A Connectionist Model of English Past-Tense and Plural Morphology

Kim Plunkett and Patrick Juola

Slides: Dana Dahlstrom

CSE 258A, UCSD

2002.04.17

English Past-Tense Inflection

- The -ed suffix forms the past tense of most English verbs, but a significant minority take an irregular form.

- Children learning English exhibit a U-shaped developmental profile.

- Symbolic theorists postulate a two-route mechanism: memorization of irregulars and a default rule system for everything else.

- Connectionists argue a single associative route is adequate.

Noun Plurals and the Past Tense: Similarities

- Both have a general rule with a small, semi-regular group of common exceptions.

- They have similar developmental profiles.

- Many phonotactic features are relevant to both.

Noun Plurals and the Past Tense: Differences

- There are far fewer irregular nouns than irregular verbs, but the irregular nouns are individually more frequent.

- Nouns are more frequent in running text.

- Irregular verbs are more irregular (e.g., all noun plurals share onsets with the singular form).

Questions

- Do regular nouns and verbs have separate inflectional routes?
- Does one associative memory system handle irregular nouns and verbs?
- How can we explain the functional separation between regular and irregular forms?

Inflection and Meaning May Be Linked

- Denominal verbs are regularized even when there is a homophonous irregular verb:

  flied out vs. *flew out*
  ringed vs. *rang*
  sticked vs. *stuck*

  Comment: The denominals here are all sports terms.

- A non-denominal verb where there are two past-tense inflections with different meanings:

  hanged vs. *hung*

Inflection and Meaning May Not Be Linked

- Slip-of-the-tongue data show (even irregular) inflection can occur positionally without regard to sentence meaning:

  days of the week becomes *weeks of the day*
  know one if I heard it becomes *hear one if I knew it*

- Comment: Anyway, people often get the wrong past-tense inflection when there are two of them with different meanings.

Problems with the Dual-Route Model

Plunkett and Juola claim the dual-route model must appeal to a complex rule system to account for certain variations:

- Words which are irregular with respect to one form but not others:

  The past tense of *go* is *went*, but the plural is *goes*.

- Some denominal verbs are derived from singular noun forms, while others are derived from plurals:

  knife, to knife, knifed vs. *half, to halve, halves*

Connectionism to the Rescue

- Form variations are more easily accommodated because the representation doesn't require systematic rule applications.

- Unlike the dual-route model, the single-route model would explain why voicing assimilation and epenthesis rules are the same for verb past-tense and noun plural inflections.

- Comment: Perhaps it would even help explain why inflections are often the same: English noun plurals, third-person singular present tense, and possessive nouns all use the -s inflection.

Experimental Design

Approach

Plunkett and Juola built a single-system connectionist model to simultaneously learn verb past tense and noun plural inflections.

- Their focus is on the time course of acquisition.
- The model's performance patterns are similar to children's.

Task Formulation

The network converts a stem to an inflected form. This formulation assumes:

- an *a priori* concept of "stem" and inflectional paradigms
- the ability to analyze language sounds correctly
- a mental lexicon, including some syntax and semantics

Network Design and Layout

- The design is a feed-forward, multilayer perceptron network.
- The input is a phonological representation of the stem plus a syntactic category.
- The output is a phonological representation of the inflected form.
- 130 input units, 160 output units, 200 hidden units

Network Configuration and Training

- The initial weights were randomly selected from $[-0.5, 0.5]$.

- The network was presented with each pattern individually in random order and trained with backpropagation for each.

- The learning rate $\eta = 0.1$, and the momentum $\alpha = 0$.

Input/Output Representation

- sixteen-bit phonetic feature vectors [Figure 4.1, Table 4.1]

- input: 8 vectors (CCCVVCCC) plus two bits for syntactic form (=130 bits)

  cat (/k&t/) represented by `##k#&##t`
  cot (/kAt/) represented by `##k#A##t`
  child (/CaIld/) represented by `##CaI#ld`

- output: (guess) 10 vectors (CCCVVCCCVC) (=160 bits)

  cats (/k&ts/) represented by `##k#&##t#s` or `##k#&#ts##`
  children (/Cil/dr@n/) represented by `##C#ildr@n`

Training Data

- Examples consist of monosyllabic noun and verb stems with no "foreign" sounds, together with their inflected forms.

- The token frequencies were taken from written, edited, adult-to-adult communication, not spoken child-directed communication.

