Language for communication: Language as rational inference

Ted Gibson Department of Brain and Cognitive Sciences MIT

Saarbruecken, Germany

April 17, 2015

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)



Language for Communication?

Contrary to Chomsky, we argue that language approximates an optimal code for human communication (Zipf, 1949).

This can potentially explain:

- the online behavior of language users (Genzel & Charniak, 2002; Aylett & Turk, 2004; Levy, 2005; Jaeger, 2006; Levy & Jaeger, 2007)
- the structure of languages themselves (e.g. Ferrer i Cancho & Sole, 2003; Ferrer i Cancho, 2006; Piantadosi, Tily, & Gibson, 2011; Gibson et al., 2013)

But what about the issue of **ambiguity**?

Ambiguity

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

Syntax: Frank shot the hunter with the shotgun.

<u>Referential</u>: 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).
- context disambiguates, e.g., word use:
 - John wanted **to** run.
 - John went **to** school.
 - John wanted **two** dollars.
 - Sam wanted some money **too**.
- Piantadosi, Tily & Gibson (2012): An information-theoretic proof that efficient communication systems will necessarily be globally ambiguous when context is informative about meaning (because short / easy items will get re-used in different contexts)

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**.

E.g., why do we re-use words? In part, to keep the code short.

In other approaches, the existence of ambiguity out of context is an <u>unexplained accident.</u>

Information theory and cross-linguistic universals and differences

• Words:

- Word length and information theory:
 - <u>Proposed universal</u>: Shorter words are more ambiguous
 - <u>Proposed universal</u>: Contextual predictability predicts word length across languages
- Information theory may help us to understand <u>cross-linguistic</u> <u>differences</u> in restricted semantic domains: **color**

• Syntax:

- Language comprehension in a noisy channel: the rational integration of noise and prior
 - language comprehension accuracy
 - on-line language comprehension: the P600 in ERPs
- <u>Proposed universal</u>: A noisy-channel proposal for aspects of word order evolution

Language / Communication:Words Piantadosi, Tily & Gibson (2011)



Zipf (1949): more frequent words are shorter:

- "Principle of least effort"
- Extension: more *predictable* words should be shorter.
 - e.g., to maintain Uniform Information Density (Aylett & Turk, 2004; Jaeger, 2006; Levy & Jaeger, 2007)
 - Estimate of predictability: n-grams (3-grams) over large corpora

Language for communication:Words Piantadosi, Tily & Gibson (2011)



More predictable words are shorter!

How does the effect arise?

- Is it just differences among broad classes of words like content vs. function words? Or within class too?
- look at long/short pairs (chimpanzee → chimp), which differ in length but are controlled for meaning



Mahowald, Fedorenko, Piantadosi and Gibson (Cognition 2013)



Info/Information theory

Using Google trigrams, we looked at average surprisal for long vs. short forms.

Mean surprisal for long forms (9.21) is significantly higher than mean surprisal for short forms (6.90) (P = .004 by Wilcoxon signed rank test)

Linear regression shows significant effect of length on surprisal (t = 2.76, P = .01) even when controlling for frequency.



Mahowald, Fedorenko, Piantadosi and Gibson (Cognition 2013)

Information theory and cross-linguistic universals and differences

• Words:

- Word length and information theory:
 - <u>Proposed universal</u>: Shorter words are more ambiguous
 - <u>Proposed universal</u>: Contextual predictability predicts word length across languages
- Information theory may help us to understand <u>cross-linguistic</u> <u>differences</u> in restricted semantic domains: **color**

• Syntax:

- Language comprehension in a noisy channel: the rational integration of noise and prior
 - language comprehension accuracy
 - on-line language comprehension: the P600 in ERPs
- <u>Proposed universal</u>: A noisy-channel proposal for aspects of word order evolution

Information theory and color words

Gibson, Jara-Ettinger, Bergen, Piantadosi, Gibson & Conway, 2015

Information theory applied to words in a restricted semantic domain: **Color**



Information theory and color words

Gibson, Jara-Ettinger, Bergen, Piantadosi, Gibson & Conway, 2015

Information theory applied to words in a restricted semantic domain: **Color**

Some languages have more color terms than other languages



- English: I I "basic" color terms: black, white, red, green, yellow, blue, brown, pink, orange, purple, grey
- Berinmo: 5 "basic" color terms (Roberson et al. 2000; Davidoff et al, 1999)
- Dani: 2 color terms (Rosch Heider 1972): dark / light or "black" / "white"

