Class 5:
Language processing over a noisy channel

Ted Gibson
9.59J/24.905J

Review from last time: Mahowald et al. 2013

Words with a long/ short form (e.g., math, mathematics) are preferred as short in a supportive context.

Question: is this because of ***audience design:*** the desire to help out our conversation participants?

Answer: it's impossible to tell based purely on corpus data.

Evidence for audience design:

Clark (1996): Asking for directions: Speakers use words that are appropriate to listeners background knowledge. But maybe this difference reflects different lexical knowledge?

Wilkes-Gibbs & Clark (1992): participants coordinating over names for objects of "tangrams", moving them around. Directors who talk to one listener first and then have to describe the same tangrams to a naïve listener use more words with the naïve person. (Probably not an effect of lexicon.)

Language for communication?

- More controversial than some might think...

"The natural approach has always been: Is it well designed for use, understood typically as use for communication? I think that's the wrong question. The use of language for communication might turn out to be a kind of epiphenomenon. ... If you want to make sure that we never misunderstand one another, for that purpose **language is not well designed**, *because you have such properties as ambiguity. If we want to have the property that the things that we usually would like to say come out short and simple, well, it probably doesn't have that property."* (Chomsky, 2002, p. 107)

Ambiguity

Syntax: Frank shot the hunter with the shotgun.

Lexicon: run (polysemy); two/to/too (homophony)

Referential: He said that we should give it to them.

Ambiguity: A communicative *benefit*

- Ambiguity is only a problem *in theory*

  - Ambiguity is not a problem in normal language use, because context disambiguates (Wasow & Arnold, 2003; Wasow et al., 2005; Jaeger, 2006; Roland, Elman, & Ferreira, 2006; Ferreira, 2008; Jaeger, 2010).

- Piantadosi, Tily & Gibson (2012):

  - An information-theoretic proof that efficient communication systems will necessarily be globally ambiguous when context is informative about meaning

  - "ambiguity" potentially allows for re-use of easy linguistic elements:

    - *John wanted **to** run.*
    - *John went **to** school.*
    - *John wanted **two** dollars.*
   - *Sam wanted some money **too**.*

Ambiguity:
A communicative *benefit*

Because context disambiguates, we don't have to say so much:

When language is constructed to be as unambiguous as possible, with no other possible interpretations, what you get is legalese:

EXCEPT FOR THE LIMITED WARRANTY ON MEDIA SET FORTH ABOVE AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, THE APPLE SOFTWARE AND SERVICES ARE PROVIDED "AS IS" AND "AS AVAILABLE", WITH ALL FAULTS AND WITHOUT WARRANTY OF ANY KIND, AND APPLE AND APPLE'S LICENSORS (COLLECTIVELY REFERRED TO AS "APPLE" FOR THE PURPOSES OF SECTIONS 8 and 9) HEREBY DISCLAIM ALL WARRANTIES AND CONDITIONS WITH RESPECT TO THE APPLE SOFTWARE AND SERVICES, EITHER EXPRESS, IMPLIED OR STATUTORY, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES AND/OR CONDITIONS OF MERCHANTABILITY, OF SATISFACTORY QUALITY, OF FITNESS FOR A PARTICULAR PURPOSE, OF ACCURACY, OF QUIET ENJOYMENT, AND NON-INFRINGEMENT OF THIRD PARTY RIGHTS.

