

(to appear in Proceedings of 30th Child Language Research Forum,
Stanford University, april 1999. Copyright CSLI.)

Draft, August 27, 1999.

Gradual Constraint-Ranking Learning Algorithm Predicts Acquisition Order¹

PAUL BOERSMA AND CLARA LEVELT

We will show that the Gradual Constraint-Ranking Learning Algorithm is capable of modelling attested acquisition orders and learning curves in a realistic manner, thus bridging the gap that used to exist between formal computational learning algorithms and actual acquisition data.

1 An Attested Acquisition Order

Levelt, Schiller, and Levelt (to appear) found that the acquisition order for syllable types for twelve children acquiring Dutch is as depicted in Figure 1.

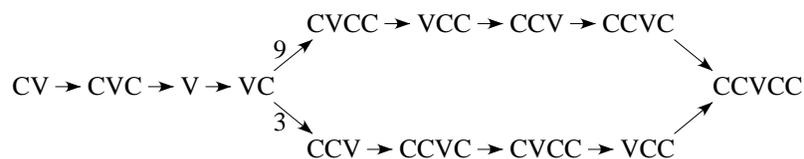


Figure 1. Acquisition order for syllable types in Dutch.

Thus, syllables with unbranching codas (-VC) are always acquired before syllables without onsets (V-), but there is variation in the order of the acquisition of complex codas (-VCC) and complex onsets (CCV-).

2 An Optimality-Theoretic Account

To account for the acquisition order in Figure 1, Levelt et al. proposed that the child's syllable forms are determined by a developing *Optimality-*

¹ This work is supported by grants from the Netherlands Organization for Scientific Research.

Theoretic grammar (Prince and Smolensky 1993) of interacting *markedness* and *faithfulness* constraints. Four markedness constraints play a role:

- *CODA “don’t produce codas (-VC or -VCC)”
- ONSET “don’t produce vowel-initial syllables (V-)”
- *COMPLEXCODA “don’t produce complex codas (-VCC)”
- *COMPLEXONSET “don’t produce complex onsets (CCV-)”

A single faithfulness constraint is involved. It militates against deleting or inserting segments:

- FAITH “realize lexical segments; don’t realize non-lexical segments”

In the initial state of the child’s grammar, all markedness constraints are ranked above the faithfulness constraint (Gnanadesikan 1995). Faithfulness then gradually rises in the hierarchy, overtaking the markedness constraints one by one: first *CODA, then ONSET, then (variably) *COMPLEXCODA or *COMPLEXONSET, and finally the remaining one. In the end, faithfulness is ranked on top and the child masters all syllable structures.

3 The frequency hypothesis

On the basis of cross-linguistic data on syllable inventories, Levelt and Van de Vijver (1998) noted that several developmental orders are possible in principle, next to the order in Figure 1. They hypothesized that language-particular orders are determined by the relative frequency of appearance of the different syllable types in adult, child-directed speech.

The attested distribution of overt syllable types in adult Dutch child-directed speech is shown in Table I.

CV	44.81 %	CCVC	1.98 %
CVC	32.05 %	CCV	1.38 %
VC	11.99 %	VCC	0.42 %
V	3.85 %	CCVCC	0.26 %
CVCC	3.25 %		

Table I. Frequencies of various syllable types in Dutch.

These data were extracted from a corpus of 112,926 primary stressed syllables (Joost van de Weijer, p.c.). We see that adults violate *CODA in 49.95 percent of the forms, ONSET in 16.26 percent, *COMPLEXCODA in 3.93 percent, and *COMPLEXONSET in 3.62 percent. Thus, the order of the frequencies of violations of markedness constraints is equal to the order in which FAITH was proposed (in §2) to overtake these constraints.

Having noted that the Gradual Constraint-Ranking Learning Algorithm (Boersma 1997; 1998: chs. 14–15) is sensitive to differences in frequencies of constraint violations, we will model Dutch syllable-type acquisition with the help of this algorithm, which we will describe next.

4 Gradual Constraint-Ranking Learning Algorithm

The algorithm consists of three ingredients:²

Continuous ranking scale. Each constraint has a *ranking value* along a continuous scale. This is in contrast with original Optimality Theory, where constraints are ranked along an ordinal scale. On the continuous scale, the distance between constraints can vary: some lie relatively close to each other, others are separated by a larger distance. This can have an effect at evaluation time, because of the following property.

