Deconstructing TD deletion

Meredith Tamminga
University of Pennsylvania

Josef Fruehwald
University of Edinburgh

N WAV 4 2 – Oct. 19, 2013
Introduction

Argument: “TD deletion” is not a unified phenomenon but rather the output of three distinct processes

Two recurring questions:
- What is a variable?
- How can any particular variable be defined?
What is a variable?

A core concept in sociolinguistics

- Probabilistic grammatical operation?
- Decision from among a set of options?
- Merely an analytic heuristic?
What is a variable?

“Sociolinguistic variation analysis is concerned with choices speakers make among the alternatives available to them regardless of the structural provenience of those choices. In many cases… there is no motivation for tying variation to rules of grammar.”

Fasold 1991 p. 12-13
How can variables be defined?

Where does one variable end and another begin?
How can variables be defined?

“...considerations of variable constraints may provide a principled basis for combining or separating linguistic processes.”

Wolfram 1975 p. 71

Influential principle of isomorphic constraints
(see Tagliamonte 2002)
Our view

- Crucial to tie observed surface variation to underlying grammatical structures
- Two new types of quantitative evidence to assess potential structural relationships.
- Delineating the grammatical underpinnings of variation also relevant to social meaning
TD basics

Variable deletion of word-final coronal stops in consonant clusters

Grammatical status of the coronal stop:

- monomorpheme > semiweak > regular past
- ‘fast’ > ‘kept’ > ‘worked’
TD history (abridged)

\[ t, d \rightarrow \emptyset / \left[ \frac{\alpha_{\text{cons}}}{\zeta_{\text{obs}}} \right] \gamma + \delta + \left[ \begin{array}{c} \varepsilon_{\text{voice}} \\ \vdash \end{array} \right] \beta \sim V \]

<table>
<thead>
<tr>
<th>work</th>
<th>kep+t</th>
<th>fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>work + ed</td>
<td>kept</td>
<td>fast</td>
</tr>
<tr>
<td>worked</td>
<td>kept</td>
<td>fast</td>
</tr>
</tbody>
</table>
Data

Buckeye Corpus (Pitt et al. 2007)
- 8,006 tokens
- 40 white speakers from Columbus

Philadelphia Neighborhood Corpus (subset from Labov & Rosenfelder 2011)
- 10,189 tokens
- 42 white speakers from South Philadelphia
Methods

- Buckeye tokens extracted using Python from hand-corrected phone tier
- PNC tokens coded auditorily using Praat script

<table>
<thead>
<tr>
<th>Deletion rates</th>
<th>Monomorph</th>
<th>Semiweak</th>
<th>Regular past</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buckeye</td>
<td>53%</td>
<td>41%</td>
<td>23%</td>
</tr>
<tr>
<td>PNC</td>
<td>55%</td>
<td>45%</td>
<td>31%</td>
</tr>
</tbody>
</table>
Clustering

Clustering of individual speakers: how similar or different are deletion rates for individuals?

- If all TD deletion is the same, expect speaker differences to show consistent range across categories
Clustering

Clustering of individual tokens: how evenly dispersed are the variants in sequences?

- Variants generally facilitate their own reuse (Poplack 1980, Scherre & Naro 1991, Cameron 1992 *inter alia*)
- If all TD deletion is the same, expect tokens to show persistence that disregards categories
Speaker clustering

First source of evidence on unity of process: patterns in individuals’ rates

TD deletion context: how do rates from different speakers cluster across grammatical categories?

Fit model to get individual speaker random intercepts within each category
Speaker clustering

Fit mixed-effects logistic models using:

\[ TD \sim \text{Gram} + \text{PreSeg} + \text{FolSeg} + \text{Freq} + \]
\[ (-1 + \text{Gram} | \text{Speaker}) + (1 | \text{Word}) \]

Captures speaker-level random error for each of the grammatical categories separately

Are inter-speaker differences consistent across categories?
Speaker clustering

Inter-Speaker Variation, Buckeye

By Speaker Random Effect

Grammatical Class

regular past  semiweak  monomorphemes

sd = 0.49  sd = 0.75  sd = 0.58
Speaker clustering

- Fruehwald & MacKenzie (2012): speakers more tightly clustered for regular past tense and monomorphemes than semiweak

