Psych 215L: Language Acquisition

Lecture 8

Grammatical Categories II

Mintz 2003

Just to remind us of the problem…

"…it is not fully known how child language learners initially categorize words. There has been recent interest in the idea that distributional information carried by the cooccurrence patterns of words in sentences could provide a great deal of information relevant to grammatical categories."

Mintz 2003, on Theorists

And what theorists initially thought…

"Pinker (1987) argued that, given sentences in (2a,b), a distributional learner would incorrectly categorize *fish* and *rabbits* together, and, hearing (2c), would incorrectly assume that (2d) is also permissible."

(2a) John ate fish. (2b) John ate rabbits.
(2c) John can fish. (2d) *John can rabbits.

"The crux of the problem…is that a given word form…can belong to multiple categories and thus occur in different syntactic contexts…potentially providing misleading category information…argued that the resulting erroneous generalizations would be common, and would render a distributional approach to categorization untenable."

Mintz 2003, Another Problem

"The fundamental issue is that lexical adjacency patterns are variable…another question is how the learner is to know *which* environments are important and which should be ignored. Distributional analyses that consider all the possible relations among words in a corpus of sentences would be computationally unmanageable at best, and impossible at worst."

One idea: local contexts

"…by showing that local contexts are informative, these findings suggested a solution to the problem of there being too many possible environments to keep track of: focusing on local contexts might be sufficient."

Experimental Evidence

Idea: Children may be attending to other kinds of distributional information available in the linguistic environment

There is evidence that children can track information that is non-adjacent in the speech stream (Santelmann & Jusczyk 1998, Gómez 2002)

he *is* run*ning*

Also, frequency of lexical frames is something children are sensitive to (Childers & Tomasello 2001: children more easily acquire novel verb meanings when the verbs occur in lexical frames that occur frequently in the input)

Frequent Frames

Idea: What categorization information is available if children track frequent frames?

Frequent frame: X___Y
where X and Y are words that frame another word and appear frequently in the child's linguistic environment

Examples:

the__is	can___him
the king is…	can trick him…
the goblin is…	can help him…
the girl is…	can hug him…

Frequent Frames vs. Bigrams

"In the present approach the word 'W' in the environment '…X W Y…' is stored as 'jointly following X and preceding Y', but such would not be the case if W occurred after X and before Y on independent occasions…bigram contexts…record only independent cooccurrence patterns (e.g. 'following X', 'preceding Y')….property of joint co-occurrence in the frame contexts involves an additional relationship..."

Experimental Support

"Another important difference…adults will categorize words in an artificial language based on their occurrence within frames…whereas bigram regularity alone has failed to produce categorization in artificial grammar experiments, without additional cues…"

Goals

"The goal of the work described here…what assumptions would be reasonable to build into [a model of grammatical categorization by learners]. Specifically, the goal was to formulate a unit to which there is some evidence that children and adults attend, and with which adults have been shown to categorize, and examine how predictive it is of category membership."

Data

CHILDES Child Language Data Exchange System

Data representing child's linguistic environment:
6 corpora of child-directed speech from the CHILDES database

Table 1
Experiment 1 session ranges for analyzed corpora, number of utterances, number of tokens and types categorized, percentage of corpus (tokens) accounted for by categorized types, and percentage of corpus (tokens) analyzed

Child	CHILDES sessions	# of utterances
Peter	peter01-peter12	19846
Eve	eve01-eve20	14922
Nina	nina01-nina23	14417
Naomi	n01-n58	6950
Anne	anne01a-anne23b	26199
Aran	aran01a-aran20b	20857
Mean		

What is a "frequent" frame?

Definition of "frequent" for frequent frames:
Frames appearing a certain number of times in a give corpus

"The principles guiding inclusion in the set of frequent frames were that frames should occur frequently enough to be noticeable, and that they should also occur enough to include a variety of intervening words to be categorized together. While these criteria were not operationalized in the present experiment, a pilot analysis with a randomly chosen corpus, Peter, determined that the 45 most frequent frames satisfied these goals and provided good categorization."

How Frequent Frames Work

Trying out frequent frames on a corpus of child-directed speech.

