Modeling the learning of the Person Case Constraint

Adam Liter[1] Naomi H. Feldman[1,2]

[1]Department of Linguistics, University of Maryland

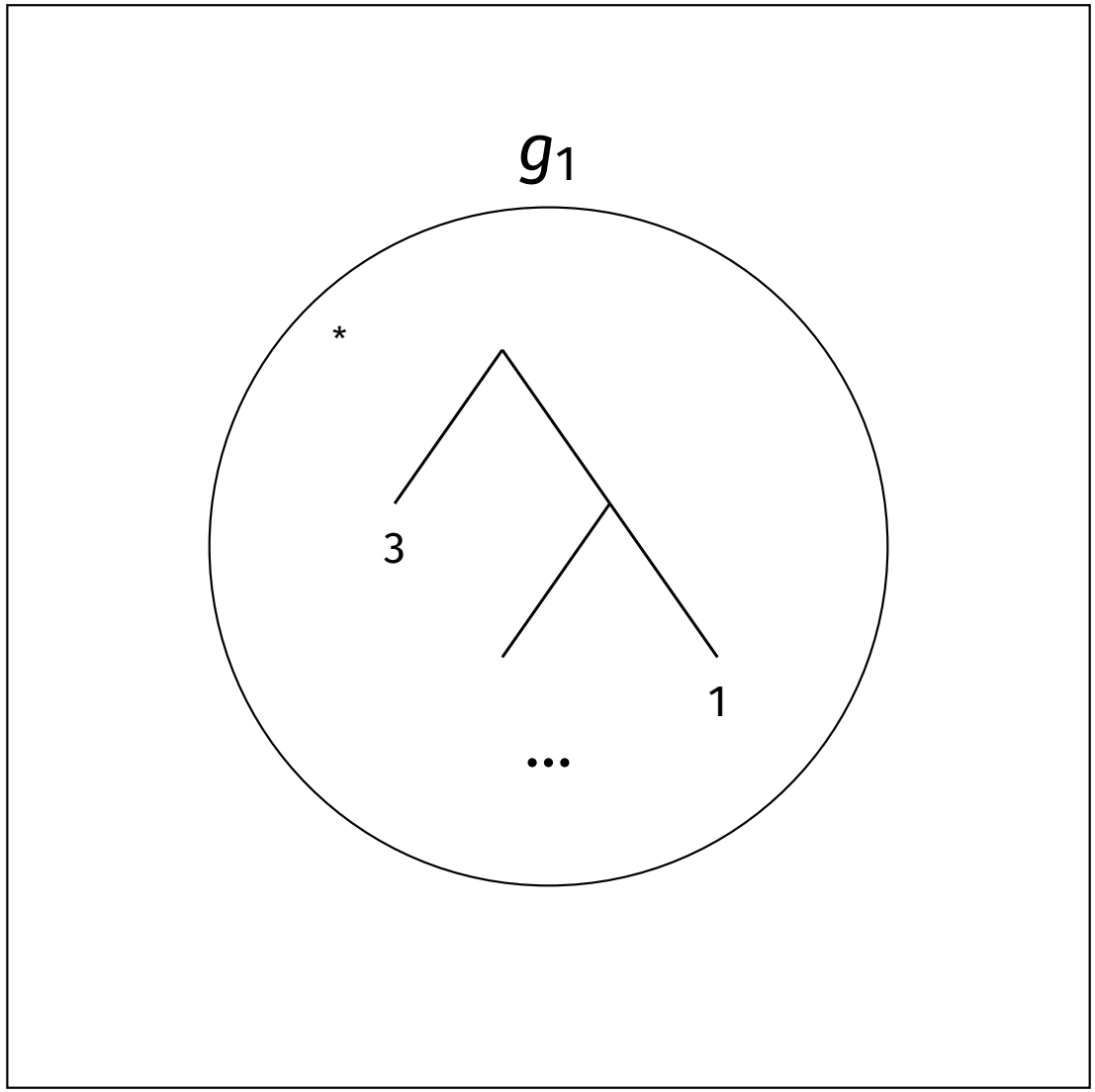
[2]UMIACS, University of Maryland

SCiL
January 2, 2020

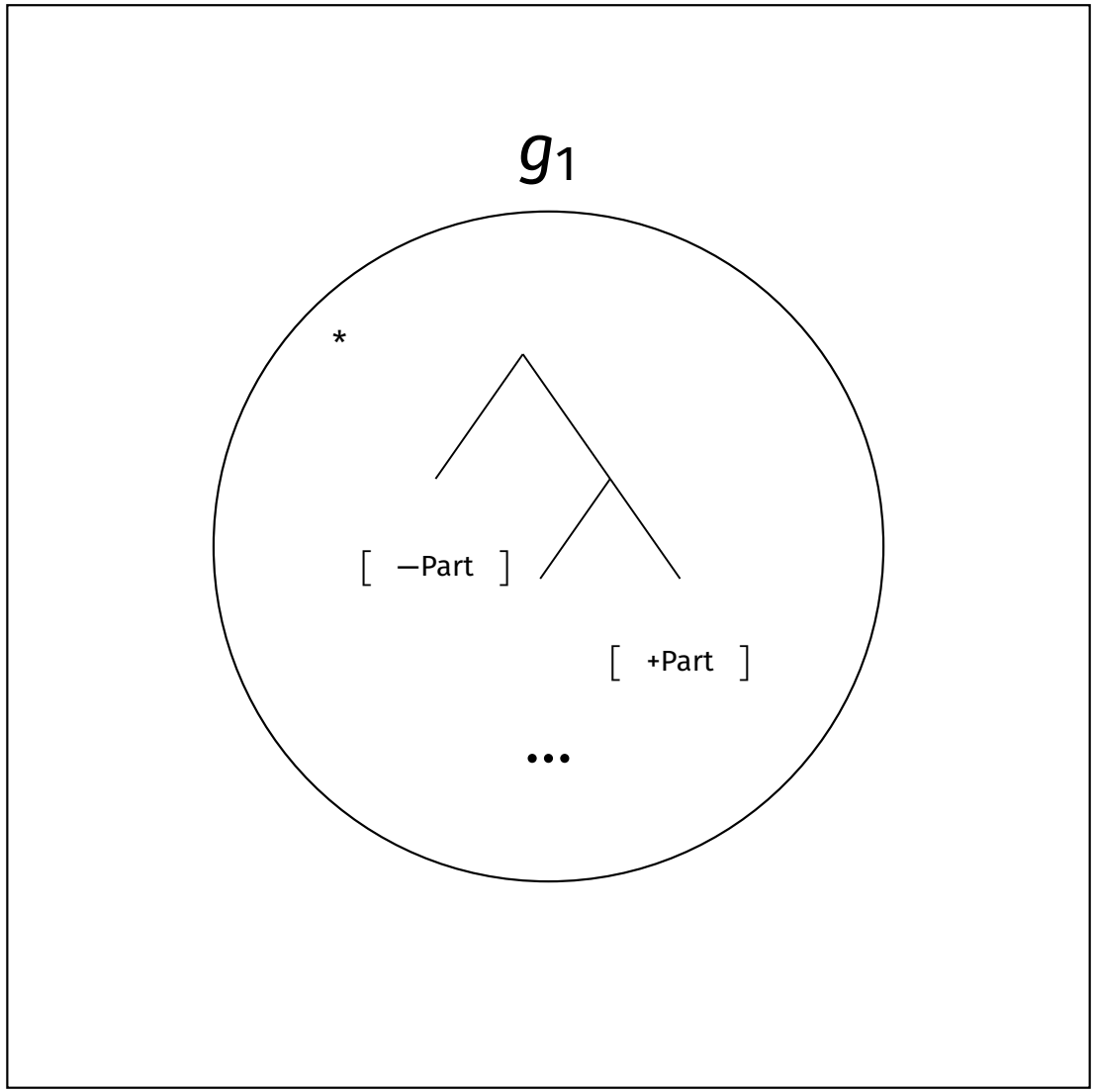
Grammars and their representations

(1) * Me le recommendó
1.SG.ACC 3.SG.DAT recommend.PST
'S/he recommended me to her/him'

