Psych 215L: Language Acquisition

Lecture 11

Morphology III: Learning Models

Words & Rules

Computational Problem: Identifying word affixes that signal meaning.
= Identify the rules for altering word forms in order to signal meaning.

Example: What do you have to change about the verb to signal the past tense in English? (There are both regular and irregular patterns.)

blink~blinked (+ed)	confide~confided (+ed)	drink~drank ("ih" --> "ey")
rub~rubbed (+ed)	hide~hid ("aye" --> "ih")	think~thought ("ink" --> "ought")

Psychological Reality of Rules: Debate

Pinker & Ullman (2002): Words and Rules
There is a rule for the regular pattern (+ed), but irregular verbs are stored in an associative memory. There is no abstraction of irregular patterns like *drink~drank* and *sink~sank*. To use an irregular past tense form, a speaker simply retrieves the appropriate irregular form from memory.

Rumelhart & McClelland (1986), McClelland & Patterson (2002): Words, No Rules
There is associative memory for everything. The mind never explicitly uses a rule to transform a verb into its past tense.

Yang (2002), Chomsky & Halle (1968): Rules, No Words
There are rules for everything, both regular and irregular patterns.

About those irregular past tense forms

Regular past tense rule: +ed

Applies to every verb

Irregular past tense rule 1: no change

Applies to: cut~cut, hurt~hurt, fit~fit, …

Irregular past tense rule 2: ink --> ank

Applies to: drink~drank, sink~sank, shrink~shrank, …

Irregular past tense rule 3: final vowel sound --> "oo"

Applies to: draw~drew, fly~flew, know~knew, …

About those irregular past tense forms

Regular past tense rule: +ed

Applies to every verb — walk, blink, sigh, … More general

Irregular past tense rule 1: no change

Applies to: cut~cut, hurt~hurt, fit~fit, …

Irregular past tense rule 2: ink --> ank

Applies to: drink~drank, sink~sank, shrink~shrank, …

Irregular past tense rule 3: final vowel sound --> "oo"

Applies to: draw~drew, fly~flew, know~knew, …

More specific: applies to just these verbs

Irregular rules

How do we know if humans really abstract across irregular verbs with neighboring (rhyming) past tense forms and store rules unconsciously in their minds the way we think they do for the regular past tense?

Competing idea 1

No Irregular Rules: Irregular past tense performance for any given verb is based on how frequently the child hears that past tense form. There may be some benefit to performance if the verb form has neighboring irregular words ("drink" benefits from "sink" and "shrink").

What matters: frequency of that verb's past tense form in the child's input

Irregular rules

How do we know if humans really abstract across irregular verbs with neighboring (rhyming) past tense forms and store rules unconsciously in their minds the way we think they do for the regular past tense?

Competing idea 1

No Irregular Rules

What matters: frequency of verb's past tense form in the child's input

Prediction for children's behavior: Children should perform the same on verb past tense forms they encounter equally often.

Irregular rules

How do we know if humans really abstract across irregular verbs with neighboring (rhyming) past tense forms and store rules unconsciously in their minds the way we think they do for the regular past tense?

Competing idea 2

Irregular Rules: Irregular past tense performance for any given verb is based on how frequently the child hears that past tense form *and* how often the child hears any irregular verbs that follow the same past tense rule (ex: draw~drew follows the same rule as fly~flew, grow~grew, know~knew, so "draw" benefits from the past tense forms of these verbs, too).

What matters: frequency of individual verb past tense form, frequency of neighboring (sometimes rhyming) past tense forms [rule frequency]

Irregular rules

How do we know if humans really abstract across irregular verbs with neighboring (rhyming) past tense forms and store rules unconsciously in their minds the way we think they do for the regular past tense?

Competing idea 2

Irregular Rules

What matters: frequency of individual verb past tense form, frequency of neighboring (sometimes rhyming) past tense forms [rule frequency]

Prediction for children's behavior: For verb past tense forms that children hear equally often, they should perform better on verbs that belong to an irregular rule class whose members appear more frequently.

Irregular rules

How do we know if humans really abstract across irregular verbs with neighboring (rhyming) past tense forms and store rules unconsciously in their minds the way we think they do for the regular past tense?

Competing ideas: Predictions

No Irregular Rules

Prediction for children's behavior: Children should be the same on verb past tense forms they encounter equally often.

