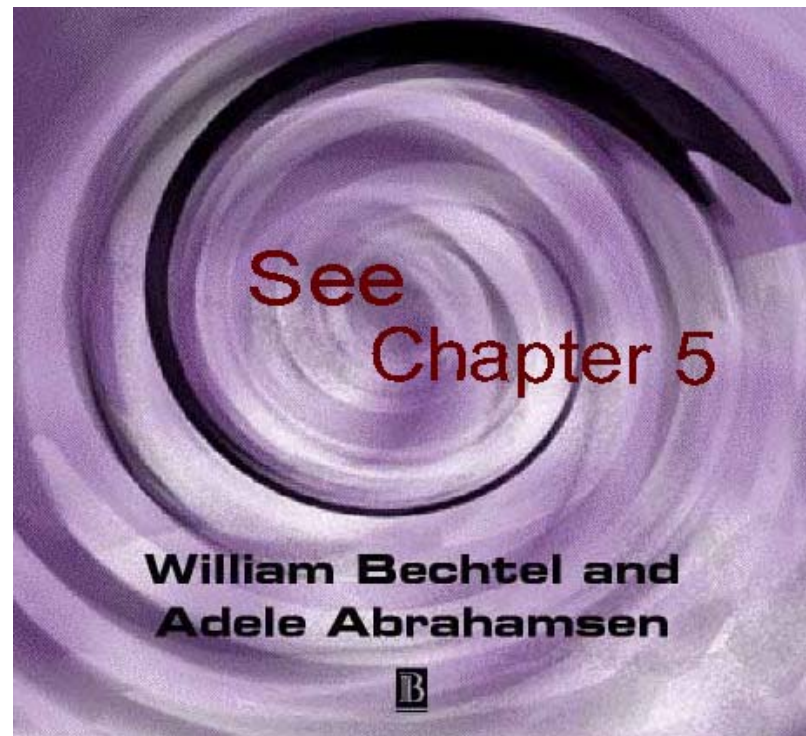


Neural Nets and Symbolic Reasoning

Is language governed by rules?

Models of past tense acquisition



Outline

- Is language use governed by rules?
- Rumelhart and McClelland's model
- Pinker and Prince's arguments for rules
- Plunkett and Marchman's simulation
- General conclusions

1 Is language use governed by rules?

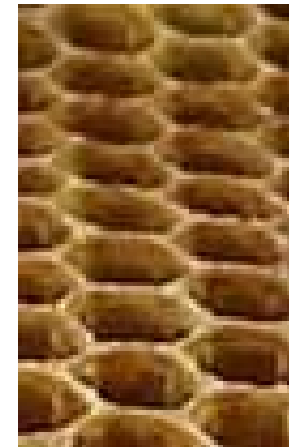


Noam Chomsky

Explicit inaccessible rule view

- Linguists stress the importance of rules and representations in describing human behaviour. Both are composed of sequences of **symbols**.
- We **know the rules** of language, in that we are able to speak grammatically, or even to make judgements of whether a sentence is or is not grammatical
- This does not mean we know the rule in a conscious, **accessible** way (like the rules of chess)
- It has been held (Chomsky, Pinker, ...) that our knowledge of language is stored **explicitly as rules**. Only we cannot describe them verbally because they are written in a special code only the language processing system can understand

- **No explicit inaccessible rules.** Our performance is characterisable by rules, but they are emergent from the system, and are not **explicitly** represented anyway
- **Honeycomb:** structure could be described by a rule, but this rule is not explicitly coded. Regular structure of honeycomb arises from interaction of forces that wax balls exert on each other when compressed
- **Parallel distributed processing view:** no explicit (albeit **inaccessible**) rules
- **Connectionism** not necessarily in conflict with the rule and representation view if rules and representations are assumed to be **emerging** at a certain level of description.



"...lawful behaviour and judgements maybe produced by a mechanism in which there is no explicit representation of the rule. Instead, we suggest that the mechanisms that process language and make judgements of grammaticality are constructed in such a way that their performance is characterizable by rules, but that the rules themselves are not written in explicit form anywhere in the mechanism..."

