John Archibald
University of Calgary

30,000,000 theories of phonological development*

1. Introduction

Two of the stated goals of this conference are to (1) see how current theories can illuminate data from phonological development, and (2) to investigate how developmental data allow us to evaluate phonological theories. These are important questions that any linguistic theory has to acknowledge. As far back as Pinker (1979) we have been reminded that one of the conditions on a formal model of grammar is that the grammar be 'learnable'. Much of the energy along these lines has been spent in looking at syntactic acquisition, but the same features are important in phonology.

In this paper, I will present some of the necessary background to reveal how we can possibly be in a situation where we have so many possible theories to account for something as 'simple' as children acquiring the sounds of their first language. Let us begin by refreshing our memories as to Pinker's (1979) conditions on formal models of language acquisition, enumerated below:

- The *learnability* condition is met if the theory can account for the fact that the language can be learned.

---

* With thanks to James McCawley for the precursor to this title. Thanks also to Piotr and Beata for inviting me to Warsaw. While I was the recipient of marvelous hospitality even by the border guards who patiently explained to me the situation of Canadian citizens needing visas while suggesting to me that I might find a return to Germany an expedient option. Luckily, I was given a two-day visa and all was well. Thanks too to Jerzy Rubach for his kindness.

- The *equipotentiality* condition is met if the theory does not succeed merely by being extremely narrow; forcing many things to be specified as innate where they can be learned. The theory must be able to account for the acquisition of all languages.
- The *time* condition is met if the theory accounts for learning in the time the learner normally takes for the acquisition of a grammar.
- The *input* condition is met if the theory accounts for language learning with the typical input available to the learner.
- The *developmental* condition is met if the theory makes correct predictions about the learner's capabilities during the course of acquisition.
- The *cognitive* condition is met if the theory agrees with the known cognitive faculties of the learner.

These serve to remind us that while, in some sense, the linguist is free to propose a model of phonological knowledge that deals only with the characteristics of the data, acquisitionists must impose a more stringent requirement on a model of acquisition which ensures that children can acquire this grammatical knowledge on the basis of the evidence available to them in the primary linguistic data.

Chomsky (1988) has also invoked general schemata to circumscribe the acquisition of linguistic knowledge. As given in (1), we see the goals of the linguist who is interested in language acquisition.

(1) Chomsky on the acquisition of knowledge
   i. Characterize the knowledge (Grammar)
   ii. Record the relevant experience (Data)
   iii. Determine how (ii) is relevant to (i) (Learning Theory)

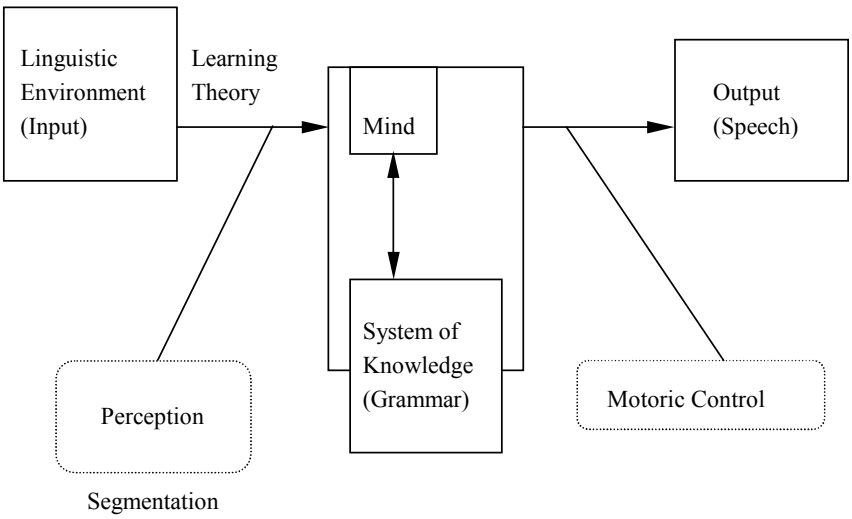
On the surface they appear to be such simple goals but, of course, a huge amount of linguistic literature is spilled on these topics.

