

Distinguishing underlying and surface variation patterns in speech  
perception

Laurel A. Lawyer<sup>1</sup>

Department of Linguistics  
Center for Mind and Brain  
University of California, Davis  
267 Cousteau Drive  
Davis, CA 95618  
(530) 297-4427  
lalawyer@ucdavis.edu

David P. Corina

Department of Linguistics &  
Department of Psychology  
Center for Mind and Brain  
University of California, Davis  
267 Cousteau Drive  
Davis, CA 95618  
(530) 297-4427  
dpcorina@ucdavis.edu

April 4, 2017

Acknowledgments: This work supported in part by the NIH under Grant R01  
DC014767-01 and two UC Davis & Humanities Graduate Research Awards.

---

<sup>1</sup>Corresponding author

## Abstract

This study examines the relationship between patterns of variation and speech perception using two English prefixes: ‘in-’/‘im-’ and ‘un-’. In natural speech, ‘in-’ varies due to an underlying process of phonological assimilation, while ‘un-’ shows a pattern of surface variation, assimilating before labial stems. In a go/no-go lexical decision experiment, subjects were presented a set of ‘mispronounced’ stimuli in which the prefix nasal was altered (replacing [n] with [m], or vice versa), in addition to real words with unaltered prefixes. No significant differences between prefixes were found in responses to unaltered words. In mispronounced items, responses to ‘un-’ forms were faster and more accurate than to ‘in-’ forms, although a significant interaction mitigated this effect in labial contexts. These results suggest the regularity of variation patterns has consequences for the lexical specification of words, and argues against radical underspecification accounts which argue for a maximally sparse lexicon.

Keywords: Speech perception, lexical access, underspecification, phonology, speech variation

## Introduction

Whether owing to differences in vocal anatomy, accent, social situation, or the linguistic context of the words themselves, variability in pronunciation is ubiquitous in natural speech. Human beings are exceptionally adept at dealing with this variation, and typically face little difficulty in understanding spoken language. In the laboratory, we often seek to control these sources of variability as nuisance variables related tangentially to the object of study. However, variation itself is a growing source of research interest, spearheaded in part by recent work in the fields of psycholinguistics, sociolinguistics, and laboratory phonology. The present paper addresses linguistic sources of variation tied to word-internal assimilation processes, and asks whether the relative frequency of the variation pattern effects speech

perception differentially.

The relationship between the phonetic and phonological processes that produce variation in pronunciation has long been a source of debate within linguistics (Fromkin, 1975; Keating, 1990; Reiss, 2007; Tobin, 1988; Trubetzkoy, 1969). In modular theories of language processing (e.g. Chomsky, 1965; Fodor, 1983), phonetics and phonology occupy two separate domains. In these theories, the phonological component is considered a true part of the ‘grammar’, taking complex word forms and altering the sounds within them to fit parameters or rules determined by each specific language. Phonetics, on the other hand, has historically been placed outside the traditional purview of ‘grammar’, and in production, covers changes wrought by translating the output of the phonological component into actionable motor programs for speech.

However, not all theories of language processing involve such strict delineation between these domains. Indeed, many have embraced a more dynamic system of linguistic organization in which phonetics and phonology are either deeply intertwined or not formally distinguished. The movement toward phonetically-informed phonology, or toward systems which collapse phonetics and phonology together, have come from a number of different research traditions, including those within linguistic theory (cf. Hayes, Kirchner, & Steriade, 2004; Lindblom, 1990; Ohala, 1990; Ohala, 2010) and those working more in experimental linguistics and psycholinguistics (Bybee, 2003; Docherty & Foulkes, 2014; Gahl, 2008; Gow & Im, 2004; Sosa & MacFarlane, 2002).

In this paper, we investigate variation in two English prefixes with assimilation patterns which are differentially productive, which we refer to as ‘underlying’ or ‘surface’ variants (following Luce, McLennan, and Chance-Luce, 2003). For the purposes here, underlying variation is a change in pronunciation which is grammatically conventionalized, characterizing something which interfaces with the phonology of the language in a traditional view. Surface variation, on the other hand, is treated as any change in the pronunciation of a word which has not been conventionalized, and hence is more likely to vary between speakers,

situations, and specific instances of a word. Below, we review this distinction in more depth and introduce a paradigm which allows for the comparison of these types of variation in speech perception.

## Underlying variation

Underlying variation results in a variety of modifications to word forms, including the addition, deletion, or modification of specific segments (sounds) in a word. In many instances, these changes are caused by the addition of phonological material created by morphological operations, such as prefixing or suffixing, in which case they are considered morphophonological alternations. One of the most common types of morphophonological variation is that of assimilation, where two sounds become more similar to one another measured by some phonological parameter.

Assimilation is observed in a number of prefixes in English, particularly those of Latinate origin (Bauer, 1983; Jespersen, 1954). One such example is found in the ‘in-’ prefix, which assimilates the final sound to match the place of articulation of the stem to which it is attached. For instance, compare the words *inarticulate*, *intolerant*, and *improbable*. When the stem begins with a vowel, English speakers produce the [in] form exclusively. However, when a stem begins with a labial sound (e.g. [b, p, m]), the final consonant in the prefix changes to match. (It should be noted that in some cases both place of articulation and manner of articulation change, resulting in forms such as *irreverent* and *illogical*. These are beyond the scope of the current investigation, and so are not discussed further.) There is no option for speakers as to whether they would prefer to produce a form like *i[m]effective* or *i[n]mediate*. This prefix alternation also applies to new forms (Baldi, Broderick, & Palermo, 1985), and as such, it is a regular (as in, exceptionless) morphophonological alternation, and can therefore be considered fully phonologized. Note that in this case, the alternation is also reflected in the orthography of the forms themselves, making the labial assimilation highly visible.

## Surface variation

Surface variation also results in changes to pronunciation which can include the addition, deletion, or modification of sounds. What makes surface variation distinct from underlying variation is that it is not grammatically required, which may result in greater variability in application of the alternation. Compare for instance the epenthetic (intrusive) [p] sound in words like *ham[p]ster* and *dream[p]t*. Many people produce these words with the additional [p] sound (Clements, 1987; Ohala, 1997), but it is not universally produced either across or within English speakers (Fourakis & Port, 1986). Assimilation, too, is commonly observed both within single words, and across word boundaries. For instance, in a situation analogous to the prefix ‘in-’ discussed above, the prefix ‘un-’ is seen to participate in an assimilation which alters the place of the final nasal segment within words. In careful speech, the ‘un-’ prefix is canonically pronounced with a final [n] regardless of the stem to which it is attached (e.g., *untried*, *unbecoming*). However, in many situations (e.g., in casual or fast speech), the ‘un-’ prefix assimilates to [um] before labial consonants (e.g., *u[m]predictable*, *u[m]bearable*) (Baldi et al., 1985).

In usage, the cumulative outcome of interacting patterns of underlying and surface variation result in differences in the relative frequency of the alternation. For instance, within the ‘in-’ prefix, the relative frequency of the assimilation is extremely high: the underlying variation pattern means it is expected in the case of all labial stems. Thus ‘in-’ can be seen to vary reliably and with high relative frequency. On the other hand, whereas many people produce [um] forms of the prefix ‘un-’, the [un] form is also acceptable and would serve as the canonical form of the prefix. The relative frequency of [um] forms is then rather less than the relative frequency of ‘in-’ forms. The surface variation in the ‘un-’ prefix is thus not conventionalized, but rather spontaneous, reflecting variation on a case-by-case basis instead of a fixed pattern.

It should be noted that the distinction between surface and underlying variation does not imply these two types of variation are in opposition. Indeed, given the emergence of

a fully stable pattern, surface variation may become phonologized. Likewise, phonological alternations are of course subject to the considerations of surface variation. For instance, an oft-cited assimilation pattern in English is in the voicing of the plural ‘-s’ suffix. Compare, for instance, the final sound in the word *dogs* to that in *cats*. In canonical usage, the plural suffix will be pronounced as a voiced [z] sound when following other voiced sounds such as [g], and will be pronounced as voiceless [s] following other voiceless sounds such as [t]. These assimilations are phonologically determined in English, with novel words obeying the same set of parameters at work within the existing grammar (e.g., the final sounds in *skorts*, *e-cigs*). However, this distinction has not only partially collapsed in some dialects (Bayley & Holland, 2014), there is also variation in the strength of the voicing of the [z] variant in general, resulting in a number of [z] tokens being realized closer to [s] (Davidson, 2016; José, 2010).

## Perception models and morphophonological variation

A robust research tradition has grown out of categorizing when and explaining why language users produce variant pronunciations. In speech perception, understanding how the speech stream can be parsed despite highly variable input forms has been of interest to psycholinguistics for decades. Despite this attention, surprisingly little work has been done with respect to underlying variation in general, and with morphophonological alternations or affixes in specific (though see Scharinger, 2009; Scharinger, Lahiri, & Eulitz, 2010). Below we undertake a brief review of major models which have been used to explain the perception of variant forms, focusing on the structure of lexical representations. Because underlying variation involves distinctions made at the phonological level, we suggest this is the proper locus of investigation. Other systems of speech perception, such as Gow’s feature parsing model (Gow Jr., 2003), or those involving inference mechanisms (Gaskell & Marslen-Wilson, 1996; Marslen-Wilson, Nix, & Gaskell, 1995), rely on mapping processes and other on-line

computations. These models deal primarily with surface variation, and thus, further discussion is omitted here. Finally, as these models have been reviewed in depth in a number of publications (cf. Ernestus, 2014; Ranbom & Connine, 2007; Sumner & Samuel, 2005), they will only be introduced in the following section as they relate to morphophonological alternations and allomorphy.

## **Sparse models**

There are two major approaches to relating variation directly to lexical storage. The first is primarily described by underspecification, wherein some amount of predictable information is omitted from the lexicon, creating a sparse representation. It has been proposed in several theories of phonology (Archangeli, 1988; Chomsky & Halle, 1968; Halle, 1959; Kiparsky, 1982; Trubetzkoy, 1969) as well as more recent work within psycholinguistics (Lahiri & Reetz, 2010) where its primary appeal has been to explain how variant pronunciations may be matched to stored lexical forms. In this case, by omitting all but the most critical phonological information in the lexicon, each lexical item is given greater leeway to match a variety of possible input forms.

