Wellington Working Papers in Linguistics

Volume 14, 2002

edited by Paul Warren

School of Linguistics and Applied Language Studies
Victoria University of Wellington
P.O. Box 600
Wellington
New Zealand

Published 2003

Printed for
Victoria University of Wellington
P.O. Box 600
Wellington
New Zealand

ISSN 1170-1978

Wellington Working Papers in Linguistics
14, 2002
Contents

Editorial

In keeping with time-honoured tradition, this volume, dated 2002, sees the light of day well into the following year, but to quote an infamous British Rail slogan "we're getting there."

Late in 1994 Victoria University of Wellington Linguistics Department (as then was, now part of the School of Linguistics and Applied Language Studies) became the first New Zealand Linguistics Department to appoint a psycholinguist. A sign of that psycholinguist's busy-ness is that I have only now – over eight years later – edited a volume of these Working Papers that takes psycholinguistics as its general theme. As is clear from their titles, the papers cover a broad range of research. "Affix-able words" started as an interesting study to demonstrate psycholinguistic techniques to undergraduates, and the experiment was run in a class session that Laurie Bauer took in order to enable me to travel to San Francisco for a conference. "Phonological awareness in young Cantonese-Mandarin bilinguals" is a report of an outstanding MA thesis completed in 2002, while "The asymmetrical change in progress of NEAR and SQUARE vowels in NZE" was born out of an Honours project carried out in 2001. Continued collaboration with a former Honours and MA student, now on the Linguistics staff at the University of Canterbury, is reflected in the paper on /r/-intrusion with Jen Hay.

Paul Warren

May 2003

Affix-able words: morphological productivity as the basis for (de)composition in processing

Laurie Bauer and Paul Warren

Abstract

Some of the key issues in research on morphological (de)composition in lexical processing include distinctions between types of affixation, levels of transparency and degrees of regularity. We argue that a central factor in the discussion of such issues is the productivity of affixation. Support for our arguments is presented in the form of results from a visual lexical decision experiment. Subjects responded to a range of forms ending in the English affix *-able*, including existing real words, nonce forms with real word bases and nonwords with nonsense bases. Our results show sensitivity to the difference between the productive combination of *-able* with verbal bases, and the non-productive combination of *-able* with nominal bases. Nonce forms in particular are much more acceptable as real words if they have a verbal base. We discuss these results in the context of models of morphological processing.

Introduction

The experimental research reported in this paper concerns the effects of morphological productivity on the processing of visually-presented morphologically complex words. Morphological productivity is defined for our purposes as the potential for a morphological process to be used in the coining of new forms.

Previous psycholinguistic research has addressed a range of issues concerning the decomposition and construction of morphologically complex forms in the processes of comprehension and production, and on consequent implications for lexical storage. A common theme to much of the research has involved the question of whether morphologically complex forms are constructed and analysed – whether by rule or by analogy – as part of the processes of speech production and comprehension. The alternative full-listing hypothesis assumes that such forms are accessed (in both production and comprehension) as whole forms. Implications for lexical storage clearly concern the representation of complex forms either as a set of separate morphemes[1] or as whole forms.

The issue of productivity is clearly pertinent to such questions, since the creation of new vocabulary through morphological processes assumes that these processes are synchronically transparent. However, although 'productivity may be the determining factor with regard to whether morphologically complex forms are composed in the processing system' (Badecker & Caramazza 1998:400), most research has concentrated on a range of other factors involved in composition (some of which are frequently also

[1] We use the term 'morpheme' as it is used in the psycholinguistic literature, although there has never been agreement among theorists on what a morpheme is (Mugdan, 1986) and although much modern theory dispenses with the notion (see Bauer, 1999a).

Wellington Working Papers in Linguistics 14 (2002) 1-17

assumed to affect productivity). In this paper, we wish to bring the focus back to the notion of productivity as a potential criterion for composition. In the experimental section of this paper we do this with reference to the form *–able* used suffixally. In the remainder of the current section, we summarise some of the key issues in research on the processing of morphologically complex forms, highlighting how they might relate to the question of productivity. These issues include the differences between inflectional and derivational affixation, between regular and irregular inflections, and between transparent and opaque derivational morphology.

In much of the psycholinguistic research, it is assumed that there is a distinction for processing between inflectional and derivational morphology. Although we shall see that further distinctions are made within each of these categories, the starting assumption seems to be that inflectional morphology is more productive than derivational, since the grammatical processes involved are more widespread. The consequence for processing is that inflectional morphology is more likely to involve composition (in production) and decomposition (in recognition). For storage, the implication is that inflected forms are not separately listed from their base forms.

The picture is rarely as simple as this. For a start, theoretical linguists do not always agree on whether and where to draw the distinction between inflection and derivation. In the case of English, inflectional morphology is most generally taken to include verbal *–s, –ed, –ing* and nominal plurals, and possibly also comparative and superlative *–er* and *–est*. However, Beard (1982) has argued that the English plural is in fact derivational, and his arguments have been extended to other languages by Baayen, Lieber & Schreuder (1997; see also references there), while Haspelmath (1996) has argued that adverbial *–ly*, generally assumed to be a derivation, is inflectional. Many linguists including Bauer (1988), Dressler (1989) and Plank (1994) have argued that 'inflection' and 'derivation' are prototypical categories, with considerable deviation possible from the prototypes. Psycholinguistic research, however, has largely assumed a straightforward distinction between inflection and derivation.

Even if there is agreement on what constitutes an inflection and what a derivation, it is clear that there are further distinctions that can be made within each of these categories. These include consideration of the regularity and/or transparency of morphological processes. Thus, experimental studies of inflections, mainly conducted using English materials, tend to distinguish between regular and irregular inflection. A number of experiments have shown that base forms are primed as much by preceding regularly-inflected forms as they are by the base forms themselves (e.g. Stanners, Neiser, Hernon & Hall 1979; Fowler, Napps & Feldman 1985). Other experiments show that this priming from inflected forms is due neither to phonological overlap (Slowiaczek & Pisoni 1986) nor to semantic overlap (Fowler et al 1985; Marslen-Wilson, Tyler, Waksler & Older 1994). This being the case, it would seem that the priming must be caused by the morphemic make-up of the word; when the inflected form is recognised, access is made to the base. The implication is that the inflected form is processed as a combination of base and inflection, that base and inflectional morpheme are separately represented in the mental lexicon, and that there is no representation for the full inflected form.

