

# Past-Tense Learning

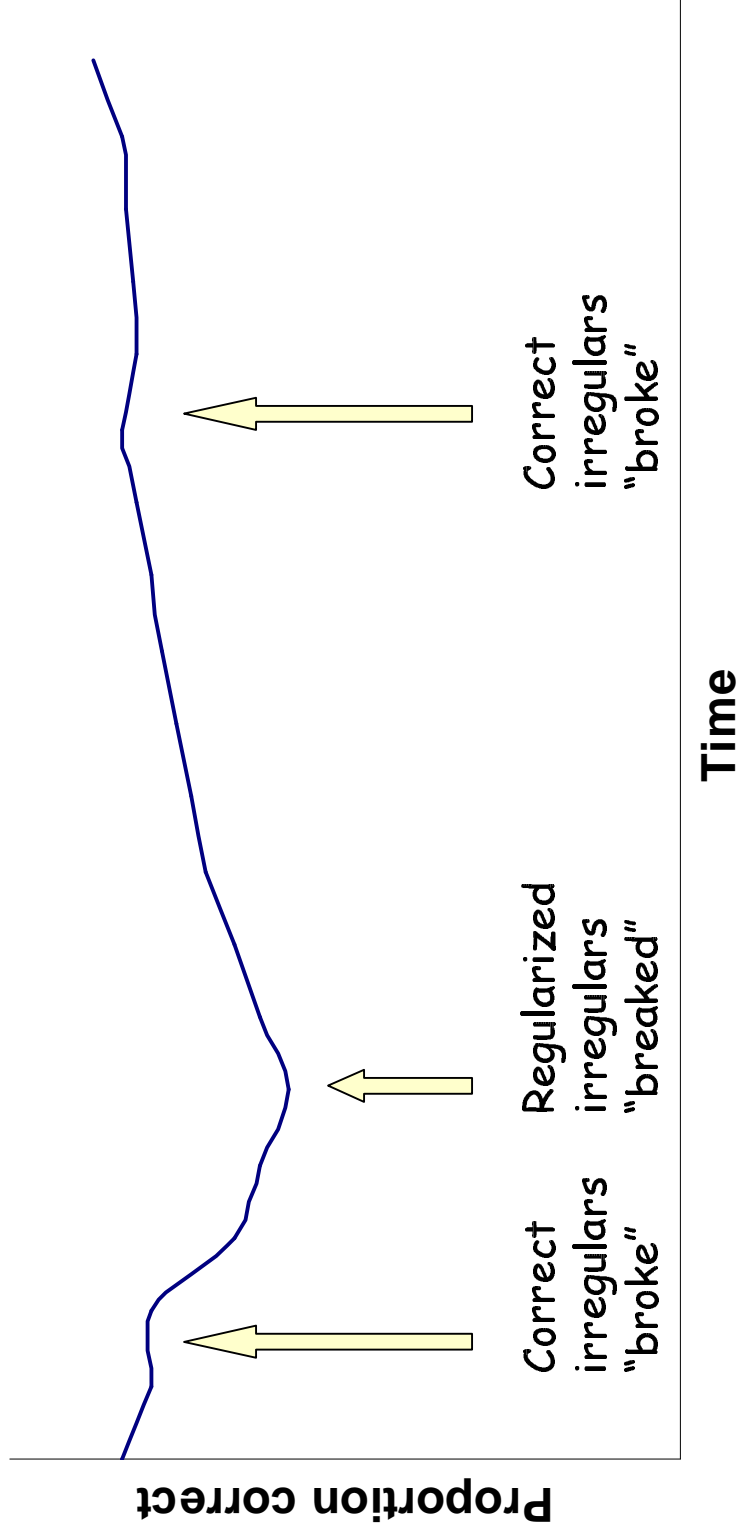
March 19, 2002

# Past Tense Learning

- Rumelhart & McClelland (1986)'s model of past-tense
- Rules or no-rules debate: generalization can be produced by a system that does not have built-in rules