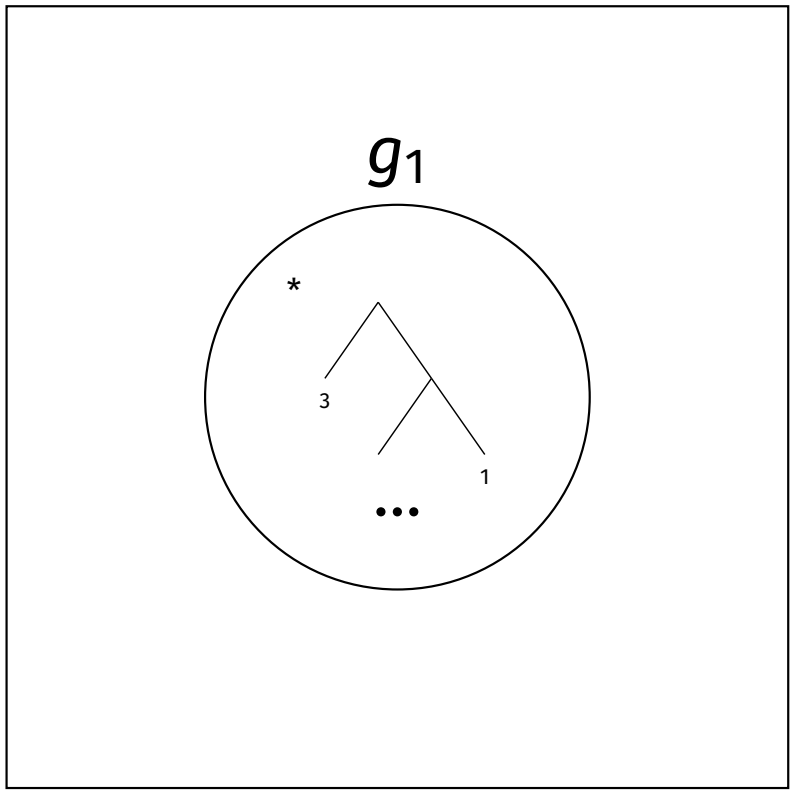
Grammars and their representations



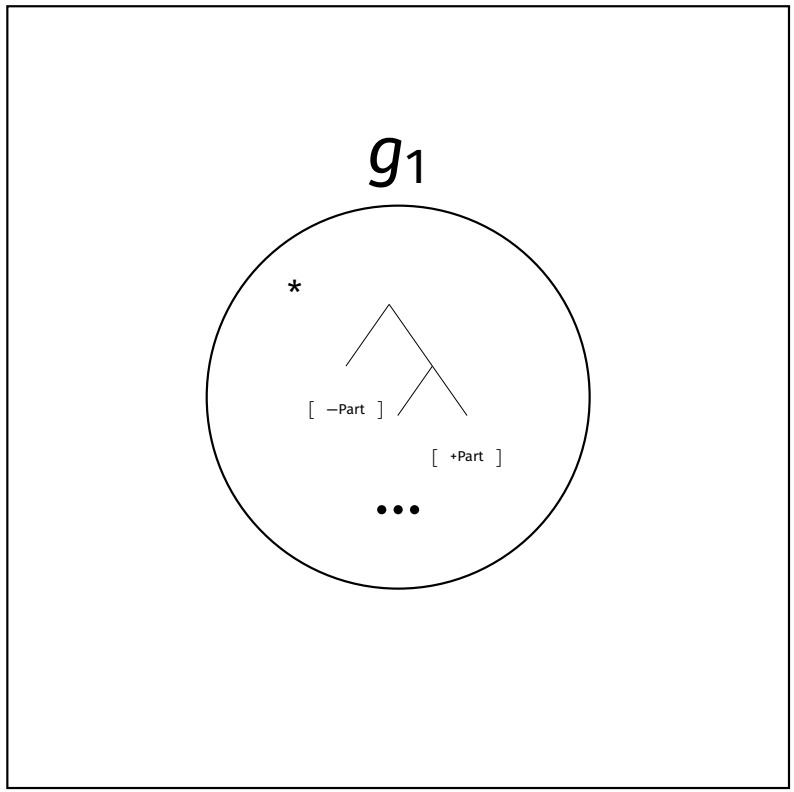
Grammars and their representations



Grammars and their representations

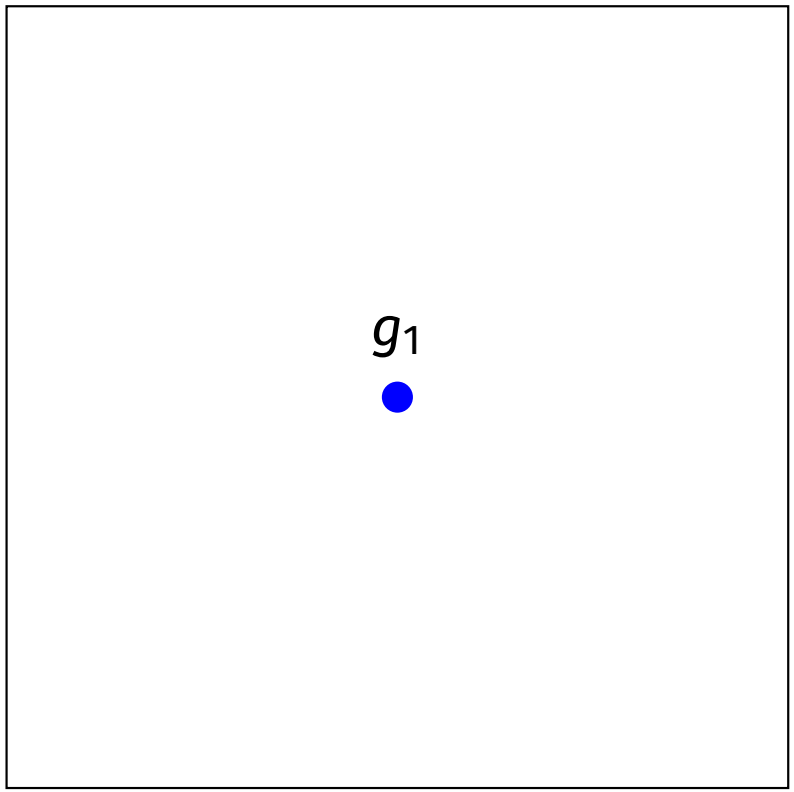


H_1

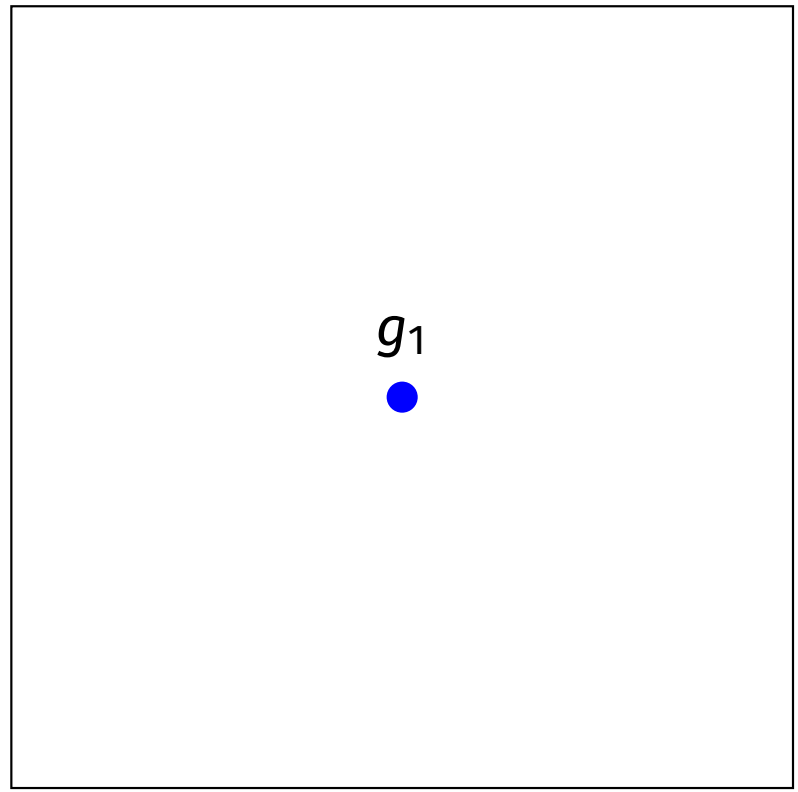


H_2

Grammars and their representations

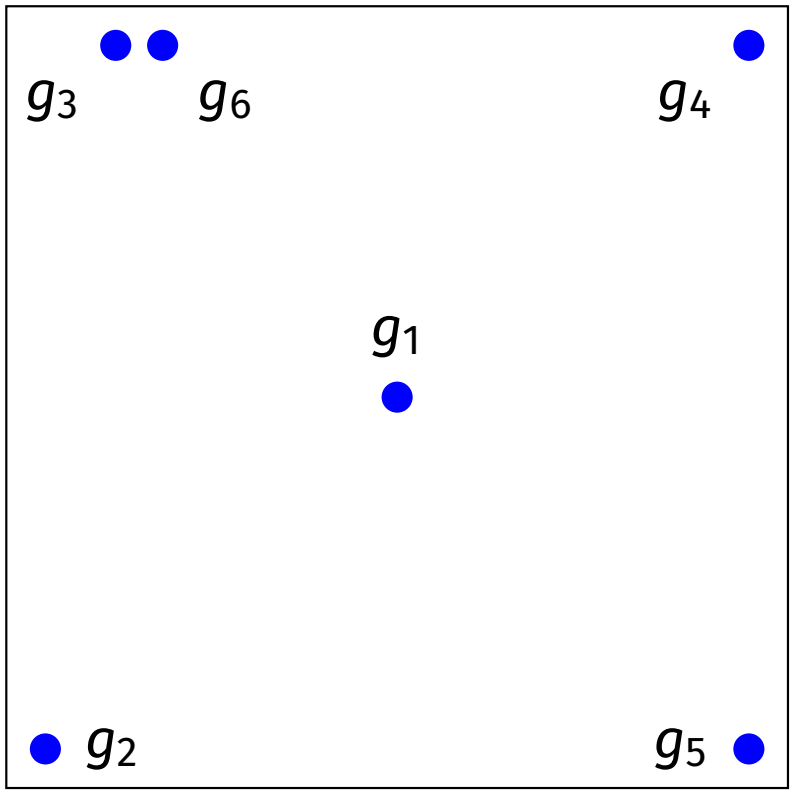


H_1

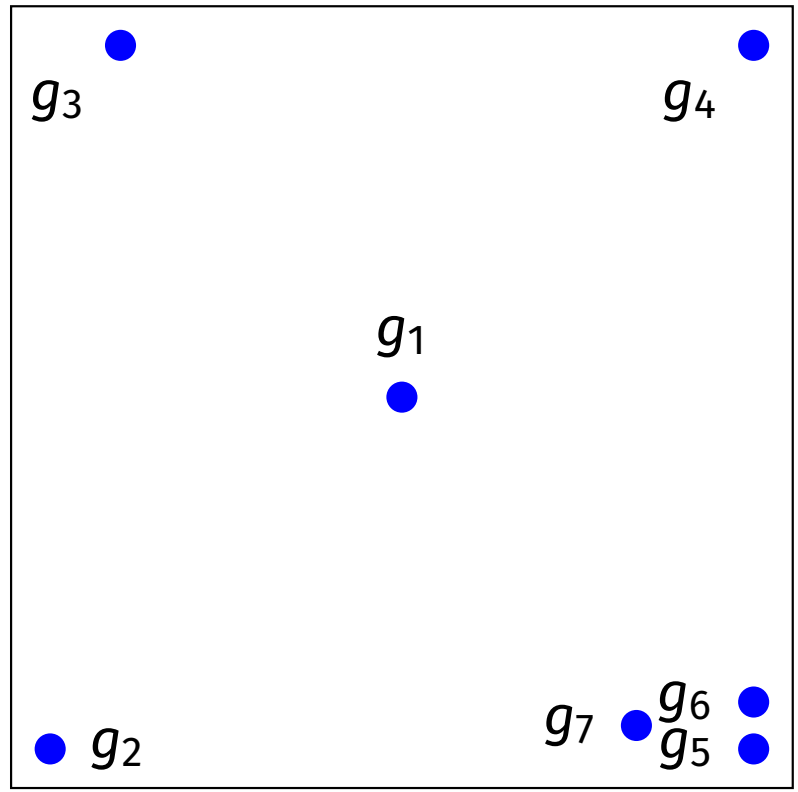


H_2

Grammars and their representations

