Language processing over a noisy channel

Ted Gibson
Department of Brain and Cognitive Sciences
MIT

http://tedlab.mit.edu
Twitter: @MITlanguagelab

CLASP, University of Gothenburg, Sweden
September 6, 2016

Language: structure, acquisition and processing

Ted Gibson

Research program

What factors affect the complexity of processing a phrase or text?
E.g., word frequency; syntactic rules; working memory resources

What pressures shape human language?
(1) communication; (2) memory; (3) culture.
Evidence: cross-linguistic universals

Methods

- Behavioral experiments (e.g., reading / listening or generation)
  - Cross-linguistic / cross-cultural experiments
- Corpus analyses
- Computational modeling
- Brain imaging

Language processing over a noisy channel

Collaborators on these projects:

- Richard Futrell, MIT
- Steve Piantadosi, U Rochester
- Kyle Mahowald, MIT
- Leon Bergen, Stanford
- Ev Fedorenko, MIT, Harvard
- Chaleece Sandberg, Boston U
- Swathi Kiran, Boston U
- Laura Stearns, MIT
- Marianna Eddy, Tufts
- Kim Brink, MIT
- Eunice Lim, UCSD
- Rebecca Saxe, MIT
- Roger Levy, MIT

Language comprehension

Language processing is probabilistic in nature, with highly variable frequencies of the relevant units:

- Words: "say" vs. "play" vs. "catch" vs. "caste"
- Meanings: *The dog bit the boy* vs. *The boy bit the dog*
- Syntax: *John was smoking* vs. *That John was smoking was annoying*

MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell & Tanenhaus, 1995; Hale, 2001; Levy, 2008; Jaeger, 2010; Ambridge, Kidd, Rowland, & Theakston, 2015; Lau, Clark & Lappin, 2016; cf. Chater, Tenenbaum, & Yuille, 2006.

Language comprehension

What happens when we encounter a low probability event?

- *The teacher gave the book the student.*

The literal syntax indicates a low probability event in the world.

People make many errors, both in production and comprehension

The most likely linguistic sequence may not be exactly what was perceived.

Given the reliance on a probabilistic linguistic sources of information, the processing system must therefore be very sensitive to the possibility of errors

Misinterpretations of spoken language in songs: Mondegreens

American writer Sylvia Wright in "The Death of Lady Mondegreen", Harper's Magazine, 1954.

17th-century ballad "The Bonnie Earl o' Moray":

Ye Highlands and ye Lowlands,
Oh, where hae ye been?
They hae slain the Earl o' Moray,
And laid him on the green.

Wright misheard the last line as "And Lady Mondegreen"

In unsupportive contexts, more frequent words and phrases are sometimes perceived instead

Mondegreens in songs

Creedence Clearwater Revival, "Bad Moon Rising"
"There's a bathroom on the right"
("There's a bad moon on the rise")

Jimi Hendrix, "Purple Haze"
"Excuse me while I kiss this guy"
("Excuse me while I kiss the sky")

Rush, "Limelight"
"living in a fish island"
("living in a fish-eye lens")

Rational inference in language: Noisy-channel models of language

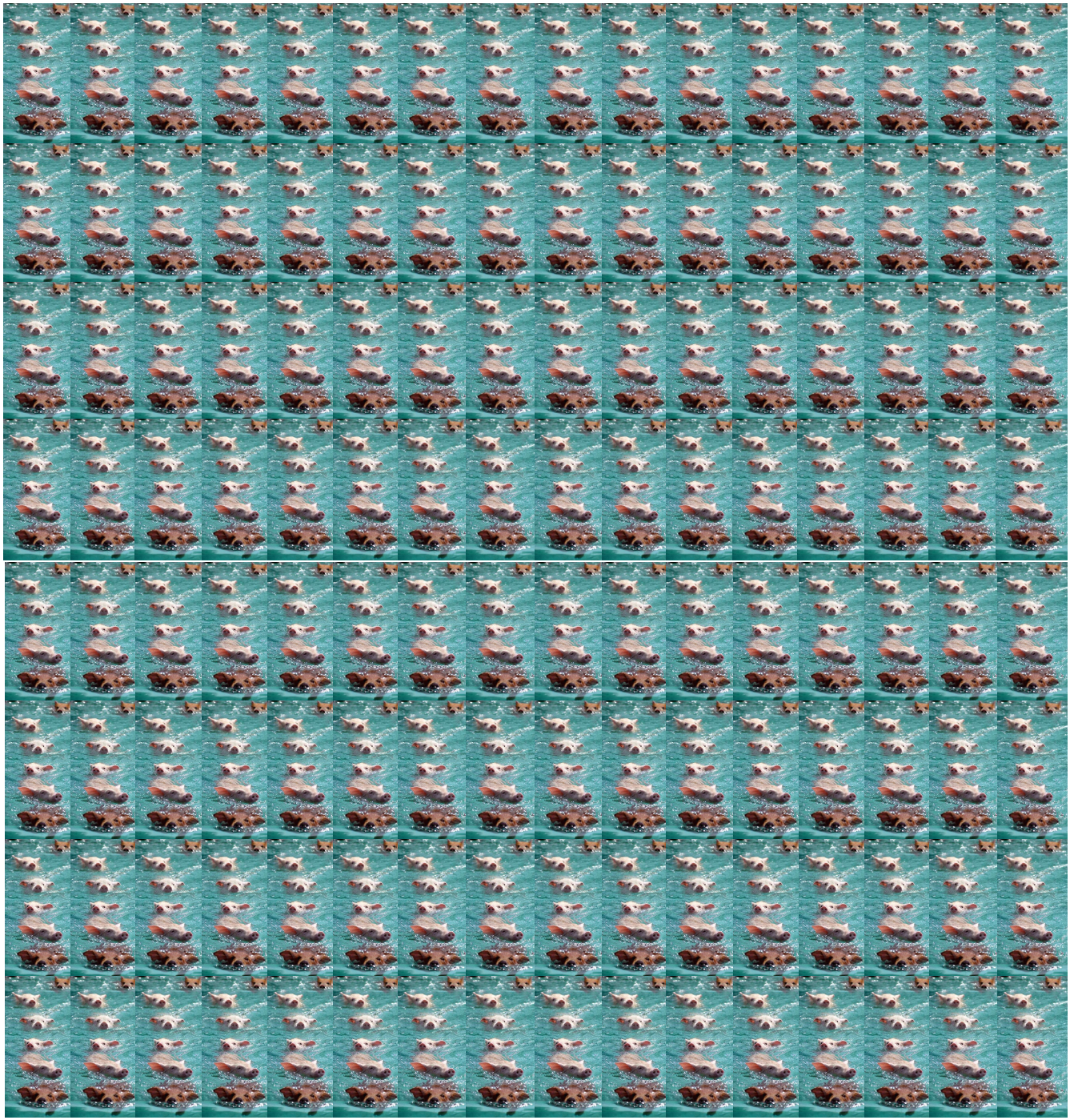
Correction

THERE was an error printed in a story titled "Pigs float down the Dawson" on Page 11 of yesterday's *Bully*.

The story, by reporter Daniel Burdon, said "more than 30,000 pigs were floating down the Dawson River".

What Baralaba piggery owner Sid Everingham actually said was "30 sows and pigs", not "30,000 pigs".

Thirty sows and pigs
in a river



Thirty thousand pigs in a river

Rational inference in language: Noisy-channel models of language

Language for communication: The rational integration of noise and prior lexical, syntactic and semantic expectation:

Maximize $P(s_i \mid s_p)$ by maximizing $P(s_i) * P(s_i \rightarrow s_p)$

All linguistic measures (e.g., reading times, acceptability ratings) reflect:

- the prior expectation of what might be produced
- the likelihood of noise changing s_i into s_p

Noisy-channel models of comprehension

- Classic assumption in sentence processing:
  input to the parser is an **error-free** sequence of words
  (e.g., Frazier & Fodor, 1978; Gibson, 1991, 1998; Jurafsky, 1996; Hale, 2001; Levy, 2008a).

- This assumption is problematic (e.g., Levy, 2008b).
  Many sources of noise:
  (a) perception errors (mis-hearing/mis-reading); the environment can be noisy
  (b) production errors (mis-speaking/mis-typing)

- Classic issue in signal processing (e.g., Shannon, 1948)

- Previous work: Speech (Jelinek, 1975; Clayards, Tanenhaus, Aslin & Jacobs, 2008); Memory (Botvinick, 2005); Reading (Levy et al., 2009)

Language processing over a noisy channel

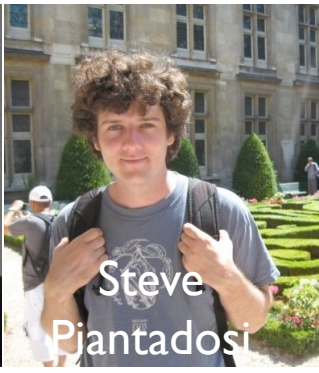
- Language comprehension in a noisy channel: the rational integration of noise and prior
  - Language comprehension accuracy
- Applications: Speaking with an accent
- Aphasic language comprehension
- Event-related potentials: The P600
- Applications in psycholinguistic phenomena: agreement error production
- Cross-linguistic word order universals: SOV and SVO word order

Noisy-channel models of comprehension

General prediction for sentence interpretation:
The ultimate interpretation of a sentence should depend on the *proximity of plausible alternatives* under the noise model.

A plausible noise model (cf. Levenshtein distance):
some cost for deletions, insertions (maybe swaps?)