Language as efficient communication:
Shorter words are more ambiguous
Piantadosi, Tily & Gibson (2012)
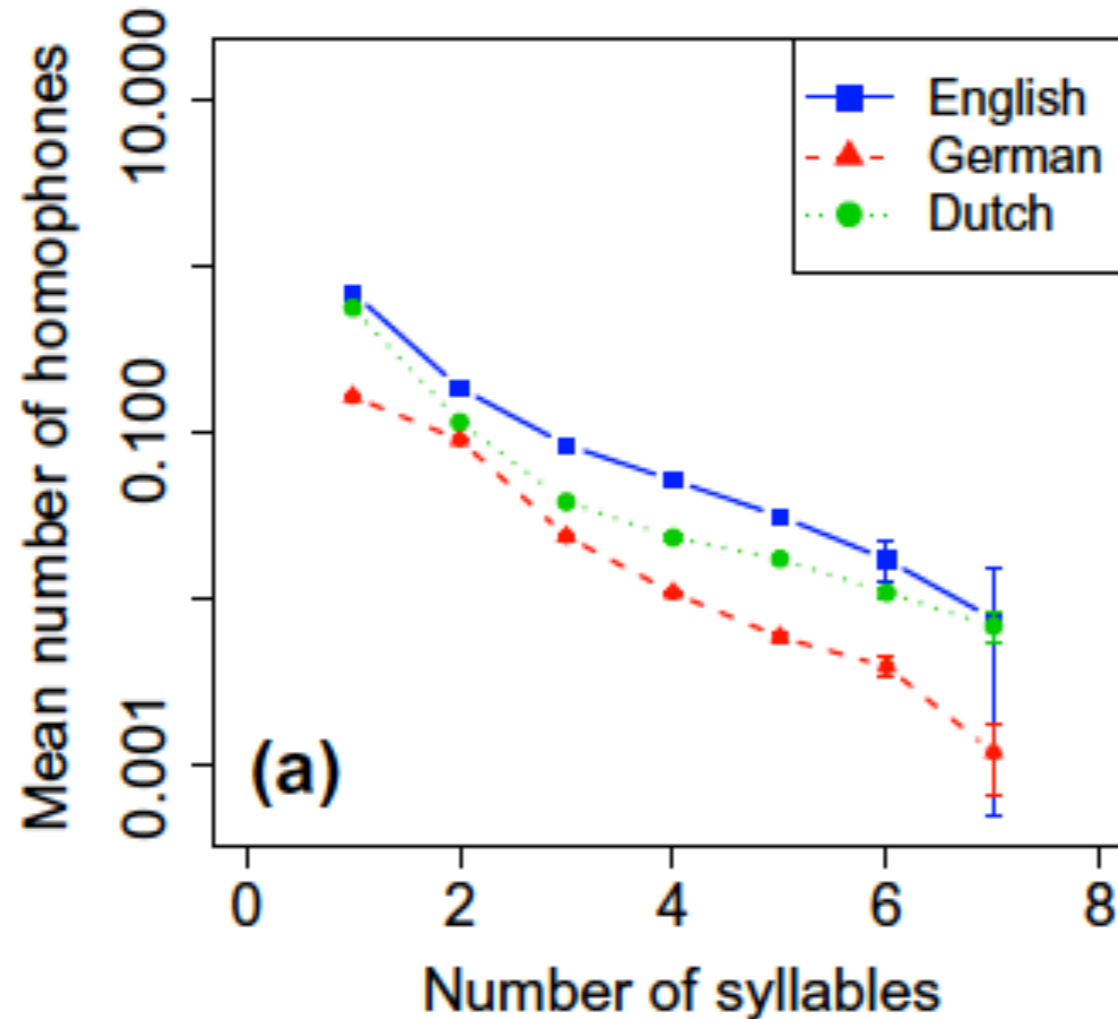
- Number of additional meanings each phonological form has, as a function of length.

- Shorter phonological forms having more homophones / meanings.

Ambiguity:
a communicative benefit

The existence of ambiguity out of context in human language (which is disambiguated by context) is explained by **information theory**.

