Reading:

Steven Pinker’s, Words and Rules, Chapter 4
Does the theory of words and rules explain the dichotomy between regular and irregular verbs?

Maybe a rule is not necessary to explain the past tense.

Maybe children simply analogize from verbs they already know (e.g., from correct forms like folded, molded, scolded to over-regularization’s like holded).

All-rules versus all-memory approach.

Children’s performance gets better as they get older. With inflectional morphology they get worse before getting better. This is what child psychologists call **U-shaped development**.

**Stage 1** children use only a small number of verbs in the past tense with very few errors.

**Stage 2** after a certain amount of time, the error rate appears to increase significantly; children add regular past tense suffix *-ed* to irregular verb stems even with verbs whose past tense forms they had previously mastered.

**Stage 3** the error rate slowly decreases, as the child gets older, until almost no errors are made.
In the 1980s, Rumelhart and McClelland promoted multilayer feed-forward perceptron as basis for cognitive modeling.

Parallel Distributed Processing (PDP for short).

Their aim was to simulate three-stage performance in the acquisition of the past tense.

Not a full-blown language processor that learns past tense from full sentences heard in everyday experience.

Model is presented with pairs of inputs: (a) the phonological structure of the stem and (b) the phonological structure of the correct past tense.

Model is tested by giving it the stem and examining what it generates as the corresponding past tense form.
Their model was pretty **radical**: no lexicon of **words**, no **rules**.

**Two-level fully-connected feed-forward** perceptron network.

And they **didn’t even use hidden units**, at least, not at first.

**Input**: verb’s base form, e.g., [dans], [sInk]

**Output**: past tense form, i.e., [danst], [sank]
The design of the input and output was crucial.

R&M followed Chomsky and Halle, and used features.

Triples of sets of features, known as Wickelfeatures.

They cover three phonemes in a row.

- St-Hv-St  stop+high vowel+stop
- N-St-[  nasal+stop+word-final

Both input and output were labeled with 460 different such triples; $460 \times 460 = 211,600$ connections (and weights).

Initially, these connections are all set to 0; learning involves modification of the strengths of these interconnections.

They trained model with 420 input/output pairs.
### Model Structure

**Input units**
(sounds in verb stem)

- St-Hv-St
- Bv-N-
- [-Cons-Cons]
- ...
- N-St-

**Output units**
(sounds in past tense form)

- St-Hv-St
- Bv-N-
- [-Cons-Cons]
- ...
- N-St-
Connectionist Model of English Past Tense

Active input pattern and corresponding output (indicated with + and −).

Matrix of weights indicating strength of connections from each input unit to each output unit.
After the 84,000 training iterations, the settled network worked well for almost all the 420 verbs in the training set. Performed adequately for separate test set of 86 other verbs. 3/4 of regular verb stems were assigned the correct past tense. Most irregular verbs stems were assigned overgeneralized regular past tense forms (e.g., dugged, catched).

Childlike Behavior

- **U-shaped learning**: after a period of outputting gave correctly, the network shifted to the incorrect gived.
- Was reluctant to stick -ed on a stem ending in [t] or [d].
- Made lots of childlike errors, e.g., cling/clang, sip/sept.
- In other words, R&M appears to capture stem-stem similarity.
R&M’s empirical observations

- children learn common verbs first, and rarer verbs later
- they tend to learn irregular verbs before regular ones
- children’s vocabulary grows very quickly all of a sudden, a few months after they start learning words, i.e. at some point they get a huge spurt of regular verbs.

R&M’s network training

- They first trained it on just 10 verbs, all at once, 8 irregular.
- And then trained it on 410 verbs, all at once, 80% regular
- Error rates increase dramatically at the start of the second training phase, before recovering gradually.
- Model started to make errors such as *breaked*.
U-shaped Learning Curve

Percent Features Correct

Trials

Regular

Irregular
U-shaped Learning Curve

The graph shows two learning curves: one for Regular and another for Irregular tasks. The Y-axis represents the percent features correct, ranging from 0.5 to 1.0. The X-axis represents the trials, ranging from 0 to 200. The Irregular curve starts higher and drops sharply before rising again, indicating a U-shaped learning curve.
Q₁: Do parents start at some point using more regular verbs when talking to their children?
A₁: Data from spontaneous conversations involving children shows no evidence for this.

Q₂: Is there a vocabulary spurt and thus a richer mixture of regular verbs when children begin to over-apply -ed?
A₂: Children’s vocabularies spurt the mid-to-late ones versus mid-to-late twos (when children start to make over-regularization errors new regular verbs are actually coming in more slowly than they were previously).

Q₃: What if we change the network’s training scheme?
A₃: Training regime is fragile. Language learning is a fairly robust process (e.g., children exhibit very similar learning curves, even with vastly different patterns of data to learn from).
Problem 1
R&M’s model only produces past-tense forms; you cannot turn the arrows around and get the model to run backwards and recognize past-tense forms. Obviously, people do both!

Problem 2
The model computes every detail of the pronunciation of the past-tense form. Many details common in other parts of the language system. Should they be duplicated in different networks?

Problem 3
Model has no notion of words, it relies exclusively on sounds to compute its past-tense form; can’t tell difference between words that sound alike (e.g., break-broke, brake-braked).
Main Problem

How do you represent an entity made of parts in a fixed arrangement such as a word? Units can only be on or off, you can’t refer to them with symbols. R&M use Wickelphones as a solution.

```
Verb stem
  prefix
    out-

Verb stem
  onset
    cons
      s
      t
      r
  rime
    nucleus
      vowel
      i
    coda
      p
```
R&M’s model illustrates the no rules, all memory extreme:

- it was sufficiently explicit to make testable predictions;
- researchers did experiments which appeared to conflict with those predictions;
- criticism led to the design of revised experiments;
- model also changed to fix flaws (e.g., different input representation, addition of hidden layer, tree-like behavior, rule-like mechanism).
Hypothesis A
Regular past tense forms are formed by a rule which is blocked for irregular verbs. Blocking principle is innate.

Hypothesis B
There are no rules, only a general associative mechanism for recognizing patterns; reason by analogy.

Hypothesis C
Regular past tense forms are formed by a rule which is blocked for irregular verbs. Blocking principle is innate. Memory is assumed to be associative a bit like R&M pattern associator. Words are linked to other words via similarity. Families of irregular verbs are easier to store and retrieve, since these verbs repeatedly strengthen their shared associations.
Children’s speech errors come with incomplete learning process.
They try to master irregular forms like *bled* and *sang*.
Learning is impossible without innately organized circuitry to do the learning.

**Charles Darwin**

Human language is an *instinctive tendency* to acquire an art. It certainly is not a true instinct for every language has to be learned. It differs, however, widely from all ordinary arts, for man has an instinctive tendency to speak, as we see in the babble of our your children; while no child has an instinctive tendency to brew, bake, or write.
Simple learning model shows the characteristics of young children learning the past tense!

Generates U-shaped learning curve for irregular verbs and exhibits tendency to overgeneralize similar to young children.

Makes empirical predictions that can be tested.

Manipulates actual data and can simulate rather than describe behavior; specific representations (e.g., Wickelfeatures).

Is connectionism the right approach to learning? The jury is still out there, at the time it revised understanding of how linguistic information is acquired and applied.

Next lecture: speech segmentation.