Stable Variation vs. Language Change and the Factors that Constrain Them

Shayna Gardiner
University of Toronto

Naomi Nagy
University of Toronto

Follow this and additional works at: https://repository.upenn.edu/pwpl

Recommended Citation

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/pwpl/vol23/iss2/10
For more information, please contact repository@pobox.upenn.edu.
Stable Variation vs. Language Change and the Factors that Constrain Them

Abstract
To better diagnose language change vs. stable variation, we must clarify their differences – a critical endeavor especially for variables that may change very slowly over long time periods, where an Apparent Time approach may not reveal clues to change in progress. Wallenberg and Fruehwald (2013) propose the Continuity Hypothesis: that stable variables should be constrained by at least one continuous factor; we provide a stringent test of this hypothesis, analyzing 38 dependent variables from articles published in Language Variation and Change. Of the 23 ‘changing’ variables analyzed, none was reported to be constrained by continuous factors; of the 8 ‘stable’ variables analyzed, only one was found not to be associated with factors that could be treated as continuous. This significant distinction (Fisher's Exact Test, p
Stable Variation vs. Language Change, & the Factors That Constrain Them

Shayna Gardiner and Naomi Nagy

1 Theorizing Language Change and Stable Variation

Distinguishing language change from variation is a core tenet of variationist sociolinguistics. Language variation is what we see when a language has more than one way of expressing the same thing. That is, variation occurs when a given linguistic element has more than one form. The element with multiple forms is said to be a “variable.” Language change necessitates variation (no language change can happen without it) but the reverse is not always true. Sometimes natural language variation instead remains stable over time. In order to achieve a full understanding of language change vs. variation, therefore, we must understand the differences between stable variables and language change variables. This paper explores data that supports one proposal of how to do this, that put forward by Wallenberg (2013) and Fruehwald and Wallenberg (2013).

Language change can be thought of as two or more forms “competing” for usage after a new form is introduced (Kroch, 2005, 1). Such competition causes the frequency of use of each form to change as the new form replaces the old.

Wallenberg (2013) and Fruehwald and Wallenberg (2013) expand this model further, using it to account for stable variation. They propose that all variation is the result of competing forms. For Wallenberg (2013) and Fruehwald and Wallenberg (2013), therefore, two outcomes are possible in variation: language change or specialization. Specialization itself has two types: specialization for different functions, and partial specialization, which can result in stable variation (Wallenberg, 2013, 2-3). In this model, the majority of variation should be change in progress: the Principle of Contrast (the idea that every two forms should contrast in meaning) means that indefinite competition between forms should be unstable over time (Wallenberg, 2013). Stable variation will be rare; Fruehwald and Wallenberg (2013) propose that the mechanism creating stability is the mathematical character of some extra-grammatical dimension with which the variation interacts. Namely, stable variation will only happen along a continuous dimension. Since there is no single environment in which one variant is favoured, instead there are many environments where the variant’s likelihood of occurring slowly increases incrementally, no full replacement occurs.

Support comes from data from several instances of variation, including the the much-studied -ing variation in modern English, which is widely accepted as stable. (Fruehwald and Wallenberg, 2013, 25) found that this variable is affected by another continuous factor: a nominal-verbal scale. That is, -ing variation is specialized along a nominal-verbal continuum. They also find that it is affected by style, which they argue also exists along a continuous dimension as well (formal-informal). Fruehwald and Wallenberg (2013) also observed historical data from Icelandic and English embedded yes/no questions. Their results indicate that English is an instance of stable variation – and that it is also affected by a continuous factor, prosodic weight. Conversely, the Icelandic data show a change in progress, and results indicate that the Icelandic variation is not conditioned by a continuous factor (Fruehwald and Wallenberg, 2013, 15). Fruehwald and Wallenberg (2013) also found that a significant continuous factor does not preclude discrete factors from influencing variant choice in stable variation scenarios.