(Gibson, Bergen & Piantadosi, 2013, *PNAS*)



Noisy-channel models of comprehension

Testing the predictions: syntactic alternations:
More changes leads to lower likelihood of inferring the alternative (cf. MacWhinney & Bates, 1989; Ferreira, 2003)

"Minor" change alternations:

PO-goal ➔ DO-goal (1 deletion):
*The mother gave the candle **to** the daughter.* ➔ *The mother gave the candle the daughter.*

DO-goal ➔ PO-goal (1 insertion):
The mother gave the daughter the candle. ➔ *The mother gave the daughter **to** the candle.*

"Major" change alternations:

Passive ➔ Active (2 deletions):
*The ball **was** kicked **by** the girl.* ➔ *The ball kicked the girl.*

Active ➔ Passive (2 insertions):
The girl kicked the ball. ➔ *The girl **was** kicked **by** the ball.*

Noisy-channel models of comprehension

Design:

- manipulate plausibility (using role reversals)
- examine interpretation

Interpretation was assessed with comprehension questions.

Examples:

a. **Sentence:** *The ball kicked the girl.*
Question: Did the ball kick something/someone?

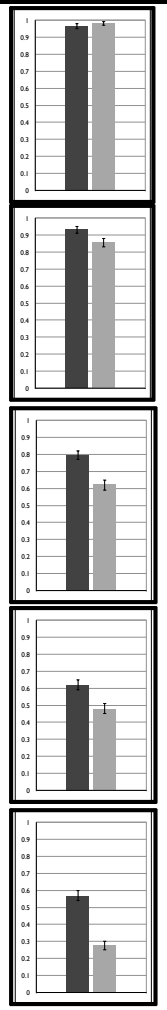
b. **Sentence:** *The mother gave the candle the daughter.*
Question: Did the daughter receive something/someone?

E.g., in (a) a "yes" answer indicates that the reader relied on syntax (surface form) to interpret the sentence; a "no" answer indicates that the reader relied on semantics. The reverse holds for (b).

(Gibson, Bergen & Piantadosi, 2013)

Results

1a. Passive -> Active:	The ball was/∅ kicked by/∅ the girl.	2 deletions
1b. Active -> Passive:	The girl ∅/was kicked ∅/by the ball.	2 insertions
2a. Subj-loc –> Obj-loc:	∅/Onto The cat jumped onto/∅ a table.	1 insertion, 1 deletion
2b. Obj-loc –> Subj-loc:	Onto/∅ the table jumped ∅/onto a cat.	1 deletion, 1 insertion
3a. Intrans –>Trans:	The tax law benefited ∅/from the businessman.	1 insertion
3b. Trans –> Intrans:	The businessman benefited from/∅ the tax law.	1 deletion
4a. DO –> PO-goal:	The mother gave the daughter ∅/to the candle.	1 insertion
4b. PO –> DO-goal:	The mother gave the candle to/∅ the daughter.	1 deletion
5a. DO –> PO-benef:	The cook baked Lucy ∅/for a cake.	1 insertion
5b. PO –> DO-benef:	The cook baked a cake for/∅ Lucy.	1 deletion



More changes lead to a greater reliance on syntax:
major changes (93.4%) vs. minor changes: (56.1%)
Deletions are perceived to be more likely than insertions, leading to lower likelihood of literal meaning for deletions:
single insertions (66.1%) vs. single deletions (46.0%)

Results

1a. Passive -> Active:	The ball was/∅ kicked by/∅ the girl.	2 deletions
1b. Active -> Passive:	The girl ∅/was kicked ∅/by the ball.	2 insertions
2a. Subj-loc −> Obj-loc:	∅/Onto The cat jumped onto/∅ a table.	1 insertion, 1 deletion
2b. Obj-loc −> Subj-loc:	Onto/∅ the table jumped ∅/onto a cat.	1 deletion, 1 insertion
3a. Intrans −>Trans:	The tax law benefited ∅/from the businessman.	1 insertion
3b. Trans −> Intrans:	The businessman benefited from/∅ the tax law.	1 deletion
4a. DO −> PO-goal:	The mother gave the daughter ∅/to the candle.	1 insertion
4b. PO −> DO-goal:	The mother gave the candle to/∅ the daughter.	1 deletion
5a. DO −> PO-benef:	The cook baked Lucy ∅/for a cake.	1 insertion
5b. PO −> DO-benef:	The cook baked a cake for/∅ Lucy.	1 deletion

Prediction: more noise should lead to greater reliance on likely meaning

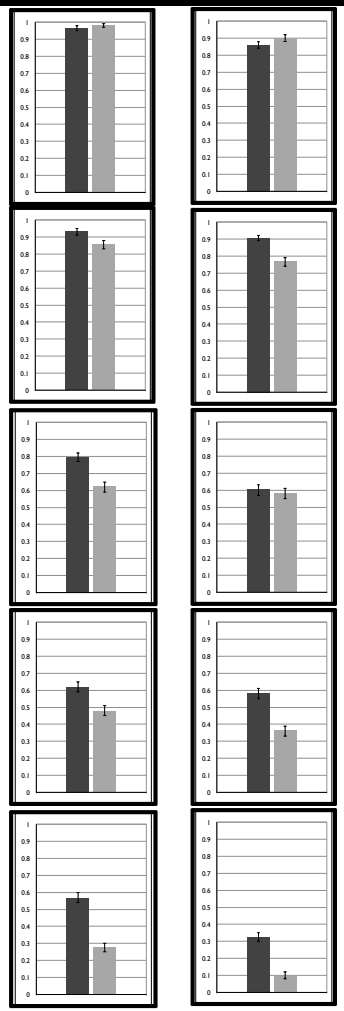
Manipulation:

add noise to 30 of the 60 fillers

10 - extra function word; 10 - missing function word; 10 - local transpositions

Results

1a. Passive -> Active:	The ball was/∅ kicked by/∅ the girl.	2 deletions
1b. Active -> Passive:	The girl ∅/was kicked ∅/by the ball.	2 insertions
2a. Subj-loc −> Obj-loc:	∅/Onto The cat jumped onto/∅ a table.	1 insertion, 1 deletion
2b. Obj-loc −> Subj-loc:	Onto/∅ the table jumped ∅/onto a cat.	1 deletion, 1 insertion
3a. Intrans −>Trans:	The tax law benefited ∅/from the businessman.	1 insertion
3b. Trans −> Intrans:	The businessman benefited from/∅ the tax law.	1 deletion
4a. DO −> PO-goal:	The mother gave the daughter ∅/to the candle.	1 insertion
4b. PO −> DO-goal:	The mother gave the candle to/∅ the daughter.	1 deletion
5a. DO −> PO-benef:	The cook baked Lucy ∅/for a cake.	1 insertion
5b. PO −> DO-benef:	The cook baked a cake for/∅ Lucy.	1 deletion



More syntactic errors decreased the reliance on syntax:

56.1% vs. 42.7 for the minor-change alternations

Noisy-channel models of comprehension

Manipulations of semantic / plausibility prior:

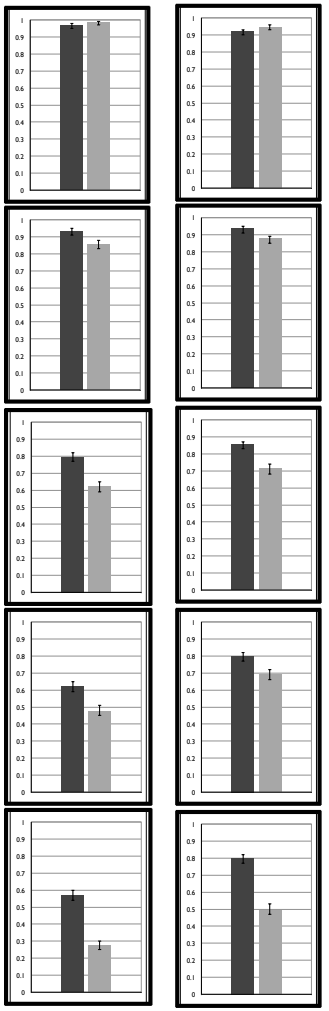
Plausibility prior: how likely it is that an implausible utterance will be generated

Expt 1a - 1e:
Each was run with 60 plausible fillers.
Implausible ratio = 1/8 (10 implaus + 70 plaus)

Expt 3a - 3e:
Each was run with 60 plausible fillers plus the materials in the other experiments.
Implausible ratio = 5/16 (50 implaus + 110 plaus)

Results

1a. Passive -> Active:	The ball was/∅ kicked by/∅ the girl.	2 deletions
1b. Active -> Passive:	The girl ∅/was kicked ∅/by the ball.	2 insertions
2a. Subj-loc −> Obj-loc:	∅/Onto The cat jumped onto/∅ a table.	1 insertion, 1 deletion
2b. Obj-loc −> Subj-loc:	Onto/∅ the table jumped ∅/onto a cat.	1 deletion, 1 insertion
3a. Intrans −>Trans:	The tax law benefited ∅/from the businessman.	1 insertion
3b. Trans −> Intrans:	The businessman benefited from/∅ the tax law.	1 deletion
4a. DO −> PO-goal:	The mother gave the daughter ∅/to the candle.	1 insertion
4b. PO −> DO-goal:	The mother gave the candle to/∅ the daughter.	1 deletion
5a. DO −> PO-benef:	The cook baked Lucy ∅/for a cake.	1 insertion
5b. PO −> DO-benef:	The cook baked a cake for/∅ Lucy.	1 deletion



More implausible materials increased the reliance on syntax:

56.1% vs. 72.6 for the minor-change alternations

Noisy-channel models of comprehension

Summary:

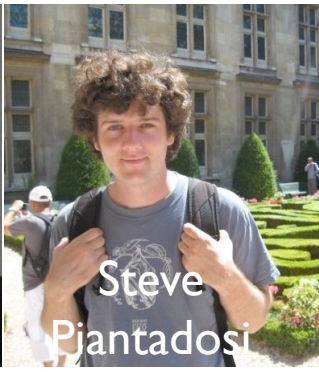
Evidence for a noise model:

1. People are more likely to infer the plausible alternative if it involves inferring fewer errors.
2. People are more likely to infer the plausible alternative if it is one deletion away compared to one insertion.
3. Increasing the noise increases the reliance on plausibility.

