

# Informatics 1 Cognitive Science

## Lecture 4: Language Acquisition

---

Frank Keller

23 January 2024

School of Informatics  
University of Edinburgh  
[keller@inf.ed.ac.uk](mailto:keller@inf.ed.ac.uk)

Slide credits: Frank Mollica, Chris Lucas, Mirella Lapata

The Past Tense in English

Evidence from Language Acquisition

The Knowledge Acquisition Problem

Preview: Neural Networks

# Recap

Human language involves two different kinds of “mental tissue”:

- a finite lexicon of words, stored in and retrieved from memory;
- a finite grammar of rules (productive, abstract, combinatorial);
- these two mechanisms produce an infinite set of sentences.

Examples in context-free grammar notation:

- **words:** Det  $\rightarrow$  the; N  $\rightarrow$  dog
- **rules:** NP  $\rightarrow$  Det N; VP  $\rightarrow$  V NP

## Evidence for Words & Rules

How can we test the Words & Rules hypothesis? How do we figure out if speakers really store words and apply rules in their heads?

## Evidence for Words & Rules

How can we test the Words & Rules hypothesis? How do we figure out if speakers really store words and apply rules in their heads?

Various ways to get evidence:

- look at large amounts of text or speech (corpus data)
- study how humans process language in real time (eye-tracking, brain imaging)
- looks at what sort of errors speakers make
- track language acquisition in children

## Evidence for Words & Rules

How can we test the Words & Rules hypothesis? How do we figure out if speakers really store words and apply rules in their heads?

Various ways to get evidence:

- look at large amounts of text or speech (corpus data)
- study how humans process language in real time (eye-tracking, brain imaging)
- looks at what sort of errors speakers make
- track language acquisition in children

The English past tense can be used as a “fruit fly” (exemplar) to study and model language acquisition.

## The Past Tense in English

---

## Regular Verbs

- Past tense: just add *-ed* to the end of present tense form.
- Set of regular verbs is open-ended (probably tens of thousands in the mental lexicon of an educated adult).
- New regular verbs enter English every year.

jog →

walk →

play →

kiss →



## Regular Verbs

- Past tense: just add *-ed* to the end of present tense form.
- Set of regular verbs is open-ended (probably tens of thousands in the mental lexicon of an educated adult).
- New regular verbs enter English every year.

jog → jogged

walk → walked

play → played

kiss → kissed

## Regular Verbs

- Past tense: just add *-ed* to the end of present tense form.
- Set of regular verbs is open-ended (probably tens of thousands in the mental lexicon of an educated adult).
- New regular verbs enter English every year.

jog → jogged

walk → walked

play → played

kiss → kissed

spam →

mosh →

pluto →

grok →

## Regular Verbs

- Past tense: just add *-ed* to the end of present tense form.
- Set of regular verbs is open-ended (probably tens of thousands in the mental lexicon of an educated adult).
- New regular verbs enter English every year.

jog → jogged

walk → walked

play → played

kiss → kissed

spam → spammed

mosh → moshed

pluto → plutoed

grok → grokked

# Irregular Verbs

buy	→
hold	→
steal	→
go	→

- Some past tense forms don't just add *-ed* to the end of the present tense form.
- Irregular past tense inflection is chaotic and idiosyncratic.
- Irregular verbs are a closed list, of 150–180 members.
- There have been no recent additions (not since *sneak–snuck* arrived during the 19th century).

Linguists use an asterisk (\*) to mark things which are not part of the language, or, at least, which make native speakers uncomfortable, or are meaningless.

# Irregular Verbs

buy → \*bayed → bought

hold → \*holded → held

steal → \*stealed → stole

go → \*goed → went

- Some past tense forms don't just add *-ed* to the end of the present tense form.
- Irregular past tense inflection is chaotic and idiosyncratic.
- Irregular verbs are a closed list, of 150–180 members.
- There have been no recent additions (not since *sneak–snuck* arrived during the 19th century).

Linguists use an asterisk (\*) to mark things which are not part of the language, or, at least, which make native speakers uncomfortable, or are meaningless.