The details of underspecification differ between theories in the degree to which lexical items are underspecified. In its mildest form, referred to as alternation-based or archiphonemic underspecification, the omission of phonological material is only motivated by regular and fully predictable morphophonological alternations Inkelas, 1995. For instance, in the ‘in-’ prefix, the place of articulation would be omitted from the stored form of the prefix representation, which would be subsequently restored or derived when the prefix is attached to a stem during production. It is important to note that this type of underspecification does not apply to the ‘un-’ prefix, because the allomorphy in this case is more variable and therefore less predictable. For the ‘un-’ prefix, the stored form would simply be /un-/, with assimilation-induced deviations from this canonical pronunciation produced spontaneously. Allomorphy in this system is thus only partially encoded: in the case that it is driven by

underlying variation patterns, underspecification makes this system of variation explicit. Variation which is not underlying has no direct representation in stored lexical forms.

Other, more complex systems which seek to reduce the information stored in the lexicon to the greatest degree possible have been suggested both within theoretical linguistics (Kiparsky, 1982) and within psycholinguistics (Lahiri & Reetz, 2010). These ‘radical’ underspecification systems sort each phonological parameter into default (‘unmarked’) and non-default (or ‘marked’) values, and posit that anything with a default value is underspecified. Of the underspecification models, only radical underspecification has been the subject of much work in experimental linguistics. Some evidence from neurophysiological and behavioral studies supporting a system of radical underspecification have been presented, primarily focused on whether place features default to coronal (Cornell, Lahiri, & Eulitz, 2011; Eulitz & Lahiri, 2004; Friedrich, Eulitz, & Lahiri, 2006, 2008; Lahiri & Marslen-Wilson, 1991; Walter & Hacquard, 2004; Wheeldon & Waksler, 2004), though there is recent work on frication (Schluter, Politzer-Ahles, & Almeida, forthcoming) and laryngeal features (Hestvik & Durvasula, 2016; Hwang, Monahan, & Idsardi, 2010).

## **Rich models**

In contrast to this drive toward sparse representations, a second tradition has moved instead toward larger and more inclusive lexical representations, primarily referred to as exemplar or usage-based models. Exemplar theory draws heavily from work in memory processing, and positions speech perception as essentially a memory matching enterprise (Johnson, 2007). Following the system laid out by Johnson (2007), to perceive any word, the input is matched against a bank of stored wordforms (exemplars) which have been previously encountered. These exemplars house vast amounts of information, including full acoustic (spectral), visual, and articulatory specifications, as well as additional information about usage. Through a system of weighting and matching algorithms which compare input forms to stored exemplars, each new form activates relevant categories (be they grammatical, semantic, meta-linguistic,



etc.) and these category activations represent the perception of the item. Categories, such as phonemes or grammatical designations, are emergent from the system and are not predetermined entities. A lexical item then could be seen as a category itself, which is activated by lower categories representing its meaning, its phonological form, its grammatical form, social usage, etc. Allomorphy in these systems would be therefore represented either as two sets of exemplars linked to highly overlapping semantic and phonological information, or as two sets of exemplars linked to an abstracted lexical node which would be analogous to a single morpheme.

As these models do not allow one to directly predict the structure of any emergent abstracted forms, there appears to be no *a priori* distinction between surface and underlying sources of variation. However, some usage-based models suggest that each a prototype for each item is produced by summing over the total set of exemplars linked to a particular node (eg. Bybee, 2003, 2010). Prototypes may have varying strengths within categories depending on the degree of congruence among the exemplars, which would allow a distinction between ‘in-’ and ‘un-’ to emerge organically. In this case, because ‘in-’ varies more regularly and therefore more frequently, the prototype representing the combined category of [in] and [im] prefixed items may be less strong. On the other hand, as the variation between ‘un-’ and ‘um-’ is less frequent and thus proportionally more [un] forms would exist, this could result in a stronger prototype effect for this prefix.

Despite being grounded opposing views of lexical richness, it is interesting to note both alternation-based underspecification and usage-based theories make similar predictions about the relationship between the ‘in-’ and ‘un-’ prefixes. Namely, that regularly alternation results in a reduction of specific information about the character of the alternating sound. In underspecification theories, this is formalized as an omission, whereas in usage-based theories this same effect can be seen as a weakening of the prototype. In both cases, these run contrary to theories of radical underspecification, wherein additional elements which do not alternate may be omitted from the lexical form of the item.

## Experimental design and methodology

The following study uses the ‘in-’ and ‘un-’ prefixes to test whether these theoretically-motivated discrepancies in underlying structure affect speech perception. The primary question of interest is whether the surface variation exhibited by the ‘un-’ prefix results in a more explicit lexical representation, which can be viewed either as a stronger prototype or a richer structural specification. This prefix is compared to ‘in-’, which by virtue of its underlying pattern of alternation is suggested to have a less explicit lexical representation, in the form of a weaker prototype or an underspecified nasal segment. These suggestions are in line both with alternation-based underspecification, as well as usage-based theories of speech perception. However, these predictions run contrary to those from radical underspecification, as both prefixes end in coronal nasals, and coronal segments are always underspecified. In a radically underspecified account, both prefixes would be predicted to have equivalently underspecified representations, namely both lacking place features for the final nasal irrespective of their pattern of variation.

The paradigm utilized relies on mispronunciation detection in a go/no-go paradigm as a means to test whether subjects are differentially sensitive to variation in the ‘in-’ and ‘un-’ prefixes. For ‘un-’, because it varies less regularly, we propose that induced changes to the final nasal of the prefix will be more salient, as the input nasal does not match the stored form of the word as strongly. For instance, for a word like *undeniable*, a change to the nasal, as in *umdeniable* should conflict with the stored form of the ‘un-’ prefix, making these items easier to detect as mispronunciations. The leeway for matching mispronounced forms to ‘in-’ on the other hand, is predicted to be greater. Here, since ‘in-’ has a weaker representation due to its conventionalized variation pattern, changes to the nasal segment should be less salient and thus more difficult to detect. Thus if *impure* is altered to *inpure*, this form should better match the stored form of the ‘in-’ prefix than an equivalent change to the ‘un-’ prefix. These predictions run contrary to those generated by radical underspecification, which would predict that changes to ‘in-’ and ‘un-’ have equivalent effects.

Previous work using mispronunciations in behavioral experiments suggests that small discrepancies, such as changes to a single feature, are not frequently reported (Cole, 1973). Of those mispronunciations that are reported, reaction times in lexical decision experiments show an inverse relationship with the degree of mispronunciation, such that the closer an item is to its original pronunciation, the longer it takes to accurately identify the item as mispronounced. Similarly, nonwords have been shown to elicit longer response times than real words (Forster & Chambers, 1973; Marslen-Wilson & Warren, 1994; Vitevitch & Luce, 1998; Whaley, 1978), with classification rates for nonwords are frequently higher than for mispronunciations. While the stimuli in this study do contain only a single feature change, in many cases (discussed at more length below) these result in phonotactically ill-formed items. That is, they result in sound patterns which are either extremely rare or completely disallowed in English. By this measure, we further predict that the mispronounced items which contain phonotactic violations will be treated more like nonwords, resulting in higher classification rates and relatively faster response times, than mispronounced items which do not result in phonotactic irregularities.

## **Subjects**

33 subjects (18 female; ages 18–23, mean age = 19.9, sd = 1.3) participated in this experiment. Subjects were drawn from a pool of undergraduates enrolled in psychology courses at UC Davis, and were given course credit for their participation. As required by the Institutional Review Board at UC Davis, informed consent was acquired from all subjects before commencing the experiment. Subjects were also screened for a history of neurological events and hearing deficits prior to participating.

## **Stimuli**

Using the Celex corpus (Baayen, Piepenbrock, & Gulikers, 1995), 60 ‘in-’ and 60 ‘un-’ prefixed items were chosen which represented an even distribution across major places of articulation:

20 labial, 20 coronal, and 20 velar plosive stems. The items were matched in length across the prefixes, as well as for overall frequency, written frequency, and spoken frequency in the Celex corpus, and in the scaled million-word Celex corpus. A set of 120 filler items were also drawn from the Celex corpus and matched to the experimental stimuli in frequency, number of syllables, lexical category, and overall morphological structure (complex derived word forms beginning with a prefix). No statistically significant differences were found when comparing frequency and length across experimental and filler sets, or between prefix sets, although there was a trend toward ‘in-’ items being slightly longer than ‘un-’ items (see Table 1).

(Table 1 about here)

Modified stimuli were created from the experimental word stimuli by changing the place of articulation in the nasal segment of the ‘in-’ and ‘un-’ prefixes. The prefix-final nasals which in real words contained an [n] were changed for an [m], and any which originally had an [m] were changed for an [n]. This results in forms such as ‘i[n]proper’ from *improper*, or ‘u[m]deniable’ from *undeniable* (see Table 2 for a full set of examples). Note that these modifications result in a phonotactic distinction between items with labial stems and items with non-labial stems. Modified items with non-labial stems have phonotactically aberrant forms which violate phonological expectations in both prefix sets (e.g., [im-d/t/g/k] and [um-d/t/g/k]). Items with labial stems (construing the [um-p/b], and [in-p/b] sequences) contain attested sequences in both prefix sets which differ somewhat between prefixes. For ‘in-’ there do exist a small number of low-frequency words begin with this sequence (i.e., *in-bound*, *input*, *in-patient*). For ‘un-’, recall that this prefix participates in an assimilation to labial stems in some speech styles, thus in this case the modification results not only in a phonotactically allowable sequence, but also an attested pronunciation variant of these items.

(Table 2 about here)

Modified filler items were created in parallel with the modified experimental stimuli by

introducing a number of alterations to a novel set of 80 real words. Half (N=40) include only a change to a single major feature category (such as ‘bilateral’ becoming ‘binateral’), and half (N=40) include a change to a single segment (such as ‘remodel’ becoming ‘rezodel’). Alternations in the modified filler stimuli affect only consonants located in prefixes or near the beginnings of the words, mimicking the structure of the experimental stimuli.

All real word and filler stimuli were recorded using an ART M-Two Cardioid FET Condenser microphone. Real-word experimental and filler items were recorded by a native speaker of Californian English familiar with the experimental paradigm. During recording, each item was placed in neutral sentence frame (“The word xxx is xxx”) with a short pause after the critical item, followed by a randomized set of adjectives to control intonation. Each sentence was repeated three times, and the best example was selected by the experimenter for use, and clipped out of the original sentence. Modified filler stimuli were practiced by the speaker prior to being recorded in the same session using the same methods.