In other approaches, the existence of ambiguity out of context is an _unexplained accident._

Noisy-channel models of comprehension

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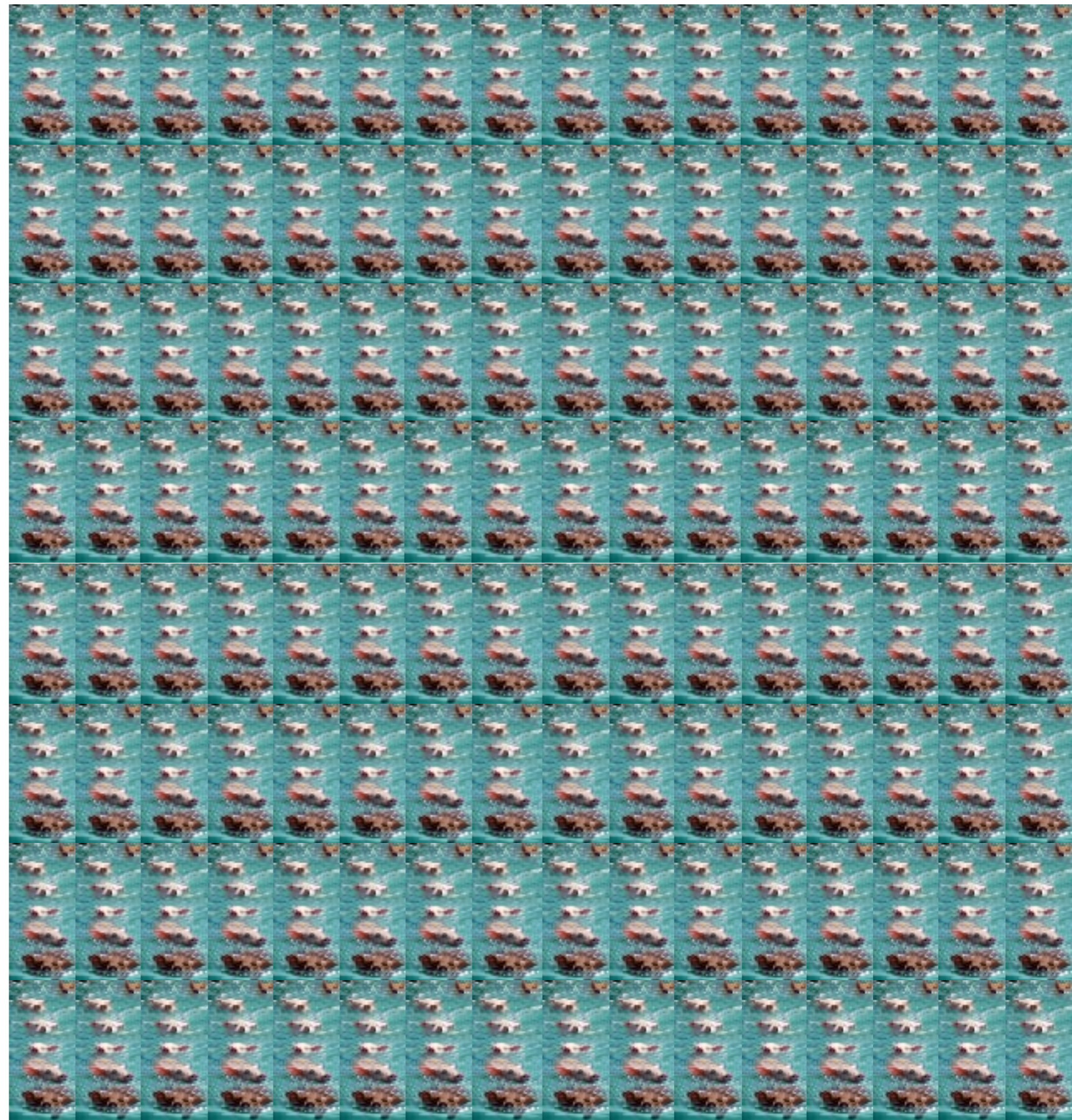
What Baralaba piggery owner Sid Everingham actually said was "30 sows and pigs", not "30,000 pigs".

The Morning Bulletin would like to apologise for this error, which was also reprinted in today's *Rural Weekly* CQ before the mistake was known.

