Aligning mispronounced words to meaning: Evidence from ERP and reaction time studies

Adam Charles Roberts, Allison Wetterlin, and Aditi Lahiri

University of Oxford

This is a pre-copyedited, author-produced PDF of an article accepted for publication in The Mental Lexicon *following peer review. The definitive publisher-authenticated version* (Roberts, A.C., Wetterlin, A., & Lahiri, A. (2013) Aligning mispronounced words to meaning: Evidence from ERP and reaction time studies. *The Mental Lexicon 8:2,* 140-163.) *is available online at: 10.1075/ml.8.2.02rob*

Abstract

Many models have been proposed to account for the role that the mental lexicon plays in the initial stages of speech perception. One fundamental disparity between these models is how speech is phonologically represented in the mental lexicon. Theories range from full specification and representation of all phonological information to sparse specification. We report on two perception experiments using context independent mispronunciations (i.e. mispronunciations not governed by phonological rules) to test the predictions of the two most divergent models. Models assuming full specification and storage of all phonological information (e.g. exemplar models) predict symmetric acceptance or rejection of mispronunciations that only differ from real words in place of articulation of the medial consonant (**temor-tenor, *inage-image*). Models assuming that only contrasting phonological information is stored (as in FUL) predict asymmetric patterns of acceptance, i.e. mispronunciations with medial coronal consonants will be better tolerated (**temor*) than mispronunciations with medial labial or velar consonants. Results of two experiments using lexical decision with semantic priming in British English reveal an asymmetry in the acceptance of mispronunciations for coronal vs. noncoronal consonants. Both reaction time latencies as well as N400 amplitudes exhibit asymmetries, supporting the notion of abstract asymmetric lexical representation.

Keywords: N400, semantic priming, word recognition, mispronunciations, ERP

Introduction

Language transmission is often a challenge in a noisy world and we frequently tax our brains to decode words since their pronunciations are variable. One could draw a difference between ‘mispronunciations’ and ‘contextually viable variants’, where the latter occur when sounds assimilate or alter in the context of adjacent sounds. Two main types of contextual variations have been chiefly examined; changes in place of articulation of word final consonants due to assimilation (*green bottle* > *gree[m] [b]ottle*), and consonant deletion or reduction word medially (*ce[nt]re* > *ce[n]er*). Not all variants are equally difficult to resolve. Most studies have argued that viable variants are retrievable in appropriate phonological or discourse contexts. The controversy that still exists is whether the acceptance of certain phonological variants is only dependent on context or whether there is an asymmetry in the phonological representation of the real word as well. This paper investigates listeners' ability to resolve certain mispronunciations, which are not rule-governed, in search of the systematic processes that determine how words and speech sounds are stored and accessed in the mental lexicon.

The majority of studies on the perception of phonologically legal and rule-based variations have focused on either place assimilation **greem bottle* (Gaskell & Marslen-Wilson, 1998; Gow, 2003; Gaskell & Snoeren, 2008; Mitterer & Blomert, 2003; Wheeldon & Waksler, 2004) or stress based reductions, e.g., *center* > *cener* (Ranbom & Connine, 2007; Mitterer & Ernestus, 2007; Pitt 2009; Zimmerer, Reetz, & Lahiri, 2011). Other studies investigated mispronunciations that are not rule-governed such as word initial place alternation **domato* for *tomato* (Connine, Blasko, & Titone, 1993; Bölte & Coenen, 2002) or word medial **Horbe* for *Horde* (Friedrich, Eulitz, & Lahiri, 2006). We discuss rule-based variations first and then turn to context free mispronunciations.

The explanations offered in the literature as to how listeners retrieve the corresponding real word for context-dependent variants focus on two aspects: representation and processing. Models of representation differ notably in the specification of phonological information stored in the mental lexicon. The two extremes are full specification versus underspecified representation. Models assuming full specification (e.g., exemplar models) suggest that very detailed acoustic information of individual words and all possible variants are stored (cf. Coleman & Pierrehumbert, 1997; Pierrehumbert, 2001; Johnson, 2007; Ettlinger & Johnson, 2009; Ranbom & Connine, 2007). Here word representations include all contextually possible variants such as flaps or assimilated forms (**greem*), as well as any token of a sound a person has ever experienced. Models assuming abstract representations suggest that not all variants are stored (cf. Connine et al., 1993: Lahiri & Marslen-Wilson, 1991, 1992; Marslen-Wilson, Nix, & Gaskell, 1995; Wheeldon & Waksler, 2004; Lahiri & Reetz, 2010).

In terms of processing of rule-based variations, many researchers assume that contextual inference constrains lexical activation (Gaskell & Marslen-Wilson, 1998; Gaskell & Snoeren, 2008). This suggests that a variant is more likely to be properly and easily recognised in a correct than in an incorrect context (i.e., **greem* is more likely to be accepted as *green* before *beans* than before *grass*). The experimental task most often used has been lexical decision of the variant form in appropriate (sometimes biasing) contexts (cf. Marslen-Wilson et al., 1995; Wheeldon & Waksler, 2004; Gaskell & Marslen-Wilson, 1996, 2001; Gow, 2002; Mitterer & Blomert, 2003) where the principal finding is that **greem* activates *green* in a viable context, e.g. *beans,* but not in an inappropriate context, e.g., *grass*. A similar pattern of results is found by Mitterer & Blomert (2003) for Dutch who also show that in a passive listening task, variants like **greem* trigger higher Mismatch Negativity (MMN) only in an unviable and not in a viable context. Although Gaskell and colleagues assume assimilation to be complete and the viability effects depend on higher level semantic integration of the nonword, Gow's research suggests that such variations always contain acoustic traces of the real word (e.g., [m] in **greem* has traces of [n]); thus, sublexical mechanisms can abstract acoustic features of variants which are never completely altered in context.

A different set of results were obtained by Wheeldon & Waksler (2004) and Gumnior, Zwitserlood, & Bölte (2007) who find that the rule-based variant **greem* primes *green* irrespective of context (e.g. **greem* activates *green* equally well before both *beans* and *grass*). However, Gumnior and colleagues find a more complex pattern of results, where the identity condition primes more than either contextually appropriate or inappropriate changes; that is, *green* is always better than its rule-based variant. We should also note that the variants like **greem* in isolation also activate the real word, both in repetition form priming (e.g., Marslen-Wilson et al., 1995) and in semantic priming with German word-final consonants (Lahiri & Reetz, 2002). Our own approach also has a goodness of fit component, where the initial process automatically extracts and maps features onto abstract underspecified representations of words and unless there is a direct conflict, the variant will be accepted by the real word (i.e., **greem,* will be initially accepted regardless of context). The same underspecification hypothesis is entertained by Wheeldon and Waksler (2004) and to some extent by Gumnior et al. (2007) who add special importance to the real word. That is although **greem* activates *green*, the real word recognition is faster.

