Individual- vs. community-level variation: New evidence from variable (t,d) in Canadian English[1]

Katharina Pabst
University of Toronto

Lex Konnelly
University of Toronto

Melanie Röthlisberger
University of Zurich

Sali A. Tagliamonte
University of Toronto

1. Introduction

The simplification of consonant clusters involving the alveolar stops /t/ and /d/ in word-final clusters in English, also known as (t,d) deletion, is one of the most frequently studied variables in variationist sociolinguistics (Labov et al. 1968; Guy and Boyd 1990; Tagliamonte and Temple 2005, *inter alia*). This research has shown that uninflected or monomorphemic words, as in (1a), undergo deletion at a higher rate than regular past tense verbs, as in (1b), with semi-weak verbs, i.e., past tense forms which have both a stem-vowel change and a past tense suffix, patterning in between, as in (1c).

(1) a. The **REST** was all fields. (wcarlsberg/F/78)[2]
b. I **FOUNØ** I'd come home all pissed off about shit. (fconnor/M/17)
c. Yeah, the girls **LOVED** him. (mmcguigan/M/32)

Researchers have proposed a variety of models to account for this pattern, ranging from variable rules (Labov et al. 1968) to lexical phonology (Guy 1991) and Optimality Theory (Kiparsky 1994). All of these approaches have one thing in

[1] This study is based on the Toronto English Archive, which was collected by Sali A. Tagliamonte through the generous funding of the Social Sciences and Humanities Research Council of Canada (2003-2006) for research grant #410-2003-0005. Preliminary analysis of variable (t,d) in these materials can be found in Tagliamonte (2012:179-187).
[2] All examples stem from the Toronto English Archive and contain three items of demographic information about the speaker: Speaker pseudonym, sex, and age at the time of recording.

common: They are based on the assumption that (t,d) deletion is a single, unified process. Recently, Tamminga and Fruehwald (2013) have questioned this assumption. Analyzing two corpora of American English, they found that semi-weak verbs have higher rates of inter-speaker and inter-lexical variation than monomorphemes and regular past tense verbs. Combining this finding with previous research suggesting that children do not attach the past tense suffix to semi-weak verbs (Guy and Boyd 1990; Roberts 1997; Smith et al. 2009), they argue that the surface variation between the different morphological forms is the result of three separate processes, two phonological (phonological deletion in monomorphemes and phonological deletion in all past tense forms) and one morphological (allomorphy in semi-weaks). Thus, while all monomorphemic, semi-weak, and regular past tense forms are subject to the more general phonological rule of consonant cluster reduction, the semi-weak verbs are also subject to allomorphy; for example, a semi-weak verb like *kept* also has a competing allomorph *kep*.

This paper replicates Tamminga and Fruehwald's (2013) analysis using data from Toronto, Canada (Tagliamonte 2003-2006). Tamminga and Fruehwald (2013) hypothesized that if (t,d) deletion is one unified process, inter-speaker differences should show a consistent range across categories – that is, their rates of deletion should be relatively similar for each morphological class, respectively. They investigated this by fitting a mixed-effects model on the full dataset that included a random slope of individual speaker by morphological class. Results showed that the range of inter-speaker differences was not consistent across these categories for their data: Speakers were more tightly clustered for regular past tense and monomorphemes than for semi-weak verbs.

Testing Tamminga and Fruehwald's (2013) hypotheses on a different data set allows us to address two issues: First, it allows us to test whether semi-weak verbs display higher rates of inter-speaker and inter-lexical variation in different varieties. If this were true, it would be further supportive to the hypothesis that there are indeed multiple distinct processes at work. Second, it allows us to further investigate "[o]ne of the enduring questions in linguistics" (Guy 1980:1): the relationship between individual- and community-level variation. Decades of research have shown that there is indeed such a thing as a "community grammar", meaning that individuals who are part of the same community tend to display similar patterns of variation. However, relatively little is known about the extent to which individuals can and do deviate from community norms (for a more detailed discussion of these issues, see Walker & Meyerhoff 2003). Taking a closer look at inter-speaker variation will allow us to address this issue in the literature and augment our understanding of variation at the community level.