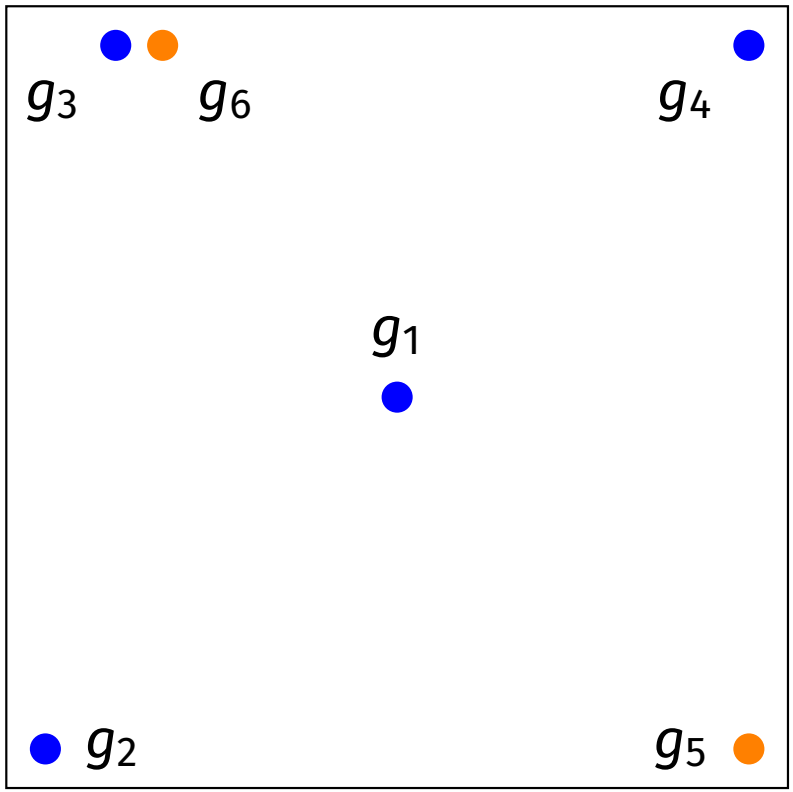


H_1

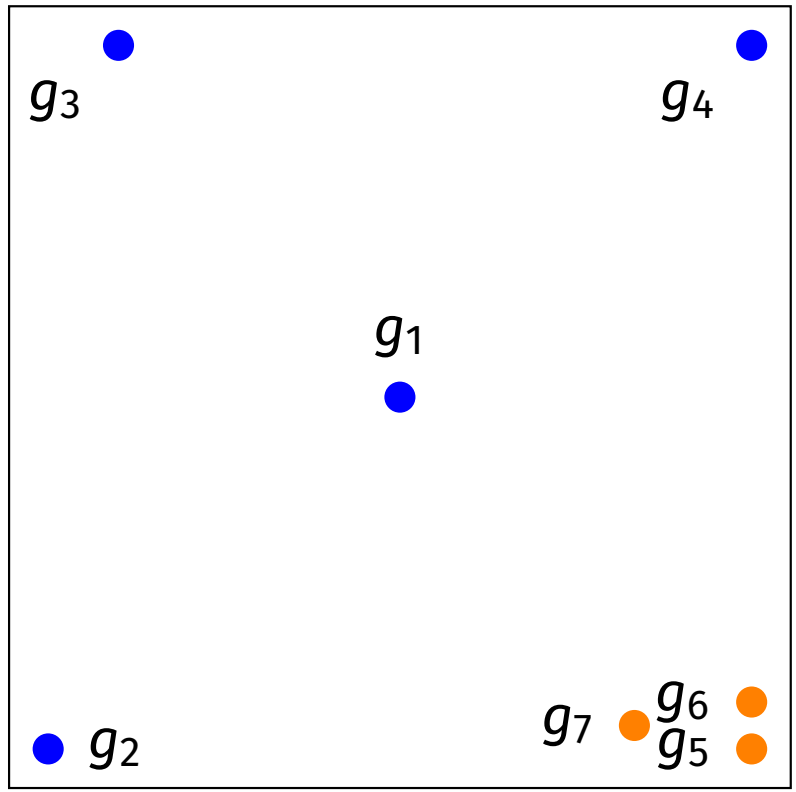


H_2

Grammars and their representations



H_1



H_2

Parsimony and acquisition implications

- "this kind of learnability evaluation [...] does *not* care about whether a [...] theory is appropriately restrictive or economical" (Pearl et al. 2017: 312).
- For other similar approaches, see also Pearl & Sprouse (2013), Rasin & Katzir (2017), Pearl & Sprouse (2019).

