Doing away with defaults: The parametric gradient hypothesis

Katherine Howitt, Meredith Lancaster and William Gregory Sakas
Hunter College and the Graduate Center, City University of New York
BUCLD 42, November 2017

Default parameter settings have been posited to simplify the learning process [1][2] and provide easy implementation of the Subset Principle [3][4]. Defaults have proven effective in overcoming obstacles to a learner which appear otherwise insurmountable. However, proposed default values are not always ideal from either a psycholinguistic or theoretical perspective (e.g., non-immediate acquisition of obligatory subjects [5], movement as default [1]).

In lieu of defaults, we propose that parameter values exist on a gradient scale of confidence where one binary value of a parameter gains or loses cogency through exposure to sentences revealing linguistic features, or *e-triggers* [1], of the child's language environment. E-triggers can be either *unambiguous*—aggressively increasing the learner's confidence in a parameter value, *ambiguous but relevant*—conservatively increasing the learner's confidence, or *irrelevant*—learner's confidence does not change.

We have simulated a No Defaults Learner (*NDL*) on four languages drawn from the CUNY-CoLAG Domain [6]. CoLAG is an artificial language domain generated by 13 syntactic parameters that contains phenomena typical of child-directed speech: null subjects, wh-movement, verb movement, word order, etc. The 13 parameters of CoLAG have been grouped into three categories [1]: (*C1*) parameters with unambiguous triggers for both values; (*C2*) parameters with unambiguous triggers for one value only; and (*C3*) parameters lacking unambiguous triggers for both values. We present data (Table 1) supporting the NDL's effective acquisition of parameter values in C1 and C2. The data suggest that the model's ability to successfully balance aggressive and conservative confidence adjustment in C2 will lead to future success in C3.

The NDL proceeds by searching a sentence for e-triggers for each parameter. When it encounters one, it adjusts its confidence for that parameter accordingly (Table 2), and then proceeds to the next parameter. The NDL, while distinctly different, is related to the variational learner (*VL*) [7][8]. The weights employed by the VL are similar to the confidence gradient of the NDL. However, the VL uses weights to conduct a non-deterministic search of a discrete parameter space, whereas the NDL's gradient scale directly embodies the parameters themselves.

A significant aspect of our parametric gradient hypothesis is that parameters can incrementally encode indirect negative evidence. Wh-Movement is a subset-superset parameter in CoLAG. A non-fronted Wh-phrase is an unambiguous e-trigger for Wh-in-situ, and when encountered, the NDL will adjust confidence aggressively. However, unambiguous e-triggers for Obligatory-Wh-Movement (ObWhM) do not exist. When the NDL encounters a fronted Wh-phrase, it conservatively adjusts its confidence toward ObWhM, although the movement could stem from optional topicalization (unrelated to the Wh-Movement parameter). For a language that doesn't contain Wh-in-situ, this conservatism enables the NDL to unfailingly converge (unlike other learners, e.g., the VL) on the subset language.

The parametric gradient hypothesis, coupled with the rejection of defaults and use of e-triggers, offers a theoretically compelling way to model an individual's syntactic competence and can further be used to simulate group consistency, differences in idiolect, and the effects of language contact. The work presented here lays the groundwork for future detailed investigation.

(499 Words)

	Subject Position		Headedness in IP		Headedness in CP		Null Subject		Null Topic	
	C-Value	# Sentences	C-Value	# Sentences	C-Value	# Sentences	C-Value	# Sentences	C-Value	# Sentences
CoLAG German	0.01	1,035	0.99	802	0.01	3,645	0.01	293,431	0.94	500,000
CoLAG French	0.01	877	0.01	598	0.01	3,230	0.01	195,525	0.01	97,716
CoLAG English	0.01	753	0.01	646	0.01	4,611	0.01	167,917	0.01	83,899
CoLAG Japanese	0.01	5,634	0.99	849	0.99	418	0.95	500,000	0.01	376,779