2. The data

The data we use to replicate Tamminga and Fruehwald's (2013) findings stems from a subset of the Toronto English Archive, a corpus consisting of sociolinguistic interviews with speakers born and raised in Toronto, Canada. The subset consists of 56 speakers (28 males, 28 females) between the ages of 17 and 92. The speakers represent a variety of educational backgrounds, occupations, and ethnicities (for a more detailed description of the data set, see Tagliamonte 2012). The size and representation of social groups within the corpus makes it the ideal tool with which to pursue the question of the role of the individual vs. the group, as it closely models the makeup of the broader speech community.

The data comprise 3,418 tokens: all lexical items with word-final consonant clusters ending in /t/ or /d/, excluding negative and interrogative constructions (e.g., *isn't*, *won't*, etc.), proper nouns, neutralization contexts, lexical items where the exact type of noun phrase could not be determined as well as the conjunction *and* (cf. Tagliamonte and Temple 2005). The data show that (t,d) deletion in the Toronto data has four variants: deletion (Ø), and three surface realization forms:

[t], [d], and the glottal stop [ʔ]. For the purposes of this study, all tokens that end in [t] or [d] will be treated as realized, while all tokens that end in Ø and [ʔ] will be treated as deleted (following Walker 2012).

The overall distribution of final consonant deletion and retention is presented in Table 1.

Table 1
Overall distribution of variable (t,d)

	N	%
Deletion	1,704	49.85 %
Retention	1,714	50.15 %
Total N	3,418	100.00 %

As the table shows, the rate of deletion in the Toronto data is 49.85%. This contrasts with data from York English in the UK, which has a much lower deletion rate of 24% (Tagliamonte and Temple 2005:287).

3. Coding and analysis

Our coding schema follows existing precedent, focusing on three linguistic factors most commonly found to be significant in conditioning rates of (t,d) deletion: preceding phonological environment, following phonological environment, and the morphological class of the word (cf. Guy and Boyd 1990; Bayley 1994; Santa Ana 1996), summarized in the table below. Tokens were also coded for age, sex, occupation, and normalized corpus frequency; however, consistent with the analysis that we replicate here, we do not consider demographic factors in the present study.

Table 2
Summary of coding protocol

Linguistic factors

Preceding phonological segment (based on Tagliamonte and Temple 2005)	Sibilants Non-sibilant fricatives Stops Liquids and glides Nasals
Following phonological segment (based on Walker 2012)	Consonant Vowel Pause
Morphological class (based on Guy and Boyd 1990; Tagliamonte and Temple 2005, *inter alia*)	Monomorphemes (e.g., *mist, raft, old*) Regular past tense (e.g., *walked, showed*) Semi-weak past tense (e.g., *left, told*)
Normalized corpus frequency	Frequency of token in corpus divided by total number of words in corpus

To address the two research questions, namely (1) what is the extent to which individual speakers deviate from the community norm and (2) what are the possible differences between (t,d) deletion at the individual- vs. the community-level, we must tease apart the effect of the various conditioning factors. For that purpose, we fitted a mixed-effects logistic regression model to the data using the lme4 package in R (Bates et al. 2015; R Core Team 2016). In the following, we present the results of the model with individual as random effect to establish the overall community norm. Next, we visualize the individual-level variation by plotting the adjustments of each individual speaker by morphological class.

4. Results

4.1 Establishing the community norm

Results of the fixed effects are presented in Table 3 below. In order to replicate the earlier study, we followed similar statistical modeling as closely as possible, including a random intercept for word and a random slope for individual by morphological class whereby the intercept was excluded. Following standard practice in multivariate modeling, we scaled numeric factors by two standard deviations and centered them around the mean (Gelman 2008).

