

The Effect of Dialect Features on the Perception of “Correctness” in English-Word Voting Patterns on Forvo.com

Jessica Grieser*

1 Introduction

“Web 2.0” is the name given to the recent boom in the creation of web-based resources easily accessible by lay internet users to share and publish information about themselves and the things that matter to them. But perhaps more interesting than the social function of Web 2.0 resources is their ability to harness the value of feedback from a great number of users to validate the authority of a source. The most famous site to make use of this concept is Wikipedia.com, which allows any user to create and edit entries on any topic of their choosing with the idea that users will censor one another’s work to ensure the accuracy of the information on the site. Forvo.com, the website considered in this project, uses such technology to gauge listener responses to word pronunciations in over 200 languages. Although its intent is to provide a user-verified online pronunciation dictionary, this process also allows for the examination of naïve listener reactions to linguistic variation, also known as perceptual dialectology (Preston 2002).

Preston (2002) calls for the perceptual study of dialects for several reasons. First, dialect features and their interpretations and language ideologies they invoke are an integral part of both the speakers’ and hearers’ culture and ethnography; language variation cannot be easily untangled from its cultural meaning. Second, there is evidence that there is much knowledge to be gained from the interaction between folk belief and empirical study, as, for example, when folk medicine practices are evaluated using the same research methods more commonly used to assess the efficacy of laboratory-based medicine. Third, and perhaps most important, any person who hopes to help others assess their own language ideologies must first understand what those ideologies are.

One of the difficulties of gathering naïve listener reactions, however, is what Labov (1972) terms the “observers’ paradox,” which argues that the gatherer of sociolinguistic data will inevitably have an effect on the data gathered. One technique to avoid the observers’ paradox is to have speakers identify regions that they feel are most “correct” by drawing dialect maps, and to compare the maps drawn by speakers of different locales (Preston 2002, Hartley 2002). A second useful tool is the matched-guise technique, where listeners are presented with identical or very similar speech in a variety of accents, often produced by the same speaker (cf. Purnell, Baugh, and Isardi 1999). Giles (1970) finds that listeners will not only identify an accent in such an exercise but are also confident making judgments about the character of the speaker based on their perceptions of the social meaning of his or her accent and its “correctness.”

It is this notion of correctness that is one of the most compelling reasons to study perceptual dialectology. Although linguists agree that no dialect is empirically better than another, naïve speakers generally do perceive that there is a correct variety of their language, even if they themselves do not speak it (Preston and Niedzielski 2003) and despite their being generally inept at using phonetic features to separate one dialect from another (Clopper and Pisoni 2004). Speakers also tend to have a bias toward their own speech; they are more likely to judge their speech as different from that of their broader region (Giles 1970) and to ignore acoustic evidence that runs counter to what they perceive as the salient phonological features of their own dialect (Niedzielski 1999).

Determining which features speakers pay attention to when evaluating the correctness of their own accents and those of others tells us a great deal about which variants are assigned prestige and which have little, if any, effect on a listener’s perception of a speaker. The present study seeks to evaluate listener norms for prestige variants as displayed when listeners are asked to rate a single English pronunciation based on its “nativeness” and its “correctness.” Many studies have shown that naïve listeners cue in to a number of phonetic features when evaluating speech for dialect region affiliation and that the particular features that the listeners most attend to can be identified

*I thank Dr. Natalie Schilling, my classmates, and the attendees of NWAV 38 for their feedback on this project. Any errors which remain are my own.

(Clopper and Pisoni 2004, Preston 2002). This study examines user votes on the pronunciation website Forvo to evaluate if the votes pattern in a way that enables the identification of individual variants which affect a given user’s likeliness to evaluate a pronunciation as “correct,” looking specifically at the realization of medial /t/ and at accent local/topic match. I ultimately conclude that both these factors influence the degree to which a given pronunciation may be scored as “correct” by naïve listener voters on the site.