With irregular inflections we find a different and somewhat inconsistent picture. Stanners et al (1979) report that irregularly inflected words only partially prime their bases (i.e. as primes they are not as effective as the base form itself, but more effective than an unrelated control word). However, Fowler et al (1985) claim full priming from irregular words, though it should be noted that the materials on which they base this claim include derivationally as well as inflectionally related words. These data notwithstanding, the general picture (which is supported by evidence from other sources, such as speech errors – cf Stemberger & MacWhinney 1986) appears to be that irregularly inflected forms are stored separately from base forms, but that the two are linked in the lexicon (giving the partial priming found by Stanners et al). PET scan data collected by Jaeger et al (1996) supports the basic contention that regularly and irregularly inflected forms are processed differently, although the findings of this research have been challenged (e.g. by Seidenberg & Hoeffner 1998, Indefrey 2000).

Within the set of regularly-inflected words, effects of token frequency have also been found, at least in speech production. Stemberger & MacWhinney (1986) report fewer production errors related to the morphology of high frequency inflected forms than for low frequency inflected forms, in both spontaneous and experimental situations. This seems to suggest that high frequency items are stored as whole forms (and thus less susceptible to errors involving their morphological structure) in addition to being stored or accessed via decomposition (as evidenced by the priming data).

These observations for inflectional morphology – that (de)composition is involved in the storage and processing of regular but not irregular forms, and that high frequency regularly inflected words may also be stored as whole forms – suggest some general efficiency model for storage and processing. That is, regularly inflected forms are analysed morphologically, for the very reason that the inflections are regular and widespread. Indeed, since a language such as Archi can have over 1.5 million regularly inflected forms for every verb lexeme (Kibrik 1998:468), storage would be uneconomical in many languages. By contrast, it makes little sense to process irregular forms in this compositional way, since the procedures for doing this would in effect be idiosyncratic and inefficient. So irregular forms are most probably fully represented. In addition, though, a commonly occurring regular base+affix combination may also be represented as an unanalysable whole (as well as compositionally) because the frequency with which that combination is met makes it more efficient to process it as a whole.

Although it is possible for patterns exhibited by irregular forms to be used productively (for example the pattern *wring, wrung* is argued to be productive in English, but it applies only to a very small set of lexemes - Bybee & Moder 1983), the regular forms in English are always productive and apply to the largest set of words. Such forms are also extendable to new forms, even in early stages of language acquisition (Berko Gleason 1958). There is an implication here that the most productive forms must be morphologically (de)composable.

Research on the processing of derivational morphology distinguishes between prefixation and suffixation, and between various degrees of transparency. At the transparent end of the scale we have cases where the

derivational process has no marked effect on the phonological and orthographic form of the base, and where the semantics of the derivative are largely compositional. Various priming effects have been attested for items with such transparent derivation. First, both prefixed forms (Stanners et al 1979; Marslen-Wilson et al 1994) and suffixed forms (Fowler et al 1985; Sandra 1994) prime their base forms. Further, Marslen-Wilson et al found that derived (prefixed or suffixed) forms generally prime other derived forms with the same base. The exception is when two suffixed forms share the same base. The absence of priming in that case is explained in terms of competition effects between two words from the same initial cohort. Segui & Zubizarreta (1985) present data showing priming between words with different bases but the same suffix (in transparent affixation), but not between words sharing the same prefix. Again, we can explain the lack of priming for the prefixed forms in terms of competition between words from the same initial cohort. While Segui & Zubizarreta argue for full-listing of derived forms in linked morphological families, it is clear that access has been possible to the features of the suffix, resulting in the priming of two words with different bases but identical suffixes.

With less transparent derivation (i.e. with Level I or +-boundary affixes which involve changes in stress patterns or important allomorphic changes) there are mixed results. One study shows partial priming of the base from the derived form (Stanners et al 1979), while another reports full priming equivalent to that of totally transparent forms (Fowler et al 1985). Finally, the case of opaque derivation (i.e. where there may be some historical but no clear synchronic basis for morphological analysis or relatedness, e.g. between *conceive* and *receive*) shows no priming.

The overall pattern appears to be that two derivationally related forms must have a clear formal and semantic relationship for priming effects to arise. Where there is partial overlap, Stanners et al argue for separate listing of base and derived forms (which is clearly also the case for synchronically opaque derivations), but with some link between the forms in the lexicon (as was argued also for irregularly inflected forms and their bases). This is compatible with Jackendoff's (1975) notion of redundancy rules in the lexicon. Fowler et al's finding of total priming for semi-transparent derivations may be compatible with this if we assume that some derivatives, like frequent regularly inflected forms, are both stored as whole forms and analysed morphologically.

The experimental evidence sketched above suggests that like regular inflectional processes, transparent derivational processes are dealt with psychologically by morphological (de)composition on-line. However, as far as English is concerned, it is usually the case that the transparent derivational processes, like the regular inflectional processes, are the most productive types. It is our view that morphological productivity is the most important factor here. Productivity cuts across, but is not identical to, both derivational transparency and inflectional regularity. It provides a more unified explanation of the effects summarised above. We hypothesise in fact that only productive morphological processes are available for decomposition, and that words formed by non-productive processes must be stored in some other way. Note, however, that our hypothesis here does not mean that all words obtained via a productive morphological process will necessarily be accessed

via decomposition. For instance, whole-word access may be characteristic of highly frequent regularly inflected forms, as observed above.

