

(0) **Some traditional questions.**

- Is language acquisition more like learning or more like growth or triggering?
- Are specific properties of human syntax innately present in every human learner?
- Are properties of learnable languages determined by every learner that can learn them?

Questions like these generate heated disputes, but they are typically posed in these vague terms that obscure the real issues. I think the questions are better formulated this way:

- How does our ability to learn languages shape the languages we speak?

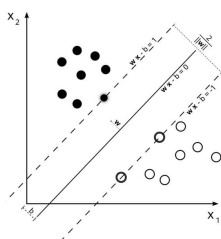
We will introduce some ways of thinking about learning that will allow much clearer and more interesting questions to be formulated and (some of them) answered.

(1) **Example 1.** Suppose we have heard these parts of a verbal paradigm [1, 4], affixes attached to the verb *li-*

	male human	female human	non-human
1	li-w-da	li-j-da	li-b-da
2	li-w-de	li-j-de	—
3	li-w	li-j	li-b

Do we expect li-b-de in the unfilled position?

- *k*-CNF (Valiant 1984). No. (no generalization)
- **Support vector machine** (Sadhara 2001). No



Maximum-margin hyperplane and margins for a SVM trained with samples from two classes. Samples on the margin are called the support vectors.

- **Maximum entropy** (Paris and Vencovská 1994). With 3×3 options, chances ≈11%
- **Monomial-preference** (Pertsova 2007). Yes.

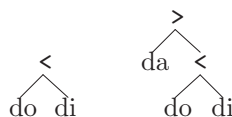
(2) **Example 2.** Suppose we have heard the following stress patterns [7, §1.6]:

´	´	´L	´H	L´
L´	´LL	´LH	´HL	´HH
L´L	L´H	LL´	LL´	L´LL
L´LH	´LLL	´LLH	´HLL	´HHL
L´HL	L´HH	´LHL	´LHH	´HHL
´HHH	LL´L	LL´H	LLL´	LLL´

Do we expect LL´LLL?

- **Neighborhood distinct** (Heinz 2007, 2008). Yes.

(3) **Example 3.** Suppose we hear the following syntactic expressions with the indicated headedness:



Is *da da do di* (where these expressions have the same categories as in the examples) in the language?

- **Rigid** (Kanazawa 1996). Yes. (contrast [2, 3, 9])

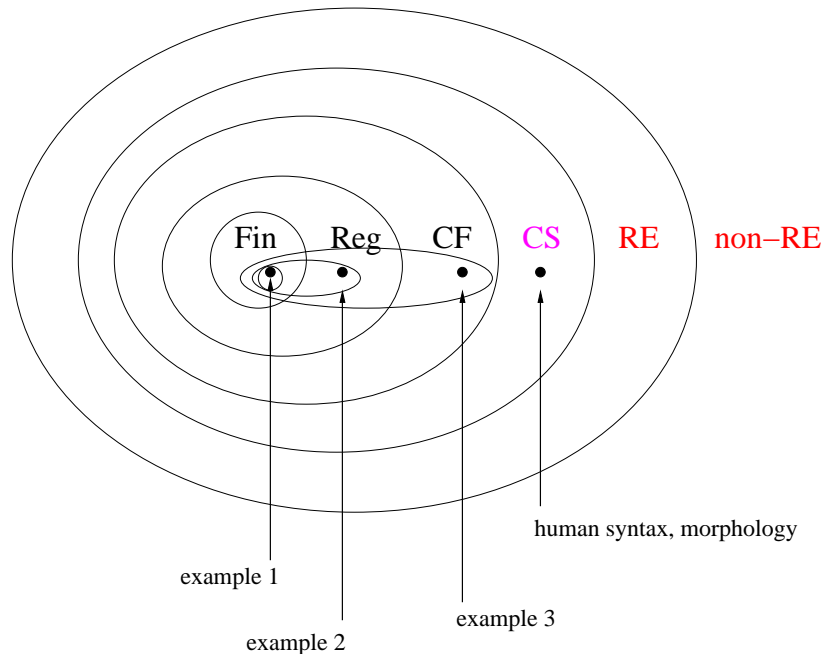
(4) **Data: exact and noisy.** First, we will look at clean data from one source. Noise is added later. **Learners as functions ϕ .** We will study two kinds of learning models too:

ϕ :data \rightarrow H, where H is the set of possible hypotheses about the data source

ϕ :data \rightarrow P, where P is the set of probability distributions on H

		exact	noisy
(5) the plan:	discrete H	1	2
	prob H	3	4

Why do ‘exact data, discrete H’ first? Because (i) it is easier, (ii) sometimes the language learner’s data *is* quite clear and exact, and (iii) this part of learning theory has surprising, important lessons about how generalization works.