Thirty sows and pigs in a river

Thirty thousand pigs in a river

Noisy-channel models of language comprehension: Mondegreens

Wikipedia: *American writer Sylvia Wright coined the term in her essay "The Death of Lady Mondegreen", published in Harper's Magazine in November 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"
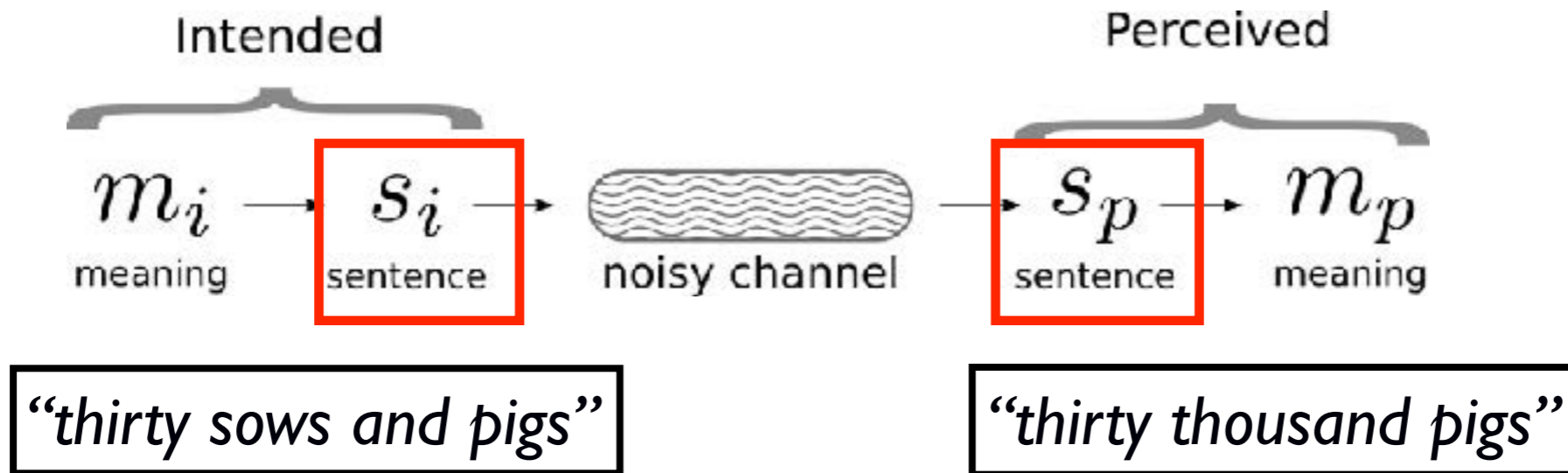*"There's a bathroom on the right"*

Mondegreens in songs

Creedence Clearwater Revival, "Bad Moon Rising"
"There's a bathroom on the right"

Manfred Mann, "Blinded by the light"
"wrapped up like a douche"

Mondegreens in songs

Creedence Clearwater Revival, "Bad Moon Rising"
"There's a bathroom on the right"

Manfred Mann, "Blinded by the light"
"wrapped up like a douche"

Jimi Hendrix, "Purple Haze"
"Excuse me while I kiss this guy"

Mondegreens in songs

Creedence Clearwater Revival, "Bad Moon Rising"
"There's a bathroom on the right"

Manfred Mann, "Blinded by the light"
"wrapped up like a douche"

Jimi Hendrix, "Purple Haze"
"Excuse me while I kiss this guy"

Rush, "Limelight"
"living in a fish island"

Rational inference in language: Noisy-channel models of language

Intended — Perceived

m_i (meaning) → s_i (sentence) → noisy channel → s_p (sentence) → m_p (meaning)

"thirty sows and pigs"   *"thirty thousand pigs"*

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)

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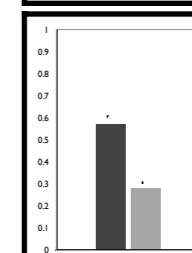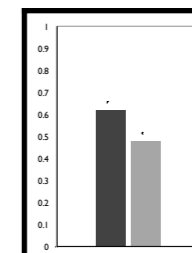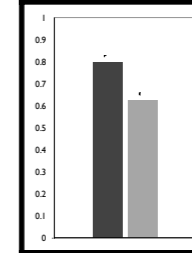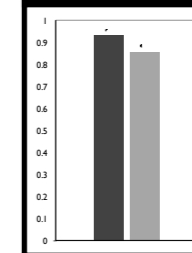   b. **Sentence:** *The mother gave the candle the daughter.*
   Question: Did the daughter receive something/someone?