Note also that we do not claim that all transparent or regular processes will necessarily be available for decomposition. While there is clearly some kind of relationship between transparency and productivity, the nature of that relationship is controversial, and no doubt indirect. The same is true with productivity and type-frequency. Mayerthaler (1981) argues that all are linked through naturalness. The precise nature of the relationship is not important here, though it is important that productivity is not simply equated with transparency or frequency.

A further complicating factor is that not all transparent affixation is productive to the same extent. This is principally because there are a number of constraints on productivity (cf Bauer 1983: 84-97, 1994), including the notion of blocking introduced by Aronoff (1976), as well as semantic, syntactic, phonological and lexical constraints. These constraints mean that even transparent processes of affixation cannot be guaranteed to be fully productive (if, indeed, any derivational process can be claimed to be fully productive). They also mean that a given process of affixation may be more productive under certain circumstances (e.g. with a particular class of base) than under others.

We further hypothesise, then, that native speakers are sensitive to constraints on the productivity of a morphological process. We argue that this can be demonstrated in their responses to carefully constructed nonce words (novel but legal forms) in which the same affix is combined with bases from two classes, one combination being productive, the other not. Our prediction is that subjects are more likely to accept as real words nonce forms which combine the affix with a base from the productive class. Although such a prediction relates chiefly to the processing of novel forms, i.e. to learning processes, we argue that experimental demonstration of such sensitivity can be taken as support for a general model of morphological processing that involves decomposition. We expand on this argument in the General Discussion.

The form *–able* is a useful one to use in the investigation of our hypotheses, since it can be analysed in the vocabulary of English as having two types of usage, only one of which is productive: whether or not it is productive depends on the class of base to which it is attached. The form *–able* can be added productively to transitive verbs to create new adjectives: Given a new transitive verb such as *koreanise*, the corresponding *koreanisable* is possible. The form *–able* is also found added to nominal bases, as in *knowledgeable*. In both cases the affixation is derivational and transparent. It is true that we have not tried to grade semantic transparency, as is done, e.g. by Derwing (1973: 124). Because the noun+*able* forms are all lexicalised, the contribution of the *–able* suffix in particular to the meaning of the entire lexeme may not always be obvious, but the bases are always transparently recoverable. For our purposes it is irrelevant whether the *–able* in *koreanisable* and that in *knowledgeable* represent different homophonous affixes (Aronoff, 1976: 121-129) or the same affix (Marchand, 1969: 229-232). What is important is that the same form has two distinguishable uses, only one of which is automatically productive.

Experiment

In our experiment, subjects are required to make lexical decisions about a series of real and nonsense words presented one at a time on a computer screen. The tokens include six sets of items with *–able* affixes. Two sets are existing real words with verb or noun bases (e.g. *conceivable, impressionable* – we will refer to items in this set as the *real* words). A further two sets are nonwords constructed by affixing *–able* to existing verbs and nouns (e.g. *entrenchable, omissionable* – these we will call the *nonce* words). In the cases of the words with verbal bases these are, we claim, possible words, but an extensive dictionary search failed to produce any evidence that they are 'existing' words established in the community. In the cases of the words with nominal bases which, we claim, are not possible words in the present state of the language system, there was no obvious reason why these forms could not have been coined at some earlier stage in the history of English, and apparently parallel forms are attested (Bauer, 1999b). The final two sets contain nonsense forms with nonword bases which themselves have derivational morphology that makes them verb-like or noun-like (e.g. *entrulchable, arissionable*, where *entrulch* is a verb-like nonsense base and *arission* a noun-like nonsense base – this set will be referred to as the *nonword* set).

Our first prediction for these materials is a basic lexicality effect: existing words will be identified as real words, and nonwords will be identified as nonwords, and this identification will be faster for the real words.

Secondly, we predict that the nonsense forms with nonword bases will be more consistently identified as nonwords than the nonce forms with real word bases. (See Bauer 1999b for a demonstration that the grammaticality of the base determines the grammaticality of the derivative.)

If the constraints on morphological productivity influence subjects' willingness to accept nonce forms as real words, then we predict that there will be more real word responses to the nonce words with existing verb bases than to those with existing noun bases.

If the nonword bases are identified as verb-like or noun-like, then the productivity hypothesis also predicts more real word responses to the subset of the nonsense words with verb-like nonword bases.

In each case, a simple full-listing hypothesis leads to no clear expectation that nonce forms with verb and noun bases (or nonwords with verb-like and noun-like nonsense bases) will differ, since none of these forms exist in the lexicon.

A compositional account without productivity, in which bases are simply listed with or connected to their affixes, and in which constraints on affixation form no part of the morphological process, predicts no difference between the verb-based and noun-based nonce forms.

Materials

Ten tokens were selected or constructed in each of the six conditions outlined above (verb and noun bases in each of the real, nonce and nonword forms). The selection of tokens was subject to the following constraints.

Real words: It was important that each *–able* word should have a base that was clearly identified as a verb or a noun, and so forms such as *valuable*, where *value* is category ambiguous, were excluded. In addition, the tokens had to be words that would clearly be recognised by our subjects as real words. For this reason, we piloted a set of *–able* words with verb and noun bases with a small sample of subjects in the same age range as the subject population to be used in the main experiment. On the basis of this informal pre-test, certain words which we had assumed would be acceptable were eliminated, such as *consumable* and *personable*. From the remaining list of *–able* words, we selected ten with verb bases and ten with noun bases that had a similar frequency of use, as measured from a word frequency list of the British National Corpus[2]. The verb words had an average frequency of 127 per 100 million (standard deviation 153), while the noun words had a frequency of 118 (s.d. 158). The range in the first case was 0 to 380, in the second 0 to 337. Note that the two examples of words with zero frequency (*deniable* and *exceptionable*) were accepted as familiar forms in our pilot, and both had *un-* derivations with attested frequencies in the BNC (*undeniable*: 208, *unexceptionable*: 30). The nouns were slightly longer than the verbs (11.2 vs. 11.0 characters).