Roadmap

Clitics

- Clitics are bound morphemes (*i.e.*, affixal morphemes).

(2) Me lo cuentas
1.SG.DAT 3.SG.ACC tell
'(You) tell it to me'

- Direct object = *lo*, indirect object = *me*

Restrictions on clitics

- When two arguments are realized as clitics, not all combinations are possible.

(1) * Me le recommendó
1.SG.ACC 3.SG.DAT recommend.PST
'S/he recommended me to her/him'

- Direct object = *me*, indirect object = *le*

The Person Case Constraint (PCC)

- These sorts of restrictions are part of a broader phenomenon called the PCC (Bonet 1991, 1994).

IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

Table: Me-First PCC (Romanian, Spanish, *etc.*)

The Person Case Constraint (PCC)

- These sorts of restrictions are part of a broader phenomenon called the PCC (Bonet 1991, 1994).

IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

Table: Me-First PCC (Romanian, Spanish, *etc.*)

(2) Me lo cuentas
1.SG.DAT 3.SG.ACC tell
'(You) tell it to me'

The Person Case Constraint (PCC)

- These sorts of restrictions are part of a broader phenomenon called the PCC (Bonet 1991, 1994).

IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

Table: Me-First PCC (Romanian, Spanish, *etc.*)

(1) * Me le recommendó
1.SG.ACC 3.SG.DAT recommend.PST
'S/he recommended me to her/him'

The Person Case Constraint (PCC)

IO↓/DO→	1	2	3
1	NA	*	✓
2	*	NA	✓
3	*	*	✓

(a) Strong PCC (Greek, Spanish, *etc.*)

IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	*	✓

(b) Ultrastrong PCC (Classical Arabic, Spanish, *etc.*)

IO↓/DO→	1	2	3
1	NA	✓	✓
2	✓	NA	✓
3	*	*	✓

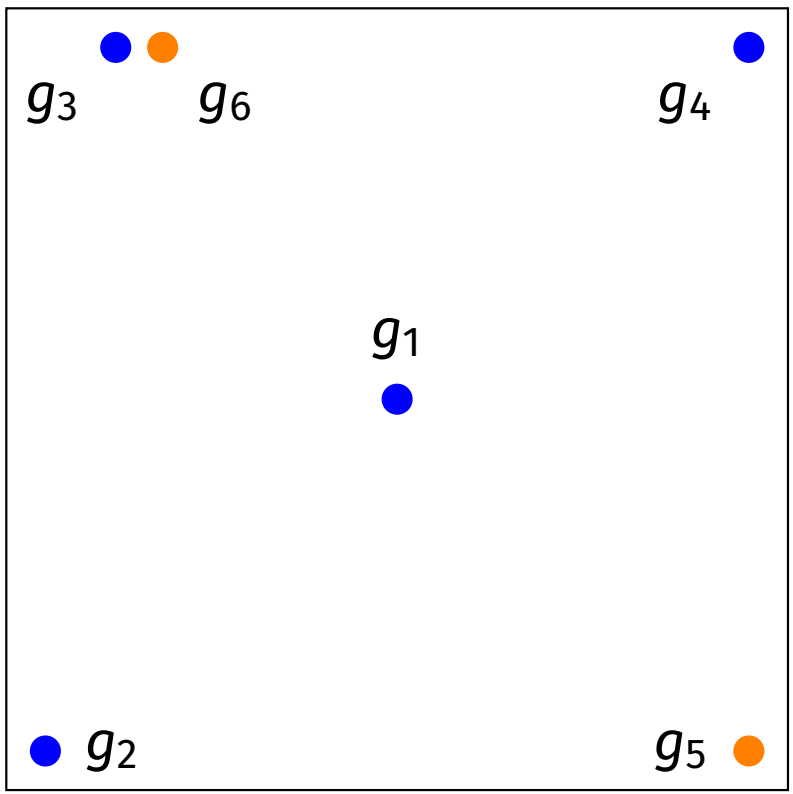
(c) Weak PCC (French, Catalan, Spanish, *etc.*)

IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

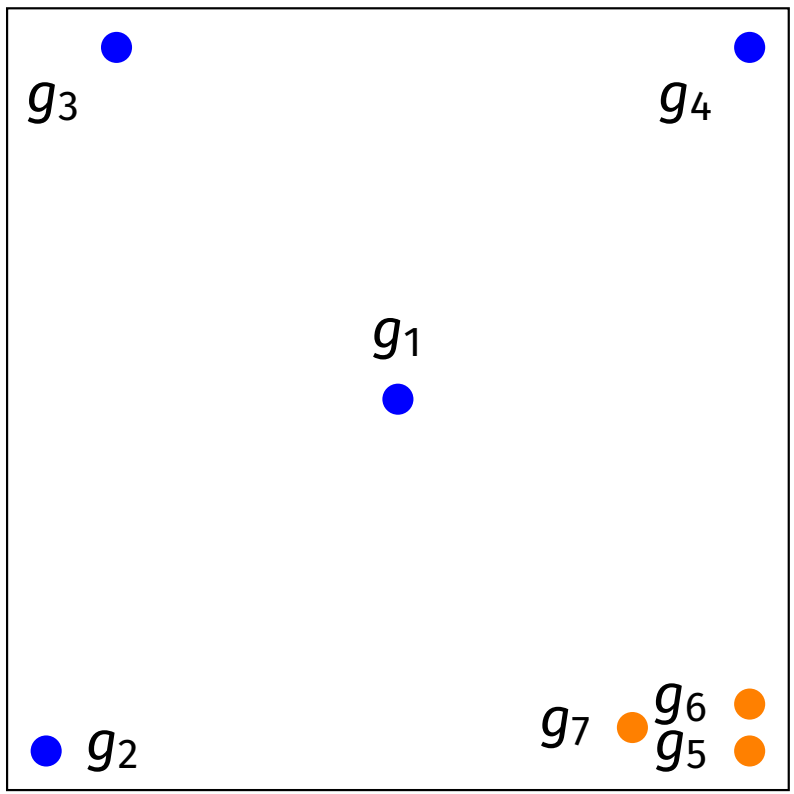
(d) Me-First PCC (Romanian, Spanish, *etc.*)

Learning evaluation

- Proof of concept that learning implications are useful for telling apart theories and their representations.



H_1



H_2

Evaluating two theories of the PCC

Simple theory

(3)
a. $1 = 1$
b. $2 = 2$
c. $3 = 3$

Feature-based theory

(4)
a. $1 = \begin{bmatrix} \text{+Auth} \\ \text{+Part} \end{bmatrix}$
b. $2 = \begin{bmatrix} \text{—Auth} \\ \text{+Part} \end{bmatrix}$
c. $3 = \begin{bmatrix} \text{—Auth} \\ \text{—Part} \end{bmatrix}$

A simple theory of the PCC

Grammar	1 2	1 3	2 1	2 3	3 1	3 2	3 3
SG_1	✓	✓	✓	✓	✓	✓	✓
SG_2	✓	✓	✓	✓	✓	✓	*
SG_3	✓	✓	✓	✓	✓	*	✓
SG_4	✓	✓	✓	✓	✓	*	*
SG_5	✓	✓	✓	✓	*	✓	✓
SG_6	✓	✓	✓	✓	*	✓	*
Weak PCC → SG_7	✓	✓	✓	✓	*	*	✓
SG_8	✓	✓	✓	✓	*	*	*
…	…	…	…	…	…	…	…
Me-First PCC → SG_{21}	✓	✓	*	✓	*	✓	✓
SG_{22}	✓	✓	*	✓	*	✓	*
Ultrastrong PCC → SG_{23}	✓	✓	*	✓	*	*	✓
…	…	…	…	…	…	…	…
Strong PCC → SG_{55}	*	✓	*	✓	*	*	✓
…	…	…	…	…	…	…	…
SG_{85}	*	✓	*	✓	*	✓	✓
SG_{86}	*	✓	*	✓	*	✓	*
SG_{87}	*	✓	*	✓	*	*	✓
…	…	…	…	…	…	…	…
SG_{128}	*	*	*	*	*	*	*

A feature-based theory of the PCC

Nevins (2007)

- Person features are feature bundles, consisting of two binary feature values.

