Psych 215L:
Language Acquisition

Lecture 7

Grammatical Categories I

Grammatical Categorization

Computational Problem: Identify grammatical categories

These will tell you how words are used in the language.

"This is a DAX."

"He is sibbing."

DAX = noun

SIB = verb

Categorization: How?

How might children initially learn what categories words belong to?

Deriving Categories from Semantic Information
Semantic Bootstrapping Hypothesis (Pinker 1984)

Children can initially determine a word's category by observing what kind of entity in the world it refers to.

objects, substance = noun (*goblins, glitter*)

action = verb (*steal, sing*)

Word's semantic category (meaning) is then linked to innate grammatical category knowledge (noun, verb)

Semantic Bootstrapping

Pinker (1984) premise: who and what are being talked about are meaning aspects of an utterance that are transparent to learners even before they have acquired much knowledge about the vocabulary and structure of their language

What is the innate knowledge children have?

"Innate linking rules" between "action-words" and "Verb", for example.

Categorization: How?

How might children initially learn what categories words belong to?

Deriving Categories from Semantic Information
Semantic Bootstrapping Hypothesis (Pinker 1984)

Children can initially determine a word's category by observing what kind of entity in the world it refers to.

Slight problem: hard to identify the referent in the world for words sometimes (like verbs)

"Look! He's *frepping*!"

frep = climb, perch, glower, grab, yell, …?

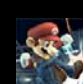
Categorization: How?

How might children initially learn what categories words belong to?

Deriving Categories from Semantic Information
Semantic Bootstrapping Hypothesis (Pinker 1984)

Children can initially determine a word's category by observing what kind of entity in the world it refers to.

Another problem: mapping rules are not perfect
Ex: not all action-like words are verbs



"active", "action"
action-like meaning, but they're not verbs

Categorization: How?

"A very different view assumes that distributional relationships among form-based cues are central to category-based abstraction….Examples of such cues are relative locations of words in strings, phonological regularities within words of a class and co-occurrence relations between classes….functor categories tend to have shorter vowel durations, weaker amplitudes and simplified syllable structure compared to lexical categories such as noun and verb…" - Gómez & Lakusta

Categorization: How?

Distributional Learning

Children can initially determine a word's category by observing the linguistic environments in which words appear:

relative location of words in an utterance: "He likes to SIB."

phonological regularities within classes of words: *the, a, an*

= short (monosyllabic) words, simple syllables

co-occurrence relations between grammatical categories:
Determiner Noun (the goblin)
= Determiners (a, the, an, …) precede Nouns (goblin)

Distributional Learning Evidence

Distributional Learning (Evidence)

Children are sensitive to the distributional properties of their native language when they're born (Shi, Werker, & Morgan 1999).

7-month-olds can recognize and track specific *functor* words (*a, the, to, will*…) in fluent speech (Höhle & Weissenborn 2003)

15- to 16-month German infants can determine novel words are nouns, based on the distributional information around the novel words (Höhle et al. 2004)

18-month English infants can track distributional information like *"is…-ing"* to signal that a word is a verb (Santelmann & Jusczyk 1998)

Categorization: How?

Idea (Gómez & Lakusta 2004)

"If infants are able to identify categories in the speech stream by means of their phonological properties, they might then use this information to learn the predictive relationships between categories."

(1) Sound properties of certain words can be tracked distributionally (ex: monosyllabic, simple syllables = noticeable to infants).

(2) Infants can group words together into categories based on these properties.

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	[data missing]
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	[data missing]

data missing

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

"the" goes with these words

"the" behavior = precedes "king", "girl", "baby", etc.

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

"a" goes with almost all the same words

Inference: "a" has almost the same distribution as "the", so "a" is the same category as "the"

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

Prediction:
"a" acts like "the",
"a" goes with "dwarf"

Conclusion:
"a dwarf" is in language

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

"will" goes with these words

"will" behavior = precedes "sing", "laugh", "steal", etc.

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

"can" goes with almost all the same words

Inference: "can" has almost the same distribution as "will", so "can" is the same category as "will"

About Categorization

Data Observed

	X1	X2	X3	X4	X5
A1 = the	king	girl	baby	goblin	dwarf
A2 = a	king	girl	baby	goblin	
	Y1	Y2	Y3	Y4	Y5
B1 = will	sing	laugh	steal	run	sneeze
B2 = can	sing	laugh	steal	run	

Prediction:
"can" acts like "will" so "can" goes with "sneeze"

Conclusion: "can sneeze" is in language

Previous studies with aX, bY paradigm

"Interestingly, although learners readily acquire the legal positions of words with respect to which occur first versus which occur second….categories and their relationships (i.e. that words belong to particular a, b, X, and Y classes, and that a-words go with Xs and not Ys) are virtually impossible to acquire unless some subset of the X- and Y-category members are marked with salient conceptual or perceptual cues." - Gómez & Lakusta

Something besides statistical learning abiltiies is needed?