Evidence for priors:

1. Plausibility Prior: Increasing the likelihood of implausible events decreases the reliance on semantics.

(Gibson, Bergen & Piantadosi, 2013, *PNAS*)



Language processing over a noisy channel

- Language comprehension in a noisy channel: the rational integration of noise and prior
  - Language comprehension accuracy

- Applications: Speaking with an accent:
  Don't underestimate the benefits of being misunderstood

Challenges faced by L2 speakers

- L2 speakers are embarrassed by their accents and the errors they make (Gluszek & Dovidio, 2010)
- L2 speakers are perceived to be:
  - less credible (Bourdieu, 1991; Lev-Ari & Keysar, 2010; Livingston et al., 2014)
  - less educated (Fraser & Kelly, 2012)
  - less intelligent (Fuertes, Potere & Ramirez, 2002; Anderson et al., 2007).

(Gibson, Tan, Futrell, Mahowald, Konieczny, Hemforth & Fedorenko, submitted)

L2: One potential advantage

Imagine you are at a cocktail party where you want to make business connections.

Suppose someone asks you about a Marketing Technologist position.

If you have an L2 accent, you could say "Marketing Technologist was hired SEO Consultant."

With a foreign accent, they may interpret this in the most plausible way. Without a foreign accent, you cannot get away with this uncertainty.

(Gibson, Tan, Futrell, Mahowald, Konieczny, Hemforth & Fedorenko, submitted)

L2: One potential advantage

Arianna Huffington, Smith College commencement address in 2013:

"*I moved to New York in 1980 and met Henry Kissinger, who told me not to worry about my accent, because you can never, in American public life, underestimate the advantages of complete and total incomprehensibility.*"

Advantage: ***Easier to bullshit***

(Gibson, Tan, Futrell, Mahowald, Konieczny, Hemforth & Fedorenko, submitted)

L2 vs. L1 Speakers: New Experiments

Interpretation of implausible materials, spoken by the same person (each of 2 Speakers), +accent or -accent

3 sets of implausible materials, from the PNAS paper:

DO/PO

- *The mother gave the candle the daughter.*
- *The mother gave the daughter to the candle.*

Transitive/intransitive

- *The businessman benefited the tax law.*
- *The tax law benefited from the businessman.*

Active/Passive

- *The ball kicked the girl.*
- *The girl was kicked by the ball.*

L2 vs. L1 Speakers: New Experiments

Interpretation of implausible materials, spoken by the same person (each of 2 Speakers), +accent or -accent

3 sets of implausible materials, from the PNAS paper

Fillers: Filler items from Gibson et al., spoken with no accent by the other speaker

Speaker 1: accented / no-accent target items

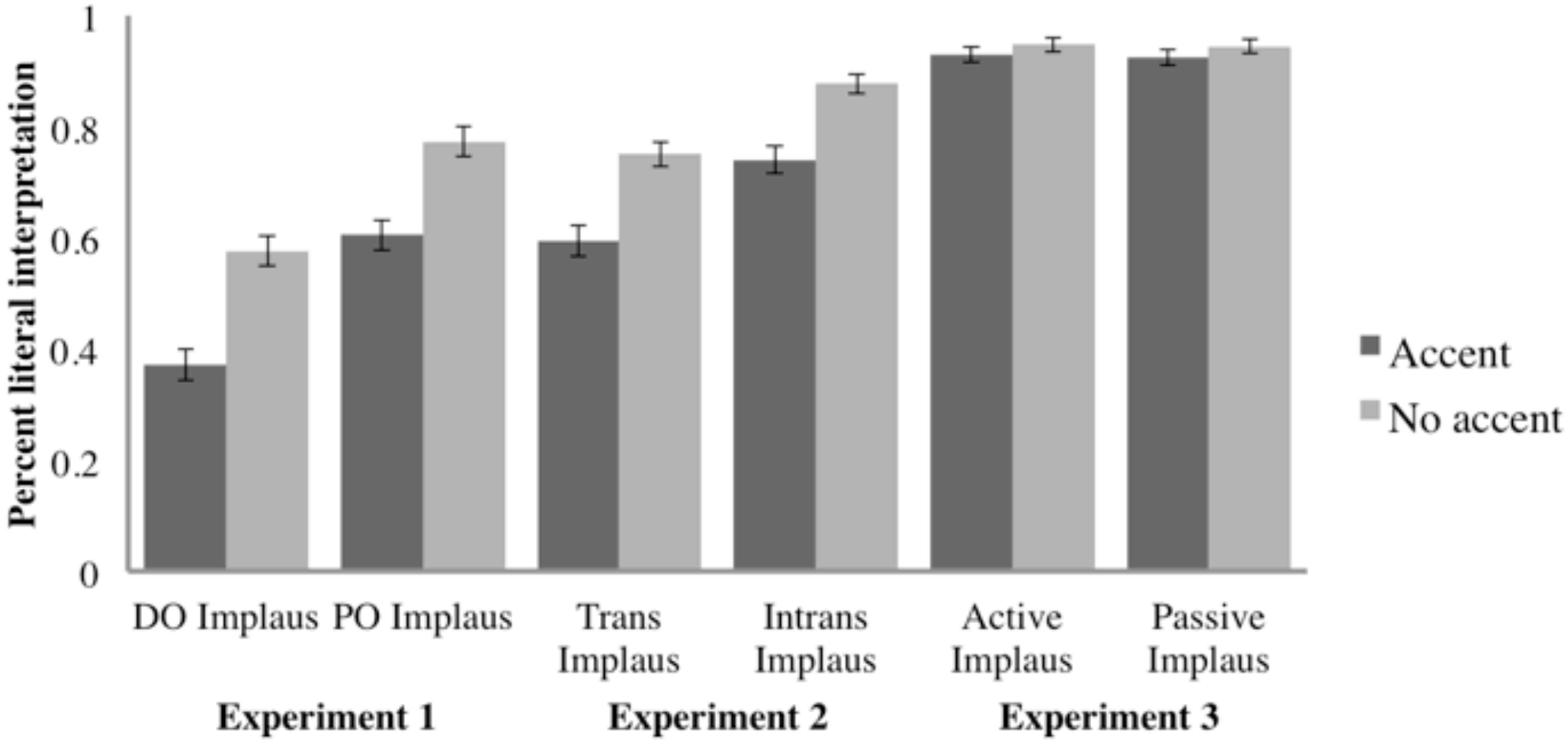
Speaker 2: no-accent filler items

(Gibson, Tan, Futrell, Mahowald, Konieczny, Hemforth & Fedorenko, submitted)

L2 vs. L1 Speakers: Results

1. DO, PO: ~20% inference effect
2. Transitive, Intransitive: ~15% inference effect
3. Active, Passive: no significant difference.

Result: ~20% Bullshit advantage in an L2 accent!



Language processing over a noisy channel

- Language comprehension in a noisy channel: the rational integration of noise and prior
  - Language comprehension accuracy

- Applications: Speaking with an accent

- Aphasic language comprehension

The noisy-channel proposal applied to aphasic comprehension

Old observation: aphasics' comprehension relies more on world knowledge than non-brain-damaged controls. (e.g., Caramazza & Zurif, 1976)

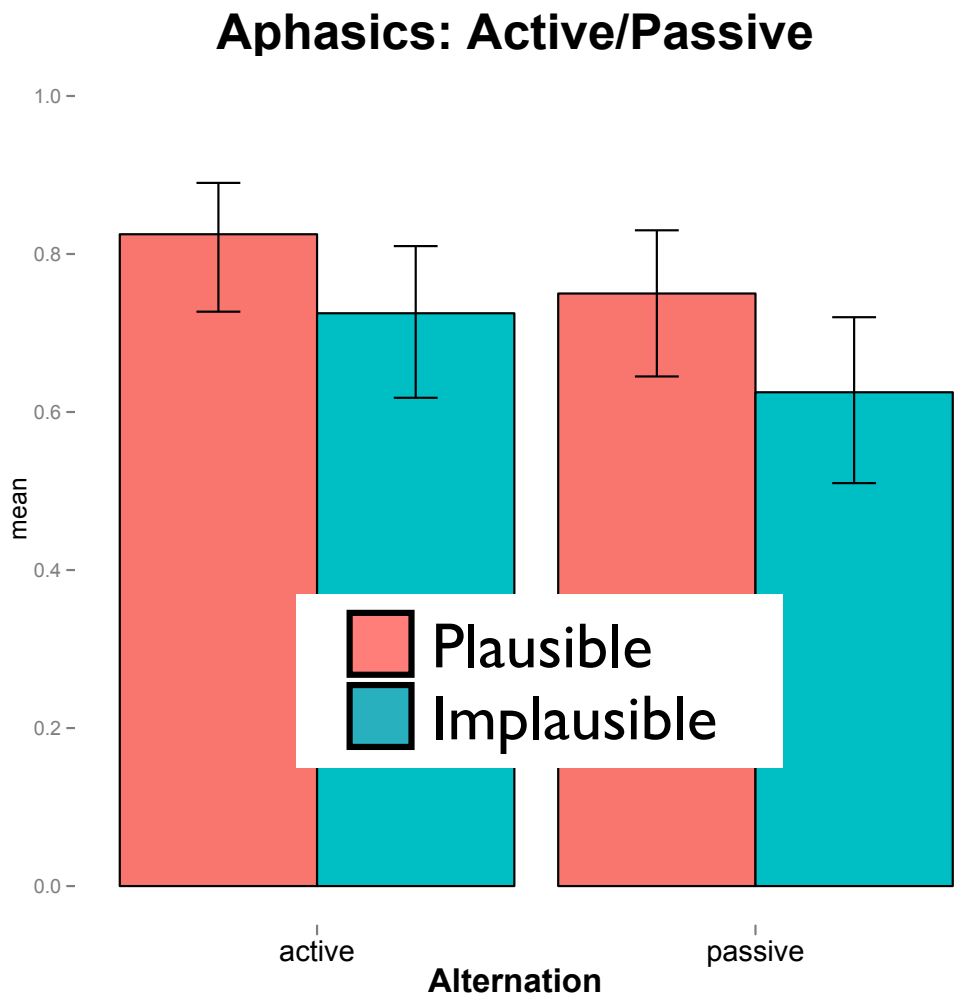
Hypothesis: Aphasics' language model is noisier than that of healthy individuals. In maximizing $P(s_i \mid s_p)$, aphasics will rely more on their prior distribution $P(s_i)$ over plausibly intended sentences.