Another type of rule-based variation is due to deletion or reduction of medial consonants (*ce[nt]re* > *ce[n]er*) which has been examined by several scholars (Ranbom & Connine, 2007; Mitterer & Ernestus, 2007; Pitt, 2009; Zimmerer et al., 2011). Results show that recognition of such word variants is possible within the legitimate phonological context (Pitt, 2009). Furthermore, Pitt (2009) also shows that participants can even learn pseudo-word variants in a correct context (e.g., **senty, *seny*), after being exposed to these items in dialogues.

Thus, a large body of research assumes that rule-based variants are better accepted as words in a viable rather than in an unviable context. However, some researchers find that these variants can also be recognised in isolation, and the reason behind this is thought to be underspecified representations rather than experience based storage, which we will discuss further below. We now turn to non rule-governed mispronunciations which by definition occur independently of any particular context.

Studies on non-contextual, non rule-governed mispronunciations confront listeners with words that have changes which are not caused by neighbouring sounds. These include word-initial changes without context (e.g., *pattern* > **kattern*) or a word medial change such as *sonnet* to **sommet*. Research examining word-initial variation indicate that a certain degree of mispronunciation of these consonants is tolerated. In form priming experiments both in English and German, a one-feature change as in *tomato* to **[d]omato* led to activation of the real word (Connine et al., 1993; Bölte & Coenen, 2002). A two-feature change (*tomato* to **[z]omato*) also enabled form priming in English, but to a lesser degree than a one-feature change and was restricted to low frequency words (Connine et al., 1993) whereas German showed no significant difference in form priming for the one and two-feature changes (Bölte & Coenen, 2002). However, using semantic priming, Bölte & Coenen (2002) found that while mispronunciations like *[d]omato, *[z]omato both primed paprika, a mispronunciation with two or more feature changes primed significantly less than with only one feature change. Connine et al. (1993) proposed that acoustic information is mapped onto the lexical representation for the best match possible (**domato* was closest to *tomato,* differing only in voicing). In a cross-modal fragment priming EEG experiment, Friedrich, Lahiri, & Eulitz (2008) also report that altering the place of articulation of initial consonants in German led to recognition of some words but not others. Specifically, a change in a coronal consonant (e.g. *Drachen* 'dragon' to **Brachen*) was acceptable, but non-coronal to coronal was not (e.g., **Drenze* for *Grenze* 'border' was not).

Word-medial consonant changes which are not rule governed have rarely been studied. Examining the neurophysiological correlates of spoken real and mispronounced German words (*Horde* ‘horde’ > **Horbe*, *Probe* ‘sample’ > **Prode*), Friedrich, et al. (2006) found that some mispronunciations were tolerated but others were not. Although N400 amplitudes for all mispronounced words were enhanced compared to real words, ERPs revealed that the N400 non-word effect started earlier for non-coronal mispronunciations than for coronal mispronunciations. Thus, coronal mispronunciations (e.g., **Horbe*) were accepted as words longer than the non-coronal mispronunciations (**Prode*).

One issue inherent in models of word recognition is asymmetry. In models where viability of contexts enhances acceptability of nonword but rule-governed variants, there is an asymmetry in processing. However, models exist which assume asymmetry in representation as well. Exemplar models, for example, assume that the most frequent tokens have the highest chance of attracting an unclear acoustic percept. Our notion of asymmetry, established in the FUL model (*Featurally Underspecified Lexicon,* Lahiri & Reetz 2002, 2010), however, assumes asymmetric feature representations which cause certain mispronunciations to be more viable than others. For instance, phonological changes that produce word variants are often asymmetric; word-final coronal consonants like [n, t, d] take on the place of the following consonant, allowing *green* to become *greem* before *beans*, but, non-coronal consonants such as [p b m k g] remain unchanged despite the context. Thus, *tame tiger* does not become **ta[n]e [t]iger* by any phonological rule. That is, in production a coronal consonant can become non-coronal, but not the reverse. We claim that this is due to coronal consonants being underspecified for place of articulation in representation. The processing system extracts the place feature from the signal (e.g., [labial] from **greem*) and matches it on to the underspecified representation of the /n/ in *green*. The feature [coronal] from **tane,* however, conflicts and mismatches with the stored feature [labial] of *tame*, and the word is not activated. Our processing results show this asymmetry, not just for word final variants which may be subject to assimilations, but also for word initial variants.

Studies on phonological variation have largely concentrated on the viability of the context and not in asymmetry in the assimilation itself. Thus, the focus has been on whether **greem* is equally as acceptable in the context of *beans* (correct context for assimilation) as in the context of *grass* (incorrect context); the fact that only a certain subset of consonants undergo the assimilation, and not others, has been largely disregarded. Thus, *tame* cannot have a phonologically governed variant. Our focus is precisely on this type of asymmetry.

Present Study and Hypotheses

The focus of this paper is on the possible asymmetry in perception of word-medial mispronunciations rather than contextual variants. If these mispronunciations show asymmetric perception concerning place of articulation, then context could not have played a role and models assuming anything other than asymmetric specification of sounds in the mental lexicon cannot account for the asymmetry. Models allowing for storage of detailed information of individual words (e.g. exemplar), hypotheses assuming contextual inference, and explanations based on acoustic similarity cannot explain the systematic asymmetric acceptance of mispronunciations. As mentioned above, the only study addressing context-free, medial consonant variability, has examined mispronunciation of place of articulation in German (*Ho**n**ig* > *ho**m**ig). In an auditory lexical decision task, Friedrich et al. (2006) found an asymmetry for medial consonants in German, demonstrating that certain mispronunciations can still activate the phonology of real words. This has not been replicated in any other language and furthermore it has not yet been tested whether the activation goes beyond the form level and accesses the semantics of the word. Semantic priming paradigms are often used to investigate lexical access. The facilitative effect ('priming') of semantic context on word recognition has been demonstrated in a large number of studies using reaction time (RT) measures. A word (e.g. *river*) is generally recognised with greater speed and/or accuracy if it is preceded by a semantically related word (e.g. *stream*) than by an unrelated context (e.g. *clock*) (Meyer & Schvanefveldt, 1971, Collins and Loftus, 1975; Neely, 1991). Since accessing single words and integrating their meaning are fundamental operations for successful language comprehension, semantic priming procedures are also to examine to what extent similar sounding mispronunciations would activate words in memory by employing lexical decision. Our expectation is that if **sommet* activates *sonnet* it should facilitate recognition of *poem*.

Performance measures such as reaction times, reflect the outcome of the activation, storage and retrieval processes involved in language and memory. Complementary techniques such as recording event-related brain potentials (ERPs) allow a more direct view of the processes taking place as stimuli are actually perceived and subjects perform various tasks (Hillyard and Picton, 1987). The semantic priming paradigm is frequently used to vary the amplitude of the N400 response (cf. Bentin, McCarthy, & Wood, 1985, Holcomb & Neville, 1990, Holcomb & Anderson, 1993) an ERP component that has been found to reflect cognitive processing in word recognition and semantic integration. The prime is considered the 'context' into which the target word must be integrated. Semantically anomalous words produce larger N400 responses than related words because integration is suggested to be more difficult when expectations are not met than when they are met (Kutas & Hillyard, 1980). This implies that any factor that facilitates lexical access should reduce the N400 amplitude. Word-like nonwords, which are associated with long reaction times in lexical decision tasks, often elicit larger N400s than real words (Deacon, Dynowska, Ritter, & Grose-Fifer, 2004).