Frame: the ___ is
"the radio is in the way…but the doll is…and the teddy is…"

radio, doll, teddy = Category1 (similar to Noun)

Frame: you ___ it
"you draw it so that he can see it… you dropped it on purpose!…so he hit you with it…"

draw, dropped, with = Category 2 (similar-ish to Verb)

Metrics for Success

Determining success with frequent frames:

(Accuracy)

Precision = <u># of words identified correctly as Category within frame</u>
# of words identified as Category within frame

(Completeness)

Recall = <u># of words identified correctly as Category within frame</u>
# of words that should have been identified as Category

Metrics for Success

Determining success with frequent frames:

Precision = <u># of words identified correctly as Category within frame</u>
# of words identified as Category within frame

Recall = <u># of words identified correctly as Category within frame</u>
# of words that should have been identified as Category

Frame: you ___ it
draw, dropped, with = Category 2 (similar-ish to Verb)

# of words correctly identified as Verb = 2
# of words identified as Verb = 3
Precision = 2/3

Metrics for Success

Determining success with frequent frames:

Precision = <u># of words identified correctly as Category within frame</u>
# of words identified as Category within frame

Recall = <u># of words identified correctly as Category within frame</u>
# of words that should have been identified as Category

Frame: you ___ it
draw, dropped, with = Category 2 (similar-ish to Verb)

# of words correctly identified as Verb = 2
# of words should be identified as Verb = many (all verbs in corpus)
Recall = 2/many = small number

Some Frequent Frame Results

Table 2
Samples of representative categories from several corpora. The number of tokens categorized for each type is in parentheses

Peter
<u>Frame</u> you__it
put (52), see (28), do (27), did (25), want (23), fix (13), turned (12), get (12), got (11), turn (10), throw (10), closed (10), think (9), leave (9), take (8), open (8), find (8), bring (8), took (7), like (6), knocked (6), putting (5), pull (5), found (5), make (4), have (4), fixed (4), finish (4), try (3), swallow (3), opened (3), need (3), move (3), hold (3), give (3), fixing (3), drive (3), close (3), catch (3), threw (2), taking (2), screw (2), say (2), ride (2), pushing (2), hit (2), hiding (2), had (2), eat (2), carry (2), build (2), brought (2), write (1), wiping (1), wipe (1), wind (1), unzipped (1), underneath (1), turning (1), touching (1), tore (1), tie (1), tear (1), swallowed (1), squeeze (1), showing (1), show (1), said (1), rip (1), read (1), reach (1), pushed (1), push (1), play (1), pick (1), parking (1), made (1), love (1), left (1), knock (1), knew (1), hid (1), flush (1), finished (1), expected (1), dropped (1), drop (1), draw (1), covered (1), closing (1), call (1), broke (1), blow (1)

Some Frequent Frame Results

<u>Frame</u> you__it
put (28), want (15), do (10), see (7), take (6), turn (5), taking (5), said (5), sure (4), lost (4), like (4), leave (4), got (4), find (4), throw (3), threw (3), think (3), sing (3), reach (3), picked (3), get (3), dropped (3), seen (2), lose (2), know (2), knocked (2), hold (2), help (2), had (2), gave (2), found (2), fit (2), enjoy (2), eat (2), chose (2), catch (2), with (1), wind (1), wear (1), use (1), took (1), told (1), throwing (1), stick (1), share (1), sang (1), roll (1), ride (1), recognize (1), reading (1), ran (1), pulled (1), pull (1), press (1), pouring (1), pick (1), on (1), need (1), move (1), manage (1), make (1), load (1), liked (1), lift (1), licking (1), let (1), left (1), hit (1), hear (1), give (1), flapped (1), fix (1), finished (1), drop (1), driving (1), done (1), did (1), cut (1), crashed (1), change (1), calling (1), bring (1), break (1), because (1), banged (1)

<u>Frame</u> the__is
moon (6), sun (5), truck (3), smoke (2), kitty (2), fish (2), dog (2), baby (2), tray (1), radio (1), powder (1), paper (1), man (1), lock (1), lipstick (1), lamb (1), kangaroo (1), juice (1), ice (1), flower (1), elbow (1), egg (1), door (1), donkey (1), doggie (1), crumb (1), cord (1), clip (1), chicken (1), bug (1), brush (1), book (1), blanket (1), Mommy (1)