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
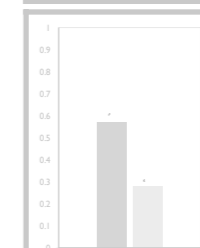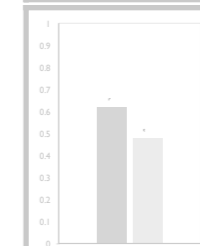4b. PO ⊡⊘ DO-goal: The mother gave the candle to/∅ the daughter.   1 deletion

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

Courtesy of Proceedings of the National Academy of Sciences. Used with permission. Source: Gibson, Edward, Leon Bergen, and Steven T. Piantadosi. "Rational integration of noisy evidence and prior semantic expectations in sentence interpretation." Proceedings of the National Academy of Sciences 110, no. 20 (2013): 8051-8056.

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
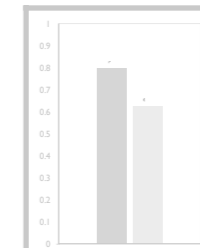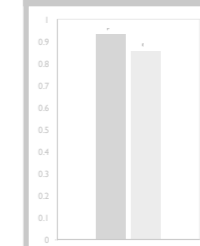4b. PO ☺☺ DO-goal: The mother gave the candle to/∅ the daughter. 1 deletion
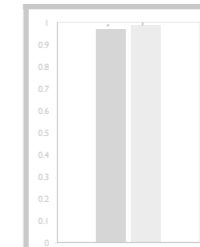
5a. DO ☺☺ PO-benef: The cook baked Lucy ∅/for a cake. 1 insertion
5b. PO ☺☺ DO-benef: The cook baked a cake for/∅ Lucy. 1 deletion

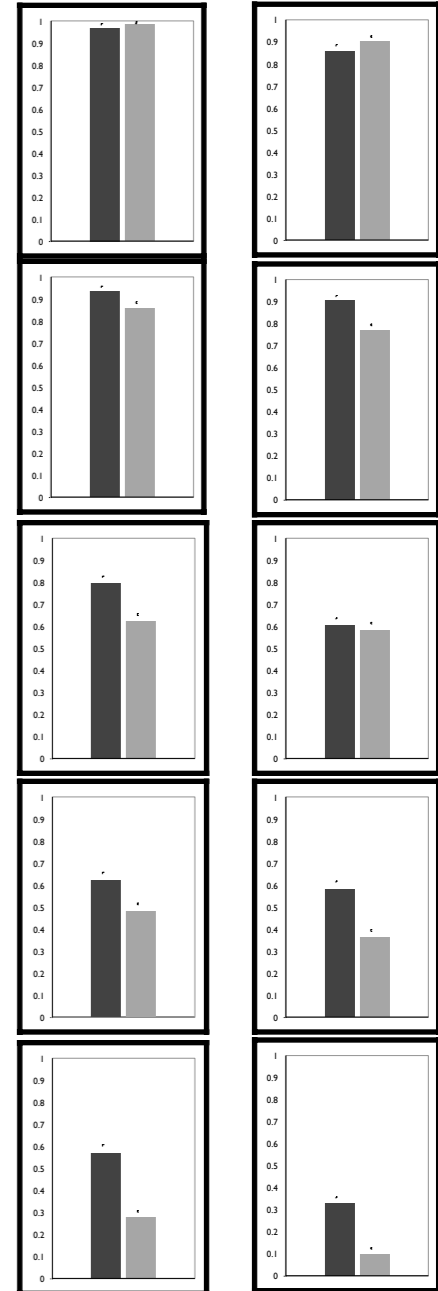**Prediction:** more noise should lead to greater reliance on likely meaning
Manipulation:
   add noise to 30 of the 60 fillers
   10 - extra function word; 10 - missing function word; 10 - local transpositions