Irregular Rules

Prediction for children's behavior: For verb past tense forms that children hear equally often, they should perform better on verbs that belong to an irregular rule class whose members appear more frequently.

Yang (2002): Irregular Rules

Evidence from CHILDES database

Children encounter "hurt" and "cut" as often as "draw", "blow", "grow", and "fly" [20 times in a given corpus of a child's experience]

Results:

Performance on "hurt" and "cut": ~80% success at correct irregular form

Performance on "draw", "blow", "grow", and "fly": ~35% success

Different performance for same frequency verbs!
Why?

Yang (2002): Irregular Rules

Evidence from CHILDES database

Children encounter "hurt" and "cut" as often as "draw", "blow", "grow", and "fly" [20 times in a given corpus of a child's experience]

Results:

Performance on "hurt" and "cut": ~80% success at correct irregular form
"No change" rule: hurt~hurt, cut~cut
Other verbs with same rule: hit, quit, split, slit, spit, bid, rid, forbid, spread, wed, let, set, upset, wet, shut, put, burst, cast, cost, thrust many!
rule frequency: > 2500

Performance on "draw", "blow", "grow", and "fly": ~35% success
"Vowel goes to 'oo'" rule: draw~drew, blow~blew, grow~grew, fly~flew
Other verbs with same rule: know, throw, withdraw, slay less!
rule frequency: < 100

Yang (2002): Irregular Rules

Evidence from CHILDES database

Children encounter "hurt" and "cut" as often as "draw", "blow", "grow", and "fly" [20 times in a given corpus of a child's experience]

Results:

Performance on "hurt" and "cut": ~80% success at correct irregular form
Many "No Change" rule verbs. These verbs have benefited from children encountering the other verbs with the same rule. Better performance.

Performance on "draw", "blow", "grow", and "fly": ~35% success
Less "Vowel goes to 'oo'" rule verbs. These verbs have not benefited, since there are not many other verbs with the same rule. Worse performance.

Yang (2002): Irregular Rules

Evidence from CHILDES database

Implication: Children seem to benefit from rule use frequencies of verbs ("cut" and "hurt" benefit from the higher frequency of "no change" rule verbs).

Support for the existence of Irregular Rules.

Another Test for Irregular rules

How do we know if humans really abstract across irregular verbs with neighboring (sometimes rhyming) past tense forms and store rules unconsciously in their minds the way we think they do for the regular past tense?

Competing ideas

No Irregular Rules
Prediction for children's behavior: Children should perform better on verbs they hear more frequently.

Irregular Rules

Prediction for children's behavior: Children could perform better on verbs they hear less frequently if those verbs follow an irregular past tense rule that many other verbs follow (freerider effect)

Yang (2002): Irregular Rules

Evidence from CHILDES database

How often children encounter certain verbs in a given corpus:
"hurt", "cut": 20 times
"caught": 36 times "threw": 31 times
"knew": 58 times

Performance on "hurt" and "cut": ~80% success
Performance on "caught": ~96% success

Performance on "threw": ~49% success
Performance on "knew": ~49% success

Yang (2002): Irregular Rules

Evidence from CHILDES database

How often children encounter certain verbs in a given corpus:
"hurt", "cut": 20 times
"caught": 36 times "threw": 31 times
"knew": 58 times

Performance on "hurt" and "cut": ~80% success
Performance on "caught": ~96% success

Performance on "threw": ~49% success
Performance on "knew": ~49% success

Better performance for less frequent verbs.

Yang (2002): Irregular Rules

Evidence from CHILDES database

How often children encounter certain verbs in a given corpus:
"hurt", "cut": 20 times
"caught": 36 times "threw": 31 times
"knew": 58 times

Performance on "hurt" and "cut": ~80% success
Performance on "caught": ~96% success

Performance on "threw": ~49% success
Performance on "knew": ~49% success

Different performance for equally frequent verbs.

