

# Negotiating lexical uncertainty and expertise with disjunction

Roger Levy and Christopher Potts



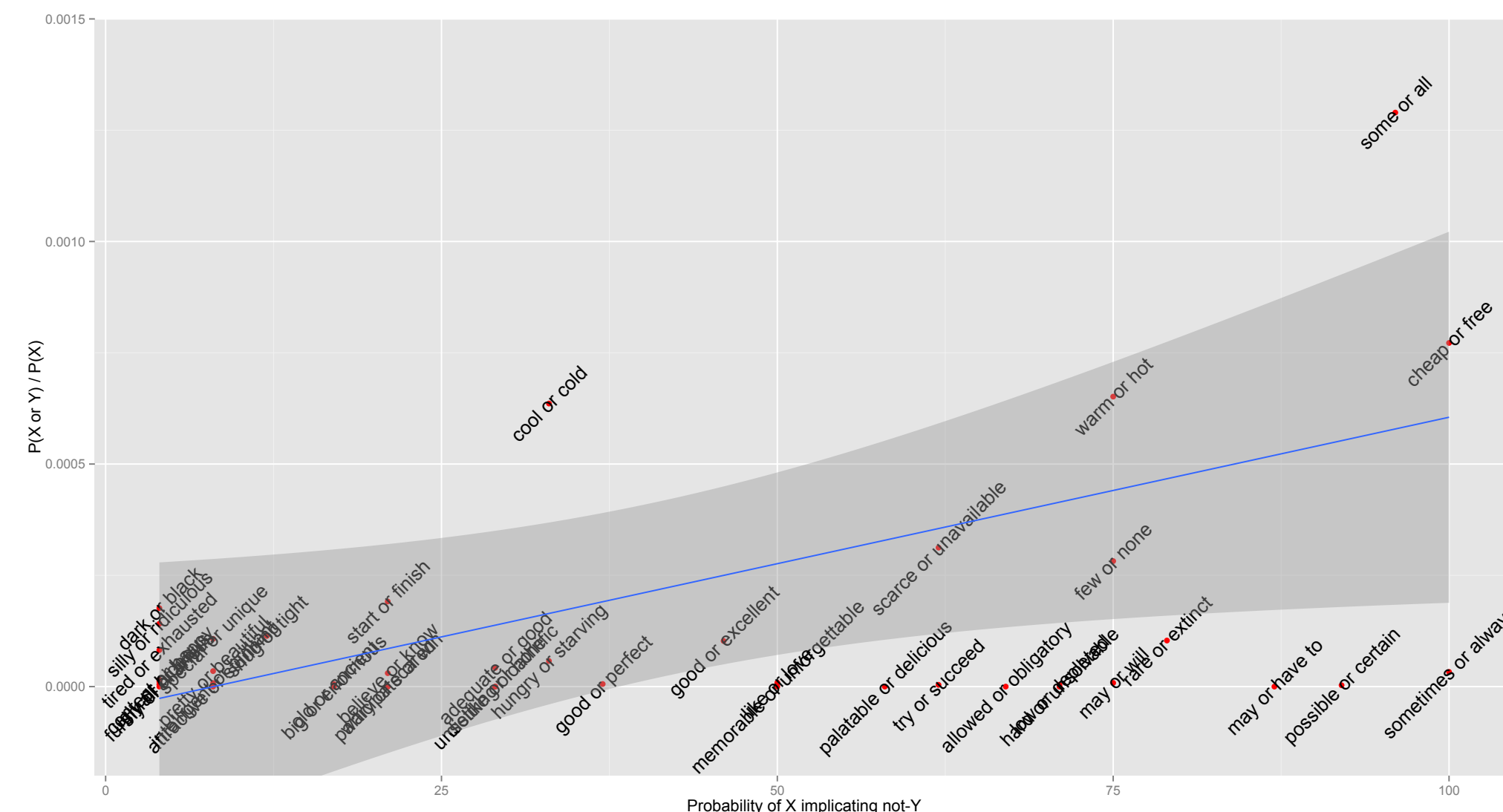
## COMMUNICATING IN LANGUAGE ABOUT LANGUAGE

