

COALESCENT ASSIMILATION ACROSS WORD- BOUNDARIES IN AMERICAN ENGLISH AND IN POLISH ENGLISH

KAMIL KAŻMIERSKI

Uniwersytet Adama Mickiewicza w Poznaniu
kamil.kazmierski@wa.amu.edu.pl

EWELINA WOJTKOWIAK

Uniwersytet Adama Mickiewicza w Poznaniu
ew56657@st.amu.edu.pl

ANDREAS BAUMANN

Universität Wien
andreas.baumann@univie.ac.at

Abstract

Coalescent assimilation (CA), where alveolar obstruents /t, d, s, z/ in word-final position merge with word-initial /j/ to produce postalveolar /tʃ, dʒ, ʃ, ʒ/, is one of the most well-known connected speech processes in English. Due to its commonness, CA has been discussed in numerous textbook descriptions of English pronunciation, and yet, upon comparing them it is difficult to get a clear picture of what factors make its application likely. This paper aims to investigate the application of CA in American English to see a) what factors increase the likelihood of its application for each of the four alveolar obstruents, and b) what is the allophonic realization of plosives /t, d/ if the CA does not apply. To do so, the Buckeye Corpus (Pitt et al. 2007) of spoken American English is analyzed quantitatively. As a second step, these results are compared with Polish English; statistics analogous to the ones listed above for American English are gathered for Polish English based on the PLEC corpus (Pęzik 2012). The last section focuses on what consequences for teaching based on a native speaker model the findings have. It is argued that a description of the phenomenon that reflects the behavior of speakers of American English more accurately than extant textbook accounts could be beneficial to the acquisition of these patterns.

Keywords: casual speech phonology, corpus phonology, foreign language acquisition, coalescent assimilation, glottalization

1. Alveolar obstruent + /j/ across word boundaries in English

1.1 Textbook descriptions

Coalescent assimilation (CA) is one of the most conspicuous types of connected speech processes in English and is included in numerous textbook descriptions of English pronunciation, foreign language teaching, and practical phonetics. There are, however, major discrepancies between them.

CA is an assimilatory process of the same kind that historically led to the coalescence of alveolars and /j/ word-internally, in words such as *nature* or *question* (Cruttenden 2014: 308). Alveolar stops and fricatives /t, d, s, z/ are notably liable to changes, and when they occur in word-final position the likelihood of them undergoing various allophonic processes is even higher (Shockey 2003; Jones 2006). The assimilatory processes occurring across word-boundaries are often perceived as attempts to make articulation as easy as possible (Carr 1999; Sobkowiak 2001) and are regarded as natural (Sobkowiak 2001: 84). Simplifying the articulatory gesture, the segment in question becomes similar to the adjacent one (Carr 1999: 16). However, in the case of CA, the directionality thereof appears to be the biggest issue among linguists and it remains unclear.

Some argue that CA is an example of progressive assimilation (Hawkins 1984; Avery and Ehrlich 1992; Roach 2009). Avery and Ehrlich (1992: 87f.) postulate that it is progressive only for stops, as in their analysis it is function words that follow a word-final alveolar stop, such as *your* or *you*, that are the targets of the process. Roach (2009: 110–113) also treats stops and fricatives separately, therefore in both these works CA of /s, z/ + /j/ is claimed to be a regressive process of assimilation of place, as in *those years*. To Carr (1999: 16), assimilation of /s, z/ to their palatal equivalents before /j/ is, indeed, similar to what happens to /t, d/ in the same environment, but there is one major difference that is the main reason for him to separate the two cases of assimilation. In the latter, the plosives are ‘coalesced’ into palato-alveolar affricates, and both features of manner and place change. Since it is not the case for /s, z/, a distinction is made between these two types of sounds.

Sobkowiak (2001) treats CA as a very radical case of regressive assimilation in sandhi context of both place and manner, which results in the production of either post-alveolar fricatives or post-alveolar affricates, and therefore analyzes stops and fricatives together.

Other times, CA is considered to be a bi-directional process, which is referred to as *reciprocal assimilation*. This term can first be found in Bronstein (1960: 212f.), who defines it as a process which occurs when both anticipatory and forward assimilations appear to take place simultaneously. He does not include alveolar stops that coalesce with /j/ in the list of his examples, focusing mostly on fricatives, and does not differentiate between intra-morphemic assimilations and those that occur across boundaries. This is relatively surprising, given that

CA is currently said to be more complete in the case of stops, not fricatives, and yet these are virtually omitted from his analysis. Collins and Mees (2013) who distinguish between place, manner, and energy assimilations, treat CA as a special form of a co-occurrence of both place and manner assimilations and, hence, also treat it as a bi-directional process. They use the term ‘reciprocal assimilation’ and define it as “a two-way exchange of articulation features” (Collins and Mees 2013: 122). For instance, in the case of /d/ + /j/ becoming /dʒ/ both place and manner assimilations apply, changing a sequence of an alveolar obstruent and a palatal approximant into a single segment, a palato-alveolar affricate (Collins and Mees 2013: 122). Reciprocal assimilation is most common with alveolar fricatives, as they are said to merge with any following word that begins with a palatal approximant; for plosives /t, d/ it is typical when the following word is *your* or *you* (Collins and Mees 2013: 125). Therefore, as opposed to Bronstein, under their umbrella term of reciprocal assimilation, they include all alveolars that coalesce with /j/. They also note that this process is purely phonetic and optional. Any other possible phonetic realizations of the sequence in question are not described by the authors. The observation that the process is optional is not an exhaustive treatment of the issue, as there are likely to be both linguistic and extralinguistic factors influencing the likelihood of the application of the process, as optionality does not mean randomness. The present study is an attempt to shed some light onto these.

Aside from problems with the directionality of CA, it is also difficult to decide what triggers this process and what is its target. Shockey (2003: 44-45) agrees with Hawkins’ (1984: 272) postulate that palatal approximants are somewhat more vulnerable to variation than the adjacent ‘stronger’ consonants, such as stops. Thus in the case of fricatives, the underlying alveolar fricative becomes a post-alveolar one because /j/ follows, whereas when it comes to stops, an underlying palatal approximant becomes a post-alveolar fricative. Her analysis suggests that it is not one segment that is the result of the process – it is a sequence of a stop + a post alveolar fricative, and the argument that affricates in English are not single segments but rather sequences of two has been present in the literature for quite some time (cf. Ladefoged 2011). She notes that frequency plays a big role in CA, providing examples where the second word in a sequence is a common word that seems to trigger this process. And indeed, most linguists argue that CA is most likely to operate when the second word in a sequence is a function word, such as *you*, *your*, or *yourself* (Collins and Mees 2013; Wells 2003; Sobkowiak 2001; Zsiga 2013: 50). Sobkowiak (2001: 84) points out that “coalescence is sensitive to the textual cohesion of the word string and function words cohere quite closely with their neighbours”, in line with the conjecture that the presence of a syntactic boundary between words diminishes the likelihood of CA. Cruttenden (2014: 320) postulates that CA is generally found in common expressions, such as *did you* or *would you*, and is in fact so ingrained in those that it may take place even in careful and slow speech. Thus, he emphasizes the role of bigram frequency of a given word sequence.

The same point is made by Collins and Mees (2013: 125), who observe that CA is most common in tag questions and in frequent expressions; so much so that it has found its way even to orthography, with *d'ya* being an informal representation of dialogue.

One other interesting question that remains unresolved as far as CA is concerned is whether it occurs in one or two stages. Although some researchers do not analyze it in too much depth and see it simply as a process that changes alveolars into post-alveolars in the context of /j/, it is not obvious whether it is that straight-forward. The two-step analysis is quite plausible, at least in the case of fricatives (Hawkins 1984, Cruttenden 2014). Acoustic studies seem to suggest that the duration of the post-alveolar fricative that is the result of CA is actually longer than the duration of its equivalent in non-derivative contexts (Cruttenden 2014: 312). Hawkins' (1984: 320) analysis shows that these segments may not be equal in length to geminates, but there's a trace of the lost /j/ that is left, indicating that it operates in the following way: [sj] → [ʃj] → [ʃ].

