Psych 156A/ Ling 150: Psychology of Language Learning

Lecture 6

Words III - Grammatical Categories

Announcements

Lecture notes from last time corrected & posted (there was an error in one of the slides on recall and precision)

Pick up HW1

Be working on HW2 and the review questions for words

Grammatical Categorization

Computational Problem: Identify the grammatical category of a word (such as noun, verb, adjective, preposition, etc.)

This will tell you how this word is used in the language, and will allow you to recognize other words that belong to the same category since they will be used the same way.

Examples of different categories in English:

noun = goblin, kitten, king, girl

Examples of how nouns are used:

I like that goblin. Kittens are adorable.

A king said that no girls would ever solve the Labyrinth.

Grammatical Categorization

Computational Problem: Identify the grammatical category of a word (such as noun, verb, adjective, preposition, etc.)

This will tell you how this word is used in the language, and will allow you to recognize other words that belong to the same category since they will be used the same way.

Examples of different categories in English:

verb = like, are, said, solve, stand

Examples of how verbs are used:

I like that goblin. Kittens are adorable.

A king said that no girls would ever solve the Labyrinth.

Sarah was standing very close to him.

Grammatical Categorization

Computational Problem: Identify the grammatical category of a word (such as noun, verb, adjective, preposition, etc.)

This will tell you how this word is used in the language, and will allow you to recognize other words that belong to the same category since they will be used the same way.

Examples of different categories in English:
adjective = silly, adorable, brave, close

Examples of how adjectives are used:
I like the silliest goblin. Kittens are so adorable.
The king said that only brave girls would solve the Labyrinth.
Sarah was standing very close to him.

Grammatical Categorization

Computational Problem: Identify the grammatical category of a word (such as noun, verb, adjective, preposition, etc.)

This will tell you how this word is used in the language, and will allow you to recognize other words that belong to the same category since they will be used the same way.

Examples of different categories in English:
preposition = near, through, to

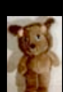
Examples of how prepositions are used:
I like the goblin near the king's throne.
The king said that no girls would get through the Labyrinth.
Sarah was standing very close to him.

Grammatical Categorization

Computational Problem: Identify the grammatical category of a word (such as noun, verb, adjective, preposition, etc.)

This will tell you how this word is used in the language, and will allow you to recognize other words that belong to the same category since they will be used the same way.

"This is a DAX."
DAX = ??



"He is SIBing."
SIB = ??

"He is very BAV."
BAV = ??

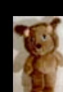
"He should sit GAR the other dax."
GAR = ??

Grammatical Categorization

Computational Problem: Identify the grammatical category of a word (such as noun, verb, adjective, preposition, etc.)

This will tell you how this word is used in the language, and will allow you to recognize other words that belong to the same category since they will be used the same way.

"This is a DAX."
DAX = noun



"He is SIBing."
SIB = verb

"He is very BAV."
BAV = adjective

"He should sit GAR the other dax."
GAR = preposition

Categorization: How?

How might children initially learn what categories words are?

Idea 1: 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*)

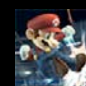
action = verb (*steal, sing*)

property = adjective (*shiny, stinky*)

The word's meaning is then linked to innate grammatical category knowledge (nouns are objects/substances, verb are actions, adjectives are properties)

Semantic Bootstrapping Hypothesis: Problem

Mapping rules are not perfect
Ex: not all action-like words are verbs



"bouncy", "a kick"
action-like meaning, but they're not verbs

Ex: not all property-like words are adjectives

"is shining", "it glitters"
seem to be referring to properties, but these aren't adjectives

Categorization: How?

Idea 2: Distributional Learning

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

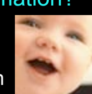
Kittens are adorable. Noun

Sarah was standing very close to him. Verb

I like the silliest goblin. Adjective

The king said that no girls would get through the Labyrinth. Preposition

Are children sensitive to distributional information?



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

15-16 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)

Mintz 2003: Is distributional information enough?

How do we know in child-directed speech (which is the linguistic data children encounter)…

(1) What distributional information children should pay attention to?