2 Data Source and Methodology

Forvo has the goal of producing user-created and user-maintained database to provide a pronunciation guide for every word in every language in the world. It lists words or phrases to be pronounced, with everything from everyday items and idiomatic phrases to celebrity and place names. Any user can record and upload a pronunciation for evaluation, even if other pronunciations have already been provided, and that pronunciation is then voted upon by other users as to its correctness.

Forvo uses a model similar to those employed by Wikipedia, YouTube, and Amazon¹ in allowing people to vote for their favorite examples, with the idea that the more users who vote a particular example as correct, the more likely that example represents the correct form. Launched in January 2008, Forvo has garnered in less than two years a registered user base of over 60,000, with new users joining the site at a rate of about forty a day. In addition, anyone with a microphone can record pronunciations on the website as an anonymous user, so the thousands of Forvo users likely represent only a small fraction of the total people who are recording their pronunciations on the site.

As of this writing, over 585,000 pronunciations are recorded on the site in 241 languages. As Forvo is an English-based² website hosted in the United States, it is unsurprising that English is the most well-represented language in the database, with over 60,000 recorded pronunciations.

This database was chosen as the source for this project for several reasons. As an entirely user-created and free database, anyone with a microphone attached to their computer can participate in the database. Although “computer-owners” does limit the sample to speakers who can afford or who have access to a computer, it nevertheless provides a much broader sample than a study forced to draw its users from a small subject pool. Additionally, like many other user-driven sites like YouTube, the voting scale on Forvo is uni-dimensional—users are asked to vote on how “correct” the pronunciation is on a scale of 1 to 5³ and are asked for no further input.⁴ Unlike studies in which users are asked to attend to a particular phonetic feature such as vowel quality (Niedzielski 1999), asked to identify the locale of the speaker (Clopper and Pisoni 2004), or asked to rate the speaker’s personal characteristics based on their speech (Giles 1970, Thakerar and Giles 1981, Bayard et al. 2001), this single dimension means that the votes themselves are indicative of the underlying understanding of what makes a pronunciation “correct” (although of course phonetic features and listener perception of the speaker’s locale and personal characteristics are part of this overall evaluation of correctness).

Lastly, a database like Forvo allows for data to be gathered in a way that virtually eliminates the observer’s paradox. The expressed goal of the site is to have speakers pronounce and evaluate words in their native tongues so that others can be guided to the pronunciations deemed most authoritative by sheer number of positive evaluations. Although the voting patterns can and do demonstrate biases toward dialects of languages and thus expose the underlying opinions of the speakers, the voters are likely unaware of their votes being used in this way. They are instead try-

¹en.wikipedia.org, www.youtube.com, and www.amazon.com

²The website was launched in English in 2008, a Spanish language version was launched in 2009, and the site now offers translations of the site into German, French, and Mandarin Chinese.

³The score descriptions provided by the website are as follows: 1-worthless, 2-no native or bad sound, 3-good, 4-great, 5-perfect!

⁴A feature allowing users to comment on pronunciations in addition to voting was added during the period that data was collected for this project, however, it has yet to see widespread use. As commenting becomes more common on the site, further study might involve examining the content of these comments in relation to a pronunciation’s score.

ing to adhere to the goals of the site by voting for the pronunciations they feel are the best representations of how they feel the word would be pronounced by a native speaker.

Of course, native speakers are likely to employ a wide range of pronunciations even for the same word, any of which is correct at least in some sense, and nearly all of which will allow for a native listener to understand the word being spoken. When asked to evaluate “correctness,” then, listeners are not truly evaluating the correctness of a pronunciation but rather the level of prestige they feel a pronunciation has. By examining which features the high-scoring pronunciations have in common, we can cue in to the features to which listeners are unconsciously assigning prestige.