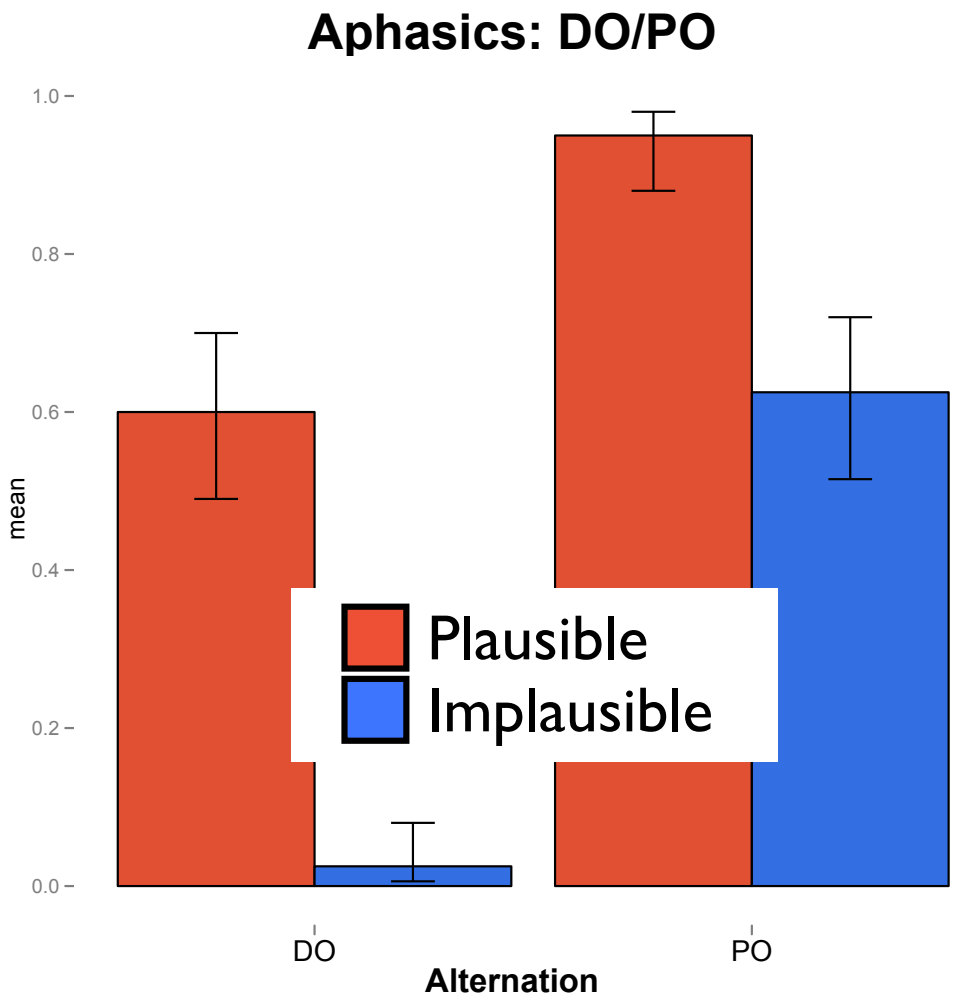
(Gibson, Sandberg, Fedorenko, Bergen & Kiran, 2015, *J of Aphasiology*)

Prediction:

Aphasics will rely on semantics more than healthy individuals, in both major-edit (active-passive) and minor-edit alternations (DO-PO).

Results: Active / Passive vs. DO / PO





Aphasics rely more on semantics in minor-edits (DO/PO) than in major-edits (active-passive): $z = 2.93$, $p < .005$

Similar results for other populations (replicating Gibson, Bergen & Piantadosi, 2013)

Language processing over a noisy channel

- Language comprehension in a noisy channel: the rational integration of noise and prior
  - Language comprehension accuracy

- Applications: Speaking with an accent

- Aphasic language comprehension

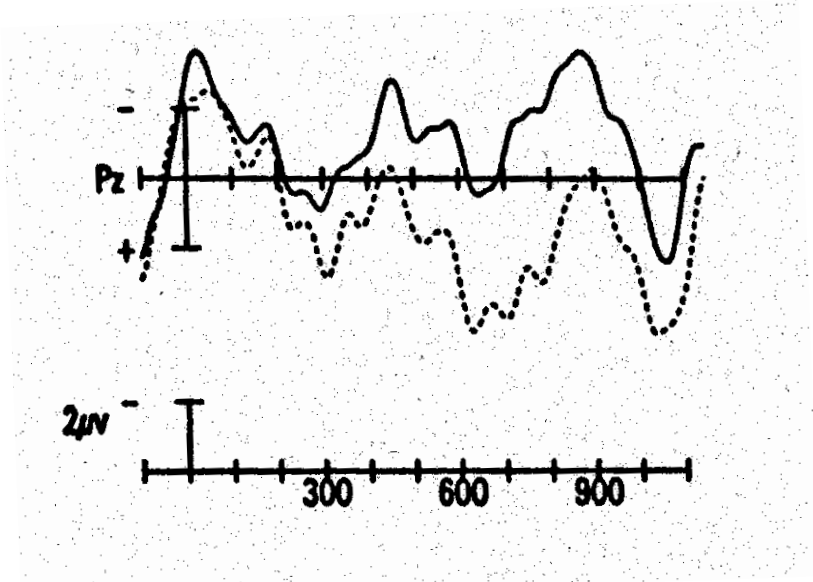
- Event-related potentials: The P600

Noisy-channel proposal for the P600

Fedorenko, Stearns, Bergen, Eddy & Gibson, subm.

The P600: Syntactic surprisal?

(Osterhout & Holcomb, 1992; Hagoort & Brown, 1993)



Every Monday he mows the lawn.
Every Monday he *mow the lawn.

Traditional interpretations:

- ungrammaticality detection
- syntactic reanalysis

Noisy-channel proposal for the P600

Fedorenko, Stearns, Bergen, Eddy & Gibson, subm.

Proposal: When a correction can be made, the P600 occurs

Time of P600: The P600 occurs relatively late because it indexes **correction**.

Corrections are not just syntactic:

- "Syntactic" violations: *Every Monday he mow / mows the lawn*
- "Semantic P600's": *The hearty meal was devouring / devoured ...*
- Orthographic errors: *fone / phone*

Not just error detection: No P600 is predicted when a correction cannot be made

- Classic N400: *I take my coffee with cream and dog / sugar*

Experiment

Idea: vary the likelihood of an error by substituting words that come from the phonological and orthographic neighborhood of the plausible target (and that are therefore likely substitutions).

Materials:

The storyteller could turn any incident into an amusing...

anecdote (control)	
antidote (critical)	P600
anecdotes (syntactic)	*P600*
hearse (semantic)	*no P600*

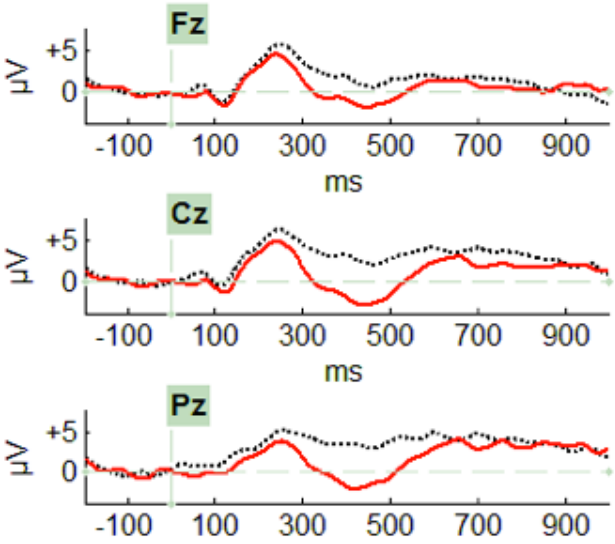
Task: reading with occasional comprehension questions.