Even though so commonly encountered, CA appears to be a very complex process, whose directionality and operation stages remain unclear and the question as to what triggers this type of assimilation and what is its target is still a matter of debates.

1.2. Empirical studies

Alveolar obstruents have been studied extensively, not least because of their liability to change. Since English allows for an overlap between adjacent consonants, causing stops to often go unreleased at the end of phrases, especially when they precede an obstruent or a nasal, Davidson (2011) analyzed the variation in stop releases in American English spontaneous speech. The long-held view that alveolars are very susceptible to being unreleased was corroborated, as he found that alveolars are especially likely to be unreleased in pre-pausal position, and out of alveolar stops, voiced variants are actually less likely to be released. Although the study concerned only the realization of stops in a pre-consonantal (pre-obstruent) position, what is of crucial importance is that it aimed to investigate what other processes stops may undergo, aside from the binary choice [±released]. It was found that alveolars in spontaneous speech can be deleted, lenited (spirantized or released as an approximant), and glottalized. The allophonic realization of word-final plosives is not without sociolinguistic significance (cf. Podesva et al. 2015, and references therein), as audibly released allophones /t/ tend to be associated with intelligence and education. One possible realization, glottalization is statistically more likely (but still optional) at prosodically significant locations such as phrase boundaries, utterance boundaries and pitch accents in English, and is characterized by a wide range of variation as far as the rate of its occurrence is concerned (Redi and Shattuck-Hufnagel 2001); the highest rate of glottalization can be found in utterance final position.

This study attempts to investigate the realization of alveolar obstruents /t, d, s, z/ followed by /j/: whether CA is the most common process that occurs in this environment, and what factors influence the likelihood of its application. Additionally, for the plosives /t, d/, other possible phonetic realizations are considered. The investigation is based on a statistical analysis of data drawn from corpora of conversational speech of American and Polish speakers.

2. Methodology: the Data

The corpus used to investigate the patterns present in the speech of native speakers of English is the freely available Buckeye Corpus (Pitt et al. 2007). It comprises over 300,000 words of speech of 40 speakers from Central Ohio, stratified for gender (20 female, 20 male) and age (20 below 30 years of age, 20 over 40 years of age). The corpus is annotated phonetically. It was searched by means of the bundled SpeechSearcher software, allowing phonemic queries of phonemes spanning word boundaries.

The corpus used to investigate the patterns present in the speech of Polish speakers of English is the spoken component of the PLEC Corpus (Peżik 2012). It comprises about 200,000 words of speech produced by both teachers and students (high school and college students). The corpus is annotated orthographically. It was searched by means of ELAN¹ (Sloetjes and Wittenburg 2008), allowing orthographic queries employing regular expressions. The Buckeye data for /t#j/ and /d#j/ were collected by the first two authors, with inter-rater agreement measured as discussed below, and the Buckeye data for /z#j/ were collected by the second author. The Buckeye data for /s#j/, as well as the PLEC data for all four environments were collected by the first author.

2.1. /t#j/ and /d#j/

All instances of /t/ followed by /j/ with an intervening word boundary were retrieved from the Buckeye Corpus using a phonemic query in SpeechSearcher. This yielded 1074 hits, three of which had to be discarded due to annotation errors. The results were exported into a spreadsheet and coded for (a) the presence of a major syntactic boundary, (b) presence of stress on the first word, (c) presence of stress on the second word and, crucially, (d) the phonetic outcome of the [tj] sequence. The presence of a syntactic boundary was determined based on the categorization in Batliner et al. (1998). It was treated as a binary variable, i.e. all sequences were categorized either as spanning a boundary (e.g. those spanning a boundary between clauses or embedded sentences/phrases) or not. The presence of stress was determined by auditory

¹ ELAN Linguistic Annotator Version 4.9.3. Max-Planck-Institute for Psycholinguistics Nijmegen, The Netherlands. <http://tla.mpi.nl/tools/tla-tools/elan/>

inspection. The phonetic outcome was based on the inspection of the spectrograms of the sequences in Praat (Boersma and Weenink 2016) according to the protocol visualized in Figure 1 (see Appendix 1). The phonetic outcome, then, was categorized as one of the following: RR (regular release), U (unreleased), D (deletion), G (glottalization) or CA (coalescent assimilation).

After an initial training session, the results of the coding of the first 100 items were compared to assess inter-rater agreement. The results are presented in Column A of Table 1 (see Appendix 2). The agreement with regard to syntactic boundary was almost perfect, but agreement with regard to the phonetic outcome was moderate, and so in need of improvement, and agreement with regard to stress was slight. Therefore, another training session was applied, and agreement was measured again. The results on the second batch of items is presented in Column B of Table 1 (see Appendix 2).

Crucially, after the second training session, agreement on **outcome** rose from ‘moderate’ to ‘substantial’. Though agreement on **boundary** fell, it was still ‘substantial’. The agreement regarding these two variables was therefore deemed sufficient to proceed with the coding. Coding of stress was abandoned due to ‘slight’ agreement both on **w1_stress** and **w2_stress**. No subsequent coding involved stress. All items were subsequently divided in half, the first half being coded by the first author and second half by the second author.

After coding, the resulting dataset was enriched with the following variables: gender and age of the speaker, the grammatical status of word 1 and word 2 (function word vs. content word), whether [t] is part of a cluster or a single coda consonant, whether word 2 is you (or one of its relatives, i.e. your, yours, yourself), frequency of word 1, frequency of word 2, and bigram frequency of the word 1 word 2 sequence. The frequency data was retrieved from the Buckeye Corpus.

When it comes to the PLEC Corpus, the sequence had to be retrieved by means of orthographic queries. For /t#j/ these were: *. *t y.**, *. *t u.**, *. *t eu.**, *. *te y.**, *. *te u.**, *. *te eu.**, *. *ed y.**, *. *ed u.**, and *. *ed eu.**. These searches yielded 879 hits. Just above half of them, however, i.e. 444 had to be discarded. The largest part of the rejectamenta is constituted by the second word being <yyy>, a Polish-style hesitation noise. Two other reasons are due to the vagaries of English spelling: <-ed> often stands for [ɪd], and not for [t], and word-initial <u> often stands for /ʌ/ rather than /ju:/. Finally, audio was misaligned with annotation in some cases.

After the results have been exported, the dataset was further coded in a manner strongly analogous to the one described for the Buckeye data above. There were three differences, however: a) with regard to extra-linguistic data, speakers were coded as either teachers or students, b) the frequency data was retrieved from the PLEC Corpus c) one more value of the **outcome** variable was added, namely V (voiced). This last differences warrants a comment. It transpired during the coding process that certain instances of [t] were released and fully voiced. Such a realization had not been conspicuous in the initial

exploration of the Buckeye Corpus, and therefore it was not included in the initial coding scheme.

The analysis proceeded in a similar fashion for /d/ followed by /j/ with an intervening word boundary, with necessary modifications: for the PLEC Corpus, the orthographic queries were adjusted to search for /d/, and the V (voiced) value of the **outcome** variable was not applicable. There were 795 hits in the Buckeye Corpus. 12 had to be rejected as they were due to annotation errors, bringing the final number down to 783. In the PLEC Corpus, the orthographic queries were: *. *d y.**, *. *d u.**, *. *d eu.**, *. *de y.**, *. *de u.**, and *. *de eu.**. They yielded 838 hits. 321 had to be rejected, as they included the hesitation noise <yyy>, <ed> occasionally stood for [t], <u> occasionally stood for /ʌ/, or they suffered from a misaligned annotation or a missing recording. This brought the final number of datapoints from PLEC down to 517. The inter-rater agreement for both **outcome** and **boundary** was ‘almost perfect’ and so deemed sufficient after the comparison of the first 100 items (see Table 2/Appendix 2 for details). Consequently, no further training was deemed necessary, and the items were divided similarly as those for /t#j/, with the first author coding the first half and the second author the second half.