- **Eliminative or integrative position?**

Rumelhart & McClelland (1986) developed a connectionist model of past tense acquisition in English which challenged the classical rule view.

Stages of past tense acquisition in children

Stage 1 (1-2 years)

Past tense of a few specific verbs, some regular (e.g. looked, needed), most irregular (came, got, went, took, gave). Children initially memorize forms

Stage 2 (2-5 years)

Evidence of general rule for past-tense, i.e. add *ed* to stem of verb. Children often overgeneralise irregulars, e.g. *came* or *comed* instead of *came*. Ability to generate past tense for an *invented* word, e.g. *rick*. Subjects say *ricked* when using the 'word' in the past-tense

Stage 3

Children produce correct forms for both regular and irregular verbs.

The U shaped learning curve

Slightly older child: *Daddy came home*

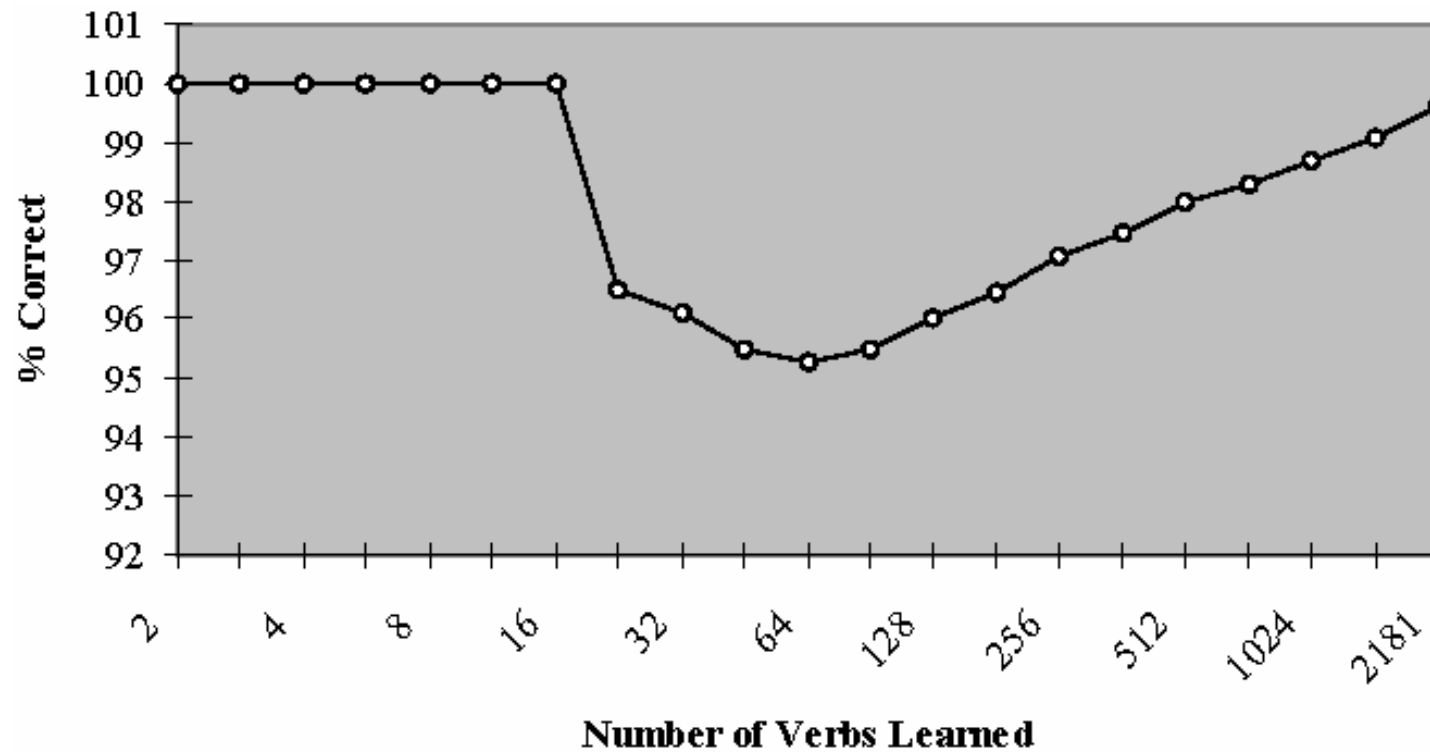