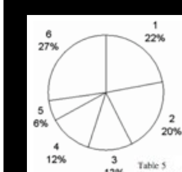
Another Look at Frequent Frame Coverage

Table 1
Experiment 1 session ranges for analyzed corpora, number of utterances, number of tokens and types categorized, percentage of corpus (tokens) accounted for by categorized types, and percentage of corpus (tokens) analyzed

Child	CHILDES sessions	# of utterances	Tokens categorized	Types categorized	Percentage of corpus accounted for	Percentage of corpus analyzed
Peter	peter01-peter12	19846	5690	446	48%	6%
Eve	eve01-eve20	14922	3513	400	46%	5%
Nina	nina01-nina23	14417	6265	469	51%	8%
Naomi	n01-n58	6950	1617	297	38%	5%
Anne	anne01a-anne23b	26199	4389	405	54%	4%
Aran	aran01a-aran20b	20857	5628	620	61%	5%
Mean			4517	439.5	50%	6%

"Frequent frames can thus focus a learner on a relatively small number of contexts that can have broad impact on how words in the input are categorized….be very useful to young language learners, who have limited memory and processing resources."

The Robustness of Frequent Frames



"…on average 45% of the frequent frames of a given corpus were frequent frames for at least three other corpora, indicating that many informative distributional contexts are shared from corpus to corpus."

Table 5
Frames that were frequent frames in at least two corpora, organized by number of corpora in which each occurred

6	5	4	3	2
do__want	a__of	I__think	I__know	I__it
put__on	put__in	I__you	are__doing	a__on
the__in	to__the	are__going	did__do	can__it
the__on		is__a	go__the	do__think
to__it		it__the	the__of	don't__it
want__to		to__a	the__one	have__got
what__you		would__like	there__is	have__look
you__a		you__that	to__to	here__are
you__it			what__it	in__box
you__me			why__you	put__back
you__the			you__with	shall__put
you__to				the__and

Precision results

Experiment 1 token and type accuracy for Standard and Expanded Labeling including baseline accuracy of random categories

Corpus	Token accuracy (Standard)		Token accuracy (Expanded)		Type accuracy (Standard)		Type accuracy (Expanded)	
	Analysis	Random	Analysis	Random	Analysis	Random	Analysis	Random
Peter	0.98	0.49	0.97	0.32	0.96	0.55	0.95	0.49
Eve	0.98	0.51	0.91	0.25	0.92	0.50	0.89	0.40
Nina	0.98	0.48	0.98	0.29	0.95	0.46	0.94	0.36
Naomi	0.97	0.48	0.96	0.30	0.94	0.49	0.93	0.41
Anne	0.98	0.37	0.84	0.24	0.94	0.41	0.90	0.31
Aran	0.97	0.44	0.80	0.23	0.89	0.42	0.87	0.33
Mean	0.98	0.46	0.91	0.27	0.93	0.47	0.91	0.38

Precision generally quite high.

Interpretation: When a frequent frame clustered words together into category, those words often did belong together. (Nouns together, verbs together, etc.)

Recall results

Experiment 1 token and type completeness for Standard and Expanded Labeling including baseline accuracy of random categories

Corpus	Token completeness (Standard)		Token completeness (Expanded)		Type completeness (Standard)		Type completeness (Expanded)	
	Analysis	Random	Analysis	Random	Analysis	Random	Analysis	Random
Peter	0.06	0.03	0.09	0.03	0.07	0.04	0.08	0.04
Eve	0.06	0.03	0.12	0.03	0.07	0.04	0.09	0.04
Nina	0.08	0.04	0.13	0.04	0.10	0.05	0.12	0.05
Naomi	0.07	0.03	0.11	0.04	0.07	0.03	0.08	0.04
Anne	0.08	0.03	0.11	0.03	0.09	0.04	0.12	0.04
Aran	0.08	0.04	0.13	0.04	0.09	0.04	0.10	0.04
Mean	0.07	0.03	0.12	0.03	0.08	0.04	0.10	0.04

Recall generally quite low.

"…there were often several noun categories and several verb categories (all very accurate), rather than one category of all the nouns, one of all the verbs, etc."