# U-Shape Learning

Overregularization



Overregularization = irregular correct / (irregular correct + irregular regularized)

# Facts to Be Explained

- U-shape learning
- Regular rule is a default rule in English
- Input distribution should obey the frequency distribution of verbs
  - Most verbs are regular (high type frequency)
  - Most frequently used verbs are irregular (high token frequency)
- Same mechanism should explain different inflections (e.g., German plural: low type frequency, low token frequency)
- Learning does not need feedback on errors
- No critical vocabulary mass? (linear increase in vocabulary size; nonlinear relation between vocabulary size and # of irregular verbs inflected for past tense)
- The nature of the distinction between regulars and irregulars

# Dual-Representation Theories

- Dual-representation theories

- Store instances
- Form rule

The rule is applied only if an instance cannot be retrieved. (blocking)

Overregularization is due to unsuccessful retrievals.

Problems: how is the rule learned, what is the justification for blocking

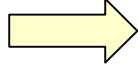
# Single-Representation Theories

- Neural-network based
  - When the vocabulary is small: store instances in the net
  - When the vocabulary is large: network learns regularization
  - The U-shape is due to the time needed to integrate regularizations with exceptions
  - Problems: unconfirmed assumptions about
    - vocabulary growth
    - input distribution
    - feedback
    - passive production
- } (leading to input-uptake distinction),

# The ACT-R (4.0) Past-Tense Model

# The Distinction Regular-- Irregular

- Regular past tense is easier to memorize, but more costly to produce:
  - Regular past tense is slightly longer than an irregular past tense
  - Regular past tense may add a phonetic load to the production of the word
- Irregular past tense is harder to memorize, but cheaper to produce



High frequency → irregular

Low frequency → regular



# U-shape Learning

- As in the dual-representation accounts, it results from a temporary imbalance between retrieving the examples and using the rule -- the system has not seen enough examples to act correctly
- ACT-R offers a mechanism for creating the rule (production compilation) and an explanation for blocking (production cost)
- No feedback is necessary

# The Model

- The model either “perceives” past-tense examples or has to produce them
- Three strategies of producing the past tense:
  1. Retrieval: retrieve it from memory
  2. Analogy: generate the past tense by using suffixes observed in other past-tense examples
  3. Zero strategy: use the stem (break)
- First strategy works only if there are examples in memory
- Last two strategies are penalized by higher effort

# The Model: Regularization

- No production rule for regularization initially
- Production compilation for the analogy productions leads to
  - a production that adds the suffix "-ed" (regularization)
  - a production that adds a blank suffix
- All past tense chunks (correct or generated by the model) are kept in declarative memory
- The model quickly learns to use the regularization production (lower cost than analogy)

## Performance: U-Shape

- It takes a while for the “-ed” rule to be learned and preferred
- No U-shape for regulars
- U-shape for irregulars is due to a change in the type of errors: zero rule is replaced with “-ed” rule
- The onset of regularization is predicted by a sudden increase of marking past tense for regulars (as in children)

# Performance

- Individual differences: children may have different input distributions (e.g., fewer examples may be offered);
  - high frequency leads to little overregularization; there is a negative correlation frequency - overregularization
- Gradual increase in vocabulary size in terms of both regulars and irregulars
- If one changes the input distribution to reflect German plural, the onset of regularization is later
- Critical mass: to form the new rule, it needs a reasonably sized and activated vocabulary

# No-Input Model

- Absence of correct past-tense examples from environment
- Additional production to generate a new past tense (high cost):
  - add a random suffix, or
  - modify the stem into something new
- The model prefers irregular forms for high frequency verbs and regular forms for low frequency verbs
- U-shape learning with onset of regularization corresponding to learning of new rules

# Questions

I noticed that the last term of Figure 10, page 45, is "squooosh." I had never heard of the word, and tried looking it up in the Oxford English Dictionary. Given the word was last in the list sorted by frequency, I wasn't particularly surprised that I wasn't familiar with the word. However, I was surprised that the query "squooosh" returned the message "there are no results."