Yang (2002): Irregular Rules

Evidence from CHILDES database

Irregular rule members:
"No Change" rule: hurt~hurt, cut~cut
hit, quit, split, slit, spit, bid, rid, forbid, spread, wed, let, set, upset, wet, shut, put, burst, cast, cost, thrust many! > 2500

"Change to 'aught'" rule: catch~caught
buy, bring, teach, think less, but very frequent verb forms > 600

"Vowel goes to 'oo'" rule: throw~threw, know~knew
draw, blow, fly, withdraw, slay less all around! < 100

Yang (2002): Irregular Rules

Evidence from CHILDES database

Irregular rule members:
"No Change" rule: hurt~hurt, cut~cut ~80% success
hit, quit, split, slit, spit, bid, rid, forbid, spread, wed, let, set, upset, wet, shut, put, burst, cast, cost, thrust many! > 2500

"Change to 'aught'" rule: catch~caught ~96% success
buy, bring, teach, think less, but very frequent verb forms > 600

"Vowel goes to 'oo'" rule: throw~threw, know~knew ~49% success
draw, blow, fly, withdraw, slay less all around! < 100

Summary: Support for Rules, No Words

Irregular past tense verb forms benefit if the child encounters many other verbs that use the same rule. The frequency of the rule influences the child's performance.

Example rule & performance:
"No Change" rule: hurt~hurt, cut~cut
children's success: ~80%
many other verbs in this class: *hit, quit, split, slit, spit, bid, rid, forbid, spread, wed, let, set, upset, wet, shut, put, burst, cast, cost, thrust*

Summary: Support for Rules, No Words

Logic of argument:
(1) children benefit from irregular rule's use
(2) this would not happen if children's minds don't have an irregular rule
Therefore, children's minds must have an irregular rule. So, irregular verbs are not just memorized individually. They have irregular rules the same way regular verbs use the regular rule.

Questions of Productivity: When do children figure out that they need a rule for certain groups of verbs?

Chomsky & Halle, 1968: "…existence of exceptions does not prevent the systematic formulation of those regularities that remain"

How does a child extract the regularity that's there?
Big question: How does a child know what's systematic/ productive?

Words To Rules?

Idea: The point of using rules for past tense forms would be that it's easier in some sense -- as opposed to simply storing each verb and its associated past tense individually.

look	looked		look	
kiss	kissed		kiss	
lurch	lurched	vs.	lurch	+ed
laugh	laughed		laugh	
dance	danced		dance	
harder			easier	

Words To Rules?

Idea: The point of using rules for past tense forms would be that it's easier in some sense -- as opposed to simply storing each verb and its associated past tense individually.

If a particular transformation (rule) occurs a lot (like +ed), it's said to be productive. Productive rules make sense to store because they're used for a lot of different verbs.

Question: What determines if a rule is productive? That is, how does a child decide that a rule is used enough to be worth storing?

What We Know From Children's Errors

The errors kids make with the past tense

I holded the rabbit

Most are **over-regularizations**: hold-holded (make up 10% of all irregular past tense forms: Marcus et al. 1992; Yang 2002)

Very rare are **over-irregularizations**: bring-brang (0.2% of irregular past tense forms: Xu & Pinker, 1995)

Cross-linguistically: most errors are **over-regularizations** or omissions of past tense morphology (Phillips 1995; Guasti 2002)

The point: "Children recognize and generalize productive rules while memorizing the restricted use of unproductive ones"

Some Definitions

Default: "when all else fails"
When more specific rules fail to apply, use this rule (which by definition is the most general).

English past tense:
+ed
kiss-kissed

Productive: "predictable" or "generalizable"
A rule automatically applies to a set of lexical items characterized by a certain context. It can extend to novel items that fit this context (though may not always)

Possible hypothesized rule:
If a verb is monosyllabic and ends in -ing, change to -ang

sing-sang,
spling-splang/splinged

A default rule is always productive, but a productive rule can exist without being the default. Neither kind of rule needs to be exception-less.

Productive Rules

Productive?

Yang (2005): Productivity of a rule depends on some kind of cost-benefit analysis for how many words follow the rule and how many words don't.

Specifically, the child keeps track of how many exceptions there are for a particular rule. If there are too many exceptions, it's easier to just not have a rule.

Rule: *ing --> *ang

Verbs that follow the rule: ring~rang, sing~sang, …

Verbs that don't follow the rule: sting~stung, bring~brought, …

Computational Complexity & Tolerance Principle

Idea: Cost-benefit analysis based on computational complexity

Empirical evidence points to time complexity as a sensible metric - how long does it take to access the right rule? (Morphological processing is oriented towards time efficiency.)