Stage 1

Older child: *Daddy comed/camed home*

Stage 2

Even older child: *Daddy came home*

Stage 3



The traditional view

- (1) The child memorizes some verbs, using memorization alone to produce correct inflected form
- (2) The child discovers grammar (e.g. $X \rightarrow Xd$), and in burst of joy and enthusiasm, produces forms like *singed*, *bringed*, *seed*, *goed*, etc.
Close temporal coincidence: overregularization kicks in when children first come to inflect regulars consistently.
- (3) Very gradually child memorizes the irregulars, to the point of producing them with adult reliability. **Exceptions block regularities!**

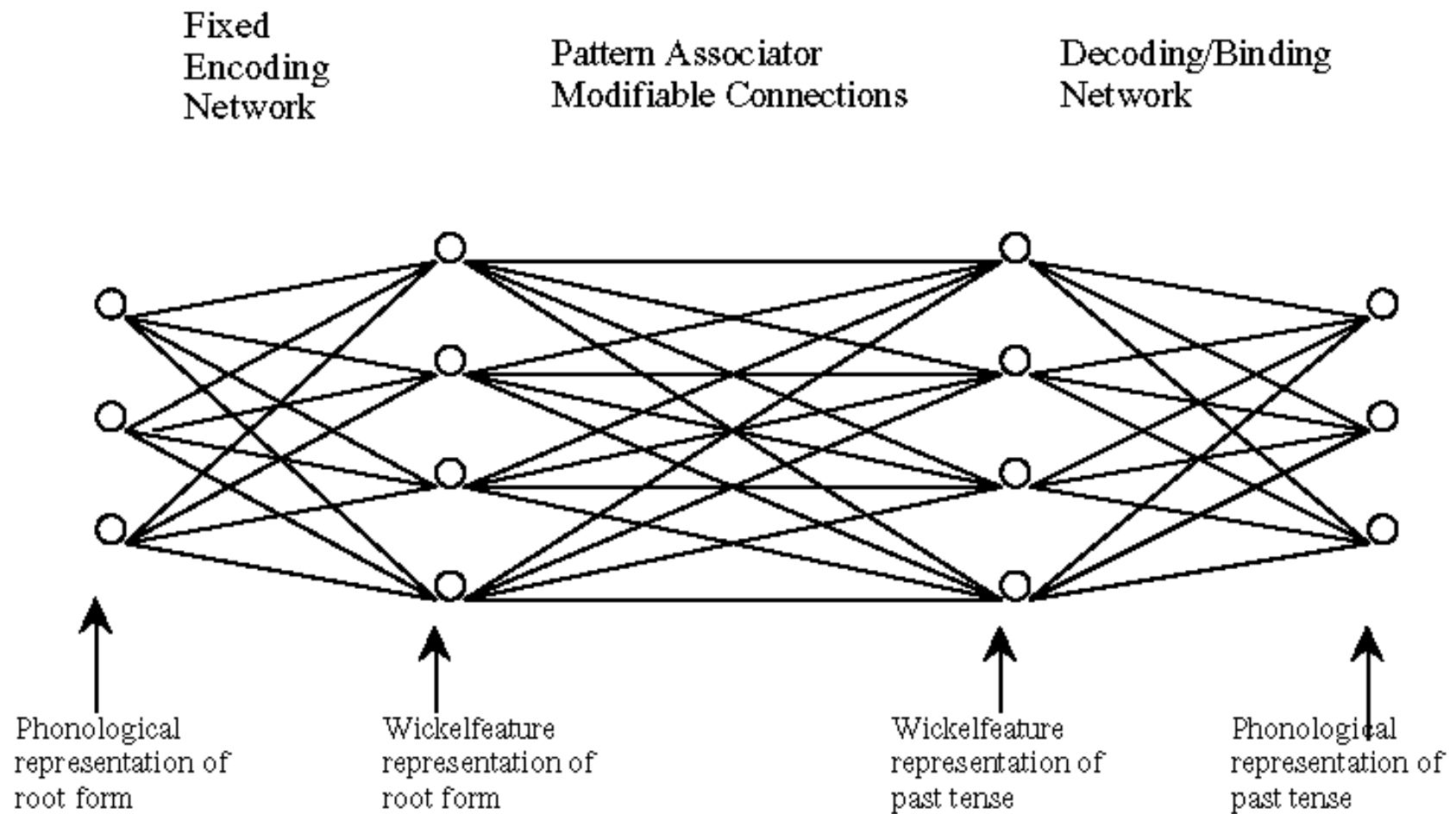
Be careful: Overregularization is vivid and interesting, so the non-careful investigator overestimates its occurrence. It occurs rather rarely (2.5% is typical, some kids higher, some lower).

