Sedigheh Moradi, Alëna Aksënova and Thomas Graf  
Stony Brook University



Contact Information:  
Department of Linguistics  
Stony Brook University  
Stony Brook, NY 11794-4376

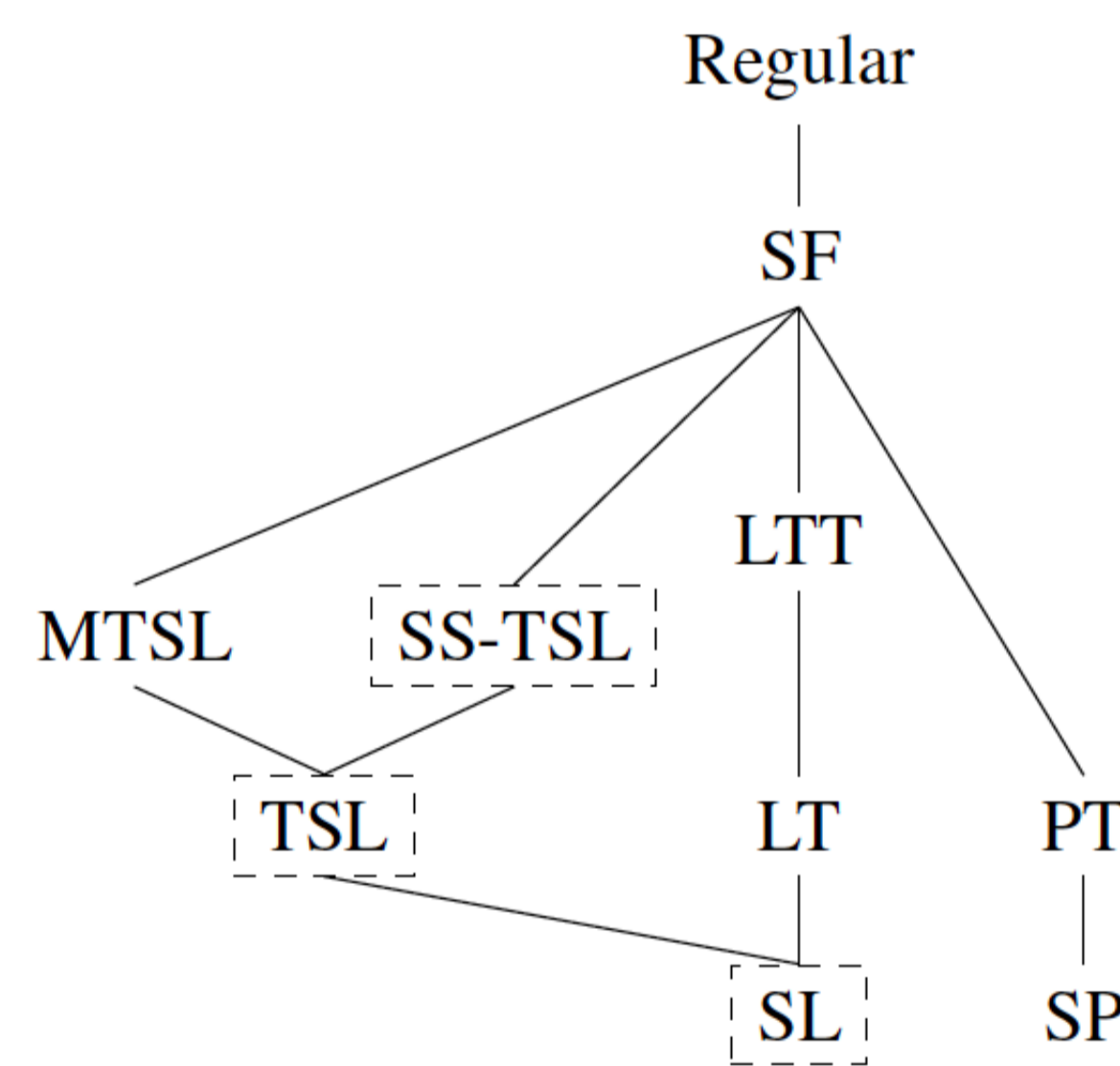
Email:  
sedigheh.moradi@stonybrook.edu  
alena.aksenova@stonybrook.edu