Nonce words: In this case, we constructed new forms by adding *–able* to bases that could only be either verbs or nouns, making any expected modifications to the spelling of the base (e.g. *electrifiable* from *electrify*). None of the forms thus created were attested in an extensive dictionary search. The average frequency of the bases (including inflected forms) was 542 per 100 million for the verbs and 621 for the nouns (s.d.: 832 and 892), with ranges of 18-2676 and 40-2830 respectively. It should be noted that these frequency values are considerably lower than those obtained for the bases of the real words (6275 for verb bases, 7573 for noun bases). A lower average frequency for the verb bases of the nonce forms compared to the bases of the existing forms is unavoidable, since the more frequently used verbs tend to have attested *–able* forms. The important point is that the verb and noun bases within each condition have comparable frequency distributions. The average character length of the verb+*able* forms was 11.7, and that of the noun+*able* forms was 11.5.

Nonwords: The nonword tokens (i.e. nonsense bases with *–able*) were based on the nonce words. In each case, two characters in the base of the nonce words were changed to produce a nonword (e.g. *entrulchable* from *entrenchable*) . The characters in question were in various positions in the

[2] The word frequency list we used contains only words that occur more than 5 times in the BNC. Note that many of the tokens selected fail to appear in the smaller Brown corpus.

word, the only constraints being that the resulting nonword should be a possible word of English in terms of its spelling and pronunciation, and that the base should keep its verb-like or noun-like character, generally by preserving morphological form that is characteristic of these categories. Of course the character lengths of the nonwords were the same as those of the nonce words, and the frequencies of the full forms and of the bases were zero.

Design

A single test list was constructed using the 60 test items and 50 fillers. 30 fillers were real words, and 20 were nonsense words. 12 real word fillers and 12 nonword fillers ended in –tion, their purpose being to provide another instance of a single ending that occurred frequently in the test list. 5 real word fillers were words ending in non-morphemic –able (e.g. monosyllable). The remaining real word and nonword fillers were a mixed bag, but all were (or could be supposed to be) morphologically complex. In addition, 15 practice items were used, representing the conditions of the main test set. The complete list of test words is given in the Appendix.

Procedure

Subjects were seated at Macintosh computers running PsyScope software (Cohen, MacWhinney, Flatt, & Provost,1993). They were given written instructions on the procedure but not the purpose of the experiment and had an opportunity to ask questions of the experimenter. The instruction set included examples of what was meant by real words and nonwords. They were then exposed to the first 10 practice items, after which accuracy scores and average response time for real and nonsense words were displayed on the computer screen. After a further opportunity to seek clarification from the experimenter, they then saw the remaining practice words and the main set of test and filler items. The order of items in the main set was individually randomised for each subject, but with the constraint that items were selected evenly from the six test sets, so that there was no bunching of items from similar conditions.

Each trial presentation was preceded by a fixation point (*) displayed for 1 second in the center of a text box in the middle of the screen. This was replaced by the test item in lowercase Geneva 18-point font. This test item remained on screen for a maximum of 600 milliseconds. Subjects had to press one of two keys on the computer keyboard to signal their decision that what they had seen was a real word (the '/' key) or a nonword (the 'z' key). Their response cleared the display (if the token was still on screen). A maximum of 2.5 seconds was allowed for the response to be made, and there was a delay of 500 milliseconds before the next presentation sequence began. The complete test session lasted some 20 minutes.

Subjects

Subjects were undergraduate students taking part in a practical demonstration as part of a course in psycholinguistics. They were naïve to the exact nature of the experimental materials, but had received some tuition on lexical access and morphology. Non-native speakers of English taking the course also took part in the demonstration class, but their data were excluded

from the analysis. In all, 20 subjects were included (18 female, 2 male), all of whom were tested together in a computer laboratory in a single session.

Results

The data gathered in this experiment includes both response choice and time taken to make that choice. The main focus in our presentation of the results, reflecting our predictions earlier, is on the analysis of response choices. As will become apparent, a complete comparison of response times across conditions is in any case not possible, because the majority response ("yes" or "no") differs across conditions, and so factors other than the nature of the stimuli (which hand is used, whether it is a positive or negative response, etc.) become involved in the time taken to make the response.

Response choice Of the 1200 possible responses to test items (20 subjects x 60 test items), 5 (0.4%) were missing (i.e. responses were timed out or not registered by the computer). The remaining responses were distributed as shown in Table 1. This table shows quite clearly that the existing words were reliably identified as real words (with 96% accuracy overall) and that the nonwords were identified as nonwords (92% accuracy), in line with our first prediction on page 6.

	Verb(-like) base		Noun(-like) base	
	Real word	Nonword	Real word	Nonword
Real words	192	5	188	12
Nonce words	141	58	57	143
Nonwords	25	174	6	194

Table 1. Response distribution by test type

Our second prediction was that nonsense forms with nonword bases would be more consistently identified as nonwords than the nonce forms with real word bases. Overall, 92% of the forms with nonword bases were identified as nonwords, but just 50% of the nonce forms with real word bases. This difference is highly significant in chi-square (χ^2=168.8, df=1, p<0.001).

For the nonce forms, we predicted that if constraints on morphological productivity influence subjects' preparedness to accept the nonce forms as real words, then there should be a difference between verb+*able* and noun+*able* forms, with the former more likely to be accepted as real words. This is confirmed in the data, with 71% real word decisions for the verb-base nonce forms, but just 28% for the noun-base forms (χ^2=69.9, df=1, p<0.001). Closer scrutiny of the items involved reveals that the real word responses for noun+*able* items are largely given to the three tokens *vacationable, statuable* and *omissionable*. If these are excluded, the nonword response rate for these noun-based nonce forms rises to 88%. The results for verb-based nonce forms were more homogeneous, except that the two forms *enlivenable* and *embodiable* received somewhat more nonword responses than the other items. Excluding these two verb-based items results in 77% real word responses overall. There was a certain amount of inter-subject variability in responses, with real word

responses to the nonce words ranging from 15% to 75%. Without exception, however, subjects gave more real word responses to verb-based nonce forms than to noun-based ones.