2 Rumelhart and McClelland's model



David E. Rumelhart

The basic idea



To capture **order information** the *wickelfeature* method of encoding words was used.

Wickelphones: represent target phoneme and immediate context.

e.g. came /kAm/: #**K**_a, k**A**_m, a**M**# (# marks word boundaries).

Hence, 3 Wickelphones are used to encode /kAm/

If we distinguish 35 different phonemes we have $35^3 = 42875$ Wickelphones. If we use one input unit and one output unit for each Wickelphone we need a connection matrix with $35^3 \cdot 35^3 = 2 \cdot 10^9$ individual weights to represent all their possible connections.

42875 Wickelphones are coarse-coded onto 1210 **wickelfeatures**, where 16 wickelfeatures correspond to each wickelphone.

| | | | | | |
|---------------|-----|---------------|------|---------|----------------------------------|
| e.g. $kA_m =$ | 1 | (Interrupted, | Low, | Voiced) | |
| | 2 | (Back, | Low, | Front) | |
| | 3 | (Stop, | Low, | Nasal) | |
| | 4 | (Unvoiced, | Low, | Voiced) | |
| | ... | 16 | | | |
| | | 11 | 10 | 11 | 1210 different wickelfeatures |

11 (10) units to represent the feature specifications of a single phoneme. These features are sufficient to represent **similarities** between phonemes.

Importance of *wickelfeatures*

- The representations generated with the help of *wickelfeatures* are distinctive enough that different words can be distinguished (using some **redundancies** instead of 1210 only 460 *wickelfeatures* are required!)
- They overlap enough to support **generalization** on the basis of the **similarity structure** of the verb stem
- **Transfer effects**: Having learned that *sing* produces *sang*, for example, the network can be presented with *ring* and produce *rang*.