Experimental mispronounced stimuli were created in Audacity (2010) by splicing sequences sourced from real words onto the relevant stems. A single sequence of each prefix type (‘in-’, ‘im-’ and ‘um-’) preceding a voiced segment was selected from additional recorded items. For ‘in-’ stimuli, spliced sequences were extracted from prefixed items (e.g., ‘imbalance’ or ‘indifferent’). For ‘un-’ items, the spliced prefixes were extracted from a familiar (but non-prefixed) [um] sequence (‘umbrella’). This was done as there was concern that recording mispronounced items naturalistically would have resulted in undesired stress or intonation patterns due to the speaker producing deliberately mispronounced or variant stimuli. All splices were made at zero-crossings during periods of relative silence preceding the onset of the stem-initial plosive consonants. Average intensity was normalized across all items using Praat Boersma and Weenink, 2011. The resultant mispronounced stimuli were assessed auditorily by the researcher for naturalness and auditory fidelity.

## Procedure

This experiment used a go/no-go paradigm to maximize potential signal detection (see Perea, Rosa, & Gómez, 2002), as a pilot version of this experiment using a simple lexical decision paradigm had resulted in low accuracy scores / poor signal detection rates for the modified experimental stimuli. In this version of the experiment, subjects were placed into word and modified word response groups. Subjects were instructed to make speeded responses by pressing a button to indicate whether a given item was a correctly pronounced word (for the ‘word’ group), or was unfamiliar by virtue of being pronounced strangely, ‘made up’, or unfamiliar (for the ‘modified word’ group). Response groups and response hand (‘right’ or ‘left’) were counterbalanced across subjects.

Subjects were seated comfortably in a private testing booth. Stimuli were presented via Presentation software (Neurobehavioral Systems Inc., 2014) on a Dell Latitude E5500 laptop over a pair of Beyerdynamic DT 770 Pro circumaural studio headphones. Each subject was presented with a pseudo-randomized list of stimuli which contained all filler items, and a subset of experimental items. Stimuli were balanced to ensure no subject heard both the original and the modified version of any experimental item, resulting in each subject hearing only half of the possible experimental stimulus items, balanced between word and modified word sets. Each trial consisted of a 1500msec silent fixation, followed by a single stimulus item presented auditorily in isolation. Following presentation, subjects were given a response window of 1000ms, after which a jittered period of silence (600-1400msec) followed to reduce anticipation. Trials were binned into three approximately 10-minute blocks.

Prior to beginning the experiment, subjects were administered an informal assessment of hearing thresholds. While this did not provide a clinical assessment of hearing acuity, it did provide a measure to compare relative hearing ability between subjects, and hearing thresholds were used in the formulation of statistical models discussed below.

## Analysis

Statistical analyses of reaction times and accuracy for words and modified words were performed in R (R Core Team, 2013) using the lme4, lmerTest, and lsmeans packages (Bates, Maechler, Bolker, & Walker, 2014; Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016; Lenth, 2016). Outliers which exceeded 2 standard deviations from the mean were removed, resulting in a loss of 4% of the available observations. Statistical analysis response latencies utilized a linear mixed effects model with log-transformed reaction times as the dependent variable. Accuracy data was analyzed using a binomial mixed logit model. For each response group ('word responders'/'modified word responders'), word and modified word responses were modeled separately for both accuracy and response latency,

Numerous factors were available for the mixed effects models. Specifically: Lexical status (word/modified), Prefix (IN, UN), Stem (labial, coronal, dorsal), Sex (M/F), Age, Handedness (L/R), ResponseHand (L/R), VocabularyScore, BilingualStatus (Y/N), Trial, Length (in msec), Frequency, StemFrequency, UniquenessPoint (in msec), and six additional factors constituting hearing thresholds at six frequencies (250, 500, 1000, 2000, 4000, and 8000 Hz). Prior to inclusion in the model, continuous factors were transformed to approximate a more normal distribution, and scaled and centered where appropriate to reduce the possibility of colinearity. Because Frequency and StemFrequency are somewhat correlated ( $r = .23$ ), StemFrequency was also residualized against Frequency prior to transforming the resultant values. The transforms as well as centering values for each continuous variable are listed in Table 3.

(Table 3 about here)

All models were initially estimated with the maximum fixed effects structure. Not all effects contributed significantly to the final models. To determine which elements remained in the final models, individual factors were removed iteratively by excluding the factor with the lowest z-value and refitting the model until only factors with a z-value above 2 remained. Each model was also initially fitted with by-subject and by-item random intercepts, as well

as by-subject random slopes, the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013). Inclusion of these random effects in the model were justified by means of log likelihood comparisons between the optimal model and a null model excluding these effects.

## Results: Response latencies

In this and the following section (discussing response accuracy), each response group is analyzed separately. This is due to the fact that words and modified words represent different response categories (e.g., hits or false alarms) for each response group.

### Word Response Group

For subjects in the word response group, mean values for log-transformed reaction times for each Prefix and Stem category are shown in Table 4.

(Table 4 about here)

### Modified Words

In the word response group, latencies for modified words are derived from ‘false alarm’ (i.e., incorrect) responses. For these items, a significant effect of Stem was observed ( $F(2, 104) = 3.44, p = .03$ ). Posthoc pairwise analysis using Tukey’s method, adjusted for multiple comparisons, shows that labial stems had faster responses than dorsal stems ( $\beta = -.14, t = -2.61, p = .03$ ), with no other significant differences were found among Stems. There was no significant main effect of Prefix in this model. These effects are shown in Figure 1.

Both Frequency and StemFrequency were significant predictors of response latency. Responses were faster to modified items derived from real words with high word frequency ( $\beta = -.09, t = -3.70, p = .0003$ ) and stem frequency ( $\beta = -.05, t = -2.09, p = .04$ ) values. No additional factors were found to be significant.



## Words

Latencies for word responses represent ‘hit’ (correct) values. Both Frequency and StemFrequency were found to contribute significantly to this model. In both cases, responses were faster to items with higher frequencies, with word frequency ( $\beta = -.09, t = -4.89, p < .0001$ ) playing a larger role than stem frequency ( $\beta = -.04, t = -2.36, p = .02$ ). A significant effect of Trial also suggested that responses slowed over time ( $\beta = .04, t = 2.20, p = .01$ ). Neither Prefix nor Stem, nor any other factors, were found to be significant predictors in this model.

(Figure 1 about here)

## Modified Word Response Group

For subjects in the modified word response group, mean values for log-transformed reaction times to each Prefix and Stem category are shown in Table 5.

(Table 5 about here)

## Modified Words

Analysis of the latencies for correctly categorized modified words (‘hits’) in this response group resulted in a main effect of Prefix, with ‘un-’ responses being faster than ‘in-’ responses ( $\beta = -.24, t = -3.56, p = .0006$ ). There was also a main effect of Stem ( $F(2, 75) = 6.05, p = .004$ ). Posthoc pairwise analysis revealed that labial stems had slower responses than both coronal stems ( $\beta = -.29, t = -3.40, p = .003$ ) and dorsal stems ( $\beta = -.22, t = -2.43, p = .04$ ), with no significant distinction between the latter two categories. There was no significant interaction between Stem and Prefix, and no other factors were significant in this model. Stem and Prefix effects are illustrated in Figure 2.

## Words

For the modified word response group, latencies for words are derived from ‘false alarm’ (i.e., incorrect) responses. In this group, Stem Frequency (but not overall Frequency) was a

significant factor. Subjects showed longer responses to items derived from real words with more frequent stems ( $\beta = .15, t = 2.89, p = .005$ ). A significant effect of Trial was also observed, showing that subjects made faster responses over time ( $\beta = -.11, t = -2.34, p = .02$ ). No other factors were significant in this model.

(Figure 2 about here)

## Discussion

In the response latency data, similar patterns are observed in both the word and modified word response groups. In responses to words, we observe that frequency is a significant predictor of latency, with responses to frequent words being faster than responses to less frequent words. These type of frequency effects are ubiquitous in studies of lexical recognition, having been demonstrated numerous times (eg. Broadbent, 1967; Goldinger, Luce, & Pisoni, 1989; Segui, Mehler, Frauenfelder, & Morton, 1982; Taft, 1979).

Of greater interest are the responses to modified items. Here, we observe an interplay between prefix, stem, and response group. For subjects making modified word responses (hits), responses to ‘un-’ items are faster than ‘in-’ items. However, prefix was not found to be a significant predictor of response latencies for subjects making word responses (false alarms). On the other hand, stem appears to mediate responses for both groups. In the modified word group, hit responses are slower for labial stems than for dorsals or coronals. In the word response group, we find false alarm responses to labial stems to be faster than dorsals, with coronals not significant differing from either labials or dorsals.

In short, it appears that when subjects correctly identify items as modified words, both ‘in-’ items as well as items with labial stems ([um-b/p], [in-b/p]) are more difficult to identify and thus generate slower responses. The distinction in prefixes supports the notion that ‘in-’ forms may tolerate greater variability due to their naturally alternating status, reflected in the slower responses to modified ‘in-’ items.

When subjects erroneously report modified words items to be real words, labial stems

are responded to more quickly, providing further evidence that modified items with labial stems are particularly difficult to identify. While this is expected behavior for labial ‘un-’ stimuli, the parallel situation with ‘in-’ was not predicted. Possible explanations for this are taken up in the General Discussion.

## Results: Classification accuracy

### Word Response Group

Mean accuracy values for each Prefix and Stem category for the word response group are shown in Table 6.

(Table 6 about here)

### Modified Words

In the word group, classification accuracy for modified words showed significant main effects of both Stem and Prefix. For Prefix, responses to ‘un-’ items were more accurate than ‘in-’ items ( $OR : 1.86, z = 2.51, p = .01$ ). Within Stems, posthoc pairwise analysis shows that labial stems were less accurate than both coronal stems ( $OR : 0.38, z = -3.24, p = .003$ ) and dorsal stems ( $OR : 0.45, z = -2.65, p = .02$ ). There was no significant difference between coronal and dorsal stems, and no significant Prefix by Stem interaction. These effects are pictured in Figure 3.

There were also significant effects of Frequency and Trial. Subjects were more likely to classify a modified word incorrectly (i.e., as a word) if the item was derived from a high frequency word ( $OR : 0.71, z = -2.66, p = .007$ ), and classification accuracy improved over time ( $OR : 1.40, z = 3.89, p = .0001$ ).

## Words

For word classification, both Frequency ( $OR : 1.97, z = 3.29, p = .001$ ) and Stem Frequency ( $OR : 2.05, z = 3.83, p = .0001$ ) were significant predictors of accuracy, resulting in higher correct classification rates as both Word and StemFrequency increase. Accuracy also increased in line with VocabularyScore ( $OR : 2.18, z = 5.13, p < .0001$ ). No other factors were found to be significant predictors of word classification accuracy.

(Figure 3 about here)

## Modified Word Response Group

Mean accuracy values for each Prefix and Stem category for the modified word response group are shown in Table 7.