2. What we agree on: I

In order to see how, in the somewhat facetious title, thirty million theories could possibly arise, consider the schematic diagram given in (2)

which is designed to be a theory-neutral, uncontroversial model of the language acquisition process. People hear something, learn something, and say something.

(2)



Regardless of the theoretical stance one takes, there is clearly *some* environmental influence on the organism. Describing the properties of the input may be viewed to be relatively theoretically neutral, but once one starts to discuss the interaction between the organism and the environment, things become more contentious. One requires a theory of perception. This theory may or may not be specific to *linguistic* perception.

Parameter 1: Linguistic Perception is Distinct from Auditory Perception YES/NO.

One also requires a theory of what is to be acquired (a target model of the grammar). Of course, there are many different models of the target grammar, many of which are construct-specific (e.g., what is the best

model of feature structure?, what is the best model of intonational phonology?), but nonetheless we can conceive of some parametric variation here.

Parameter 2: Phonological patterns emerge from constraints on output YES/NO

Parameter 3: Lexical entries are underspecified YES/NO

We think that all researchers would agree that learners draw on grammatical knowledge when initiating the motoric command sequences to produce linguistic utterances. But again, theorists may differ when attempting to account for accepted patterns in data such as whether the behaviour of coronals is due to phonological or phonetic factors.

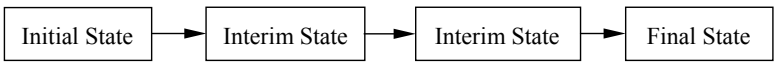
Parameter 4: Coronals are unmarked due to the smaller muscle mass of the tip of the tongue YES/NO

It is clear then that different theories of phonological development can emerge based on the different theoretical stances to be taken on each of the boxes in (2) above.

3. What we agree on: II

There is a second set of facts that we can all agree on: (1) learners begin the language acquisition task with a well-defined cognitive state, (2) the end state of language acquisition is also a well-defined state, (3) individuals acquiring a language proceed through developmental stages that are systematic. A schematic which summarizes these beliefs is given in (3):

(3)



4. Trouble in paradise

However, now we must move beyond generalities and attempt to make these constructs specific. The three questions in (4) illustrate the range of debate that can result.

(4) i. What is the nature of the initial state?
ii. Do the interim states matter?
iii. What is the nature of the final state?

When it comes to the nature of the initial state, we can immediately refer to the views of modular and non-modular architecture. A modular architecture would assume that there is a language faculty which is dedicated to linguistic knowledge. A non-modular architecture would assume a structure something like a neural network in which the knowledge is represented in a format that is common across cognitive domains.

Parameter 5: The language faculty is modular YES/NO

The question of whether the interim states are relevant invokes two distinct theoretical debates. First of all, Chomsky is famous for using the phrase "instantaneous acquisition". Often this phrase is misinterpreted as evidence for Chomsky supposedly believing that language acquisition for children *is*, in fact, instantaneous. However, what is actually adopted here is a theoretical stance which states that the end point of the language acquisition process is not affected by the interim states, and therefore, the theoretical linguist when attempting to model the final state may proceed with the assumption that it is *as if* language acquisition were instantaneous.

Parameter 6: The interim states affect the end state YES/NO

Secondly, there is the debate about continuity versus maturation. The Continuity Hypothesis maintains that the interim grammars are not *qualitatively* different from the end-point grammars. The units of mental

representation for child grammars would be the same as for adult grammars. This is contrasted with the Maturation Hypothesis which maintains that child grammars may contain structures which are not found in adult grammars; child grammars may be qualitatively different from adult grammars.