Table 3

Main effects of factors in the model

(model predictions are for deletion; all factors shown)

Factor	β	p
Following segment		
vowels → consonants	2.81	***
vowels → pauses	0.55	***
Preceding segment		
liquids → nasals	0.95	***
liquids → sibilants	0.58	***
liquids → fricatives	0.00	
liquids → stops	-0.14	
Morphological class		
regular → monomorphemes	0.55	***
regular → semi-weak	0.31	
Frequency	0.06	

Estimates of the coefficients are given on a logit-scale in the column labelled β. Positive values indicate a preference for deletion, and negative values a preference for retention. The reference levels of the model are given on the left side of the arrow and the predicted level on the right side of the arrow. The table can therefore be interpreted as follows: If the following phonological segment is a consonant instead of a vowel or a pause instead of a vowel, deletion becomes more likely. Further, if the preceding phonological segment is a nasal or sibilant instead of a liquid, deletion also becomes more likely. Finally, if the word is a monomorpheme (e.g., *mist*) instead of regular past tense (e.g., *showed*), deletion also becomes more likely. Frequency does not have a statistically significant impact on deletion. Predictor importance was then assessed with the Anova() function in the car package (Fox and Weisberg 2011): Results show that following phonological segment is the most important predictor, followed by preceding phonological segment and then morphological class. This indicates that the patterning of (t,d) deletion is consistent with other varieties, stable in the ranking of its constraints (Guy 1980). Having established the overall community norm, we can return to the question of how rates from different speakers cluster across morphological categories.

4.2 Visualizing individual variation

Following Tamminga and Fruehwald (2013), we fitted a model with a random slope of individual by morphological class in order to capture by-individual adjustments to the mean of the effect of morphological class. This allows us to capture the speaker-level random error for each class separately. A boxplot visualizing the by-speaker adjustments to the three morphological classes is provided in Figure 1.

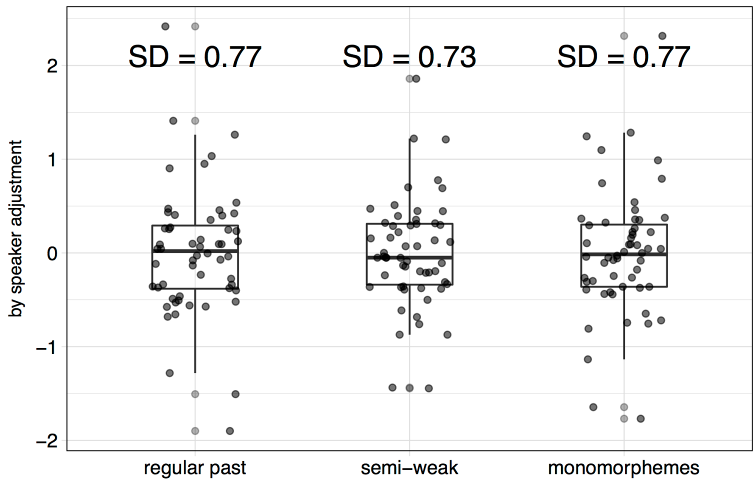

Figure 1. Clustering of inter-speaker variation by morphological category in the Toronto data.

Outliers that are located towards the top of the plot are the individuals who have the greatest rates of retention in the data set, while outliers located towards the bottom are those who have greater rates of deletion. Comparing the standard deviation across each category shows that, unlike in Tamminga and Fruehwald's (2013) data, the inter-speaker variation is largely stable across morphological categories. This is confirmed by the results of a Levene's test, which indicates that there is no statistically significant differences between the standard deviations (F = 0.2452, p = .7828).

Our findings contrast substantially with Tamminga and Fruehwald (2013). Recall that they observed the tightest clustering for regular past tense forms, followed by monomorphemes, and the greatest variation for semi-weak forms. We, on the other hand, do not find any differences across morphological categories.