(5)

a. $1 = \begin{bmatrix} \text{+Auth} \\ \text{+Part} \end{bmatrix}$

b. $2 = \begin{bmatrix} \text{—Auth} \\ \text{+Part} \end{bmatrix}$

c. $3 = \begin{bmatrix} \text{—Auth} \\ \text{—Part} \end{bmatrix}$

- PCC variants arise based on searching hierarchical syntactic representation for these features and restrictions on this search.

A feature-based theory of the PCC

Nevins (2007)

	Grammar	1 2	1 3	2 1	2 3	3 1	3 2	3 3
	FG_1	✓	✓	✓	✓	✓	✓	✓
Weak PCC →	FG_2	✓	✓	✓	✓	*	*	✓
Me-First PCC →	FG_3	✓	✓	*	✓	*	✓	✓
Ultrastrong PCC →	FG_4	✓	✓	*	✓	*	*	✓
	FG_5	*	*	*	*	*	*	✓
Strong PCC →	FG_6	*	✓	*	✓	*	*	✓
→	FG_7	*	✓	*	✓	*	*	✓
	FG_8	*	*	✓	*	✓	*	✓
	FG_9	*	*	*	*	*	*	✓

Evaluating two theories of the PCC

Simple theory

(6)
a. $1 = 1$
b. $2 = 2$
c. $3 = 3$

Feature-based theory

(7)
a. $1 = \begin{bmatrix} \text{+Auth} \\ \text{+Part} \end{bmatrix}$
b. $2 = \begin{bmatrix} \text{—Auth} \\ \text{+Part} \end{bmatrix}$
c. $3 = \begin{bmatrix} \text{—Auth} \\ \text{—Part} \end{bmatrix}$

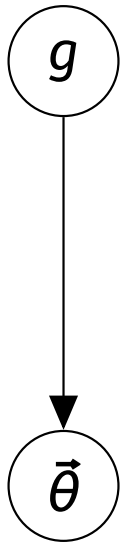
Bayesian learning model

The generative model

IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

Bayesian learning model

The generative model

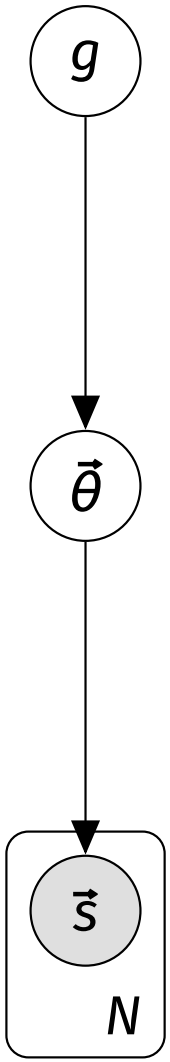


IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

IO↓/DO→	1	2	3
1	NA	0.0004	0.1852
2	0	NA	0.5475
3	0	0.0152	0.2518

Bayesian learning model

The generative model



IO↓/DO→	1	2	3
1	NA	✓	✓
2	*	NA	✓
3	*	✓	✓

IO↓/DO→	1	2	3
1	NA	0.0004	0.1852
2	0	NA	0.5475
3	0	0.0152	0.2518

IO↓/DO→	1	2	3
1	NA	0	50
2	0	NA	148
3	0	4	68

Bayesian learning model

Inferring the grammar

- Given $\vec{s}$, our learning model uses Bayes' rule to infer $p(g \mid \vec{s})$.
- In doing so, we integrate over $\vec{\theta}$; importantly, this leads to higher likelihoods for grammars that allow fewer clitic combinations (cf. Tenenbaum & Griffiths 2001).

Data for simulations

Aguirre (2003)

- Using the Aguirre Corpus (Aguirre 2003) from CHILDES (MacWhinney 2000), we estimated the frequency of clitic combinations in child-directed speech for a dialect of Spanish from Spain.
- 13,411 child-directed utterances extracted with `PyLangAcq` (Lee et al. 2016).
- Utterances parsed with `spaCy` (Honnibal & Montani 2017).
- 2% of utterances contained two clitics.
- Smoothing was applied to estimate $\vec{\theta}$ for simulations.

Data for simulations

Aguirre (2003)

- These are speakers of a Me-First PCC language.

IO↓/DO→	1	2	3
1	NA	0	50
2	0*	NA	148
3	0*	4	68

Data for simulations

Aguirre (2003)

- With smoothing:

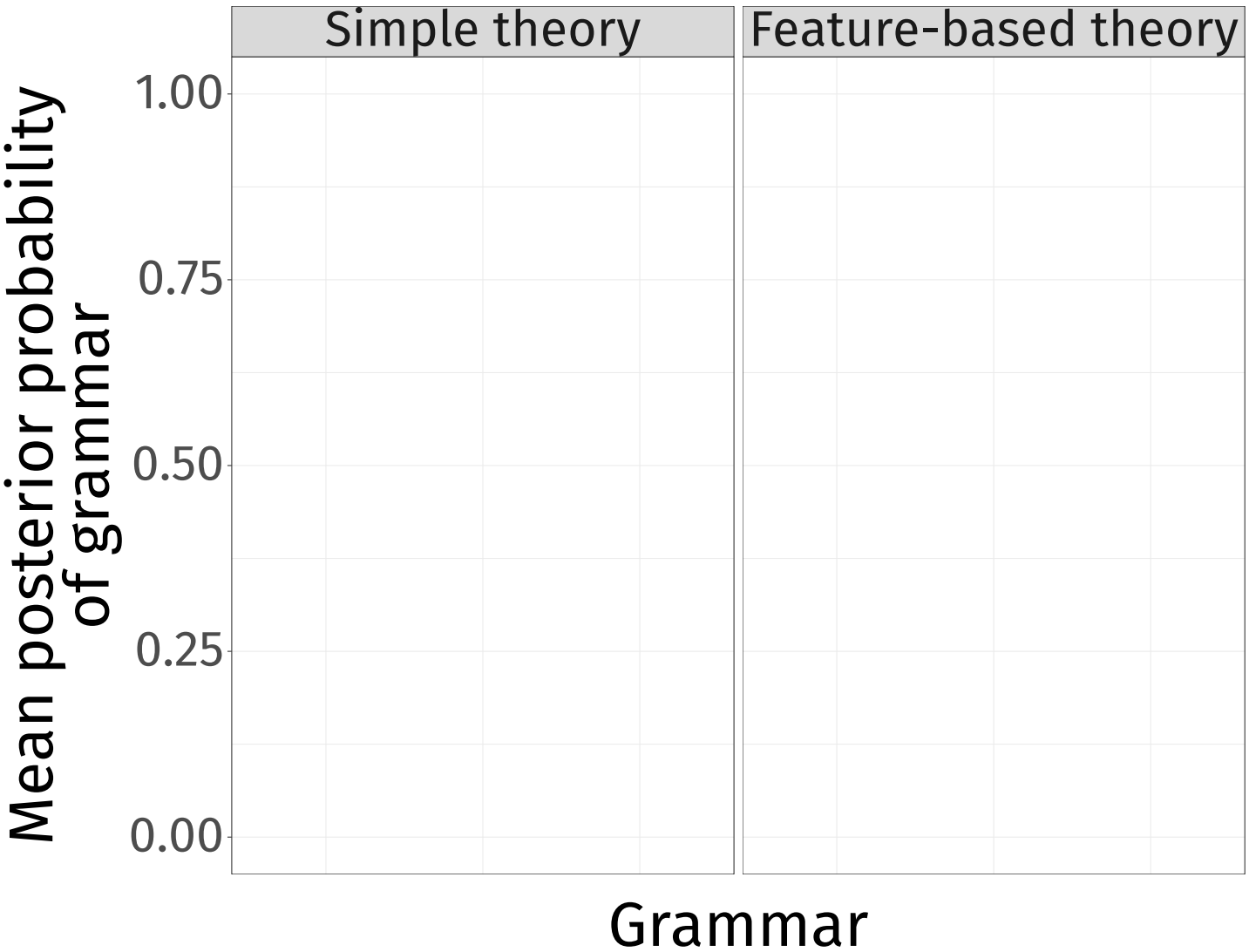
IO↓/DO→	1	2	3
1	NA	0.0004	0.1852
2	0	NA	0.5475
3	0	0.0152	0.2518

Simulations

- Hart & Risley (1995) estimate that children hear 1 million utterances in first 3 years of life, and $\approx 2\%$ of utterances contain two clitics.
- The extracted counts, with smoothing, were used to simulate corpora with n PCC constructions for $n = 66$, $n = 666$, and $n = 6,666$.
- We trained Simple learning models and Feature-based learning models with 1,000 replications for each corpus size.