Data was gathered for this study over a period of six weeks in late 2008 by monitoring the website for changes in the most frequently listened pronunciations (those likely to have been rated by multiple listeners). Pronunciations selected for rating analysis were limited to those that had received three ratings or more. One difficulty encountered in gathering data was that only registered users’ profiles listed the number of users who had rated a given pronunciation.⁵ Therefore, pronunciations recorded by anonymous users were excluded from this study as it was impossible to determine whether they met the three-vote threshold.⁶ The collected pronunciations were then grouped into low-scoring (1-1.9), medium-scoring (2.0-3.9) and high-scoring (4.0+) pronunciations and analyzed for features which affected this score.

122 tokens from 26 speakers were used in this study. Of the 26 speakers, most identified their dialect⁷ as from the United States and Canada (18 and 2, respectively), two identified as from other English-speaking nations (one each from England and Australia), and the remaining four from other locales around the globe.

The tokens collected were provided by 22 men and 4 women, representing an apparent general trend among Forvo users.⁸ However, the women in this sample were responsible for 29% of the total tokens, nearly twice what would be expected by their representation among the total speakers in this study.

3 Discussion

Two primary factors seemed to influence voters’ perceptions of the correctness of a pronunciation most heavily: perceived hypercorrection and topic association. We will look at these both in turn. Given the overwhelming percentage of speakers with a dialect locale in one of the four English-speaking countries represented and the U.S.-centric features of Forvo itself, it seemed most salient to focus on the ways voters might be assessing differences between U.S. English varieties and non-U.S. English varieties.

3.1 Medial /t/ Release as Evidence of Hypercorrection

Wardhaugh (1999) and Bayard et al. (2001) point to the realization of intervocalic /t/ as a flap as a strong marker of American English, one that Australian and New Zealand youth attend when contrasting American and British pronunciations (Bayard et al. 2001). Wardhaugh points out that reduced pronunciation is a natural part of the speech of most native English speakers; it is a byproduct of speaking the language quickly and fluidly. One way this manifests itself for American speakers is the realization of intervocalic /t/ as a flap, so that words like *water*, instead of being

⁵In the course of this project I made a request to the site managers that this feature be added to all votes, however, it has not yet occurred as of this writing. Further study of anonymous user pronunciations would be useful if and when this feature is added.

⁶Although it is possible to tell that a score has at least 3 votes (if the average score ends in .33, .67 or some other fraction other than .5 for example), this would bias the evaluation of anonymous pronunciations against those that may have received identical scores from multiple users.

⁷Forvo asks users to pinpoint their “dialect location” on a map. Although the site specifies that this should be the location of the user’s dialect, it is certainly possible that many users indicate their present living location.

⁸Although there is no way on the current site to gather an overall statistic about male vs. female users, a comparison of the 20 top users for the week of 12/1/08 and the top 20 users of all time reveals that on average, 72% of the top users are male.

pronounced /watər/ is instead realized as /warər/. In fact, pronunciations with the /t/ fully released sound hypercorrect to most American speakers and many would consider such a pronunciation not as authentic as the flapped variant.⁹ British speakers, on the other hand, are far more likely to consider a fully-released intervocalic /t/ acceptable. Because this feature both represents a difference between U.S. English varieties and non-U.S. varieties, and is also subject to phonological reduction in casual speech (for U.S. speakers), it emerged as a salient variable for this study.

Given that the focus of Forvo is to generate the most accurate pronunciations possible, it makes sense that occasionally a speaker will go overboard in the carefulness of his/her pronunciation, producing a hypercorrect variant that lacks the natural stress pattern, intonation, or phonetic reduction generally present in a more natural pronunciation of the word. Additionally, the task of producing words in isolation rather than in connected speech means that speakers are even less likely to use a phonologically reduced form. Wardhaugh points out that when pronouncing words carefully, speakers sometimes make a variety of changes to the way that they might produce a word in conversation, such as reintroducing consonants which have a long history of being silent (e.g., /aftm/ for *often*) or altering the stress pattern of a word (e.g., com`parable vs. `comparable). For U.S. listeners on Forvo, the release of intervocalic /t/ where a less careful U.S. pronunciation would use a flap seems to be one of the features most affecting the scores which they assign to given pronunciations.