In this paper, our interest lies in investigating both performance measures as well as ERPs. Since brain-imaging studies require repetitive presentation of the stimuli, the designs of the two experiments were different. In Experiment 1, we used a latin square design with no repetition of stimuli, while in Experiment 2 a typical blocked EEG design was used (see Methods). Reaction time measures in an EEG experiment are notorious for lack of sensitivity given the repeated stimuli.

We predict an asymmetry between types of mispronunciations, as certain words have fewer phonological features in their lexical representation and would therefore better tolerate mispronunciations. Specifically, we expect that mispronunciations of words with medial coronal consonants would be tolerated and consequently prime corresponding targets, but that words with mispronounced non-coronal medial consonants would not be tolerated and thus would not prime. The asymmetry in representation should be manifested both in reaction time (RT) latencies (Experiment 1) and N400 amplitude (Experiment 2). We anticipate that words with mispronounced coronal consonants (**temor*) will activate the lexical representation of the corresponding real words (*tenor*), and thus their synonyms (e.g. SINGER), which translates into shorter reaction time latencies and similar sized N400 as the real words. Furthermore, mispronunciations of words with non-coronal medial consonants should show no activation, i.e. the mispronunciation **inage* would be rejected and show a larger N400 peak and longer reaction times than *image* for the target PICTURE (Table 1).

Table 1 - Examples of stimuli and predictions

Place	**Prime-type**	**Wordness**	**Prime**	**Target**	**RT priming**	**N400**
coronal » non-coronal change	Related	real-word	tenor	SINGER	✓	X
		nonword	*temor			
	Unrelated	real-word	muffin		X	✓
		nonword	*muppin			
non-coronal » coronal change	Related	real-word	image	PICTURE	✓	X
		nonword	*inage		X	✓
	Unrelated	real-word	bootie			
		nonword	*boolie			

Methods

Design and Selection of Stimuli

Primes.

Primes were differentiated by Place (*coronal*, *non-coronal*), Relatedness (Prime-Type: *related, unrelated*), and Wordness, (*real-words*, i.e. correct pronunciations and *nonwords*, i.e. mispronunciations) (see Table 1). All real-word primes were disyllabic English nouns with initial stress and medial coronal or non-coronal consonants ([t, d, n] e.g., *tenor* vs. [p, b, k, g, m] e.g., *image*). The nonword primes were created by changing the place of the middle consonant to non-coronal or coronal respectively (e.g., **temor,* **inage*), while ensuring that they were all possible English words following English phonotactics. The control words were phonologically and semantically unrelated to the test primes (e.g., *bootie*). The unrelated nonwords were also created by changing the medial consonants of the controls by one or more features (e.g. **boolie*) and did not necessarily match the place modification in their respective test words. Using the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995), the two groups of semantically related prime words were matched for written frequency and cohort (see table 2).

Table 2 - Matching criteria for prime words (mean values)

Place	written freq	log written freq	spoken freq	log spoken freq	word length	cohort
Coronal	330.39	2.11	123.57	1.53	5.07	15.35
Non-coronal	328.24	1.93	91.43	1.33	5.41	13.24

Targets.

Related word primes and their mispronunciations were matched to semantically related targets (e.g., *tenor, *temor* to SINGER). Prime-target word pairs were selected from a pool of possible word pair combinations by four native speakers of British English, and only pairs that were selected by all speakers were used. All prime-target pairs were then checked for relatedness by questionnaire (see *Analysis of prime-target relatedness*). Since priming effects can vary depending on the type of semantic priming (Moss, Ostrin, Tyler, & Marslen-Wilson, 1995), we chose targets that were not just associatively related to the primes but instead were as close to synonymous as they could be, by ensuring they shared as many semantic features as possible. Each of the two sets (coronal, non-coronal) had 46 word targets (92 total), each paired with a related and unrelated correctly pronounced word prime and a related and unrelated mispronounced prime. The two groups of targets were matched for written frequency (mean frequency: coronal 699.87; non-coronal 549.15) and word length (mean number of letters: coronal 5.07; non-coronal 5.41).

In addition to the 92 real-word targets, 92 non-word targets were created. Target nonwords were matched with disyllabic real-word and nonword primes with initial stress as in the test condition (e.g. *haggis*, **habbis,* *BEW). There were 55 monosyllabic and 37 disyllabic nonword targets.

Analysis of prime words.

Prime word groups were compared for log written frequency (Celex corpus; t(90) = -1.20, p = .233), and log spoken frequency (BNC corpus; t(90) = -1.26, p = .210). To control for effects of competition where multiple cohort items are activated (e.g. as often found in fragment priming, Marslen-Wilson & Zwitserlood 1989), the uniqueness point / deviation point (DP), of each prime word and corresponding mispronunciation was measured in milliseconds (coronal real-word 216.17, nonword 186.41; non-coronal real-word 219.56, nonword 201.85). Effects on DP were analysed using a univariate Analysis of Variance (ANOVA) procedure. There was a significant main effect of *Wordness*, $F(1,180) = 6.24$, $p = .013$, with mispronunciations having an earlier DP than correct-pronunciations, but no effect for *Place*, $F(1,180) < 1$, *ns*, and no interaction of *Wordness x Place*, $F(1,180) < 1$, *ns*.

Analysis of word-targets.

The two groups of targets were compared for log written frequency (coronal: 2.56; non-coronal: 2.36; $t(90) = -1.65$, $p = 0.102$) and word length (mean number of letters: coronal 5.07; non-coronal 5.41; $t(90) = -1.30$, $p = .197$).

Analysis of prime–target relatedness.

To ensure that prime-target pairs were appropriately chosen, and that the semantic priming effect would be comparable between conditions, both related and unrelated prime–target pairs were analysed using a five-point Likert scale of semantic relatedness. Two questionnaires were created to ensure participants saw each target only once. Within each questionnaire items were randomised by participant. Seventy-nine native speakers of British English took part in the questionnaires (Q1: 37, Q2: 42), and participants were excluded if their mean rating fell beyond 2 standard deviations of the group mean (Q1: 3, Q2: 3). Effects on Relatedness were analysed using a Linear Mixed Effects Model (LMM) procedure in SAS for the fixed effects *Prime-type* and *Place*, with *participant* (nested under *Prime-type* and *Place*) included as a random effect. There was a significant main effect of *Prime-type*, with related words judged as more semantically related to the target than unrelated words, $F(1,142) = 2180.61$, $p < .001$, but no effect of *Place*, $F(1,142) < 1$, *ns*, and no interaction of *Prime-type x Place*, $F(1,142) > 1$, *ns* (see table 3).

