Explain me this:
Learning what not to say
Adele Goldberg
adele@princeton.edu

Need to categorize

Form ~ function pairings: constructions

Constructions at varying levels of complexity and abstraction

<table>
<thead>
<tr>
<th>Word</th>
<th>e.g., Bosnia, break, understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word (partially filled)</td>
<td>e.g., pre-N, V-ing</td>
</tr>
</tbody>
</table>

Constructions at varying levels of complexity and abstraction

<table>
<thead>
<tr>
<th>Word</th>
<th>e.g., Bosnia, break, understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word (partially filled)</td>
<td>e.g., pre-N, V-ing</td>
</tr>
<tr>
<td>Idiom (filled)</td>
<td>e.g., Got milk?, give the Devil his due</td>
</tr>
<tr>
<td>Idiom (partially filled)</td>
<td>e.g., Jog &lt;someone&gt;'s&gt; memory, send &lt;someone&gt; to the cleaners</td>
</tr>
</tbody>
</table>
Constructions at varying levels of complexity and abstraction

<table>
<thead>
<tr>
<th>Level</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>e.g., lost, break, misunderstanding</td>
</tr>
<tr>
<td>Word (partially filled)</td>
<td>e.g., pre-N, V-ing</td>
</tr>
<tr>
<td>Idiom (filled)</td>
<td>e.g., Got milk, give the Devil his due</td>
</tr>
<tr>
<td>Idiom (partially filled)</td>
<td>e.g., jog &lt;someone&gt; - money, sold &lt;someone&gt;</td>
</tr>
<tr>
<td>Unusual constructions</td>
<td>The Xer the Yer (e.g., The more you think about it, the less you understand)</td>
</tr>
<tr>
<td>(partially filled)</td>
<td>Scream construction (e.g., What an I/&lt;**king Jiminy Cricket?)</td>
</tr>
<tr>
<td>(unfilled)</td>
<td></td>
</tr>
<tr>
<td>Ditransitive construction</td>
<td>e.g., He gave her a fish taco; He baked her a muffin.</td>
</tr>
<tr>
<td>Subj V Obj Obj</td>
<td></td>
</tr>
<tr>
<td>Passive: Subj aux VPpp (PPhy)</td>
<td>e.g., The armadillo was hit by a car.</td>
</tr>
</tbody>
</table>

Certain things, one just shouldn’t say

I haven’t seen you for a long time—are you pregnant?
I only care about my grade in this course.
You look so old!
Who circumcised YOU!?

Positive evidence: types of expressions that are witnessed.

Varied type frequency
Semantic variability
Semantic similarity of target sentence to closest attested instance

Assume you can say these sentences:

Scraper--nu the vip the hug.
Load--nu the yih the vork.
Flip--nu the loof the mln.

How likely is it that you can also say:

Rumple--nu the phob the jirn.

Subjects gave likelihood estimate from 0-100%

4 Studies (N = 55 distinct MT participants for each)
Results analyzed using mixed linear models, subjects & items as random effects.

Our quarry: how do speakers learn when they can generalize a construction and when they cannot?

How do we learn what not to say?
Cat g o r izat io n

Similarity space

Cov erage: Degree to which the category has been attested.


---

Increased type frequency increases coverage

Coverage is low

Similarity determined by Latent Semantic Analysis (Landauer 1998)


---

Lack of semantic variability of attested tokens inhibits centralization if no instance falls outside cluster.


Semantic (and phonological) categorization goes a long way to explaining which new instances sound acceptable.

Still, there remain odd gaps in what we can say (e.g., Braine 1970; Lakoff 1970; Baker 1979; Bever 1988; Pinker 1989; Goldberg 1995; Ambridge et al. 2008; 2009; 2011; 2012)

??She explained him the story.
She told him the story.

??He vanished the rabbit.
He hid/banished the rabbit.

??She considered to go.
She wanted/hoped/planned to go.

??The asleep boy
The astute/sleeping boy

Reliable direct feedback is not available

Me loves you, Mommy.

I have just completed a colorful mural on my bedroom wall with indelible markers.

Is this how speakers learn to avoid saying:

??She disappeared something.?

Conservatism via Entrenchment

People use verbs only in ways they've heard them used before.

High overall verb frequency predicts grammatical inflexibility

Verb Frequency Effect

??The magician vanished the rabbit.
>

??The magician disappeared the rabbit.
But speakers can be creative

“I grabbed the guy by the collar of his flowered shirt, popped him a punch” (COCA corpus)

“A 15-year-old googled his way to revolutionizing cancer detection…” (http://pixelsandpills.com)

“Mike couldn’t believe she had managed to flirt his wallet open once again…” (COCA corpus)

How do speakers learn what not to say?