To study this feature, the 18 tokens with intervocalic /t/ from the data were isolated for further analysis. They were then grouped by score range (low, medium, high) and each range was evaluated for whether the /t/ was flapped or fully released. This initial set of data shows that t-release is indeed affecting the score; for medium-scoring words, the /t/ was almost three times as likely to have been released as flapped, where for the high scoring words, it was slightly more likely to have been flapped. This data indicates that there is indeed at least a slight bias toward the American flapped pronunciation for intervocalic /t/. This data is presented in Table 1.

Score range	Tokens	Realized as /r/	Realized as /t/
Low (<1.9)	0	-	-
Medium (2.0-3.9)	11	3	8
High (>4.0)	7	4	3

Table 1: /t/ release scores.

The data gets even more interesting, however, when we take into account the location of the speaker. As Table 2 shows, although at the medium scoring level, U.S. speakers were producing the released variant for this feature, for the high scoring level, only the British and Australian speakers' pronunciations with released /t/ were scored highly.

Score range	Tokens	Realized as /r/, U.S.	Realized as /r/, Non-U.S.	Realized as /t/, U.S.	Realized as /t/, U.S.
Low (<1.9)	0	-	-	-	-
Medium (2.0-3.9)	11	3	0	2	2
High (>4.0)	7	4	0	0	3

Table 2: /t/ release and speaker locale.

This suggests two things. First, there is indeed a bias among Forvo voters toward the American English standard pronunciation which usually realizes intervocalic /t/ as a flap, as Australian

⁹Too few tokens were available to analyze other medial stops in a meaningful way, however, isolated instances seem to indicate that overly careful pronunciation of other medial stops is viewed by Forvo voters as hypercorrect American pronunciation as well. For example a pronunciation of *Barack Obama* with a very audible aspiration of the /k/ in *Barack* was given a medium score with 4 votes.

and British speakers releasing the /t/ were also given medium-range scores along with the American speakers. However, the data also indicates that the dialect that the speaker is using has an effect on whether the released /t/ is judged as a hypercorrect (and therefore not as good) American English pronunciation, or a standard British/Australian pronunciation. It is unclear whether voters are using other clues from the word, such as vowel quality, to classify a speaker as non-U.S. or if they are using the dialect location map, but it is nevertheless evident that the identification of the speaker as non-U.S. affects the voters' perception of the "correctness" of this variant.

The results for /t/ release are instructive because they show that not only are voters listening to and attending to /t/ release in their evaluations of a given pronunciation, but that their interpretations of this variable seem to change based on other linguistic information about the speaker. The judgment of /t/ release is not uniform; whether a released variant is perceived as correct or hypercorrect is based on other linguistic information available to the listener. If a listener perceives a speaker to be a non-U.S. speaker, it seems they change their expectations accordingly for what makes a correct pronunciation. But what if it is not the speaker who is from the location, but the topic at hand?

3.2 Dialect Locale and Topic Sensitivity

It has been well-documented that given extra-linguistic information about a speaker, listeners are apt to make different judgments about the qualities of the speech they hear than they might in the absence of such information. Listeners have been known to misjudge vowel quality based on which vowels they feel a speaker from their location ought to pronounce rather than which ones he/she did (Niedzielski 1999), and to change impressions about speech rate, speaker character, and importantly, standardness of pronunciation based on whether they are told the speaker is highly intelligent or somewhat slow (Thakerar and Giles 1981). On Forvo, the locale of a registered user is documented and appears on a "dialect map" so that when a word has multiple recorded pronunciations a listener can click on different parts of the map to hear the variation. As shown above in the voters' treatment of released or flapped intervocalic /t/, this information about speaker locale is used by the voters in distinguishing between standard pronunciations in one dialect versus hypercorrect pronunciations in another. A second piece of non-linguistic information seems to also be relevant to Forvo voters' perception of pronunciation, however, and that is the topic of the word or phrase being pronounced. Tables 3 and 4 show score ranges for U.S. and Non-U.S. topics, respectively, with an additional breakdown by the locale of the speakers.