Question: What is the threshold for determining if a rule is productive or not?

We want some way a child could calculate this, some algorithm based on the time it takes to access the correct rule. This is what the Tolerance Principle is supposed to do.

The computational process of morphologically derived words: executed sequentially (Carmazza 1997; Levelt et al. 1999)

1) Word search (look up the word stem in the lexicon: *dance*)
2) Rule selection (find the right rule to use: *dance + ed*)
3) Rule application (apply the rule to get the derived form: *danced*)

Productivity Assessment/Tolerance Principle deals with this part

Serial Search

Rule selection: Lexical Search Theory (Rubenstein et al. 1970; Forster 1976)

Lexical processing involves serial search that is sensitive to the token frequencies of the words.

Idea: Rule selection also involves serial search, listed by token frequency.

Elsewhere Condition Serial Search (ECSS)
Rule: *ing-*ang

If word = *sting* then *stung* (freq 100)
Else if word = *swing* then *swung* (freq 80)
Else if word = *ding* then *dinged* (freq 10)
Else if word = *cling* then *clung* (freq 8)
Else Apply *ing --> *ang

Serial Search

Rule selection: Lexical Search Theory (Rubenstein et al. 1970; Forster 1976)

Lexical processing involves serial search that is sensitive to the token frequencies of the words.

Idea: Rule selection also involves serial search, listed by token frequency.

Elsewhere Condition Serial Search (ECSS)
Rule: *ing-*ang

If word = *sting* then *stung* (freq 100)
Else if word = *swing* then *swung* (freq 80)
Else if word = *ding* then *dinged* (freq 10)
Else if word = *cling* then *clung* (freq 8)
Else Apply *ing --> *ang

swing?
--> swung

Time units: 2

Serial Search

Rule selection: Lexical Search Theory (Rubenstein et al. 1970; Forster 1976)

Lexical processing involves serial search that is sensitive to the token frequencies of the words.

Idea: Rule selection also involves serial search, listed by token frequency.

Elsewhere Condition Serial Search (ECSS)
Rule: *ing-*ang

If word = *sting* then *stung* (freq 100)
Else if word = *swing* then *swung* (freq 80)
Else if word = *ding* then *dinged* (freq 10)
Else if word = *cling* then *clung* (freq 8)
Else Apply *ing --> *ang

ring?
--> rang

Time units: 5+rule application

When to Bother With a Rule?

Trade off: Storing individual exceptions + rules vs. exceptions only

If there are few enough exceptions, then it's more efficient to store the exceptions and then have the rule as an "elsewhere" option.

Elsewhere Condition Serial Search (ECSS)
Rule: *ing-*ang

If word = *sting* then *stung* (freq 100)
Else if word = *swing* then *swung* (freq 80)
Else if word = *ding* then *dinged* (freq 10)
Else if word = *cling* then *clung* (freq 8)
Else Apply *ing --> *ang

If there are too many exceptions, then it's more efficient to store the exceptions alone and not have a rule.

Elsewhere Condition Serial Search (ECSS)
*Rule: *ake-*ade (make-made)*

If word = *bake* then *baked* (freq 600)
Else if word = *take* then *took* (freq 400)
Else if word = *shake* then *shook* (freq 200)
Else if word = *rake* then *raked* (freq 100)
…
Else if word = *slake* then *slaked* (freq 1)

Tolerance Principle In Action

Tolerance Principle: How many is too many exceptions?

N = number of items that fit the context the rule applies to
M = number of items that are exceptions to the rule

T(M, N) = time it takes to find out if a rule applies to a given word when there are M exceptions and N items that have the rule's context

T(N,N) = time it takes to find out if a rule applies to a given word when all words are stored as exceptions

Tolerance Principle In Action

Tolerance Principle: How many is too many exceptions?

When it takes longer if exceptions are stored along with a rule (T(M,N)) than it does if all words are stored as exceptions (T(N,N)), don't bother storing the rule. The rule is not productive.

