Computational Psycholinguistics

Learning the English Past Tense

Garance PARIS

Winter Semester 2011/2012

Exam Date:
Wednesday, 8th February

- Location: U.15
- 120 minutes from 14:00 to 16:00
- Be a little early, so we can start on time
- Don't forget your student ID + pocket calculator
- No mobile phones ;-) (Bring a watch if desired)

Past Tense Formation

Regulars: Stem + -ed

Three allomorphs of *-ed* depending on the final phoneme of the stem:

- */d/* if the stem's final phoneme is voiced, as in *arm/armed*,
- */t/* if the stem ends in an unvoiced phoneme, as in *wish/wished*,
- and */ɪd/* if the vowel ends in */t/* or */d/*, as in *pit/pitted*.

Past Tense Formation

Irregulars

- Arbitrary like *go/went*: Very few
- Many small subgroups show some pattern
  - No-change verbs like *hit/hit*: All end in *t* or *d* (Must be distinguished from *-/ɪd/* verbs)
  - Vowel-change verbs *(come/came, ring/rang, sing/sang):* Groups often have vowel + final consonant in common (they “rhyme”)
  - Blends *(sleep/slept, weep/wept):* Vowel change + ending

Children Behavior

U-Shaped Learning Curve

- Children make mistakes over the course of learning
- Surprisingly, mistakes often occur **after** producing the correct form at an earlier stage

  Example: A child might correctly say *went* at age three, but then produce *go-ed* a year later

Symbolic Account

Two Mechanisms: Dual-Route Model

- Early: Children simply memorize past tenses forms
- Later: They discover the *-ed* rule
  Apparently no more need to memorize forms,
  but leads to overgeneralization
- Finally: Learn which verbs are regular
  and which are exceptions

Suggested model: Dual-Route
Accounts for U-shaped learning and
double-dissociations

Symbolic Account

Criticism

- No explicit account of how the “+ed” rule is learned (except for reliance on frequency, see below)
- How much do we need to assume is innate?
  - Does this include the notion of inflection (only language specific “parameters” need to be learned)
  - or only a more general learning mechanism?
- Given the high frequency of irregulars in child directed speech, how are good/spurious rules distinguished and selected?
  Why not select a rule based on some subgroup of irregulars?

Symbolic Account

Criticism

- Frequency is an important factor in the recovery-phase in U-shaped learning

  English has an unusually large class of regular verbs: Only 180 irregulars
  What about other languages?
  - Only 20 % of plurals in Arabic are regular
  - Norwegian has 2 regular forms for past tense verbs: 3-route model?

Connectionist Models

Advantages

- No distinct mechanisms for regular and irregular forms
- Account for the actual learning process
- Requires positing less “innate” mechanisms
- Instead exploits the learning environment (data + frequency)

Plunkett and Marchman (1993; 1996)

Architecture and Training

- One hidden layer, trained with backprop
- Model maps verb stems to their past tense (both phonological representations)
- Initially training using 10 regulars and 10 irregulars (estimates of children's early vocabulary)
- Gradually increased to mimic child learning
- Total: 500 verbs, 90 % regular (as in English)
- Higher frequency verbs introduced earlier, and so seen more often
- Irregulars more frequent in English: Presented more often during training (This is essential, otherwise the regulars swamp the network; arguably reflects child directed speech)

Connectionist Models

Evaluation

- Final model successfully learned task for all verbs, but made errors during training
- Early acquisition: Error free
- Overall, low rate (5-10%) of overregularization
- No overregularization for common irregulars (as in child speech: *go-ed* is rare)
- Very small number of irregularization errors (treating a regular verb as an irregular one) e. g. *pick/puck*

Connectionist Models

Evaluation

The model is highly sensitive to the training data and frequency effects:

- The onset of overregularization is closely bound to a "critical mass" of regular verbs entering the child vocabulary
- Frequent presentation of arbitrary irregulars (go/went) is necessary to good performance
- Simulates the frequency $\times$ regularity interaction:
  - Faster reaction time for high frequency irregulars than low frequency ones
  - No advantage for regulars

Connectionist Models

Evaluation

- This suggests it is dangerous to infer dissociations in mechanisms due to observed dissociations in behavior:
  Critical mass effect can have the appearance of a distinct mechanism
- However: We know multi-layered networks can learn such mappings in general
  This does not prove that children use the same type of mechanism

Connectionist Models

Evaluation

- Lesioning leads to different behavior for regulars and irregulars
- But the model offers no explanation of the double-dissociation observed by Pinker (1994)

Connectionist Models

Evaluation

Pinker & Prasada argue that the statistical properties of English help the model:

- Regulars have low token frequency but high type frequency:
  Facilitates generalisation across this class of items
- Irregulars have low type frequency but high token frequency:
  Facilitates rote learning mechanism for these words
- They argue a connectionist model could not model the case with both low type and token frequency,
  e. g. Inflection of plural nouns in German