(Table 7 about here)

## Modified Words

Classification of modified items by the modified word response group showed significant main effects of Prefix ( $OR : 4.30, z = 4.77, p < .0001$ ) and a significant Prefix x Stem interaction. Pairwise posthoc testing within each Prefix revealed that Stem was a significant factor for UN items only. Within the IN items, there was no significant differences in classification accuracy by Stem (all  $p > .7$ ). Within UN, coronal items were more likely than labial items to be correctly classified as modified words ( $OR : 3.15, z = 3.88, p = .001$ ). There was no significant difference between labial items and dorsals, or between dorsal items and coronals within the UN prefix (all  $p > .2$ ). Stem and Prefix effects are shown in Figure 4.

Other significant contributions to modified word classification accuracy included Frequency, whereby items derived from high frequency words were more likely to be classified incorrectly as words ( $OR : .80, z = -2.30, p = .02$ ). There was also a significant interaction between VocabularyScore and BilingualStatus ( $OR : 5.90, z = 2.50, p = .01$ ), showing

that subjects with higher vocabulary scores were more likely to correctly categorize modified words, but this effect was restricted to monolingual subjects.

## Words

The only significant predictor in of classification accuracy for words was StemFrequency. Here, the more frequent the stem, the more likely subjects were to correctly classify these items as words ( $OR : 1.91, z = 3.16, p = .002$ ). No other factors were significant in this model.

(Figure 4 about here)

## Discussion

Analysis of the classification accuracy data echoes the findings observed with latency data. Responses to real words are again mediated primarily by frequency, such that more frequent words, or words with more frequent stems, are more likely to be categorized as real words. Interestingly, we also observe frequency effects in the modified word classification accuracy, where accuracy for modified words derived from more frequent words is reduced, reflecting the tendency for higher frequency items to be treated as real words regardless of response context.

For modified words, both response groups show a similar pattern, wherein more modified ‘un-’ items were correctly categorized than modified ‘in-’ items. In both groups, word stem also plays a determining role in response accuracy, with labials again standing out against the other items. In the word group, fewer labial stems were correctly categorized for both ‘in-’ and ‘un-’ prefixes. In the modified word response group, there was no difference in categorization within the ‘in-’ prefix, but fewer labial ‘un-’ items were correctly classified than other ‘un-’ items. Taken together, these results provide additional support for the notion that ‘in-’ prefixed items tolerate greater degrees of variation in perception, which is observed here as a reduced ability for subjects to report modified ‘in-’ items as modified

words. Poor classification of labial ‘un-’ items was predicted, as these items are frequently encountered in natural speech. The relatively poorer classification of labial ‘in-’ items in the word response group warrants further discussion, which will be taken up below.

## General Discussion

### Prefixes

The main goal of this research was to investigate whether ‘un-’ and ‘in-’ prefixed items are differentially tolerant to mispronunciations due to distinctions in their natural patterns of variation. Indeed, we observe that responses to mispronounced ‘in-’ and ‘un-’ stimuli produced different results, both in response speed and response accuracy. In both response groups, modified ‘un-’ forms were more likely to be identified than ‘in-’ forms, and hit responses were faster to ‘un-’ forms than ‘in-’ forms. Taken together, this suggests that the ‘un-’ forms were less confusable with real words than their ‘in-’ counterparts.

This is consistent with lexical accounts which directly incorporate alternation, including both usage-based accounts and alternation-based underspecification. In both cases, ‘in-’ was suggested to have a weaker prototype or less rich structural specification which would allow for greater matching tolerance. That is indeed what we observe. As ‘un-’ varies less predictably, it was suggested to have a stronger prototype or richer structural specification, which would result in less tolerance for deviations from the standard [un] form. This is what we observe, with participants able to make faster and more accurate decisions about items with modified [un] prefixes.

Modeling these results with respect to alternation-based underspecification is relatively straight forward, as a structural model of this variety makes specific claims about the storage of specific items, including prefixes. Here, we predict simply that the prefixes are stored as /un-/, with a fully specified nasal segment, and as /iN-/, with a nasal segment underspecified for place.

With respect to usage-based models, it is worth noting that this analysis requires some assumptions about the ways in which an exemplar-based model would have to be structured. First, it requires a separate prototype for each prefix. Second, this data also requires that prefixes are processed separately from the wholly-composed word. If this were not the case, there would be no a priori reason to expect that modified words with ‘in-’ prefixes such as *imtolerant* would match any better to their related real word forms than for instance *umtidy*. Thus the perception mechanism would require access to stored lexical knowledge about the prefixes themselves.

Whether we pursue a more structural or more usage-based analysis, what is clear is that this data does not support a radically underspecified lexicon. In radical accounts, coronals often serve as the prime example of underspecified segments. Therefore, both prefixes would be predicted to provide equivalent matches to modified stimuli, and thus no distinctions between them would be expected. This is not what the data shows. Instead we find distinctions between ‘in-’ and ‘un-’ in both response latencies and accuracy, consistent with accounts which can incorporate variation into the structure of the lexical representations.

## **The influence of orthography**

In addition to questions of lexical specification, it is also worth exploring the potential influence of orthography in these results. Previous studies have shown effects in the auditory domain which can be tied to issues of orthographic regularity (Chéreau, Gaskell, and Dumay (2007), Ventura, Morais, Pattamadilok, and Kolinsky (2004), Ziegler and Ferrand (1998), though see also Mitterer and Reinisch (2015)). In the present study, one may point out that all of the ‘un-’ stimuli contain the [um] sequence, which is not an orthographic variant of the ‘un-’ prefix itself. In contrast, the [in] and [im] sequences both have a direct orthographic representation. However, caution is warranted as the orthographic situation in this study is complex, particularly as we have employed ‘mispronounced’ stimuli which include not only phonotactically irregular sequences, but orthographically unattested sequences as well (if one

were to make a direct mapping of the heard sequences). Given prior research, it is not clear whether the orthographically unattested [um-t] forms should be expected to differ from the orthographically unattested [im-t] forms by virtue of a system which recognizes the spelling variant of the ‘in-’ prefix itself.

However, the role orthography may itself play in the structure of the lexical representations of the ‘in-’ and ‘un-’ prefixes is not irrelevant. Given that some usage-based models allow for links between lexical items with their visual/orthographic forms, the discrepancies in orthography may serve to amplify differences between the two prefixes. Orthography and variation patterns are, however, not strictly independent. One can easily imagine that the differing orthographic status of these prefixes could be a secondary mechanism to highlight the status of the alternations of these prefixes. Thus the more conventionalized alternation pattern of ‘in-’ is codified in the orthographic representation. This suggestion, while speculative, does provide a way to unify orthographic and phonological influences in these two prefixes. It should be noted, however, that the relationship between alternation and orthography in English in general are much more varied (for instance, the previous example of plural ‘-s-’ alternation is conventionalized but does not have an orthographic alternation). Orthographic correspondences in particular have roots deep in the stylistic and linguistic choices made in the evolution of the English language itself. Any statistical tendencies relating the reliability of variation and the likelihood of orthographic representation warrants its own investigation.

## **The behavior of labials**

One surprising finding in this data is that items with labial stems elicited distinct responses within both prefix groups. Particularly in the reaction time data, we observe labial stems take longer to correctly identify (as hits in the modified word response groups), and responses are quicker when labial stems are mistaken for real words (as false alarms in the word response group). Items with labial stems were also found to be less accurate, though this



was primarily true for the ‘un-’ items, as the accuracy for labial ‘in-’ items was shown to be significantly different from the other stems only in the word response group. This separation of labial items from other stems was the anticipated behavior in the ‘un-’ stimuli, as the labial forms contained an assimilation which is frequently observed and should be familiar to the participants in the study. However, no prediction was made regarding the behavior of labial ‘in-’ items.

Some explanation may lie in the phonotactics of the sequences used in this study. Both labial ‘in-’ and ‘un-’ items are phonotactically well-formed, albeit in both cases relatively infrequently encountered in standard usage. The [ump/b] sequence is testified in a handful of forms (eg. *umpire*, *umbrella*, *umpteenth*) as is the [inp/inb] sequence (eg. *input*, *in-bound*, *in-patient*). Phonotactic regularity has been shown to play a role in perception (Breen, Kingston, & Sanders, 2013; Dupoux, Pallier, Kakehi, & Mehler, 2001; Steinberg, Jacobsen, & Jacobsen, 2016), which may be reflected in this data as slower response times for correct modified word identification.<sup>1</sup> However, we note that phonotactics alone cannot explain the full pattern of responses observed in this study. In particular, in non-labial items, we still observe a distinction between responses to ‘in-’ and ‘un-’ prefixes which cannot be driven by phonotactics. In both cases, forms such as *umdetered* or *imdelicate* have equivalent phonotactic violations.

Another source of potential difference within the ‘in-’ prefix set is a distinction in assimilation type. There is a growing acknowledgment that the phonetic details of assimilation are much more complicated than a simple exchange of one sound for another. A number of studies have demonstrated that dynamic assimilation, as is observed across word boundaries, results in an incomplete assimilation whereby some phonetic cues to a sound’s pre-assimilated form remain, and that listeners use these cues to uncover the original identity of the segment (Gow Jr., 2002). To our knowledge, this phenomena has not been studied with underlying sources of variation, including morphophonological alternations of the type used here,

---

<sup>1</sup>Note that this holds true even when responses are adjusted relative to the uniqueness points of each word.

though there has been suggestion that assimilation in this case is ‘complete’ (Jun, 2004). More work is needed to determine whether assimilation within prefixed words behaves in the same manner as the surface assimilation observed across word boundaries.

## Mispronunciation

Finally, the data presented above show that modified items used in this experiment are treated in large part as real words. Across both prefix categories, subjects reported an average of 72% of the modified stimuli as real words. These high false alarm rates are in line with previous literature (Cole, 1973) which shows equivalent identification rates for items in which a single feature was altered (approximately 70%). Modified filler items, which were mispronounced by one or more features within a single segment, showed rather higher identification rates, with subjects reporting only 23% of these items as real words.