In a similar vein to the Fruehwald and Wallenberg (2013) study, Pintzuk and Taylor (2011) previously found support for a similar hypothesis using data from Icelandic and Old English. They proposed that stable variation is influenced by information structure constraints – including the continuous variable phrase complexity – but that change over time is not. That is, they show that syntactic change over time proceeds independently of the type of syntactic variation that occurs due to information status and phrase complexity.

These findings indicate that there is still a gap in our knowledge of stable variation: an answer to the famous “transition” problem, how does a language change – or not – from point A to point B (Weinreich et al., 1968)? There is a lack of consensus on what it means for a variable to be “stable” over time (how do we know it isn’t an extremely slow change, for example), on how to measure and
predict stability vs. change (useful in case where we do not have 100+ years of observable data), and on what kinds of social and linguistic factors can block language change and cause two (or more) forms to stabilize.

This study is a meta-analysis to discover how well this hypothesis, henceforth the Continuity Hypothesis, is supported by additional data. It examines a sample of previous variationist sociolinguistic studies and shows that the predictions of Wallenberg (2013) and Fruehwald and Wallenberg (2013) are supported. Their Continuity Hypothesis is supported; we see stable variables being constrained by factors that exist (or could exist) on a continuum, though these variables may also be governed by discrete/categorical factors. We also see that language change variables are governed only by discrete/categorical factors.

To our knowledge, such a meta-analysis has never been conducted. This one will therefore provide a jumping-off point for creating a fuller theory of the mechanisms behind language change and variation, allowing us to make more numerous and more accurate predictions about what variables will lead to partial specialization (stable variation), full specialization, or complete replacement. We will be better-equipped to predict which scenarios will lead to language change, and which will lead to stable variation.

2 Methods

We examined issues of Language Variation and Change (LVC) 22-27 (2010-2015). By restricting this study to the last five years, we allow time for increased usage Rbrul (Johnson, 2015), a statistical package for multivariate analysis. Rbrul can model continuous factors, while its predecessors such as Goldvarb (Sankoff et al., 2012) could only model categorical factors. Researchers can thus compare whether continuous or categorical models of independent variables better fit their data.

After excluding articles that did not conform to type requirements, the result was 35 entries (tokens) from Vols 24-27. These tokens included only 6 instances where the variable was considered stable. In order to obtain more stable variables to compare to language change variables, we extracted entries from Vols. 22 and 23, only for those papers where the variation was said to be stable, adding two additional stable variable tokens. One additional language change entry is included as well, due to it being from one of the articles from which we retrieved a stable variable. We obtained 38 tokens in total.

We coded each dependent variable in each article for whether it was observed by the LVC authors to be stable variation (partial specialization that never ends in specialization or replacement), change in progress (specialization in categorical contexts, or complete replacement in all contexts), or unknown (including those where the stability of the variable is not discussed at all).

We coded each article’s dependent variable based on whether the paper explicitly referred to it as stable or changing. The only exceptions to this were -ing and t/d deletion, which are widely accepted as stable in Standard North American varieties of English. However, if the article studied a variety of English in which either of these was not stable, it was coded as changing instead.

We prepared tokens for all articles containing a multivariate logistic regression analysis of one or more linguistic variables. Each variable in a paper received its own entry.

For each token, we coded:

1. the variable’s significant independent variables

2. whether any independent variable (factor) was modelled as continuous, or could be modelled as continuous

3. whether style was a significant factor

A separate category was made for the factor style because Wallenberg (2013) and Fruehwald and Wallenberg (2013) argue for a continuous model of style, while many authors treat it as discrete. We excluded the factors age, year of birth, and year of recording, which are by nature continuous. This choice was made because a change over time should by nature be correlated with year of birth and age: older people (i.e., those with earlier birth years) will likely retain more of the old variant,
while the younger generation will bring in more of a new variant. In effect, if there is generational change, age and year of birth will be significant factors, despite being continuous. We also excluded frequency: previous research indicates that it has two apparently contradictory effects on language change. First, frequency can lead to phonetic reduction and/or grammaticalization Phillips (2001, 1984), and therefore language change. Second, many frequently-used forms resist structural change (Diessel, 2007; Bybee and Thompson, 1997; Hopper and Traugott, 1993). That is, frequency has been hypothesized both to cause change and to block it, making it difficult to set a prediction for its effect in this meta-analysis.