Berlin & Kay (1969): The World Color Survey (WCS)

330 colors in WCS color grid: Approximately **a subset relation** among sets of color terms across languages:



Berlin & Kay discuss the distribution of color terms in terms of "basic" color terms: basic color terms are thought to be **visual-perception** based: the most salient colors in the color space (e.g., Kay & Maffi, 1999)

These are the **modal colors** in the WCS

Cluster analysis based on a particular perceptual space can predict the subsets, using modal colors from the WCS (Regier, Kay & Khetarpal, 2007)

The approximate subset relationship across languages is suggestive evidence for the perceptual hypothesis

Puzzles for the perception-based hypothesis

I. Why are there exceptions to the subset ordering (e.g., Berinmo vs. English)?



1

1

3

Kel

Puzzles for the perception-based hypothesis

2. Why do more industrialized cultures have more color words?

Kay & Maffi, 1999: "As technology develops, the increased importance of color as a distinguishing property of objects appears to be an important factor in causing languages to add basic color terms, i.e., to refine the lexical partition of the color domain (Casson 1997)."

We have to appeal to culture to explain color distributions anyway. So maybe we can do away with the notion of "basic" color term: Color terms are just experience-based (Deutscher, 2010; cf. Gladstone, 1860)

People with different experiences with colors and color labeling will be different at their ability to use color words. E.g., painters, interior designers

Information theoretic account: the mode doesn't matter: the *distribution* matters

An information-theory account of color terms

Two ways to improve the information content of color labels:

- More color labels (Berlin & Kay)
- More consistent use of color terms, independent of the number of labels

Case study: Tsimane', in the Bolivian Amazon vs. English

It turns out that Tsimane' has about the same number of "basic" terms as English, but people are much more variable in using these terms

The Amazon basin



Color naming in Tsimane' & English

Task: "What color is this chip?"

N=61;80 colors

31 total colors were given; median of 11 / subject

Highly variable answers across people

green-knowers call leaves, grass and the sky "green" blue-knowers call leaves, grass and the sky "blue"

Tsimane': 4.88 bits of entropy in 80 color grid English: 3.78 bits (N=35)

If 10 color words were used optimally to convey 80 colors, 3 bits of entropy $(80/10 = 2^3)$ would remain.









Comparing the **consistency of term use** independent of the number of colors

- Compute the entropy for each possible number of labels, keeping the most frequent n color terms in each language, and entering "unknown" for all less frequent terms.
- English has less entropy than Tsimane' for any number of labels, e.g., 0.9 bits less for 8 labels



What accounts for differences in color-term informativeness? Culture, industrialization?

People in industrialized cultures have more experience with arbitrarily colored artifacts

Color terms should therefore be more common in industrialized cultures. No corpus of Tsimane' to test this yet.

Object-naming experiment:

Color term usage on a contrastive labeling task (Sedivy, 2003), using familiar natural objects (e.g., bananas) and synthetic colored objects (e.g., cups)

For both kinds of objects, the Tsimane' hardly ever label colors: 5.5% overall vs. 53.5% for the same objects in English.



Information-theory and color terms

- I. The notion of "basic" color term is probably problematic (Saunders, 2000)
- 2. An information theoretic account allows us to see finer grained differences among languages than restricting attention to modal / basic color terms
- 3. Perhaps culture is sufficient to explain cross-linguistic differences, with no appeal to perception (Deutscher, 2010; cf. Gladstone, 1860)

Information theory and cross-linguistic universals and differences

• Words:

- Word length and information theory:
 - <u>Proposed universal</u>: Shorter words are more ambiguous
 - <u>Proposed universal</u>: Contextual predictability predicts word length across languages
- Information theory may help us to understand <u>cross-linguistic</u> <u>differences</u> in restricted semantic domains: **color**

• Syntax:

- Language comprehension in a noisy channel: the rational integration of noise and prior
 - language comprehension accuracy
 - on-line language comprehension: the P600 in ERPs
- <u>Proposed universal</u>: A noisy-channel proposal for aspects of word order evolution

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

gery 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.