2.2. /s#j/ and /z#j/

For the fricative + glide sequences, the **outcome** variable was binary. The phonetic outcome was treated as either undergoing coalescent assimilation or not. The coding is agnostic about whether the glide is fully merged with the fricative or whether it only causes the preceding fricative to assimilate and remains in place. All instances where the realization of the alveolar fricative differed from canonical (confirmed both acoustically, by the concentration of high intensity in lower frequencies than canonically, and auditorily) were coded as showing assimilation.

The phonemic query for /s#j/ in the Buckeye Corpus yielded 373 hits. Five of them had to be discarded due to annotation errors, bringing the final number down to 368. The orthographic queries in the PLEC Corpus were *. * s y.**, *. *s u.**, *. *s eu.**, *. *ce y.**, *. *ce u.**, *. *c eu.**, *. *se y.**, *. *se u.**, *. *se eu.**, *. *x y.**, *. *x u.**, *. *x eu.**, *. *xe y.**, *. *xe u.**, and *. *xe eu.**. The number of hits was 679. A number of queries, namely those starting with *. *s* and *. *se*, yielded a large number of words ending in /z/ which had to be discarded. Together with rejections due to reasons already mentioned for plosives above, the final number of datapoints sunk to 368.

The phonemic query for /z#j/ in the Buckeye Corpus yielded 532 hits. Eight of them had to be discarded due to annotation errors, bringing the final number down to 524. The orthographic queries in the PLEC Corpus were *. * s y.**, *. *s u.**, *. *s eu.**, *. *se y.**, *. *se u.**, *. *se eu.**, *. *z y.**, *. *z u.**, *. *z eu.**, *. *ze y.**, *. *ze u.**, and *. *ze eu.**. The number of hits was 825. After the rejection due to

spelling standing for [s] rather than [z], as well as other, already discussed reasons, the final number of datapoints sunk to 177.

Outcome and **boundary**, as well as the remaining information, was provided in the same way as for the stop + glide sequences described above.

2.3. Summary

An overview of the results of coding with regard to **outcome** is presented in Figure 8-15 (see Appendix 1).

3. Statistical analysis: procedure

As described in the previous section, the phonetic outcome is expected to depend on multiple factors. Hence, we adopted a multivariate regression-modeling approach. For each sequence type (/t#j/, /d#j/, /s#j/, /z#j/) and each corpus (Buckeye and PLEC) a separate model was fitted to the respective data. Since the phonetic outcome of /s#j/ and /z#j/ did not show any variation in the PLEC-based data, a total of 6 models were computed. All computations were done in R (R Core Team 2013). In the following three subsections, we shall have a closer look at the statistical modeling procedure. The respective results will be presented in the subsequent section.

3.1. Predictor variables and data transformation

In all models a single dependent variable, the phonetic outcome (**outcome**), subject to multiple predictor variables and interactions among predictor variables was implemented. The range of values of the dependent variable obviously differs depending on the sequence type. The phonetic outcome of both /t#j/ and /d#j/ is multinomial with the possible realizations **rr** (regular release), **g** (glottalization), **d** (deletion), **u** (lack of audible release, or unreleased), and **ca** (coalescent assimilation), as described previously. Regular release was treated as baseline category in the model. In contrast, the phonetic outcome of /s#j/ and /z#j/ represents a binary variable, since the fricative can be either palatalized (**p**) or not (**n**, baseline category). This obviously determines the model family to be worked with. For the former sequence types multinomial logistic regression models were employed, while in the latter case binary logistic regression models were sufficient. This shall be covered in more detail in Section 3.1 below.

The predictor variables are **gender** (binary, baseline: male), **w1_grammar** (binary, baseline: lexical), **w1_freq** (continuous), **w2_grammar** (binary, baseline: lexical), **w2_freq** (continuous), **w2_you** (binary, baseline: not *you*), **w1_w2_freq** (continuous), **boundary** (binary, baseline: no boundary), and **cluster** (binary, baseline: no cluster). The Buckeye-based data also include the variable **age** (binary, baseline: young), while the PLEC-based data include **role**

(binary, baseline: student). Both variables allow for a similar, albeit not identical, interpretation in that students are typically not only younger than teachers, but also less proficient in English. In addition to these variables, **speaker** was included as a cluster variable whenever possible.

Frequency data typically exhibit skewed distributions, which may have unfavourable effects on the validity of the model, and the present data sets made no exception. Hence, all frequency variables (**w1_freq**, **w2_freq**, **w1_w2_freq**) were Box-Cox transformed (Hyndman and Khandakar 2008, Box and Cox 1964). The respective transformation coefficients are shown in Table 3 (see Appendix 2). A visual inspection of pre-transformation and post-transformation normal quantile-quantile plots (not shown) revealed that the transformations lead to more normal-like distributions in all relevant cases. For notational simplicity, **w1_freq**, **w2_freq** and **w1_w2_freq** shall denote the respective transformed variables for the remainder of this paper.

3.2. Collinearity

Due to the nature of the selected predictor variables, collinearity effects were to be expected, which have a negative impact on the appropriateness of the fitted multivariate models. For example, **w2_you** and **w2_freq** in almost all cases are closely related in the present data, since the function word *you* scores high token frequencies. In order to avoid collinearity effects among sets of predictor variables, we proceeded as follows. First, for each set of variables involved in the model, pairwise correlation coefficients (Pearson's r) were determined. The correlation coefficients were used to employ a hierarchical clustering technique, in which highly correlated variables ($r > 0.7$, cf. Booth et al. 1994) were defined as forming a 'correlated cluster' (Ward clustering based on correlation similarity). For each correlated cluster a representing variable was chosen by determining the center of the cluster (in terms of correlation similarity) and selecting the one variable closest to the center (Dorman et al. 2012). If there was no unique choice, i.e. if multiple variables in a cluster had an equal distance from the cluster center, a best-regressor (AIC data-snooping) approach was applied to the candidate variables (*ibid.* 33). This was done in the following way. For each candidate, a univariate model together with its AIC (Akaike information criterion) was computed based on the data. The candidate which scored the lowest AIC was then chosen as the representing variable for the whole cluster to be included in the multivariate model. All other variables from the cluster were discarded and hence not included into the multivariate model. The advantage of this approach over PCA (principal component analysis) clearly is the more accessible interpretation of the predictor variables at work, although PCA arguably produces less biased analyses. Plots of the computed cluster trees are shown in Figure 2-7 (see Appendix 1).

3.3. Model selection

For the analysis of /t#j/ and /d#j/ multinomial multivariate mixed-effects regression models were used (**multinom** function from the **nnet** package; Venables and Ripley 2002). In contrast, cumulative logit-link mixed models were fitted to the /s#j/ and /z#j/ data (**clm** and **clmm** functions from the **ordinal** package; Christensen 2015), which are formally equivalent with mixed-effects logistic regression models in the case of binary outcome variables. The advantage of the latter approach is that random effects can be easily included into the model. This allows for an implementation of **speaker** as a random-intercept cluster variable.

We opted for a top-down model nesting approach. This means that after removing strongly correlated variables from the analysis (see previous subsection), all remaining variables and plausible interactions are built into the model in order to be subsequently removed until the model represents a (local) optimum with respect to a goodness-of-fit criterion. For pairwise model comparisons, a somewhat conservative model-selection procedure was employed. The primary criterion for model selection was AIC (Akaike information criterion), which weights the amount of explained variation against the number of predictor variables in the model. In a nested pair, the model with the lower AIC value is favored. However, if the smaller model had a lower AIC than the larger model but did not show a significant improvement in terms of a likelihood-ratio test (at a canonical significance level of 0.05), the larger model was preferred. In the present analysis, the initial models included all variables that were not removed by the collinearity-prevention procedure described above plus interactions among **freq_w1** and **grammar_w1**, and **freq_w2** and **grammar_w2**, which were considered as plausible, as fixed effects. The significance of the **speaker** variable as random effect was tested analogously.

4. Statistical analysis: results

In the following, the resulting regression models are presented. For the sequence types /t#j/ and /d#j/ two models were fitted, respectively, one with the Buckeye data and one with the PLEC data. For the other two sequence types (/s#j/, /z#j/) only the Buckeye data could be made use of.