What the child has to do

"…there are two essential steps in an aX bY category abstraction. Learners must first associate a- and b-elements with cues differentiating X and Y categories. They can then categorize a- and b-elements based on their co-occurrence…In the second step, learners group (or categorize) a- and b-elements by merit of their joint association with particular distinguishing cues. Once a- and b-categories are formed, learners can rely on memory for a pair they have heard…to make inferences about a pair they have not heard…" - Gómez & Lakusta (2004)

What the child has to do

"By this view, Step 1 learning is evidenced by the ability to discriminate correct from incorrect pairings of functional and lexical test items with distinguishing cues present. Step 2 learning is evidenced by discrimination of test items in the absence of distinguishing cues." - Gómez & Lakusta (2004)

17-month-olds can do both steps, and…

"We know that by 7 and 12 months of age, infants are able to abstract patterns from artificial grammars as evidenced by their ability to discriminate grammatical from ungrammatical strings in new vocabulary…also know from Gerken et al. (2003) that 12-month-olds do not show Step 2 learning….[but] might be able to engage in a more preliminary form of category-based abstraction."

Gómez & Lakusta 2004: Categorization Experiment

Testing 12-month-olds, using artificial language paradigm (so children couldn't have any experience with the categories beforehand)

On the validity of artificial language experiment designs (how much are they really like language for the children tested): Lany et al. (2007) show this knowledge persists for at least 24 hours and enables learning of related artificial language constructions

General procedure:
Infants exposed to one of two training languages (L1 or L2).
Used same set of vocabulary (all novel words).

L1 generalization: a goes with X, b goes with Y (aX, bY language)
L2 generalization: a goes with Y, b goes with X (aY, bX language)

Gómez & Lakusta 2004: Categorization Experiment

L1

	X1	X2	X3	X4	X5	X6
A1 = alt	coomo	fengle	kicey	loga	paylig	wazil
A2 = ush	coomo	fengle	kicey	loga	paylig	wazil
	Y1	Y2	Y3	Y4	Y5	Y6
B1 = ong	deech	ghope	jic	skige	vabe	tam
B2 = erd	deech	ghope	jic	skige	vabe	tam

Gómez & Lakusta 2004: Categorization Experiment

L1

	X1	X2	X3	X4	X5	X6	
A1 = alt	coomo	fengle	kicey	loga	paylig	wazil	Disyllabic words
A2 = ush	coomo	fengle	kicey	loga	paylig	wazil	
	Y1	Y2	Y3	Y4	Y5	Y6	
B1 = ong	deech	ghope	jic	skige	vabe	tam	Monosyllabic words
B2 = erd	deech	ghope	jic	skige	vabe	tam	

Gómez & Lakusta 2004: Categorization Experiment

L1

	X1	X2	X3	X4	X5	X6	
A1 = alt	coomo	fengle	kicey	loga	paylig	wazil	Disyllabic words
A2 = ush	coomo	fengle	kicey	loga	paylig	wazil	
	Y1	Y2	Y3	Y4	Y5	Y6	
B1 = ong	deech	ghope	jic	skige	vabe	tam	Monosyllabic words
B2 = erd	deech	ghope	jic	skige	vabe	tam	

Association: alt/ush (a1,a2) go with these words (X1-X6)
Abstraction: alt/ush (a1,a2) go with disyllabic words
Categorization: alt/ush are a category whose behavior is to go with disyllabic words

Gómez & Lakusta 2004: Categorization Experiment

L1

	X1	X2	X3	X4	X5	X6	
A1 = alt	coomo	fengle	kicey	loga	paylig	wazil	Disyllabic words
A2 = ush	coomo	fengle	kicey	loga	paylig	wazil	
	Y1	Y2	Y3	Y4	Y5	Y6	
B1 = ong	deech	ghope	jic	skige	vabe	tam	Monosyllabic words
B2 = erd	deech	ghope	jic	skige	vabe	tam	

Association: ong/erd (b1,b2) go with these words (Y1-Y6)
Abstraction: ong/erd (b1,b2) go with monosyllabic words
Categorization: ong/erd are a category whose behavior is to go with monosyllabic words

Gómez & Lakusta 2004: Categorization Experiment

L2

	X1	X2	X3	X4	X5	X6	
A1 = alt	deech	ghope	jic	skige	vabe	tam	Monosyllabic words
A2 = ush	deech	ghope	jic	skige	vabe	tam	
	Y1	Y2	Y3	Y4	Y5	Y6	
B1 = ong	coomo	fengle	kicey	loga	paylig	wazil	Disyllabic words
B2 = erd	coomo	fengle	kicey	loga	paylig	wazil	

Gómez & Lakusta 2004: Categorization Experiment

General procedure:
Infants exposed to one of two training languages (L1 or L2).
Used same set of vocabulary (all novel words).