## Abstract

Morphotactics has been argued to be limited to the formal class of tier-based strictly local languages [1]. 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. The traditional segmentation of Noon affixes [8] yields a 3-TSL grammar, while the same pattern is 3-SS-TSL under the perspective of micromorphology [9]. 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 Languages

The formal class of regular languages can be decomposed in smaller subclasses that together form the *subregular hierarchy* [7, 5, i.a.].



## Strictly Local

Strictly local (SL) languages evaluate a string based on the  $n$ -grams it contains [6]. SL grammars capture local dependencies by **blocking or allowing substrings of a certain length**. As a result, it is not possible to capture a long-distance dependency with a SL grammar.

### SL language

Language L1:  $ab, abab, ababab$ , etc.  
Rules of language L1:

- well-formed words start with  $a$ ;
- well-formed words end with  $b$ ;
- the symbols  $a$  and  $b$  should alternate.

Tier alphabet:  $\Sigma = \{a, b\}$

Negative SL grammar:  $G_{NegSL} = (*\times ab, *aa, *bb, *a\times)$

## Tier-based SL

Tier-based strictly local (TSL) languages evaluate strings by looking for allowed or prohibited substrings while evaluating its tier [4]. Only the symbols present in the *tier alphabet*  $T$  are projected allowing the **representation of long-distance dependencies locally over the tier**.

### TSL language

Language L2:  $b, aaab, aaba, baa$

Rules of language L2:

- $a$  might be present or not;
- there must always be a single  $b$ .

Tier alphabet:  $T = \{b\}$

Positive TSL grammar:  $G_{PosTSL} = (\times b, b\times)$

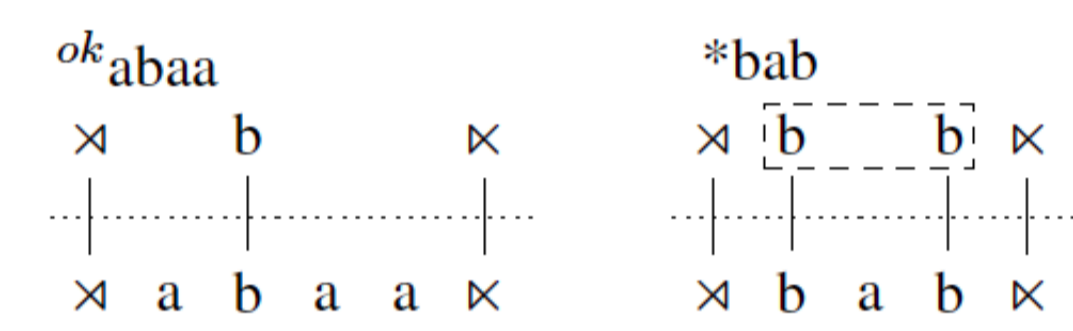


Figure 2: Examples of the TSL evaluation

## Structure Sensitive TSL

Structure sensitive tier-based strictly local grammars (SS-TSL) allow the projection of items on the tier only if they satisfy certain local condition(s) [3, 2]. SS-TSL languages **encode long-distance dependencies that interact with local restrictions**.

### SS-TSL language

Language L3:  $bnxxmxx, bxxnxxx, b$

Rules of language L3:

- $x$  and  $b$  might be present or not;
- $b$  that is immediately followed by  $n$  must be followed by  $m$ .