My question of the week is, what does "[squooosh](#)" mean and how should I interpret the results returned by that word?

(Peter Hu)



One important area of language learning is learning a second language. Without having any data on that topic, I would intuitively guess that it would go through different processes than children learning their first language. First, in most foreign language schools, students are taught about the rules (e.g., blocking) explicitly before they have a chance to induce it from exposure. Secondly, feedback is not only present, but frequent and carries a larger benefit/cost (e.g., in terms of grades and penalties on exams and assignments.) Would these environmental constraints make 2nd language learning 1) follow a different learning process than the one described here? 2) following a different learning curve than the U-shape?

(Junlei Li)

I think that this model minimizes an important alternative phonetic association strategy that, given a novel word, would retrieve a word that sounds similar and use it as a template for generating the past tense for the novel example. Has this been considered or implemented and did it change the model's behavior? I predict that it would make the model perform even closer to human performance while making the model more plausible or complete.

(Mark McGregored)

I just realized where my confusion came from on the original experiment...it was actually done both ways. Oops. I'm now thinking of doing both versions of the experiment, but I'm not sure if it might be too much. Do you think it would be better to just do the one where they report the color (of either the word, or a box), or would I be able to do both? I'd like to do both if I can, but I'm not sure if I'll be able to do both successfully.

(Andy Herrman)

I assume that there will be three different LEARNED-REGULAR-RULES for each of the following examples:

- 1) want -> wanted, help -> helped
- 2) cry -> cried, copy -> copied
- 3) stop -> stopped, plan -> planned

The reason that this ACT-R model could learn these different rules is because there are following chunks previously learned and stored in the declarative memory.

- 1) (PAST-TENSE-GOAL101 isa PAST of WANT stem WANT suffix ED)
- 2) (PAST-TENSE-GOAL102 isa PAST of CRY stem CR suffix IED)
- 3) (PAST-TENSE-GOAL103 isa PAST of STOP stem STOP suffix PED)

These chunks are critical for the model to gain new production rules. However, how these chunks could be learned is not clear, because the perceiving component was not modeled. Without the ability to learn new ways to decompose (current-tense, past-tense) into (current-tense, stem, suffix), there is no way to really gain new production rules. How do people learn this?

Again, I assume that, eventually there would be a model that could learn new ways of decomposition, which are represented by different newly learned rules.

Recursively, these rules will be learned by combining other rules. Where are those rules from? Are there inborn primitive rules? Is there any other production-rule learning mechanism in addition to production compilation?

(Mon-Chu Chen)

Would you please explain a little more about why the "Broke" model had activation noise decline for individual chunks, and what effect this had on the models performance? How was this implemented?

(Philip I. Pavlik)

Will the system learn any other rules for generalization other than the -ed rule?? In the German plural experiment, the system was able to learn a rule for regulars, even though these words have a type frequency below 10% and token frequency below 5%. Are there any other plural "rules" in English that occur with frequencies comparable to the regulars in the German plurals?, and if so are these rules learned as well??

(Jonathan Giloni)

In this paper, months of learning are simulated. How was this accomplished in ACT-R? Did running this model actually involve the firing of millions of productions, or are there modeling shortcuts used for these kinds of long running simulations?

(Adam Goode)

i am wondering which parts of the model the theory is claiming are innate and which are learned? that is, humans are said to possess some sort of specialized mechanisms for dealing with language but the model in this paper doesn't seem to use any specialized mechanisms besides some existing production rules. does that mean those initial production rules are what is claimed to be innate? how is language processing a distinct aspect of cognition?

(Eli Silk)



Merriam-Webster's  
**COLLEGIATE® DICTIONARY**

Merriam-Webster

Collegiate® Dictionary Collegiate® Thesaurus Unabridged

One entry found for **squoosh**. Main Entry:

**squoosh**

Pronunciation: 'skwush, 'skwüsh

Function: *verb*

Etymology: by alteration

Date: 1942

: **SQUASH**