Conservatism via Entrenchment:
People only repeat what they’ve heard

Statistical Preemption:
People learn to avoid certain novel formulations by systematically witnessing a competing alternative

Ambridge et al. (2014)
Boyd & Goldberg (2011)
Boyd, Ackerman & Kutas (2013)
Brooks & Tomadillo (1999)
J. Claus (2014)
Goldberg (1995; 2006; 2011)

Verb Frequency Effect

?The magician vanished the rabbit.
> ??The magician disappeared the rabbit.

Brooks & Tomadillo (1999)
Theoktistos (2004)
Ambridge et al. (2008)
Robenalt & Goldberg (to appear, Cognitive Linguistics; to appear Language Learning)

Statistical Preemption

People can be creative.
They can extend verbs to novel uses as long as there isn’t a readily available alternative

Disappear NP

sounds bad because every time it would have been appropriate, made NP disappear occurred in its place.

Make NP vanish has occurred less often.

Causative construction NP V NP periphrastic causative construction [make NP VP]

Many verbs can appear in either, with corresponding difference in function
Each construction has its own function (e.g., Bolinger 1971; Clark 1987; Langacker 1987; Kemmer & Verhagen 1994; Goldberg 1995).

Is that a problem for statistical preemption (Bowman 1988; Pinker 1989)?

Is an advantage:
There will be contexts in which CxA should be preferred over CxB for verb.
If CxB is consistently witnessed instead… can learn that CxA is not possible for verb.

(Goldberg 1995)
Are higher frequency verbs always less flexible?

Conservatism via entrenchment: yes

Statistical preemption: only if another formulation is used in the same general context.

Study: Generate novel sentences with high and low frequency verbs (novelty confirmed via COCA)

Robenalt & Goldberg 2015: Cognitive Linguistics; 2016: Language Learning

Binned sentences via norming study

The editorial embarrassed the poor man out of town. Alternative (No-CA)
The editorial mortified the poor man out of town.

The chef coated ranch dressing over the salad. Has-competing Alternative (Has-CA)

The chef doused ranch dressing over the salad.

Acceptability Ratings

N=108 from MT

<table>
<thead>
<tr>
<th>Rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

The chef coated ranch dressing over the salad. Has-competing Alternative (Has-CA)

The magician vanished the rabbit.

Ashley was terribly mortified. 

Two competing constructions

Provides evidence that instances are not in second category
Higher type frequency provides *better* evidence that instances are not in second category.

If usage is not in competition, frequency doesn’t provide evidence of a restriction.

At the same time, it *is* also true that familiar formulations are preferred to novel formulations.

Quantifying evidence for statistical preemption

**Probability of CxB statistically preempting CxA for verb:**

\[ P(CxB) \text{ a discourse context in which the learner might expect to hear CxA[verb]} \]

**Confidence of statistical preemption for verb**

In frequency (CxB & [CxA would be at least as appropriate]}

**Interim summary**

People prefer to use familiar formulations.

But we are willing to use verbs creatively.

Frequency of verb in competing constructions is relevant to unacceptability.

A *pronominal recipient and lexical theme* favors the ditransitive construction over the dative.

*She gave me the ball.*  
*She gave the ball to me.*
"Alternating" verbs: Probability of a DATIVE given a context appropriate for a DITRANSITIVE.

| Alternating verbs: | Ditransitive: | P (dative | discourse context in which the learner might expect to hear ditransitive) |
|-------------------|---------------|---------------------------------------------------------------|
| Tell               | 36            | 3713                                                          |
| Give              | 111           | .01                                                           |
| Show              | 35            | 932                                                           |
| Send              | 146           | 1098                                                          |
| Sell              | 40            | 152                                                           |
| Bring             | 111           | 415                                                           |
| Read              | 81            | 275                                                           |
| Send              | 7             | 176                                                           |
| Total             | 507           | 1743                                                          |

P(CxB | a discourse context in which the learner might expect to hear CxA[verb])

"NON-alternating" verbs: Probability of a DATIVE given context appropriate for a DITRANSITIVE.

| Non-alternating verbs: | Ditransitive: | P (dative | discourse context in which one might expect to hear ditransitive) |
|------------------------|---------------|---------------------------------------------------------------|
| Explain                | 120           | 1.99                                                          |
| Whisper                | 16            | .94                                                           |
| Transfer               | 20            | 1.0                                                           |
| Return                 | 74            | .88                                                           |
| Entreat                | 13            | 1.0                                                           |
| Deliver                | 33            | .65                                                           |
| Present                | 43            | 37                                                           |
| Repeat                 | 26            | 0.1                                                           |
| Total                  | 345           | 69                                                           |

"CONFIDENCE" of statistical preemption for verbs:

Strength of statistical preemption: a function of PROBABILITY and CONFIDENCE of the preemtping construction.