- Languages are neither fixed across time nor identically reproduced in all speakers, but rather continually renegotiated during interactions.
- People accommodate to each other's usage patterns, form temporarily lexical pacts, and instruct each other about their linguistic views.
- Some of this communication in language about language is direct, as with explicit definitions, but much of it arrives via secondary pragmatic inferences.
- Disjunction supports what appear to be opposing inferences about language.
  - Hurfordian pressure:**  $X$  or  $Y$  conveys that  $X$  and  $Y$  are disjoint
  - Definitional inference:**  $X$  or  $Y$  conveys that  $X$  and  $Y$  are synonymous
- This pattern is cross-linguistically robust, so we seek a single pragmatic model that can derive both of these meanings from the semantics of disjunction given different contextual assumptions.

## HURFORDIAN PERCEPTIONS AND INTENTIONS

**Generalization:**  $X$  or  $Y$  conveys that the speaker is using a lexicon where  $X$  and  $Y$  are disjoint, or addresses a speaker concern that the listener is using such a lexicon.

- the nuptials will take place in either **France** or **Paris**
- the **canoe** or **boat** will be held by the stream's current
- In 1940, 37% of us had gone to a **church** or **synagogue** in the last week.



$X$  or  $Y$  usage correlates with  $X$  implicating *not*  $Y$

Our corpus	
'general or specific'	75
'specific or general'	86

## DISJUNCTIVE DEFINITION AND IDENTIFICATION

**Generalization:**  $X$  or  $Y$  can convey  $\llbracket X \rrbracket = \llbracket Y \rrbracket$  when the speaker is mutually, publicly known to be an expert or would like to establish expertise.

- wine lover or *oenophile*
- A Geological History of Manhattan or New York Island
- New Haven or "the Elm City"
- woodchuck or "land beaver"

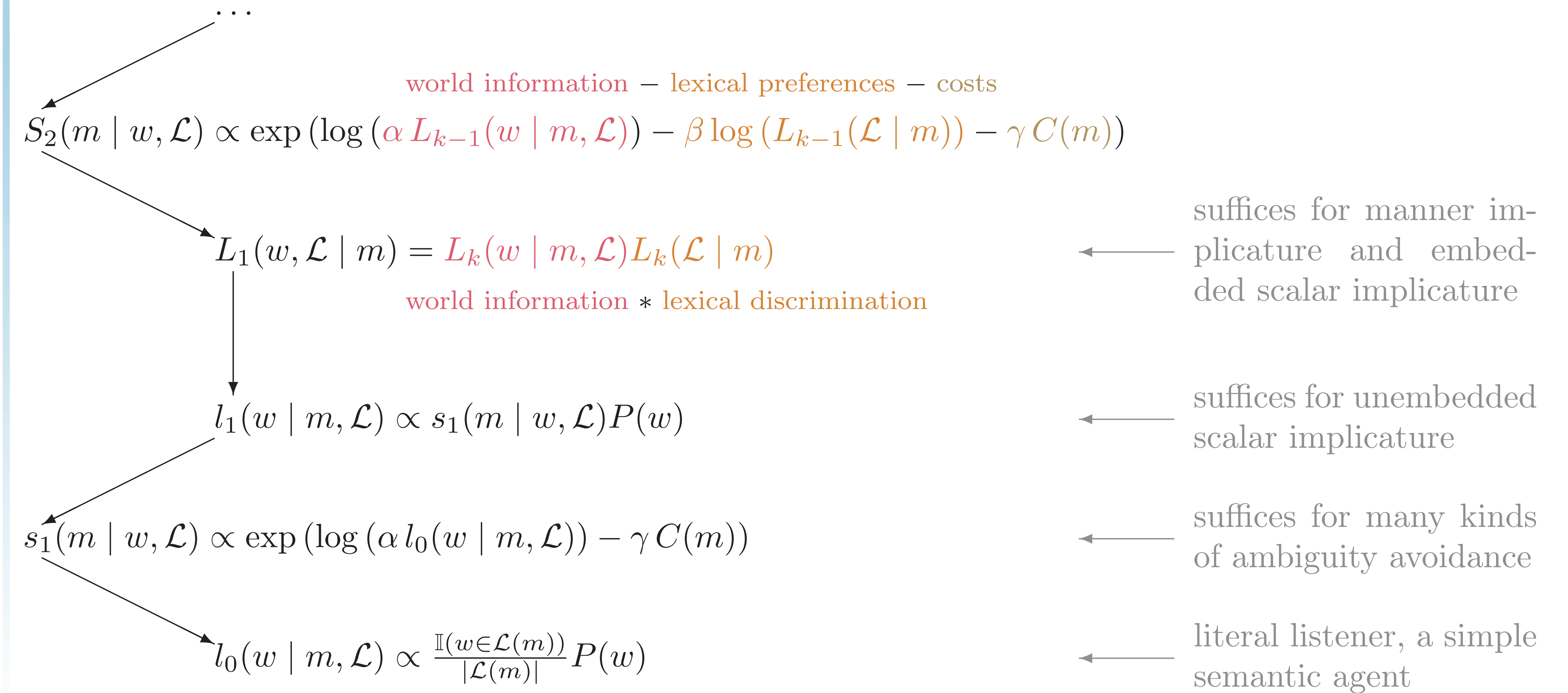
- Motivation: speaker is a known 'instructor'; listener is a known non-expert
- Motivation: speaker wishes to display expertise to another expert
- Motivation: speaker sees value in (temporarily or permanently) defining a term

