Bibliography for Subregular Linguistics

Kenneth Hanson

September 6, 2023 (DRAFT)

This document is an overview of work in subregular linguistics, focusing on the use of subregular stringsets, treesets, and maps to analyze the complexity of linguistic patterns. A selection of works on mathematical foundations and learnability is also included.

In many cases there is more than one perspective to be taken on any given linguistic phenomenon. For example, patterns in phonology can be treated either as stringsets or maps, and syntactic dependencies can be treated either using tree grammars or constraints on path strings. Not all phenomena have been properly examined from each perspective, making for some obvious research topics.

Contents

1 Intro to Subregular Linguistics **1**

2 Phonology **2**
- 2.1 SL, SP, and TSL . . . 2
- 2.2 The Full Hierarchy . . . 3
- 2.3 SS-TSL and MTSL . . . 3
- 2.4 Phonological Maps . . . 4

3 Syntax **4**
- 3.1 Minimalist Grammars . . . 4
- 3.2 Tree tiers . . . 5
- 3.3 Command strings . . . 6
- 3.4 Syntactic Maps . . . 6

4 Logic and Learning **6**

5 Other Topics **7**
- 5.1 Morphology . . . 7
- 5.2 Probabilistic Grammars . . . 7

1 Intro to Subregular Linguistics

Start with Graf (2022a) for a broad introduction. Heinz (2018) focuses on phonology, but goes into greater depth. Rogers, Heinz, et al. (2013) provide a more formal, but accessible treatment of stringsets.

- Thomas Graf (2022a). "Subregular linguistics: bridging theoretical linguistics and formal grammar". In: *Theoretical Linguistics* 48.3–4. DOI: doi:10.1515/tl-2022-2037
  - Overview of subregular classes, application in phonology and syntax.

  - Emphasizes how subregular complexity results strike a balance between precise data coverage and robustness to changes in linguistic theory.
  - Part of a series of articles including responses by other authors and a response-to-the-responses by Thomas.
- Jeffrey Heinz (2018). "The computational nature of phonological generalizations". In: *Phonological Typology, Phonetics and Phonology*
  - Overview of subregular phonology, including both phonotactics and phonological maps.
  - Introduces classes of stringsets in terms of their correspondence to different types of mathematical logic.
  - Compares the computational perspective with more popular modes of explanation such as optimization and phonetic naturalness, and argues that it provides a better model of phonological typology.

2 Phonology

Subregular linguistics started with phonotactic patterns, analyzed using stringsets. This is the most straightforward topic both conceptually and mathematically, and so is still the natural starting place. The key subregular classes to be familiar with are SL (strictly local), SP (strictly piecewise), and TSL (tier-based strictly local). I've also included some works that deal with more expressive classes.

Most work to date focuses on segmental phonology, especially harmony patterns. I am not aware of any comprehensive analyses of tone or stress from a phonotactic perspective on par with the work on consonant and vowel harmony, though see Heinz (2014) and Rogers and Lambert (2019) regarding stress and Graf (2017) for a treatment of unbounded tone plateauing. See Section 2.4 for some work on tone patterns analyzed as phonological maps.

2.1 SL, SP, and TSL

- Jeffrey Heinz (2010a). "Learning Long-Distance Phonotactics". In: *Linguistic Inquiry* 41.4
  - Readable introduction to learning of phonotactics. Presents a model centered on SL and SP, as it predates the introduction of TSL. The learning model, called *string extension learning*, remains relevant.
- Jeffrey Heinz (2014). "Culminativity Times Harmony Equals Unbounded Stress". In: *Word Stress: Theoretical and Typological Issues*. Chap. 8
  - Shows how complex constraints can be factored into multiple simple constraints falling into classes such as SP, which are efficiently learnable.
- Jeffrey Heinz, Chetan Rawal, et al. (2011). "Tier-based strictly local constraints for phonology". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human language technologies*
  - Introduces the class TSL, which we now believe to subsume the vast majority of local and long-distance linguistic dependencies.
  - Provides some basic mathematical results. As far as math papers go, this one is rather easy to read.
- Kevin McMullin and Gunnar Ólafur Hansson (2016). "Long-Distance Phonotactics as Tier-Based Strictly 2-Local Languages". In: *Proceedings of the Annual Meetings on Phonology*. Vol. 2. DOI: 10.3765/amp.v2i0.3750

  - Discussion of long-distance phonotactics, focusing on consonant harmony. Almost all harmony patterns are TSL-2, but consonant harmony is more obviously long-distance in nature—many vowel harmony patterns are actually within SL (possibly with $k > 2$ if consonant clusters are allowed).