While this experiment replicates the main findings of Cole (1973), there are some discrepancies particularly with respect to subject performance on filler items. Cole (1973) shows that items with a mispronunciation in the initial syllable are easier to detect than mispronunciations in subsequent syllables. This is not the case with the data presented here, as experimental items all contained mispronunciations within the first syllable, and yet in comparison to filler items which contained mispronunciations in either the first or second syllable, were much more difficult to accurately classify. One source of discrepancy between the experimental and filler items, and indeed between the experimental items and those items used in Cole (1973), is the fact that the experimental items contain mispronunciations in the coda of the syllable. This is contrary to a majority of the filler items, and all of the items used in Cole’s set of initial-syllable mispronunciations. In both cases, these items contain mispronunciations in the onset (beginning) of the syllable. Given the wide literature reporting the privileged status of the initial phoneme in both linguistic typology (e.g. Jakobson, 1962; Prince & Smolensky, 1993), and in speech perception (e.g. Marslen-Wilson & Welsh, 1978; Redford & Diehl, 1999), the fact that the mispronunciations used in this ex-

periment are in coda position of the initial syllable may render them particularly difficult to perceive. Additional work is needed to explore the relationship between mispronunciations and syllable positions in general.

## Conclusions

The study presented here used a go/no-go lexical decision paradigm to test the prediction that the stored forms of the ‘in-’ and ‘un-’ prefixes differ due to distinctions in their patterns of variation. Because ‘in-’ participates in an underlying variation pattern which alters the place of the nasal segment, it was suggested that the stored form of this prefix would contain specific information about the place of final nasal. The ‘un-’ prefix exhibits only surface variation, therefore the stored form of this prefix was suggested to have a richer specification, including more information about the identity of the final nasal. Results from the lexical decision experiment, particularly with reference to classification accuracy, support this analysis, showing that listeners have difficulty discriminating modified forms of the ‘in-’ prefix from their canonical forms. This stands in contrast to a majority of the modified ‘un-’ forms, which were classified more quickly and more accurately than ‘in-’ forms. However, the data also revealed an interaction with stem consonants, such that ‘un-’ prefixed words with labial stems were particularly very difficult to classify. As this subset of ‘un-’ stimuli naturally participate in a familiar and frequent surface assimilation, this behavior was expected. Taken together, the data presented here demonstrate that the perceptual system is sensitive to the source or degree of regularity in variation, and that these patterns of variation have an effect on lexical specificity. These results are consistent both with alternation-based accounts of underspecification, as well as usage-based accounts such as exemplar theories. However, the data presented here conflicts with other, more radical, views of underspecification, such as suggested by Lahiri and Reetz (2010).

## Acknowledgments

The authors wish to thank Meghan Sumner for helpful comments and suggestions. Additional thanks are owed to Chris Graham for assistance with stimulus development, and Todd LaMarr, Diane Alshouse, Elizabeth Cole, Jennifer Luevano, Kaitlin Murray, Julie Ngo, Joshua Petracich, Annie Welch and Yingxi Yang for assistance with data collection. This work supported in part by the NIH under Grant R01 DC014767-01 and two UC Davis & Humanities Graduate Research Awards.

## References

- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(2), 183–207. doi:10.1017/S0952675700002268
- Audacity Team. (2010). Audacity, Version 1.3.4. <http://audacity.sourceforge.net/>. Retrieved from <http://audacity.sourceforge.net/>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The celex lexical database. CD-ROM. Linguistic Data Consortium, University of Pennsylvania. Philadelphia, PA. Retrieved from [http://www ldc.upenn.edu/Catalog/readme\\_files/celex.readme.html](http://www ldc.upenn.edu/Catalog/readme_files/celex.readme.html)
- Baldi, P., Broderick, V., & Palermo, D. S. (1985). Prefixal negation of english adjectives: psycholinguistic dimensions of productivity. In J. Fisiak (Ed.), *Historical semantics historical word-formation* (Vol. Studies and Monographs vol. 29, pp. 33–58). Trends in Linguistics. The Hague: Mouton.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2014). lme4: linear mixed-effects models using eigen and s4. ArXiv e-print; submitted to *Journal of Statistical Software*. Retrieved from <http://arxiv.org/abs/1406.5823>
- Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.
- Bayley, R. & Holland, C. (2014). Variation in Chicano English: the case of final (z) devoicing. *American Speech*, 89(4), 385–407. doi:10.1215/00031283-2908200
- Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer, Version 5.2.35. <http://www.praat.org/>. Retrieved from <http://www.praat.org/>
- Breen, M., Kingston, J., & Sanders, L. D. (2013). Perceptual representations of phonotactically illegal syllables. *Attention, Perception, & Psychophysics*, 75(1), 101–120. doi:10.3758/s13414-012-0376-y

- Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological review*, 74(1), 1. doi:10.1037/h0024206
- Bybee, J. (2003). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: orthographic effects in auditory priming. *Cognition*, 102(3), 341–360. doi:10.1016/j.cognition.2006.01.001
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English* (4th). Cambridge, MA: Cambridge: MIT Press.
- Clements, G. N. (1987). Phonological feature representation and the description of intrusive stops. In *Proceedings of the parasession on autosegmental and metrical phonology, chicago linguistics society* (Vol. 23, pp. 29–51).
- Cole, R. A. (1973). Listening for mispronunciations: a measure of what we hear during speech. *Perception & Psychophysics*, 13(1), 153–156. doi:10.3758/BF03207252
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2011, June). "what you encode is not necessarily what you store": evidence for sparse feature representations from mismatch negativity. *Brain research*, 1394, 79–89. doi:10.1016/j.brainres.2011.04.001
- Davidson, L. (2016). Variability in the implementation of voicing in american english obstruents. *Journal of Phonetics*, 54, 35–50. doi:10.1016/j.wocn.2015.09.003
- Docherty, G. J. & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*, 142, 42–56. SI: Usage-Based and Rule-Based Approaches to Phonological Variation. doi:10.1016/j.lingua.2013.01.011
- Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and cognitive processes*, 16(5-6), 491–505.
- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, 142, 27–41. doi:10.1016/j.lingua.2012.12.006

- Eulitz, C. & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*(4), 577–583. doi:10.1162/089892904323057308
- Fodor, J. A. (1983). *Modularity of mind: an essay on faculty psychology*. Cambridge: MIT Press.
- Forster, K. I. & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635. doi:10.1016/S0022-5371(73)80042-8
- Fourakis, M. & Port, R. (1986). Stop epenthesis in english. *Journal of Phonetics*, *14*, 197–221.
- Friedrich, C., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: an erp study. *Behavioral and Brain Functions*, *2*(36), 1–10. doi:10.1186/1744-9081-2-36
- Friedrich, C., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1545–1559. doi:10.1037/a0012481
- Fromkin, V. A. (1975). The interface between phonetics and phonology. In *Ucla working papers in phonetics* (Vol. 33, pp. 104–107). Los Angeles: UCLA.
- Gahl, S. (2008). "time" and "thyme" are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496. doi:10.1353/lan.0.0035
- Gaskell, M. G. & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 144–158. doi:10.1037/0096-1523.22.1.144
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: effects of competition and inhibition. *Journal of memory and language*, *28*(5), 501–518. doi:10.1037/0033-295X.105.2.251

- Gow Jr., D. W. (2002). Does english coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 163. doi:10.1037/0096-1523.28.1.163
- Gow Jr., D. W. (2003). Feature parsing: feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65(4), 575–590. doi:10.3758/BF03194584
- Gow, D. W. & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51(2), 279–296. doi:10.1037/0096-1523.28.1.163
- Halle, M. (1959). *The Sound Pattern of Russian*. Description and Analysis of Contemporary Standard Russian. The Hague: The Hague: Mouton.
- Hayes, B., Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically-based phonology*. Cambridge: Cambridge University Press.
- Hestvik, A. & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in english. *Brain and Language*, 152, 28–43. doi:10.1016/j.bandl.2015.10.007
- Hwang, S.-O. K., Monahan, P. J., & Idsardi, W. J. (2010, August). Underspecification and asymmetries in voicing perception. *Phonology*, 27, 205–224. doi:10.1017/S0952675710000102
- Inkelas, S. (1995). The consequences of optimization for underspecification. In *Proceedings-nels* (Vol. 25, pp. 287–302). Boston: University of Massachusetts. Retrieved from <http://roa.rutgers.edu/files/40-1294/40-1294-INKELAS-0-0.PDF>
- Jakobson, R. (1962). *Selected writings 1: phonological studies*. The Hague: Mouton.
- Jespersen, O. (1954). *A modern english grammar on historical principles*. London: George Allen & Unwin Ltd.
- Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 25–40). Oxford University Press.
- José, B. (2010). The Apparent-Time Construct and stable variation: final /z/ devoicing in northwestern Indiana. *Journal of Sociolinguistics*, 14(1), 34–59. doi:10.1111/j.1467-9841.2009.00434.x



- Jun, J. (2004). Phonetically-based phonology. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), (Chap. Place assimilation, pp. 58–86). Cambridge: Cambridge University Press.
- Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of phonetics*, 18(3), 321–334.
- Kiparsky, P. (1982). Lexical Morphology and Phonology. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 3–91). Seoul: Hanshin Publishing Company.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *Lmertest: tests in linear mixed effects models*. R package version 2.0-32. Retrieved from <https://CRAN.R-project.org/package=lmerTest>
- Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245–294. doi:10.1016/0010-0277(91)90008-R
- Lahiri, A. & Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *Journal of Phonetics*, 38, 44–59. doi:10.1016/0010-0277(91)90008-R
- Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1–33. doi:10.18637/jss.v069.i01
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (Vol. 55, pp. 403–439). NATO ASI Series. Amsterdam: Springer. doi:10.1007/978-94-009-2037-8\_16
- Luce, P. A., McLennan, C. T., & Chance-Luce, J. (2003). Rethinking implicit memory. In J. Bowers & C. Marsolek (Eds.), (Chap. Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. pp. 197–214). Oxford: Oxford University Press.

- Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63. doi:10.1016/0010-0285(78)90018-X
- Marslen-Wilson, W., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: abstractness, inference and english place assimilation. *Language and Cognitive Processes*, *10*(3-4), 285–308. doi:10.1080/01690969508407097
- Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological review*, *101*(4), 653. doi:10.1037/0033-295X.101.4.653
- Mitterer, H. & Reinisch, E. (2015). Letters don't matter: no effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, *85*, 116–134. doi:10.1016/j.jml.2015.08.005
- Neurobehavioral Systems Inc. (2014). Presentation software (ver. 14.04). www.neurobs.com.
- Ohala, J. J. (1990). There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics*, *18*, 153–171.
- Ohala, J. J. (1997). Emergent stops. In *Proceedings of the 4th seoul international conference on linguistics* (pp. 84–91).
- Ohala, J. J. (2010). The relation between phonetics and phonology. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (Second, Chap. 17, pp. 653–677). Oxford: Blackwell Publishing. doi:10.1002/9781444317251.ch17
- Perea, M., Rosa, E., & Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition*, *30*(1), 34–45. doi:10.3758/BF03195263
- Prince, A. & Smolensky, P. (1993). Optimality theory: constraint interaction in generative grammar. ROA-537. Rutgers Optimality Archive. Retrieved from <http://roa.rutgers.edu>

- R Core Team. (2013). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Ranbom, L. J. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. doi:10.1016/j.jml.2007.04.001
- Redford, M. A. & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in cvc syllables. *Journal of the Acoustical Society of America*, *106*(3), 1555–1565. doi:10.1121/1.427152
- Reiss, C. (2007). Modularity in the sound domain: implications for the purview of universal grammar. In G. Ramchand & C. Reiss (Eds.), *The oxford handbook of linguistic interfaces* (pp. 53–80). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199247455.001.0001
- Scharinger, M. (2009). Minimal representations of alternating vowels. *Lingua*, *119*(10), 1414–1425. doi:10.1016/j.lingua.2007.12.009
- Scharinger, M., Lahiri, A., & Eulitz, C. (2010, July). Mismatch negativity effects of alternating vowels in morphologically complex word forms. *Journal of Neurolinguistics*, *23*(4), 383–399. doi:10.1016/j.jneuroling.2010.02.005
- Schluter, K., Politzer-Ahles, S., & Almeida, D. (forthcoming). No Place for h: an ERP investigation of English fricative place features. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2016.1151058
- Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, *20*(6), 615–627. doi:10.1016/0028-3932(82)90061-6
- Sosa, A. V. & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: collocations involving the word of. *Brain and Language*, *83*(2), 227–236. doi:10.1016/S0093-934X(02)00032-9
- Steinberg, J., Jacobsen, T. K., & Jacobsen, T. (2016). Repair or violation detection? pre-attentive processing strategies of phonotactic illegality demonstrated on the constraint

- of g-deletion in German. *Journal of Speech, Language, and Hearing Research*, 1–15. doi:10.1044/2015-JSLHR-H-15-0062
- Sumner, M. & Samuel, A. G. (2005). Perception and representation of regular variation: the case of final/t. *Journal of Memory and Language*, 52(3), 322–338. doi:10.1016/j.jml.2004.11.004
- Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 7(4), 263–272. doi:10.3758/BF03197599
- Tobin, Y. (1988). Phonetics versus phonology. In Y. Tobin (Ed.), *The Prague school and its legacy* (pp. 49–70). Amsterdam: John Benjamins. doi:10.1075/llsee.27.07tob
- Trubetzkoy, N. S. (1969). *Principles of phonology*. Baltaxe, Christiane A. M., Trans. (Original work published 1935). Berkeley and Los Angeles: University of California Press.
- Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, 19(1), 57–95. doi:10.1080/01690960344000134
- Vitevitch, M. S. & Luce, P. A. (1998). When words compete: levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325–329. doi:10.1111/1467-9280.00064
- Walter, M. A. & Hacquard, V. (2004). MEG evidence for phonological underspecification. In E. Halgren, S. Ahlfors, M. Hämäläinen, & D. Cohen (Eds.), *Proceedings of the 14th international conference on biomagnetism*. Boston: Biomag.
- Whaley, C. (1978). Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17(2), 143–154. doi:10.1016/S0022-5371(78)90110-X
- Wheeldon, L. & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and language*, 90(1-3), 401–12. doi:10.1016/S0093-934X(03)00451-6
- Ziegler, J. C. & Ferrand, L. (1998). Orthography shapes the perception of speech: the consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, 5(4), 683–689. doi:10.3758/BF03208845

## Tables and captions

Table 1: Summary of stimulus metrics. Note that all columns are abbreviated with their Celex designations. These are: frequency (Cob), scaled frequency from the 1 million words Celex corpus (CobMln), written frequency (CobW), scaled written frequency from the 1 million words Celex corpus (CobWMln), spoken frequency (CobS), scaled spoken frequency from the 1 million words Celex corpus (CobSMln), and syllable count. None of these measures provide statistically significant differences between prefix sets, or between the experimental and filler stimuli.

		Cob	CobMln	CobW	CobWMln	CobS	CobSMln	SyllCnt
'in-'	$\mu =$	63.33	3.57	61.08	3.72	2.25	1.82	3.93
	$\sigma =$	65.09	3.58	62.12	3.76	4.41	3.41	.86
'un-'	$\mu =$	58.67	3.30	56.68	3.38	1.98	1.63	3.58
	$\sigma =$	89.22	5.02	85.15	5.12	4.75	3.68	.31
	F(2,119) =	.068	.068	.062	.097	.263	.193	2.662
	$p < =$	.93	.93	.94	.90	.76	.83	.07
Fillers	$\mu =$	62.62	3.52	60.14	3.65	2.48	1.96	3.64
	$\sigma =$	80.31	4.48	76.24	4.57	5.05	3.90	.98
	F(2,239) =	.270	.261	.225	.228	1.428	1.313	.858
	$p <$	.89	.90	.92	.92	.22	.27	.49

Table 2: Experimental stimulus categories with examples.

Prefix	Stem	Word	Modified word
<b>in-</b>	labial	i[m]precise	i[n]precise
	coronal	i[n]decent	i[m]decent
	dorsal	i[n]capable	i[m]capable
<b>un-</b>	labial	u[n]prepared	u[m]prepared
	coronal	u[n]dying	u[m]dying
	dorsal	u[n]crossed	u[m]crossed

Table 3: Transforms, center, and scale values for continuous factors included in the statistical models.

Factor	Transform	Center	Scale
RT	log - 500ms	—	—
Trial	—	160.00	92.09
Length	—	9.11	2.03
VocabularyScore	—	64.27	8.53
Frequency	log	3.59	1.04
StemFrequency	log	4.38	2.16
UniquenessPoint	log	1.61	0.43
H250	—	32.91	8.17
H500	—	24.94	6.78
H1000	—	20.88	5.56
H2000	—	21.81	6.24
H4000	—	25.09	7.26
H8000	—	12.94	5.89

Table 4: Summary of log-transformed reaction time data for subjects in the ‘word’ response group, for words (hits) and modified words (false alarms).

		Words		Modified Words	
		mean	sd	mean	sd
IN	labial	6.62	0.53	6.60	0.59
	coronal	6.62	0.58	6.72	0.60
	dorsal	6.68	0.45	6.71	0.48
UN	labial	6.58	0.52	6.63	0.52
	coronal	6.66	0.49	6.67	0.52
	dorsal	6.68	0.49	6.72	0.50
Filler		6.58	0.51	6.97	0.53



Table 5: Summary of log-transformed reaction time data for subjects in the ‘modified word’ response group, for words (false alarms) and modified words (hits).

		Words		Modified Words	
		mean	sd	mean	sd
IN	labial	7.54	0.54	7.49	0.43
	coronal	7.16	0.63	6.87	0.85
	dorsal	7.05	0.41	7.10	0.45
UN	labial	7.06	0.60	7.03	0.50
	coronal	7.16	0.53	6.82	0.50
	dorsal	7.40	0.22	6.72	0.51
Filler		7.08	0.53	6.90	0.51

Table 6: Summary of accuracy data for subjects in the ‘word’ response group, for words (hits) and modified words (correct rejections).

		Words		Modified Words	
		mean	se	mean	se
IN	labial	87.01	2.53	20.22	2.98
	coronal	88.40	2.39	25.14	3.25
	dorsal	87.01	2.53	24.04	3.17
UN	labial	92.00	1.92	22.50	3.31
	coronal	80.36	3.07	39.58	3.54
	dorsal	95.48	1.57	34.43	3.52
Filler		93.18	0.54	83.54	0.98

Table 7: Summary of accuracy data for subjects in the ‘modified word’ response group, for words (correct rejections) and modified words (hits).

		Words		Modified Words	
		mean	sem	mean	sem
IN	labial	91.60	2.43	19.46	3.25
	coronal	91.97	2.33	18.88	3.28
	dorsal	86.47	2.98	25.85	3.62
UN	labial	93.20	2.08	24.06	3.72
	coronal	90.71	2.46	45.71	4.23
	dorsal	95.39	1.71	35.94	4.26
Filler		94.84	0.54	68.39	1.39

## Figures captions

Figure 1: Word Response Group: mean response latency for words (hits) and modified words (false alarms), broken out by prefix and stem. There is no significant difference in responses to words. In modified words, labial stems have shorter response times than dorsal stems in both prefix sets ( $p = .03$ ).

Figure 2: Modified Word Response Group: mean response latency for modified words (hits) and words (false alarms), broken out by prefix and stem. There is no significant difference in responses to words. In modified words, labial stems have shorter response times than coronal stems ( $p = .003$ ) and dorsal stems ( $p = .04$ ) in both prefix sets. Responses to ‘un-’ items are faster than ‘in-’ items ( $p = .0006$ ).