Parameter 7: Child grammars are qualitatively different from adult grammars YES/NO

The third point in (4) above has to do with the theoretical assumptions of the model of the target grammar. In the phonological literature today, there are several models of phonological competence. Whether one supports a Principles and Parameters model, or Government Phonology, or Beats and Binding, or Optimality theory will have a decisive influence on the questions of "what is to be acquired?". As the field of learnability has made clear, before we can come to an understanding of how 'something' is acquired, we need to have a good model of what that 'something' is.

Parameter 8: Phonological competence includes parameters YES/NO

Parameter 9: Universal Grammar includes information about the cues to the setting of those parameters YES/NO

Up to now, we have considered different sources of variation in theories of phonological development. We have illustrated these points with nine parameters. If all the parameters are independent, we have sanctioned 2^9 (512) grammars.

Another theoretical distinction that can lead to diverse schools of thought is the Rationalist/Empiricist debate. For many hundreds of years, philosophers have been debating the answer to the question "how we come to know what we know?". The table in (5) illustrates some of the differences between Rationalist and Empiricist approaches to language acquisition.

(5) Rationalist versus Empiricist Theories

Empiricism	Rationalism
• powerful innate learning procedures	• innate structure
• pattern association	• domain specificity
• environment as shaper	• environment as trigger
• learning by induction	• learning by deduction
• mind as tablet of hot wax	• mind as dark museum
• feedback crucial	• no negative evidence
• hypothesis construction	• hypothesis selection

Rationalist theories tend to assign more weight to innate structures of knowledge in humans. For example, children might be born with the knowledge that all human languages have things like consonants and vowels, and onsets and nuclei. However, they have to learn, on the basis of exposure to the primary linguistic data available in the ambient language, which segments can be assigned to onsets or codas in their language, or whether onsets can branch or not. Some children will be exposed to a language which allows branching onsets, while other will not. Some children will be exposed to a language which allows obstruents in codas while others will not.

Parameter 10: Environment acts as a trigger YES/NO

Parameter 11: Feedback is essential to learning YES/NO

5. A cornucopia of knowledge- (or skill-) based theories

We see below a list of some of the theories we have already discussed plus a few others. For each of these, I will give some examples to illustrate the properties of the theories.

(6) **Modular Theories** → Chomsky
- examples of modules: language, morals, music
- selective deficits used as evidence

(7) **Non-modular Theories** → Piaget, Slobin, O'Grady
- cause and effect, time (apply to all domains)
- operating principles: e.g., pay attention to the ends of words

(8) **Emergent Modularity Theories** → Elman
- modules may emerge from a dynamic self-organizing system

Parameter 12: Modules are an emergent cognitive structure YES/NO

(9) **Functional Theories** → Boersma, Dziubalska-Kołaczyk
- preference theories
- conflicting demands of perception and articulation

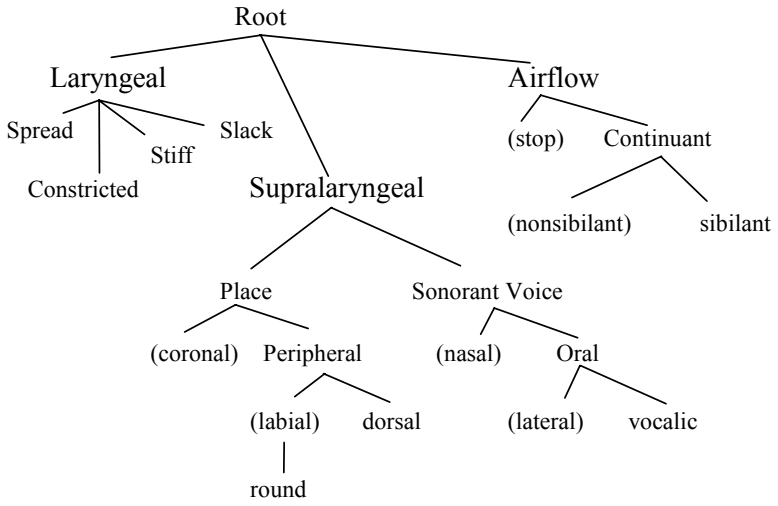
(10) **Principles and Parameters Theories** → Dresher & Kaye, Fikkert
- all languages have syllables
- all languages have [PLACE]
- codas are allowed [Yes/No]
- feet are built from the [Left/Right]