Table 3 - Relatedness ratings of prime-target pairs

Place	**Prime-type**	**Mean Relatedness 1-5 (SEM)**
Coronal	**Related**	4.01 (0.02)
	Unrelated	1.11 (0.02)
Non-coronal	**Related**	4.02 (0.02)
	Unrelated	1.19 (0.02)

Stimulus recordings

All words and nonwords were spoken by a male native speaker of British English. The stimuli were recorded in a sound-attenuating chamber with an M-Audio Microtrack recorder (sampling rate 44.1kHz). Nonwords were checked against their corresponding real-words to match for duration, medial consonant length and first syllable length (see table 4).

Table 4 - Matching criteria for acoustic stimuli

Place	**Prime-type**	**Duration (ms)**	**Medial consonant length (ms)**	**1st syllable length (ms)**
Coronal	**Related**	441	94	144
	Unrelated	442	96	142
Non-coronal	**Related**	490	98	158
	Unrelated	491	99	159

Experiment 1

Methods

As mentioned earlier, we employed a cross-modal lexical decision task with semantic priming, using Latin Square design (ISI of 0ms) such that no target was ever seen more than once by any participant.

Participants.

33 native speakers of British English (18 female, 3 left handed, average age 19) with normal or corrected-to-normal vision and no hearing or other impairments volunteered for the study.

Procedure.

Participants sat in front of individual 17" CRT monitors in a quiet room with button boxes to record their responses and responded “Yes” with their dominant hand. Reaction times and response accuracy were recorded. Groups of up to 8 were tested simultaneously, while each participant was seated in a separate booth and was unable to see the responses of others.

Disyllabic correctly pronounced words and mispronunciations were presented to the participants as auditory primes. The visual target appeared on the screen 0 ms after the offset of the auditory prime. Participants were asked to judge whether what they saw on the screen was a word in English. They undertook a training block of 7 trials before the experiment which consisted of a block of 184 trials. A block contained every target word only once, such that each participant saw 92 real-words and 92 nonwords pseudorandomised across conditions, with related real-word primes occurring in 12.5% of trials.

Results & Discussion

One participant had an accuracy of less than 90% and was excluded. Out of 92 words, four targets showed errors above 10% and were excluded from analyses. Moreover, RTs < 300 ms and > 1000 ms were also removed (5% of data). Effects on RT were analysed using a LMM procedure for the fixed effects *Prime-type (*related vs unrelated*), Wordness* (real-word / nonword) and *Place* (coronal / non-coronal). The LMM included both *target* (nested under *Prime-type*) and *participant* as random effects. Effects on response accuracy were analysed using logistic regression.

An overview of the data can be seen in Figure 1, where priming effects are illustrated by a *(unrelated word – related word)* reaction time plot. Longer bars indicate greater priming effects in relation to unrelated words. Significant *Difference-of-Differences* t-tests are indicated with *s in the figures. Mean RTs and accuracy are shown in Table 5.

Figure 1 - Experiment 1 Unrelated – Related RT Priming Results
Difference graph showing the degree of priming for coronal (e.g. prime *tenor, temor*; target SINGER) and non-coronal (e.g. prime *image, inage*; target PICTURE) words and mispronunciations when subtracted from the unrelated condition. Higher bars = more priming, significant difference-of-differences contrasts indicated by *.

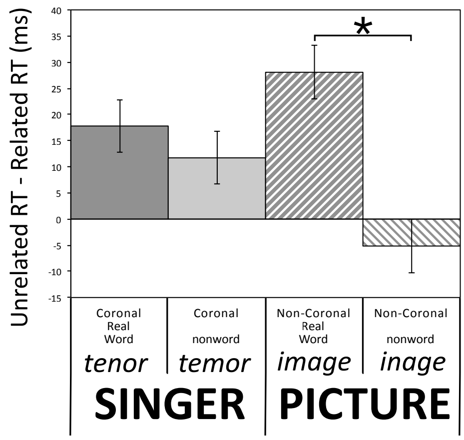


Table 5 – Experiment 1 Reaction Time and Accuracy Results

Place	Wordness	Prime-type	Accuracy	RT(ms)	Priming (ms)	Sig
coronal	real-word	related	98.56%	506.21	17.79	-
		unrelated	97.39%	524.00		
	nonword	related	97.69%	521.23	11.70	
		unrelated	95.32%	532.93		
non-coronal	real-word	related	97.30%	506.57	28.07	*
		unrelated	95.50%	534.64		
	nonword	related	97.90%	537.85	-5.15	
		unrelated	96.03%	532.70		

Errors were low across all conditions, with a significant effect for Prime-type only (related = 2.14%, unrelated = 3.94%), $\chi^2(1, N = 2554) = 6.83$, p = .009, where participants made more errors in unrelated cases than in related cases.

Faster responses were elicited by real-words than nonwords (*Wordness*), F(1,2488) = 19.64, *p* < .001; and by related primes in comparison to unrelated primes (*Prime-type*), F(1,2480) = 19.16, *p* < .001. We found a significant two-way interaction of *Prime-type x Wordness*, F(1,2485) = 10.72, *p* = .001 as well as a significant three-way interaction of *Prime-type x Wordness x Place*, F(1,2485) = 5.09, *p* = .024. Two sets of planned comparisons were performed on the three-way interaction of *Prime-type x Wordness x Place*, the first (*Priming*) looking at the priming effect within *Prime-type* (coronal real-word / coronal nonword / non-coronal real-word / non-coronal nonword), defined as: {*unrelated word – related word}*. *Priming* tests showed significant effects in the coronal real-word condition, t(2482) = 2.93, *p* = .003; coronal nonword condition, t(2482) = 1.98, *p* = .048; and non-coronal real-word condition, t(2482) = 4.24, *p* < .001;

but not in the non-coronal nonword condition, t(2482) = -0.86, $p = .390$ indicating these mispronunciations did not semantically prime their targets.

The second set of planned comparisons (*Difference-of-Differences*) looked at the difference of priming effects within *Place* (coronal / non-coronal), defined as:

{*(unrelated real-word – related real-word) – (unrelated nonword – related nonword)*}

This showed a significant non-coronal effect, with real-words priming more than nonwords, t(2485) = 3.91, $p < .001$, but no significant effect was found for the coronal t-test, t(2485) = 0.72, $p = .474$. An asymmetry in the acceptance of nonwords was evident — coronal words and nonwords successfully primed the semantically related target, but only non-coronal words and not the corresponding non-words primed their targets.

Our results indicate an asymmetry between acceptance of coronal and non-coronal mispronunciations, with coronal mispronunciations eliciting the same degree of priming and acceptance as real-word coronal primes, whereas non-coronal mispronunciations behaved like unrelated controls. There was no significant difference between the real coronal word and its variant. Thus, **temor* and *tenor* primed SINGER equally well, while **inage* failed to prime PICTURE although there was significant priming when PICTURE was preceded by *image*.