Noisy evaluation. Every time an Optimality-Theoretic tableau has to be evaluated, an amount of normally distributed noise is temporarily added to the ranking value of each constraint. The constraints in the tableau are then ordered on the basis of the resulting effective ranking values, after which the familiar Optimality-Theoretic principle of strict domination determines the winning candidate. If two constraints A and B are at a relatively close distance from each other (not more than a few noise standard deviations), the effective ranking value will sometimes be higher for A, sometimes for B, which can lead to variation in the surface form, with the relative probabilities depending on the difference between the ranking values.

Error-driven learning. The child's grammar gradually changes as she compares her own forms with adult forms. Specifically, if she notices a difference between her form and the adult form, she will lower the ranking of all constraints in her grammar that are violated in the correct adult form by a small value (*plasticity*) along the ranking scale, and she will raise the ranking of all constraints violated in her own incorrect form. Gradually, the child will become more likely to produce the adult form.

5 Modelling the Acquisition Process

5.1 Modelling the Initial State

To model the initial dominance of markedness over faithfulness, we simulate the child's initial state by arbitrarily setting the ranking value of all markedness constraints to 100, and that of FAITH to 50.

² The algorithm is available in the Praat program, <http://www.fon.hum.uva.nl/praat/>

5.2 Modelling the Language Environment

For the distribution of syllable types as inputs to the learning algorithm, we took the attested distribution of overt syllable types in adult Dutch child-directed speech (Table I). Thus, our simulated learner was presented with thousands of syllables, drawn from a distribution equal to the one in Table I.

5.3 Modelling Error-Drivenness

Every time our simulated learner is presented with (“hears”) an adult surface form, she will compare it to a surface form that would be generated by her current grammar from an underlying form whose phonological representation is equal to the adult surface form. If the two surface forms are different, she will take action by changing some constraint rankings.

5.4 Modelling the Noise

Throughout the simulation, the noise standard deviation was fixed at 2.0. This entails that if two constraints are ranked by a distance of about 10 or more, the output is nearly categorical, and that if the distance is much smaller than 10, there may be variation and optionality in the output.

5.5 Modelling the Plasticity

The error-driven ranking change was fixed at 0.1, which means, for instance, that for *CODA to fall to a ranking value of 80, the learner would have to produce 200 violations of *CODA in forms in which the adult correctly produces a coda.

5.6 The Results of the Simulation

Figure 2 summarizes the result of our simulation. The paths followed by the constraint rankings as functions of time confirm the proposed account (§2), with FAITH overtaking first *CODA, then ONSET, then the remaining two.

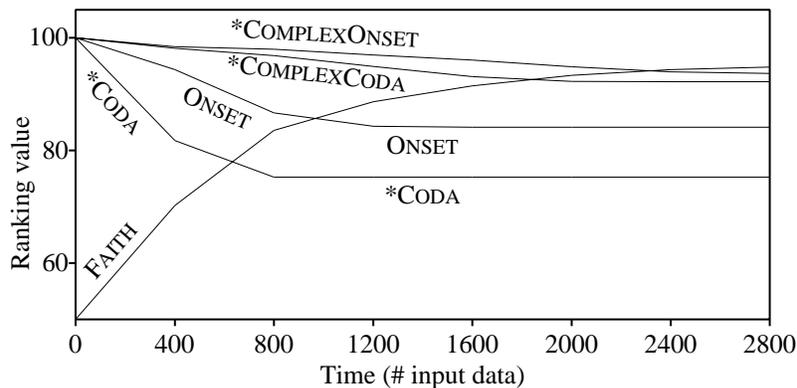


Figure 2. Constraint rankings as functions of time.

5.7 A Detailed Look into What Happens

Suppose that our learner is in the stage of the learning process that corresponds to having heard 400 data, and is presented with the adult surface form [a:p]. Tableau I shows the details of what happens.

/a:p/ ‘monkey’	*COMCONS	*COMPCODA	ONSET	*CODA	FAITH
√ [a:p]			*!→	*→	
☞ [pa:]					←**
[pa:p]				*!	*
[a:]			*!		*

Tableau I. After 400 data.