(11) **Connectionist Theories** → Stemberger, Rumelhart & McClelland
- knowledge is the state of a network
- spreading activation of nodes; weights of connections

Parameter 13: There is no cognitive difference between regular and irregular phenomena YES/NO

All of the above theories (we are up to 2^{13} = 8,192 now) refer to quite broad properties. However, not surprisingly, there are also differences within theoretical approaches. Even in models of phonological acquisition which share many beliefs and assumptions on the nature of features, we can find differences. An input representation model is shown in (12).

(12) **Input Representation Theories** → Brown, Rice & Avery



Models such as these assume that the acquisition of a complex feature geometry proceeds in a monotonic and deterministic fashion. By this we mean that features can only be added one at a time in the order determined by Universal Grammar. Even here there can be a difference of opinion as to whether children begin with fully specified representations and prune those features not required by the ambient language (e.g., Stampe), or whether children begin with impoverished representations that are elaborated on the basis of positive evidence.

Parameter 14: Children begin with impoverished representations YES/NO

Such theories also must take a stance of the issue of underspecification of the lexical entry. Many researchers assume that phonological (i.e., underlying) representations are stored without predictable phonological information while phonetic representations are fully specified. Therefore a choice must be made as to the theory of underspecification theory.

Parameter 3 (repeated): Lexical entries are underspecified YES/NO

Parameter 15: Underspecification is contrastive only YES/NO

Other theories argue that the locus of explanation for phonological phenomena lies not in the stored representations but rather in the patterns found in the output. Such a theory is shown in (13).

(13) **Output Form Theories** → Pater, Gnanandesikan

	*COMPLEX	FAITH
please: pliz → pliz	*!	
☞ pliz → piz		*
peas: ☞ piz → piz		
piz → iz		*!

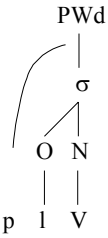
If we assume simple constraints like being faithful to an input form and preferring simple (not complex) structures, we could account for the fact that a child may produce words like *please* and *peas* as [piz]. These models also accept that forms may not be either perfectly well-formed or perfectly ill-formed but rather may be an optimal form for a particular context.

Parameter 16: Patterns in production determine phonological patterns YES/NO

Parameter 17: Constraints are violable YES/NO

There have been recent proposals of hybrid theories with adopt much of the machinery of Optimality Theory but introduce a prosodified input representation, as shown in (14).

(14) **Hybrid Representation Theories** → Goad & Rose

	*COMPLEX	*APP-LEFT	MAX HEAD	MAX
a. 	*!			
b. 		*!	*!	
☞ c. 				*

Parameter 18: Input representations are prosodified YES/NO

It is not just the grammar that suffers from a multiplicity of theoretical positions (we are now up to 2^{18} = 262,144). Learning theories face the same fate, as shown in the next section.

6. A plethora of learning theories

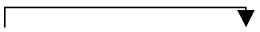
Some researchers have proposed that learning proceeds in a manner reminiscent of natural selection. Candidates for a mental representation compete with one another and the hypothesis that is the 'best fit' is the one that will survive. Hypotheses with a poor fit will be purged from the system.

(15) **Genetic Algorithms** → Pulleyblank & Turkel
- a grammar string
- a cross-over operator
- a mutation operator
- a fitness metric
- a reproductive mechanism

Parameter 19: Learning proceeds via a genetic algorithm YES/NO

Somewhat more specific to Optimality Theory (but not completely) is the learning algorithm of Constraint Demotion shown in (16).

(16) **Constraint Demotion** → Tesar & Smolensky
"For any constraint C assessing an uncanceled winner mark, if C is not dominated by a constraint assessing an uncanceled loser mark, demote C to immediately below the highest-ranking constraint assessing an uncanceled loser mark."