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 $P(s_i | s_p)$ 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

- 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): noisy environment;
 - (b) production errors (mis-speaking/mis-typing)
- Classic issue in signal processing (e.g., Shannon, 1948)
- Previous work: Noisy-channel effects on reading (Levy et al., 2009)

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)

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

"Major" change alternations:

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

Active \rightarrow Passive (2 insertions): The girl kicked the ball. \rightarrow The girl <u>was</u> kicked <u>by</u> the ball.

"Minor" change alternations:

PO-goal \rightarrow DO-goal (I deletion): The mother gave the candle <u>to</u> the daughter. \rightarrow The mother gave the candle the daughter.

DO-goal \rightarrow PO-goal (1 insertion): The mother gave the daughter the candle. \rightarrow The mother gave the daughter <u>to</u> the candle.



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)



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

 1a. Passive → Active: 1b. Active → Passive: 	The ball was/Ø kicked by/Ø the girl. The girl Ø/was kicked Ø/by the ball.	2 deletions 2 insertions
2a. Subj-loc → Obj-loc: 2b. Obj-loc → Subj-loc:	\emptyset /Onto The cat jumped onto/ \emptyset a table. Onto/ \emptyset the table jumped \emptyset /onto a cat.	insertion, 1 deletion deletion, 1 insertion
3a. Intrans → Trans: 3b. Trans → Intrans:	The tax law benefited Ø/from the businessma The businessman benefited from/Ø the tax lav	n. 1 insertion v. 1 deletion
4a. DO → PO-goal: 4b. PO → DO-goal:	The mother gave the daughter \emptyset /to the candle The mother gave the candle to/ \emptyset the daughte	e. 1 insertion r. 1 deletion
5a. DO ➔ PO-benef: 5b. PO ➔ DO-benef:	The cook baked Lucy Ø/for a cake. The cook baked a cake for/Ø Lucy.	1 insertion 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

1a. Passive → Active: 1b. Active → Passive:	The ball was/ \emptyset kicked by/ \emptyset the girl. The girl \emptyset /was kicked \emptyset /by the ball.	2 deletions 2 insertions	
2a. Subj-loc ➔ Obj-loc: 2b. Obj-loc ➔ Subj-loc:	\emptyset /Onto The cat jumped onto/ \emptyset a table. Onto/ \emptyset the table jumped \emptyset /onto a cat.	1 insertion, 1 deletion 1 deletion, 1 insertion	
3a. Intrans → Trans: 3b. Trans → Intrans:	The tax law benefited \emptyset /from the businessma The businessman benefited from/ \emptyset the tax la	an. 1 insertion aw. 1 deletion	
4a. DO ➔ PO-goal: 4b. PO ➔ DO-goal:	The mother gave the daughter \emptyset /to the cand The mother gave the candle to/ \emptyset the daughte	le. 1 insertion er. 1 deletion	
5a. DO ➔ PO-benef: 5b. PO ➔ DO-benef:	The cook baked Lucy Ø/for a cake. The cook baked a cake for/Ø Lucy.	1 insertion 1 deletion	

More syntactic errors decreased the reliance on syntax: 56.1% vs. 42.7 for the minor-change alternations

Manipulations of semantic / plausibility prior:

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

Expt la - le: 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)

 1a. Passive → Active: 1b. Active → Passive: 	The ball was/Ø kicked by/Ø the girl. The girl Ø/was kicked Ø/by the ball.	2 deletions 2 insertions	
2a. Subj-loc ➔ Obj-loc: 2b. Obj-loc ➔ Subj-loc:	\emptyset /Onto The cat jumped onto/ \emptyset a table. Onto/ \emptyset the table jumped \emptyset /onto a cat.	1 insertion, 1 deletion 1 deletion, 1 insertion	
3a. Intrans → Trans: 3b. Trans → Intrans:	The tax law benefited Ø/from the businessma The businessman benefited from/Ø the tax la	n. 1 insertion w. 1 deletion	
4a. DO ➔ PO-goal: 4b. PO ➔ DO-goal:	The mother gave the daughter \emptyset /to the candle The mother gave the candle to/ \emptyset the daughter	e. 1 insertion er. 1 deletion	
5a. DO ➔ PO-benef: 5b. PO ➔ DO-benef:	The cook baked Lucy Ø/for a cake. The cook baked a cake for/Ø Lucy.	1 insertion 1 deletion	1 3 0 0 0 0 0 0 0 0 0 0 0 0 0

