Psych 215L: Language Acquisition

Lecture 18

Complex Systems

Complex Linguistic Systems

What is the generative system that creates the observed (structured) data of language (ex: metrical phonology)?

Observable data: stress contour EMphasis

Complex Linguistic Systems

What is the generative system that creates the observed (structured) data of language (ex: metrical phonology)?

Observable data: stress contour EMphasis

(S S) S
EM pha sis

(H L) H
EM pha sis

(H L L)
EM pha sis

(S S S)
EM pha sis

Why look at English stress?

Legate & Yang (2011)

- English stress is a volatile area theoretically (a hard system to capture because of all the exceptions); developmental work could shed light on the target state for adults (e.g., see also Pearl (2011), who worries about this)

- Some empirical data on children's stress knowledge is available

- Because of the need to capture both "core" and "exceptional" data, the system is very interesting from a developmental point of view

A quantitative description of productivity

Legate & Yang (2011)

Idea: Want to describe productive data more compactly by using abstract rules, and leave less productive data to be memorized

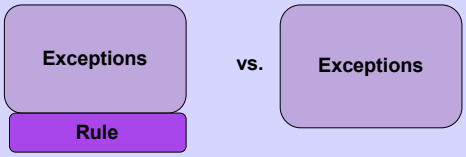
Elsewhere Condition, using a serial search procedure
Basic format: Does it match this exception, or this one, or this one? If not, go to the ELSEWHERE condition because it must follow the normal rule.

{order by type frequency} ← more frequent exceptions accessed faster
IF word is *elephant*, then…
IF word is *octopus*, then…
IF word is *teddy bear*, then…
…
ELSE {do something else}

A quantitative description of productivity

Legate & Yang (2011)

Logic: If a rule has too many exceptions, it takes a really long time to find the right thing to do. In fact, it may take longer than simply listing out each word explicitly (that is, treating everything as if it were an "exception" that had to be memorized).



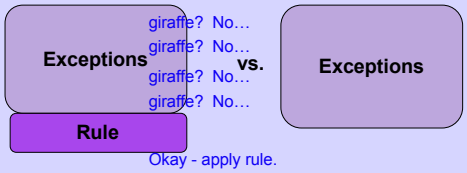
A quantitative description of productivity

Legate & Yang (2011)

Logic: If a rule has too many exceptions, it takes a really long time to find the right thing to do. In fact, it may take longer than simply listing out each word explicitly (that is, treating everything as if it were an "exception" that had to be memorized).

What to do for *giraffe*?

Time = # exceptions + rule application



giraffe? No…
giraffe? No…
giraffe? No…
giraffe? No…

Okay - apply rule.

A quantitative description of productivity

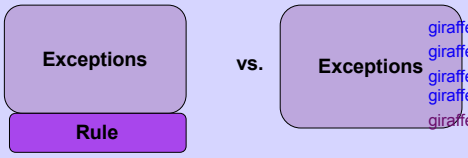
Legate & Yang (2011)

Logic: If a rule has too many exceptions, it takes a really long time to find the right thing to do. In fact, it may take longer than simply listing out each word explicitly (that is, treating everything as if it were an "exception" that had to be memorized).

What to do for *giraffe*?

Time = # exceptions + rule application

Time = # exceptions

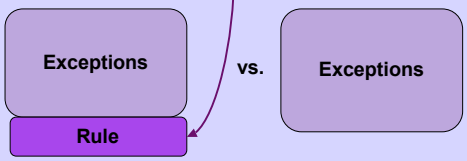


giraffe? No…
giraffe? No…
giraffe? No…
giraffe? No…
giraffe? Yes!

A quantitative description of productivity

Legate & Yang (2011)

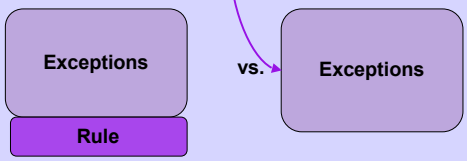
Suppose we have N lexical items to deal with, with m exceptions.
T(N, m) = time to wait to find what to do, on average, if rule is productive.
Normal items (N-m of them) will have to wait for m exceptions to be looked at before the rule can be applied. Exceptional items processed faster (<m).



A quantitative description of productivity

Legate & Yang (2011)

Suppose we have N lexical items to deal with, with m exceptions.
T(N, N) = time to wait to find what to do, on average, if rule is not productive.
All items (N of them) will be searched for (and eventually found) in the list of "exceptions".



A quantitative description of productivity

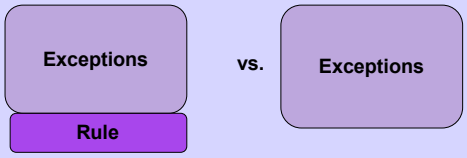
Legate & Yang (2011)

Tolerance Principle:
A rule will be represented as productive if T(N,m) < T(N,N).