Although the performance measures led to our predicted asymmetry, we wished to complement these findings with a more direct measure of when and how these processes were taking place. We therefore ran a separate experiment using the same cross-modal priming task while recording the ERPs of the subjects. However, though the stimuli were identical, the presentation procedure was adjusted to confer to an EEG experiment. ERP studies typically use a large number of trials to allow for averaging within conditions, and involve blocked repeats of target stimuli. Blocked designs are sub-optimal for priming tasks particularly those which predict asymmetries, where the effect of repetitions can severely reduce the magnitude of priming. However in experiment 2, the focus was on differences in ERPs rather than in reaction time.

Experiment 2

Eliciting N400 in lexical decision tasks with semantic priming is well established. We expect higher amplitude N400 responses to unrelated words than related words (Kutas & Hillyard, 1980), and for nonwords in comparison to real-words (Deacon, Dynowska, Ritter, & Grose-Fifer, 2004).

Methods

Participants.

Twenty-two right-handed native speakers of British English (11 female; average age 30) with normal or corrected-to-normal vision and no hearing or other impairments volunteered for the study. Handedness was assessed using the Edinburgh Handedness Inventory (Oldfield, 1971). All participants were appropriately reimbursed.

Procedure.

Participants sat in an electrically shielded, sound attenuated booth and made lexical decisions to stimuli presented on a 20” TFT monitor. The same presentation procedure was used as earlier except that the visual target appeared on the screen 500ms after the offset of the auditory prime. Response mapping was changed for half of the participants, so that 50% pressed “yes” with their dominant hand and 50% with their non-dominant hand to compensate for response preparation effects in the contralateral hemisphere (Kutas & Donchin, 1977). Speed and accuracy were equally emphasised. Participants undertook a training block of 7 trials before participating in the experiment which consisted of four blocks of 184 trials, with 2 minute breaks between blocks to reduce fatigue. A block contained every target word only once, such that in each block a participant saw 92 real-words and 92 nonwords pseudorandomised across conditions, with related word primes occurring in 12.5% of trials.

EEG Recordings.

EEG recordings were taken from 61 sintered Ag/AgCl ring electrodes placed in an equidistant montage, online referenced to Cz (Easycap M1 montage). The electrooculogram (EOG) was recorded from IO1, IO2 and Nz. A cheek electrode served as ground. All electrode impedances were kept below 10 kΩ and signals were digitized at 256Hz.

ERP Analysis.

The continuous EEG data were re-referenced offline to the average of all electrodes, and filtered with a 0.1Hz high-pass and 30Hz low-pass filter. Pre-experimental eye movement data were used to capture characteristic scalp topographies of eye artifacts, which were then applied to an EOG correction algorithm (Ille, Berg, & Scherg, 2002) across the experimental data. To remove other sources of non-EEG noise, trials were rejected if they exceeded an amplitude of 90µV or a gradient of 75µV/division (Scharinger & Felder, 2011). Electrode sites were interpolated to a standard 81 channel 10-10 coordinate space.

Trials with incorrect responses and filler trials (non-word targets) were excluded. In the course of data analyses, three participants were excluded from the study because of too many EEG artifacts (>25% trials rejected). The EEG epochs were averaged with respect to the target stimulus onset for all targets that were classified correctly as words, with a pre-stimulus baseline period of 200ms and a window of 1000ms from target onset. N400 amplitudes were computed on the basis of the signals obtained in four regions of interest (ROI) based on previous ERP research (Scharinger & Felder, 2011; Friedrich et al 2008, 2009; Friedrich 2005; Kutas & Federmeier, 2000; Kutas & Hillyard, 1980), chosen to show effects of N400 and P350 components: Left Fronto-Temporal (C5, F5, FC5, FT7, TP7, T7), Right Fronto-Temporal (C6, F6, FC6, FT8, TP8, T8), Centro-Parietal (Pz, P1, P2, CPz, PO1, PO2) and Fronto-Central (Fz, FCz, F1, F2, FC1, FC2). Visual inspection of the waveforms could not differentiate a clear P350 effect, but the N400 response was determined by visual inspection and set at 250-400ms for all ROIs, taking the mean amplitude in the time window. The N400 is not always negative in absolute terms, so is typically examined using difference waveforms. Here we computed difference waveforms using a point by point subtraction of the related from the unrelated response average within each participant. ERP data were analysed using a LMM procedure with *participant* as a random effect and fixed effects of *ROI* (Fronto-Central / Centro-Parietal / Left Fronto-Temporal / Right Fronto-Temporal), *Wordness* (real-word / nonword) and *Place* (coronal / non-coronal). Planned comparisons (*Difference-of-Differences*) were performed on the three-way interaction of *ROI x Wordness x Place*, looking at the word / nonword difference within *Place* for each *ROI*, defined as:

real-word difference wave – nonword difference wave

Results & Discussion

RTs below 300 ms and greater than 1000 ms were removed (9% of data). Errors were analysed using logistic regression, and were low across all conditions, with significant effects for *Prime-type*, $\chi^2(1, N = 7678) = 4.14$, p = .042, where participants made more errors in unrelated cases than related cases, *Wordness*, $\chi^2(1, N = 7678) = 4.18$, p = .041, where nonword primes caused more errors, and *Place*, $\chi^2(1, N = 7678) = 28.80$, p < .001 where primes in the non-coronal category gave more errors (table 6).

Figure 2 - Experiment 2 Unrelated – Related RT Priming Results
Difference graph showing the degree of priming for coronal (e.g. prime *tenor, temor*; target SINGER) and non-coronal (e.g. prime *image, inage*; target PICTURE) words and mispronunciations when subtracted from the unrelated condition. Higher bars = more priming, significant difference-of-differences contrasts indicated by *.

