Information-theoretic characterization of the Subregular Hierarchy

Huteng Dai[1] and Richard Futrell[2]

Rutgers University[1] and University of California, Irvine[2]

Goal

To **link two notions of complexity**:

- **Formal language theory** (FLT) (Hopcroft and Ullman, 1979; Heinz, 2018)
- **Statistical complexity theory** (Feldman and Crutchfield, 1998)

Why?

Statistical complexity theory quantifies the **memory requirements** for generating a sequence incrementally. Limitations of human working memory may explain why natural language occupies the formal complexity class that it does.

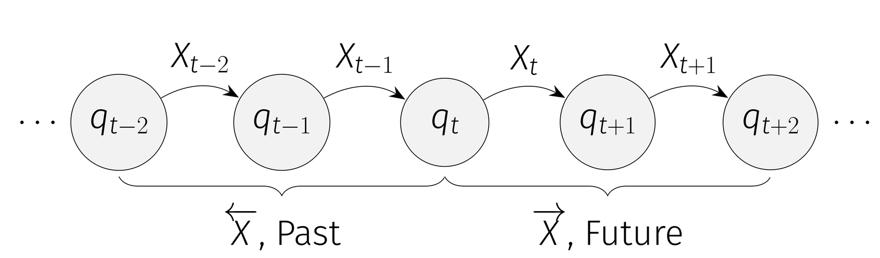
Probabilistic Deterministic Finite-State Automata (PDFAs)

A Probabilistic Deterministic Finite-State Automaton (PDFA) has:

- **Internal states** $\mathcal{Q}$, an **alphabet** Σ,
- an **emission distribution** $O : \mathcal{Q} \to \Sigma$ (a stochastic function), and
- a **transition function** $T : \mathcal{Q} \times \Sigma \to \mathcal{Q}$ (a deterministic function).

To construct a stochastic process from a PDFA: whenever the PDFA emits an end-of-word symbol, let it transition back into its initial state. The resulting infinite stream of symbols is a **stationary ergodic stochastic process**.

- Let X be any stochastic process generating sequences of symbols indexed as $\ldots X_{t-2}, X_{t-1}, X_t, X_{t+1}, \ldots$.



Statistical Complexity

The statistical complexity of X is the minimal amount of information about the past which must be stored in incremental memory in order to correctly generate the future:

$$S \equiv \min_{M : D_{KL}[\overrightarrow{X} | \overleftarrow{X} || \overrightarrow{X} | M] = 0} H[M],$$

$$\text{where } H[M] \equiv -\sum_x p_M(x) \log p_M(x).$$

Statistical complexity breaks into two terms (where $I[X : Y]$ is **mutual information**):

$$S = H[M] = I[M : \overrightarrow{X}] + H[M | \overrightarrow{X}]$$
$$= \underbrace{I[\overleftarrow{X} : \overrightarrow{X}]}_{E} + \underbrace{H[M | \overrightarrow{X}]}_{C},$$

Some Example PDFAs

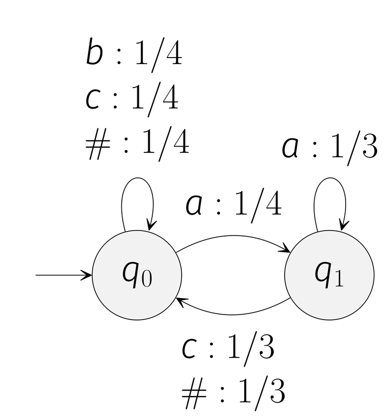


Figure 1: SL_2 PDFA of $\neg ab$, $\Sigma = \{a, b, c\}$

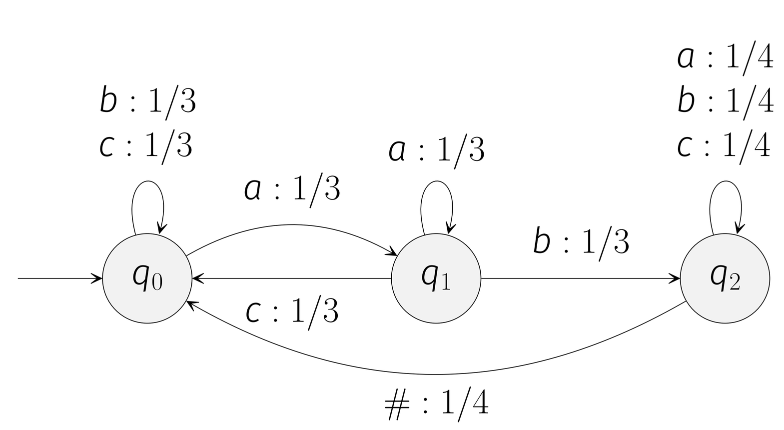


Figure 2: LT_2 PDFA of Some-ab, $\Sigma = \{a, b, c\}$

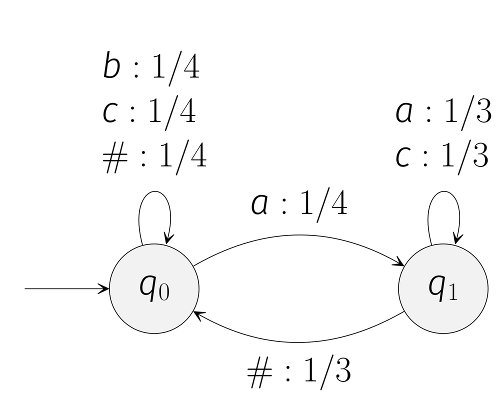


Figure 3: SP_2 PDFA of $\neg a \ldots b$, $\Sigma = \{a, b, c\}$

Calculation of Quantities for PDFAs

- The **statistical complexity** in a minimal PDFA is the entropy of the stationary distribution Q over states, given by the first left eigenvector of the state transition matrix, whose entries are given by:

$$p(q_{t+1} | q_t) = \sum_{x_t \in \Sigma} p_O(x_t | q_t) \delta_{q_{t+1} = T(q_t, x_t)}.$$