4.1. /t#j/

In the Buckeye data (see Table 4/Appendix 2), two highly correlated clusters were observed: first **w1_w2_freq** and **w1_freq**, and second **w2_you**, **w2_freq**, and **w2_grammar**, of which **w1_freq** and **w2_you** were chosen as representing variable. The model reveals a significant negative impact of **boundary** and a significant positive impact of **w2_you** (and its correlates) on coalescent

assimilation (**ca**). In addition, boundary exhibits a negative impact on glottalization (**g**) and deletion (**d**), while **cluster** has a discriminating effect: it favors deletion, but inhibits glottalization and lack of audible release (**u**).

The PLEC data show a slightly different picture (see Table 5/Appendix 2). Two highly correlated clusters were observed: first **w1_w2_freq**, **w1_grammar**, and **w1_freq**, and second **w2_you**, **w2_freq**, and **w2_grammar**, of which **w1_grammar** and **w2_you** were selected. In line with the AE data, the model shows a significant and substantial negative impact of **boundary** and a significant positive impact of **w2_you** (and its correlates) on coalescent assimilation (**ca**). However, in contrast to the AE data, **cluster** also has a positive impact on coalescent assimilation as well as on deletion, but shows no other significant influence on the outcome. Interestingly, **u** seems to be correlated with **gender** and **role**. Apparently, voicing (**v**) does not depend on any of the included variables.

4.2. /d#j/

The highly correlated clusters in the AE /d#j/ data (Table 6/Appendix 2) were first **w1_w2_freq**, **w1_grammar**, and **w1_freq**, and second **w2_you**, **w2_freq**, and **w2_grammar**, of which **w1_freq** and **w2_you** were selected. The results do not differ much from the AE /t#j/ model, save for a frequency driven tendency of deletion in the first word, and slight gender effects (women favoring **ca** and men favoring **u**).

The clusters in the PLEC /d#j/ data (Table 7/Appendix 2) were found to be almost the same as in the case of /t#j/ (**w1_w2_freq** substituted by **w1_freq**), so that the respective representing variables remain **w1_grammar** and **w2_you**. Only three possible outcomes were observed, all of which show interesting interactions with the predictor variables. Coalescent assimilation is to a large extent positively influenced by **w2_you** and negatively influenced by **boundary**. Slight negative effects on **ca** can again be observed by the presence cluster. Deletion seems to be largely driven by the grammar status, in that grammatical items impede deletion. Nevertheless, it can be seen from the significant interaction coefficient that this only holds for rare grammatical items. Two remarks are appropriate at this point. First, the outcome **u** shows a number of significant interactions with the predictor variables; however, these should be treated with caution since only 3 observations were made in this category. Second, the intercept of the **ca** outcome outweighs any of the coefficients in strength. While this is not in itself problematic, it might simply reflect the general weakness of linear models that they cannot be entirely satisfactorily applied to non-monotonous problems.

4.3. /s#j/

The /s#j/ data (Table 8/Appendix 2) only showed a single correlated cluster consisting of **w2_grammar**, **w2_freq**, and **w2_you**, of which again the last variable was chosen as representing factor. As discussed above, the only possible alternative to no adaptation was palatalization (**p**). Older speakers tend to palatalize more frequently, and bigram frequency positively increases the rate of palatalization. As in the case of coalescent assimilation in the previous models, boundary has a negative impact on palatalization. The random intercept was shown to improve the model significantly (likelihood-ratio test, $p < .001$) from which it can be concluded that the speakers did not behave homogeneously.

4.4. /z#j/

Finally, the analysis of the /z#j/ data (Table 9/Appendix 2) looks as follows. The highly correlated clusters were **w1_w2_freq**, and **w1_freq**, and again **w2_you**, **w2_freq**, and **w2_grammar**, of which **w1_w2_freq** and **w2_you** were selected. As in the previous case, palatalization is more likely in the speech of older speakers, and significantly less likely if there is a boundary. In line with most of the previously presented models, **w2_you** promotes assimilation. Keep in mind that this might also be caused by the grammatical status and/or frequency of *you*, both of which are highly correlated with this function word. As in the previous model, a likelihood-ratio test ($p < .001$) revealed that the **speaker** grouping-variable has a significant effect on the phonetic outcome, so that there are speaker specific base levels of applying palatalization.

5. Discussion

5.1. General discussion

For all four models fitted to the Buckeye data, the presence of a syntactic boundary decreases the likelihood of application of CA. It uniformly makes a difference, then, whether the two words are in the same syntactic unit or not. The likelihood that CA will occur is increased, on the other hand, for three out of four Buckeye datasets (with the exception of /s#j/) if the second word in the sequence is *you* (or one of related words). In other words, CA does indeed go hand in hand with *you*. A factor which has made it into all models, and whose impact has reached significance for the plosive + /j/ sequences is the presence or absence of a cluster. It is a factor that usually goes unmentioned in discussions of CA, but is not without relevance here. For both /t#j/ and /d#j/, if the first word ends in a cluster, then deletion is likely to occur. This deletion, then, takes apart the context for CA to take place. The frequency and grammatical status of the

first word in the sequence, pace Avery and Ehrlich (1992), do not seem to impinge on CA. It does influence the likelihood of deletion, though (the word which fed this result is *and*). A sociolinguistic factor which has an influence on the application of CA with both fricatives is age. Being over 40 increases the likelihood of a Buckeye speaker to apply it. An interesting finding is the prevalence of glottalization in the context in which CA could occur. Glottalization was in fact by far the most frequent outcome for /t#j/ in Buckeye. A vast number of contexts, then, in which CA could occur results in a [ʔ] sequence instead. And so if for a given sequence CA does not apply, glottalization is more likely than regular release.

Comparing the models fitted to the plosive + /j/ sequence in Buckeye to the results in the PLEC corpus, the similarity of behavior of two factors can be noticed. The presence of a syntactic boundary decreases the likelihood of application of CA in the two datasets from PLEC, just as it does in Buckeye. The other similarity is that the identity of the second word, i.e. whether it is *you* or not increases the likelihood of occurrence of CA in PLEC.

One of the most striking differences between the Buckeye data and the PLEC data is that the outcome that appears most often in PLEC is regular release. This stands in stark contrast to Buckeye, where glottalization was the most frequent. Consequently, Buckeye and PLEC speakers differ a great deal as to what they do when they do not opt for CA. In Buckeye, glottalization is the most likely outcome, and in PLEC, it is regular release. Additionally, there is one more process in PLEC, unattested in Buckeye, namely voicing of /t/. This could be seen as evidence of Poznań-Cracow voicing, i.e. of replacing voiceless obstruents with their voiced counterparts before sonorants, in the speech of PLEC speakers.

For /s#j/ and /z#j/ in PLEC no statistical modeling could be applied as there was very little (for /s#j/) or no (for /z#j/) variance in the outcome that could be modeled. This is not an uninteresting finding in its own right, though. PLEC speakers do not assimilate the sequences with fricatives, something that Buckeye speakers do very often. This is a very marked contrast.

5.2. Implications for SLA/Pronunciation teaching

To begin with, there is no consensus as to whether connected speech processes should be taught to foreign language students. On one end of the spectrum Ladefoged and Johnson (2011: 111) proclaim that “[f]oreigners who make insufficient use of them sound stilted”, and so they advocate teaching students to use the processes in production. Others stress the importance of training to recognize connected speech processes in perception, rather than production, as this is necessary to easily understand native speakers (Bowen 1975: 163; Sobkowiak 2001). And on the other end of the spectrum are those who do not see it as important in teaching. Roach (2009: 113) says “[m]uch more could be said about assimilation but, from the point of view of learning or teaching

English pronunciation, to do so would not be very useful. It is essentially a natural phenomenon that can be seen in any sort of complex physical activity [...]. The appeal to naturalness suggests that connected speech processes do not have to be taught because they will be ‘naturally’ picked up by students anyway. Another line of argumentation against teaching them is that they arguably make speech comprehension more difficult, at least for other non-native speakers, and as English learners interact with other non-native speakers more often than with native speakers, employing features such as assimilations would be counterproductive (Jenkins 2000).