	WH-Movement		Pied Piping		Topic Marking		VtoI Movement		Affix Hopping	
	C-Value	# Sentences	C-Value	# Sentences	C-Value	# Sentences	C-Value	# Sentences	C-Value	# Sentences
CoLAG German	0.99	32,618	0.01	7,611	0.01	7,309	0.97	500,000	0.01	1,055
CoLAG French	0.99	21,759	0.01	6,577	0.01	4,895	0.96	500,000	0.01	1,438
CoLAG English	0.99	31,026	0.99	5,508	0.01	4,180	0.01	3,359	0.99	3,818
CoLAG Japanese	0.03	500,000	0.01	9,353	0.99	458	0.01	12,845	0.01	6,420

Table 1: NDL Simulation Data. Averages of 100 simulation runs of the NDL with a maximum of 500,000 sentences for four CoLAG languages. In the table, each language is named for the natural language it most resembles. *C-Value* is a confidence point on the learner's gradient scale for each parameter over the interval (0, 1). C-values are initialized to 0.5. *# Sentences* is the average number of sentences consumed by the NDL before reaching an imposed *C-value* threshold of 0.01 (or 0.99) or reaching the maximum 500,000 sentences.

C-Value Adjustment for:	`Learner consumes sentence s ⇒`

Unambiguous Triggers:	`if s has an e-trigger for Pi(0) or Pi(1):`
	`  adjust confidence at rate R toward 0 or 1 respectively`
	`else:`
	`  do nothing (s doesn't contain an e-trigger for Pi)`

Ambiguous Triggers:	`if s has an ambiguous e-trigger for Pi(0) or Pi(1):`
	`  adjust confidence at conservative rate r towards 0 or 1 respectively`
	`else:`
	`  do nothing (s does not contain an e-trigger for Pi)`

Mutually Exclusive Parameter Values P_i and P_j:	`if s has an unambiguous e-trigger for Pi(1):`
	` adjust confidence towards 1 at rate R AND adjust confidence for Pj toward 0 at rate R`
	`else if s has an unambiguous e-trigger for Pi(0):`
	` adjust confidence towards 0 at rate R AND adjust confidence or Pj toward 1 at rate R`

Table 2: C-value Adjustment for Unambiguous and Ambiguous E-triggers and Mutually Exclusive Parameter Values. Following [7], we employ the L_{r-p} scheme [9] for adjusting the learner's confidence, where adjusting toward 0 is calculated using: $P_i = P_i - \check{R}\,(P_i)$ and adjusting toward 1 is calculated using: $P_i = P_i + \check{R}\,(1 - P_i)$. $\check{R}$ can be either R (aggressive rate) or r (conservative rate). For the simulation study reported in Table 1, $R = .02$ and $r = .001$.

References

[1] Sakas, William G. & Janet D. Fodor. 2012. Disambiguating Syntactic Triggers. *Language Acquisition*. 19(2). 83-143.

[2] Hyams, Nina M. 1986. *Language acquisition and the theory of parameters.* Dordrecht: Reidel.

[3] Manzini, M. Rita & Kenneth Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18(3). 413-444.

[4] Fodor, Janet D. & William G. Sakas. 2005. The Subset Principle in syntax: Costs of compliance. *Journal of Linguistics* 41(03). 513-569.

[5] Hyams, Nina M. & Kenneth Wexler. 1993. On the grammatical basis of null subjects in child language. *Linguistic Inquiry* 24(3), 421-459.

[6] CUNY CoLAG Projects. *Colag.cs.hunter.cuny.edu/projects.html.* Web access on 11 May 2017.

[7] Yang, Charles. 2002. *Knowledge and Learning in Natural Language*. Oxford University Press: Oxford, UK.
[8] Yang, Charles. 2012. Computational models of syntactic acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science* 3, 205-213.
[9] Bush, Robert & Frederick Mosteller. 1958. *Stochastic Models for Learning.* New York: Wiley.