L1 generalization: a goes with X, b goes with Y (aX, bY language)
L2 generalization: a goes with Y, b goes with X (aY, bX language)

Test phase:
Infants exposed to *new* phrases from their training language
L1 children: new aX, bY examples
L2 children: new aY, bX examples

Gómez & Lakusta 2004: Categorization Experiment

L1 test

	X1	X2	X3	X4	X5	X6	
A1 = alt	beevit	meeper	gackle	roosa	nawlup	binnow	Disyllabic words
A2 = ush	beevit	meeper	gackle	roosa	nawlup	binnow	
	Y1	Y2	Y3	Y4	Y5	Y6	
B1 = ong	vot	pel	tood	rud	biff	foge	Monosyllabic words
B2 = erd	vot	pel	tood	rud	biff	foge	

The point: Children needed to complete association, abstraction, and categorization in order to realize that these new instances of aX and bY were part of the artificial language L1.

Gómez & Lakusta 2004: Categorization Experiment

L1 process

	X1	X2	… X6
A1 = alt	coomo	fengle	…. wazil
A2 = ush	coomo	fengle	…. wazil
	Y1	Y2	…Y6
B1 = ong	deech	ghope	…tam
B2 = erd	deech	ghope	…tam

Gómez & Lakusta 2004: Categorization Experiment

L1 process

	X1	X2	… X6	
A1 = alt	coomo	fengle	…. wazil	Association
A2 = ush	coomo	fengle	…. wazil	Association
	Y1	Y2	…Y6	
B1 = ong	deech	ghope	…tam	Association
B2 = erd	deech	ghope	…tam	Association

Gómez & Lakusta 2004: Categorization Experiment

L1 process

	X1	X2	… X6	
A1 = alt	coomo	fengle	…. wazil	Abstraction: disyllabic words
A2 = ush	coomo	fengle	…. wazil	Abstraction: disyllabic words
	Y1	Y2	…Y6	
B1 = ong	deech	ghope	…tam	Abstraction: monosyllabic words
B2 = erd	deech	ghope	…tam	Abstraction: monosyllabic words

Gómez & Lakusta 2004: Categorization Experiment

L1 process

	X1	X2	… X6	
A1 = alt	coomo	fengle	…. wazil	Categorization based on similar distribution: disyllabic words
A2 = ush	coomo	fengle	…. wazil	
	Y1	Y2	…Y6	
B1 = ong	deech	ghope	…tam	Categorization based on similar distribution: monosyllabic words
B2 = erd	deech	ghope	…tam	

Gómez & Lakusta 2004: Categorization Experiment

L1 process

	X1	X2	… X6	
A1 = alt	coomo	fengle	…. wazil	Extension to new examples: alt beevit
A2 = ush	coomo	fengle	…. wazil	
	Y1	Y2	…Y6	
B1 = ong	deech	ghope	…tam	Extension to new examples: ong pel
B2 = erd	deech	ghope	…tam	

Test Items

Table 4 *Test materials*

Test sets			
L1-test1	L1-test2	L2-test1	L2-test2
[alt nawlup] [ong pel]	[alt roosa] [ong tood]	[alt biff] [ong meeper]	[alt rud] [ong gackle]
[alt gackle] [ong foge]	[alt meeper] [ong biff]	[alt vot] [erd gackle]	[alt pel] [ong nawlup]
[alt bevit] [erd tood]	[alt binow] [erd rud]	[alt tood] [ong binnow]	[alt foge] [erd roosa]
[ush binnow] [erd vot]	[ush gackle] [erd foge]	[ush rud] [ong roosa]	[ush vot] [erd meeper]
[ush roosa] [ong rud]	[ush nawlup] [ong vot]	[ush pel] [erd nawlup]	[ush tood] [erd binnow]
[ush meeper] [erd biff]	[ush bevit] [erd pel]	[ush foge] [erd bevit]	[ush biff] [ong bevit]
[ong vot] [alt roosa]	[ong rud] [ush roosa]	[ong gackle] [ush vot]	[ong roosa] [ush rud]
[ong tood] [ush bevit]	[ong pel] [alt bevit]	[ong bevit] [alt rud]	[ong meeper] [alt vot]
[ong biff] [ush gackle]	[ong foge] [ush meeper]	[ong nawlup] [ush tood]	[ong binnow] [ush pel]
[erd foge] [alt meeper]	[erd biff] [ush binnow]	[erd binnow] [alt pel]	[erd nawlup] [ush foge]
[erd pel] [alt binow]	[erd tood] [alt gackle]	[erd meeper] [alt foge]	[erd gackle] [alt tood]
[erd rud] [ush nawlup]	[erd vot] [alt nawlup]	[erd roosa] [ush biff]	[erd bevit] [alt biff]

Note: Test strings were presented in sets (e.g. L1-test1). Notice, X- and Y-categories consist of new one- and two-syllable words. The strings in the test sets were presented in a random order.

