Strict Locality in Morphological Derivations

Keywords: morphology, computational linguistics, subregular languages

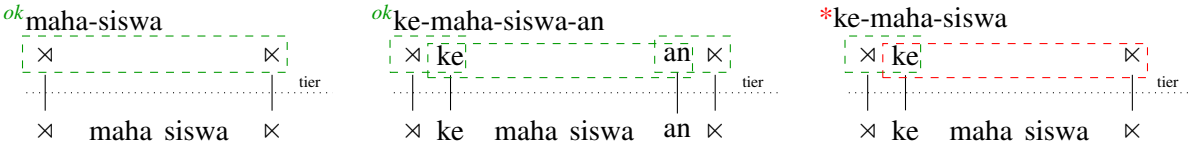
Overview It is well-known that the computational complexity of dependencies in phonology and morphology fits in the class of regular languages (Kaplan and Key 1994, Beesley and Kartunnen 2003). Recently, it has been suggested that most of the phonological patterns occurring in natural language do not need the full power of regular languages, but can in fact be captured by classes in the *subregular hierarchy* (Heinz and Idsardi 2013). Similarly, Chandlee (2014) has claimed that morphological mappings can be analyzed as subregular functions, and Aksënova et al. (2016) have argued that morphotactics does not require more power than phonology. In this paper, we present a pattern posing a problem for a subregular account of morphology. We then argue for a derivational analysis of morphological dependencies, and show how this problematic case can indeed be reduced to a subregular process, if considered over the sequence of morphological operations.

(Tier-Based) Strictly Local Morphotactics Particularly relevant to the following discussion are the classes of *Strictly Local* (SL) and *Tier-based Strictly Local* (TSL) languages. SL grammars enforce local dependencies by allowing (or banning) certain substrings listed in the grammar. For example, a SL process in English is the addition of the progressive *-ing* suffix (cf. Chandlee 2014), which is only allowed immediately after the verbal stem. The simplified SL grammar capturing this pattern then permits only strings where *-ing* is immediately preceded by a verb, and might be followed by a word-final marker ⋉ (*running*), plural affix *-s* (*meetings*), or adverbial suffix *-ly* (*interestingly*):

1. G_{ing} = {VERB-PROG, PROG-⋉, PROG-PL, PROG-ADV, ...}

Aksënova et al. (2016) argue that many morphotactic dependencies are at most TSL: a *tier* is defined as the projection of a subset of the segments of the input string, and the grammar permits or blocks certain substrings only over the tier representation of the original string. As an example, consider affixation in Indonesian: the circumfix *ke-an* derives abstract nouns from other stems (*tinggi* 'high' → *ke-tinggi-an* 'altitude'), and the amount of material in-between its two parts can be unbounded because of productive compounding in Indonesian (*maha-siswa* 'big pupil, student' → *ke-maha-siswa-an* 'student affairs'). Since *ke-* and *-an* might be arbitrarily far from each other, it is not possible to force them to only occur simultaneously using solely SL constraints. This long-distance dependency can be captured instead by projecting *ke-* and *-an* on a tier. Then, the TSL grammar allows the tier to either be empty (⋊⋉), or contain *ke-* and *-an* together, while ruling out everything else (cf. 2).

2. G_{ke-an} = {⋊⋉, ⋊-KE, KE-AN, AN-⋉}



Russian Nominalization However, there are examples that do not fit the picture outilined above. Consider the Russian nominalization suffix *-nie*, which can attach only to atelic verbs (cf. 3 vs 4).

3. umira-tʲ 'to be dying'	umira-nie 'dying'	4. umere-tʲ 'to die'	*umere-nie
die$_{atelic}$-INF	die$_{atelic}$-NMN	die$_{telic}$-INF	*die$_{telic}$-NMN

Russian has also prefixes and suffixes changing the telicity of a verb, and the occurrence of a telic prefix can disallow *-nie* affixation, even if the word also contains an atelic suffix (see Pazelskaya 2012). Crucially, whether *-nie* is allowed in these cases depends on the order of application of the telic/atelic affixes. Consider (5), showing a bare atelic verb. Prefix *ot-* in (6) makes it telic, so *-nie* cannot be applied.

5. [kry]-tʲ 'to cover'	[kry]-tie 'covering'
[cover$_{atelic}$]-INF	[cover$_{atelic}$]-NMN
6. [ot-kry]-tʲ 'to open'	*[ot-kry]-nie
[OT$_{telic}$-cover$_{atelic}$]-INF	*[OT$_{telic}$-cover$_{atelic}$]-NMN

The atelic suffix *-va* (7) is impossible to use with an atelic stem by itself, but it can be added to its telic version with the prefix *ot-* (8). In this case the last applied prefix is atelic, therefore *-nie* can be used.

7. *[kry-va]-t^{j} *[kry-va]-nie
   *[cover$_{atelic}$-VA$_{atelic}$]-INF *[cover$_{atelic}$-VA$_{atelic}$]-NMN
8. [[ot-kry]-va]-t^{j} 'to open' [[ot-kry]-va]-nie 'opening'
   [[OT$_{telic}$-cover$_{atelic}$]-VA$_{atelic}$]-INF [[OT$_{telic}$-cover$_{atelic}$]-VA$_{atelic}$]-NMN

Similarly to (7), (9) shows how it is not possible to add a telic prefix to a telic stem. Again, as expected, if one more telic prefix *na-* is added to the atelic stem, *-nie* cannot be applied anymore (10).