Yang (2005): This happens when m < N / ln N

Productive representation

Non-productive representation



English Past Tense Example

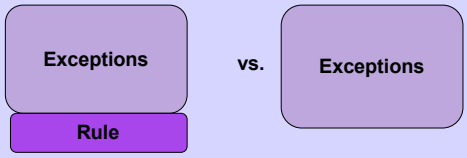
Legate & Yang (2011)

Regular past tense: +ed kissed hugged missed tugged
But also many irregulars: catch-caught, go-went, drink-drank, …

How many regulars needed to make +ed a productive rule?

Productive representation

Non-productive representation



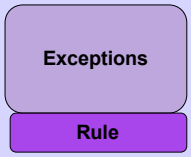
English Past Tense Example

Legate & Yang (2011)

Regular past tense: +ed kissed hugged missed tugged
But also many irregulars: catch-caught, go-went, drink-drank, …

If irregulars (m) < N (total) / ln N (total), then it's productive.

Productive representation



vs.

Non-productive representation

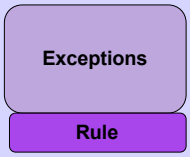
English Past Tense Example

Legate & Yang (2011)

Regular past tense: +ed kissed hugged missed tugged
But also many irregulars: catch-caught, go-went, drink-drank, …

If there are 120 irregulars, then need 800 total (680 regulars) for that to be true: 120 =~ 800 / ln 800. Turns out there are way more than this.

Productive representation



vs.

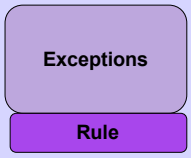
Non-productive representation

Tolerance Principle Success

Legate & Yang (2011)

Also works to explain why there is German feminine plural (+n) productivity, and why there is irregular English unproductivity (e.g. drink-drank, sink-sank, …), despite there being several verbs following irregular patterns.

Productive representation



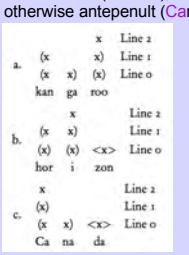
vs.

Non-productive representation

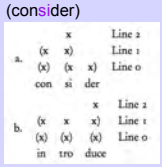
English Stress: Nouns vs. Verbs

Legate & Yang (2011), based on Halle & Vergnaud (1987) & Hayes (1995)

Nouns: main stress final if VV (kangaroo), otherwise next to last if VV/VC (horizon), otherwise antepenult (Canada)



Verbs: main stress final if VV/VC (introduce), otherwise next to last (consider)

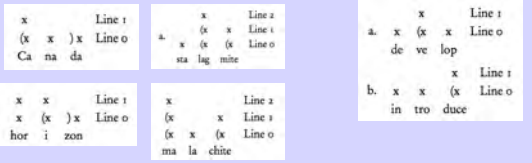


English Stress: Nouns vs. Verbs

Legate & Yang (2011), based on Halle (1998)

Nouns: main stress antepenult if two light syllables follow (Canada), otherwise penult if VV/VC penult (horizon, stalagmite), otherwise final (malachite)

Verbs: main stress final if VV/VC (introduce), otherwise penult (develop)



Learning English Stress

Legate & Yang (2011): Learner Assumptions

- phonemes are known (experimental evidence that this happens by age 1: Werker & Tees 1983, Kuhl et al. 1992)

- phonotactic rules are in place that allow for native-like syllabification (experimental evidence this happens by 9 months: Jusczyk et al. 1992)

- word segmentation is online (experimental evidence this could happen as early as 7.5 months: Jusczyk & Aslin 1995)

- detection of main stress in a word is online (evidence that basic stress patterns (foot-headed-left (trochaic) vs. foot-headed-right (iambic)) are known by 7.5 months: Jusczyk, Cutler & Redanz 1993, Jusczyk, Houston & Newsome 1999.)

Learning English Stress

Legate & Yang (2011): Not cue-based learning (Dresher 1999)

- cue-based approach: metrical parameters are set in an ordered sequence, based on cues observed in the linguistic input

- L & Y argue: "A major motivation for learning as a sequence of decisions is to uphold the idealization of the child as a deterministic learner. For instance, suppose the child has not yet determined the quantity sensitivity of the language: if he proceeds to the stress placement parameters in a quantity sensitive language such as English, he might as well need to retreat from these parameters. But this idealization of deterministic learner is both empirically problematic and formally unnecessary…More important, and more general to the theory of language and language learning, is the issue of balancing generalizations with exceptions."

Learning English Stress

Legate & Yang (2011): How Tolerance Principle learning works

- "we envision the learner experimenting and evaluating the core metrical hypotheses in an incremental fashion as he processes linguistic data, and the learner chooses the grammar most highly valued with respect to the present data…"

a. If a grammar fails to reach productivity as prescribed by the Tolerance Principle (4), it is rejected.
b. If there are multiple grammars meeting the Tolerance threshold, the learner selects the one with fewest exceptions (i.e., most productive).
c. If no grammar is productive, then the stress patterns of words are memorized as a lexicalized list.[13]

"Each grammar G_i, then, can be associated with a tuple (N_i, m_i), the number of words (N_i) it could apply to, and the number of words that contradict it (m_i)."