Gómez & Lakusta 2004: Categorization Experiment

Expt 1 Results:
"A Wilcoxon Signed Ranks Test showed that infants listened significantly longer to strings from their training language than to strings from the other language…Eighteen out of 24 infants showed this pattern…" - G & L (2004)

This suggests that 12-month-olds were able to complete association, abstraction, and categorization for this artificial language - based only on the distributional information available.

Specifically, the distributional information was the occurrence of one item next to another one in the training phase (L1: aX, bY).

Real Categorization…?

"The ability to discriminate legal from illegal marker-feature pairings…reflects sensitivity to the co-occurrence relations between markers and X- and Y-categories based on their distinguishing features…The fact that infants were able to generalize the novel X- and Y-elements suggests that learning was to some degree abstract (involving grouping of the X- and Y-elements according to syllable number)."
-G&L

"Does such grouping count as *categorization*? We would argue 'yes' to the extent that categorization involves distinguishing elements according to some features…" -G&L

Experiment 2: Real Life Ain't Pretty

"…whether young learners are able to separate more probable from less probable structure by exposing them to artificial languages with varying degrees of probabilistic structure." - G&L

Table 5 *Median differences and 95.1% confidence intervals for listening times (in seconds) to strings from the predominant versus the non-predominant training language. Infants showed significant discrimination even when 17% of the strings encountered during training were from the non-predominant training language*

Probability ratio	Median difference	95.1% confidence interval
Exp. 1: 100/0	1.618*	(0.665–2.605)
Exp. 2: 83/17	1.248*	(0.130–2.850)
Exp. 2: 67/33	−0.125	(−0.925–0.835)

Note: * Listening time differences in these conditions were statistically significant, $p \leq .05$.

G&L on the applicability of their results to real life

Expt 1
"…infants in this study were not simply learning associations…they were generalizing based on abstract feature of syllable number, demonstrating they are capable of categorizing at a level at least one step removed from physical identity…Such generalization is an important precursor…by 17 months old, [they] can form a- and b-categories comprised of elements with no common features other than their co-occurrence patterns…"

G&L on what Expt 2 means

Expt 2
"…important for determining whether infant learners are equipped to tolerate some degree of inconsistency in their linguistic input…were indeed able to focus on the predominant patterns in their training data…appear to be limits on such learning, however….in Condition 67/33…"

G&L on explaining Expt 2 results

"Were infants in Condition 83/17 learning two forms of structure simultaneously or only the more predominant abstract structure?"

"Because infants were tested on their ability to generalize to new marker-word phrases…we are unable to distinguish these explanations in the present studies."

G&L on explaining Expt 2 results

"What about learning in Condition 67/33? Infants…were clearly not generalizing the marker-word pairing. Nor were they engaged in learning two forms of structure simultaneously…or they would have shown discrimination on the test…"

Possibility 1: "disrupted learning entirely" (nothing to generalize)

Possibility 2: "…infants learned *only* specific marker-word phrases from the non-predominant language…" (why not from the predominant language?)

Possibility 3: Infants learned associations probabilistically (67/33) and forced choice test won't distinguish that from chance

"…we are unable to distinguish these possibilities with the present data because we did not test infants on marker-word phrases from training."

G&L on explaining Expt 2 results

Favoring disrupted learning…

"Infants show some selectivity in terms of their tendency to focus on different types of structure. Given two sources of statistical information, infants will favor the source of greater statistical regularity….it is reasonable to hypothesize that learners will only focus on a particular source of information to the extent that it yields some degree of statistical regularity."

G&L on the noise threshold

"…the question of whether learning degrades gradually or catastrophically with increases in noise. The present findings suggest that learning degrades gradually in that there were no significant decreases in learning from the 100/0 to the 83/17 conditions, and then a marginal decrease between Conditions 83/17 and 67/33."

Marginal decrease from 83/17 to 67/33?

Probability ratio	Median difference	95.1% confidence interval
Exp. 1: 100/0	1.618*	(0.665–2.605)
Exp. 2: 83/17	1.248*	(0.130–2.850)
Exp. 2: 67/33	−0.125	(−0.925–0.835)

Another study on inconsistent input

Hudson Kam & Newport (2005):

Artificial language study with variable input (45%, 60%, 75%, 100% of one type)

Children behave differently from adults

- children tended to show categorical behavior with 60% of one type (pick one option or the other most of the time, even if one appeared 60% of the time)
- adults tended to probability match (pick one option 60% of the time if it appeared 60% of the time)