(2) If the available distributional information will actually correctly categorize words?

Mintz 2003: What data should children pay attention to?

"…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."

Mintz 2003: 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…

Mintz 2003: Samples of Child-Directed Speech

Data representing child's linguistic environment:
6 corpora of child-directed speech from the CHILDES database, which contains transcriptions of parents interacting with their children.

CHILDES Child Language Data Exchange System

Corpus (sg.), corpora (pl). = a collection of data
[from Latin *body*, a "body" of data]

Mintz 2003:
Defining "Frequent"

Definition of "frequent" for frequent frames:
Frames appearing a certain number of times in a 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…. a pilot analysis with a randomly chosen corpus, Peter, determined that the 45 most frequent frames satisfied these goals and provided good categorization."

Set of frequent frames = 45 most frequent frames

Mintz 2003:
Defining "Frequent"

Example of deciding which frames were frequent:

Frame	How often it occurred in the corpus
(1) the___is	600 times
(2) a___is	580 times
(3) she__it	450 times
…	
(45) they__him	200 times
(46) we___have	199 times
…	

These frames considered "frequent"

Mintz 2003:
Testing the Categorization Ability of Frequent Frames

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

Frame (1): the___is
Transcript: "…the radio is in the way…but the doll is…and the teddy is…"

radio, doll, teddy are placed into the same category by the___is

Frame (13): you___it
Transcript: "…you draw it so that he can see it… you dropped it on purpose!…so he hit you with it…"

draw, dropped, with are placed into the same category by you___it

Mintz 2003:
Determining the success of frequent frames

$$\text{Precision} = \frac{\text{\# of words identified correctly as Category within frame}}{\text{\# of words identified as Category within frame}}$$

$$\text{Recall} = \frac{\text{\# of words identified correctly as Category within frame}}{\text{\# of words that should have been identified as Category}}$$

Mintz 2003: Determining the success of frequent frames

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

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

Frame: you___it
Category: draw, dropped, with (similar to Verb so compare to Verb)

# of words correctly identified as Verb = 2 (draw, dropped)
# of words identified as Verb = 3 (draw, dropped, with)
Precision for you___it = 2/3

Mintz 2003: Determining the success of frequent frames

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

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

Frame: you___it
Category: draw, dropped, with (similar to Verb so compare to Verb)

# of words correctly identified as Verb = 2 (draw, dropped)
# of words should be identified as Verb = all verbs in corpus (play, sit, draw, dropped, ran, kicked, …)

Mintz 2003: Determining the success of frequent frames

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

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

Frame: you___it
Category: draw, dropped, with (similar to Verb so compare to Verb)

# of words correctly identified as Verb = 2
# of words should be identified as Verb = 100
Recall = 2/100 (much smaller number)

Mintz 2003: Some actual frequent frame results

Frame: you___it

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

Mintz 2003: Some actual frequent frame results

Frame: the___is

Category includes:
moon, sun, truck, smoke, kitty, fish, dog, baby, tray, radio, powder, paper, man, lock, lipstick, lamb, kangaroo, juice, ice, flower, elbow, egg, door, donkey, doggie, crumb, cord, clip, chicken, bug, brush, book, blanket, Mommy

Mintz 2003: How successful frequent frames were

Precision: Above 90% for all corpora (high) = very good!

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

Recall: Around 10% for all corpora (very low) = maybe not as good…

Interpretation: A frequent frame made lots of little clusters, rather than being able to cluster all the words into one category. (So, there were lots of Noun-ish clusters, lots of Verb-ish clusters, etc.)

Mintz 2003: Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: 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

What about putting clusters together that have a certain number of words in common?

Mintz 2003: Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: 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

Mintz 2003: Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: Many frames overlap in the words they identify.