- Alëna Aksënova, Jonathan Rawski, et al. (n.d.). "The computational power of harmonic forms". To appear in *Oxford Handbook of Vowel Harmony*.
  - Overview of vowel-harmony patterns, including those requiring multiple tiers.

2.2 The Full Hierarchy

- James Rogers, Jeffrey Heinz, et al. (2013). "Cognitive and Sub-Regular Complexity". In: *Formal Grammar*. Vol. 8036. Lecture Notes in Computer Science
  - Overview of the 'local' and 'piecewise' branches of the classic subregular hierarchy, with logical definitions and intuitions in terms of the general properties of cognitive mechanisms needed to process or generate patterns of each class.
  - Examples of stress patterns that fall within each class.
- James Rogers and Dakotah Lambert (2019). "Some classes of sets of structures definable without quantifiers". In: *Proceedings of the 16th Meeting on the Mathematics of Language*. URL: https://aclanthology.org/W19-5706
  - Defines the SPL and PLT classes, which are hybrids of SP/SP and LT/PT, respectively.
  - Mentions co-SL and co-SP, which are equally simple to SL and SP, respectively.
  - Most up-to-date survey of the typology of stress patterns (brief as it may be).
- Thomas Graf (2017). "The Power of Locality Domains in Phonology". In: *Phonology* 34. URL: https://dx.doi.org/10.1017/S0952675717000197
  - Presents a generalization of SP called IBSP, intended as a more empirically adequate and linguistically informed model of certain long-distance phenomena such as unbounded tone plateauing.

2.3 SS-TSL and MTSL

Structure-sensitive TSL (SS-TSL) is a catch-all for various extensions of TSL with somewhat more complex tier projection mechanisms. Multi-TSL (MTSL) is simply the conjunction of constraints on multiple tiers. These classes are meant to handle some patterns that go beyond the capabilities of SL/SP/TSL while retaining the idea of forbidden substrings on a tier.

- Aniello De Santo and Thomas Graf (2019). "Structure Sensitive Tier Projection: Applications and Formal Properties". In: *Formal Grammar*. DOI: 10.1007/978-3-662-59648-7_3
  - Formal definitions and properties of the classes ITSL, MTSL, and MITSL.
- Thomas Graf and Connor Mayer (2018). "Sanskrit n-Retroflexion is Input-Output Tier-Based Strictly Local". In: *Proceedings of SIGMORPHON 2018*
  - Uses IO-TSL (TSL with structure-sensitive tier projection) to analyze one Sanskrit *n*-retroflexion, one of the most complex known phonological patterns.
- Alëna Aksënova and Sanket Deshmukh (2018). "Formal Restrictions On Multiple Tiers". In: *Proceedings of the Society for Computation in Linguistics 2018*. URL: https://aclanthology.org/W18-0307

  - Presents a generalization about MTSL patterns in vowel harmony, which is that the tiers never overlap partially, only completely or not at all. If true, this greatly reduces the number of possible combinations of tiers, making learning easier.

- Daniel Gleim and Johannes Schneider (2023). "Phonological processes with intersecting tier alphabets". In: *Proceedings of the Society for Computation in Linguistics 2023*. DOI: 10.7275/6EMR-8380
  - Presents three counter-examples to the empirical generalization in Aksënova and Deshmukh (2018). It is an open question whether it is possible to reanalyze all such cases either with SP constraints, or with structure-sensitive tier projection.

2.4 Phonological Maps

- Jane Chandlee and Jeffrey Heinz (2018). "Strict Locality and Phonological Maps". In: *Linguistic Inquiry* 49.1. URL: https://muse.jhu.edu/article/683698
- Jane Chandlee, Jeffrey Heinz, and Adam Jardine (2018). "Input Strictly Local Opaque Maps". In: *Phonology* 35.2. DOI: 10.1017/S0952675718000027
- Jeffrey Heinz and Regine Lai (2013). "Vowel Harmony and Subsequentiality". In: *Proceedings of the 13th Meeting on the Mathematics of Language (MoL 13)*. URL: https://aclanthology.org/W13-3006
- Adam Jardine (2016). "Computationally, tone is different". In: *Phonology* 33.2. DOI: 10.1017/S0952675716000129
- Phillip Alexander Burness et al. (2021). "Long-distance phonological processes as tier-based strictly local functions". In: *Glossa: a journal of general linguistics* 6.1. URL: https://www.glossa-journal.org/article/id/5780/

3 Syntax

Subregular syntax to date has used variants of Minimalist Grammars (MGs, Stabler 1997) to define the tree language to be analyzed. A basic understanding of classic MGs is enough to get started.