The ranking values that can be read off Figure 2 (at 400 data) will probably give rise to the effective constraint ordering shown along the top row of Tableau I. On hearing the adult surface form [a:p], the child will recognize it as the underlying form /a:p/ ‘monkey’, which she then takes as an input to her own grammar, as shown in the top left cell of Tableau I. The tableau shows four relevant candidates for the child’s output form. According to the temporary ranking in the tableau, the form [pa:] will win, as is indicated by the pointing finger (☞). However, the child notices that the adult surface form is [a:p], and that this form is different from her own surface form. Since the adult form is available among the candidates, we can indicate this correct form with a check mark (√). Likewise, we indicate the incorrectness of the child’s own form by putting two asterisks around the pointing finger.

Since the child’s surface form is incorrect, the child will take action by raising the ranking values of all constraints violated in that form. In this case, only FAITH will have to be promoted, and this is indicated by the leftward arrow in Tableau I. But the child will take another action. Since the correct form occurs in the tableau, too, she will lower the ranking values of the constraints violated in that form (ONSET and *CODA), as indicated by the rightward arrows. If the child repeatedly says [pa:] for /a:p/, she will eventually manage to rank FAITH above ONSET and *CODA, and become more likely to produce the adultlike form [a:p].

Having seen the details of the learning algorithm, we can return to the child’s initial stage. In the beginning, the constraint ranking causes the child to produce CV syllables only. In 44.81 percent of the cases, the adult form will be CV as well, so nothing happens. In 49.95 percent of the cases, though, the adult form will contain one or more coda consonants. The child takes this as her underlying form, but still generates a CV surface form

herself, and notices the difference. As a result, she will lower *CODA and raise FAITH. After 400 data, *CODA has moved down the ranking scale by a distance of approximately $49.95\% \cdot 400 \cdot 0.1 = 20.0$, and FAITH has risen to about 72. At that time, the constraints will be ranked as in Tableau I.

After about 800 data, *CODA has fallen far below FAITH, so that the child will make few errors in pronouncing simple codas. Thus, there will be no differences between the number of *CODA violations in the adult and learner forms, so that *CODA will stop moving through the hierarchy. However, ONSET still outranks FAITH, so that the child may now produce /a:p/ with an epenthesis as [pa:p], which is a form attested in one of the twelve live subjects. As Tableau II shows, this error will cause gradual demotion of ONSET, and further raising of FAITH.

/a:p/ 'monkey'	*COMCONS	*COMPCODA	ONSET	FAITH	*CODA
√ [a:p]			*!→		*
[pa:]				**!	
*  * [pa:p]				←*	*

Tableau II. After 800 data.

After 1200 data, ONSET is dominated most of the time, so the child begins to sound more adultlike again. She will still have trouble, however, with complex onsets and codas, as witnessed by her production of underlying /e:nt/ 'duck' as [e:t] (Tableau III). Again, [e:t] is a form attested in reality.

/e:nt/ 'duck'	*COMCONS	*COMPCODA	FAITH	ONSET	*CODA
√ [e:nt]		*!→		*	*
*  * [e:t]			←*	*	*
[te:t]			**!		*

Tableau III. After 1200 data.

This proceeds until faithfulness has overtaken the constraints against complex onsets and codas. As can be guessed from Figure 2, however, the rankings will continue to diverge until FAITH is ranked by a distance of 10 above all the others. The cause of this *safety margin* is noisy evaluation: if FAITH is ranked above *COMPLEXCODA by a distance of only 4.0, the probability of /e:nt/ being produced as [e:t] is still 7.9 percent. The curves of the rankings as functions of time get gradually flatter, because the learner will produce fewer errors as her rankings approach the adult's grammar.

6 Replicating the Acquisition Order

6.1 Predicted and attested learning curves

After every 100 data, we measured the performance of our learner by feeding her 10,000 underlying CVC syllables, having her stochastic grammar generate the corresponding surface forms, and seeing what percentage of these surfaced faithfully as CVC. We did the same for four other syllable types. The resulting learning curves are in Figure 3.