Results

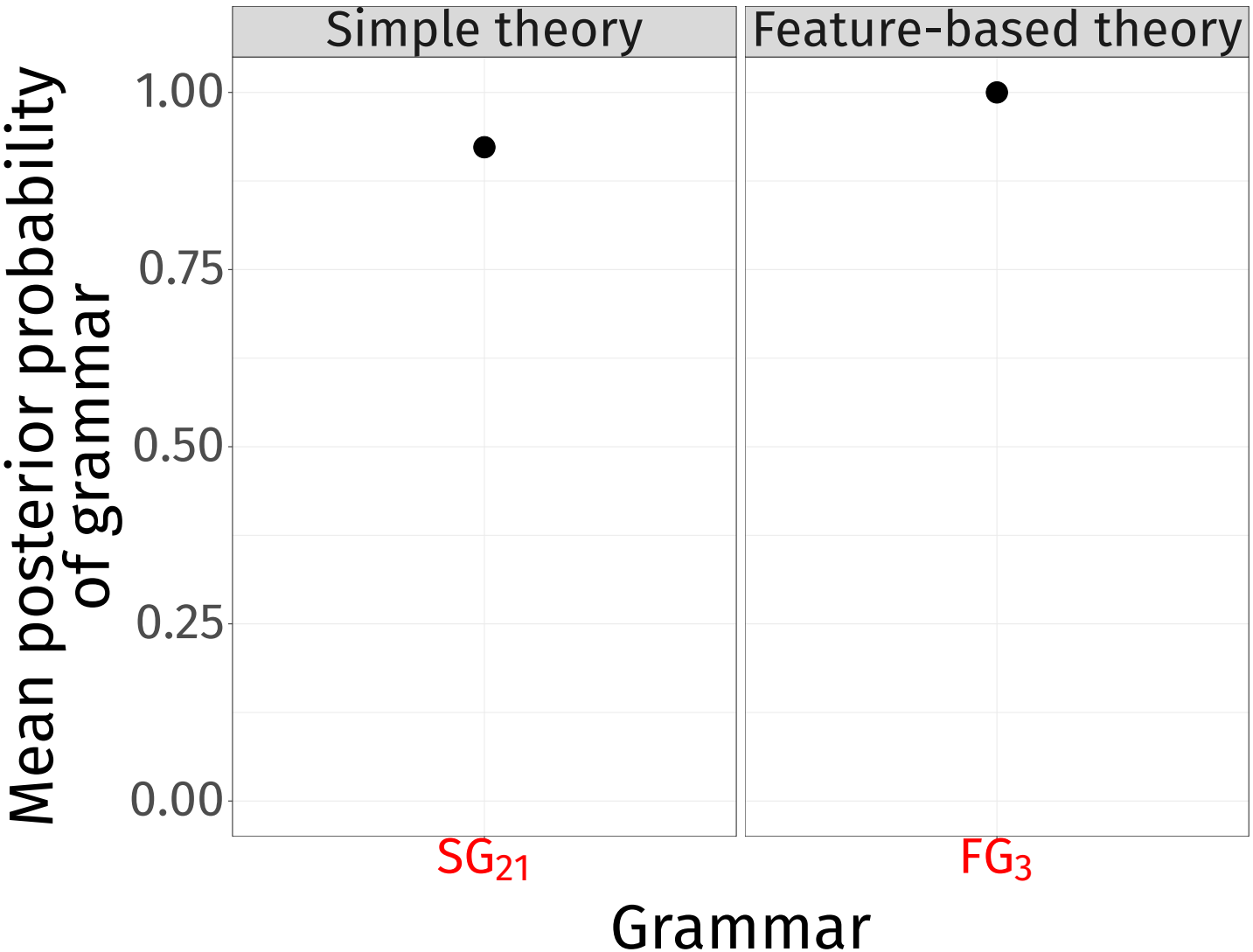
- red = target grammar; black = non-target grammar



Results

Corpuse size: 6,666

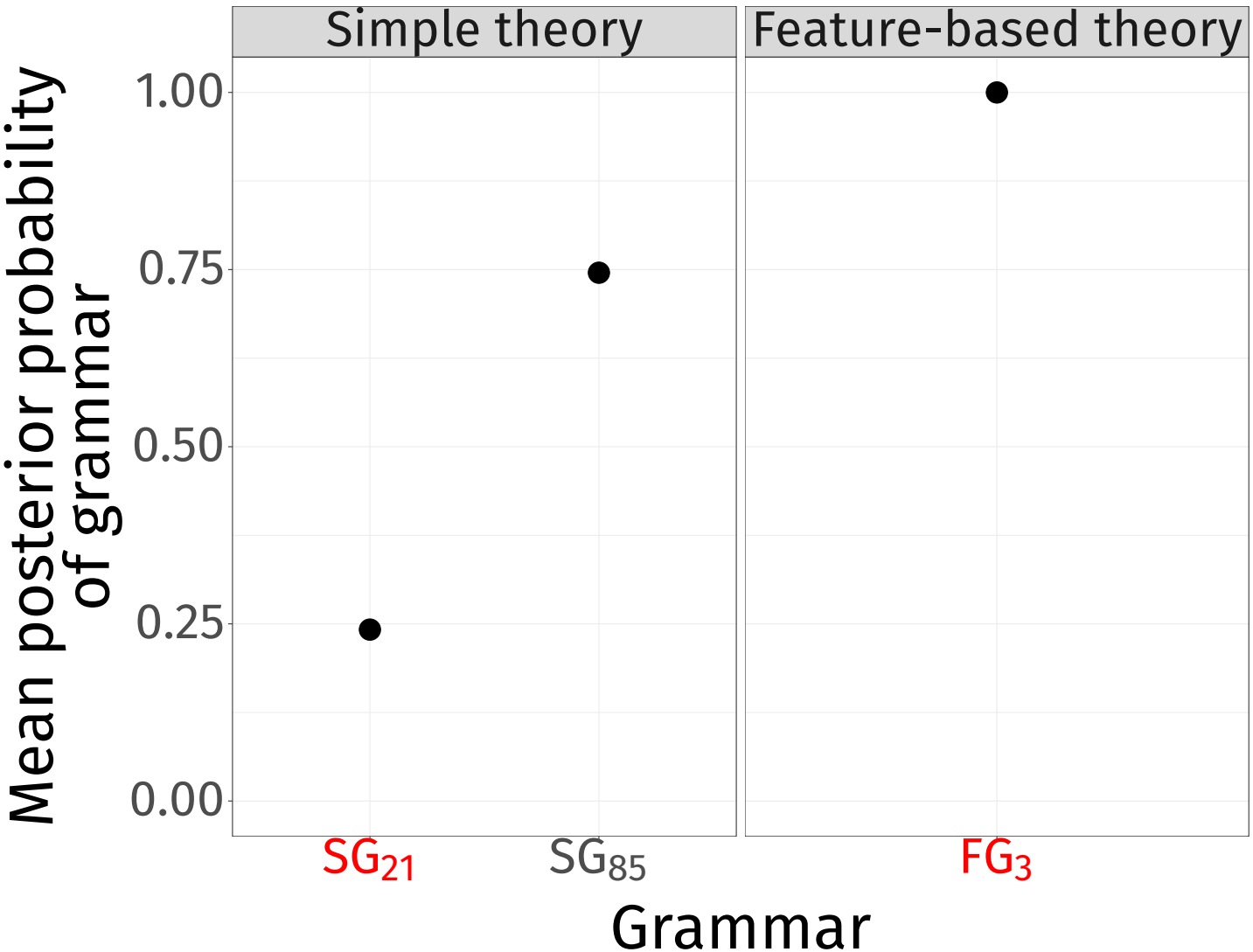
- red = target grammar; black = non-target grammar



