The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis

Anne S. Hsu [a,*], Nick Chater [b], Paul M.B. Vitányi [c]

[a] Department of Cognitive, Perceptual and Brain Sciences, University College London, 26 Bedford Way, London, WC1H 0AP, UK
[b] Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK
[c] National Research Center for Mathematics and Computer Science in the Netherlands (CWI), Science Park 123, 1098 XG Amsterdam, The Netherlands

ABSTRACT

There is much debate over the degree to which language learning is governed by innate language-specific biases, or acquired through cognition-general principles. Here we examine the probabilistic language acquisition hypothesis on three levels: We outline a novel theoretical result showing that it is possible to learn the exact *generative model* underlying a wide class of languages, purely from observing samples of the language. We then describe a recently proposed practical framework, which quantifies natural language learnability, allowing specific learnability predictions to be made for the first time. In previous work, this framework was used to make learnability predictions for a wide variety of linguistic constructions, for which learnability has been much debated. Here, we present a new experiment which tests these learnability predictions. We find that our experimental results support the possibility that these linguistic constructions are acquired probabilistically from cognition-general principles.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Much research suggests children can learn language from mere exposure, without relying on other's feedback about their own utterances, i.e., from positive evidence alone. How children thus learn language has been a heavily researched topic. Two main perspectives on language acquisition can be understood through the distinction between *discriminative* and *generative* learning models (Hsu & Griffiths, 2009). A discriminative model learns by establishing a boundary between categories by mapping inputs to categories from a set of input-category pairs. For language, these are categories of grammatical and ungrammatical sentences. From the discriminative perspective, the ability to learn from only positive examples seems puzzling: with only positive examples, i.e., gram-

matical sentences, a discriminative learner has no basis on which to determine the boundary between grammatical and ungrammatical sentences. Under discriminative based perspectives, theoretical analyses of learnability from only positive examples tend to be discouraging (Gold, 1967; Nowak, Komarova, & Niyogi, 2002). Indeed, thinking about language learning as a classification problem has led many theorists to conclude that language acquisition faces fundamental "logical" problems (Baker & McCarthy, 1981; Horning, 1969).

The Bayesian approach to cognitive development, explored in this special issue, and the cognitive sciences more generally, e.g., Griffiths, Chater, Kemp, Perfors, and Tenenbaum (2010), suggest a different perspective on learning: generative models learn by making inferences about the probability distribution that produces the language input. Thus, from a generative perspective, language acquisition is not a matter of discriminating "good" from "bad" linguistic forms; instead the aim is to model the underlying regularities that give rise to the language. The key assumption

Table 1
Sentences used in experiment: quadruplets illustrating restriction rule.