- For SL_k languages, **excess entropy**

$$E = I[X_{t-k+1}, \ldots, X_{t-1} : X_t, \ldots, X_{t+k-1}].$$

- For other languages, an SL_k approximation always gives a **lower bound on** E (Travers and Crutchfield, 2011). We compute lower bounds on E from SL_7 approximations.

Quantities

- **Excess entropy** E is the amount of information in the past which is actually used to predict the future.
- **Crypticity** C is the amount of information which must be stored, but which ends up being useless.
- **E/S ratio** is the proportion of bits stored in memory which end up being useful.

Main Result

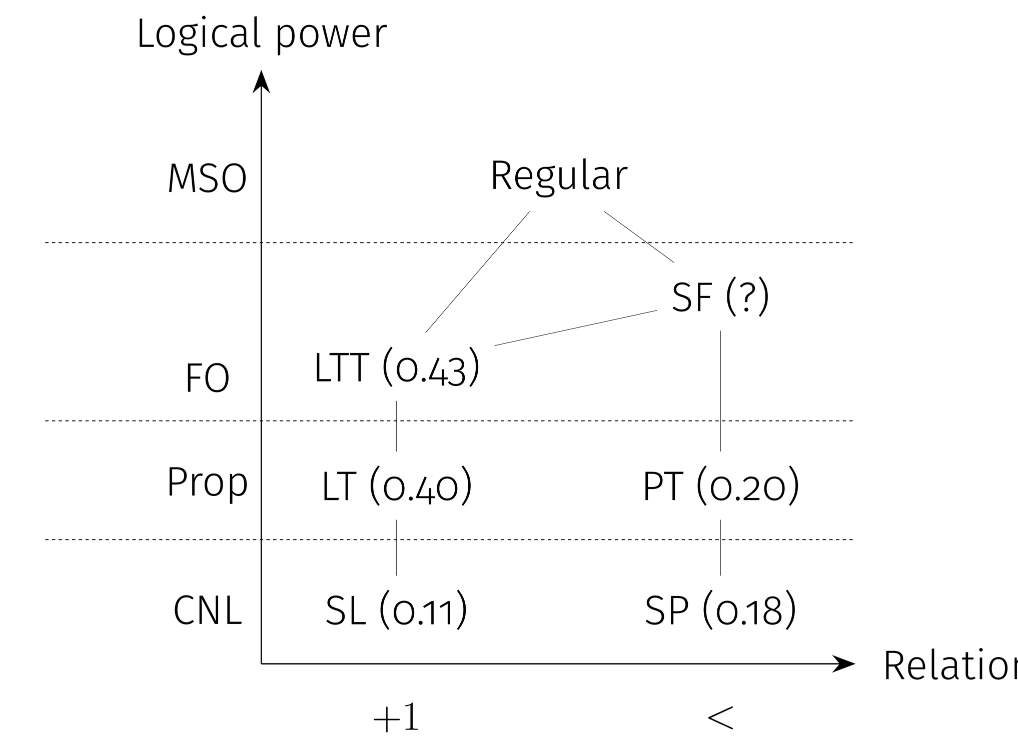
In a **series of example languages**, we find that increasing FLT complexity in the subregular hierarchy correlates with **E/S ratio**.

	SL_2	LT_2	LTT_2	SP_2	PT_2
S	0.97	1.53	1.94	0.99	1.53
E	0.09	$\geq$0.61	$\geq$0.83	$\geq$0.18	$\geq$0.30
C	0.75	$\leq$0.91	$\leq$1.10	$\leq$0.80	$\leq$1.22
E/S	0.11	$\geq$0.40	$\geq$0.43	$\geq$0.18	$\geq$0.20

Table 1: Information quantities for PDFAs

Conclusion

Based on our examples, we **conjecture** that crypticity correlates with logical power in the subregular hierarchy. In future work, we will formalize this idea.



References

Feldman, D. P. and Crutchfield, J. P. (1998). Measures of statistical complexity: Why? *Physics Letters A*, 238(4-5):244–252.

Heinz, J. (2018). The computational nature of phonological generalizations. *Phonological Typology, Phonetics and Phonology*, pages 126–195.

Hopcroft, J. E. and Ullman, J. D. (1979). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley.

Travers, N. F. and Crutchfield, J. P. (2011). Asymptotic synchronization for finite-state sources. *Journal of Statistical Physics*, 145(5):1202–1223.

Acknowledgements

We thank Jim Crutchfield, Jeff Heinz, Adam Jardine, and anonymous reviewers for their comments and insights.

RUTGERS UCI

Take a picture to download the extended abstract