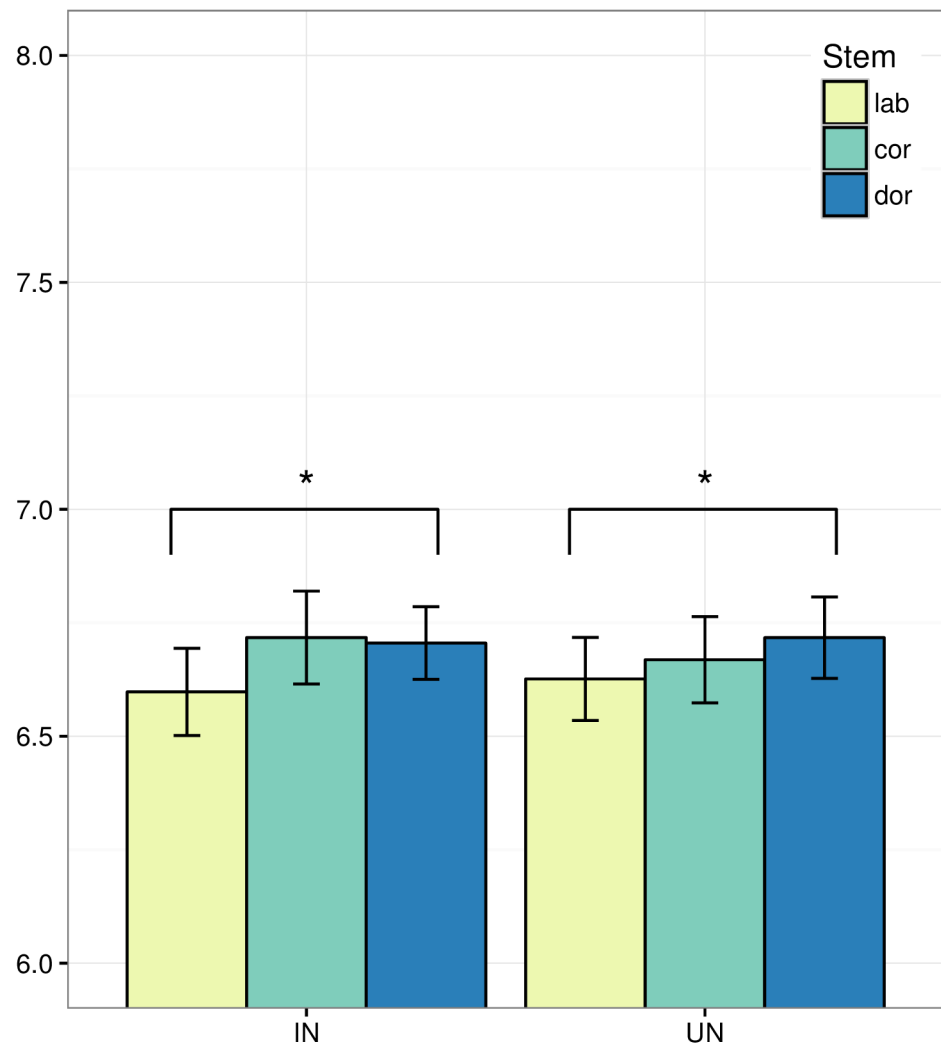
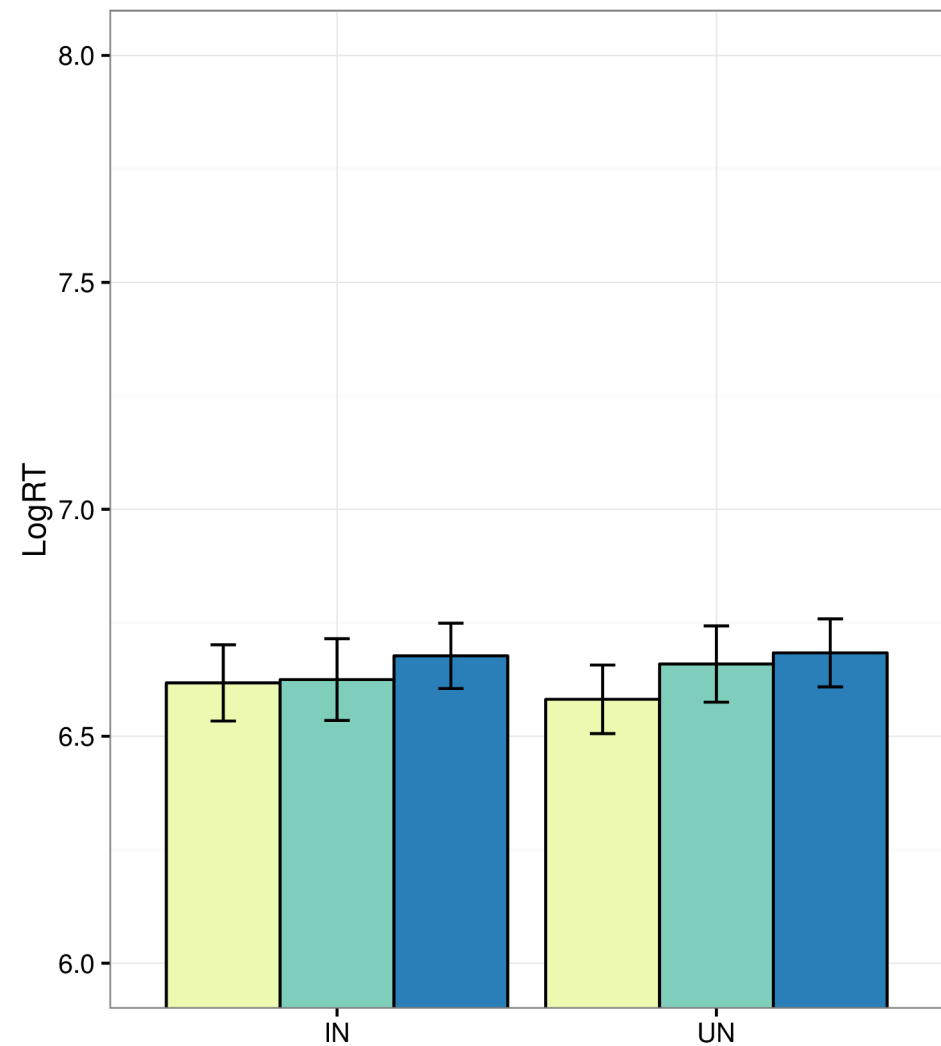
Figure 3: Word Response Group: mean accuracy for words and modified words, broken out by prefix and stem. There is no significant difference in responses to words. In modified words, responses to ‘un-’ stimuli were more accurate than responses to ‘in-’ stimuli ( $p = .01$ ). Responses to labial stems were less accurate than coronal stems ( $p = .003$ ) and dorsal stems ( $p = .02$ ) in both prefix sets.

Figure 4: Modified Word Response Group: mean accuracy for words and modified words, broken out by prefix and stem. There is no significant difference in responses to words. In modified words, responses to ‘un-’ stimuli were more accurate than responses to ‘in-’ stimuli ( $p = .0001$ ). For ‘un-’, responses to labial stems were less accurate than coronal stems ( $p = .001$ ).

Word Response Group

Words (Hits)

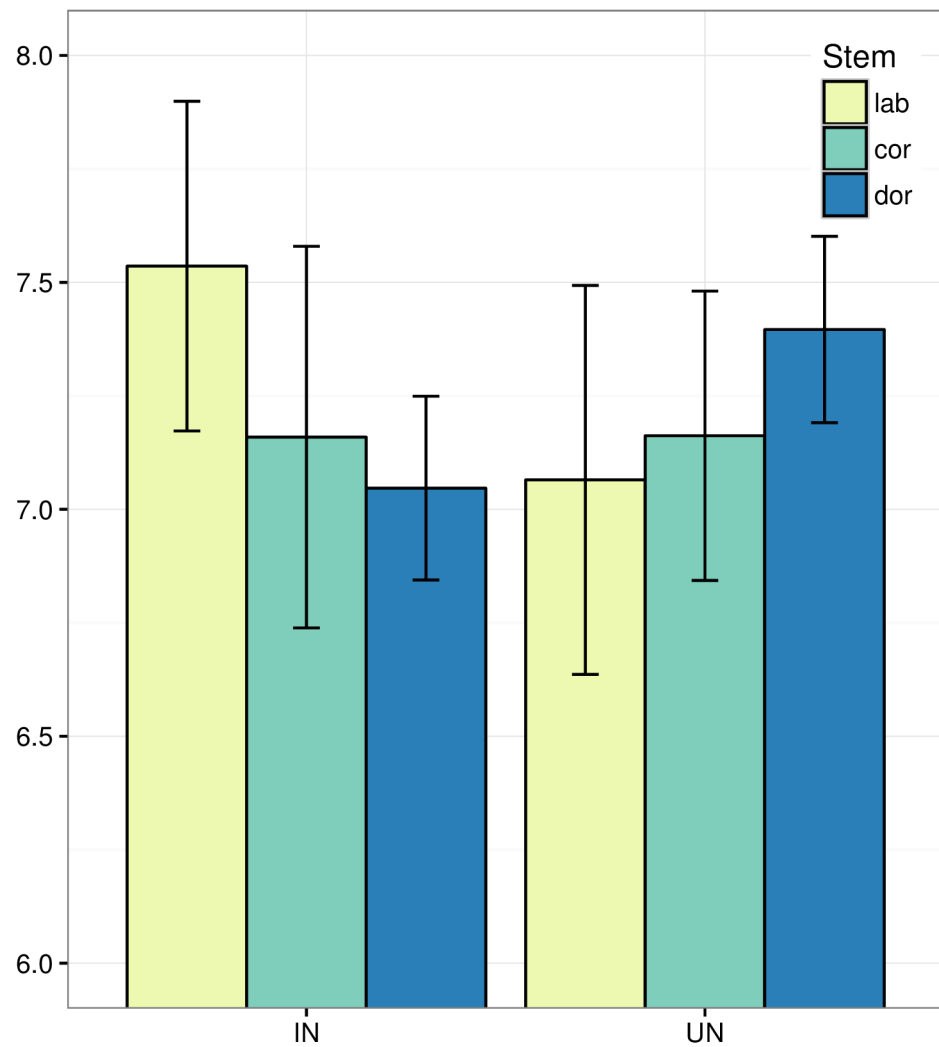
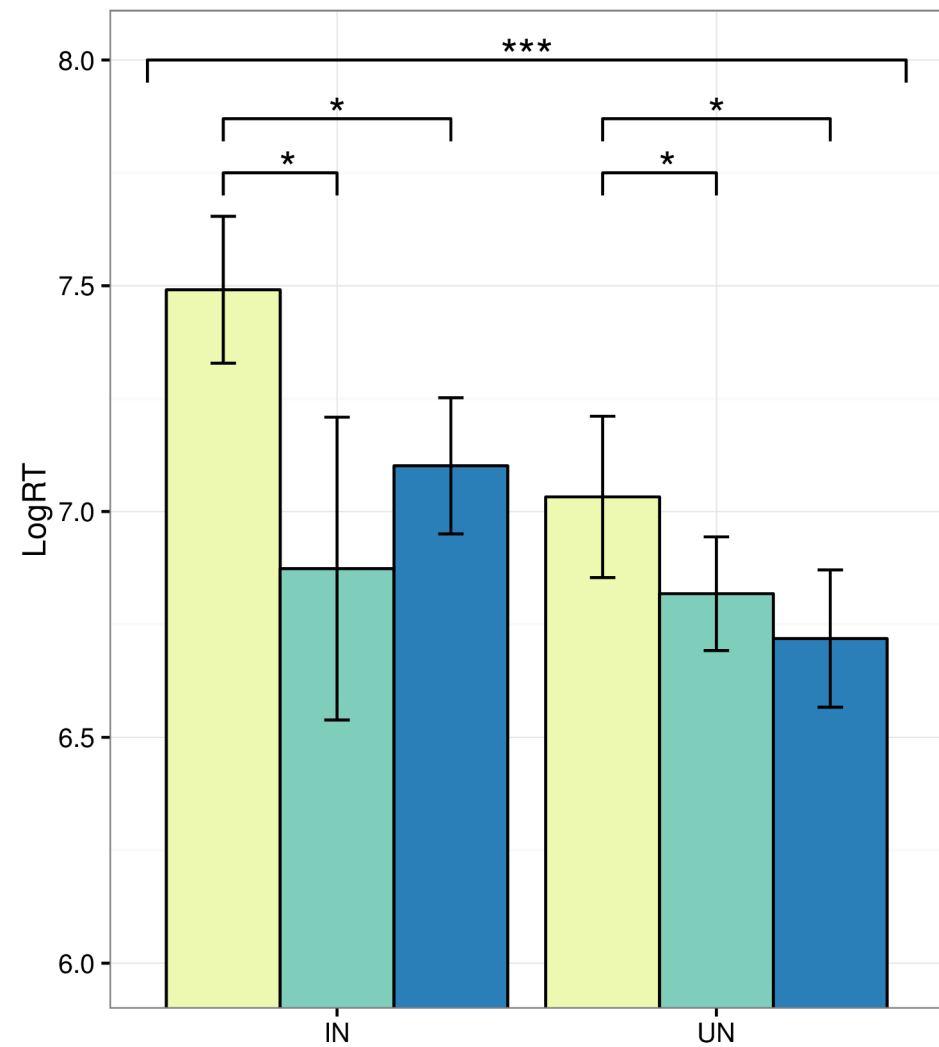
Modified Words (False Alarms)