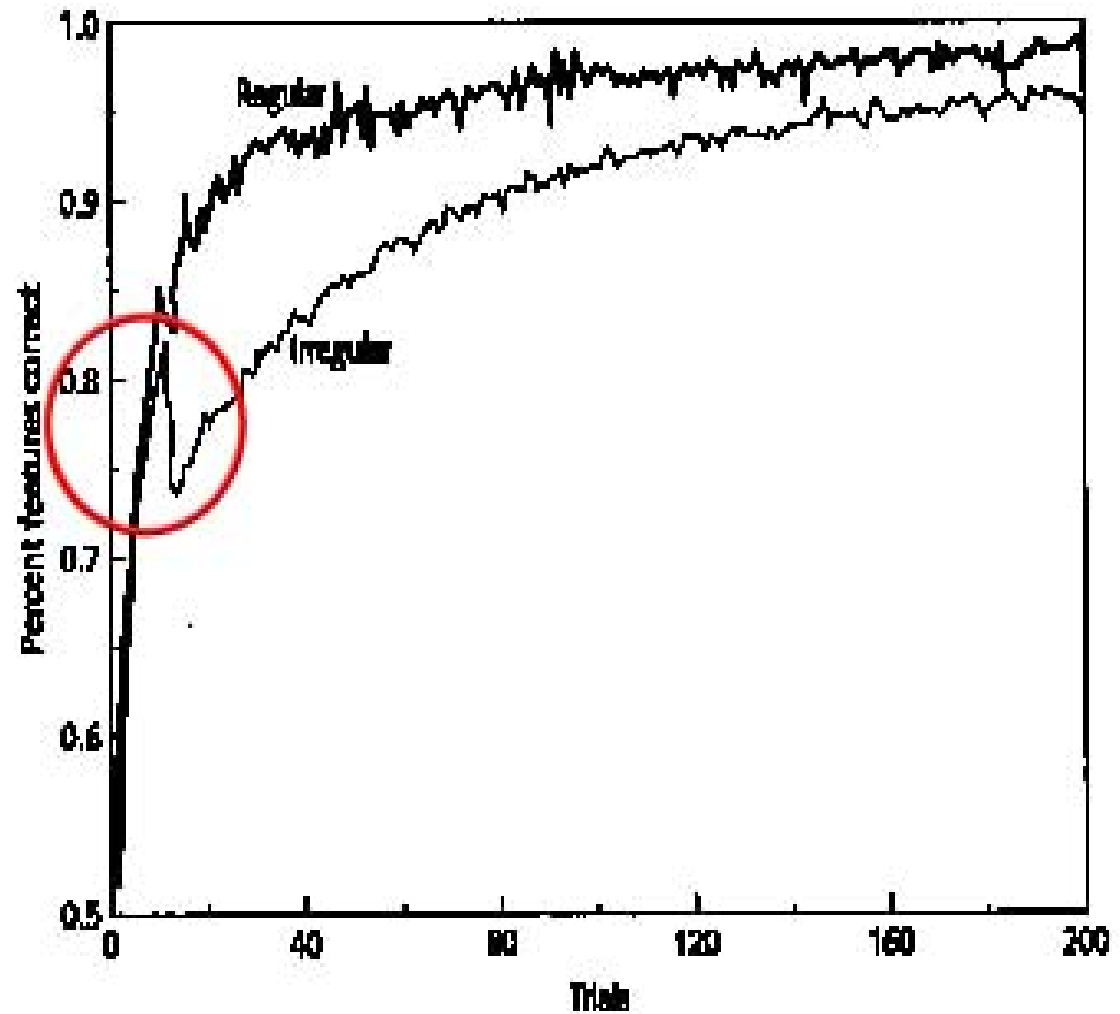
- 506 verbs divided into three sets:
 - 10 high-frequency verbs (8 irregular; 2 regulars: *live*, *look*)
live – lived, look – looked, come – came, get – got, give – gave, make – made, take – took, go – went, have – had, feel – felt
 - 410 medium-frequency verbs (76 irregular)
 - 86 low-frequency verbs (14 irregular)
- **Training I:** 10 high-frequency verbs for 10 epochs (Delta rule)
- **Training II:** 410 medium-frequency verbs added, for 190 epochs
- **Testing:** During learning the performance of the presented 420 verbs was registered. Afterwards, the 86 lower-frequency verbs were presented and the transfer responses were recorded.

Results (Overview)

- Network effectively learned the past tense of both regular and irregular verbs. The overall degree of transfer was 91% correctly generated *wickelfeatures* (92 % for regular, 84% for irregular)
- Matched human performance in learning and error patterns
 - U-shaped curve
 - Regular before irregular
 - Overregularization
- Matched the observed differences between different verb classes.