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
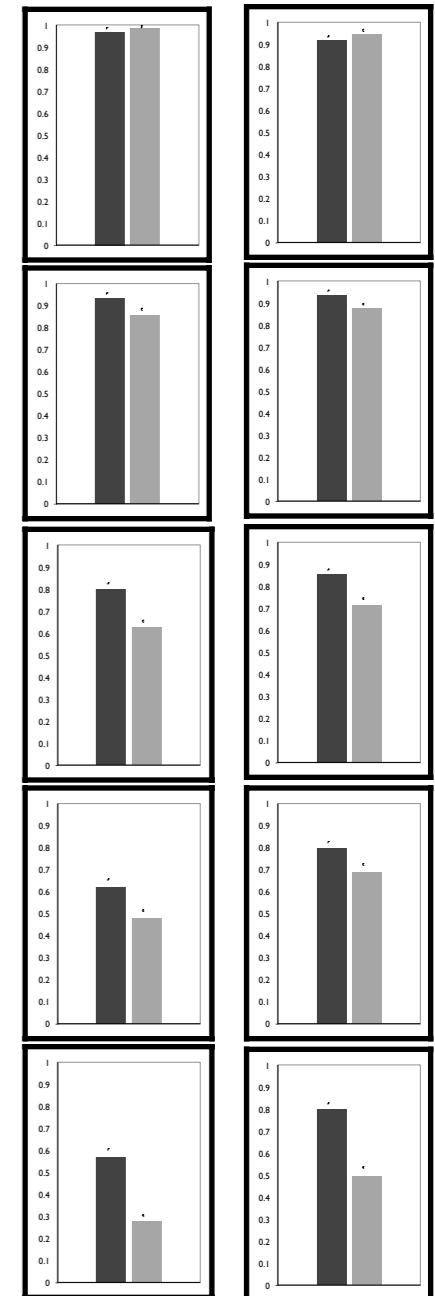4b. PO ⬛⊛ DO-goal: The mother gave the candle to/∅ the daughter.   1 deletion

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*)

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

A classic finding in the sentence production literature (Bock & Miller, 1991) and comprehension (Pearlmutter, Garnsey & Bock, 1999):

Agreement error _asymmetry_:

1. The key to the cabinets was / **were** on the table.
(Many errors for plural local noun)

2. The keys to the cabinet were / ??**was** on the table.
(Very few errors for singular local noun)

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

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Note it is currently unclear if either kind of attraction error is more common in natural speech / text.

It is easy to find examples of both plural and singular attraction errors in people's speech / writing:

Singular attractions:

The stairwells on the BCS Headquarters side of the building is in the process of being painted. (email to BCS department April 2013)

The consequences of that is ... (talk at MIT, Oct, 2013)

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

<u>Our claim:</u>

Agreement errors result from **rational misidentification of the preamble.**

The asymmetry between singular and plural head-nouns is explained by 2 factors:

- Deletions are much more likely than insertions (Gibson et al, 2013). *Thus agreement errors will occur more often when the head noun is singular.*
- Prior distribution of NP sequences: the singular-singular is much the most common sequence. *Thus there will be few errors confusing sing-sing as plural-sing.*

Plural-head/Singular-local   *The keys to the cabinet…*
Given the plural head noun, it is unlikely that the comprehender will infer that the plural-marking was produced by mistake, so unlikely to be pulled to the sing-sing.
Singular-head/Plural-local   *The key to the cabinets…*
Given the singular head noun, it is possible that the comprehender will think that the producer intended a plural / plural, hence producing an error.

9.59J/24.905J Lab in Psycholinguistics
Spring 2017