More implausible materials increased the reliance on syntax: 56.1% vs. 72.6 for the minor-change alternations

Summary:

Evidence for a noise model:

I. More changes from one alternative to another leads to lower likelihood that the alternative will be considered.

- 2. Deletions are preferred to insertions.
- 3. Increasing the noise increases the reliance on semantics.

Evidence for priors:

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



Information theory and cross-linguistic universals and differences

• Words:

- Word length and information theory:
 - <u>Proposed universal</u>: Shorter words are more ambiguous
 - <u>Proposed universal</u>: Contextual predictability predicts word length across languages
- Information theory may help us to understand <u>cross-linguistic</u> <u>differences</u> in restricted semantic domains: **color**

• Syntax:

- Language comprehension in a noisy channel: the rational integration of noise and prior
 - language comprehension accuracy
 - on-line language comprehension: the P600 in ERPs
- <u>Proposed universal</u>: A noisy-channel proposal for aspects of word order evolution

The noisy-channel proposal applied to aphasic comprehension

Hypothesis: Aphasic patients' perception is noisier than that of healthy individuals. In maximizing $P(s_i | s_p)$, aphasics will rely more on their prior distribution $P(s_i)$ over plausibly intended sentences.

(Gibson, Sandberg, Fedorenko, Bergen & Kiran, submitted)

Prediction:

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

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.

- The P600 occurs relatively late because it indexes correction.
- A P600 is predicted when a **correction** can be made
 - "Syntactic" violations: Every Monday he mow / mows the lawn
 "Semantic P600's": The hearty meal was devouring / devoured ...
 Orthographic errors: fone / phone
- No P600 is predicted when a **correction** cannot be made

•Semantic violations: I take my coffee with cream and dog / sugar (?)

Information theory and cross-linguistic universals and differences

• Words:

- Word length and information theory:
 - <u>Proposed universal</u>: Shorter words are more ambiguous
 - <u>Proposed universal</u>: Contextual predictability predicts word length across languages
- Information theory may help us to understand <u>cross-linguistic</u> <u>differences</u> in restricted semantic domains: **color**

• Syntax:

- Language comprehension in a noisy channel: the rational integration of noise and prior
 - language comprehension accuracy
 - on-line language comprehension: the P600 in ERPs
- <u>Proposed universal</u>: A noisy-channel proposal for aspects of word order evolution

Syntax:

Word order across the world's languages

\bigcirc	1. Subject-Object-Verb (SOV)	497
	2. Subject-Verb-Object (SVO)	435
\bigcirc	3. Verb-Subject-Object (VSO)	85
\diamond	4. Verb-Object-Subject (VOS)	26
\diamond	5. Object-Verb-Subject (OVS)	9
\diamond	6. Object-Subject-Verb (OSV)	4
\bigcirc	7. Lacking a dominant word order	172
	total	1228

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

- SO is a near universal:Almost no OS languages
 - Plausible explanation: people assume that the agents (subjects) occur before the patients (objects) (MacWhinney, 1975).
- OV / VO are almost equally balanced:
 - SOV most common: 47.1% of languages with a dominant word order
 - SVO is next: 41.2% of languages with a dominant word order

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





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.





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





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

- Languages are attempts to create codes that are both natural and robust to noise (Shannon, 1948; Shannon, 1951; Aylett & Turk, 2004; Levy, 2008; Jaeger, 2010; Piantadosi et al., 2011).
- **SOV** is the default, most natural order.
 - The agent is likely to come before the patient (e.g., MacWhinney, 1975);
 - An old-before-new bias means that the verb is likely final (Goldin-Meadow et al. 2008; Schouwstra et al., 2011; Paul, 1880; Jackendoff, 1972; Gibson et al. 2013)
- Switch to **SVO** when needed to allow robust communication over a noisy channel.