The magic number of frequency…

"It would be desirable to analyze the corpora using a frequency threshold for each corpus that is based on a relativized frequency criterion, as the salience of frequent frames to human learners is more likely to be a factor of relative frequency than absolute number."

Experiment 2

"The set of frequent frames was…selected to include all frames whose frequency in proportion to the total number of frames in the corpus surpassed a predetermined threshold of 0.13%…this specific threshold was determined based on the frequent frames for each corpus in Experiment 1….frequent frame selection method for Experiment 2 provided a kind of normalization of the method used in Experiment 1."

Relativized Frequent Frame Coverage

Experiment 2 session ranges for analyzed corpora, number of utterances, number of tokens and types categorized, percentage of corpus (tokens) accounted for by categorized types, and percentage of corpus (tokens) analyzed

Child	CHILDES sessions	# of utterances	Tokens categorized	Types categorized	Percentage of corpus accounted for	Percentage of corpus analyzed
Peter	peter01-peter12	19846	5086	437	47%	5%
Eve	eve01-eve20	14922	3380	398	43%	4%
Nina	nina01-nina23	14417	4309	387	42%	6%
Naomi	n01-n58	6950	1319	294	34%	4%
Anne	anne01a-anne23b	26199	4839	512	60%	5%
Aran	aran01a-aran20b	20857	6172	676	66%	6%
Mean			4184.2	450.7	49%	5%

Similar coverage to non-relativized frequent frames

Relativized Frequent Frame Precision

Experiment 2 token and type accuracy for Standard and Expanded Labeling including baseline accuracy of random categories

Corpus	Token accuracy (Standard)		Token accuracy (Expanded)		Type accuracy (Standard)		Type accuracy (Expanded)	
	Analysis	Random	Analysis	Random	Analysis	Random	Analysis	Random
Peter	0.98	0.51	0.97	0.32	0.95	0.59	0.95	0.53
Eve	0.98	0.56	0.91	0.27	0.92	0.52	0.89	0.38
Nina	0.98	0.52	0.97	0.32	0.95	0.48	0.94	0.36
Naomi	0.96	0.56	0.96	0.39	0.94	0.51	0.93	0.42
Anne	0.98	0.37	0.82	0.23	0.94	0.40	0.90	0.34
Aran	0.97	0.45	0.80	0.22	0.91	0.42	0.88	0.33
Mean	0.98	0.49	0.91	0.29	0.94	0.50	0.92	0.39

Relativized Frequent Frame Recall

Experiment 2 token and type completeness for Standard and Expanded Labeling including baseline accuracy of random categories

Corpus	Token completeness (Standard)		Token completeness (Expanded)		Type completeness (Standard)		Type completeness (Expanded)	
	Analysis	Random	Analysis	Random	Analysis	Random	Analysis	Random
Peter	0.06	0.03	0.08	0.03	0.06	0.04	0.07	0.04
Eve	0.07	0.04	0.13	0.04	0.07	0.04	0.09	0.04
Nina	0.08	0.04	0.13	0.04	0.08	0.04	0.10	0.04
Naomi	0.07	0.04	0.11	0.05	0.07	0.04	0.08	0.04
Anne	0.10	0.04	0.13	0.04	0.10	0.04	0.13	0.05
Aran	0.10	0.05	0.17	0.05	0.11	0.05	0.13	0.05
Mean	0.08	0.04	0.13	0.04	0.08	0.04	0.10	0.04

Getting Better Scores

Getting better precision (which was already high)
"…one way to circumvent the erroneous classifications…would be to filter out extremely low frequency targets."

Getting better recall (which was pretty low)
"It is a prevalent characteristic of these frame-based categories that there is considerable overlap in the words they contain.…two frame-based categories could be unified if they surpass a threshold of lexical overlap. This possibility was tested on the results from one of the corpora, Peter, using a criterion of 20% overlap. The outcome was that 17 different verb categories were joined to form one category of 261 word types, 99.3% of which were verbs."

Unification

"Accuracy was not adversely affected by the unification of categories, remaining at 0.90 or above…indicating that the unification procedure did not join together frame-based categories containing words from different grammatical categories. Furthermore, type completeness reached 0.91…indicating that, as expected, the distributional categories that had been fragments of grammatical categories were merged by the unification procedure…it appears that a very simple conglomeration procedure based on lexical overlap could be used to join accurate smaller categories together into a more complete category."