Our final prediction was that if the nonword base in the two nonword conditions are identified as verb-like and noun-like, then there should be more realword responses to the subset with verb-like bases. The results show that although the overall error rate for the nonword condition is low (cf. the "real word" choices in the last row of Table 1), four times as many real word responses were given to the nonwords with verb-like bases as to those with noun-like bases (χ^2=11.4, df=1, p<0.001), supporting our further prediction that the identification of the base with one syntactic category rather than another will influence the likelihood of accepting the combination of the base with the *-able* suffix as a potential real word.

A further observation from the data is that the verb-based real words were somewhat more reliably identified as real words than the noun-based words, though this difference was not significant (χ^2=2.12, df=1, p > 0.10).

Response times On the basis of a distributional analysis of the data, 17 data points (1.4% of the total) were excluded from the response time analysis as extreme values (greater than 1600 milliseconds). The exclusions affected all conditions, with rather more nonce forms eliminated (12, of which 9 had noun bases) than real words (1) or nonwords (4).

Mean response times across items and across subjects were computed for the majority response category for each of the experimental conditions. That is, we calculated average real word responses to real words and to nonce words with verb bases and average nonword responses to nonwords and to nonce words with noun bases. The alternative responses could not be used in this analysis by both items and subjects, either because they were too infrequent or, in the case of the somewhat larger set of such response to nonce forms, because they were too unevenly distributed across items or subjects. Figure 1 displays the mean responses, with standard error bars, for each condition. Note that the majority decision for the three data sets on the left is a real word response, while for those on the right it is a nonword response.

The data in the first and last pairs of columns in Figure 1 show quite clearly that, as indicated as part of the first prediction on page 6, positive decisions to real words were faster than negative decisions to nonwords. This unsurprising result is confirmed in a planned comparison of these four conditions in Analysis of Variance (ANOVA), with stimulus Type (real word, nonword) and Base (verb, noun) as factors. Separate ANOVAs were run on means calculated over items (F1) and over subjects (F2; in this subject analysis the two factors constituted within-subject variables). In the item analysis there was a significant main effect of Type (F1[2,54]=43.61, p<0.0001). The subject analysis similarly gave a significant main effect of Type (F2[1,19]=36.22, p<0.0001), but also a significant interaction of Type with Base (F2[1,19]=5.09, p<0.04). This interaction reflects the fact that the real words with verb bases were accepted as real words more rapidly than those with noun bases, while the nonsense words with verb-like bases took longer to reject as not being real words than those with noun-like bases (though neither of these individual comparisons is significant in t-tests).

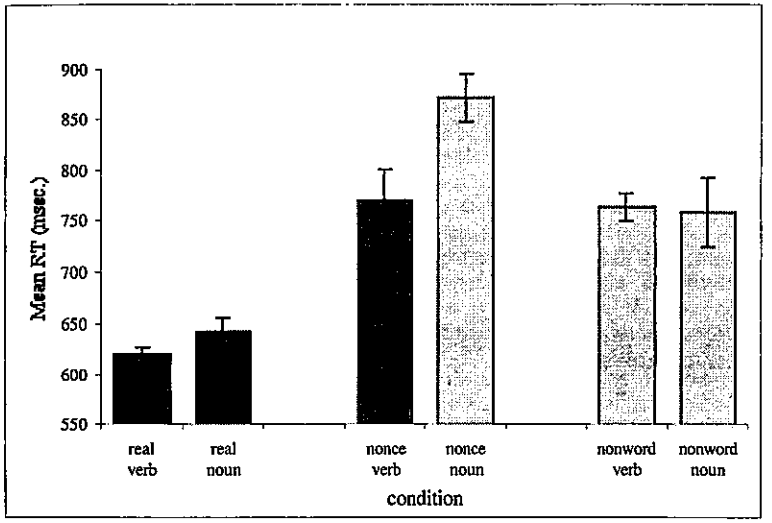


Figure 1. Mean response times for majority decisions (milliseconds). Note that responses for the three conditions on the left are "yes" (real word) responses, while those for the three conditions on the right are "no" (nonword) responses (see text).

Since the nonce form data in Figure 1 confound lexical status and base form, a comparison was made of the real word and nonword responses to the nonce forms, for both verb- and noun-based items. Note that since the nonword responses to verb-*able* forms and the real word responses to noun-*able* forms are somewhat erratically distributed across items and subjects, it is not possible to carry out separate item and subject analyses for these data. Instead, the complete set of nonce word data was entered into a single ANOVA, with response category (real word, nonword) and base type (verb, noun) as factors. The mean data from this analysis are shown in Table 2. Only the main effect of response category was significant (F[1,383]=10.76, p<0.01), with nonword responses overall some 100 milliseconds slower than real word responses. The difference in response times for the majority decisions for verb- and noun-based items in this category (Figure 1) is therefore a difference in their treatment as real words and nonwords respectively.

	Verb base	Noun base
Real word responses	765 (*N:138*)	775 (*N:55*)
Nonword responses	873 (*N:58*)	864 (*N:136*)

Table 2. Mean response times for nonce forms (milliseconds)

Further comparisons are possible between conditions that receive the same type of majority response. These show that the average real word response time to nonce forms with verb bases is significantly longer than the equivalent average for real word forms with verb bases ($t=7.34$, $df:18$, $p<0.0001$), and the average nonword response time to nonce forms with noun bases is longer than the average for nonword forms with noun-like bases (marginally significant at $t=2.03$, $df:18$, $p<0.06$). Combined with the response choice data, these results confirm that the nonce forms with verb bases are generally considered to be possible real words, but are not as familiar to our participants as the existing real word forms, while the nonce forms with noun bases are not considered to be real words, but take longer to reject than the forms with nonsense bases, presumably because the real word status of their bases means that they are only recognised as nonwords once the affix is encountered.