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

Mintz 2003: Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: Many frames overlap in the words they identify.

the__is/was	a___is	that___is …
dog	dog	cat
cat	goblin	goblin
king	king	king
girl	girl	teddy
teddy		

Mintz 2003: Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: Many frames overlap in the words they identify.

the__is/was, a___is		that___is …
dog	goblin	cat
cat		goblin
king		king
girl		teddy
teddy		

Mintz 2003: Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: Many frames overlap in the words they identify.

the/a__is/was		that___is …
dog	goblin	cat
cat		goblin
king		king
girl		teddy
teddy		

Mintz 2003:
Getting better recall

How could we form just one category of Verb, Noun, etc.?

Observation: Many frames overlap in the words they identify.

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

Recall goes up to 91% (very high) = very good!
Precision stays above 90% (very high) = very good!

Mintz 2003: Recap

Frequent frames are non-adjacent co-occurring words with one word in between them. (ex: the___is)

They are likely to be information young children are able to track, based on experimental studies.

When tested on realistic child-directed speech, frequent frames do very well at grouping words into clusters which are very similar to actual grammatical categories like Noun and Verb.

Frequent frames could be a very good strategy for children to use.

Wang & Mintz 2008:
Simulating children using frequent frames

"…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…"

Wang & Mintz 2008:
Simulating children using frequent frames

"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."

Computational model: a program that simulates the mental processes occurring in a child. This requires knowing what the input and output are, and then testing the algorithms that can take the given input and transform it into the desired output.

Wang & Mintz (2008): 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

Wang & Mintz (2008): How the model works

(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

Frames:
you___the, read___story, the___to, story___mommy

Wang & Mintz (2008): How the model works

If memory is not full, a newly-encountered frame is added to the memory and its initial activation is set to 1.

Memory	Activation

Processing Step 1

Wang & Mintz (2008): How the model works

If memory is not full, a newly-encountered frame is added to the memory and its initial activation is set to 1.

Memory	Activation
you___the	1.0

Processing Step 1 (you___the)

Wang & Mintz (2008): How the model works

The forgetting function is simulated by the activation for each frame in memory decreasing by 0.0075 after each processing step.

Memory	Activation
you___the	0.9925

Forgetting function

Wang & Mintz (2008): How the model works

When a new frame is encountered, the updating depends on whether the memory is already full or not. If it is not and the frame has not already been encountered, the new frame is added to the memory with activation 1.

Memory	Activation
read___story	1.0
you___the	0.9925

Processing Step 2 (read___story)

Wang & Mintz (2008): How the model works

When a new frame is encountered, the updating depends on whether the memory is already full or not. If it is not and the frame has not already been encountered, the new frame is added to the memory with activation 1.

Memory	Activation
read___story	0.9925
you___the	0.9850

Forgetting function

Wang & Mintz (2008): How the model works

When a new frame is encountered, the updating depends on whether the memory is already full or not. If it is not and the frame has not already been encountered, the new frame is added to the memory with activation 1.

Memory	Activation
the___to	1.0
read___story	0.9925
you___the	0.9850

Processing step 3 (the___to)

Wang & Mintz (2008): How the model works

When a new frame is encountered, the updating depends on whether the memory is already full or not. If it is not and the frame has not already been encountered, the new frame is added to the memory with activation 1.

Memory	Activation
the___to	0.9925
read___story	0.9850
you___the	0.9775

Forgetting function

Wang & Mintz (2008): How the model works

When a new frame is encountered, the updating depends on whether the memory is already full or not. If it is not and the frame has not already been encountered, the new frame is added to the memory with activation 1.

Memory	Activation
story___mommy	1.0
the___to	0.9925
read___story	0.9850
you___the	0.9775

Processing step 4 (story___mommy)

Wang & Mintz (2008): How the model works

When a new frame is encountered, the updating depends on whether the memory is already full or not. If it is not and the frame has not already been encountered, the new frame is added to the memory with activation 1.

Memory	Activation
story___mommy	0.9925
the___to	0.9850
read___story	0.9775
you___the	0.9700

Forgetting function

Wang & Mintz (2008): How the model works

If the frame is already in memory because it was already encountered, activation for that frame increases by 1.

Memory	Activation
story___mommy	0.9925
the___to	0.9850
read___story	0.9775
you___the	0.9700

Processing step 5: (you____the)

Wang & Mintz (2008): How the model works

If the frame is already in memory because it was already encountered, activation for that frame increases by 1.