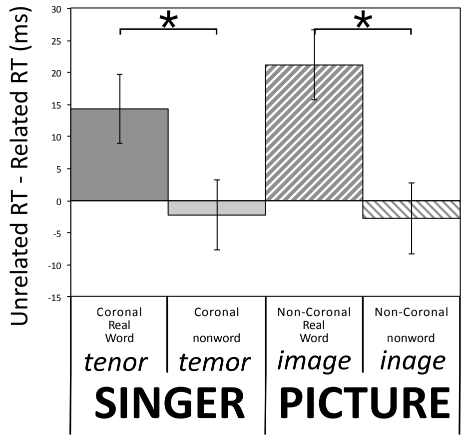


Table 6 – Experiment 2 Reaction Time and Accuracy Results

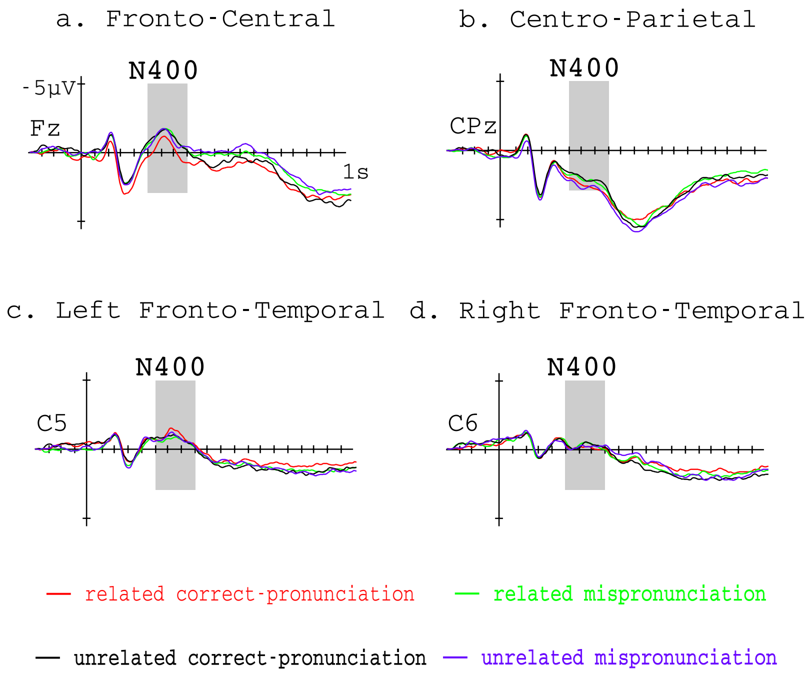
Place	Wordness	Prime-type	Accuracy	RT(ms)	Priming (ms)	Sig
coronal	real-word	related	97.58%	624.19	14.33	*
		unrelated	97.23%	638.52		
	nonword	related	96.44%	635.67	-2.25	
		unrelated	96.46%	633.43		
non-coronal	real-word	related	96.02%	630.85	21.18	*
		unrelated	93.42%	652.03		
	nonword	related	94.84%	644.02	-2.75	
		unrelated	93.15%	641.27		

Effects on RT were again analysed using a LMM procedure in SAS for the fixed effects *Prime-type, Wordness* and *Place*. The LMM included *participant* as a random effect. Faster responses were elicited by related primes in comparison to unrelated primes (*Prime-type*), F(1,7211) = 8.51, *p* = .004, and by coronal rather than non-coronal primes (*Place*), F(1,7211) = 10.42, p = .001. We found a significant two-way interaction of *Prime-type x Wordness*, F(1,7211) = 12.98, *p* < .001. The three-way interaction of

Prime-type x Wordness x Place was not significant, F(1,7211) < 1, ns. However, the two sets of planned comparisons were again performed on the three-way interaction, *Priming* tests showed significant effects in the coronal real-word condition, t(7211) = 2.66, *p* = .008 and non-coronal real-word condition, t(7211) = 3.88, *p* < .001; but not in the coronal nonword condition, t(7211) = -2.25, *p* = .680 or non-coronal nonword condition, t(7211) = -0.27, *p* = .786 indicating both sets of mispronunciations did not semantically prime their targets. *Difference-of-Differences* showed significant effects for both coronal and non-coronal primes (see figure 2), with real-words priming more than nonwords, coronal t(7211) = 2.17, *p* < .030; non-coronal t(7211) = 2.92, *p* = .004. The lack of asymmetry shown here is likely caused by the repetition of stimuli throughout the experiment.

Figure 3 - Coronal ERP waveforms

Grand average ERPs for related and unrelated coronal real-words and mispronunciations at example electrodes in the a) Fronto-Central, b) Centro-Parietal, c) Left Fronto-Temporal and d) Right Fronto-Temporal ROIs. Time windows used for analyses are indicated with grey shading.



Grand average ERPs for selected electrodes in the ROIs are shown in figures 3 (coronal) and 4 (non-coronal), with ERP difference-of-differences waveforms shown in figure 5. In general, the waveform during the N400 time period was more negative for unrelated words than related words, and for mispronunciations compared to real words. The N400 response was maximal in the Centro-Parietal ROI.

The N400 amplitudes showed a significant effect of *ROI*, F(3,1885) = 3.86, *p* = .009 with the Centro-Parietal regions having an overall positive amplitude and Frontal regions an overall negative amplitude, and a significant effect of *Wordness*, F(1,1885) = 6.78, *p* = .009 with real-words having a more positive amplitude. There were significant interaction effects of *Wordness x ROI*, F(1,1885 = 12.66, *p* < .001; and *ROI x Wordness x Place*, F(3,1885) = 2.63, *p* = .049.

Figure 4 - Non-coronal ERP waveforms
Grand average ERPs for related and unrelated non-coronal real-words and mispronunciations at example electrodes in the a) Fronto-Central, b) Centro-Parietal, c) Left Fronto-Temporal and d) Right Fronto-Temporal ROIs. Time windows used for analyses are indicated with grey shading.

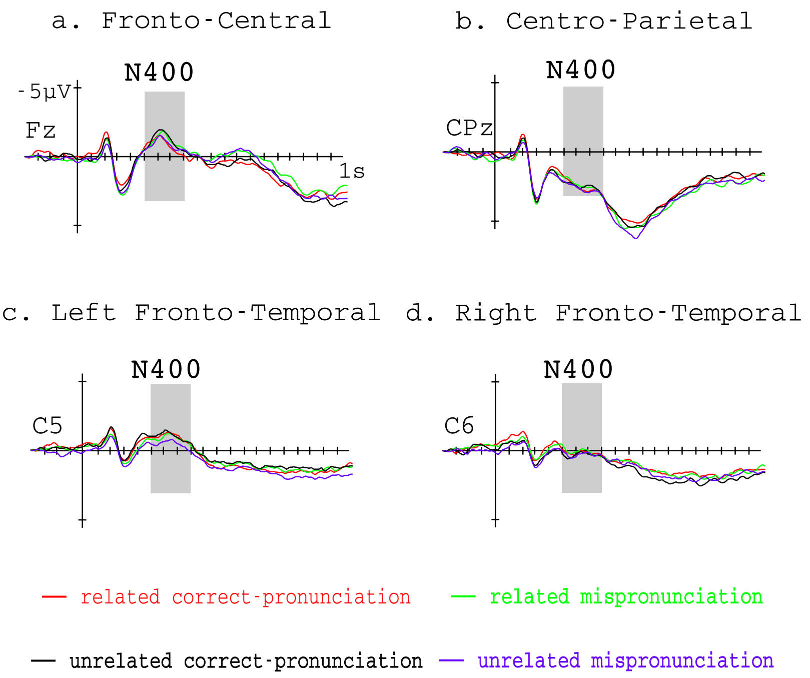
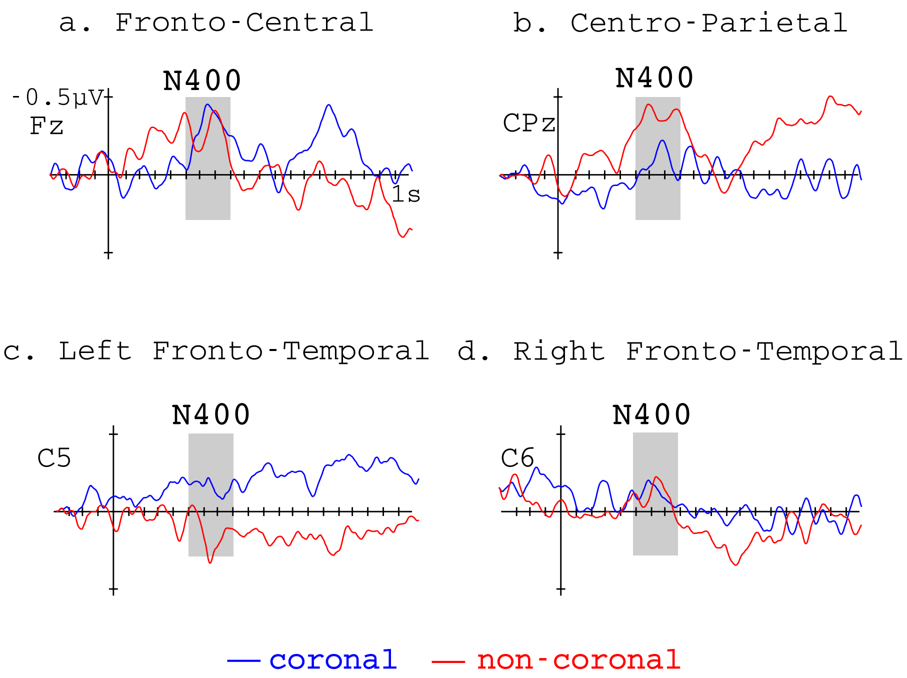