Summary

The results presented above support the predictions we set out earlier. First, the basic lexicality effect was observed. Existing real words were identified as words and the nonwords with nonsense bases were consistently identified as nonwords, with the real word identification more rapid than the nonword identification.

Our second prediction was that the nonwords, with their nonsense bases, would be identified more consistently as nonwords than the nonce forms with real word bases. The pattern of response choice in Table 1 clearly confirms the greater consistency of nonword identification for the nonwords. In addition, the comparison of response times for nonce and nonsense forms with noun(-like) bases shows faster nonword decisions for the forms with nonsense bases.

Our main prediction concerning productivity was that subjects would be sensitive to the greater productivity of *–able* with verb bases than with noun bases, and so would more readily accept the verb-based nonce forms as real words. This is clearly confirmed by the response choices, with 71% of verb+*able* nonce forms identified as real words but only 28% of the noun+*able* forms. As the data in Table 2 show, when a noun-based form is accepted as a real word, this acceptance is no slower than for a verb-based form, i.e. the only response time effect for the nonce words is that real word responses were faster than nonword responses.

We also predicted that recognition of the verb-like or noun-like nature of the nonsense base of the nonwords would affect responses. There is support for this in the significantly greater number of real word responses to the nonwords with verb-like bases, and the slightly (but non-significantly) faster nonword response times for the forms with noun-like bases.

General discussion

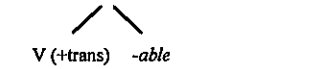
Our experimental data show a striking effect of features of the base form on the acceptability of a derived nonce word ending in the affix *–able*. We claim that this finding is best explained if we assume that native speakers not only know (implicitly) which morphological processes are productive, but also the constraints under which they are productive. Our additional claim is that our

results are best understood in a framework that allows morphological decomposition in lexical access. That is, subjects in our experiment identify the test items as having an *–able* affix, and attempt to access these items as a base and an *–able* affix.

Considering first the access of the base, it might be tempting to argue that the faster nonword response times to nonwords than to nonce words shows that morphological decomposition has made a real word base form available in the latter case only. However, our result is not direct evidence for morphological decomposition and subsequent recognition of the lexical status of the base, since any model of left-to-right analysis of the forms displayed on screen in the experiment would predict earlier detection of nonword status for the nonword items, because this nonword status becomes clear within the base. One test of this prediction would be to look for a correlation of nonword decision times with nonword point (defined as the number of letters into the string at which it ceases to be a real word). Since however, the nonword point varied only from 3 to 6 letters across the items, it is unsurprising that such a test for correlation proved inconclusive.

Access of the *–able* affix makes available featural information about the affix, including its morphological constraints, i.e. that it is productive only with verb bases. Segui and Zubizarreta (1985: 761) propose the affix frame in (1), by which *–able* combines with a transitive verb, creating an adjective with a passive interpretation (i.e. the object of the verb serves as subject when the adjective is in the predicate: *she loves the teddy → the teddy is lovable*). This frame is an adequate illustration for our purposes, though we will return shortly to a discussion of Segui and Zubizarreta's use of it.

(1)



We claim that in the process of constructing a representation, subjects match features of the base with the specification in the affix frame. This frame allows new verb+*able* combinations, which are consequently accepted as real words, but not new noun+*able* ones, which are not.

The fact that nonce verb+*able* forms take significantly longer than existing words to accept as real words suggests differences in the access and representation of these forms. We contend that the process of learning or the lexicalisation of real words in an individual's mental lexicon establishes links between bases and affixes, in a manner that is effectively no different from the "hard-wiring" of other collocations. These links allow more rapid recognition of the existing real word combinations of base and affix.

Of course, it is a short step from arguing that certain connections between affixes and bases are stronger than others to a claim that the existing real word forms that have these stronger connections are in fact fully listed. For

instance, we might assume that existing morphologically complex words are represented as whole forms, and that bases and affixes are also separately listed. The input is initially matched against whole word representations. If one is found, the subject registers a real word response. If not, the internal structure of the input is analysed, and the affix identified. The base and affix combination is then checked against the productivity constraints on the affix, and a lexical decision made. Some further factor would have to account for the additional time taken to complete the nonword response compared to the real word responses for nonce forms (Table 2). This might involve differences in using right and left hands for the two responses, or some extra component involved in making a "no" decision. Similar accounts, but with parallel searches involving both full-listing and decompositional approaches, are offered under hybrid models such as Frauenfelder and Schreuder's (1992) Morphological Race Model.

While our data do not allow us to distinguish between an account in which base-affix connection strengths reflect lexicalisation processes and an account with both full-listing and the listing of affixes, they do nevertheless argue for the availability of morphological information in lexical access, at least for our nonce words. And if it is available for these nonce words, then there is no reason to suppose that it is not generally available. Note though that we make quite different assumptions about the use of information such as that encapsulated in the frame in (1) from those made by Segui and Zubizaretta (1985). These authors claim that such frames are useful in the acquisition of the lexicon, rather than in post-acquisition processing. When an affix is learned, every item in the lexicon is scanned, and those that match the specifications (i.e. transitive verbs for the *–able* frame in (1) above) are inserted into a frame with the affix. This happens automatically at some point in the acquisition process. It results in a new set of lexical entries being generated automatically and without the need to encounter all specific examples in the linguistic environment (Segui & Zubizarreta 1985: 762). However, we suggest that this process is not simply part of vocabulary learning, but is an active element of morphological productivity. The compelling argument is that new bases that meet the relevant conditions will be open to this affixation, and not just bases that the learner happened to already know at the point in time when the affix and its specifications were learned.