Attested in Chinese, German, Hebrew, Ilokano, Japanese, Russian, and Tagalog. Seems to survive even where the language has a dedicated definitional disjunction morpheme (e.g., Finnish, Italian).

## FURTHER INFORMATION

Paper, references, model code, corpus data: <http://github.com/cgpotts/pypragmods/>

## MODELING COMMUNICATION WITH ANXIOUS EXPERTS



## DEFINITIONAL CONTEXTS

Require low disjunction costs and high  $\beta$ : the speaker is invested in communicating about the lexicon and can tolerate the cost of a disjunction that is synonymous with one of its disjuncts.

$L_2$ hears $A$ or $X$	$w_1$	$w_2$	$w_1 \vee w_2$
$\mathcal{L}^*[A: \{w_1\}, B: \{w_2\}, X: \{w_1, w_2\}]$	0	0	.08
$\mathcal{L}_1[A: \{w_1\}, B: \{w_2\}, X: \{w_2\}]$	.07	0	.08
$\mathcal{L}_2[A: \{w_1\}, B: \{w_2\}, X: \{w_1\}]$	.77	0	.06

$$\alpha = 5; \beta = 7; C(or) = .01$$

<div><math>S_2</math> observes <math>\langle \mathcal{L}_2, w_1 \rangle</math></div>																															
<div><math>A</math> 0 <math>X</math> 0 <math>A \text{ or } X</math> .05</div>																															
<div><math>\downarrow</math></div>																															
<div><table><tr><td colspan="4"><math>L_1</math> hears <math>A \text{ or } X</math></td><td><math>w_1</math></td><td><math>w_2</math></td><td><math>w_1 \vee w_2</math></td></tr><tr><td><math>\mathcal{L}^*</math></td><td><math>A: \{w_1\}, B: \{w_2\}, X: \{w_1, w_2\}</math></td><td>0</td><td>0</td><td>.23</td></tr><tr><td><math>\mathcal{L}_1</math></td><td><math>A: \{w_1\}, B: \{w_2\}, X: \{w_2\}</math></td><td>0</td><td>0</td><td>.38</td></tr><tr><td><math>\mathcal{L}_2</math></td><td><math>A: \{w_1\}, B: \{w_2\}, X: \{w_1\}</math></td><td>.38</td><td>0</td><td>0</td></tr></table></div>										$L_1$ hears $A \text{ or } X$				$w_1$	$w_2$	$w_1 \vee w_2$	$\mathcal{L}^*$	$A: \{w_1\}, B: \{w_2\}, X: \{w_1, w_2\}$	0	0	.23	$\mathcal{L}_1$	$A: \{w_1\}, B: \{w_2\}, X: \{w_2\}$	0	0	.38	$\mathcal{L}_2$	$A: \{w_1\}, B: \{w_2\}, X: \{w_1\}$	.38	0	0
