Reconciling Minimum Description Length with Grammar-Independent Complexity Measures

Jon Rawski, Aniello De Santo, and Jeffrey Heinz

Department of Linguistics @ SBU

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

Email:
jonathan.rawski@stonybrook.edu
aniello.desanto@stonybrook.edu
jeffrey.heinz@stonybrook.edu

Introduction

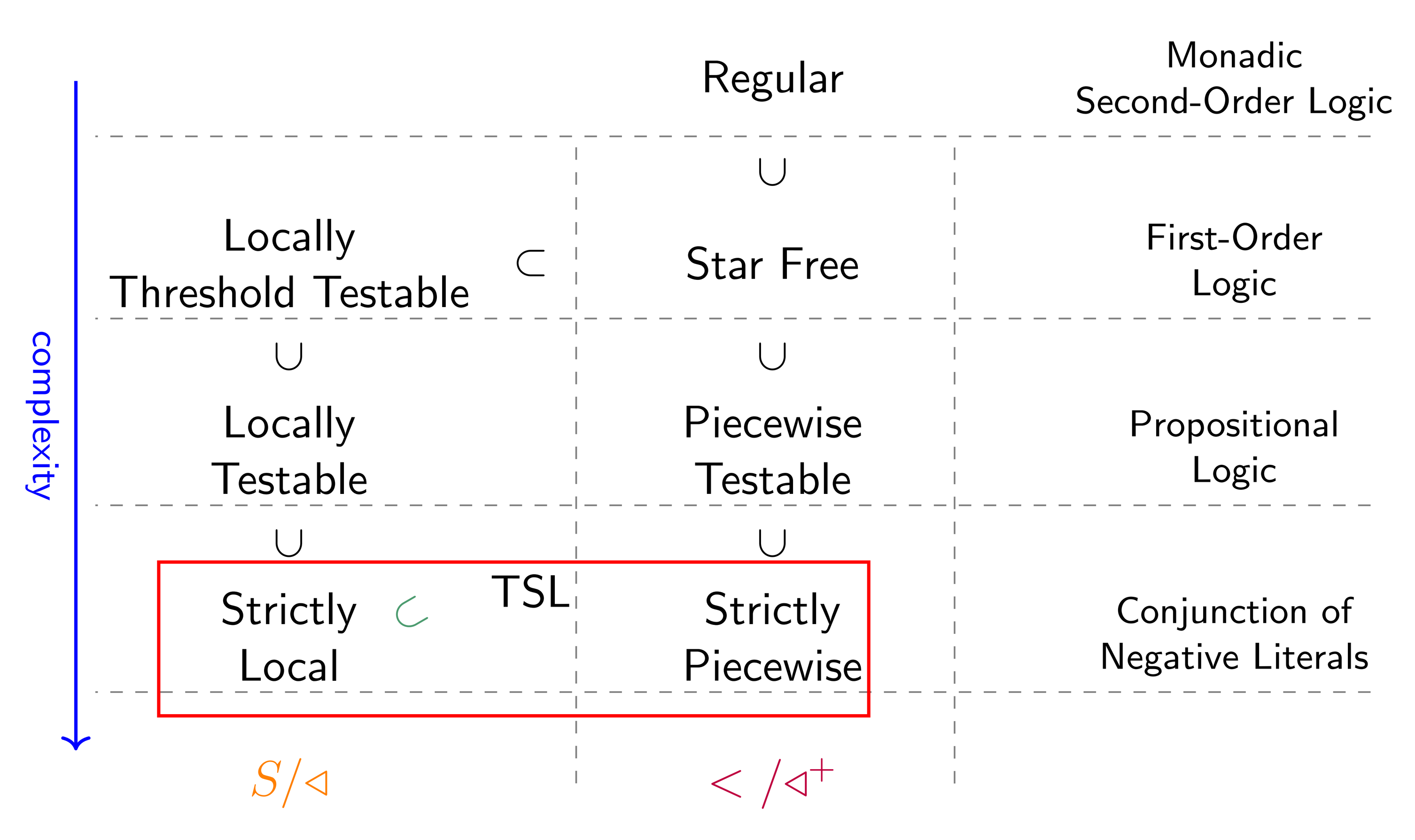
The Chomsky Hierarchy divides all logically possible linguistic patterns into nested regions of complexity, and provides grammar-independent characterizations that highlight necessary properties of any grammar/device that aims to recognize, generate, or encode a given linguistic pattern. Here we:

- review recent results in support of the so-called subregular hypothesis within this hierarchy;
- highlight contrasts between these results and Minimal Description Length (MDL) approaches;
- specify questions that need to be addressed in order to reconcile the two frameworks.