The basic idea of subregular syntax is that subregular classes of string languages, when generalized to trees, can be used model syntactic dependencies in much the same way as in phonotactics. For syntax in particular, it appears that SL and TSL are the right characterization for the vast majority of syntactic patterns. But the mathematics of trees is much more complex than that of strings, leaving a lot of uncertainty in the details.

As a practical matter, SL constraints are not an issue, but TSL can be generalized to trees in multiple ways. To date two approaches have been used: tree tiers (Graf 2018), in which the tier itself is a tree, and constraints over command strings (Graf and Shafiei 2019). As subregular syntax is much younger than subregular phonology, most phenomena have only been treated in one system or the other.

3.1 Minimalist Grammars

The MG feature calculus can be used to generate a variety of structures. For our purposes, the standard approach based on derivation trees is all we need.

- Thomas Graf (n.d.). “Minimalism and Computational Linguistics”. To appear in *Cambridge Handbook of Minimalism*.
  - Introduction to MGs for syntacticians. Comparison with Minimalist Program theories.
- Thomas Graf (2013). “Local and Transderivational Constraints in Syntax and Semantics”. PhD thesis. UCLA. URL: http://thomasgraf.net/doc/papers/Graf13Thesis.pdf
  - Chapter 1 is a general introduction to MGs. The remainder is not directly relevant but includes some of the mathematical background for subregular syntax, as it deals with constraints on derivation trees.

3.2 Tree tiers

Start with Graf (2022b), which uses the more recent approach involving dependency trees to analyze movement. Hanson (2023b) uses the same system to analyze case. Detailed definitions can be found in Graf and Kostyszyn (2021).

See Graf (2018) for the original definition of tree tiers. This paper uses a traditional MG derivation trees with internal nodes corresponding to Merge and Move operations. It also has a slightly different focus, analyzing the complexity of the MG Merge and Move, and one version of Adjoin. Vu et al. (2019) uses this system in their analysis of case dependencies.

- Thomas Graf (2022b). “Typological implications of tier-based strictly local movement”. In: *Proceedings of the Society for Computation in Linguistics 2022*. URL: https://aclanthology.org/2022.scil-1.15
  - Analyzes a wide range of movement patterns and restrictions on movement using tiers over MG dependency trees. A good entry point, since most recent works use dependency trees rather than traditional derivation trees.
- Kenneth Hanson (2023b). “A TSL Analysis of Japanese Case”. In: *Proceedings of the Society for Computation in Linguistics 2023*. URL: https://scholarworks.umass.edu/scil/vol6/iss1/4
  - In-depth analysis of case patterns in Japanese, providing a unified analysis of local and long-distance case assignment.
  - In contrast to Vu et al. (2019), uses dependency trees rather than derivation trees, and mostly avoids structure-sensitive tier projection.
- Thomas Graf and Kalina Kostyszyn (2021). “Multiple Wh-Movement is not Special: The Subregular Complexity of Persistent Features in Minimalist Grammars”. In: *Proceedings of the Society for Computation in Linguistics 2021*. URL: https://aclanthology.org/2021.scil-1.25
  - Presents an analysis of multiple wh-movement, later summarized in Graf (2022b).
  - Contains the definition of tree tiers over dependency trees.
- Thomas Graf (2018). “Why movement comes for free once you have adjunction”. In: *Proceedings of CLS* 53. URL: https://thomasgraf.net/output/graf18cls.html
  - First introduction of tree tiers, including input-sensitive tier projection.
  - Shows that in standard MGs Merge is SL and Move is TSL, but also that adding adjunction can change this. Depending on the model of adjunction and choice of representation (derivation trees in this case), both Adjoin and Merge become TSL as well.
  - Adjunction (and therefore also Merge) is probably SL for dependency trees regardless of analysis, though no one has looked into this closely.

- Mai Ha Vu et al. (2019). "Case assignment in TSL syntax: A case study". In: *Proceedings of the Society for Computation in Linguistics 2019*. URL: https://aclanthology.org/W19-0127
  - First attempt at using tree tiers to model distribution of morphological case. Focuses mainly on English, with some discussion of German.
  - Uses derivation trees and structure-sensitive tier projection.