If T(N,N) < T(M,N), rule is not productive. Don't store rule.
(This happens when M ≈ N/ln N)

Yang (2005): Productivity

Tolerance Principle: Main Idea

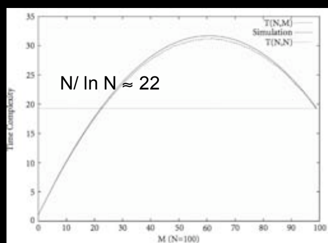
If the child knows a rule whose context fits N words, the child should only store the rule explicitly if the number of exceptions M is less than N/ln N. Otherwise, the child should store the words the rule applies to on an individual basis.

M <= N/ln N

M >= N/ln N

How long access takes on average

Tolerance Principle in Action, N = 100, N/ln N = 22



If more than about 22 words are exceptions, then it's faster to just store all the words as exceptions (because 78 words have to wait 22 time units before the rule can be applied).

Predictions for English Past Tense

Default +ed rule can only be productive if it applies to the vast majority of types it could apply to. There are 150 irregular verbs (M=150), so there need to be at least 1000 regular verbs (N=1000) for it to be faster to have a rule + exceptions. This seems to be true (we have a lot of regular verbs).

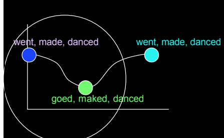
Tolerance Principle for children learning

1) Child identifies possible rule. (*ing --> *ang)
2) Child (unconsciously) checks current vocabulary with Tolerance Principle to see if it's better to store a rule + exceptions, or just exceptions.
3) Child repeats with each new word type encountered. (Productivity of rules can change.)

Predictions for English Past Tense

By the time the child has a productive rule (like +ed), the child should know a good deal more regular verbs than irregular verbs. This seems to be true (Marcus et al. 1992).

U-shaped development (in some children) - or at least the initial dip:



U-shape based solely on child's vocabulary input (how many exceptions they're exposed to)

1) Initially, irregular verbs learned first because they're frequent.
2) Only a few regular verbs required to posit +ed rule (20-30).
3) At this point, kids may have rule but it may not be productive because they haven't learned enough regulars. (Too many exceptions.) [initial stage]
4) Once they do see enough (M < N/ln N), then they use the rule productively. [dip of U-curve]

Predictions for English Plural Nouns

English plural nouns: Many regular nouns initially, few irregulars.

+s rule (goblin-goblins) becomes productive very quickly. No initial good performance with irregulars.

Should never see U-shaped curve in development - only an increase in performance. This seems to be true (Brown 1973, Falco & Yang 2005).

For another (compatible) take on why there should be no U-shaped curve for plurals based on the idea that type frequency matters, see Maslen et al. (2004)

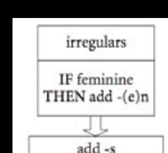
Predictions for German Plural Nouns

German plural nouns: many "irregular" regular rules
  Ex: +en for feminine nouns (Frau - Frauen)

M = 80 exceptions
Tolerance Principle predicts at least N where N/ln N >= 80 to have a productive rule. There must be N = 500 feminine nouns (and 420 that follow the +en rule). There are at least 3600.

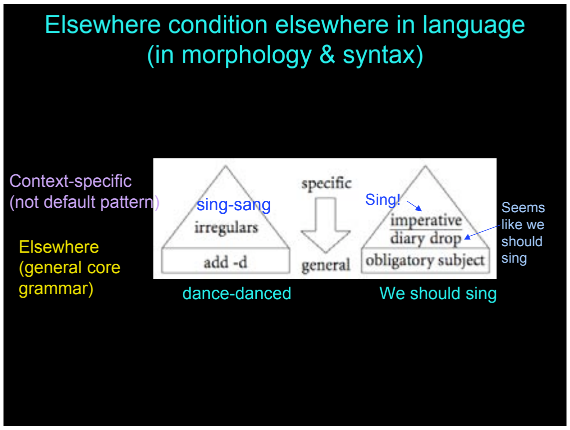
Therefore, this rule should be productive (and seems to be): Wiese 1996; Dressler 1999; Wunderlich 1999

Process for German Plural Formation



More context-specific rule (feminine)

Elsewhere condition rule


Elsewhere condition elsewhere in language
(in morphology & syntax)
Context-specific
(not default pattern)
Elsewhere
(general core
grammar)
sing-sang
irregulars
add -d
specific
general
Sing!
imperative
diary drop
obligatory subject
Seems
like we
should
sing
dance-danced
We should sing