	(DEM)	SUP	DEM	MID	ORG	BET
winner	(*)		*			*
loser		*			*	

Parameter 20: Learning proceeds via an algorithm of Constraint Demotion YES/NO

Still others have proposed that learning proceeds by a random one-step learning procedure known as the Triggering-Learning algorithm shown in (17).

(17) **Triggering-Learning Algorithm** → Gibson & Wexler
"Given an initial set of values for *n* binary-valued parameters, the learner attempts to syntactically analyze an incoming sentence *S*. If *S* can be successfully analyzed then the learner's hypothesis regarding the target grammar is left unchanged. If, however, the learner cannot analyze *S*, then the learner uniformly selects a parameter *P*... changes the value associated with *P*, and tries to reprocess *S* using the new parameter value. If analysis is now possible, then the parameter value change is adopted. Otherwise, the original parameter value is retained."

Parameter 21: Learning proceeds via a Triggering-Learning algorithm
YES/NO

A little more idiosyncratic is the proposal that cognitive development is best modeled by theories proposed to account for the behaviour of non-organic systems. When considering the process of crystal formation we note that very complex structures emerge with infinite variation in morphology. Perhaps language acquisition proceeds in the same way as children acquire more and more complex knowledge.

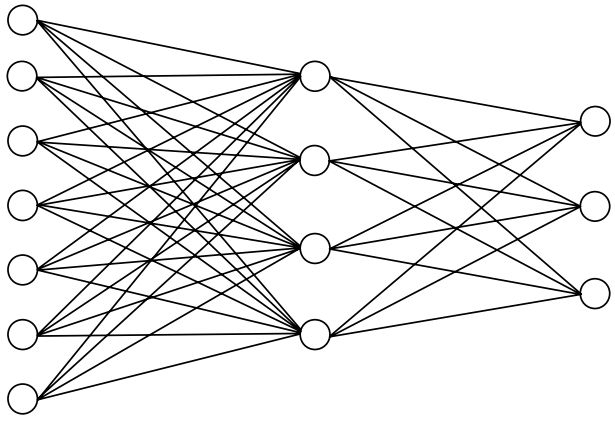
(18) **Crystal Formation** → Mohanan

- UG has the properties of a "strange attractor" like a planet orbiting two stars; some states are more likely than others.
- "The current debate on the choice between rules and constraints has the same status as the question of whether gravity is a rule or constraint."

Parameter 22: Learning proceeds via procedures of non-organic complexity YES/NO

Connectionist learning algorithms assume that learning is a stochastic modeling of the input from the ambient language, as shown in (19).

(19) **Connectionist Algorithms** → Stemberger



Parameter 23: Learning proceeds via a statistical modeling of the input YES/NO

Others have suggested that Universal Grammar provides the learner with built-in knowledge as to the kinds of evidence which will tell the learner that the current grammar-state is incorrect. An example of this is shown in (20).

(20) **Cue-Based Learning** → Dresher
- Boundedness of Constituents:
  Parameter: Line 0 constituents are {unbounded/bounded}.
  Default: Assume Line 0 constituents are unbounded.
  Cue: The presence of a stressed nonedge light syllable indicates bounded constituents.
- Extrametricality:
  Parameter: A syllable on the {right/left} {is not/is} extrametrical.
  Cue: Stress on a peripheral syllable rules out extrametricality on that side.

Parameter 24: UG provides innate cues for the resetting of parameters YES/NO

Finally, there are learning theories which assume that the reranking of constraints is not an instantaneous, all-or-nothing process. The algorithm given in (21) assumes that a ranking is a continuous function that can be adjusted minimally based on the environmental input.

(21) **Gradual Learning Algorithm** → Boersma, Levelt, Hayes
"Lower the rankings of all the constraints violated in the adult form, and raise the rankings of all the constraints violated in the learner's form (by a little amount)."