Results

Corpuse size: 666

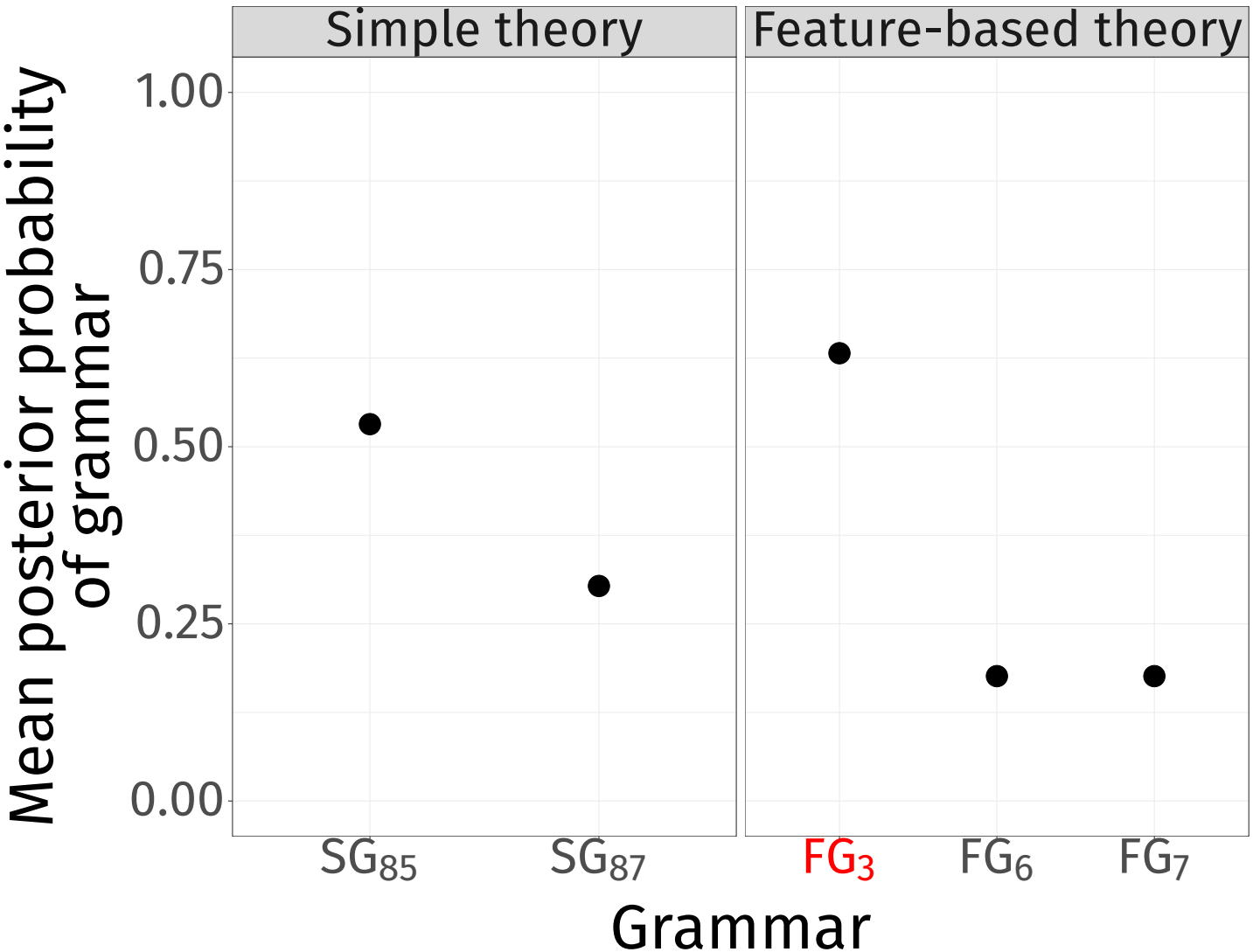
- red = target grammar; black = non-target grammar



Results

Corpuse size: 66

- red = target grammar; black = non-target grammar



Discussion

- Both class of hypotheses learn the target grammar, but the simple model is led astray when there is less data.
- Simple theory does have larger hypothesis space, but even with only 66 data points, each theory is only assigning probability to a few grammars (*i.e.*, not just about size of hypothesis space).

Discussion

- Both class of hypotheses learn the target grammar, but the simple model is led astray when there is less data.
- Simple theory does have larger hypothesis space, but even with only 66 data points, each theory is only assigning probability to a few grammars (*i.e.*, not just about size of hypothesis space).
- Modeling learning in this way can therefore be informative for telling apart theories and their representations.
- Especially true with information on age of acquisition and/or patterns of variation across dialects.

Discussion

Age of acquisition

- Unfortunately little is known about PCC age of acquisition.
- Tsakali & Wexler (2010) report that 5-year-old Greek-acquiring children know the PCC.
- Blasco (2000) show that Spanish-acquiring children were correctly producing accusative and dative clitics by 2;2.

Discussion

Key takeaways

- Proof of concept for using learning considerations to tell apart different theoretical and representational assumptions in domain of person features and PCC.

Discussion

Key takeaways

- Proof of concept for using learning considerations to tell apart different theoretical and representational assumptions in domain of person features and PCC.
- Similar learning models can be run for other more restrictive theories of the PCC (*e.g.*, Béjar & Rezac 2003, Pancheva & Zubizarreta 2018, Graf 2019) and for other PCC variants.
- Would want to see if other restrictive theories are ever led astray toward unattested variants.

Acknowledgments

Thanks to:

- Norbert Hornstein, Jeff Lidz, and Omer Preminger for helpful discussion;
- the University of Marylnad CNL Lab for helpful discussion; and
- support from the NSF NRT grant (NSF: #1449815).

References I

Aguirre, Carmen. 2003. Early verb development in one Spanish-speaking child. In Dagmar Bittner, Wolfgang U. Dressler & Marianne Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition: A cross-linguistic perspective* (Studies on Language Acquisition 21), 1–26. Berlin, Germany: Mouton de Gruyter. DOI: 10.1515/9783110899832.1.

Béjar, Susana & Milan Rezac. 2003. Person licensing and the derivation of PCC effects. In Ana Teresa Pérez-Leroux & Yves Roberge (eds.), *Romance linguistics: Theory and acquisition* (Current Issues in Linguistic Theory 244), 49–62. Amsterdam, The Netherlands: John Benjamins Publishing Company. DOI: 10.1075/cilt.244.07bej.

Blasco, Maria. 2000. *The acquisition of pronominal object clitics in Spanish*. New York, NY: The City University of New York dissertation.

References II

Bonet, M. Eulàlia. 1991. *Morphology after syntax: Pronominal clitics in Romance*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Bonet, M. Eulàlia. 1994. The person-case constraint: A morphological approach. In Heidi Harley & Colin Phillips (eds.), *The morphology-syntax connection* (MIT Working Papers in Linguistics 22), 33–52. Cambridge, MA.

Graf, Thomas. 2019. Monotonicity as an effective theory of morphosyntactic variation. *Journal of Language Modelling* 7(2). 3–47. DOI: 10.15398/jlm.v7i2.211.

Hart, Betty & Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

Honnibal, Matthew & Ines Montani. 2017. spaCy 2: natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

References III

Lee, Jackson L., Ross Burkholder, Gallagher B. Flinn & Emily R. Coppess. 2016. *Working with CHAT transcripts in Python*. Tech. rep. TR-2016-02. Department of Computer Science, University of Chicago.