Overlap in Action

Many frames overlap in the words they identify.

the__is	the__was	a___is	that___is …
dog	dog	dog	cat
cat	cat	goblin	goblin
king	king	king	king
girl	teddy	girl	teddy

↓

the/a/that__is/was	
dog	teddy
cat	goblin
king	
girl	

Cross-linguistic Application?

Some work done for French, and a pilot study in Cantonese:
Chemla, Mintz, Bernal, & Christophe (2009)

Very similar results: high accuracy, low completeness

Corollaries from Chemla et al.:
Reiterating the importance of the frame over the bigram or trigram

Finding it's important that frames consists of individual lexical items rather than categories made up of multiple words

Cross-linguistic Application?

"The fundamental notion is that a relatively local context *defined by frequently co-occurring units* can reveal a target word's category…[here] the units were words and the frame contexts were defined by words that frequently co-occur. In other languages, a failure to find frequent word frames could trigger an analysis of co-occurrence patterns at a different level of granularity, for example, at the level of sub-lexical morphemes. The frequently co-occurring units in these languages are likely to be the inflectional morphemes which are limited in number and extremely frequent."

Western Greenlandic

Aliikusersuillammassuaanerartassagaluarpaalli.
aliiku-sersu-i-llammas-sua-a-nerar-ta-ssa-galuar-paal-li
entertainment-provide-SEMITRANS-one.good.at-COP-say.that-REP-FUT-sure.but-3.PL.SUBJ/3SG.OBJ-but
'However, they will say that he is a great entertainer, but ...'

Wang & Mintz (2008): Dynamic FFs

"…the frequent frame analysis procedure proposed by Mintz (2003) was not intended as a model of acquisition, but rather as a demonstration of the information contained in frequent frames in child-directed speech…Mintz (2003) did not address the question of whether an actual learner could detect and use frequent frames to categorize words…"

"This paper addresses this question with the investigation of a computational model of frequent frame detection that incorporates more psychologically plausible assumptions about the memor[y] resources of learners. In addition, it implements learning as a dynamic process that takes place utterance by utterance as a corpus is processed, rather than 'in a batch' over an entire corpus."

Considering Children's Limitations

Memory Considerations

(1) Children possess limited memory and cognitive capacity and cannot track all the occurrences of all the frames in a corpus.

(2) Memory retention is not perfect: infrequent frames may be forgotten.

The Model's Operation

(1) Only 150 frame types (and their frequencies) are held in memory

(2) Forgetting function: frames that have not been encountered recently are less likely to stay in memory than frames that have been recently encountered

Dynamic Procedure

(1) Child encounters an utterance (e.g. "You read the story to mommy.")

(2) Child segments the utterance into frames:

	You	read	the	story	to	mommy.
(1)	You	X	the			
(2)		read	X	story		
(3)			the	X	to	
(4)				story	X	mommy

Dynamic Procedure

(3) If memory is not full, a newly-encountered frame is added to the memory and its initial activation is set to 1. The forgetting function is simulated by the activation for each frame in memory decreasing by 0.0075 at each processing step.

Memory			Activation
You	X	the	1.0

Processing Step 1

Dynamic Procedure

(3) If memory is not full, a newly-encountered frame is added to the memory and its initial activation is set to 1. The forgetting function is simulated by the activation for each frame in memory decreasing by 0.0075 at each processing step.

Memory			Activation
You	X	the	0.9925
read	X	story	1.0

Processing Step 2: frame *read X story*

Dynamic Procedure

(4) If the frame already exists in memory, its activation is increased by 1.

Memory			Activation
I	X	it	3.885
You	X	the	0.8945
read	X	story	0.8805
the	X	to	0.8735
story	X	mommy	0.8625
…			

Processing Step 27: frame *you X the*

Dynamic Procedure

(4) If the frame already exists in memory, its activation is increased by 1.

Memory			Activation
I	X	it	3.885
You	X	the	1.8945
read	X	story	0.8805
the	X	to	0.8735
story	X	mommy	0.8625
…			

Processing Step 27: frame *you X the*

Dynamic Procedure

(5) Since the memory buffer only stores 150 frames, it becomes full very quickly (after ~50 utterances). When memory is full, a newly-encountered frame replaces the least active frame with activation less than 1.