Learning English Stress

Legate & Yang (2011): Input

4.5 million tokens (26,700 types) pulled from all of CHILDES (2000)
-only use the 20,000 types that are either nouns or verbs

"Since nouns and verbs have somewhat different stress patterns, considering them together will pose a realistic test for any model that seeks systematic regularity amidst a heterogeneous mix of patterns."

Assume morphological knowledge:
"Based on the consistent developmental evidence that the inflectional morphology is acquired relatively early–in some languages very early–we assume that the learner is capable of parsing words into morphological structures and considering their roles in the acquisition of stress."
- necessary if dividing nouns from verbs, but unclear children have this until about 3 years old in English…

Learning English Stress

Legate & Yang (2011): What's being learned

"For the present paper, we only consider the placement of the main stress. Since the pronunciation dictionary marks primary, secondary, as well as no stress, we mark the former as 1 and collapse the latter two as 0. For instance, the word animals will be represented syllabically as LLC with the stress contour of 100."
- "introduce" would be marked as 001
- "stalagmite" would be marked as 010

"For simplicity, we only consider two specific points of stress development, one designed to capture the child's stress system under a very small vocabulary and the other when the child has already learned enough words to potentially match the target state."

Learning English Stress

Legate & Yang (2011): The input distribution from an extrapolated early vocabulary

contour	counts
1	287
10	107
100	13
01	7
010	3
1000	3

Table 1: Stress patterns for words with frequency ≥ 1 in 10,000.

"The distribution in Table 1 is clearly consistent with a quantity insensitive trochaic [ft-hd-left] system [assuming feet of size 2 syllables]. A total of 402 words can tolerate 402/ ln 402 = 67 exceptions where in fact there are 26 [01, 010, 100, 1000] . Interestingly, children learning English and similar languages go through an initial stage, which terminates at about 2;0, during which the child is limited to a maximum bisyllabic template with the primary stress falling on the first [syllable]…"

Learning English Stress

Legate & Yang (2011): The input distribution for an extrapolated later vocabulary

Considering words that occur at least once per million:

4047 nouns, 2402 verbs [some overlap, so only 5763 distinct types total]

How many follow a quantity-insensitive, trochaic grammar? 4533
This leaves 5763-4533 = 1230 exceptions.
How many exceptions can 5763 types tolerate? 5763/ln(5763) =~ 668.
This grammar is no longer productive.

Option? Try quantity-sensitive grammars instead.

Learning English Stress

Legate & Yang (2011): The morphology assumption plays out

"…we assume that the learner is capable of relating inflectional forms of verbs to their stem forms, but is incapable of parsing derivational forms into decomposable pieces (words such as *growth* and *government* will be treated as morphologically simplex)…"

"…we assume that the learner has correctly learned that inflectional suffixes do not trigger stress shift - a task easily accomplished, again, by the use of the productivity model: there are no exceptions to the lack of stress shift with inflectional morphology. In other words, the child treats all inflectional forms of walk (i.e., *walk*, *walks*, *walked* and *walking*) as *walk* for the purposes of stress acquisition…"

[Worth thinking about?: what age the child is able to recognize this and when stress acquisition happens]

Learning English Stress

Legate & Yang (2011): Evaluating adult target grammars

lex	stem	HV_{87}	H_{98}
-	-	no	no
-	+	no	no
+	-	no	yes[a]
+	+	no	yes[b]

Table 2: Evaluation of stress grammars for words with frequency ≥ 1 per million. a. with 515 exceptions. b. with 355 exceptions.

+lex = distinguish nouns from verbs (perhaps using semantic knowledge?)
+stem = inflectional decomposition of verbs (walking = walk + ing)

HV_{87} = Halle & Vergnaud 1987, Hayes 1995 grammar
H_{98} = Halle 1998 grammar

Basic results: H_{98} grammar is productive while HV_{87} is not, but has more exceptions if inflectional decomposition isn't allowed

Learning English Stress

Legate & Yang (2011): Other useful results

Why the diminuitive "ee" doesn't get stressed (kitty, doggie/doggy):
If lexical decomposition is allowed (kitty = kit + diminuitive "ee"), child can decide on productive rule that never stresses the diminuitive, based on experience with diminuitive forms in the input.

Learning when morphological affixes cause main stress shifting:
(document vs. documentary, tone vs. tonic, ridicule vs. ridiculous)

suffix	shifting	N	m	valid
-ment	no	201	0	yes
-ary	no	41	8	yes[a]

Table 3: The validity of stress preservation for certain derivational suffixes that are factually stress preserving. a. $8 < 41/\ln 41 = 11$.

suffix	shifting	N	m	valid
-ic	no	135	120	no
-ous	no	90	30	no[a]

Table 4: The validity of stress preservation for certain derivational suffixes that are factually stress shifting. a. $30 > 90/\ln 90 = 20$.