MacWhinney, Brian. 2000. *The CHILDES project: tools for analyzing talk*. 3rd. Mahwah, NJ: Lawrence Erlbaum Associates.

Nevins, Andrew. 2007. The representation of third person and its consequences for person-case effects. *Natural Language & Linguistic Theory* 25(2). 273–313. DOI: 10.1007/s11049-006-9017-2.

Pancheva, Roumyana & Maria Luisa Zubizarreta. 2018. The person case constraint: The syntactic encoding of perspective. *Natural Language & Linguistic Theory* 36(4). 1291–1337. DOI: 10.1007/s11049-017-9395-7.

References IV

Pearl, Lisa, Timothy Ho & Zephyr Detrano. 2017. An argument from acquisition: Comparing English metrical stress representations by how learnable they are from child-directed speech. *Language Acquisition* 24(4). 307–342. DOI: 10.1080/10489223.2016.1194422.

Pearl, Lisa & Jon Sprouse. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20(1). 23–68. DOI: 10.1080/10489223.2012.738742.

Pearl, Lisa & Jon Sprouse. 2019. The acquisition of linking theories: A tolerance principle approach to learning UTAH and rUTATH. Ms., University of California, Irvine, CA and University of Connecticut, Storrs, CT. https://ling.auf.net/lingbuzz/004088.

Rasin, Ezer & Roni Katzir. 2017. A learnability argument for constraints on underlying representations. Ms., Leipzig University and Tel Aviv University. Available at https://ling.auf.net/lingbuzz/002260.

References V

Tenenbaum, Joshua B. & Thomas L. Griffiths. 2001. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences* 24(4). 629–640. DOI: 10.1017/S0140525X01000061.

Tsakali, Vina & Kenneth Wexler. 2010. The acquisition of Person Case Constraint in Greek. Paper presented at the *19th International Symposium on Theoretical and Applied Linguistics*, Thessaloniki, Greece.

A feature-based theory of the PCC

Nevins (2007)

(8) $FG_1 = \begin{bmatrix} \end{bmatrix}$

(9) $FG_2 = \begin{bmatrix} \text{+Part} \end{bmatrix}$

(10) $FG_3 = \begin{bmatrix} \text{+Auth} \end{bmatrix}$

(11) $FG_4 = \begin{bmatrix} \text{+Part} \\ \text{+Auth} \end{bmatrix}$

(12) $FG_5 = \begin{bmatrix} \text{Auth/[+Part]} \\ \text{Part/[—Auth]} \end{bmatrix}$

(13) $FG_6 = \begin{bmatrix} \text{Auth/[+Part]} \end{bmatrix}$

(14) $FG_7 = \begin{bmatrix} \text{Auth/[+Part]} \\ \text{+Part} \end{bmatrix}$

(15) $FG_8 = \begin{bmatrix} \text{Part/[—Auth]} \end{bmatrix}$

(16) $FG_9 = \begin{bmatrix} \text{Part/[—Auth]} \\ \text{+Auth} \end{bmatrix}$

A feature-based theory of the PCC

Nevins (2007)

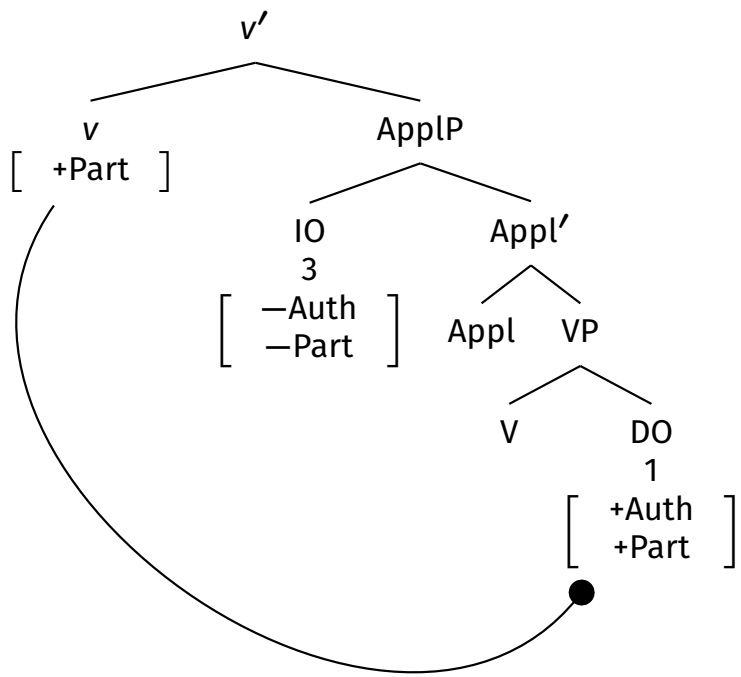
- Search is subject to two conditions:
  1. Any argument that occurs in between the probe that initiates the search and the target of the search must itself also be a target.
  2. All arguments in the domain of the search must share the same value ($+$ or $-$).

A feature-based theory of the PCC

Nevins (2007)

- Consider the grammar $FG_2 = v[\ \text{+Part}\ ]$ (*i.e.*, the Weak PCC) and *3 1.
- This violates the first condition.

(17) *

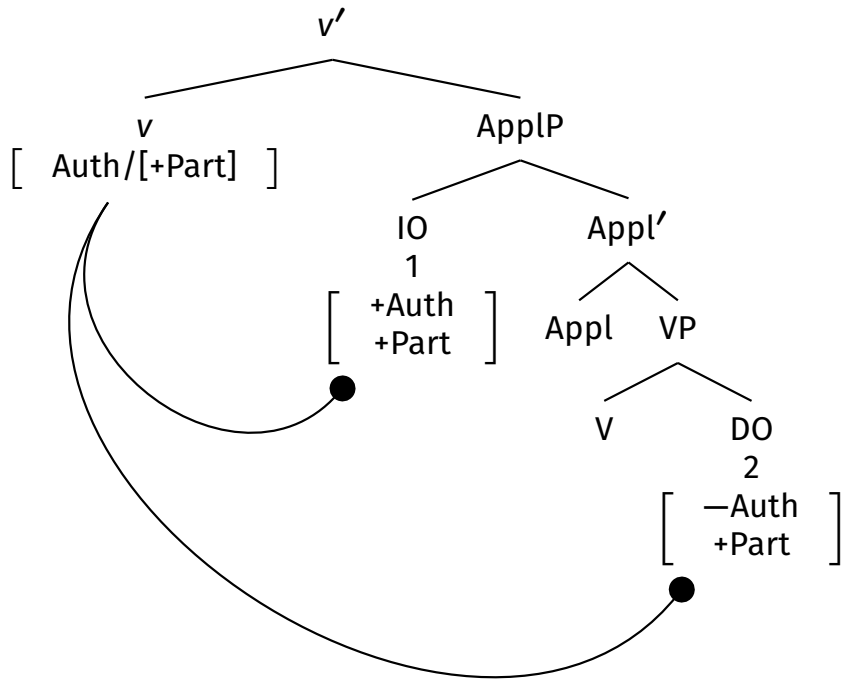


A feature-based theory of the PCC

Nevins (2007)

- Consider the grammar $FG_6 = v[\ \text{Auth}/[+\text{Part}]\ ]$ (*i.e.*, the Strong PCC) and *1 2.
- This violates the second condition.

(18) *



Aguirre corpus parsing

- *se* was treated as a third person pronoun.
- The case information assigned by `spaCy` to each clitic was the main basis for classifying the clitic as the direct or indirect object.

Aguirre corpus examples

- 1 3 ($n = 50$)

(19) ésta me la he comprado yo.

(20) oye, ese pez luego nos lo podemos subir para la bañera, vale?

(21) bueno nos la ponemos, vale?

...

- 2 3 ($n = 148$)

(22) esto te lo has mojado.

(23) te la vas a comer a la gallina?

(24) a bañar, que en el baño te lo pasas muy bien también.

...

- 3 2 ($n = 4$)

(25) ahora te le pongo.

(26) ése te le pongo mañana.

(27) te le vas a llevar el hipopótamo a la oficina.

(28) te le vas a meter el cepillo en el agua?

- 3 3 ($n = 68$)

(29) se la has comprado tú?

(30) a dónde se la has dado?

(31) a ver como se lo dices tú.

...