On the inadequacy of Gibson’s 1992 review of the Competition Model or Functionalist Linguistics and Language Universals

Ted Gibson
Department of Brain and Cognitive Sciences
MIT

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

CMU, June 8, 2019
In this landmark book, M&B propose a research program which provides a uniform framework for explaining all manifestations of psycholinguistic effects, ranging from language acquisition to sentence processing to sentence production;

and they test this framework in a variety of languages, something not previously attempted to this degree.
Gibson (1992) review

Summary: Quite a negative review.

Disappointing: Not constructively engaging with the insights and the vast amount of interesting data and insights that this group had gathered.
Re-evaluating Gibson’s (1992) criticisms

1. The Competition Model does not address Chomsky’s Poverty of the Stimulus argument adequately: How can language be learned like this?

2. The method of examining how people comprehend ungrammatical sentences like “Kicks the pencil the cow” doesn’t tell us about the grammar of a language.

Later: My lab’s research, inspired by the Competition Model: its functional roots and its methods

1. Information processing and cross-linguistic universals
2. Language processing in a noisy-channel
Syntactic islands: Arguments for Innate Universal Grammar

There is a relationship between the declarative (non-extracted) version and the extracted / long-distance one

E.g., Extraction from objects:
The baseball player loved the color of the sportscar.

Which sportscar did the baseball player love the color of __?

How could we learn this relationship?

**Functionalist account:** Construction hypothesis. We learn the meanings of linguistic forms (Declaratives; Wh-questions; Relative clauses; Clefts; etc.) from simple exposure

A wh-question asks about the identity of the extracted element
Syntactic islands: Arguments for Innate Universal Grammar

Meaning hypothesis: Unacceptable extractions have strange meanings (Ertesick-Shir 1977; Erteschik-Shir & Lappin 1979; Kuno 1987; Ambridge & Goldberg, 2008; Ambridge et al., 2014; Goldberg, 2016): Conflict in information structure.

That Chris liked Alex might bother Jerry.

E.g., *Who might [that Chris liked] bother Jerry? (Goldberg 2016)

Background constituent constraint: it is infelicitous to extract an element out of a backgrounded constituent (Goldberg, 2006, 2016).
Syntactic islands: Arguments for Innate Universal Grammar

General prediction of meaning constraints: If meaning is the source of the unacceptability of long-distance extractions, then different extraction-types (with different meanings) might vary in their acceptability, depending on their meaning (e.g., RC vs. WHQ)

Chomsky (1973): Functionalist accounts will not work, because some extraction configurations are impossible, independent of the meaning of the construction.

These are “island configurations” (Ross 1967), E.g., Extraction from subjects: “subject islands”

?*Which sportscar did the color of _ delight the baseball player? (cf. The color of the sportscar delighted the baseball player)
Syntactic islands: Arguments for Innate Universal Grammar

Chomsky’s claim: If the meaning of the construction was relevant to the acceptability of the extraction, then other constructions (with different meanings) would allow the extraction.

E.g., WHQs have focused meaning; RCs have Backgrounded meaning (RC)

All constructions with long-distance dependencies are unacceptable in island configurations; So their meaning can’t be the reason they are bad.

Therefore, island constraints must be innate, part of UG
Syntactic islands: Problems with the UG account

1. The only evidence for these purported cross-linguistic islands as part of UG comes from English and a few closely related languages.

If these are true universals, then we should see evidence in all languages.
2. The generalization turns out not to hold, when we look for comparable cross-construction evidence in WHQs and RCs, **even in English:**

Extraction from subject is not good in WHQs:
WHQ object: Which sportscar did the baseball player love the color of?
WHQ subject: ?? Which sportscar did the color of delight the baseball player?

**But extraction from subject is fine for Relative clauses, contrary to claim:** (Abeillé, Hemforth, Winckel & Gibson, 2018)

Relative clauses:
object: The dealer sold a sportscar, of which the baseball player loved the color.
subject: √ The dealer sold a sportscar, of which the color delighted the baseball player.

Same for French & Italian
Chomsky’s claim: All constructions with long-distance dependencies are unacceptable in island configurations.

This is **false**: On the contrary, as predicted by the functionalist account, when we change the construction, the acceptability of the extraction may change.