1 Subregular Complexity

The class of regular languages can be decomposed into a hierarchy of nested classes of decreasing complexity — the **subregular hierarchy** (Rogers et al. 2013; McNaughton and Papert 1971; i.a.).

The Subregular Hierarchy



The Subregular Hypothesis: natural language constraints are captured by the weak complexity classes at the bottom of the hierarchy.

Example: Intervocalic *s* voicing in German is SL

- In GERMAN, [s] is not allowed in-between two vowels:
  (1) fa[z]er 'fiber'
  (2) rei[z]en 'to.travel'
- G_{SL} = {*ase, *ise, *ese, *isi, ...}

ok r e i z e n   * r e i s e n

Example: Long distance sibilant harmony in Aari is SP

- In Aari, all sibilants agree in anteriority.
  (3) baʔse 'he brought'
  (4) ʒaʔʃit 'I arrived'
- G_{SP} = {*ʒs, *sʒ, *sʃ, *ʃs }

ok ʒ a ʔ ʃ i t   * ʒ a ʔ s i t

2 Evidence for the Subregular Hypothesis

The complexity differences highlighted by the subregular hierarchy are independent of any particular representation — and thus of the implementation details of the underlying cognitive mechanism. What's the evidence that this is a plausible metric for the complexity of human language?

Typological Coverage

A variety of phenomena are subregular in approximately the same way and these characterizations – SL, TSL, SP – go beyond simple phonotactic patterns:

- Phonotactics (Heinz 2010; Heinz&Idsardi 2013; Heinz forthcoming)
  → even across articulatory systems (Rawski 2017)
- Phonological mappings (Chandlee 2014; Chandlee et al. 2014, 2015)
- Morphotactics/Morphology (Aksënova et al. 2016; Aksënova & De Santo 2017; Chandlee 2016)
- Morpho-Semantics (Graf 2017; De Santo et al. 2017)
- Syntax/Tree Languages (Graf 2012; Graf&Heinz 2015)

Typological Gaps

The subregular hypothesis predicts that no pattern exceeding the expressivity of the bottom classes in the hierarchy is expected to arise in natural language. This tight fit provides a principled explanation for cross-linguistic typological gaps such as:

- first-last harmony, sour-grapes in phonotactics (Heinz 2015; Lai 2015);
- unbounded circumfixation in morphology (Aksënova et al. 2016).

Learning Results

- The simpler classes in the hierarchy – SL, SP, TSL – have more efficient learning algorithms (Heinz et al. 2012; Jardine and McMullin 2017) ...
- ... and appear to be more easily learnable by humans in Artificial Grammar Learning (AGL) experiments (Lai 2015; Hwangbo 2015; Avcu 2017).

3 MDL and Subregularity

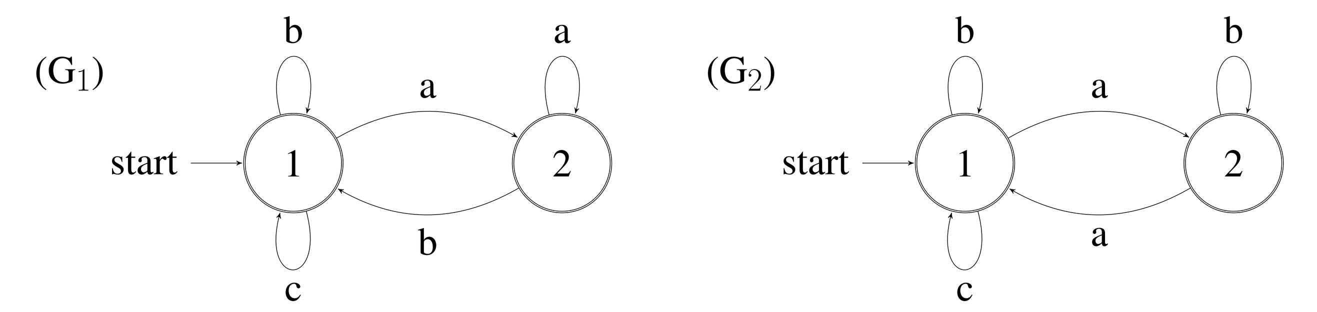
Hypothesis:

- Classes in the lower part of the hierarchy describe most natural language patterns;
- if MDL were the right basis for learning generalization, then we might expect that patterns which belong to these classes have shorter description lengths;

Two Puzzles (Heinz & Idsardi 2013; Rogers et al. 2013)

The above-mentioned correlation does not seem hold. Consider two grammars:

1. G_1 = {*ac}
2. G_2 = {*c given an odd number of a's in the left context.}



- The FSAs above are the smallest descriptions of the patterns encoded by the grammars, but G_1 corresponds to an SL pattern, while G_2 corresponds to a regular one.

Consider now a language that is recognized by both automata:

$$L_{12} = L(G_1) \cup L(G_2) = c^* \cup \{aabc, aaaabc, aaaaaabc, \ldots\}$$

- Subregular complexity predicts that a learner given a subset of L_{12} as an input will infer G_1, as it is the simplest grammar generating the language. What does MDL predict?

Reconciling MDL and Subregularity

The subregular hierarchy seems to describe computational complexity in different terms than MDL. Three fundamental questions must be addressed in order to reconcile them:

1. There are multiple ways to encode (sub-)regular languages & relative length of generalization is not preserved across these formalisms: is there some general encoding scheme where MDL generalizations match subregular predictions?
2. How can MDL-based approaches account for typological gaps?
3. How relevant to human learning are the computational distinctions highlighted by the subregular hierarchy and/or MDL? → AGL and neurolinguistic experiments are a promising way to shed light on this question.

References

Aksënova, A., T. Graf, and S. Moradi. 2016. Morphotactics as tier-based strictly local dependencies. In *Proceedings of SIGMorPhon/*. • **Aksënova**, A and A. De Santo. 2017. Strict Locality in Morphological Derivations. In *Proceedings of CLS 53*. • **Avcu**, E. 2017. Experimental investigation of the subregular hierarchy. In *Proceedings of PLC41*. • **Chandlee**, J. 2014. *Strictly local phonological processes*. PhD Thesis, U. of Delaware. • **Chandlee** J., Eyraud R., and Heinz J. 2015 Output strictly local functions. In *Proceedings of MoL*. • **Chandlee** J., Eyraud R., and Heinz J. 2014 Learning strictly local subsequential functions. *Transactions of the ACL*. • **De Santo**, A., T. Graf, and J. E Drury. 2017. Evaluating subregular distinctions in the complexity of generalized quantifiers. Talk given at ESSLLI 2017 Workshop on Quantifiers & Determiners (QUAD 2017), Toulouse, France. • **Graf**, T. 2017. The subregular complexity of monomorphemic quantifiers. Ms., Stony Brook University. • **Graf**, T. 2012. Locality and the complexity of minimalist derivation tree languages. In *Proceedings of FG 2010/2011*. • **Graf**, T., and J. Heinz. 2015. Commonality in disparity: The computational view of syntax and phonology. Talk given at GLOW 2015. • **Heinz**, J.. *forthcoming*. The computational nature of phonological generalizations. In *Phonological Typology* • **Heinz**, J. 2010 Learning Long-Distance Phonotactics. *Linguistic Inquiry*, 41. • **Heinz**, J., and William I. 2013. What complexity differences reveal about domains in language. *Topics in Cognitive Science*. • **Heinz**, J., A. Kasprzik, and T. Koetzing. 2012. Learning with lattice-structure hypothesis spaces. *Theoretical Computer Science*. • **Hwangbo**, H. J. 2015. Learnability of two vowel harmony patterns with neutral vowels. In *The third Annual Meeting on Phonology (AMP 2015)*. • **Jardine**, A., and K. McMullin. 2017. Efficient learning of tier-based strictly k-local languages. In *Language and Automata Theory and Applications, 11th International Conference*. • **Lai**, R. 2015. Learnable vs. unlearnable harmony patterns. *Linguistic Inquiry* 46. • **McNaughton**, R., and Seymour P. 1971. *Counter-free automata*. Cambridge: MIT Press. • **Rawski**, J. 2017. Subregularity across modalities: Evidence from sign phonology. In *Proceedings of CLS 53*. • **Rogers**, J., J. Heinz, M. Fero, J. Hurst, D. Lambert, and S. Wibel. 2013. *Cognitive and sub-regular complexity*, FG.