Type Breakdown

category	stems	noun plural	past tense	total types
noun only	1,680	1,680	0	1,680
verb only	346	0	346	346
homonymous	600	600	600	1,200
all	2,626	2,280 (70.68%)	946 (29.32%)	3,226 (100.00%)
regular		2,254 (69.87%)	824 (25.54%)	3,078 (95.41%)
irregular		26 (0.81%)	122 (3.78%)	148 (4.59%)

Token Breakdown

category	noun plural	past tense	total tokens
all	13,045 (76.16%)	4,084 (23.84%)	17,129 (100.00%)
regular	12,841 (74.97%)	3,087 (18.02%)	15,928 (92.99%)
irregular	204 (1.19%)	997 (5.82%)	1,201 (7.01%)

- they applied function $\log_2(\text{freq}^2 + 1)$ to flatten the distribution

- they also tried $\log_2(\text{freq}^2 + 1)^2$

Test Methodology

- Simulations were run 5 times with random initial weights, and the results were averaged.

- The system's performance was evaluated after each training epoch.

Experiments

Simulation 1

- The network was trained on the entire corpus at each epoch.
- The goal was to establish a baseline performance, not to model real-life learning. [Figure 4.2]
- Comment: The ranking of forms by final performance is the same as the ranking by prevalence in the training set.
- Comment: Humans can learn this way, but they tend not to retain it.

Simulation 2 Setup

- They began with a small training set comprising the 20 most frequent forms.

- Forms were added in decreasing token frequency order every 5 epochs.

- Irregulars are prevalent among frequent forms.

Simulation 2 Results

- The U-shaped curve appears. [Figure 4.4]

- Again, performance corresponds to training set prevalence.

- Why are the plot lines after epoch 115 in Figures 4.3 and 4.4 not precisely the same? (They are close.)

Analysis

Frequency Compression

- The $\log_2(\text{freq}^2+1)^2$ compression scheme resulted in severely diminished performance on all forms but regular nouns.

- The aim of frequency compression is to capture *saliency*; they point out other factors such as meaning are important.

Overregularization

- The overregularization rates for nouns and verbs over the course of one simulation are depicted in Figure 4.5.

- The overall overregularization rates are "broadly similar" to those found in children (within an order of magnitude).

- Nouns were overregularized earlier and more frequently; Marchman *et al.* also found this phenomenon in children.

Categorial Error Analysis

- Overregularization was significantly more frequent than irregularization, as in children.

- The most frequent irregular words were immune to overregularization.

- No-change errors were more likely on stems ending in an alveolar consonant (presumably alveolar stops: /t/ and /d/).

- No-change verbs were half as likely to be overregularized.

Generalization

- Novel nouns were regularized earlier than verbs. [Figure 4.6]

- The network learned to null inflect novel verbs ending in an alveolar stop (/d/ or /t/) and to regularize novel verbs ending in a dental fricative (/D/ or /T/). This is rule-like behavior. [Table 4.4]

Critical-Mass Effects

- Novel-word regularization increases with vocabulary size.

- Nouns and verbs show approximately the same novel-word regularization rate vs. vocabulary size curve. [Figure 4.7]

- There are fewer verbs in the training set, so they take longer to reach critical mass. [Figure 4.6]

Assimilation Effects

- Pinker and Prince proposed a "downstream" voicing module; Plunkett and Juola aim to show the effects can be captured in a single system.
- No single input feature strongly predicted output voicing, though nasality and labialism inhibited an epenthetic schwa.
- Nevertheless, the network made very few voicing errors even on novel stems with "don't care" syntactic units.

Denominal and Deverbal Forms

- The network treated cross-paradigmatic inflections nearly identically to novel forms. [Figure 4.8] (*cf.* Figure 4.6)

- They added 50-bit random "pseudosemantic" vectors to the input types, then presented the network with known irregular forms but novel semantics.

- The network regularized the novel-semantics forms 4 times as often as it produced the original irregular inflection.

- Regularization increases with the number of semantic units.

Summary and Evaluation

(Secondary) Data Contact

- All simulations learned the training corpus to near perfection.
- Overregularization displayed a U-shaped development profile.
- The development of verb past tense trailed noun plurals.

More (Secondary) Data Contact

- The network generalized strongly at around 100 forms.
- No-change past-tense forms resisted overregularization.
- Overregularization was the most frequent verb inflection error, followed by no change, then blend.

Task Veridicality

- Human probably map from a thought to a correctly inflected expression.

- The network maps from a phonological representation of a stem to a phonological representation of an inflected form.

Input Representativeness

- The raw frequencies came from written, published, adult-to-adult communication, not child-directed communication.

- On the other hand, children are certainly exposed to adult communication.

- Frequency compression is somewhat open to criticism.

Conclusions

- Plunkett and Juola rebutted a few criticisms of connectionism from the symbolic crowd.
- Their model shows good performance and significant generalization, even to novel and cross-categorial forms.
- It makes good contact with a lot of data about morphology acquisition in children, and makes a few empirical predictions.