Score range	Tokens ¹⁰	U.S. Speaker	Non-U.S. Speaker
Low (<1.9)	1	0	1
Medium (2.0-3.9)	10	3	7
High (>4.0)	8	7	0

Table 3: Topic sensitivity with U.S. topics.

Score range	Tokens	U.S. Speaker	Non-U.S. Speaker
Low (<1.9)	3	3	0
Medium (2.0-3.9)	4	2	2
High (>4.0)	1	0	1

Table 4: Topic sensitivity with Non-U.S. topics.

For this part of the study, 27 tokens¹¹ were selected from the total set whose topics were readily identifiable as particularly U.S. or non-U.S. in nature. Examples of words classified as U.S.

¹⁰In addition to pronouncing words, registered users can also suggest words which they are interested in hearing pronounced. Given the overwhelming bias in the sample toward American users, it is thus unsurprising that more U.S. topics are represented among the tokens gathered.

¹¹Four pronunciations from a single speaker were excluded from this section of the study because of the problematic existence of two pronunciations in the same recording; this data is explained below.

were locations in the U.S. such as *Nevada*, American actors such as *Tom Cruise*, and U.S. presidents. Words classified as non-U.S. topics were locations in Britain or Australia such as *Melbourne* or *Middlesbrough*, and British celebrities such as footballer *Ashley Cole* or the band *Siouxie and the Banshees*.

It is quite striking to look at the high and low scoring sections of both Tables 3 and 4. For U.S. topics, only U.S. speakers’ pronunciations are represented in the high-scoring range, and no U.S. speakers’ pronunciations are represented in the low scoring range. The same is true of non-U.S. speakers’ pronunciations for non-U.S. topics. This is evidence that where the topic is American in nature, a U.S. English accent is preferred for the word’s pronunciation, and the opposite is true where the topic is notably British or Australian.¹² The contrast between the two sets of middle scores, while still supporting the conclusion that dialect locale preference is dependent on topic, seems also to continue to support the apparent general bias on Forvo towards Standard American accents with the non-US speakers being dispreferred for U.S. topics at a greater than 2:1 ratio versus the 1:1 ratio for non-US topics.

Preference for a U.S. standard pronunciation for a U.S.-centric topic extends further than merely U.S. versus non-U.S. pronunciations. Although there was not sufficient data yet to compare a great deal of pronunciations of non-U.S. English accents that are not from primarily Anglophone countries, one particular example deserves mention. One speaker lists his dialect region as Chicago, IL, but explains in his profile that he grew up in Tokyo with English-speaking parents and as such claims native accented Japanese and English. Among many other Japanese and English pronunciations, this speaker provides pronunciations for four Japanese brands extremely popular in the U.S.: Kawasaki, Toyota, Honda, and Suzuki. In his single recording for each, he pronounces each twice; once using Japanese phonology and stress, the second time, Standard American English. His pronunciations are listed as “English” on the site. Despite the similarity in consonant production and vowel quality of his English pronunciations to other highly-rated American English pronunciations, the presence of his Japanese pronunciations in the same recording seems to be enough to cause his average score for the four recordings to plunge to 1.5. This indicates that Forvo voters likely process these four brands as U.S. companies despite their origin, making this user the only U.S.-located speaker to get low-range scores for words on a U.S. topic.

When we evaluate listener dialect bias, it makes sense to look at particular dialect features such as /t/ release. At the same time, however, the results from this portion of the Forvo study indicate that we must pay attention not only to where the speakers and listeners are from, but also what they are talking about.