Relevant data is available to learners/analysts in large corpora.

Another subtle restriction: "a-adjectives":

- The/asleep child
- The/afloat ship
- The/alive monster
- The/aghast audience
- The/ablaze building
- The/afraid child

Semantic near-synonyms:
- a. the/a sleeping child
- b. the/a scared man
- c. the/a living monster
- d. the/a burning building

Phonologically related "NON-a-adj:
- a. the/an adult male
- b. the/an acute sickness
- c. the/an aspice comment
- d. the/an aloeof woman
Explanation for the distribution:

Historical “persistence”:

- asleep < in sleep
- abloom < in bloom
- adrift < on drift
- afloat < on float
- ablaze < on blaze

As PPs, “the on drugs man

afraid <-> “on fraid”

Instead afraid < p. ppl of affray (v): “to startle”

Speakers have assimilated afraid to category of a-adjecives.

Synchronically: Requires usage-based model: speakers are aware of which adjectives they’ve heard in which constructions.

Constructionist (usage-based) view

- Learners record statistics about particular items’ distribution.
- Constructional generalizations emerge from learners categorizing over the input.

A-adjective experiment

- Undergraduate native speakers of English (n = 32)
- Production task

Will novel a-adjecives get assimilated to the category? (exp. #1)

Do speakers make use of statistical preemption? (exp. #2) Wisely? (exp. #3)

(Boyd and Goldberg, 2011, Language)
An example trial

Here are two foxes.

Which fox moved to the star?

(a) star

vigilant
sleepy
Here are two cows.

(d) star

chammy
zoopy
Here are two chipmunks.

(b) star

vigilant
asleep
Here are two cows.

(c) star

adax
zedgy
Here are two lizards.
24 counterbalanced fillers:
Prototypical adjectives: likely to elicit prenominal use: the bad dog
bad, good, smart, dumb, fast, slow, old, young, rich, poor, strong, weak

Present tense verbs: likely to elicit predicative use: the dog that bites
bites, camps, cooks, cries, gambles, made, ran, smokes, snowboards, travels, votes, we
• Exposure Block
  – 6 "practice trials": 3 attributive and 3 predicative examples using non-experimental words.
• Test Block
  – 16 critical trials interleaved with 16 filler trials

Results

Speakers do in fact avoid using real a-adj (afraid, asleep) prenominally

Results

Speakers also avoid using nonsense a-adj (ablim, afek) prenominally

Restriction on real adjectives is stronger than restriction on nonsense adjectives.

Study #2: introduce preemptive context for novel a-adj

An example trial

Here are two foxes.
“The fox that’s afek moved to the star”

Recall: experiment #1 results

Results of Exp. #2: with preemptive context (n=20)

No difference between real a-adjacents (14%) and novel (24%), t(19) = 1.25, p = .23.
Exp. 3: pseudo-preemptive context (conservatism?)

Are people savvy enough to know when a context is truly preemptive?

Exposure to pseudo-preemptive context:

*The fox that's adax and proud of itself…

Notice that prenominal attributive construction is unavailable:

*The proud of itself fox…
*The afek and proud of itself fox….

Interim summary

- Learners record the statistics of their language.
- Learners categorize their input into patterns.
- **Statistical preemption**: learners learn to avoid certain formulations because an alternative formulation is systematically used in the appropriate context.
- Learners are smart about what counts as a preemptive context.
Recent challenge to statistical preemption account of *Run afraid boy

Yang 2015 (*Language*), Bruening 2011:

- Children do not witness enough evidence for statistical preemption before they are 3.

- Learners don’t need indirect negative evidence because a-adjectives pattern like particles, e.g., off, on, out.

- The suggested evidence: a-adjectives are claimed to occur with adverbs right, straight, well, far, which also apply to particles.

BUT, many a-adjectives *never* occur with any of the key adverbs in all of COCA corpus:

*far asleep, *straight asleep, *well asleep
*far awake, *straight awake, *well awake

(REGULAR adjectives occur much more commonly:

e.g.,

far greater (1300)
well pleased (70) (Yang classifies across, away, around, ahead as a-adjectives)

straight winning (14)

Finally, children do not learn the restriction until they are 10 years old:

Summary

Yang 2015 (*Language*), Bruening 2011:

- Children do not witness enough evidence for statistical preemption before they are 3.