## Modified Word Response Group

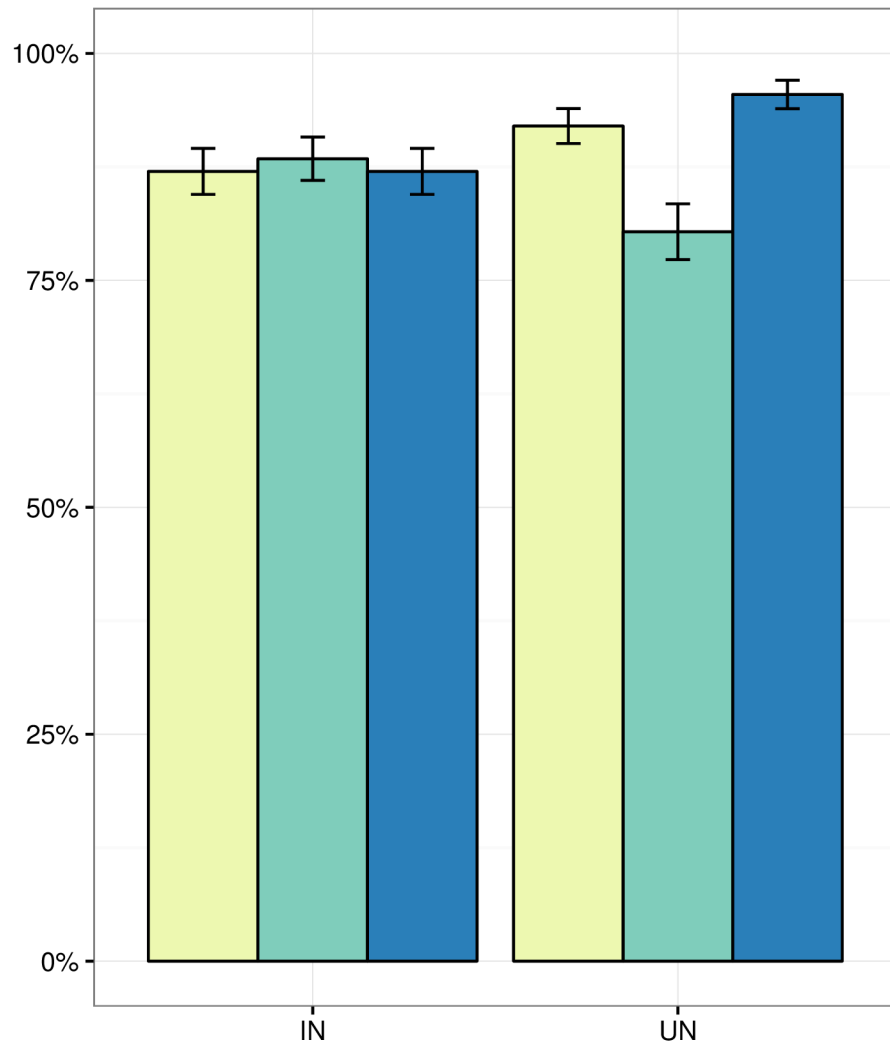
Modified Words (Hits)

Words (False Alarms)

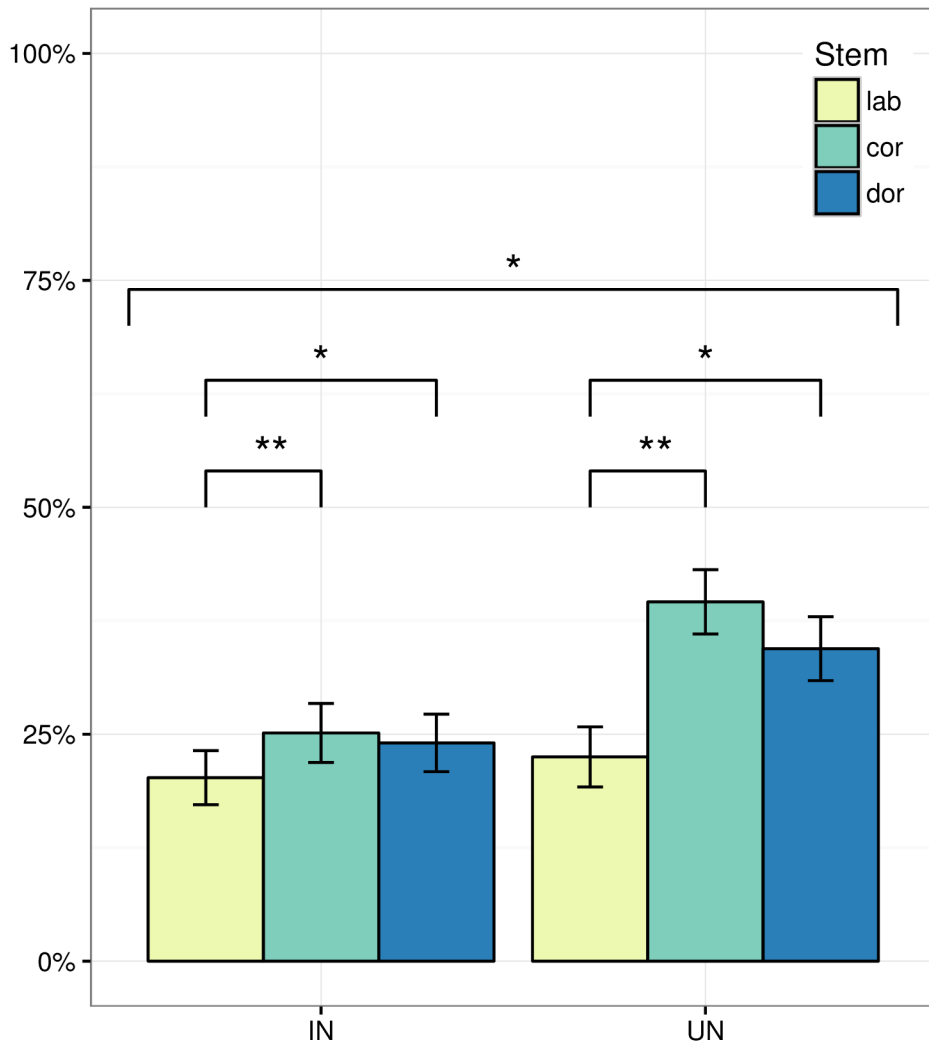


# Word Response Group

## Word Classification

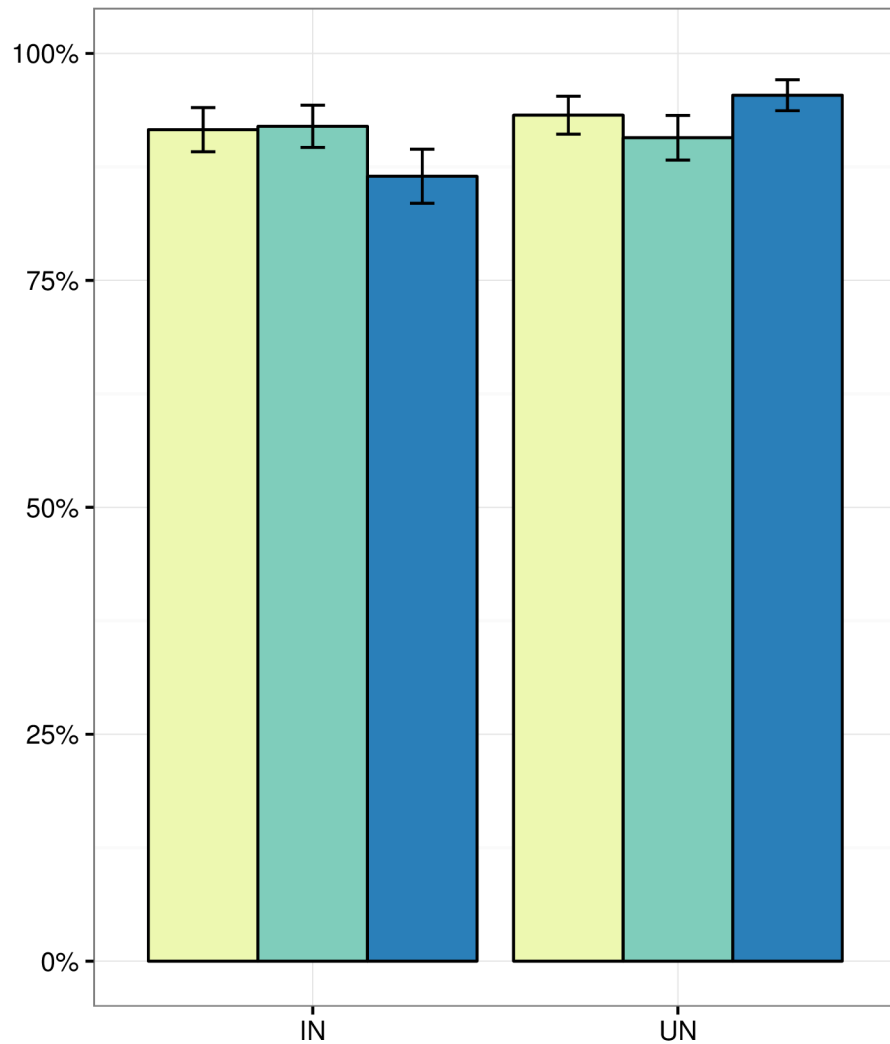


## Modified Word Classification



# Modified Word Response Group

## Word Classification



## Modified Word Classification