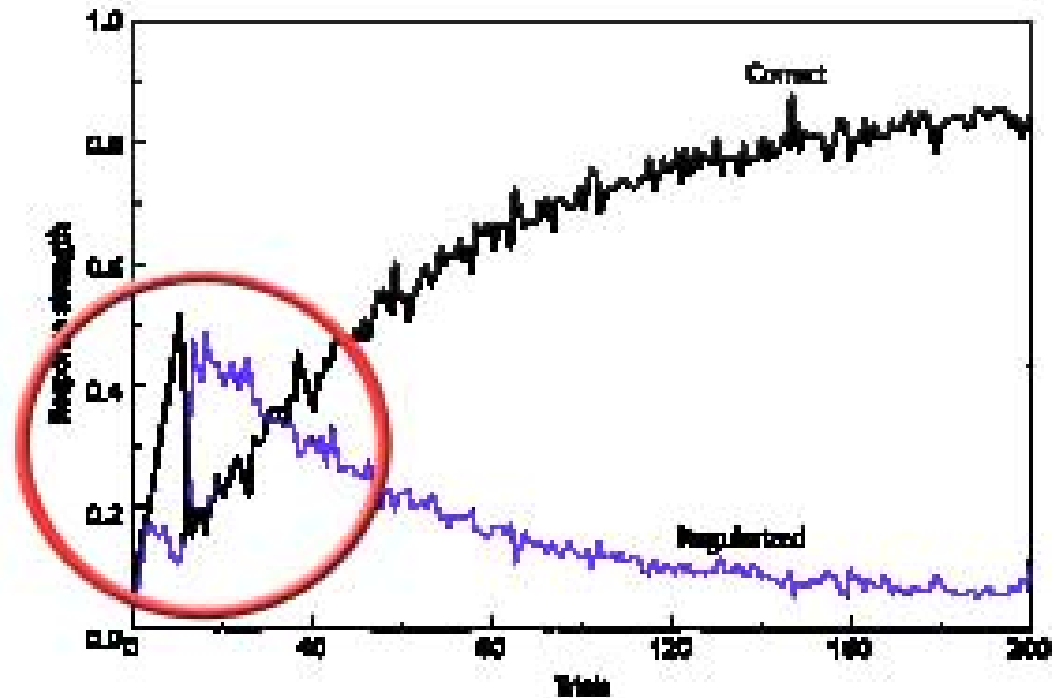
The three-stage learning curve

1. By epoch 10: 85% correct (both regular and irregular)
2. Performance correct on the irregular verbs dropped approximately 10 %.
3. The irregular verbs began to improve again by epoch 20 (gradually increasing to 95% by epoch 160).



Overregularization of irregular verbs

Response strength for high-frequency irregular verbs. The *response strength* reflects the proportion of a certain answer type compared with competing alternatives (e.g. for *come* the possible *Past Tense* forms are *came*, *comed*, *camed*, *come*).



Interestingly, the response strength increases considerably during phase 2 (epoch 10-30) **for wrongly regularized forms** (like *comed* & *camed*).

Differences between different verb classes

- **No-change verbs** (*beat, fit, set, spread, ...*): Bybee & Slobin found that verbs not ending in *t/d* were predominantly regularized and verbs ending in *t/d* were predominantly used as no-change verbs. Interestingly, the model had a propensity *not* to add an additional ending to verbs already ending in *t/d*! (already after 15 epochs of learning)
- **Verbs that undergo a vowel change**: 2 types of **overregularization error**:
 - (a) stem+*ed* (*comed, singed*)
 - (b) past+*ed* (*camed, sanged*)

Kuszaj (1977): Errors of type (b) are most frequent in older children. This is predicted by the model!

3 Pinker and Prince's arguments for rules



1. **The u-shaped learning problem:** “Rumelhart and McClelland's actual explanation of children's stages of regularization of the past tense morpheme is demonstrably incorrect.”
2. **The “ated” problem:** “Their explanation for one striking type of childhood speech error is also incorrect.”
3. **Errors are not based on sounds.** Elementary linguistic facts are not taken into account
4. **Wickelfeatures are not appropriate.** Different demonstrations clearly rule out *wickelfeatures*
5. **The phonological regularities problem:** “The model fails to capture central generalizations about English sound patterns.”

The U-shaped learning problem

- In training phase I, the model was given an input set that was **very small and rich in irregular forms**. Presumably, the failure to over-generalize the regular rule at this point was due not only to the high proportion of irregulars, but also to the small size of the learning set.
- In training phase II, Rumelhart & McClelland **shifted the nature of the input** radically and included a full complement of regular verbs. This shift led to the onset of overgeneralization of the regular rule
- One can argue that this sort of **fiddling with the input data** is an illegitimate way of deriving the desired phenomenon
- **Proportion of regular verbs** in parental speech is constant throughout relevant period (30%).