References

- [1] ANDERSON, S. R. *A-Morphous Morphology*. Cambridge University Press, NY, 1992.
- [2] BOD, R. Is the end of supervised parsing in sight? In *Proceedings ACL 2007* (2007), pp. 400–407.
- [3] BOD, R. A linguistic investigation of unsupervised DOP. In *Proceedings of Workshop on Cognitive Aspects of Language Acquisition* (2007), pp. 1–8.
- [4] CYSOUW, M. *The Paradigmatic Structure of Person Marking*. Oxford University Press, NY, 2003.
- [5] HEINZ, J. *Inductive Learning of Phonotactic Patterns*. PhD thesis, University of California, Los Angeles, 2007.
- [6] HEINZ, J. On the role of locality in learning stress patterns. University of Delaware, 2008.
- [7] IDSARDI, W. *The Computation of Prosody*. PhD thesis, Massachusetts Institute of Technology, 1992.
- [8] KANAZAWA, M. Identification in the limit of categorial grammars. *Journal of Logic, Language, and Information* 5 (1996), 115–155.
- [9] KLEIN, D., AND MANNING, C. D. Natural language grammar induction with a generative constituent-context model. *Pattern Recognition* 38 (2005), 1407–1419.
- [10] PARIS, J., AND VENCOVSKÁ, A. A note on the inevitability of maximum entropy. *Journal of the Approximate Reasoning* 13 (1990), 570–581.
- [11] PERTSOVA, K. *Learning Form-Meaning Mappings in the Presence of Homonymy*. PhD thesis, University of California, Los Angeles, 2007.
- [12] SADOHARA, K. Learning of Boolean functions using support vector machines. In *International Workshop on Algorithmic Learning Theory, ALT’01* (NY, 2001), Lecture Notes in Artificial Intelligence, Springer, pp. 106–118.
- [13] VALIANT, L. G. A theory of the learnable. *Communications of the Association for Computing Machinery* 27, 11 (1984), 1134–1142.

Background on learning:

- [1] CRISTIANINI, N., AND SHAWE-TAYLOR, J. *Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, NY, 2000.
 - [2] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, NY, 2001.
 - [3] JAIN, S., OSHERSON, D., ROYER, J. S., AND SHARMA, A. *Systems that Learn: An Introduction to Learning Theory (second edition)*. MIT Press, Cambridge, Massachusetts, 1999.
 - [4] KEARNS, M. J., AND VAZIRANI, U. V. *An Introduction to Computational Learning Theory*. MIT Press, Cambridge, Massachusetts, 1994.
 - [5] PARIS, J. *The Uncertain Reasoner's Companion*. Cambridge University Press, NY, 1994.
-
- [6] SADOHARA, K. Learning of Boolean functions using support vector machines. In *International Workshop on Algorithmic Learning Theory, ALT'01* (NY, 2001), Lecture Notes in Artificial Intelligence, Springer, pp. 106–118.
 - [7] ANGLUIN, D. Inference of reversible languages. *Journal of the Association for Computing Machinery* 29 (1982), 741–765.
 - [8] KANAZAWA, M. Identification in the limit of categorial grammars. *Journal of Logic, Language, and Information* 5 (1996), 115–155.

For kernel methods like SVMs, there is a website <http://www.kernel-machines.org> with many tutorials.

Background in formal language theory:

- [1] HOPCROFT, J. E., MOTWANI, R., AND ULLMAN, J. D. *Introduction to Automata Theory, Languages and Computation (2nd Edition)*. Addison-Wesley, Reading, Massachusetts, 2000.
 - [2] LEWIS, H., AND PAPADIMITRIOU, C. *Elements of the Theory of Computation*. Prentice-Hall, Englewood Cliffs, New Jersey, 1981.
 - [3] SIPSER, M. *Introduction to the Theory of Computation*. PWS Publishing, Boston, 1997.
-
- [4] ROZENBERG, G., AND SALOMAA, A., Eds. *Handbook of Formal Languages*. Springer, NY, 1997.

Background in probability theory, linear algebra, etc.:

- [1] ISAAC, R. *The Pleasures of Probability*. Springer-Verlag, NY, 1995.
-
- [2] DURRETT, R. *Probability: Theory and Examples*. Duxbury, Boston, 1996.
 - [3] JACOD, J., AND PROTTER, P. *Probability Essentials*. Springer, NY, 2000.
 - [4] RESNICK, S. *A Probability Path*. Birkhäuser, Boston, 1999.
 - [5] ROSENTHAL, J. S. *A First Look at Rigorous Probability Theory*. World Scientific, Hackensack, New Jersey, 2006.
-
- [6] HOWARD, R. A. *Dynamic Probabilistic Systems*. Dover, NY, 2007.
 - [7] KARLIN, S., AND TAYLOR, H. M. *A First Course in Stochastic Processes*. Academic Press, San Diego, California, 1975.
 - [8] RESNICK, S. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.
-
- [9] STRANG, G. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, Massachusetts, 1998.
 - [10] HIRSCH, M. W., AND SMALE, S. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, NY, 1974.

Gibert Strang's undergrad text *Introduction to Linear Algebra* is very readable, and videos of Strang's lectures are available on the web here <http://ocw.mit.edu/OcwWeb/Mathematics/> – you can get a good review of basic linear algebra, from the first few (or more!) of these lectures.