$L_1$ hears $A \text{ or } X$				$w_1$	$w_2$	$w_1 \vee w_2$																									
$\mathcal{L}^*$	$A: \{w_1\}, B: \{w_2\}, X: \{w_1, w_2\}$	0	0	.23																											
$\mathcal{L}_1$	$A: \{w_1\}, B: \{w_2\}, X: \{w_2\}$	0	0	.38																											
$\mathcal{L}_2$	$A: \{w_1\}, B: \{w_2\}, X: \{w_1\}$	.38	0	0																											
<div><div><div><math>\swarrow</math></div><div><math>\downarrow</math></div><div><math>\searrow</math></div></div></div>																															
$l_1$	$\mathcal{L}^*$	$w_1$	$w_2$	$w_1 \vee w_2$	$\mathcal{L}_1$	$w_1$	$w_2$	$w_1 \vee w_2$	$\mathcal{L}_2$	$w_1$	$w_2$	$w_1 \vee w_2$																			
	$A$	1	0	0	$A$	1	0	0	$A$	1	0	0																			
	$X$	.02	.02	.96	$X$	0	1	0	$X$	1	0	0																			
	$A \text{ or } X$	.02	.02	.96	$A \text{ or } X$	.01	0	.98	$A \text{ or } X$	1	0	0																			
<div><div><math>\downarrow</math></div><div><math>\downarrow</math></div><div><math>\downarrow</math></div></div>																															
$s_1$	$\mathcal{L}^*$	$A$	$X$	$A \text{ or } X$	$\mathcal{L}_1$	$A$	$X$	$A \text{ or } X$	$\mathcal{L}_2$	$A$	$X$	$A \text{ or } X$																			
	$w_1$	.98	0	0	$w_1$	.99	0	0	$w_1$	.33	0	0																			
	$w_2$	0	0	.2	$w_2$	0	.33	0	$w_2$	.33	0	0																			
	$w_1 \vee w_2$	0	0	.2	$w_1 \vee w_2$	0	0	.33	$w_1 \vee w_2$	.33	0	0																			
<div><div><math>\downarrow</math></div><div><math>\downarrow</math></div><div><math>\downarrow</math></div></div>																															
$l_0$	$\mathcal{L}^*$	$w_1$	$w_2$	$w_1 \vee w_2$	$\mathcal{L}_1$	$w_1$	$w_2$	$w_1 \vee w_2$	$\mathcal{L}_2$	$w_1$	$w_2$	$w_1 \vee w_2$																			
	$A$	1	0	0	$A$	1	0	0	$A$	1	0	0																			
	$X$	.33	.33	.33	$X$	0	1	0	$X$	1	0	0																			
	$A \text{ or } X$	.33	.33	.33	$A \text{ or } X$	.33	.33	.33	$A \text{ or } X$	1	0	0																			