Memory			Activation
I	X	it	8.75
you	X	the	6.995
read	X	story	5.65
the	X	to	5.45
story	X	mommy	5.35
…			
you	X	it	0.9925
with	X	and	0.7965

Processing Step 101: new frame *with X by*

Dynamic Procedure

(5) Since the memory buffer only stores 150 frames, it becomes full very quickly (after ~50 utterances). When memory is full, a newly-encountered frame replaces the least active frame with activation less than 1.

Memory			Activation
I	X	it	8.75
you	X	the	6.995
read	X	story	5.65
the	X	to	5.45
story	X	mommy	5.35
…			
with	X	by	1.0
you	X	it	0.9925

Processing Step 101: new frame *with X by*

Dynamic Procedure

(6) If all activations are greater than 1, no change is made other than the forgetting function (activation - .0075)

Memory			Activation
I	X	it	8.75
you	X	the	6.995
read	X	story	5.65
the	X	to	5.45
story	X	mommy	5.35
…			
you	X	it	1.9925
with	X	and	1.7965

Processing Step 101: new frame *with X by*

Dynamic Procedure

(6) If all activations are greater than 1, no change is made other than the forgetting function (activation - .0075)

Memory			Activation
I	X	it	8.7425
you	X	the	6.9875
read	X	story	5.6425
the	X	to	5.4425
story	X	mommy	5.3425
…			
you	X	it	1.9850
with	X	and	1.7890

Processing Step 101: new frame *with X by*

Input & Performance Gauge

Using same corpora for input as Mintz (2003) (6 from CHILDES)

Model's performance was evaluated every 100 frames.
Metric used: accuracy/precision (not recall)

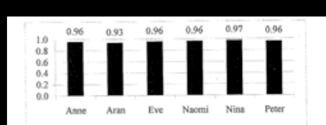


Figure 1 Accuracies after processing the six corpora

How many of the overall most frequent frames were in the model's top 45? Eve corpus

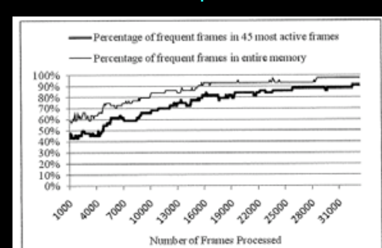


Figure 4 Percentage of Frequent Frames of Eve

How many of the overall most frequent frames were in the model's top 45? Peter corpus

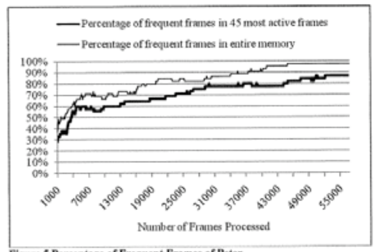


Figure 5 Percentage of Frequent Frames of Peter

What about the ones that weren't frequent frames? Are they still good categorizers?

Table 4 Ten most active frames that are not frequent frames (Mintz, 2003) in that corpus

Nina	Eve	Peter
who_X_you	no_X_not	you_X_ta
you_X_what**	a_X_bit	ta_X_it
what_X_these	it_X_you	do_X_have**
what_X_we**	what_X_of	it_X_you
a_X_on**	where_X_you**	the_X_and**
we_X_the	are_X_gonna**	the_X_room
the_X_to**	on_X_floor**	do_X_see
put_X_in**	you_X_some**	on_X_floor**
I_X_the	is_X_a**	what're_X_gonna
what_X_to	out_X_the	I'll_X_it**

** means this frame is a frequent frame in other corpora.

Table 5 Intervening word and number of tokens of *we_x_the* frame

put	36	set	1	fed	1
make	4	fill	1	see	1
bring	2	hang	1	attach	1
fix	2	comb	1	take	1
building	1	build	1	saw	1
have	1	chased	1		
give	1	open	1		

Wang & Mintz (2008) Conclusions

"…our model demonstrates very effective categorization of words. Even with limited and imperfect memory, the learning algorithm can identify highly informative contexts after processing a relatively small number of utterances, thus yield[ing] a high accuracy of word categorization. It also provides evidence that frames are a robust cue for categorizing words."