Results (24 subjects; 160 items)

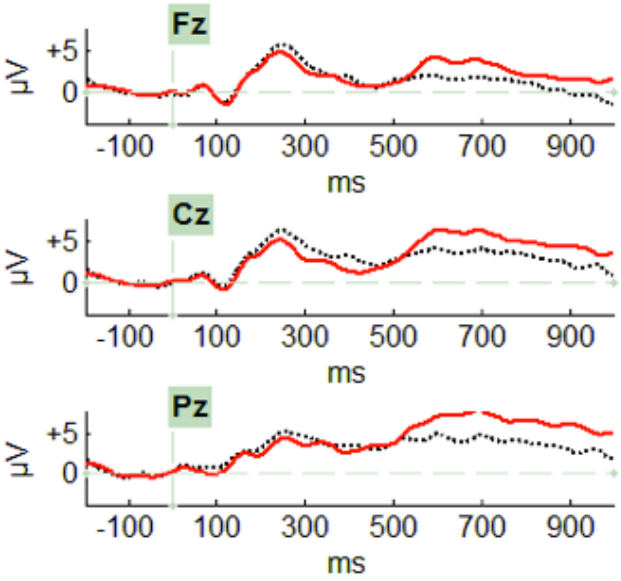
The storyteller could turn any incident into an amusing...

hearse **anecdotes** **antidote**

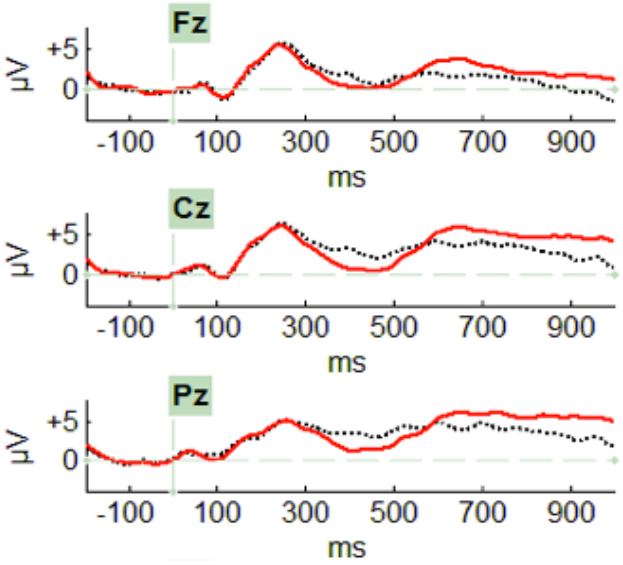
(c f . a n e c d o t e)



N400 (p<.001)



P600 (p<.001)



P600 (p<.001)
also: N400 (p<.001)

Positive is plotted up.

Language processing over a noisy channel

- Language comprehension in a noisy channel: the rational integration of noise and prior
  - Language comprehension accuracy
- Applications: Speaking with an accent
- Aphasic language comprehension
- Event-related potentials: The P600
- Applications in psycholinguistic phenomena: agreement error production

Agreement errors in sentence completions (Bergen & Gibson, 2013)

Asymmetry in agreement errors in a sentence completion task (Bock & Miller, 1991 among many others):

(1) *The key to the cabinets…*	was on the table	Correct agreement
	were on the table	Error, but COMMON

(2) *The keys to the cabinet…*	were on the table	Correct agreement
	was on the table	Error and rare

Standard explanation: memory retrieval, and there is a *markedness* difference between singular vs. plural nouns, in memory retrieval / sentence planning. **Stipulation**

Agreement errors in sentence completions (Bergen & Gibson, 2013)

Asymmetry in agreement errors in a sentence completion task (Bock & Miller, 1991 among many others):

(1) *The key to the cabinets…*	was on the table	Correct agreement
	were on the table	Error, but COMMON

(2) *The keys to the cabinet…*	were on the table	Correct agreement
	was on the table	Error and rare

Noisy channel explanation: Rational misidentification of preamble.

Maybe the producer meant:
"The **keys** to the cabinets" in (1) (*a deletion from a plural*)
But not "The **key** to the cabinet" in (2) *(an insertion from a singular)*

Deletions are much more likely than insertions: Thus agreement errors will occur more often when the **head noun** is singular.

Agreement errors in sentence completions (Bergen & Gibson, 2013)

Prediction of the noisy channel account: *The asymmetry should disappear with extra cues to singular agreement:*

(1) *Several keys to the cabinet…*	were on the table	Correct agreement
	was on the table	Error, and rare
(2) *A key to the cabinets…*	was on the table	Correct agreement
	were on the table	**Error and rare !!**

Other correct predictions of the noisy channel approach:

- The phenomenon is not tied to agreement (Bergen, Levy & Gibson, 2014)
- Misidentification of the sentence preamble also leads to repetition errors in the preamble, not just completion errors

Language processing over a noisy channel

- Language comprehension in a noisy channel: the rational integration of noise and prior
  - Language comprehension accuracy
- Applications: Speaking with an accent
- Aphasic language comprehension
- Event-related potentials: The P600
- Applications in psycholinguistic phenomena: agreement error production
- Cross-linguistic word order universals: SOV and SVO word order

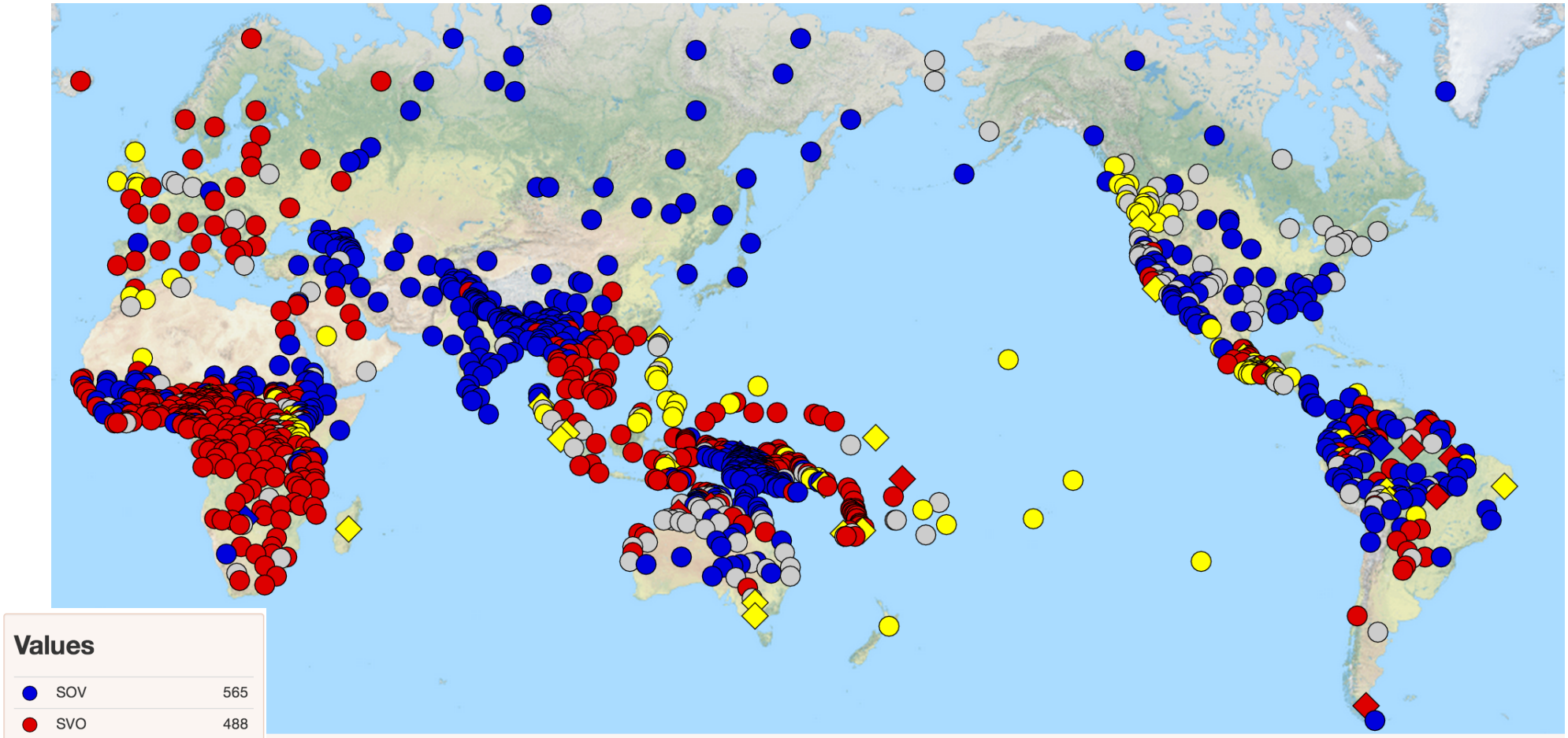
Proposal: Elements of word order could be driven by noisy-channel considerations

The Chomskyan perspective (e.g., Chomsky, 1965; 1986):
Grammars are independent of communicative and performance factors, determined by an innate U(niversal) G(rammar)

Because it's hard to find a communicative function to word order, syntactic differences are proposed to consist of differences in parameter-settings of biologically innate parameters, like Object-Verb / Verb-Object (e.g., Gibson & Wexler, 1994)

The performance-grammar correspondence hypothesis (Hawkins, 2004):
Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance (Haspelmath, 1999; Bybee & Hopper, 2001; Kirby, 1999; Kirby, Cornish & Smith, 2008; Culbertson, Smolensky & Legendre, 2012; Futrell, Mahowald & Gibson, 2015)