If one starts with the assumption that the goal of pronunciation teaching is to assist learners in attaining pronunciation patterns as close to a native-speaker model as possible, then the comparison of the results in Buckeye and in PLEC has certain consequences. When it comes to both plosive + /j/ sequences, Polish and American speakers are somewhat similar in their sensitivity to factors such as syntactic boundary and identity of the second word in the application of CA. Where the most conspicuous differences transpire is in the alternatives to CA for the realization of these sequences. The major difference between results in the two corpora when it comes to /t#j/ is the prevalence of regular release in the speech of Polish speakers, in stark contrast to native speakers. The native speakers in the present study are speakers of American English, and for many learners of English in Poland, especially those majoring in English (which is true of many speakers in the PLEC corpus) the reference accent is General British. Strictly speaking, then, an investigation of a corpus representative of British English would be necessary to enrich the picture. Since glottalization is even better attested for British English than for American English accents, however, there is reason to believe that the rates of regularly released /t/ exhibited by Poles would be higher than those of British speakers. The major differences when it comes to /d#j/ is the prevalence of deletion in the speech of Americans, where Poles again consistently go for regular release. At any rate, since CA does not occur whenever the relevant sound sequence is available, and since there are other outcomes possible in this environment, then students should be made aware of both these facts. That CA is more likely in some contexts than in others has to be taught, so that learners do not apply CA too often, and the alternatives to CA have to be taught, since not applying CA results for Polish speakers largely in regular release, which does not match the pattern of native speakers.

When it comes to the fricative + /j/ sequences, there actually is a marked difference with regard to the application of CA between Americans and Poles, in that Poles essentially do not apply it. The implications are somewhat complicated, though, by the fact that CA in fricative + /j/ sequences is associated with old age. In educational contexts with the predominance of students below the age of 40, drawing attention to CA might not be desirable.

If, following Jenkins, one were to look on assimilations as deleterious for communication in international contexts, then there are consequences with regard to /t#j/ and /d#j/, in that Poles would have to suppress the application of

CA which they do exhibit. For example, special attention would have to be paid to avoid CA in sequences such as *would you* or *could you*, which yield high rates of CA in the PLEC corpus. The speech of Polish learners with respect to /s#j/ and /z#j/ would require no intervention whatsoever.

References

- Avery, P. and S. Ehrlich. 1992. *Teaching American English pronunciation*. (Oxford handbooks for language teachers). Oxford, New York: Oxford University Press.
- Batliner, A., Kompe, R., Kießling, A., Mast, M., Niemann, H. and E. Nöth. 1998. M[^] Syntax + Prosody: A syntactic±prosodiclabelling scheme for large spontaneous speech databases. *Speech Communication* 25, 193–222.
- Booth, G. D., Niccolucci, M. J. and E. G. Schuster. 1994. *Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation*. – US Dept of Agriculture, Forest Service.
- Bowen, J. D. 1975. *Patterns of English pronunciation*. Rowley: Newbury House.
- Box, G. E. P. and D. R. Cox 1964. An analysis of transformations. *Journal of the Royal Statistical Society Series B* 26(2). 211–252.
- Bronstein, A. J. 1960. *The pronunciation of American English*. New York: Appleton-Century-Crofts.
- Carr, P. 1999. *English phonetics and phonology. An introduction*. Malden: Blackwell.
- Christensen, R. H. B. 2015. ordinal - Regression Models for Ordinal Data. R package version 2015, 1-21. Available from: <http://www.cran.r-project.org/package=ordinal/>.
- Collins, B. and I. M. Mees 2013. *Practical phonetics and phonology: A resource book for students*. New York: Routledge.
- Cohen, J. 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70, 213–220.
- Cruttenden, A. 2014. *Gimson's pronunciation of English*. London: Routledge.
- Davidson L. 2011. Characteristics of stop releases in American English spontaneous speech. *Speech Communication* 53, 1042–1058.
- Hyndman R. J. and Y. Khandakar. 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 26(3), 1–22.
Available from: <http://ideas.repec.org/a/jss/jstsof/27i03.html>.
- Jenkins, J. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.
- Ladefoged, P. and K. Johnson. 2011. *A course in phonetics*. [6th edition]. Boston: Wadsworth.
- Peżik, P. 2012. Towards the PELCRA Learner English Corpus. In P. Peżik (ed.), *Corpus data across languages and disciplines*, 33–42. Frankfurt am Main: Peter Lang.
- Pitt, M.A., Dille, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and E. Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University.
- Podesva, R. J., Reynolds, J., Callier, P. and J. Baptiste. 2015. Constraints on the social meaning of released /t/: A production and perception study of U.S. politicians. *Language Variation and Change* 27, 59–87.
- Redi L. and S. Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29, 407-429.
- Boersma, P. and D. Weenink. 2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.14. Available from <http://www.praat.org/> [Accessed:] 11 February 2016

- Roach, P. 2009. *English phonetics and phonology: a practical course*. Cambridge: Cambridge University Press
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Shockey, L. 2003. *Sound patterns of spoken English*. Malden: Blackwell Publishing.
- Sloetjes, H. and P. Wittenburg. 2008. Annotation by category — ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Sobkowiak, W. 2001. *English phonetics for Poles*. Poznań: Wydawnictwo Poznańskie.
- Venables, W. N. and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Zsiga, E. C. 2013. *The sounds of language. An introduction to phonetics and phonology*. Malden: Wiley-Blackwell.

See Appendices on the following pages.