A possible alternative to our decompositional account would be an analogical, spreading activation account. Let us assume first that some of the existing words in the mental lexicon receive activation on the basis of formal overlap with a nonce form presented in the experiment. One of these words will of course be the base used in constructing the nonce form, but in addition there will be a set of activated words containing the same affix as the nonce form, since these too share orthographic or segmental material with the nonce form. Let us further assume that these affixed words pass activation to other words from their morphological family, including base forms. In addition, there are activation links between words in the lexicon which reflect the properties that distinguish the productive and non-productive bases. Finally, we have to assume that there are stronger links within the set of productive bases (e.g. via a common link to a node for 'transitive verb' at a further level of representation within the network), or at least that this set is more heavily populated than the non-productive set(s). If a lexical decision about a nonce form is made on the basis of the level of activity in the system, then both types

of nonce word will cause activation on the basis of overlap between the affix and existing words, but a nonce form with a productive base will result in more activation overall because of the assumed greater activation within the set of productive bases. This will result in more real word responses for the form with the productive base.

There are a number of grounds on which we believe such a full-listing analogical explanation is less plausible. One is that it is based on a series of assumptions about the relative patterns of activation among productive and non-productive base sets and their influence on the decision process. These assumptions need to be carefully examined before the model as a whole could be accepted. More importantly, the explanation fails to distinguish between genuinely productive morphological processes such as verb+*able* and processes that are no longer productive but which have in the past thrown up a large set of *base* and *base+affix* combinations, such that current productivity would not be well predicted by analogy. An example is the use of *–ment* to produce nominalisations from transitive verbs with the approximate meaning 'state of being Ved', such as *embarrassment*. Lehnert (1971) lists several hundred words ending in this morpheme. Yet there is considerable evidence that *–ment* is no longer productive in present-day English, or is at best only marginally so (Bauer 1983:55, Baayen & Lieber 1991:830). On an analogical account, we might expect nonce forms such as *memorisement* and *electrifyment* to be just as acceptable as *memorisable* and *electrifiable*, yet intuition suggests this will not be the case.

Clearly, the precise mechanism of the processing of morphologically complex forms remains undetermined. However, our experimental data strongly support a role for productivity and for constraints on productivity in this process. They also support our claim that productive morphological processes are the most likely to be available for decomposition.

In conclusion, we have argued in this paper that morphological analysis during word recognition shows sensitivity not just to which morphological processes are productive, but also to the constraints under which they are productive. Our lexical decision results for tokens ending in the English affix -*able* demonstrate that comprehenders are sensitive to the fact that this affix is productive in combination with verb bases but not with noun bases. We argue that our results support models of word recognition that include a process of morphological decomposition and in which information about base-affix combinations must also be included, such as in the form of affix-frames such as those suggested for the learning process by Segui and Zubizarreta (1985). We note that recent papers such as Frauenfelder & Schreuder (1992), Niemi, Laine, & Tuominen (1994) and Hagiwara, Sugioka, Ito, Kawamura, & Shiota (1999) also focus on the importance of productivity for (de)composition. Our paper adds further support to this general view.

References

Aronoff, Mark 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baayen, R. Harald & Rochelle Lieber 1991. Productivity and English derivation: a corpus-based study. *Linguistics, 29*, 801-844.

Baayen, R. Harald, Rochelle Lieber & Robert Schreuder 1997. The morphological complexity of simplex nouns. *Linguistics, 35*, 861-877.

Badecker, William, & Alfonso Caramazza 1998. Morphology and aphasia. In Andrew Spencer & Arnold M. Zwicky (eds), *Handbook of Morphology*. Oxford and Malden, MA: Blackwell, 390-405.

Bauer, Laurie 1983. *English Word-formation*. Cambridge: Cambridge University Press.

Bauer, Laurie 1988. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.

Bauer, Laurie 1994. Productivity. In R.E. Asher (ed), *The Encyclopedia of Language and Linguistics*, Vol 6. Oxford: Pergamon, 3354-3357.

Bauer, Laurie 1999a. Is the morpheme dead? *Acta Linguistica Hafniensia, 31*, 7-25.

Bauer, Laurie 1999b. The fat owl of the remove meets the ness peril. *English Today, 15*, 39-42.

Beard, Robert 1982. The plural as lexical derivation. *Glossa, 16*, 133-148.

Berko Gleason, Jean 1958. The child's learning of English morphology. *Word, 14*, 150-177

Bybee, Joan, & Carol Lynn Moder 1983. Morphological classes as natural categories. *Language, 59*, 251-270

Cohen Jonathan D., Brian MacWhinney, Matthew Flatt, & Jefferson Provost 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Derwing, Bruce 1973. *Transformational Grammar as a Theory of Language Acquisition*. Cambridge: Cambridge University Press.

Dressler, Wolfgang U. 1989. Prototypical differences between inflection and derivation. *Zeitschrift für Phonetix, Sprachwissenschaft und Kommunikationsforschungen, 42*, 3-10.

Fowler, Carol A., Shirley E. Napps & Laurie Feldman 1985. Relations among regular and irregular morphologically related words as revealed by repetition priming. *Memory and Cognition, 13*, 241-255.

Frauenfelder, Uli H. & Robert Schreuder 1992. Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In Gert Booij and J. van Marle (eds), *Yearbook of Morphology 1991*. Dordrecht: Kluwer Academic, 165-183.

Hagiwara, H., Y. Sugioka, M. Kawamura & J. Shiota 1999. Neurolinguistic evidence for rule-based nominal suffixation. *Language, 75*, 739-763.

Haspelmath, Martin 1996. Word-class-changing inflection and morphological theory. In Gert Booij and J. van Marle (eds), *Yearbook of Morphology 1995*. Dordrecht: Kluwer Academic, 43-66.

Indefrey, Peter 2000. Neural correlates of regular and irregular inflection. Paper presented at the *9th International Morphology Meeting*, University of Vienna, Feb 2000.

Jackendoff, Ray 1975. Morphological and semantic regularities in the lexicon *Language, 57*, 639-671.

Jaeger, Jeri J., A.H. Lockwood, D.L. Kemmerer, R.D. van Valin. Jr, B.W. Murphy & H.G. Khalak 1996. A positron emission tomographic study of regular and irregular verb morphology in English. *Language, 72*, 451-497.

