Psych 229: Language Acquisition

Lecture 8

Category Models & Speech Perception

Mintz 2003: Frequent Frames

6 corpora *Do we believe these are a realistic representation of data children hear outside the laboratory?

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%

evaluation metrics

"precision"

$$\text{Accuracy} = \frac{\text{hits}}{\text{hits} + \text{false alarms}}$$

"recall"

$$\text{Completeness} = \frac{\text{hits}}{\text{hits} + \text{misses}}$$

Using 45 as absolute cut-off for "frequent" frame

[redacted] Next, a subset of these frames was selected as the set of frequent frames. 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. Hence, the frames analyzed for each corpus were the 45 most frequent frames for that corpus.

The distributional information provided by frequent frames was robust. The word types that were categorized constituted, on average, 50% of the tokens in a given corpus. This coverage was achieved by analyzing only about 6% of the tokens and their contexts. That is, the tokens of the categorized types making up the 6% contained in frequent frames constituted half the tokens in a given corpus. 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. The efficiency and accuracy provided by frequent frames could thus be very useful to young language learners, who have limited memory and processing resources.

Mintz 2003: Frequent Frames

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

Peter
Frame 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)

Frame 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), crushed (1), change (1), calling (1), bring (1), break (1), because (1), banged (1)

Frame 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)

Mintz 2003: Frequent Frames

"precision"

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

"recall"

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

Mintz 2003: Frequent Frames

What's with the low recall?

Although the categories formed were unquestionably accurate, 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.[4] This outcome is reflected in the comparatively low completeness scores. Nevertheless, it is clear from Table 2 that the categories, in general, are relatively large (by token or type counts); thus, it was not the case that low completeness was due to numerous accurate categories with only a few members each.

What might be done about it…

Although for the most part, words in a given frame-based category belonged to the same grammatical category, there were some categorization errors. As mentioned above, differences in accuracy between Standard and Expanded Labeling indicate that nouns and pronouns were occasionally grouped together, as were auxiliaries and main verbs. In addition, in some cases prepositions and verbs were grouped together. For example, the frame it__the was a frequent frame for prepositions in four of the corpora, however in three of those corpora the frame contained some verbs as well. Similarly, go__the occurred in four corpora as a frequent frame containing prepositions, but contained some verbs as well in three of those corpora. In the it__the case, the verbs that occurred were infrequent (occurring once or twice in the analyzed samples). Thus, one way to circumvent the erroneous classifications such as these would be to filter out extremely low frequency targets.

Mintz 2003: Frequent Frames

The robustness of frequent frames

As the figure shows, 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 lists all the frequent frames that occurred in at least two corpora, organized by the number of corpora in which they occurred.

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

[redacted] These results are especially impressive when one considers the restricted distributional contexts used here – the 45 most frequent frames – and they are compelling given the evidence that infants and adults naturally attend to this type of unit.

Mintz 2003: Frequent Frames

Getting away from an absolute threshold

A limitation in the present experiment is that the set of frequent frames was selected by the same absolute threshold for all corpora (the 45 most frequent frames). 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.

How about one relative to the data seen?

The purpose of Experiment 2 was to examine the categorization outcome when a frame selection method is used that is sensitive to the frequency of frames relative to the total number of frames in a corpus. An additional goal was to ensure that high accuracy scores in Experiment 1 were not due to very small categories with only a few member types, as such categorization, although accurate, is not linguistically interesting.

The magic percentage: 0.0013

As in Experiment 1, within each corpus, all frames were tallied and ranked by frequency. The set of frequent frames was then 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%. That is, a given frame in a corpus was included as a frequent frame just in case its frequency was at least 0.13% of the total number of frames in the corpus. This specific threshold was determined based on the frequent frames for each corpus in Experiment 1. In particular, the frequent frames in Experiment 1 were analyzed, corpus by corpus, by tallying the frequency of the least frequent member of the set of frequent frames, and expressing that frequency as a proportion of the total number of frames for that corpus, yielding a different proportional threshold for each corpus. These thresholds were then averaged, yielding 0.0013, or 0.13%, and this was the threshold used for all corpora in Experiment 2. Thus, the frequent frame selection method for Experiment 2 provided a kind of normalization of the method used in Experiment 1.