Appendix 1

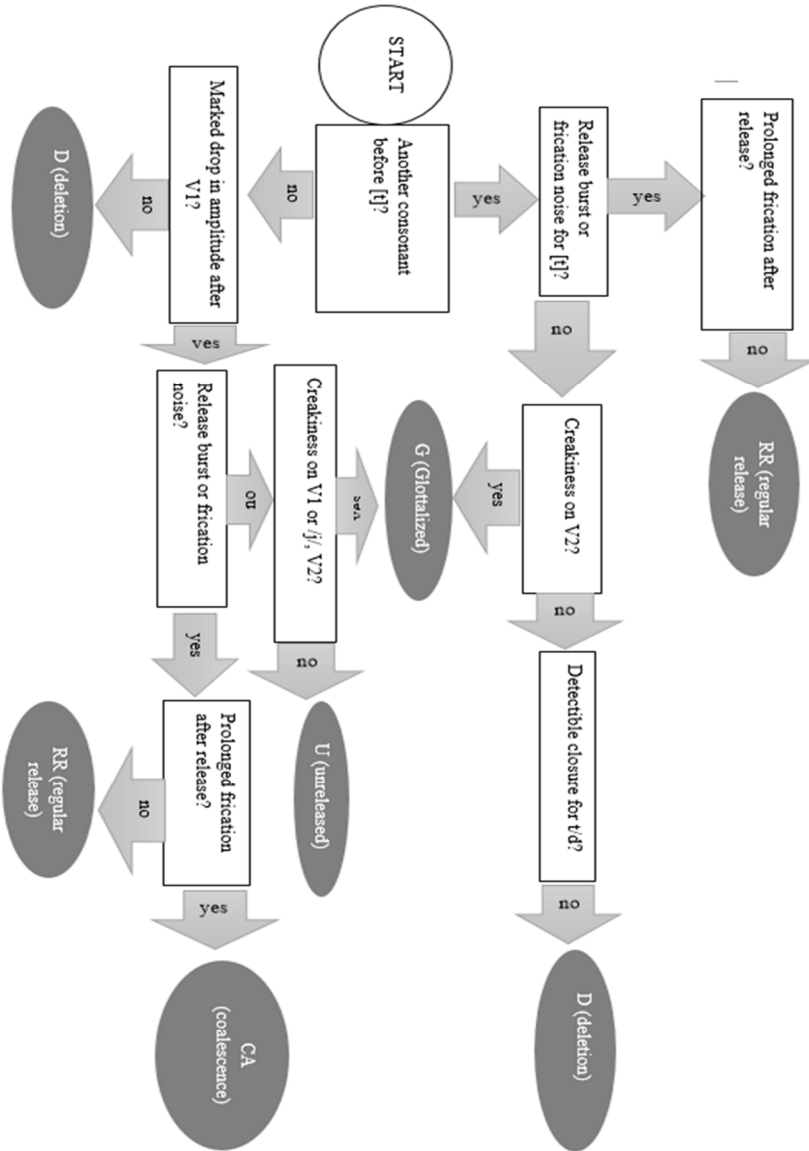


Figure 1. Data coding protocol for **outcome** for /t#j/

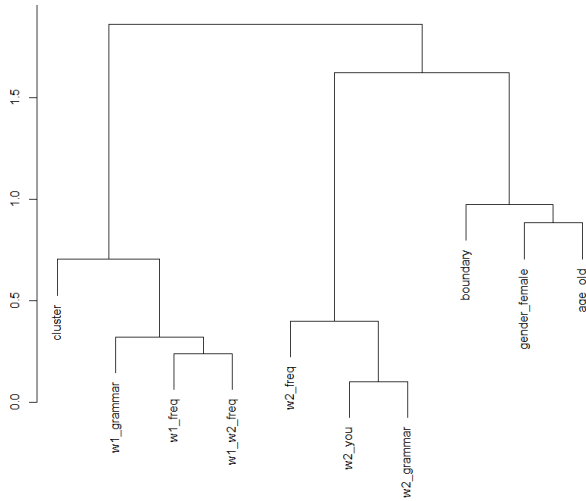


Figure 2. Hierarchical clustering of the predictor variables for outcome of /t#j/ based on the Buckeye data with respect to correlation similarity.

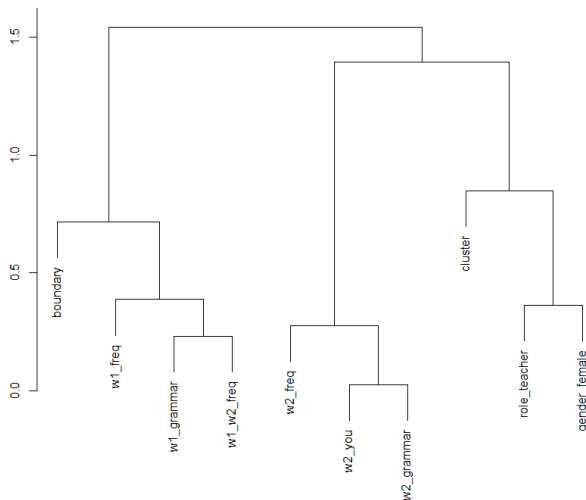


Figure 3. Hierarchical clustering of the predictor variables for outcome of /t#j/ based on the PLEC data with respect to correlation similarity.

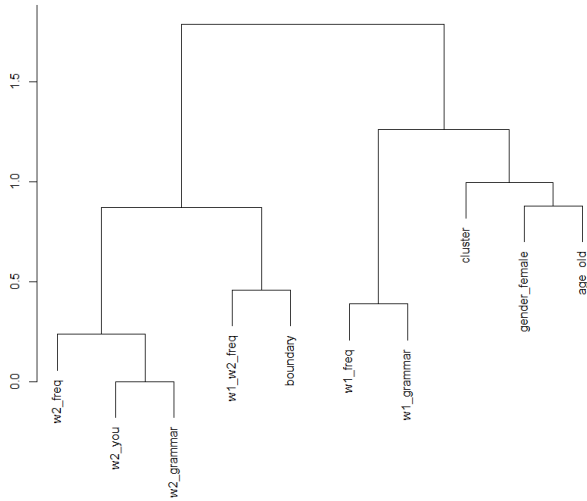


Figure 6. Hierarchical clustering of the predictor variables for outcome of /s#j/ based on the Buckeye data with respect to correlation similarity.

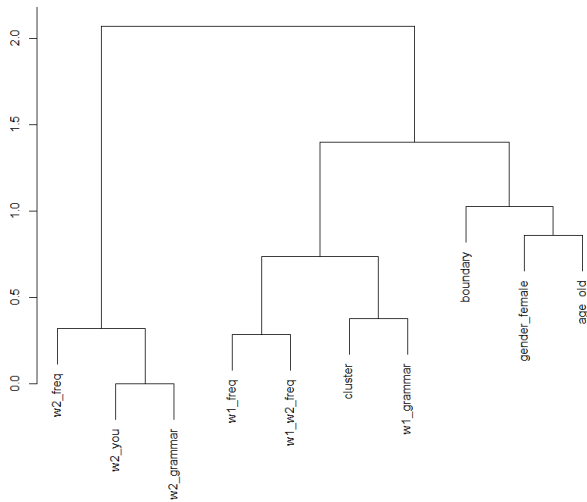


Figure 7. Hierarchical clustering of the predictor variables for outcome of /z#j/ based on the Buckeye data with respect to correlation similarity.

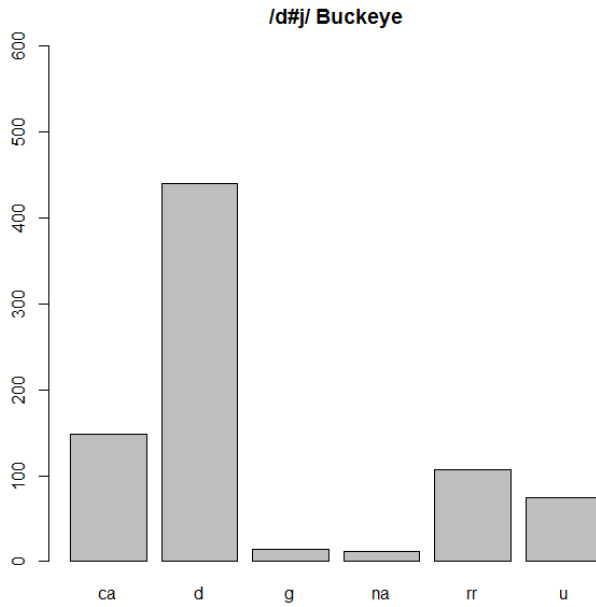


Figure 8. An overview of the results of coding with regard to outcome of /d#j/ (Buckeye data).

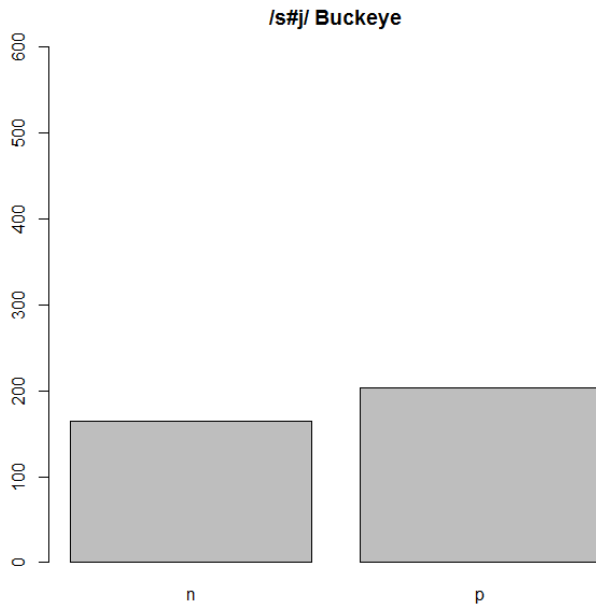


Figure 9. An overview of the results of coding with regard to **outcome** of /s#j/ (Buckeye data).

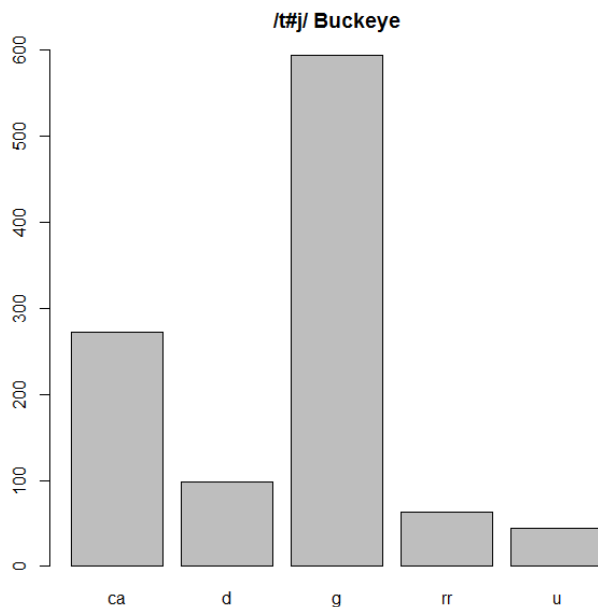


Figure 10. An overview of the results of coding with regard to **outcome** of /t#j/ (Buckeye data).