- Fruehwald (2012): morphological variation in the semiweak verbs (like Guy and Boyd 1990) combined with the phonological variation across the board

\[
\text{keep} + T_{\text{past}} \rightarrow \text{kep} \sim \text{kept}
\]
Speaker clustering

Inter-Speaker Variation, PNC

By Speaker Random Effect

Grammatical Class

regular past  semiweak  monomorphemes

sd = 0.67  sd = 0.7  sd = 0.43

NWAV 42, 10/19/13
Tamminga & Fruchwald
Speaker clustering

Attribute wider variance to overlapping processes

Semiweaks subject to two variable processes:

- Variable morphology (allomorphy):
  \[ T_{+\text{past}} \rightarrow -\emptyset \quad \text{or} \quad T_{+\text{past}} \rightarrow -t \]
- Variable phonological deletion
Deconstruction so far

**morphology**
- **work + \( T_{\text{past}} \)**
  - Insert -t
  - [wɜːkt]

**phonology**
- Del
  - [wɜːkt]

**morphology**
- **keep + \( T_{\text{past}} \)**
  - Insert -t
  - Insert ∅
  - [kɛpt]
  - [kɛpt]

**phonology**
- Del
  - [kɛpt]

**morphology**
- **fast**
  - Insert -t
  - Insert ∅
  - [fæst]

**phonology**
- Del
  - [fæst]
Token clustering

Second source of evidence on unity of process: patterns in sequences of observations

In the TD context: how do tokens of the variants cluster across grammatical categories?

Code each token for values of previous token, by category, test for persistence
Token clustering

For each token, code the previous token’s:

- variant (deletion or retention)
- grammatical status (monomorpheme or past tense)
- distance (in seconds)

Are tokens preceded by deletion more likely to be deleted, and vice versa?
Token clustering

Past trigger, past target

Mono trigger, mono target

Probability of retention vs. Seconds since previous token
Token clustering

Mono trigger, past target

Past trigger, mono target

Probability of retention

Seconds since previous token

Previous_TD
Deletion
Retention

NWAV 42, 10/19/13  Tamminga & Fruchwald  23
Token clustering

Grammatically-matched trigger/target pairs show persistence, but mismatched pairs don’t.

Suggests distinct processes of TD deletion in monomorphemes and regular past tense verbs at the phonological level
The full deconstruction

work + $T_{\text{past}}$

keep + $T_{\text{past}}$

fast

morphology

Insert -t

Insert /t/

Insert $\emptyset$

phonology

Del

Del

Del

Insert -t

Insert /t/

Insert $\emptyset$
Summary

Two results:

• Greater deletion rate variance across speakers in semiweaks than in monomorphemes or regular past tense verbs

• Grammatically matched trigger-target pairs show persistence, but mismatched ones don’t
Summary

Three distinct sources of TD variability:

• Allomorphy in semiweaks
• Phonological deletion in all past tense forms
• Phonological deletion in monomorphemes
Discussion

Methodologically, we:

• Dispute the irrelevance of grammatical representations for defining the variable

• Offer two new types of quantitative evidence on unity of process in linguistic variation

These types of evidence can be considered in conjunction with constraint hierarchies.
Discussion

Socio-stylistic payoff for attention to grammatical underpinnings of variation?

A sneak peek at some work in progress…

- Interview with Celeste, two distinct styles
Discussion

Celeste: Style-shifting and grammar

![Graph showing the relationship between style and retention]

- Careful: Irregular 147, Mono 64, Regular 27
- Casual: Irregular 17, Mono 64, Regular 27

Legend:
- Irregular
- Mono
- Regular

Retention values for each style and grammar type are indicated on the graph.
Conclusion

Variable phenomena that look similar on the surface may have different grammatical origins.

Suggested model involves three processes.

New types of quantitative evidence may lead to similar revisions of other variables.
Thank you!

Thanks to Bill Labov, Dave Embick, Laurel MacKenzie, audiences at PLC 35 and NWAV 41

Contact us:

Meredith: tamminga@ling.upenn.edu
Joe: josef.frueh@ed.ac.uk
References