3.3 Command strings

Command strings are first introduced in Graf and Shafiei (2019), but their analysis is a bit peculiar as it uses structure sensitive tier projection in order to attain finite constraints on a tier. Hanson (2023a) may be a better starting place is it uses plain TSL in a way that intentionally mirrors the typical treatment of harmony in phonology.

- Thomas Graf and Nazila Shafiei (2019). "C-command dependencies as TSL string constraints". In: *Proceedings of the Society for Computation in Linguistics 2019*. URL: https://aclanthology.org/W19-0121
- Nazila Shafiei and Thomas Graf (2020). "The Subregular Complexity of Syntactic Islands". In: *Proceedings of the Society for Computation in Linguistics 2020*. URL: https://aclanthology.org/2020.scil-1.49
- Thomas Graf and Aniello De Santo (2019). "Sensing Tree Automata as a Model of Syntactic Dependencies". In: *Proceedings of the 16th Meeting on the Mathematics of Language*. URL: https://www.aclweb.org/anthology/W19-5702
- Kenneth Hanson (2023a). *A Computational Perspective on the Typology of Agreement*. Unpublished manuscript.
- Kenneth Hanson (2023c). "Strict Locality in Syntax". In: *Proceedings of CLS 59*

3.4 Syntactic Maps

Not as much has been done with subregular tree transductions (maps) as with treesets.

- Thomas Graf (2020). "Curbing Feature Coding: Strictly Local Feature Assignment". In: *Proceedings of the Society for Computation in Linguistics 2020*. URL: https://aclanthology.org/2020.scil-1.27
- Thomas Graf (2023). "Subregular Tree Transductions, Movement, Copies, Traces, and the Ban on Improper Movement". In: *Proceedings of the Society for Computation in Linguistics 2023*. DOI: 10.7275/TK1N-Q855
- Jing Ji and Jeffrey Heinz (2020). "Input Strictly Local Tree Transducers". In: *Proceedings of the 14th International Conference on Language and Automata Theory and Applications (LATA 2020)*. Lecture Notes in Computer Science

4 Logic and Learning

This section contains some works focusing on the mathematical properties of subregular languages, in particular through the lens of their logical characterizations, as well as learnability results. The two are intimately related, as the properties of subregular classes provide a natural search space.

- Jeffrey Heinz (2010b). "String Extension Learning". In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*
- Dakotah Lambert et al. (2021). "Typology emerges from simplicity in representations and learning". In: *Journal of Language Modelling* 9.1
- Adam Jardine, Jane Chandlee, et al. (2014). "Very efficient learning of structured classes of subsequential functions from positive data". In: *Proceedings of the Twelfth International Conference on Grammatical Inference (ICGI 2014)*. Vol. 34. URL: http://proceedings.mlr.press/v34/jardine14a.html
- Adam Jardine and Jeffrey Heinz (2016). "Learning Tier-based Strictly 2-Local Languages". In: *Transactions of the Association for Computational Linguistics* 4. URL: https://aclanthology.org/Q16-1007
- Adam Jardine and Kevin McMullin (2017). "Efficient Learning of Tier-Based Strictly k-Local Languages". In: *Language and Automata Theory and Applications*

5 Other Topics

5.1 Morphology

There has been comparatively little work on morphology compared to phonology and syntax (and hardly any on semantics).

- Alëna Aksënova, Thomas Graf, et al. (2016). "Morphotactics as Tier-Based Strictly Local Dependencies". In: *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*. URL: https://www.aclweb.org/anthology/W/W16/W16-2019.pdf
- Jane Chandlee (2017). "Computational locality in morphological maps". In: *Morphology* 27.4

5.2 Probabilistic Grammars

While subregular linguistics generally works with categorical models, it is generally possible to extend subregular string languages to weighted (or even stochastic) languages, just as it is possible to extend them to maps. Those who know a bit of automata theory will understand that this is because weighted and probabilistic automata are basically types of transducers.

- Connor Mayer (2021). "Capturing gradience in long-distance phonology using probabilistic tier-based strictly local grammars". In: *Proceedings of the Society for Computation in Linguistics 2021*. URL: https://aclanthology.org/2021.scil-1.4
- Charles Torres et al. (2023). "Modeling island effects with probabilistic tier-based strictly local grammars over trees". In: *Proceedings of the Society for Computation in Linguistics 2023*. URL: https://scholarworks.umass.edu/scil/vol6/iss1/15