Memory	Activation
story___mommy	0.9925
the___to	0.9850
read___story	0.9775
you___the	1.9700

Processing step 5: (you____the)

Wang & Mintz (2008): How the model works

If the frame is already in memory because it was already encountered, activation for that frame increases by 1.

Memory	Activation
you___the	1.9700
story___mommy	0.9925
the___to	0.9850
read___story	0.9775

Processing step 5: (you____the)

Wang & Mintz (2008): How the model works

If the frame is already in memory because it was already encountered, activation for that frame increases by 1.

Memory	Activation
you___the	1.9625
story___mommy	0.9850
the___to	0.9775
read___story	0.9700

Forgetting function

Wang & Mintz (2008): How the model works

Eventually, since the memory only holds 150 frames, the memory will become full.

Memory	Activation
story___mommy	4.6925
the___to	3.9850
read___story	3.9700
you___the	2.6925
…	…
she___him	0.9850
we__it	0.7500

Memory after processing step 200

Wang & Mintz (2008): How the model works

At this point, if a frame not already in memory is encountered, it replaces the frame with the least activation, as long as that activation is less than 1.0.

Memory	Activation
story___mommy	4.6925
the___to	3.9850
read___story	3.9700
you___the	2.6925
…	…
she___him	0.9850
we__it	0.7500

Processing step 201: because___said

Wang & Mintz (2008): How the model works

At this point, if a frame not already in memory is encountered, it replaces the frame with the least activation, as long as that activation is less than 1.0.

Memory	Activation
story___mommy	4.6925
the___to	3.9850
read___story	3.9700
you___the	2.6925
…	…
she___him	0.9850
~~we__it~~	~~0.7500~~

Processing step 201: because___said

Wang & Mintz (2008): How the model works

At this point, if a frame not already in memory is encountered, it replaces the frame with the least activation, as long as that activation is less than 1.

Memory	Activation
story___mommy	4.6925
the___to	3.9850
read___story	3.9700
you___the	2.6925
…	…
because___said	1.0000
she___him	0.9850

Processing step 201: because___said

Wang & Mintz (2008): How the model works

Eventually, however, all the frames in memory will have been encountered often enough that their activations are greater than 1.

Memory	Activation
story___mommy	9.6925
the___to	8.9850
read___story	8.9700
you___the	5.6925
…	…
we___her	3.9700
she___him	2.9850

Memory after processing step 5000

Wang & Mintz (2008): How the model works

At this point, no change is made to memory since the new frame's activation of 1 would be less than the least active frame in memory.

Memory	Activation
story___mommy	9.6925
the___to	8.9850
read___story	8.9700
you___the	5.6925
…	…
we___her	3.9700
she___him	2.9850

Processing step 5001 (because___him)

Wang & Mintz (2008): How the model works

The forgetting function is then invoked.

Memory	Activation
story___mommy	9.6850
the___to	8.9775
read___story	8.9625
you___the	5.6850
…	…
we___her	3.9625
she___him	2.9775

Forgetting function

Wang & Mintz (2008): How the model did

Using same corpora for input as Mintz (2003)
(6 from CHILDES: Anne, Aran, Even, Naomi, Nina, Peter)

The model's precision was above 0.93 for all six corpora.
This is very good!
When the model decided a word belonged in a particular category (Verb, Noun, etc.) it usually did.

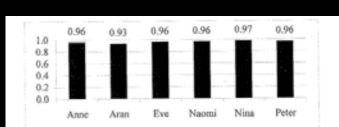


Figure 1 Accuracies after processing the six corpora

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."

Wang & Mintz (2008): Recap

While Mintz (2003) showed that frequent frame information is useful for categorization, it did not demonstrate that children - who have constraints like limited memory and cognitive processing power - would be able to effectively use this information.

Wang & Mintz (2008) showed that a model using frequent frames in a psychologically plausible way (that is, a way that children might identify and use frequent frames) was able to have the same success at identifying the grammatical category that a word is.

Questions?