Parameter 25: Constraints can be reranked gradually YES/NO

7. Conclusion

Given this variety of possible learning algorithms in addition to the variety of possible grammars, we should not be surprised to discover that we are now up to 2^{25} theories of phonological development. That means 33,554,432. Of course, not each of the parameters I have mentioned in this paper are completely independent, but I hope that, as the field of phonological development continues to progress, we do not forget some of the lessons we have learned upon the way.

References

Boersma, P. (1998). *Functional Phonology*. Holland Academic Graphics.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.). *Second Language Acquisition and Linguistic Theory*. Blackwell.

Chomsky, N. (1984). *Modular Approaches to the Study of Mind*. San Diego State University Press.

Clark, R. (1992). The selection of syntactic knowledge. *Language Acquisition* 2.1, 83-150.

Dresher, B.E. (1999). Charting the learning path: cues to parameter setting. *Linguistic Inquiry*.

Dresher, E. and J. Kaye (1990). A computational learning model for metrical phonology. *Cognition* 34, 137-95.

Dziubalska-Kołaczyk, K. (1997). 'Syllabification' in first and second language. In J. Leather and A. James (eds). New Sounds 1997. University of Klagenfurt Press.

Elman, J. et al. (1996). *Rethinking Innateness*. MIT Press.

Escudero, P. and P. Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and Gradual Learning Algorithm. In S. Arunachalam, E. Kaiser and A. Williams (eds). *Proceedings of the 25th Annual Penn Linguistics Colloquium*. (*Penn Working Papers in Linguistics* 8.1), 71-85.

Fikkert, P. (1994). *On the Acquisition of Prosodic Structure*. HIL Dissertation.

Gibson, E. and K. Wexler (1994). Triggers. *Linguistic Inquiry* 25, 407-54.

Gnanadesikan, A. (1996). Child phonology in optimality theory. In A. Stringfellow et al. (eds). *BUCLD 20 Proceedings*, 237-48.

Goad, H. and Y. Rose (2000). Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. Manuscript.

Mohanan, K.P. (1992). Emergence of complexity in phonological development. In C. Ferguson et al. (eds). *Phonological Development: Models, Research, Implications*. York Press.

O'Grady, W. (1987). *Principles of Grammar and Learning*. University of Chicago Press.

Pater, J. (1997). Minimal Violation and Phonological Development. *Language Acquisition* 6:3, 201-53.

Piaget, J. (1929). *The Child's Conception of the World*. Routledge and Kegan Paul.

Piatelli-Palmirini, M. (1980). *Language and Learning*: The Debate Between Jean Piaget and Noam Chomsky. Harvard University Press.

Pinker, S. (1979). Formal models of language learning. *Cognition*: 1, 217-83.

Pulleyblank, D. and W.J. Turkel (2000). Learning phonology: genetic algorithms and Yoruba tongue-root harmony. In J. Dekkers et al. *Optimality Theory: Phonology, Syntax and Acquisition*.

Rice, K. and P. Avery (1995). Variability in a deterministic model of language acquisition: a theory of segmental elaboration. In J. Archibald (ed.). *Phonological Acquisition and Phonological Theory*. 23-42. Hillsdale, N.J.: Lawrence Erlbaum and Associates.

Rumelhart, D. and J. McClelland (eds) (1986). *Parallel Distributed Processing*. MIT Press.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson and D. Slobin (eds), *Studies of Child Language Development*, 175-208. Holt, Rinehart, and Winston.

Stampe, D. (1972). *How I Spent My Summer Vacation* [A Dissertation on Natural Phonology]. Garland Press.

Stemberger, J. (1992). A connectionist view of child phonology. In C. Ferguson et al. (eds). *Phonological Development: Models, Research, Implications*. York Press.

Steriade, D. (2001). The phonology of perceptability effects. Ms. MIT.

Tesar, B. and P. Smolensky (2001). *Learnability in Optimality Theory*. MIT Press.