Syntax:
Word order across the world's languages



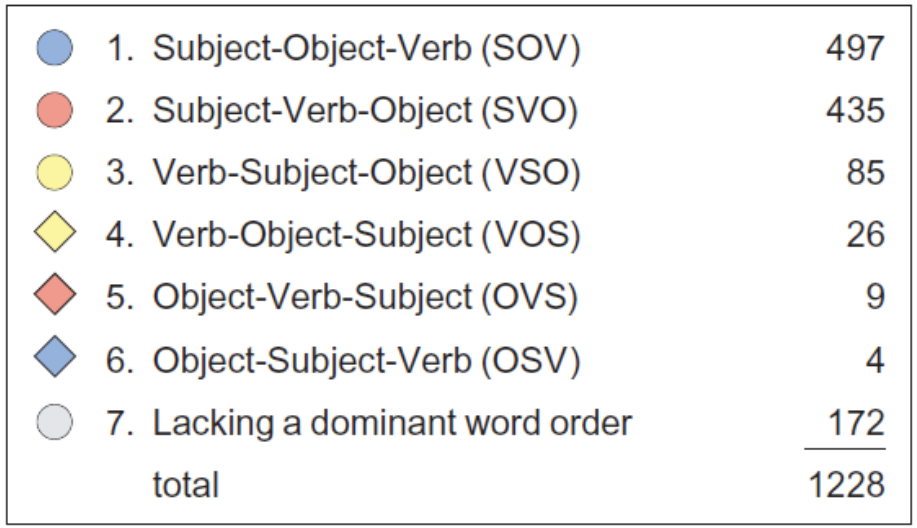
Values	
SOV	565
SVO	488
VSO	95
VOS	25
OVS	11
OSV	4
No dominant order	189

WALS Map 81: Dryer (2013)

Syntax:
Word order across the world's languages

Orders of Subject, Verb, and Object (WALS: Dryer, 2005)

- SO is a near universal: Almost no OS languages
- OV / VO are almost equally balanced:
  - SOV: 47.1% of languages with a dominant word order
  - SVO: 41.2% of languages with a dominant word order

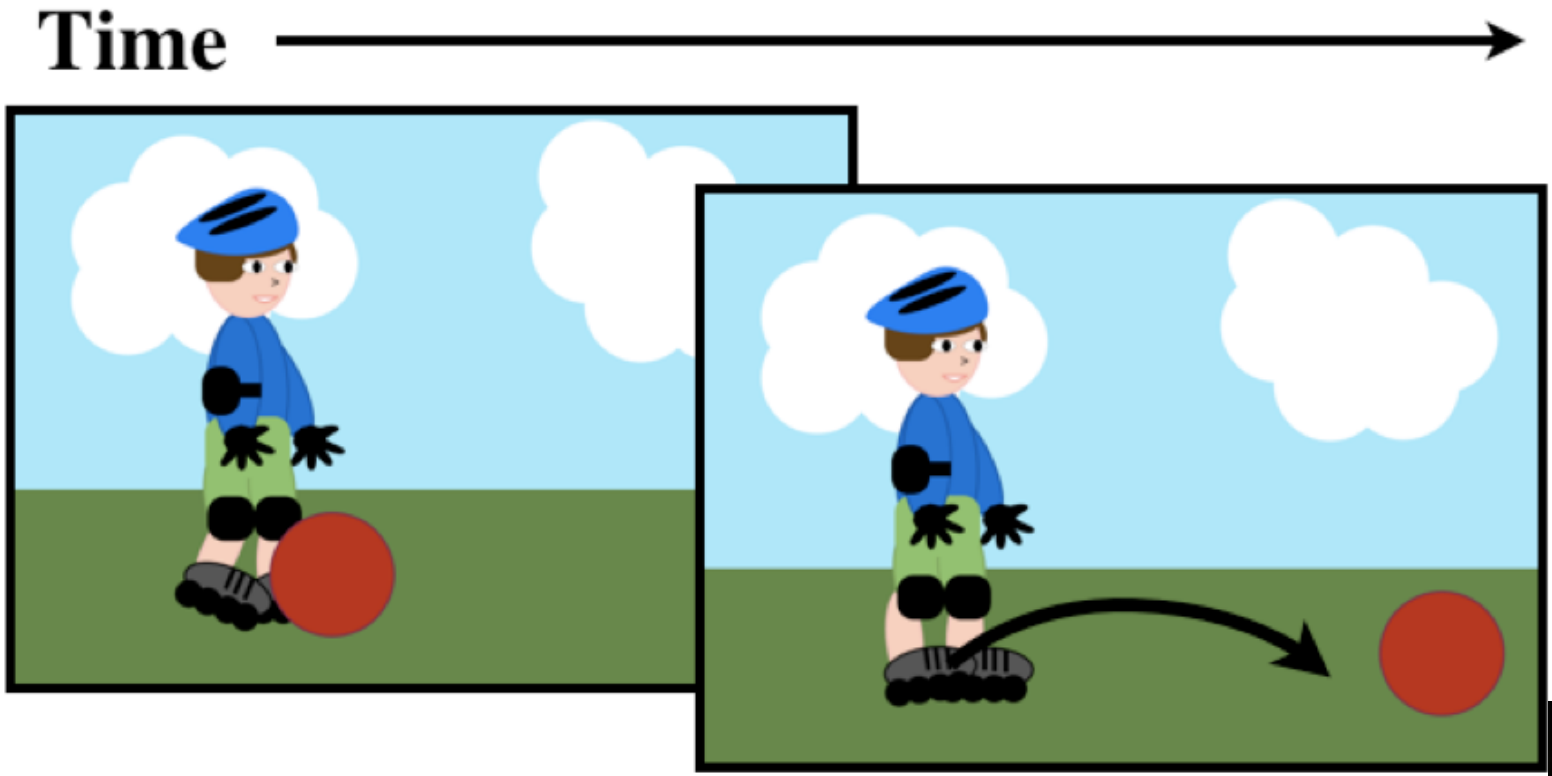


(1) Cognitive universals: SOV
- Subjects before objects (Greenberg, 1963; MacWhinney, 1977)
- Verbs at the end: ontologically-required early: need the objects before they can interact with each other: "old before new" (Jackendoff, 1972; Goldin-Meadow et al., 2008; Schouwstra et al., 2011)

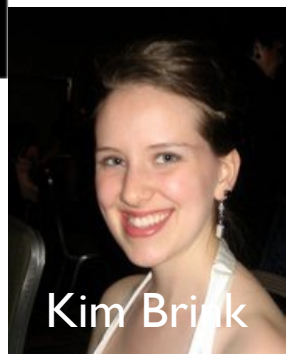
(2) Noisy channel model of communication (Shannon, 1949)

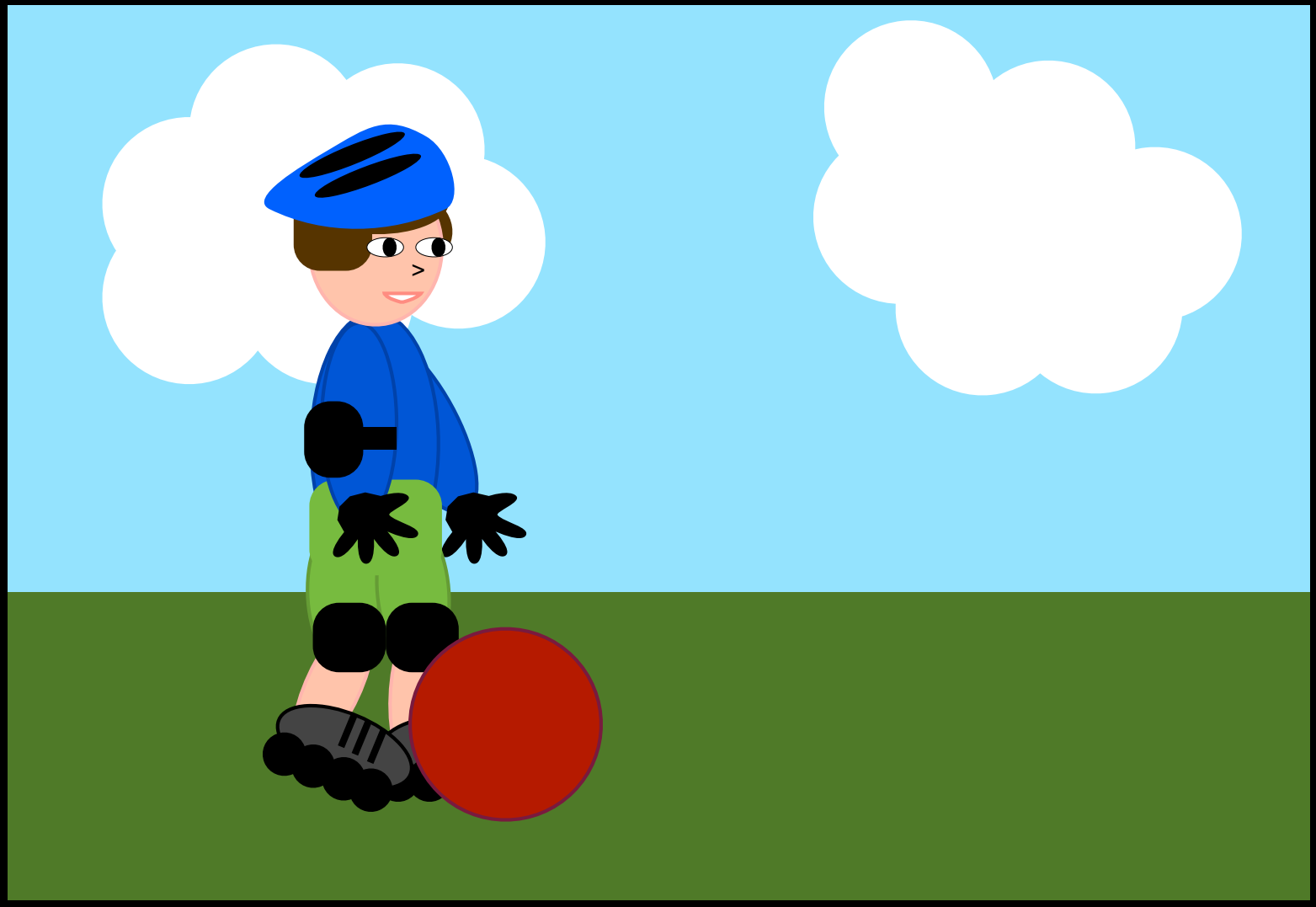
Gesture as a window onto the origin of syntax
SOV may be the most basic word order

- Participants watch animations, and then describe the scenes in words. Later, after watching them again, they gesture meanings for the animations



"The roller skater kicks the ball."