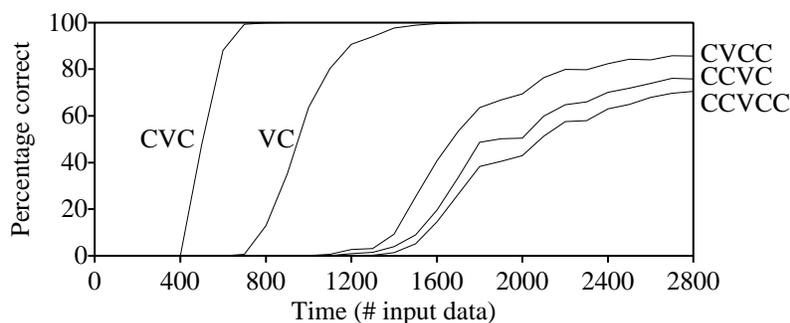


Figure 3. Five learning curves for our simulated learner.

Let us compare this to the behaviour of an actual child. Figure 4 shows the percentage of underlying CVC forms that he produced faithfully (we ignored forms with final liquids, which are often vocalized).

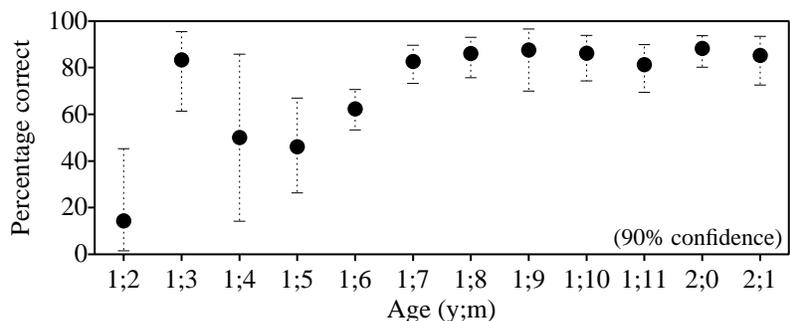


Figure 4. CVC learning curve for Tom.

Both the simulated learner and the actual child show gradual learning. For instance, Jarmo (at 1;9.9) pronounced /bo:m/ ‘tree’ as [po:], [bɔ], [bo:χ], [pɔ̃om], variably violating and satisfying *CODA during a single recording session. Such realistic modelling is not possible with learning algorithms based on ordinal ranking, like that by Tesar and Smolensky (1998).

6.2 Replicating variation in acquisition order

In our first simulation (Fig. 3), complex codas were acquired before complex onsets, but we repeated the whole experiment 30,000 times and found the reverse order in 31 percent of the cases. This variability is due to the proximity of the rates of adult *COMPLEXONSET and *COMPLEXCODA violations (§3). This result matches the behaviour of the twelve live subjects, three of whom acquired complex onsets before complex codas (Fig. 1).

7 Conclusions

The things that we modelled realistically were:

- The fixed order of acquiring syllables with codas, then vowel-initial syllables, then complex codas and onsets.
- The variable order of acquisition of complex codas and onsets.
- The graduality of the learning curves: no one-shot learning.
- The rapid initial rise and slow approach to 100 percent correctness.

There is also room for improvement. We could model on-line acquisition more precisely by taking more segmental details into consideration, e.g. by not regarding [sp-], [k1-], and [kn-] indiscriminately as ‘complex onsets’. Also, instead of making the simplifying assumptions in §5.3, we could take into account the development of perception and lexicalization as well. The learning algorithm is already well equipped to handle these refinements.

References

- Boersma, P. (1997). How we Learn Variation, Optionality, and Probability. *Proc. Institute of Phonetic Sciences of the University of Amsterdam* 21:43–58.
- Boersma, P. (1998). *Functional Phonology*. Doctoral dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
- Gnanadesikan, A. (1995). *Markedness and Faithfulness Constraints in Child Phonology*. Ms, University of Massachusetts, Amherst. Rutgers Optimality Archive 67. <http://ruccs.rutgers.edu/roa.html>
- Levelt, C., N. Schiller, and W. Levelt (to appear). The Acquisition of Syllable Types. *Language Acquisition*.
- Levelt, C. and R. van de Vijver (1998). *Syllable Types in Cross-Linguistic and Developmental Grammars*. Rutgers Optimality Archive 265.
- Prince, A. and P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.
- Tesar, B. and P. Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* 29:229–268.