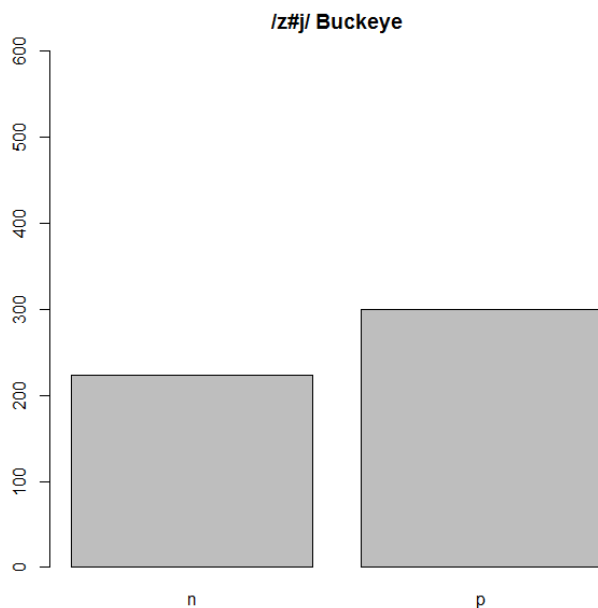


Figure 11. An overview of the results of coding with regard to **outcome** of /z#j/ (Buckeye data).

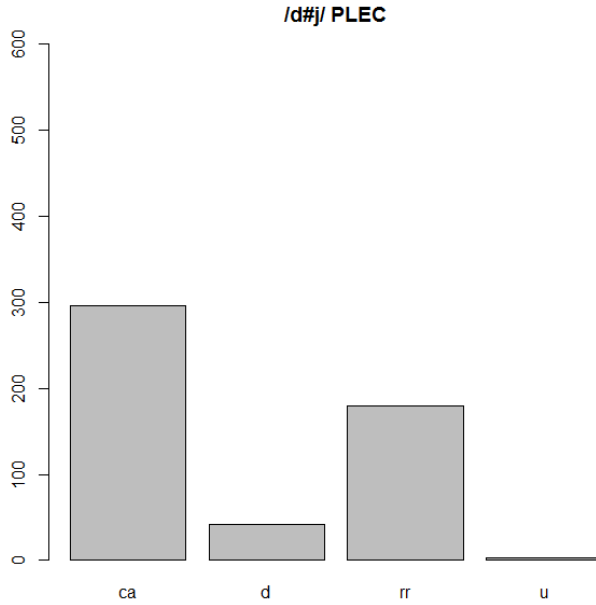


Figure 12. An overview of the results of coding with regard to **outcome** of /d#j/ (PLEC data).

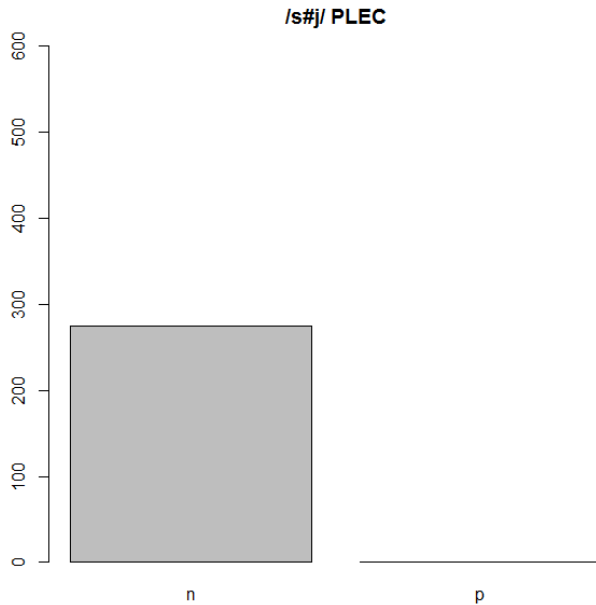


Figure 13. An overview of the results of coding with regard to **outcome** of /s#j/ (PLEC data).

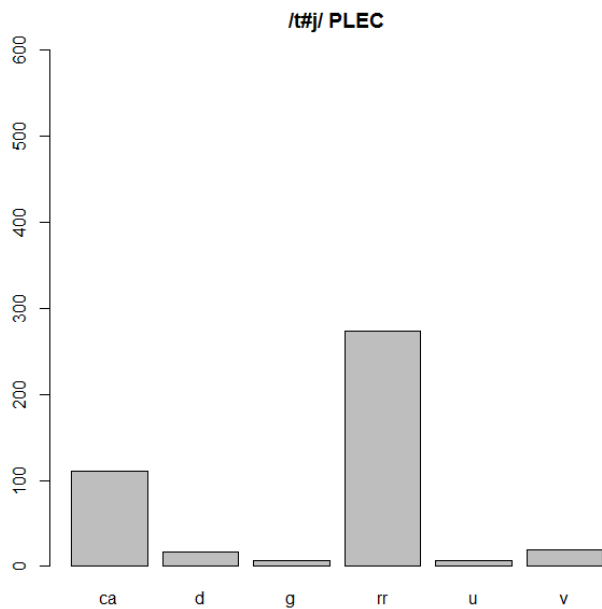


Figure 14. An overview of the results of coding with regard to **outcome** of /t#j/ (PLEC data).

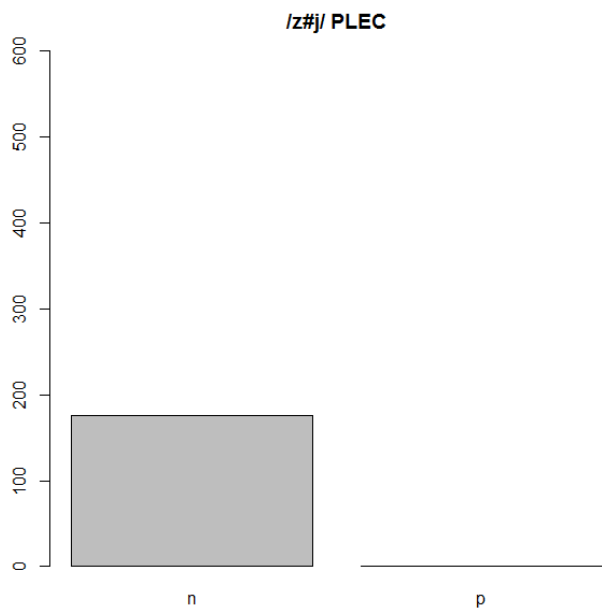


Figure 15. An overview of the results of coding with regard to **outcome** of /z#j/ (PLEC data).

Appendix 2

Table 1. Inter-rater agreement for /t#j/ for items 1-100 (Column A) and 101-200 (Column B). Joint probability of agreement (in percentages) is followed by Cohen's Kappa (Cohen 1968). The second of these two measures was taken as decisive, as it is a more robust measure, factoring in the agreement that may occur due to chance. (Note that the joint probability of agreement for **w1_stress** in Column A is higher than for **outcome**, but the tables are turned when it comes to Kappa. This is the case since **w1_stress** is a binary choice, so 50% joint probability agreement would mean agreement at chance level). Following Landis and Koch (1977), Kappa values were interpreted in the following way: $\kappa < 0$: no agreement; 0–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1 almost perfect.