	Set 1	Set 2
	(a) restricted form of construction with restriction*	(a) restricted form of construction with restriction*
	(b) un-restricted form of construction with restriction	(b) un-restricted form of construction with restriction
	(c) restricted form of related construction	(c) restricted form of related construction
	(d) un-restricted form of related construction	(d) un-restricted form of related construction
1	(a) I think it is fine where it's.*	(a) Do you know who she's?*
	(b) I think it is fine where it is.	(b) Do you know who she is?
	(c) I think it's fine for now.	(c) Do you know who she's waiting for?
	(d) I think it is fine for now.	(d) Do you know who she is waiting for?
2	(a) He arrived the plane safely.*	(a) The captain arrived the ship on a small island.*
	(b) The plane arrived safely.	(b) The ship arrived on a small island.
	(c) He landed the plane safely.	(c) The captain landed the ship on a small island.
	(d) The plane landed safely.	(d) The ship landed on a small island.
3	(a) I am glad he came the helicopter in time.*	(a) The pilot came the plane ten minutes ahead of schedule.*
	(b) I am glad the helicopter came in time.	(b) The plane came ten minutes ahead of schedule.
	(c) I am glad he landed the helicopter in time.	(c) The pilot landed the plane ten minutes ahead of schedule.
	(d) I am glad the helicopter landed in time.	(d) The plane landed ten minutes ahead of schedule.
4	(a) Tom donated the hospital a large amount of money.*	(a) Charles donated the library some very valuable books.*
	(b) Tom donated a large amount of money to the hospital.	(b) Charles donated some very valuable books to the library.
	(c) Tom gave the hospital a large amount of money.	(c) Charles gave the library some very valuable books.
	(d) Tom gave a large amount of money to the hospital.	(d) Charles gave some very valuable books to the library.
5	(a) He fell the crumbs on the ground.*	(a) During the earthquake, he fell the wine glass.*
	(b) The crumbs fell on the ground.	(b) During the earthquake, the wine glass fell from his hands.
	(c) He dropped the crumbs on the ground.	(c) During the earthquake, he dropped the wine glass.
	(d) The crumbs dropped on the ground.	(d) During the earthquake, the wine glass dropped from his hands.
6	(a) She disappeared her money behind the curtains.*	(a) Cathy was able to disappear her anger.*
	(b) She disappeared behind the curtains.	(b) Cathy was able to disappear.
	(c) She hid behind the curtains.	(c) Cathy was able to hide her anger.
	(d) She hid her money behind the curtains.	(d) Cathy was able to hide.
7	(a) That is an unusual object What's it?*	(a) I know something is bothering you. What's it?*
	(b) That is an unusual object. What is it?	(b) I know something is bothering you. What is it?
	(c) That is an unusual object. What's it used for?	(c) I know something is bothering you. What's wrong?
	(d) That is an unusual object. What is it used for?	(d) I know something is bothering you. What is wrong?
8	(a) I poured the truck with gravel.*	(a) Kate poured the fish tank with pebbles.*
	(b) I poured the gravel into the truck.	(b) Kate poured the pebbles into the fish tank.
	(c) I loaded the truck with gravel.	(c) Kate loaded the fish tank with pebbles.
	(d) I loaded the gravel into the truck.	(d) Kate loaded the pebbles into the fish tank.
9	(a) He vanished the treasure inside a cave.*	(a) Ben vanished his toys behind the door.*
	(b) He vanished inside a cave.	(b) Ben vanished behind the door.
	(c) He hid the treasure inside a cave.	(c) Ben hid his toys behind the door.
	(d) He hid inside a cave.	(d) Ben hid behind the door.
10	(a) Susan created her daughter a new dress.*	(a) The chef will create you a special dish.*
	(b) Susan created a new dress for her daughter.	(b) The chef will create a special dish for you.
	(c) Susan made her daughter a new dress.	(c) The chef will make you a special dish.
	(d) Susan made a new dress for her daughter.	(d) The chef will make a special dish for you.
11	(a) Someone is at the door. Who's it?*	(a) I heard you found a date for the party. Who's it?*
	(b) Someone is at the door. Who's there?	(b) I heard you found a date for the party. Who's the lucky girl?
	(c) Someone is at the door. Who is it?	(c) I heard you found a date for the party. Who is it?
	(d) Someone is at the door. Who is there?	(d) I heard you found a date for the party. Who is the lucky girl?
12	(a) Dan is gonna a picnic in the park.*	(a) My sister is gonna Disneyland for her school trip.*
	(b) Dan is going to a picnic in the park.	(b) My sister is going to Disneyland for her school trip.
	(c) Dan is gonna attend a picnic in the park.	(c) My sister is gonna be at Disneyland for her school trip.
	(d) Dan is going to attend a picnic in the park.	d) My sister is going to be at Disneyland for her school trip.
13	a) James suggested his manager the plan.*	a) Rachel suggested the client her new idea.*
	b) James suggested the plan to his manager.	b) Rachel suggested her new idea to the client.
	c) James told his manager the plan.	c) Rachel told the client her new idea.
	d) James told the plan to his manager.	d) Rachel told her new idea to the client.
14	(a) That is an unusual object. What's it?*	(a) Who do you think that is meeting with the boss?*
	(b) That is an unusual object. What is it?	(b) Who do you think is meeting with the boss?
	(c) That is an unusual object. What's it used for?	(c) Who do you think that the boss is meeting with?
	(d) That is an unusual object. What is it used for?	(d) Who do you think the boss is meeting with?
15	(a) Which player does does Richard wanna win?*	(a) Who do you wanna win the championships?*