We then examined the relationship between whether these LVC studies reported a linguistic change or stable variation and whether any of the independent variables were (or could be) modelled as continuous. Finally, we considered the studies in which the status of the dependent variable, as stable vs. representing a change in progress, was unclear, and suggest how to use our existing results to better understand these less-frequently studied variables.

3 Results

We first discuss the LVC studies which report stable variation. Then we discuss the studies which report change in progress. Third, we consider how best to interpret the studies where the authors did not indicate whether the variable was changing or stable. Finally, we consider options regarding stylistic variation: whether style should be considered as a discrete or continuous variable, or whether it depends on the particular context.

3.1 Stable Variables

8 of the 38 tokens were stable. These are shown in Table 1. Only a single factor from one of these studies was explicitly modelled as continuous: an attitudinal index score, in Haddican et al.’s (2013) study of Northern England vowel-fronting. They found that a more positive attitude towards the region did significantly correlate with increased vowel-fronting (Haddican et al., 2013). Happily for the Continuity Hypothesis, Haddican et al. (2013) also find that vowel variables in this region that are not correlated with the continuous factor are change variables rather than stable.

No other stable variable studies modelled any factors as continuous. But the fact that these studies did not investigate any continuous factors is not necessarily an indication that continuous factors do not exist for these variables. As research continues, we may find more factors to be relevant. Additionally, we may be able to model some currently discretely modelled factors along a continuum: for example, the factor following/preceding sound could be modelled in future work as a continuous range, perhaps along a sonority scale. This avenue is promising: preceding sound was significant in Korean /y/ variation (Jin, 2012), following sound was a significant factor in Raleigh English -ing (Forrest, 2015), following and preceding sound were both significant factors in Canadian t/d deletion (Walker, 2012; Hoffman and Walker, 2010), and following sound was the most influential factor in Appalachian English t/d deletion (Hazen, 2011), all stable variables.

3.2 Language Change Variables

23 out of 38 tokens were described as language change, either completed or in progress, in real and/or apparent time.¹ No factors for any of these variables were treated as continuous. This is not surprising, as the vast majority of the factors studied for these variables were evidently categorical, including noun animacy, gender/number, tense/aspect/mood, region, and sex.

However, three factors from two tokens have the potential to be modelled continuously; we might find evidence for or against the Continuity Hypothesis in these. Becker (2014) does find

<table>
<thead>
<tr>
<th>Variable</th>
<th>Linguistic Factors</th>
<th>Social Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raleigh Eng -ing (Forrest, 2015)</td>
<td>lexical category following sound</td>
<td>education sex</td>
</tr>
<tr>
<td>Northern Eng vowels (Haddican et al., 2013)</td>
<td>preceding sound following sound</td>
<td>sex attitudinal index score</td>
</tr>
<tr>
<td>Korean /y/ (Jin, 2012)</td>
<td>preceding segment syllable context syllable onset (y/n)</td>
<td>gender SES</td>
</tr>
<tr>
<td>English -ing (Evans Wagner, 2012)</td>
<td>following segment</td>
<td>style post-high school transition ethnicity SES</td>
</tr>
<tr>
<td>Hebrew gender (Levon, 2012)</td>
<td>constituent type neutralization number</td>
<td>speaker group speaker sex</td>
</tr>
<tr>
<td>Canadian Eng t/d deletion (Walker, 2012)</td>
<td>preceding phonology morphological status following phonology</td>
<td></td>
</tr>
<tr>
<td>Appalachian Eng t/d deletion (Hazen, 2011)</td>
<td>following phonology preceding phonology morphological type</td>
<td>gender region education ethnicity social class</td>
</tr>
<tr>
<td>Toronto English Chinese &amp; Italian community t/d deletion (Hoffman and Walker, 2010)</td>
<td>phonological context morphological status</td>
<td>ethnicity sex</td>
</tr>
</tbody>
</table>