	A (Hits #1 - #100)	B (Hits #101 - #200)
outcome	67%, $\kappa=0.56$	82%, $\kappa=0.7$
boundary	100%, $\kappa=1$	87%, $\kappa=0.72$
w1_stress	70%, $\kappa=0.14$	60%, $\kappa=0.19$
w2_stress	56%, $\kappa=0.14$	74%, $\kappa=0.08$

Table 2. Inter-rater agreement for /d#j/ for items 1-100. Joint probability of agreement (in percentages) is followed by Cohen's Kappa. $\kappa < 0$: no agreement; 0–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1 almost perfect.

	(Hits #1 - #100)
outcome	87%, $\kappa=0.84$
boundary	92%, $\kappa=0.82$

Table 3. Lambda values of the Box-Cox transformations applied to frequency data.

Corpus	Sequence	λ (w1_freq)	λ (w2_freq)	λ (w1_w2_freq)
Buckeye	/s#j/	-0.067	-0.123	-0.173
Buckeye	/z#j/	-0.067	-0.045	-0.430
Buckeye	/t#j/	-0.081	-0.023	-0.107
Buckeye	/d#j/	-0.052	2.000	-0.084
PLEC	/t#j/	-0.016	-0.044	-0.271
PLEC	/d#j/	-0.078	-0.090	-0.271

Table 4. Estimated coefficients of the multinomial regression model of the outcome of /tʃ/ based on the Buckeye data; $N = 1071$ observations; $AIC = 2050.06$; standard errors shown in brackets. Significance codes: ‘*’ $p < .05$; ‘**’ $p < .01$; ‘***’ $p < .001$.

out come	intercept	gender_fem	age_old	w2_you	cluster	w1_freq	w1_grammar	boundary	w1_freq: w1_grammar
ca	0.72 (0.78)	-0.40 (0.31)	-0.32 (0.34)	2.55 *** (0.43)	-0.05 (0.40)	0.05 (0.08)	1.34 (1.26)	-3.71 *** (0.40)	-0.15 (0.14)
d	-0.25 (0.84)	-0.56 (0.35)	-1.02 ** (0.37)	0.50 (0.42)	1.88 *** (0.45)	-0.03 (0.09)	-1.22 (1.44)	-1.52 *** (0.384)	0.29 (0.16)
g	2.73 *** (0.68)	0.17 (0.28)	-0.60 (0.32)	0.24 (0.35)	-1.86 *** (0.35)	0.04 (0.07)	0.52 (1.18)	-0.67 * (0.31)	-0.01 (0.13)
u	-0.56 (1.07)	-0.84 (0.44)	-0.08 (0.47)	0.10 (0.53)	-1.78 ** (0.59)	0.07 (0.12)	0.74 (1.84)	0.83 (0.50)	-0.10 (0.20)

Table 5. Estimated coefficients of the multinomial regression model of the outcome of /tʃ/ based on the PLEC data; $N = 435$ observations; $AIC = 780.53$; standard errors shown in brackets. Significance codes: ‘*’ $p < .05$; ‘**’ $p < .01$; ‘***’ $p < .001$.

out come	intercept	gender_fem	role_teacher	w2_you	cluster	w1_grammar	boundary	w1_freq: w1_grammar
ca	-5.96 *** (1.13)	0.54 (0.56)	0.12 (0.49)	4.55 *** (0.99)	3.04 *** (0.71)	-2.42 (1.95)	-17.57 *** (0.00)	0.50 (0.29)
d	-7.12 *** (2.04)	1.60 (1.69)	2.90 (1.73)	0.84 (0.94)	3.41 *** (0.94)	0.85 (3.12)	-0.32 (0.88)	-0.25 (0.52)
g	-3.65 * (1.80)	1.40 (1.66)	1.85 (1.79)	-1.37 (1.12)	-0.16 (0.99)	2.18 (4.15)	-1.40 (1.16)	-0.53 (0.69)
u	-27.82 *** (1.11)	-44.08 *** (0.00)	22.36 *** (1.11)	-0.54 (1.24)	3.18 (1.86)	-21.48 (13.46)	1.64 (1.10)	3.39 (2.02)
v	-3.29 ** (1.06)	-0.39 (0.79)	-0.90 (0.68)	1.73 (0.89)	0.45 (0.81)	1.84 (2.48)	-0.66 (0.70)	-0.28 (0.39)

Table 6. Estimated coefficients of the multinomial regression model of the outcome of /d#j/ based on the Buckeye data; $N = 783$ observations; AIC = 970.44; standard errors shown in brackets. Significance codes: ** $p < .05$; *** $p < .01$; **** $p < .001$.

out come	intercept	gender_fem	age_old	w2_you	cluster	w1_freq	boundary	w1_freq:w1_grammar
ca	-0.93 (0.88)	0.95 * (0.35)	-0.02 (0.37)	3.60 *** (0.71)	-0.43 (0.38)	-0.26 (0.14)	-3.19 *** (0.43)	0.07 (0.084)
d	-6.49 *** (1.01)	0.16 (0.37)	0.66 (0.40)	1.23 * (0.59)	4.56 *** (0.52)	0.60 ** (0.22)	-2.82 *** (0.45)	0.15 (0.12)
g	-7.40 ** (2.58)	0.54 (0.64)	0.51 (0.72)	-0.45 (0.99)	2.39 * (1.20)	0.42 (0.65)	-0.45 (0.75)	0.34 (0.36)
u	0.52 (0.68)	-0.79 * (0.34)	-0.55 (0.38)	-0.34 (0.45)	-0.09 (0.36)	-0.06 (0.13)	0.56 (0.44)	-0.03 (0.08)

Table 7. Estimated coefficients of the multinomial regression model of the outcome of /d#j/ based on the PLEC data; $N = 521$ observations; AIC = 416.96; standard errors shown in brackets. Significance codes: ** $p < .05$; *** $p < .01$; **** $p < .001$.

out come	intercept	gender_fem	role_teacher	w2_you	cluster	w1_grammar	w1_w2_freq	boundary	w1_freq:w1_grammar
ca	-52.34 *** (0.56)	0.11 (1.16)	1.28 (0.86)	50.12 *** (0.56)	-2.39 *** (0.64)	4.30 (2.34)	3.23 *** (0.73)	-39.09 *** (0.00)	-1.87 *** (0.48)
d	-4.51 *** (1.29)	0.56 (0.64)	0.63 (0.53)	0.86 (0.85)	1.10 (1.17)	-84.69 *** (0.02)	-0.22 (0.39)	0.08 (0.52)	14.17 *** (0.14)
u	-15.02 *** (1.15)	11.68 *** (0.96)	12.24 *** (0.92)	0.12 (1.59)	-9.51 (56.24)	-91.25 *** (7.92)	-1.12 (1.17)	0.99 (1.41)	16.62 * (7.90)

Table 8. Estimated coefficients and respective standard errors of the binary cumulative-link mixed model (logit link) of the outcome of /s#j/ based on the Buckeye data; $N = 368$ observations; $AIC = 416.96$; **speaker** variable as random intercept (40 speakers, $sd = 0.980$). Significance codes: $** p < .05$; $*** p < .01$; $**** p < .001$.

Predictor	Coefficient estimate	Standard error
gender_female	0.087	0.427
age_old	1.198 **	0.437
w2_you	0.432	0.393
cluster	0.330	0.302
w1_freq	-0.123	0.165
w1_grammar	0.454	1.713
w1_w2_freq	0.583 **	0.218
boundary	-1.980 ***	0.371
w1_freq:w1_grammar	-0.199	0.345
threshold n p	0.195	0.706

Table 9. Estimated coefficients and respective standard errors of the binary cumulative-link mixed model (logit link) of the outcome of /z#j/ based on the Buckeye data; $N = 524$ observations; $AIC = 543.31$; **speaker** variable as random intercept (40 speakers, $sd = 0.929$). Significance codes: $** p < .05$; $*** p < .01$; $**** p < .001$.

Predictor	Coefficient estimate	Standard error
gender_female	0.192	0.392
age_old	0.843 *	0.393
w2_you	0.877 *	0.321
cluster	0.226	1.149
w1_grammar	0.106	0.218
w1_w2_freq	0.250	0.218
boundary	-2.609 ***	0.274
w1_freq:w1_grammar	-0.081	0.219
threshold n p	0.125	0.513