The "ated" problem

- The right prediction of errors such as *ated* or *wented* is not enough. The mechanism which produced them matters
- In the Rumelhart & McClelland model, the form *ated* was produced by activating a **vowel change pattern** and the final *ed-* pattern
- These errors are really produced by a **coding error**. The fact that children produce errors such as *ating* or *wenting* is good evidence that children occasionally **fail to code the irregular past as clearly past**
 - Evidence 1: **Reduplications** such as *jumpeded* appear
 - Evidence 2: Comparison of **experimentally elicited forms** and spontaneously produced errors: When children are asked to produce the past tense directly from the present tense *eat* errors of the *ated* type nearly totally disappear.

Errors are not based on sounds

- **Homophonous verbs** can have different past tense forms

ring-rang,

wring-wrung

ring-ringed (secondary sense of “to form a ring about something”)

Since the verb-learning model takes a **single phonological form** as its input, it will not know when to produce “rang,” “wrung,” or “ringed.”

- *Do, have, be* never overregularized as **auxiliaries**, but are overregularized as **main verbs**
- **Denominal/deadjectival** verbs are always regular, even when based on irregular verbs (*grandstanded, high-sticked*).

Wickelfeatures are not appropriate

- **The “algalgal” problem:** “The model is incapable of representing certain kinds of words.” Same set of *wickelfeatures* for words like *algalgal* “ramrod straight” and *algal* “straight” in the Australian language *Oykangand*
- **The “slit-silt” problem:** “It is incapable of explaining patterns of psychological similarity among words.” *Wickelphonology* cannot explain the high similarities between *slit* and *silt*, for example
- **The “pit-tip” problem:** “It easily models many kinds of rules that are not found in any human language.” No real **transformation** connects a string with its mirror image. Unfortunately, such transformations are simple to learn using *wickelfeatures*: ${}_A B_C \rightarrow {}_C B_A$.

The phonological regularities problem

- An important criteria against which any model should be judged is its ability to capture “significant generalizations.” The verb-learning model fails in this regard
- An English speaker who knows that “Bach” should be pronounced as */bax/* would also automatically realize that the past tense of the neologistic verb “*to Bach*” would be */baxt/* and not */baxd/* or */baxId/*
- The present model has trouble producing */baxt/* because it has no clear featural representation of the English sound system
- However, this is a problem that can be addressed merely through a change in the phonological representation. What is needed is a clear *segmental feature representation*.

Although the Past-tense model can be criticised, it is best to evaluate it in the context of the time (1986) when it was first presented. At the time, it provided a tangible demonstration that

- it's possible to use neural net to model an aspect of human learning
- it's possible to capture apparently rule-governed behaviour in a neural net
- past-tense forms can be described using a few general rules, but can be accounted for by a connectionist net which has no **explicit** rules.
- Both regular and irregular words can be handled by the same mechanism.

4 Plunkett and Marchman's simulation

I can't imagine how language could be learned



It must be innate

- The real mechanism of learning is important: backpropagation + making use of **hidden units** (in order to find powerful generalizations)
- Use less controversial representations (no *Wickelfeatures*)
- Respond to criticism of **inaccurate data set**
- Show that **U-shaped curves** can be achieved without abrupt changes in input. Trained on all examples together (using a backpropagation net).

1. Regular verbs that add one of three allomorphs of the */-ed/* morpheme to the stem to form the past tense:

(i) *arm* → *arm-[d]* (ii) *wish* → *wish-[t]* (iii) *pit* → *pit-[id]*

2. No change verbs: *hit* → *hit*

3. Vowel change verbs where the vowel in the stem is changed while the past tense retains the same consonants as the stem form:

sing → *sang*, *ring* → *rang*

4. Arbitrary verbs where is no apparent relation between stem and past tense form: *go* → *went*

The verb stems were *artificial* with three phonemes in length. However, all were phonologically possible in English, some corresponding to real English stems.