- Learners don’t need indirect negative evidence because a-adjectives pattern like particles, e.g., off, on, out.

- The suggested evidence: a-adjectives are claimed to occur with adverbs right, straight, well, far, which also apply to particles. Actually these adverbs almost never occur with a-adjectives, while they do occur with regular adjectives.

Evidence that a-adjectives are adjectives

- Semantics
- They appear after bare nouns (Lakoff 1970; Jackendoff 1972)
  - Pat seemed afraid/alone/alive.
- They are conjoinable with other adjectives
  - Chris was afraid but happy.
  - She was alive and healthy.

(see corpus analysis in Goldberg & Boyd, to appear, Language)

Statistical preemption is consistent with available evidence, while the suggested positive evidence is not (par Yang 2015; Breuning 2011)

Error-driven learning

We anticipate (predict) what others will say as they speak.


Competition-dependent learning

Partial activation of competing form leads to learned dissociation (Anderson & Spellman 1995; Anderson; Green & McGloth 2000; Dete 2010; Newman and Norman 2010).

Anderson & Spellman 1996
Subjects learned paired associations, e.g.,
- Fruit-Apple
- Fruit-Pear
- Fruit-Kiwi
- Furniture-Table
- Furniture-Phone

Then cued with a subset of these pairs such as:
Fruit-Pe

Note “Pear” is only partially cued, therefore subjects partially activate other prototypical associates, e.g., Apple.
- memory for Fruit-Apple was weakened
- memory for Fruit Kiwi was not weakened.
- memory for other uncued pairs, e.g., Furniture-Table was not weakened either.

If speakers anticipate (1), it is partially activated. If the competing form (2) is witnessed instead, subsequently, (1) is harder to retrieve (forgotten/suppressed).

1. ?? She explained him something.
2. She explained something to him.

(2) preempts (1).
### Relationship to preemption

A construction that is in competition is weakened whenever another form “wins” (is used).

If a competing construction is *not* partially activated, there is no suppression.

---

### When do children generalize?

As soon as the tokens are recognized to form a pattern. Different ages for different types of patterns.

- Elena Lieven *et al.* (1997)
- Mike Tomasello (2000; 2003)

Much to do here: which dimensions of similarity are relevant to which constructions and how do children determine this? (Perek & Goldberg, to appear, JML)

---

### Why do children overgeneralize?

They do not have the alternative readily available (in general, or at the moment of speaking).

- Huttenlocher 1979; Benedict 1979; Graudenzi 1977, Gershkoff-Stowe 2001

---

### How do children eventually recover from overgeneralizations?

The more conventional alternative becomes more readily available (through more exposure & greater fluency).

---

### Why might L2 learners find it harder to learn arbitrary restrictions?

- “could you recommend me some [place] to apply?”
- “maybe it’s better to explain me first”
- “have you ever considered to go climbing in Ecuador? We would love for you to come and discover it!”
- “the afraid boy hiding behind the board”

---

### “The RAGE hypothesis”

(Grüter & Rohde; Les-Williams & Fernald 2010)

- Non-native speakers have reduced ability to generate expectations during language processing,
- particularly based on grammatically encoded distinctions.
If non-native speakers don’t anticipate upcoming utterances to the same extent as native speakers, they will have less opportunity for competition-dependent learning.

L2 processing is a cost on WM

Individual differences, degree of proficiency, and task can play a role in the extent to which processing is affected.

The dictator flooded information into the people.
Nancy shot a look to him.
The workers sealed concrete into the entrance.
The editor grinned a reporter into his office.
The teacher frowned a warning to the students.
The grocer hollered the man out of the store.

I2 learners tended to treat novel sentences as novel, without taking into account whether a readily available competing alternative existed or not.

Not clear that I2 learners make use of statistical preemption to the same extent as L1.
Conclusion: short version

It’s categorization.

Conclusion: more specifically

- Learners record statistics of their language.
- We actively categorize the input, recognizing patterns of form and function.
- Productivity is generally determined by general principles of induction (coverage & similarity)
- Learners learn to avoid certain formulations because an alternative formulation is systematically used instead: statistical preemption (i.e., competition driven learning)
- Adults are smart about what counts as a preemptive context.

Statistical preemption is a slow, gradual process; some arbitrary restrictions are learned very late.

L2 learners may take less able to take full advantage of statistical preemption.

Thanks to:

Jeremy Boyd
Laura Suttle
Clarice Robemak
Andrew Conway
Sam Glucksberg
Matt Johnson
Phil Johnson
Phil Johnson-Laid
Harvey Koslin

Labex, Pollet Samuelein!