4 Conclusions

It is evident that Forvo voters are sensitive to issues of dialect in their evaluations of “correctness” of a given English pronunciation. They attend to dialect features both at the micro level, when evaluating the significance of a released or flapped intervocalic /t/, and also at the macro level by preferring a dialect that matches the locale associated with the topic of the word being pronounced. While the overwhelming presence of U.S.-located voters has so far failed to result in large numbers of votes that discount non-American pronunciations, a slight overall bias toward American English is evident when examining more subtle distinctions, especially among middle-scoring pronunciations.

What is most instructive about the data gathered in this project is that it further indicates the fluidity with which naïve listeners evaluate prestige. In both the instance of intervocalic /t/ release and of topic sensitivity, there was not a single evaluation of the correct pronunciation, but rather scores were heavily influenced by the perceived dialect locale of the speaker. Where there was a match between expectation and pronunciation, such as U.S. speakers flapping intervocalic /t/ or non-U.S. speakers pronouncing “The Beatles,” higher scores were given. This indicates that it is

¹²This same process may be taking place within the voting on American English pronunciations with respect to the Southern dialect. However, the only clear example of this possibly occurring in the data was a high-scoring pronunciation of *South Carolina* where the /au/ diphthong in *South* had been monophthongized to /a/. Because of the lack of tokens, this question was not examined in this paper, but could be the subject of future study.

imperative that the prestige of a given variable not be evaluated without context, as users are likely to assign prestige relative to how right they feel the match is between the dialect they hear and the one they expect.

Further study of English pronunciations within Forvo might focus more heavily on topic sensitivity, particularly as the site's usership continues to increase and more and more words are added and tagged as relevant to U.S., U.K., or Australasian culture. In addition, as more pronunciations are added to the database, more words will inevitably be associated with multiple pronunciations, and the differences in scores for variants of the same word would provide some of the experimental control that is difficult to obtain at this time. If and when the feature counting votes for anonymous pronunciations is added, it would also be worthwhile to examine vote scores for pronunciations where the speaker's location is not readily available on a map; this would make more apparent when a voter is relying on clues within the pronunciation itself to classify the speaker's locale.

Web 2.0-based corpora like Forvo are beginning to offer new opportunities for the perceptual dialectologist to study naïve listener and speaker attitudes in a controlled and empirical way while accessing a huge and diverse subject pool. Given the rate at which Forvo is currently growing and trends in usership growth rates demonstrated by other Web 2.0 sites like YouTube, it seems apparent that, as this database continues to expand, the possibilities for linguistic analysis of its data are virtually limitless.

References

- Bayard, Donn, Ann Weatherall, Cynthia Gallois, and Jeffrey Pittam. 2001. Pax Americana? Accent evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics* 5:22–49.
- Clopper, Cynthia, and David Pisoni. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32:111–140.
- Forvo. URL www.forvo.com.
- Giles, Howard. 1970. Evaluative reactions to accents. *Educational Review* 22:212–227.
- Hartley, Laura. 2002. A View from the West: Perceptions of U.S. Dialects by Oregon Residents. In *Handbook of Perceptual Dialectology*, ed. D. R. Preston, 315–332. Amsterdam: John Benjamins Publishing Company.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Niedzielski, Nancy, and Dennis Preston. 2003. *Folk Linguistics*. New York: Walter de Gruyter.
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18:62–85.
- Preston, Dennis, ed. 2002. *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins Publishing Company.
- Purnell, Thomas, William Idsardi, and John Baugh. 1999. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18:10–30.
- Thakerar, Jitendra, and Howard Giles. 1981. They are - so they spoke. *Language and Communication* 1:255–261.
- Wardhaugh, Ronald. 1999. *Proper English: Myths and Misunderstandings about Language*. Malden, MA: Blackwell Publishers.
- Wolfram, Walt. 1974. *The Study of Social Dialects in American English*. Englewood Cliffs, N.J: Prentice-Hall.

Department of Linguistics
 Georgetown University
 Intercultural Center 479
 Washington, D.C. 20057-1051
jag257@hoyamail.georgetown.edu