Table 1: Stable variables and factors studied. Potentially continuous factors are in italics; the factor modelled as continuous is in bold.
that preceding phonological context is significant in her study of the change towards rhoticity in New York English. She is referring to preceding vowel, which can be modelled continuously, e.g., along a sonority scale. However, if we observe the constraint rankings for this data, it is evident that the results are not ordered along a sonority hierarchy: the BURR vowel (high) has the highest factor weight (0.83), then BEER (high; FW 0.53), BAR (low; FW 0.42), BEAR (mid; FW 0.38), and finally BORE (mid; FW 0.29). Similarly, Hoffman and Walker (2010) find that following phonology was significant in the Canadian Vowel Shift in Chinese and Italian communities speaking Toronto English, but in this case “following phonology” was specifically manner of articulation, which is generally seen as discrete.

A second example of a factor from the language change tokens that could be modelled on a continuum is word length, either by number of segments or number of syllables. In this meta-analysis, word length was used to investigate only one variable: the change towards rhoticity in New York City English (Becker, 2014). While it was not modelled as continuous, it was found to be the only internal factor that was not significant in New York City rhoticity.

These findings do not contradict the Continuity Hypothesis: indeed, it would seem from these studies that factors behaving continuously do not constrain change variables at all.

3.3 Making Predictions

The remaining seven tokens are our unknown variables: it was unclear whether they were language change or stable variation. Indeed, several of the papers’ authors explicitly state that their variables’ trajectories remain uncharted, either due to the newness of the variable or a dearth of research on it. There are three “unknown” variables where the authors test continuous or potentially continuous factors.

Weatherholtz et al. (2014) studied to- and for-dative variation, as well as variation between these and non-overt datives. This study did model some factors as continuous: these social factors were represented by Likert scales: interpersonal similarity, participant’s political ideology, speaker’s accent standardness, participant avoids conflict, participant compromises during conflict, speaker’s political ideology, participant dominates during conflict, speaker sounds smart, participant’s accent ideology. Once a clearer picture emerges of this variable’s stable or changing status, we will be able to tell whether it supports the Continuity Hypothesis or not. Since these social factors are continuous and significant, the hypothesis predicts that this factor should be stable over time.

One unknown variable with a potentially-continuous factor is /p/- and /k/-lenition in Murrinh Patha, an Australian Aboriginal language, which was found to be constrained by preceding segment (Mansfield, 2015). As previously suggested, this could be treated as a sonority continuum. Assuming that this factor remains significant when modelled as continuous, the Continuity Hypothesis predicts that this variable will be stable over time. Similarly, a second variable has a potentially continuous factor: in unreleased /t/ in the speech of various’ politicians, preceding sound and following sound were both significant, but modelled as discrete variables (Podesva et al., 2015b). Again, if the sonority continuum is a significant factor for this variable, the Continuity Hypothesis predicts that it will be stable over time. Thirdly, Brown and Riveras (2012) find that Spanish haber verb morphology in Puerto Rican Spanish is constrained by proportion of noun use as subject. This factor was modelled in a ternary way in the paper: high, low, and mid. However, if raw numbers were used instead of binning, this factor could be modelled on a continuum. Since it was found to be significant, the Continuity Hypothesis predicts that this variable should be stable.

3.4 Style as a Factor

Now that we have discussed the potential for several factors to be modelled continuously, we must discuss how to model the factor style. A discrete model of style might be binary, e.g., formal vs. informal, while Biber (1995) and Jankowski (2013) provide several methods by which register can be measured along a continuum. Three studies in our dataset included style as a factor.