Figure 5 - ERP difference waveforms
Difference of difference plots for example electrodes in the a) Fronto-Central, b) Centro-Parietal, c) Left Fronto-Temporal and d) Right Fronto-Temporal ROIs. Time windows used for analyses are indicated with grey shading.



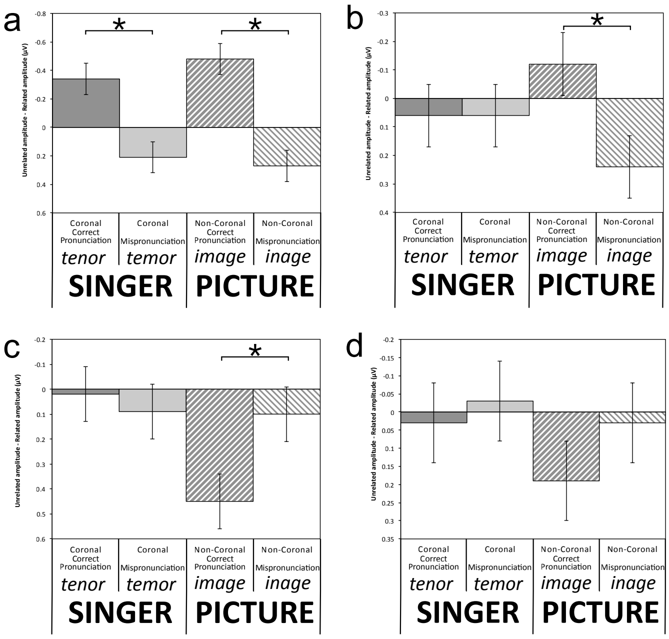
Planned comparisons *Difference-of-Differences* revealed asymmetries in *Place* in both the Left Fronto-Temporal and Centro-Parietal regions, giving significant differences in the non-coronal but not in the coronal tests (see figure 6), where non-coronal nonwords showed a more negative difference wave. There was a clear *real-word– nonword* effect in the Fronto-Central region, with significant differences in both non-coronal and coronal tests (see table 7).

Table 7 – Experiment 1 N400 difference-of-difference contrasts

ROI	**Place**	**t(1885)**	***p***
FC	*coronal*	-3.71	<.001*
	non-coronal	-5.03	<.001*
LFT	*coronal*	-0.41	.680
	non-coronal	2.41	.016*
RFT	*coronal*	0.72	.469
	non-coronal	1.10	.274
CP	*coronal*	0.01	.993
	non-coronal	-2.45	.014*

Figure 6 - Unrelated – Related N400 Results

Difference graph showing the overall eN400 responses for coronal (e.g. prime *tenor, temor*; target SINGER) and non-coronal (e.g. prime *image, inage*; target PICTURE) real-words and non-words when subtracted from the unrelated condition for each ROI. Significant *Difference-of-Differences* contrasts indicated by *. a) Fronto-Central, b) Centro-Parietal, c) Left Fronto-Temporal, d) Right Fronto-Temporal.



The ERP results for Left Fronto-Temporal and Centro-Parietal regions both show an asymmetry between coronal and non-coronal nonwords corresponding to the latency results in experiment 1. As expected, the RT results in this experiment showed priming for real words but no asymmetry, unlike experiment 1. However, the lack of asymmetry is due to the repeated presentations of the target as is usual in an EEG design. Coronal nonwords do not trigger a more negative response for the N400 than coronal real-words, suggesting that they are tolerated as variants of their corresponding words. However, the more negative N400 elicited by non-coronal nonwords in comparison to non-coronal real-words, suggests that they are rejected as nonwords.

General Discussion

Listeners always try to make sense of what is said, even when words are mispronounced. Thus, when faced with mispronunciations caused by changing the place of articulation of medial consonants, which never happens in the regular phonology, listeners still attempt to understand them. Words are represented in the mental lexicon by their phonological form and meaning, and word recognition entails accessing both. If a mispronunciation is automatically accepted, and thus recognized as a variant of an existing word, its meaning should also be accessed. If representations of words are asymmetric such that not all sounds are fully specified, we expect that certain mispronunciations will automatically activate the underspecified representations and hence be more acceptable than others.

Our results confirm the premise that not all mispronunciations are equally difficult to process. Certain mispronunciations gave rise to longer RT latencies and more negative N400s implying a lack of semantic integration. Reaction times showed significant priming by all related real-words and coronal nonwords, but not for non-coronal nonwords. The N400 effect in the Centro-Parietal and Left-Fronto-Temporal ROIs show more negative N400 responses in the semantically related non-coronal nonwords than in the non-coronal words, but no difference between coronal nonwords and words (figure 5). These results suggest that mispronunciations of coronal consonants are automatically accepted as variants as predicted in the FUL model and activate their corresponding semantic neighbours, but that the mispronunciations of non-coronal consonants are not accepted as variants. The asymmetry observed in our results correspond to recent EEG studies on word-initial deviations in fragment priming tasks (Schild, Röder, & Friedrich, 2012) as well as in MMN differences in studies contrasting place and manner of medial consonants in nonwords (Cornell, Lahiri, & Eulitz, 2012).

Lexical storage vs. abstract representation

We have argued that the asymmetry in processing reflected in differential N400 and RTs supports an asymmetric account of phonological representations of words in the mental lexicon. More precisely, words with coronal consonants tolerate more variation than non-coronals. A different view of representation is prescribed by Gaskell & Marslen-Wilson (1998), Locasto & Connine (2002), Pitt (2009), who claim that both phonological inferencing and lexical representational processes operate in conjunction with each other to allow for accurate recognition of word variants. They vary as to which of these two play a more important role. Locasto & Connine (2002) and Pitt (2009) claim that lexical representation has a more dominant role to play than phonological inferencing. As Pitt points out this may be explained by the fact that the two latter papers examine word medial variation, particularly deletion, while Gaskell and his colleagues focus on word final variation.