Tier alphabet:  $T = \{b^{before\ n}, m\}$

Positive SS-TSL grammar:  $G_{PosSSTSL} = (\times\times, \times b, bm, m\times)$

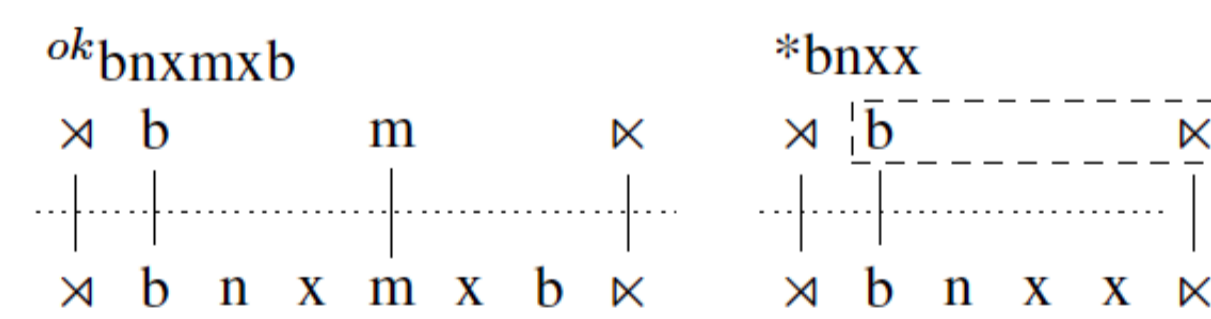


Figure 3: Examples of the SS-TSL evaluation

## Noon Adjective Inflection

The inflectional system of adjectives in Noon is very complex, but we restrict our focus on the attributive prefix and the definite suffix.

### Traditional Approach

The Noon **attributive prefix** and **definite suffix** are single morphemes regardless of their internal structures [8]. Both of these morphemes express the class of the dependent noun as well.

- (1) waas wi-yak  
road ATTR-big  
'a big road (near you)'
- (2) waas-um wi-yak-um  
road-DEF ATTR-big-DEF  
'the big road (near you)'

### Class Indefinite Definite

C1	wi-yak	wi-yak-wum
C2	fi-yak	fi-yak-fum

## Micromorphology

On the contrary, micromorphology states that a morphological unit may be morphologically complex, i.e. an affix may be a combination of other affixes [9]. According to this approach, the same adjective *wiyakwum* is segmented as shown in (3). (CM – class markers, PF – prefixal formative, SF – suffixal formative, RT – stem.)

- (3) w-i-yak-w-um  
CM-PF-RT-CM-SF  
'the big (one)'

In this case, the class markers can appear only in two positions: immediately preceding the prefixal or the suffixal formation. Moreover, **the class marker needs to be exactly the same in both positions**.

### Class Indefinite Definite

C1	w-i-yak	w-i-yak-w-um
C2	f-i-yak	f-i-yak-f-um

## Formal Analysis of Noon Pattern

Complexity of Noon pattern:

- **Micromorphology:** 3-SS-TSL  $\rightsquigarrow$  less restrictive  $\odot$ ;
- **Traditional:** 3-TSL  $\rightsquigarrow$  more restrictive  $\ominus$ .

We assume the length of the stem to be potentially unbounded, and use the marker # in order to indicate the edges of the stem.

## Micromorphological Perspective

- (4) Indefinite adjectives

CM-PF-#-RT-#

- (5) Definite adjectives

CM-PF-#-RT-#-CM-SF

### Noon pattern: 3-SS-TSL analysis

$T = \{CM, PF^{after\ cm}, SF^{after\ cm}, \#\}$

$G_{SSTSL} = (\times\text{-CM-PF}, \text{CM-PF-}\#, \text{PF-}\#\text{-}\#, \#\text{-CM}, \#\text{-CM-SF}, \text{CM-SF-}\#, \#\text{-}\times)$