# Irregular Verbs

buy → \*bued → bought

hold → \*holded → held

steal → \*stealed → stole

go → \*goed → went

sing → sang

think → \*thank → thought

ring → rang

cling → \*clang → clung

- Some past tense forms don't just add *-ed* to the end of the present tense form.
- Irregular past tense inflection is chaotic and idiosyncratic.
- Irregular verbs are a closed list, of 150–180 members.
- There have been no recent additions (not since *sneak–snuck* arrived during the 19th century).

Linguists use an asterisk (\*) to mark things which are not part of the language, or, at least, which make native speakers uncomfortable, or are meaningless.

# Irregular Verbs

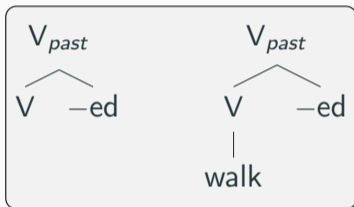
Time for a short quiz on Wooclap!



<https://app.wooclap.com/GXLBID>

# A Simple Theory of Regular and Irregular Verbs

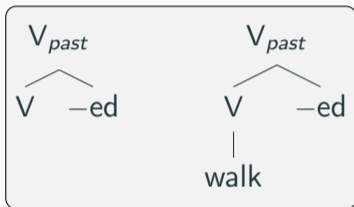
Regular past tense forms are created by a **rule**.





# A Simple Theory of Regular and Irregular Verbs

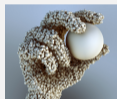
Regular past tense forms are created by a **rule**.



Irregular past tense forms are stored and retrieved as **words**.

sound:

*hold*

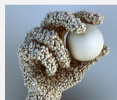


meaning:

part of speech:  $V$

sound:

*held*



meaning:

part of speech:  $V_{past}$

We have two independent mechanisms for past tense formation:

- **irregular** past tense forms stored as **words**;
- a productive **rule** for **regular** past tense forms;
- **Why don't they get in each other's way?**  
held vs. \*holded, stole vs. \*stealed

## Simple Answer

If a past tense verb form is stored in memory as a word, the rule is **blocked**. If no past tense form is stored, then the rule may be applied (e.g., *snarfed*, *mashed*, *ricked*).

# Evidence from Language Acquisition

---

## Evidence from Language Acquisition

Errors preschool children make in their spontaneous speech!

It was neat – you should have sawn it!

Doggie bat me [bit].

The cheerios got aten by the Marky.

I know how to do that. I truck myself [tricked].

This is the best place I ever sot [sat].

- Most children make at least some errors of this kind.
- Such errors persist well into their school-age years.
- Children have never heard adults using past tense forms like *swang* or *shuck*. Must be constructing these forms creatively, by *analogy* with verbs they already know.

## Overzealous Grammarians

Children don't just overgeneralize from regular past tense forms!

- they overuse the plural suffix *-s* (*mans, foots, tooths, mouses*)
- they overuse the third person sing suffix *-s* (*haves, do's, be's*)
- they overuse the comparative *-er* and superlative suffixes *-est* (*specialer, powerfulest, gooder*)
- they overuse the ordinal suffix *-th* on numerals (*oneth, twoth*)
- Children find regularity in the oddest places.

Parent: No booze in the house!

Child: What's a "boo"?

Child: "It did! It snew!"

[After being told it was going to snow.]

## U-Shaped Learning

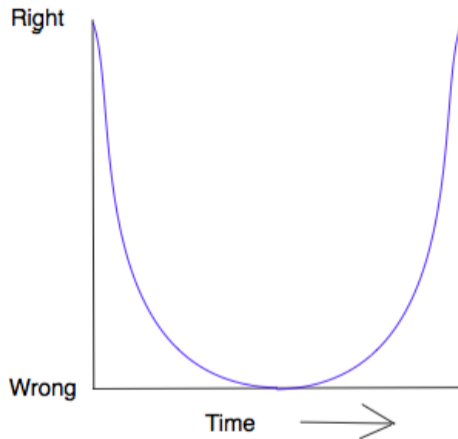
Children's performance gets better as they get older. With inflectional morphology they get worse before getting better. This is what child psychologists call **U-shaped development**.

**Stage 1** children produce both regular and irregular past tense forms 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.