Our results suggest that it is indeed lexical representation rather than inference from context at play in variant recognition, because we looked at non-contextual word-medial variants which cannot be due to any kind of "inference". Our claim is that representations are abstract and underspecified to the extent that some feature alternations are more tolerable and acceptable than others (cf. Lahiri & Reetz, 2010; Friedrich et al, 2006). In our results, we find an asymmetry that cannot be explained in terms of storing lexical representations of all variants, as Connine and her colleagues, or Coleman & Pierrehumbert (1997) would suggest, because the likelihood of changing coronals to labials is just as feasible as the reverse since they are mispronunciations and not rule-governed alternations.

In conclusion, the pervasiveness of mispronunciations in conversational speech is not trivial. A combination of experimental measures, including both brain imaging and behavioural data, provides us with a clearer idea of how humans respond to non-ideal speech. The initial activation of a word when encountering certain mispronunciations is automatic and cannot be prevented, and this activation allows for the complete lexical representation to be accessed; both form and meaning. However, this does not occur for all mispronunciations. An asymmetry occurs, even in mispronunciations that aren't contextually viable, where for some words a mispronunciation is immediately apparent, but for others it takes time to distinguish it from the real word.

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L., (1995). The CELEX Lexical Database (Release 2). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [http://web.phonetik.uni-frankfurt.de/simplex.html].

Bentin, S., McCarthy, G., & Wood, C. C., (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343–355.

Bölte, J., & Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner? *Brain and Language*, *81*, 384-397.

Coleman, J., & Pierrehumbert, J. (1997). Stochastic Phonological Grammars and Acceptability. *Third Meeting of the ACL Special Interest Group in Computational Phonology*. ACL: Somerset, NJ. 49-56.

Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing *Psychological Review, 82,* 407-428.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193-210.

Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). What you encode is not necessarily what you store: Evidence for sparse feature representations from mismatch negativity. *Brain Research 1394*, 79-89.

Cornell, S. A., Lahiri, A., & Eulitz, C., (2012, December 31). Inequality across consonantal contrasts in speech perception: Evidence from Mismatch Negativity. *Journal of Experimental Psychology: Human Perception and Performance*, Advance online publication, doi: 10.1037/a0030862.

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J., (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*, 60-74.

Ettlinger, M., & Johnson, K., (2009). Vowel discrimination by English, French and Turkish speakers: Evidence for an exemplar-based approach to speech perception. *Phonetica, 66*, 222-242.

Friedrich, C.K., 2005. Neurophysiological correlates of mismatch in lexical access. *BMC Neuroscience* 6, 64.

Friedrich, C. K., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions*, *2*, 1-36.

Friedrich, C. K., 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*, 1545-1559.

Gaskell, M. G., & Marslen-Wilson, W. D., (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human perception and performance*, *22*, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D., (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 380-396.

Gaskell, M. G., & Marslen-Wilson, W. D., (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325-349.

Gaskell, M. G., & Snoeren, N. D., (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1632-1647.

Gow, D. W., (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 163-179.

Gow, D. W., (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*, 575-590.

Gumnior, H., Zwitserlood, P., & Bölte J. (2007). Assimilation in existing and novel German compounds. *Language and Cognitive Processes*, *20,* 465-488.

Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. In F. Plum (Ed.), Handbook of physiology: The nervous system (Vol. 5, pp. 519-584). Bethesda, MD: American Physiological Society.

Holcomb, P. J., & Anderson, J. E., (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*, 379-411.

Holcomb, P. J., & Neville H. J., (1990). Auditory and visual semantic processing: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*, 281-312.

Ille, N., Berg, P., & Scherg, M., (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*, 113-124.

Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In M. Solé, P. Beddor Speeter & M. Ohala (Eds.), *Experimental Approaches to Phonology* (pp. 25-40). Oxford: Oxford University Press.

Kutas, M., & Donchin, E., (1977). The effect of handedness, the responding hand and response force on the contralateral dominance of the "readiness" potential. In D. J. Basel & S. Karger (Eds.), *Attention, Voluntary Contraction and Event-Related Cerebral Potentials* (pp. 189-210). Basel: Karger.

Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences* 4, 463–472.

Kutas, M., & Hillyard, S. A., (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Lahiri, A., & Marslen-Wilson, W., (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 245-294.

Lahiri. A., & Marslen-Wilson, W., (1992). Lexical processing and phonological representation. In G. J. Docherty & R. D. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 229-260). Cambridge: Cambridge University Press.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics 38*, 44-59.

LoCasto, P. C., & Connine, C. M., (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics*, *64*, 208-219.

Marslen-Wilson, W., Nix, A., & Gaskell, M. G. (1995). Phonological variation in lexical access: abstractness, inference and English place assimilation. *Language and Cognitive Processes*, *10*, 285-308.

Marslen-Wilson, & W., Zwitserlood, P., (1989). Accessing Spoken Words: The Importance of Word Onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576-585.

Meyer, D. E., & Schvaneveldt, R. W., (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234.

Mitterer, H., & Blomert, L., (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, *65*, 956-969.

Mitterer, H., & Ernestus, M., (2007). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73-103.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D., (1995). Accessing Different Types of Lexical Semantic Information: Evidence From Priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 863-883.

Neely, J. H. (1991). Semantic Priming Effects in Visual Word Recognition: A Selective Review of Current Findings and Theories. In D. Besner & G. Humphreys (Eds.), *Basic Processes in Reading: Visual Word Recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Oldfield, R. C., (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97-113.

Pierrehumbert, J., (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137-157). Amsterdam: John Benjamins.

Pitt, M. A., (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, *61*, 19-36.

Ranbom, L. J., & Connine, C. M., (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* *57*, 273-298.

Scharinger, M., & Felder, V., (2011). ERP signatures of cross-modal semantic fragment priming: Early context effects in speech perception. *International Journal of Psychophysiology, 80*, 19-27.

Schild, U., Röder, B., & Friedrich, C. K., (2012). Neuronal spoken word recognition: The time course of processing variation in the speech signal. *Language and Cognitive Processes*, *27*, 159-183.

Wheeldon L., & Waksler, R., (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, *90*, 401-412.

Zimmerer, F., Scharinger, M., & Reetz, H., (2011). When beat becomes house: Factors of word final /t/-deletion in German. *Speech Communication*, *53*, 941-954.

Acknowledgements

This work has been supported by the EC within the 7th framework programme under grant agreement no. FP7-IST-269670 (Advanced Investigator's Grant, PI Lahiri), the John Fell Oxford University Press (OUP) Research Fund and the University of Oxford Higher Studies fund.

We would like to give special thanks to Verena Felder for collaborating on the design of the experiments and to Henning Reetz for creating the software that allowed us to run multiple subjects and for many discussions on statistics. We also would like to thank the members of the Oxford Language and Brain Laboratory: Franziska Hack, Steven Kaye, Holly Kennard, Sandra Kotzor, Ben Moulineaux for assisting us in conducting the experiments and recruiting participants.