Replay

Goldin-Meadow et al. (2008)

SOV is the dominant word order in a task in which participants gesture sentence meanings. *The gesture-production task plausibly reflects people's word order preferences independent of their native language.*

OV

VO

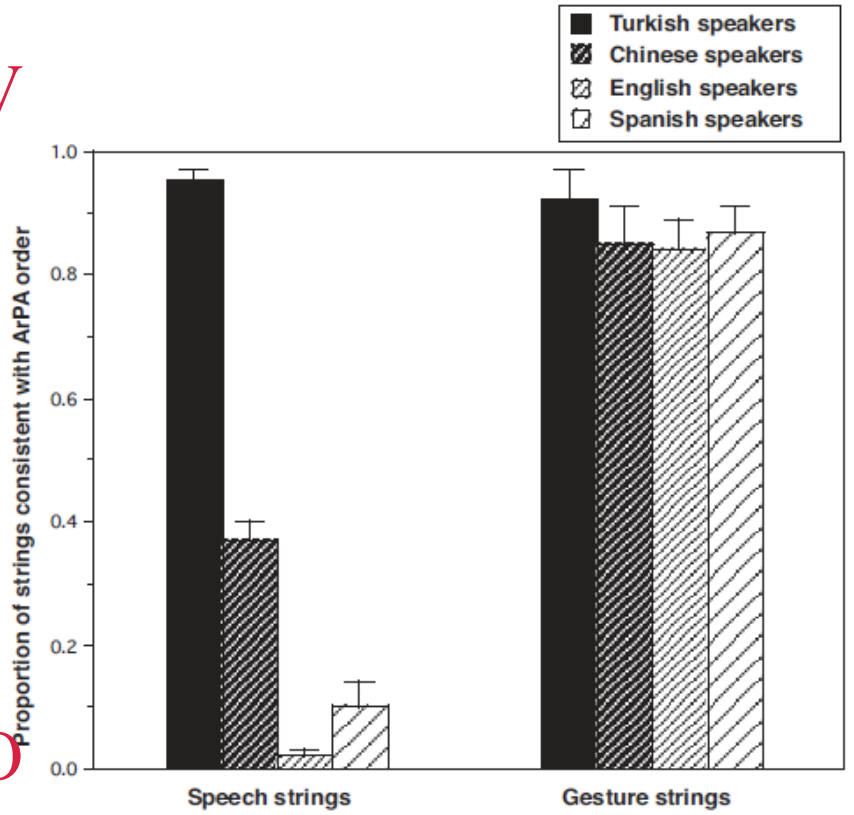
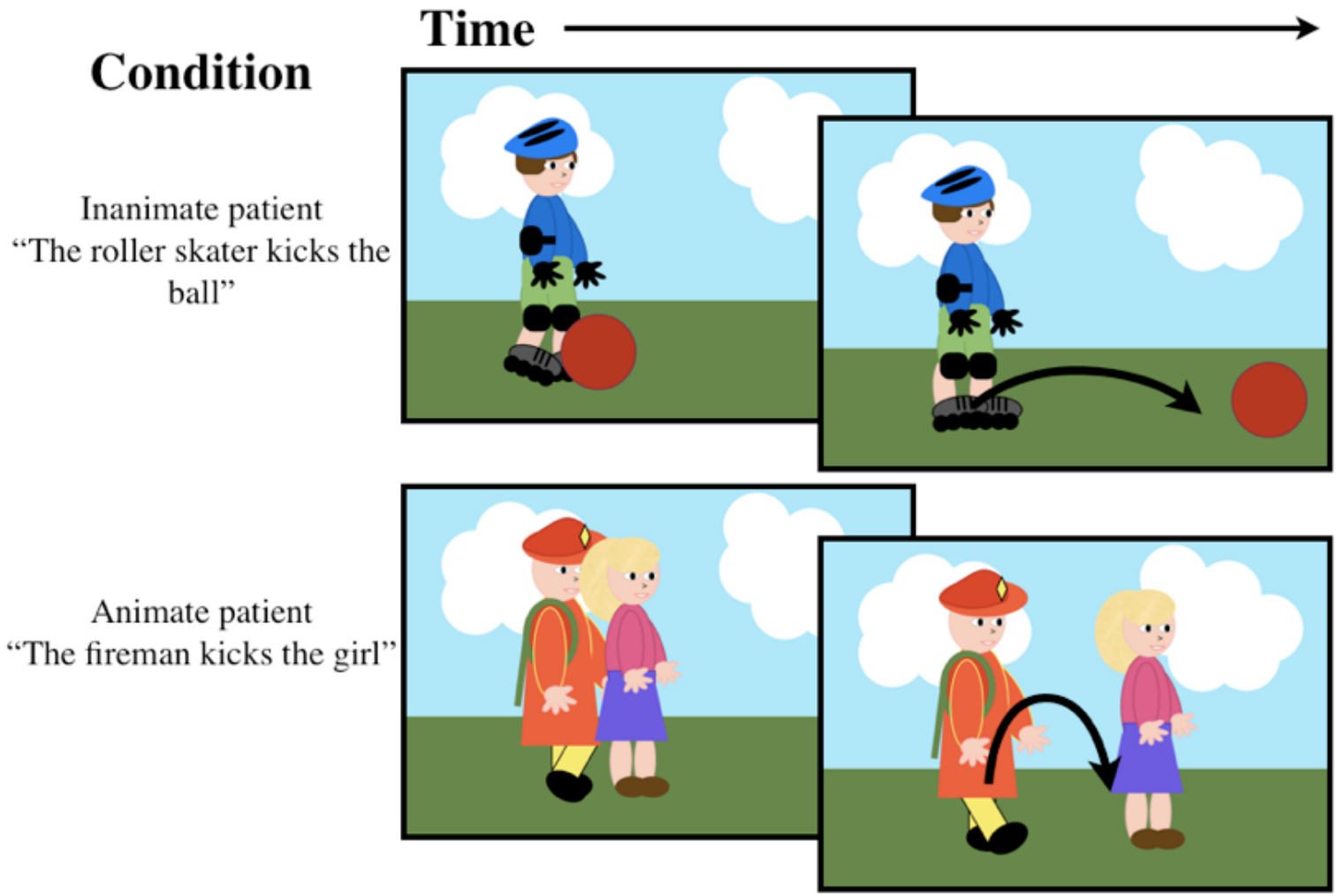


Fig. 2. Proportion of speech (*Left*) and gesture (*Right*) strings produced by speakers of Turkish, Chinese, English, and Spanish to describe transitive actions that were consistent with the ArPA order. Included are both in-place and crossing-space transitive actions.

Gibson et al. 2013: Reversible vs. Non-reversible events

Varying the similarity between the subject and the object NP: human subjects vs. inanimate / human objects

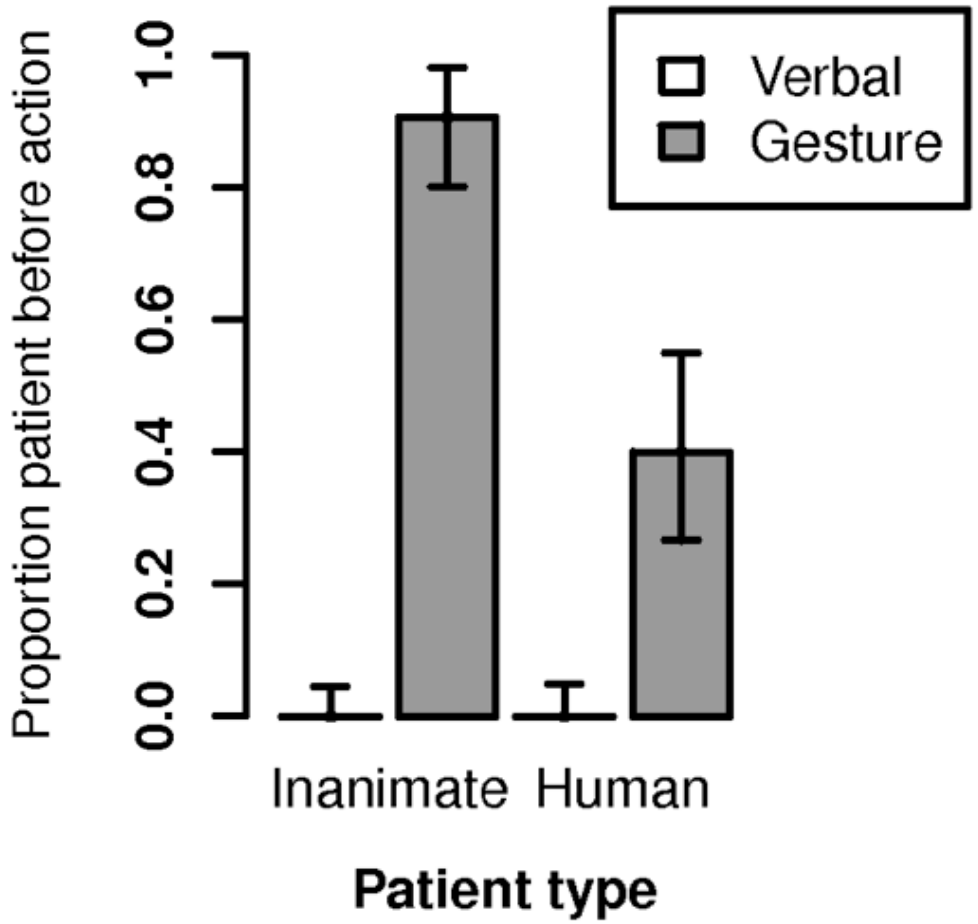


Replay

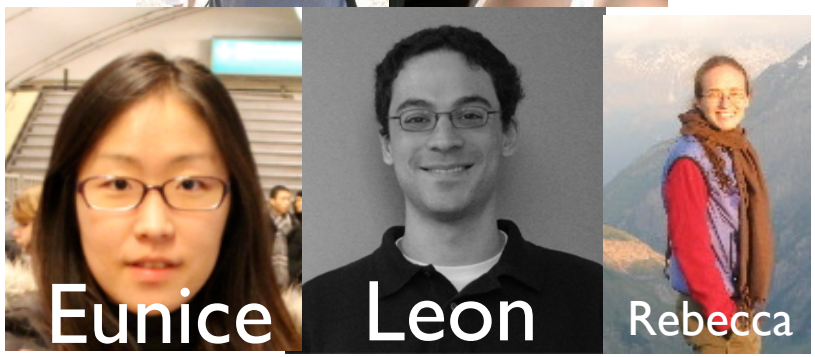
Gibson et al. 2013

Varying the similarity between the subject and the object NP: human subjects vs. inanimate / human objects:

***Preference reversal*: SVO**



The SOV / SVO also switch occurs ***for all other languages*** *that have been investigated: Russian; Tagalog; Irish; Japanese; Korean*



Why SVO?
An Information-Theoretic Account

(1) Cognitive universals: SOV

- Subjects before objects (Greenberg, 1963; MacWhinney, 1977)
- Verbs at the end: ontologically-required early: need the objects before they can interact with each other: "old before new" (Jackendoff, 1972; Goldin-Meadow et al., 2008; Schouwstra et al., 2011)

(2) Noisy channel model of communication (Shannon, 1949)

Suppose we want to convey "girl-agent boy-patient kiss" (*the girl kissed the boy)*

Noise in the channel: Likely loss of information

SOV: girl kiss: Is this girl-agent? Or girl-patient?
SOV: boy kiss: Is this boy-agent? Or boy-patient?

SVO: girl kiss: girl is agent
SVO: kiss boy: boy is patient

SVO word order is more robust to noise than SOV

Ramifications of a noisy-channel approach to cross-linguistic word order

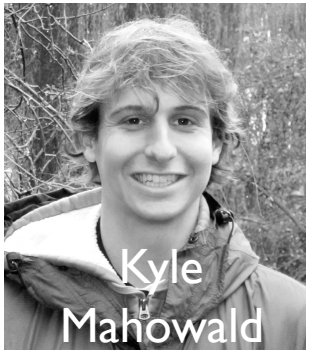
Why aren't all languages SVO? **Case-marking** SOV languages tend to be *case-marked*, while SVO languages need not be Dryer (2002) (cf. Greenberg, 1963):

	SOV	SVO
% languages	72% (181/253)	14% (26/190)

Other ramifications:

1. Because morphological endings are hard for second language learners (Lupyan & Dale, 2012), *Creoles are SVO*, even when some of the contact languages are SOV
2. Languages shift from SOV, case-marking to SVO during language contact: *Old English to modern English*
3. *Case-marking can be animacy-dependent:* Differential Object Marking languages. E.g. Farsi
4. *Word order can be animacy-dependent:* "Word order freezing", when case does not disambiguate semantic roles: SVO word order: e.g., Russian

Current work: Quantitative cross-linguistic corpora Futrell, Mahowald & Gibson



Corpora from 42 languages parsed into dependencies

- **Result 1:** All SOV languages in our set are case-marked, e.g., Japanese. Korean, Persian, Hindi, Turkish, Basque, Tamil

- **Result 2**: All un-case-marked languages in our set are SVO

  (Note that some SVO languages are case-marked: robust)

Conclusion: Language processing over a noisy channel

Suppose that language approximates an optimal code for information processing. This can potentially explain:

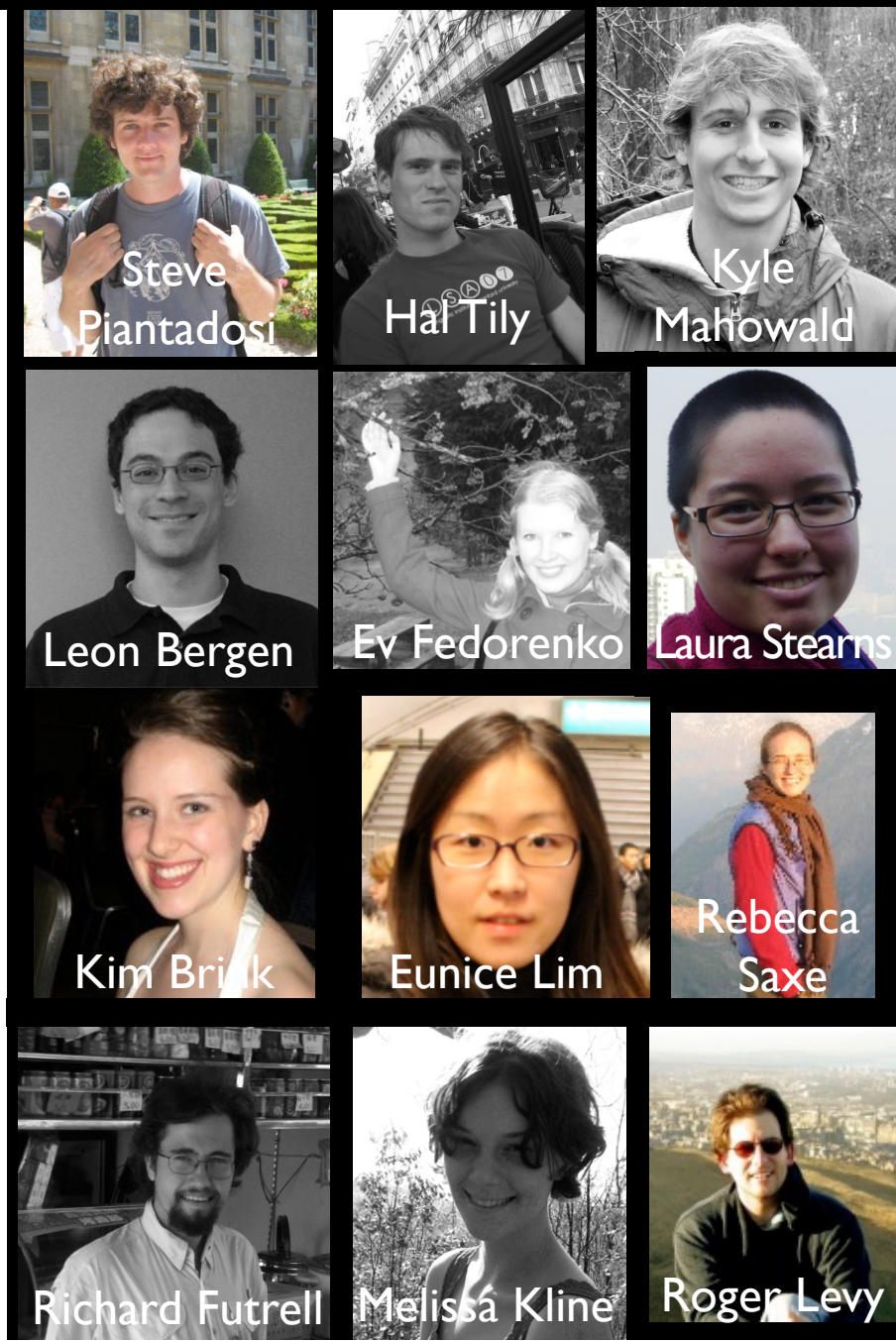
- Language use
  - Sentence interpretation (Gibson, Bergen & Piantadosi, 2013; Bergen & Gibson, 2013; Fedorenko, Stearns, Bergen, Eddy & Gibson, submitted; Gibson, Sandberg, Fedorenko, Bergen & Kiran, 2015)

- The evolution of language:
  - Syntax (Gibson, Piantadosi, Brink, Lim, Bergen & Saxe, 2013; Futrell, Hickey, Lee, Lim, Luchkina & Gibson., 2014)

Acknowledgments

- *National Science Foundation Grants from the linguistics program 0844472 (until* 2013); *1534318 (2015-2018)*

- *Collaborators:*
  - Sentence interpretation: Leon Bergen, Steve Piantadosi
  - L2 noisy channel: Ev Fedorenko, Richard Futrell, Kyle Mahowald, Caitlin Tan
  - Aphasia: Chaleece Sandberg, Ev Fedorenko, Swathi Kiran
  - ERPs: Ev Fedorenko, Laura Stearns, Marianna Eddy
  - Origin of word order: Steve Piantadosi, Kim Brink, Leon Bergen, Eunice Lim, Rebecca Saxe, Richard Futrell, Melissa Kline, Tina Hickey, Aldrin Lee, Elena Luchkina



Language processing over a noisy channel

Collaborators on these projects:

- Richard Futrell, MIT
- Steve Piantadosi, U Rochester
- Kyle Mahowald, MIT
- Leon Bergen, Stanford
- Ev Fedorenko, MIT, Harvard
- Chaleece Sandberg, Boston U
- Swathi Kiran, Boston U
- Laura Stearns, MIT
- Marianna Eddy, Tufts
- Kim Brink, MIT
- Eunice Lim, UCSD
- Rebecca Saxe, MIT
- Roger Levy, MIT