5. Discussion and concluding remarks

The results presented in the previous section have two possible interpretations:

1) Semi-weaks are not actually subject to two distinct processes, and (t,d) deletion is indeed a unified phenomenon.
2) Perhaps semi-weak verbs are subject to two processes, but inter-speaker variation is not an accurate piece of evidence to support this.[3]

As Tamminga and Fruehwald (2013) argue, if deletion is one uniform process and not three distinct ones, speaker differences should show a consistent range across categories. This is exactly what we see in the Toronto data, which suggests that, at least in Toronto English, (t,d) deletion is one single, unified process and not multiple distinct ones.

This leads to the question of why we find such different results from Tamminga & Fruehwald (2013). It is possible that this is due to differences in sampling. However, it should be noted that all three corpora are based on interviews with native speakers. In Tamminga & Fruehwald's case, all of the speakers are white; our corpus includes speakers from a wide variety of ethnic backgrounds, but if anything, we might expect this to lead to *more* rather than *less* variation. To understand what rates of inter-speaker variation can tell us about (t,d) deletion, we clearly need to examine how these rates cluster in other varieties.

[3] As discussed in Tamminga and Fruehwald (2013), alternative evidence for a multiple process analysis may come from persistence effects in morphological structure; for a more detailed discussion, see Tamminga (2016).

In conclusion, this study continues the building tradition of replication and comparison in variationist sociolinguistics and adds insights from new statistical techniques. Despite the established knowledge base on variable (t,d) in the variationist literature, this phenomenon still has mysteries to explore. As we have demonstrated here, examination of variation at the individual level according to grammatical category provides new insights into an already much-studied variable process.

References

Bayley, R. (1994). Consonant cluster reduction in Tejano English. *Language Variation and Change*, *6*(3), 303–326.

Bates, D., Mächler, M., Bolker, B. M. & Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Fox, J. & Weisberg, S. (2011). An R companion to applied regression. Second edition. Thousand Oaks, CA: Sage.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine, 27*(15), 2865-2873.

Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (ed.), *Locating language in time and space* (pp. 1-36). New York, NY: Academic Press.

Guy, G. R. & Boyd, S. (1990). The development of a morphological class. *Language Variation and Change, 2*(1), 1-18.

Kiparsky, P. (1994). *An OT perspective on phonological variation*. Paper presented at New Ways of Analyzing Variation in English, Palo Alto, CA, Oct 17-19.

Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). *The Study of the Non-standard English of Negro and Puerto Rican Speakers in New York City: Phonological and grammatical analysis: Report on co-operative* *research project 3288*. New York, NY: Columbia University Press.

R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria.

Roberts, J. (1997). Acquisition of variable rules: a study of (-t, d) deletion in preschool children. *Journal of Child Language*, *24*(2), 351–72.

Santa Ana, O. (1996). Sonority and syllable structure in Chicano English. *Language Variation and Change 8*(1), 63-89.

Smith, J., Durham, M., & Fortune, L. (2009). Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change*, *21*(1), 69–95.

Tagliamonte, Sali A. 2003–2006. Linguistic changes in Canada entering the 21st Century. Research grant. Social Sciences and Humanities Research Council of Canada. #410-2003-0005.

Tagliamonte, S. A. & Temple, R. (2005). New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change*, *17*(3), 281-302.

Tagliamonte, Sali A. (2012). *Variationist Sociolinguistics: Change, observation, interpretation*. Malden, MA and Oxford: Wiley-Blackwell.

Tamminga, M. (2016). Persistence in phonological and morphological variation. *Language Variation and Change*, *28*(3), 335-356.

Tamminga, M. & Fruehwald, J. T. (2013). *Deconstructing TD deletion*. Paper presented at New Ways of Analyzing Variation 42, Pittsburgh, PA, Oct. 17-20.

Walker, J. A. (2012). Form, function, and frequency in phonological variation. *Language Variation and Change*, *24*(3), 397-415.

Walker, J. & Meyerhoff, M. (2003). Studies of the community and the individual. In R. Bailey, R. Cameron & C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics* (pp. 175-194). New York, NY: Oxford University Press.