## U-Shaped Learning



- U-shaped learning in early childhood cognitive development.
- Child uses *spoke*, then *spoked*, and later again *spoke*.

Time for a short quiz on Wooclap!



<https://app.wooclap.com/GXLBID>



# The Knowledge Acquisition Problem

---

## Where does (linguistic) knowledge come from?

The knowledge acquisition problem:

**Q<sub>1</sub>:** How do we acquire knowledge? Clearly, we are not born knowing everything! Else we wouldn't have to go to school!

**Q<sub>2</sub>:** But are we born knowing anything at all?

**Q<sub>3</sub>:** Is the mind completely blank or do we start with some rudimentary understanding of the world?

- These questions reflect the **nature-nurture** debate.
- Debate centers on relative contributions of biology and experience in determining any particular capacity.
- Nature: traits that are **genetically** or **biologically** determined.
- Nurture: traits that are **learned through experience** and interaction with the environment.

# Rationalism versus Empiricism

DESCARTES

*Discourse  
on Method*

*Meditations*

SPINOZA

*The Ethics*

LEIBNIZ

*The Monadology*

*Discourse  
on Metaphysics*

THE RATIONALISTS



Versus

LOCKE

*An Essay Concerning  
Human Understanding  
(Abridged)*

BERKELEY

*Principles of  
Human Knowledge*

*Three Dialogues*

HUME

*An Enquiry Concerning  
Human Understanding*

*Dialogues Concerning  
Natural Religion*

THE EMPIRICISTS



# Rationalism versus Empiricism

## Rationalism

- Intelligence arises from the **manipulation of symbols** by **rules**.
- Associated with Leibniz and Descartes, Noam Chomsky.
- The human mind has lots of **innate structure (nativism)**.
- Knowledge comes from logical deduction (i.e., “calculation”).

## Empiricism

- Intelligence arises from the mind **connecting together things** that were experienced together or that look alike.
- Associated with John Locke and David Hume.
- More recently with behaviorism (Ivan Pavlov, B.F. Skinner) and even more recently with neural networks.
- The human mind starts out as a **“blank slate”**. Knowledge comes by **generalizing from observations**.

# Evaluating the Knowledge Acquisition Debate



Grasping reflex



Sucking reflex



Step reflex

- Some forms of procedural knowledge are innate.
- Newborn infants come into the world with a variety of different skills; **reflexes** are important for survival.
- Many innate abilities are **domain-specific**, i.e., attuned to perform special operations only on a certain type of information.
- *Some* innate knowledge (or assumptions) necessary for inductive generalization; modern debate is about the nature and extent of that knowledge

# Evaluating the Knowledge Acquisition Debate

Time for a short quiz on Wooclap!



<https://app.wooclap.com/GXLBID>

## Preview: Neural Networks

---

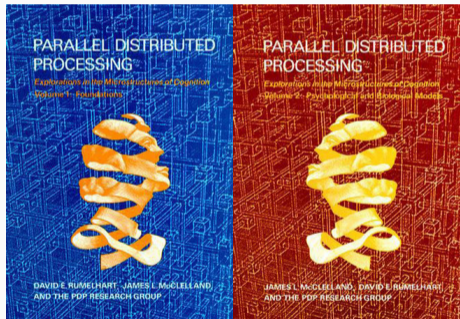
# Neural Networks

Neural networks (aka deep learning, connectionism) were invented in the 1980s, under the name of Parallel Distributed Processing (PDP):

- **Parallel:** simultaneous, independent, and simple computations.
- **Distributed:** information is represented across processing units.

Neural networks assume that everything is learned from data using a powerful, general learning algorithm: **radical empiricism**.

Crucially, neural networks impose **architectural constraints on theories**.





## Summary

- English past tense offers a simple, constrained way of analyzing and modeling language acquisition
- children's acquisition is in three stages: produce both regular and irregular forms; overgeneralization errors; errors decrease
- U-shaped curve in terms of accuracy
- wider problem: knowledge acquisition: innate vs. learned, nature vs. nurture, rationalism vs. empiricism
- Neural networks: a framework that provides algorithmic and representational constraints; radical empiricism