\footnote{These tokens were from Scherre and Naro (2014), Weatherholtz et al. (2014), Owens et al. (2013), Podesva et al. (2015b), Schleef (2013), Mansfield (2015), Brown and Riveras (2012).}
Evans Wagner (2012), a study of -ing in adolescents, was the only stable variable study to include style in the logistic regression analysis. It was found to be a significant factor in variant choice. Evans Wagner (2012) treated style as binary in her multivariate analysis (casual vs. careful), but if modelled on a continuum and still significant, these results could support the Continuity Hypothesis. Style was also included as a factor in the FINISH variable in Australian Sign Language (Auslan). This variable is an instance of language change, and style was also modelled as binary (monologue vs. dialogue) (Johnston et al., 2015). However, in this case it was found not to be significant. If style is underlyingly continuous in Auslan as well, this finding would fit in with the observation from our language change data above: that continuous factors do not appear to constrain change variables. However, style was also included as a factor in the study of South African English r-fronting, which is observed to be language change. Here, it was found to be significant. However, as with the previous two studies, style was modelled in a discrete way, narrative vs. casual in this case (Mesthrie, 2012). If style is actually a discrete factor, then this finding does not go against our observation that language change variables do not appear to be constrained by continuous factors. Style being discrete would also not be problematic for the -ing finding either: -ing could still be constrained by another continuous factor.

There is also a third option, of course: that style can be either binary or continuous, depending on the variable. Under this assumption, style is free to affect some variables in a binary or ternary way, and others in a continuous way. Until further investigation is conducted into whether to treat style as binary, continuous, or some combination thereof, we cannot say for certain that style is a valid factor choice for testing the Continuity Hypothesis.

4 Summary and Conclusions

We conclude that our data support the Continuity Hypothesis: all but one of the stable variables are constrained by factors that are or could be modelled as continuous. The only stable variable that is not constrained similarly was not tested for any potentially continuous factors (Levon, 2012). Conversely, we find that no language change variables are constrained by potentially continuous factors (excluding style). As the data indicate, in cases where a potentially continuous factor is tested, it does not constrain the change variable (e.g., word length in (Becker, 2014). Even in those cases where language change variables are constrained by a potentially continuous factor like following/preceding phonology, the factor itself affects the variable in a discrete way rather than specializing along a continuum, bringing additional variable in line with the Continuity Hypothesis. These findings are shown in Table 2; a Fisher’s exact test shows the difference to be statistically significant ($p < 0.0001$).

<table>
<thead>
<tr>
<th>Stable Variables</th>
<th>Change Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>at least one significant continuous factor</td>
<td>7</td>
</tr>
<tr>
<td>no significant continuous factors</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Stable and change variables with continuous factors (excluding style).

There is potential for the Continuity Hypothesis. We find support for the hypothesis in the stable variables’ results: stable Northern vowel variation is constrained by the continuous factor attitudinal index score (Haddican et al., 2013), and there is potential for modelling following/preceding phonology along a sonority continuum. The results from the language change variables are also promising: the single study in this set that used the potentially continuous factor word length (Becker, 2014) showed that it was not affected by that factor, and change variables affected by preceding/following phonology were affected by this factor in a discrete way rather than a continuous one (Becker, 2014; Hoffman and Walker, 2010).

The results of the unknown variables also indicate fertile testing ground. While they do not specifically inform the Continuity Hypothesis, they might do so in the future: any continuous factors tested on variables whose stable or changing nature we do not understand has the potential to
help us make hypotheses regarding the trajectories of those variables. This could be a valuable aid in better understanding little studied languages/variables, e.g., Mansfield (2015). Likewise, the style puzzle opens up another area of research: the nature of factors themselves, and how they interact with and inform the Continuity Hypothesis. This study shows that very few factors are ever modelled as continuous: only two studies in this data set have done so, despite the potential for modelling continuous factors in programs such as Rbrul. Both of these were social factors, but we have indicated two linguistic factors that could be modelled as continuous in future work: word length and following/preceding phonology.

These kinds of contributions, along with new studies that model at least one factor as continuous, would be an excellent contribution to our knowledge and our ability to use the Continuity Hypothesis to make predictions about any variable’s trajectory for stability or change over time. We invite the authors of the LVC articles cited herein to test our predictions and let us know what happens.

References
Johnson, Daniel Ezra. 2015. Rbrul version 2.29: A variable rule application in R.