9. *[na-[ot-kry]]-t^{j} *[na-[ot-kry]]-nie
   *[NA$_{telic}$-[OT$_{telic}$-cover$_{atelic}$]]-INF *[NA$_{telic}$-[OT$_{telic}$-cover$_{atelic}$]]-NMN
10. [na-[[ot-kry]-va]]-t^{j} 'to open a lot of things' *[na-[[ot-kry]-va]]-nie
    [NA$_{telic}$-[[OT$_{telic}$-cover$_{atelic}$]-VA$_{atelic}$]]-INF *[NA$_{telic}$-[[OT$_{telic}$-cover$_{atelic}$]-VA$_{atelic}$]]-NMN

This pattern is not SL, because *-nie* can be separated from the last atelic prefix applied to the stem by multiple elements. In a TSL account, we would need to disallow *-nie* affixation in all cases in which a telic prefix was applied last, we cannot just project atelic affixes on the tier. However, once both types of affixes are projected, we can't make the dependency between the last atelic affix and *-nie* local on the tier. The figure below shows that this cannot be fixed by increasing the segments considered by the tier-grammar from 2 to 3, since all the substrings in the ill-formed words are also present in the well-formed ones.

In general, it seems that *-nie* can be added only when the amount of atelic affixes is bigger than the amount of telic ones. In the best case, this can be accounted for by a TSL grammar evaluating tier sequences of length at least equal to all the (a)telic affixes plus the *-nie* suffix – basically memorizing all possible tier strings. This is hardly a computationally appealing solution. In the worst case, if the number of possible telic/atelic alternation is potentially unbounded, the pattern would be not even regular.

Strict Locality of Derivational Strings What is essential in the process described above is the order of telic-atelic applications. Now, assume we encode the order of morphological operations in derivational *strings*. Such representations for ex. (5-10) are provided below. '$X \leftarrow Y$' stands for 'Y applies to X'.

11. V$_{atelic}$ ← INF V$_{atelic}$ ← NMN
12. V$_{atelic}$ ← OT$_{telic}$ ← INF *V$_{atelic}$ ← OT$_{telic}$ ← NMN
13. *V$_{atelic}$ ← VA$_{atelic}$ ← INF *V$_{atelic}$ ← VA$_{atelic}$ ← NMN
14. V$_{atelic}$ ← OT$_{telic}$ ← VA$_{atelic}$ ← INF V$_{atelic}$ ← OT$_{telic}$ ← VA$_{atelic}$ ← NMN
15. *V$_{atelic}$ ← OT$_{telic}$ ← NA$_{telic}$ ← INF *V$_{atelic}$ ← OT$_{telic}$ ← NA$_{telic}$ ← NMN
16. V$_{atelic}$ ← OT$_{telic}$ ← VA$_{atelic}$ ← NA$_{telic}$ ← INF *V$_{atelic}$ ← OT$_{telic}$ ← VA$_{atelic}$ ← NA$_{telic}$ ← NMN

This nominalization pattern can now be captured by simple SL constraints over the derivational strings, since nothing can intervene in-between the (a)telic affixes, *-nie* and INF. All we need is a grammar only allowing sequences of affixes of opposite telicity {AF$_{atelic}$-AF$_{telic}$, AF$_{telic}$-AF$_{atelic}$}. Then, the *-nie* suffix can be applied iff the previous operation resulted in an atelic stem {AF$_{atelic}$-NIE}. Thus, allowing our grammars to judge the wellformedness of morphological operations over derivations significantly reduces the computational power needed to capture apparently complex dependencies. Interestingly, this is similar to Graf and Heinz's (2016) proposal that syntactic dependencies are TSL over derivation trees.

Discussion In this paper, we presented a process potentially challenging for recent works that suggest that morphology is subregular. Then we argued for a derivational analysis of morphological operations, and showed how a pattern computationally heavy over the derived string, is simply SL if computed over an encoding of the sequence of morphological operations. This result draws interesting parallels between morphology and syntax, and opens the path for future work on derivational representations.

Aksënova, A., T. Graf, and S. Moradi. 2016. Morphotactics as Tier-based Strictly Local Dependencies. In *Proceedings of SIGMORPHON 14*, 121–130. ● **Beesley**, K., and L. Kartunnen. 2003. *Finite state morphology*. Stanford, CA: CSLI Publications. ● **Chandlee**, J. 2014. *Strictly Local Phonological Processes*. Ph.D. thesis, University of Delaware. ● **Graf**, T., and J. Heinz. 2016. *Tier-based strict locality in phonology and syntax*. Ms., Stony Brook University and University of Delaware. ● **Heinz**, J. and Idsardi W. 2013. What Complexity Differences Reveal About Domains in Language. *Topics in CogSci* 5(1):111–131. ● **Kaplan**, R. M., and M. Kay. 1994. Regular models of phonological rule systems. *CL* 20:331–378. ● **Pazelskaya**, A. 2012. Verbal prefixes and suffixes in nominalization: grammatical restrictions and corpus data. In *The Russian Verb*, 245–261.