Mintz 2003: Frequent Frames

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	1519	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%

"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	0.98	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	0.86	0.33
Mean	0.98	0.49	0.91	0.29	0.94	0.50	0.92	0.39

"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

Mintz 2003: Frequent Frames

Back to that low "recall" score problem…

There is at least one simple way to unify distinct frame-based categories that contain words from the same grammatical category. It is a prevalent characteristic of these frame-based categories that there is considerable overlap in the words they contain. For example, the verb categories defined by frames you__to, she__to, you__the, etc., will generally have a number of member words in common because many of the same verbs can appear in each environment. Hence, 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. The non-verb items were from disparate grammatical categories and only occurred once or twice in the frames that constituted the composite category.[6] Accuracy was not adversely affected by the unification of categories, remaining at 0.90 or above (for all combinations of type/token and Standard/Expanded Labeling), indicating that the unification procedure did not join together frame-based categories containing words from different grammatical categories. Furthermore, type completeness reached 0.91 (compared to 0.07 before unification!), indicating that, as expected, the distributional categories that had been fragments of grammatical categories were merged by the unification procedure. Although further research is needed to better understand the effectiveness and limitations of this technique, 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.

Good results with a simple unification procedure!

Mintz 2003: Frequent Frames

Distributional analysis = not so bad

What is remarkable given the potential pitfalls for distributional analyses is that with frequent frames, the misclassifications are quite rare, as indicated by the high accuracy scores and the representative categories shown in Table 2.[8]

Working with semantic bootstrapping

One way to view the utility of the distributional information described here is as a way of bootstrapping into a parameterized universal grammar that contains category distinctions, such as noun and verb, specifications of whether the category is a phrasal head, specifications of options of ways in which phrases could be combined, etc. Fitting the distributional categories into the grammar would then amount to labeling the distributional categories as noun, verb, adjective, etc. One might call such a procedure Distributional Bootstrapping, as this brief sketch resembles in some respects Pinker's (1984) Semantic Bootstrapping proposal. However, for Pinker the foundational "boot-strap" categories were derived from semantic information rather than distributional information (see also Grimshaw, 1981; MacNamara, 1972).

Using distributional analysis without innate semantic biases

It is also conceivable, although the mechanisms are less clear, that properly constrained distributional information, perhaps in concert with prosodic information (e.g. Fisher & Tokura, 1996), could be used to induce higher order grammatical relations such as phrasal constituency and hierarchical structure that are pre-given in most bootstrapping accounts. The frequent frames investigated here capture some local information that might be relevant for positing higher order structural relationships (for example, argument structure for verbs with pronominal arguments: I__you, etc.). Thus, frames might be the seeds for growing higher order trees that would effectively make distributional categories syntactic. The question would then be whether the mechanisms required to motivate the construction of higher order generalizations, and to constrain the induction mechanisms to focus on the right kinds of distributional facts, would be theoretically discernable from the kinds of innate knowledge generally assumed under most bootstrapping accounts.

Mintz 2003: Frequent Frames

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.

In the procedures explored 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 are extremely frequent.

What about an actual learning model that uses frequent frames over real data?

BUCLD 32
Boston University Conference on Language Development

A Dynamic Learning Model For Categorizing Words Using Frames

Hao Wang, Toben Mintz
Department of Psychology, University of Southern California

Mintz 2003: Frequent Frames

Discussion Questions

Relation to Gambell & Yang: Word segmentation based on words having one strong syllable only vs. frequent frames made up of unstressed monosyllables?

Frequent frames in other languages? Ex: French, where le/la becomes l' before word beginning with vowel. Ex: Languages without fixed word order - what are the frequent "frames" there?

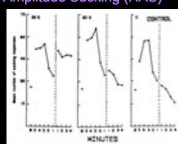
How much more "bookkeeping" does tracking frequent frames cause over tracking conditional probabilities?

Gerken & Aslin (2005): Jusczyk Speech Perception Research Review

The first phase of his research program on infant speech perception consisted of basic descriptive studies of the infants' ability, using mostly the high-amplitude sucking (HAS) technique, to discriminate phonetic contrasts instantiated in simple consonant–vowel (C–V) syllables or vowel–consonant (V–C) syllables (Jusczyk, 1977; Jusczyk, Copan, & Thompson, 1978) as well as in multisyllabic tokens (Jusczyk & Thompson, 1978). However, after these initial successes in demonstrating infants' sophisticated phonetic discrimination abilities, it became apparent to the field that the auditory correlates of phonetic contrasts may mediate this performance. That is, unlike adults, infants may perceive phonetic contrasts at the level of auditory tokens rather than as linguistic units.