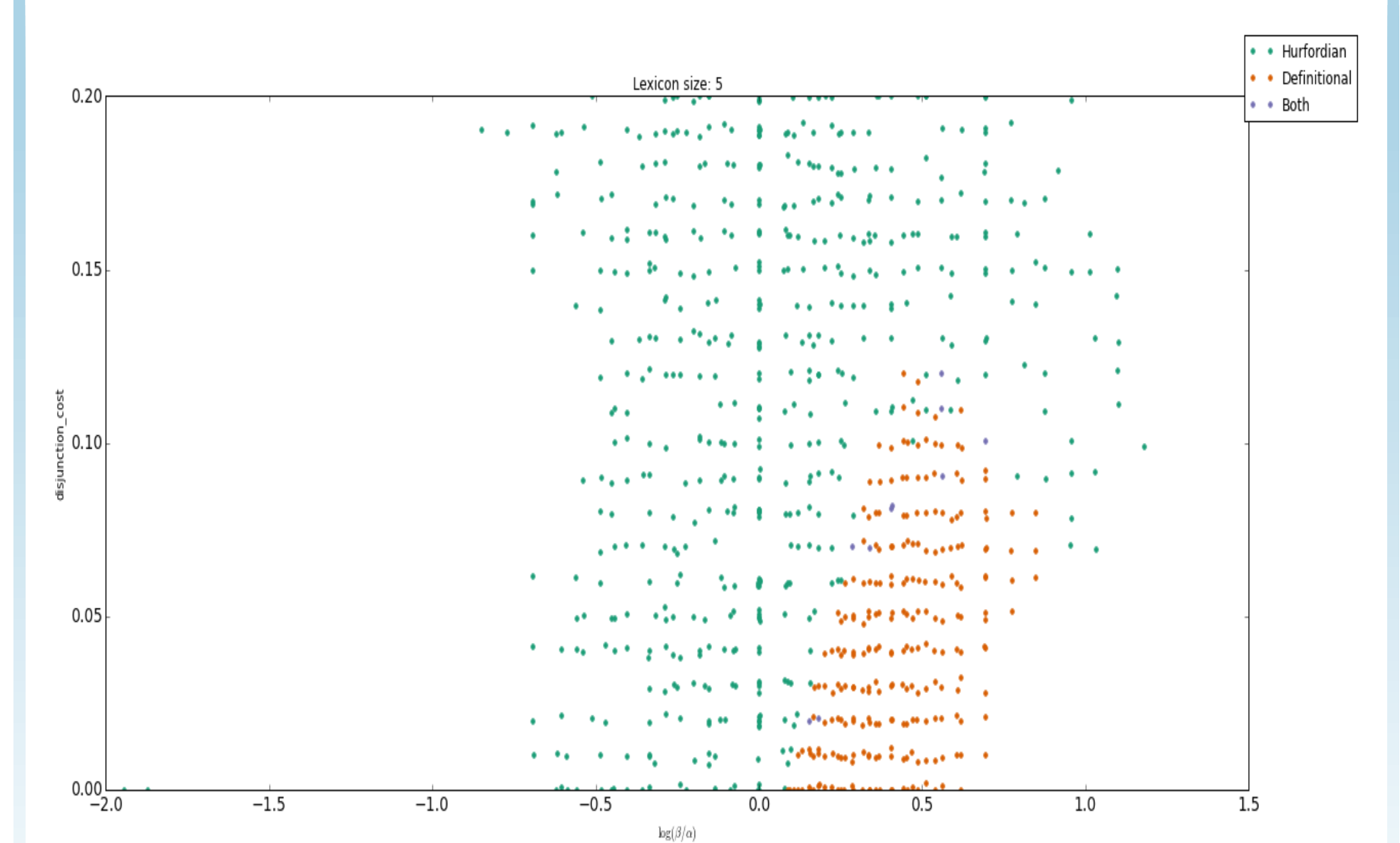
## HURFORDIAN CONTEXTS

With high disjunction costs, exclusivization maximizes the justification for the long form; the Hurfordian instinct is a rational response to a disjunction that is unduly prolix for many lexica.

$L_2$ hears $A$ or $X$	$w_1$	$w_2$	$w_1 \vee w_2$
$\mathcal{L}^*[A: \{w_1\}, B: \{w_2\}, X: \{w_1, w_2\}]$	.03	0	.14
$\mathcal{L}_1[A: \{w_1\}, B: \{w_2\}, X: \{w_2\}]$	.04	0	.45
$\mathcal{L}_2[A: \{w_1\}, B: \{w_2\}, X: \{w_1\}]$	.02	0	.32

$\alpha = 2; \beta = 1; C(or) = 1$

## CHARACTERIZATION



Summarizes a search over many parameter settings using a large lexicon and large world space.