Extractions from subjects are not good for WHQs (because this violates the meaning constraint of a WHQ: can’t focus a topic: discourse violation)

Extractions from subjects are ok for Relative clauses (because there is no meaning violation when forming an RC from a topic).
Consequences for language learnability

If other islands can be explained as meaning constraints (with potential differences across constructions) then there is no learning puzzle:

Form-meaning associations can be learned by exposure for syntactic constructions just as for words

WHQ: Seeking focused information from interlocutor
RC: Specifying backgrounded / topicalized information

2. The method of examining how people comprehend ungrammatical sentences like “Kicks the pencil the cow” doesn’t tell us about the grammar of a language.

The potential issue is that there might a categorical effect of grammaticality: grammatical vs. ungrammatical

Then studying ungrammatical things may tell us nothing about the category of grammatical things

But there is no evidence of a categorical effect of grammaticality: rather, there appears to be a probabilistic continuum from very common to very rare (like words)
Each row represents a 2-condition experiment from Linguistic Inquiry (2000-2010); 12 items; 20 subjects.

Method: acceptability judgement 1-7, z-scored within individuals. Obtain mean z-score for each item in each contrast, and averaged these to give an overall z-score for the ‘acceptable’ sentence and for the ‘unacceptable’ sentence in each contrast.

The effect size is the difference between these two z-scores.
Following up Sprouse et al. (2013): Mahowald, Graff, Hartman & Gibson (2016)

\[ z\text{-score:}\ (value - mean) / sd \]
Re-evaluating Gibson’s (1992) criticisms

To the extent that we can find a theory that explains both rare and common linguistic materials, we have a better theory.
Information processing and cross-linguistic universals

My lab’s research is inspired by both the functional roots and the methods in the Competition Model:

1. Information processing and cross-linguistic universals (see Gibson et al., Trends in Cognitive Science, 2019)

2. Language processing in a noisy-channel (see Levy et al., 2009; PNAS; Gibson, Bergen & Piantadosi, 2013, PNAS).
### Information processing and cross-linguistic universals

#### Words: Language as communication

1. **Proposed universal**: Contextual predictability predicts word length across languages (Piantadosi, Tily & Gibson, 2011, PNAS)

#### Words: Communicative efficiency:

2. Invent words for properties of objects that differentiate them from other objects
   - **Domain of investigation**: Color terms
   - **Color naming across languages reflects color use** (Gibson et al., 2017, PNAS)

#### Syntax: Information processing / communication:

3. **Proposed universal**: Languages minimize dependency lengths (Futrell, Mahowald & Gibson, 2015, PNAS)
Correction

THERE was an error printed in a story titled "Pigs float down the Dawson" on Page 11 of yesterday's Bulletin. The story, by reporter Daniel Burdon, said more than 30,000 pigs were floating down the Dawson River.

What Baralaba pig farmer owner Sid Evans said was "30,000 pigs not 30,000 pigs."
Rational inference in language: Noisy-channel models of language

Intended

\( m_i \)  \( \rightarrow \)  \( s_i \)  \( \rightarrow \)  noisy channel

Perceived

\( s_p \)  \( \rightarrow \)  \( m_p \)

“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 \( P(s_i) \times P(s_i \rightarrow s_p) \)

“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.
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’ perception 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.


**Prediction:**

Aphasics will rely on semantics more than healthy individuals, in both major-edit (active-passive) and minor-edit alternations (DO-PO).
Word order across the world’s languages:
SOV and SVO are most common

- 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
Why SOV and SVO so dominant?

(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*
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):

<table>
<thead>
<tr>
<th></th>
<th>SOV</th>
<th>SVO</th>
</tr>
</thead>
<tbody>
<tr>
<td>% languages</td>
<td>72% (181/253)</td>
<td>14% (26/190)</td>
</tr>
</tbody>
</table>

Other ramifications:

1. Languages shift from SOV, case-marking to SVO during language contact: *Old English* to *modern English*
2. **Case-marking can be animacy-dependent**: Differential Object Marking languages. E.g. Farsi
3. **Word order can be animacy-dependent**: “Word order freezing”, when case does not disambiguate semantic roles: SVO word order: e.g., Russian
Conclusions: Re-evaluating Gibson’s (1992) criticisms

1. Chomsky’s Poverty of the Stimulus doesn’t pan out: The evidence for innate syntactic knowledge doesn’t yet exist.

A simpler hypothesis is the functionalist view: Humans acquire form-meaning pairs: Words, combination rules. No need for innate syntax.

2. The method of examining how people comprehend ungrammatical sentences like “Kicks the pencil the cow” plausibly tells us a lot about the rules of language.

Thank you

(And sorry again for the hassle from 27 years ago)