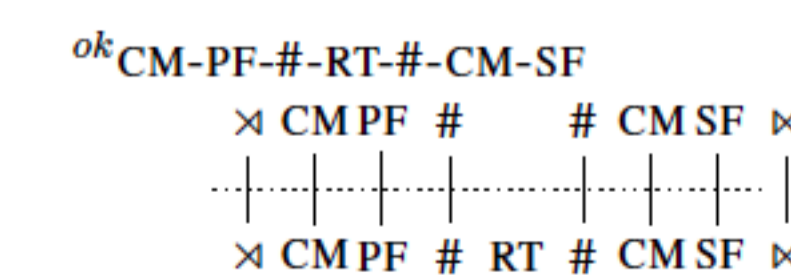


Figure 4: SS-TSL analysis of Noon adjectives

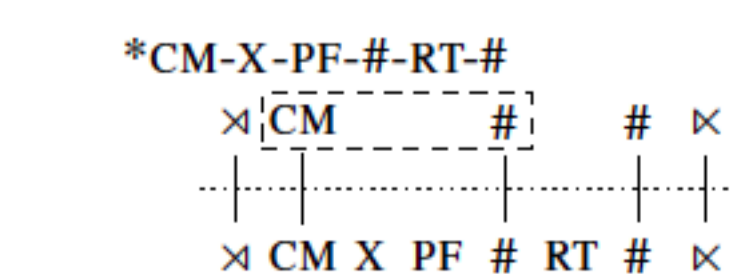


Figure 5: SS-TSL analysis of Noon adjectives [cont.]

## Traditional Perspective

Under the traditional perspective, the CM-PF and CM-SF sequences are single morphemes ATTR and DEF, respectively.

- (6) Indefinite adjectives

ATTR-#-RT-#

- (7) Definite adjectives

ATTR-#-RT-#-DEF

### Noon pattern: 3-TSL analysis

$T = \{ATTR, DEF, \#\}$

$G_{TSL} = (\times\text{-ATTR-}\#, \text{ATTR-}\#\text{-}\#, \#\text{-}\times, \#\text{-}\#\text{-DEF}, \#\text{-DEF-}\times)$

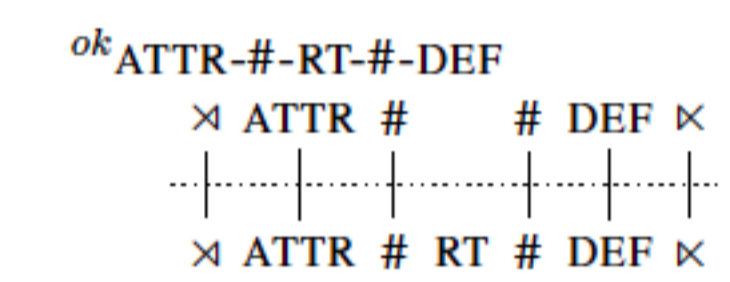


Figure 6: TSL analysis of Noon adjectives

## Conclusion

Noon inflectional morphotactic pattern varies in its computational complexity based on how it is analyzed.

- **Traditionally**, it falls into the subregular class of TSL languages.

- **Micromorphologically**, the same pattern is computationally more complex, and needs a **SS-TSL** grammar in order to be captured.

We are not discriminating one morphotactic approach over the other. Instead, 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**.

## References

- [1] Alëna Aksënova, Thomas Graf, and Sedigheh Moradi. Morphotactics as tier-based strictly local dependencies. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 121–130. Association for Computational Linguistics, 2016.
- [2] Aniello De Santo. Extending tsl to account for interactions of local and non-local constraints, 2018. Poster presented at the Society for Computation in Linguistics (SCIL) 2018, Salt Lake City, Utah.
- [3] Aniello De Santo and Thomas Graf. Structure sensitive tier projection: Applications and formal properties, 2017. Manuscript. Stony Brook University.
- [4] Jeffrey Heinz, Chetan Rawal, and Herbert G. Tanner. Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 58–64, Portland, USA, 2011. Association for Computational Linguistics.
- [5] Robert McNaughton and Seymour A. Papert. *Counter-Free Automata (M.I.T. Research Monograph No. 65)*. The MIT Press, 1971.
- [6] James Rogers and Geoffrey Pullum. Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information*, 20:329–342, 2011.
- [7] Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
- [8] Maria Soukka. *A descriptive grammar of Noon: A Cangin language of Senegal*. LINCOM Europa, Munich, 2000.
- [9] Greg Stump. Rule conflation in an inferential-realizational theory of morphotactics. *Acta Linguistica Academica*, 64(1):79–124, 2017.