Infants don't use linguistic units the same way adults do for making sound contrasts

High Amplitude Sucking (HAS)



A Brief Foray into Infant Experimental Techniques

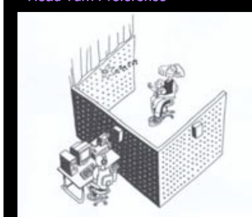
High Amplitude Sucking

Figure 1 High-amplitude sucking technique

Infants must be awake in a quiet alert state before the beginning of the study. Infants are placed in a comfortable reclined bath chair, and offered a sterilized pacifier that is connected to a pressure transducer and a computer via a piece of rubber turbing. Once the infant has begun sucking, the computer measures the infant's average sucking amplitude, or the strength of the infant's sucks. Following this baseline period, a sound is presented to the infant every time a strong or "high amplitude" suck is delivered. Infants quickly learn that their sucking controls the sounds, and they will suck more strongly and more often to hear the sounds they like the most. Infants' sucking rate over time can also be measured, to see if an infant "notices" when new sounds are played.

A Brief Foray into Infant Experimental Techniques

Head Turn Preference



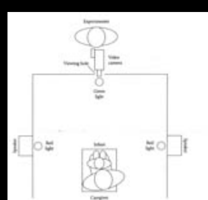


Figure 2 Head turn preference procedure

Stimulus continues until infant looks away for 2 seconds.

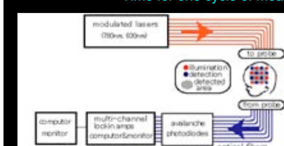
A Brief Foray into Infant Experimental Techniques

Optical Topography

Silent, noninvasive. Used for infant auditory testing. (Infant preference for forward native speech over backward native speech.)

Measure penetration of infrared light to gauge hemoglobin (blood) absorption in the brain. When brain area is activated, blood volume changes quickly.

Time for one cycle of measurement: 0.1 seconds. (real-time)



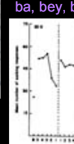
Gerken & Aslin (2005): Jusczyk Speech Perception Research Review

This hypothesis—that, in early infancy, speech may not be perceived at a phonemic level—was supported by studies demonstrating losses in sensitivity to phonetic contrasts, as infants had accumulated more language-specific information by the second half of the first year (e.g., Werker & Tees, 1984).

Young infants don't use phonemes (b, p, t, o, u, …)

A subsequent approach to the question of whether infants' early discriminations involved phonetic segments or some other unit of analysis consisted of studies that asked whether infants could recognize the similarity of phonemes despite variability in their surrounding context. Jusczyk and Derrah (1987) used the HAS technique to determine whether 2-month-olds could recognize the similarity of C–V syllables that share an initial consonant (e.g., /b/). If infants could treat acoustically variable phonetic segments as members of the same category, then a phonemic level of analysis would be supported. The results suggested that infants at this age could not in fact extract the common initial consonant from multiple exemplars that had variable vowels. Further studies (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988) of newborns and 2-month-olds showed that neither vowels nor consonants are extracted as phonetic segments from multiple exemplars. These results added support to the hypothesis that the syllable, rather than the phonetic segment, is the basic unit of speech perception in early infancy.

ba, bey, bi, bo, boo…



Young infants use syllables

Gerken & Aslin (2005): Jusczyk Speech Perception Research Review

Evidence for tracking distributional properties

The final phase of Peter's work on speech discrimination showed that young infants, who seemingly do not analyze speech at a subsyllabic level, are nevertheless sensitive to coarticulation and context effects. Levitt, Jusczyk, Murray, and Carden (1988) showed that 2-month-olds' discrimination of a /ba/–/da/ contrast, like that of adults, is influenced by the surrounding phonetic context—in this case the presence or absence of frication noise. These results are consistent with other evidence, reviewed below, that infants are keeping track of the distributional properties of their native-language input, although at the level of diphones rather than individual phonetic segments.

…over syllable context

Prosody, the melody of language

The team of Jusczyk, Hirsh-Pasek, and Kemler Nelson and their colleagues used this technique to present infants with passages, either from a child's book or as spontaneous speech from a parent to a young child (Hirsh-Pasek et al., 1987; Jusczyk, Hirsch-Pasek, et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989). Pauses at clause boundaries (Jusczyk, Hirsch-Pasek et al., 1992) were spliced out, and new pauses of a uniform duration were inserted at the correct boundaries or at nonboundary positions. Infants as young as 6-months were able to distinguish correctly "paused" passages from those incorrectly paused when the correct pauses were at clause boundaries.

Infants sensitive to pauses at syntactic clause boundaries