Probabilities of Error: SOV



Probabilities of Error: SVO



The noisy channel hypothesis: Why aren't all languages SVO?

 SOV languages should tend to be <u>case-marked</u>, while SVO languages need not be Dryer (2002) (cf. Greenberg, 1963):

	SOV	SVO	VSO
% languages	72% (181/253)	4% (26/190)	47% (28/59)

The noisy channel hypothesis: Cross-linguistic predictions

- <u>Case-marking can be animacy-dependent</u>: approximately 300 languages that have Differential Object Marking (DOM) (Aissen, 2003) where animate, but not inanimate, direct objects are case-marked (cf. Fedzechkina, Jaeger & Newport, 2012)
- <u>Word order</u> can be animacy-dependent. Among languages with relatively free word order many demonstrate "word order freezing": when case does not disambiguate semantic roles, SVO word order is preferred. E.g., Russian (Jakobson, 1936; Bouma, 2011) and Kata Kolok, a sign language in northern Bali, Indonesia (Marsaja, 2008; Meir et al., 2010).
- Fixed word order should primarily be found in SVO languages, and non-SVO languages should generally have free-er word order.
 According to Dryer (personal communication) this appears to be true.

The communicative basis for syntax Explaining the origin of different word orders

- The origin of word order
 - When inventing a communication system, people either
 - invent case-marking to keep the NPs distinct, and then produce SOV word order
 - don't invent case-marking. Then they produce SVO to minimize confusion



Current work: Cross-linguistic corpus collection *Futrell, Mahowald & Gibson*



 Corpora from 34 languages parsed into dependencies, from NLP sources: the HamleDT and UDT; cf. WALS (Dryer 2005)

• Family / Region

Indo-European (IE)/West-Germanic; IE/North-Germanic; IE/Romance; IE/Greek; IE/West Slavic; IE/South Slavic; IE/East Slavic; IE/Iranian; IE/Indic; Finno-Ugric/Finnic; Finno-Ugric/Ugric; Turkic; West Semitic; Dravidian; Austronesian; East Asian Isolate (2); Other Isolate (1)

- **Result I**:All languages minimize dependency distances (c.f. Hawkins, 1994; Gibson, 1998)
- **Result 2**: Many case-marked languages have low variability wordorder: these are all verb-final languages (as predicted by the noisychannel hypothesis); Higher variability word-order languages also have case marking, but tend to be SVO.

Conclusion: Language for communication

Language approximates an optimal code for human communication (Zipf, 1949). This can potentially explain:

- The evolution of language:
 - Words (Piantadosi, Tily, & Gibson, 2011, 2012; Gibson, Jara-Ettinger, Bergen, Piantadosi, Gibson & Conway, in preparation)
 - Syntax (Gibson, Piantadosi, Brink, Lim, Bergen & Saxe, 2013; Futrell, Hickey, Lee, Lim, Luchkina & Gibson., 2014)
- Language use
 - Sentence interpretation (Gibson, Bergen & Piantadosi, 2013; Bergen & Gibson, 2013; Fedorenko, Stearns, Bergen, Eddy & Gibson, submitted; Gibson, Sandberg, Fedorenko, Bergen & Kiran, submitted)

Acknowledgments

- National Science Foundation Grant from the linguistics program 0844472
- Collaborators:
 - Words & ambiguity: Steve Piantadosi, Hal Tily, Kyle Mahowald
 - <u>Color words</u>: Julian Jara-Ettinger, Leon Bergen, Steve Piantadosi, Mitchell Gibson, Bevil Conway.
 - Sentence interpretation: Leon Bergen, Steve Piantadosi, Ev Fedorenko, Laura Stearns, Marianna Eddy, Roger Levy, Chaleece Sandberg, Swathi Kiran
 - Origin of word order: Steve Piantadosi, Kim Brink, Leon Bergen, Eunice Lim, Rebecca Saxe, Richard Futrell, Melissa Kline, Tina Hickey, Aldrin Lee, Elena Luchkina