Kibrik, Aleksandr E. 1998. Archi. In Andrew Spencer & A.M. Zwicky (eds), *Handbook of Morphology*. Oxford and Malden, MA: Blackwell, 455-476.

Lehnert, Martin 1971. *Rückläufiges Wörterbuch der englischen Gegenwartssprache / Reverse Dictionary of Present-Day English*. Leipzig: VEB.

Marslen-Wilson, William, Lorraine Komisarjevsky Tyler, Rachelle Waksler & Lianne Older 1994. Morphology and meaning in the English mental lexicon. *Psychological Review, 101*, 3-33.

Mayerthaler, Willi 1981. *Morphologische Natürlichkeit*. Wiesbaden: Athenaion.

Mugdan, Joachim 1986. Was ist eigentlich ein Morphem? *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschungen, 39*, 29-43.

Niemi, Jussi, Matti Laine & Juhani Tuominen 1994. Cognitive morphology in Finnish: foundations of a new model. *Language and Cognitive Processes, 9*, 423-446.

Plank, Frans 1994. Inflection and derivation. In R.E. Asher (ed), *The Encyclopedia of Language and Linguistic*, Vol 3. Oxford: Pergamon, 1671-1678.

Sandra, Dominiek 1994. The morphology of the mental lexicon: internal word-structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes, 9*, 227-269.

Segui, Juan & M-L Zubizaretta 1985. Mental representation of morphologically complex words and lexical access. *Linguistics, 23*, 759-774.

Seidenberg, Mark S., & J.H. Hoeffner. 1998. Evaluating behavioural and neuroimaging data on past-tense processing. *Language, 74*, 104-122.

Slowiaczek, Louisa M., & David B. Pisoni 1986. Effects of phonological similarity on priming in auditory lexical decision. *Memory and Cognition, 14*, 230-237.

Stanners, R. F., J.T. Neiser, W.P. Hernon & R. Hall 1979. Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18*, 399-412.

Stemberger, Joseph, & Brian MacWhinney 1986. Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition, 14*, 17-26.

APPENDIX

Test words used in the experiment

Real words		Nonce words		Nonwords	
verb-base	noun-base	verb-base	noun-base	verb-like base	noun-like base
conceivable	impressionable	scandalisable	omissionable	scoldalisable	arissionable
recognisable	exceptionable	beautifiable	sewerageable	begatifiable	leberageable
forgivable	sizeable	electrifiable	reportageable	omectrifiable	remartageable
collectable	marriageable	strengthenable	corridorable	streanthenable	purridorable
maintainable	fashionable	bewilderable	squalorable	bepalderable	sphalorable
classifiable	knowledgeable	entrenchable	vacationable	entrulchable	nalationable
interpretable	companionable	enlivenable	statuable	envitenable	smetuable
detectable	peaceable	memorisable	chaliceable	mamurisable	sheliceable
preventable	seasonable	embodiable	solsticeable	empediable	sinsticeable
deniable	palatable	enactable	juncturable	enoftable	durcturable

Phonological awareness in young Cantonese-Mandarin bilinguals[1]

Guiling Hu

Abstract

The present research studied the phonological awareness in Mandarin of 31 Cantonese-Mandarin preliterate simultaneous bilingual children and 30 Mandarin preliterate monolingual children. The experiment tested the children's tone awareness, Mandarin initial awareness, Mandarin final awareness, and Mandarin rime awareness. Analyses of variance revealed that the Cantonese-Mandarin bilingual children were found to have significantly better awareness than the Mandarin monolingual children only in Mandarin tones and Mandarin initials. In awareness of Mandarin finals, the monolingual children displayed a marked advantage over the bilingual children; and no significant difference was observed between the two groups in awareness of Mandarin rimes. The results of initial, final and rime awareness are explained within an attention-based framework proposed by the present researcher, which presumes that in daily language use bilinguals and monolinguals learn to pay different amounts of attention to different phonological units depending on the saliency of the unit.

Introduction

Studies of the cognitive effects of bilingualism have revealed that knowing two languages promotes children's phonological awareness of the languages. However, most of the languages involved in those studies are non-tonal languages (e.g. Bruck and Genesee 1995; Campbell & Sais 1995; Eviatar and Ibrahim 2000). Tonal languages have largely been ignored. To test phonological awareness among tonal language bilinguals, the present research investigated segmental awareness and tone awareness in Mandarin in pre-literate Cantonese-Mandarin bilingual children. Both Cantonese and Mandarin are tonal languages.

Phonological awareness, an individual's ability to manipulate the component sounds of the sound sequences of a language, has been intensively studied in literature dealing with language acquisition among monolingual children. However, it has received less attention in studies of bilingualism.

In research on literacy acquisition, phonological awareness is found to be one of the best early predictors of children's performance in learning to read alphabetic languages (e.g. Juel, Griffith & Gough 1986; see Adams 1990, and

[1] This article is based on my MA thesis submitted in 2002 to Victoria University of Wellington. I am grateful to my MA supervisor, Dr. Paul Warren, for guidance and help throughout the whole process of my thesis writing. He also gave me invaluable comments on the draft of this article. I am also grateful to the children for their participation in this study and to the children's parents, teachers, and principals for their gracious support. This study was partially funded by a research grant (Grant No: 1/3009/62795/ R 01 168) from the Faculty Research Committee of the Faculty of Humanities and Social Sciences in the Victoria University of Wellington.

Goswami & Bryant 1990 for reviews). To the extent that alphabetic letters represent phonemes in alphabetic languages, the ability to perceive and manipulate individual phonemes and phoneme sequences is helpful in learning to read those languages.

In a 4-year longitudinal study, Bradley and Bryant (1983) found high correlations between children's scores in phonological awareness tests and their scores in standard reading and spelling tests. In a second part of their study, these researchers ran a two-year phonological awareness training programme for the children who did not perform well in the phonological awareness tests. The children were divided into four groups with different training regimes. The results showed that there was some progress in reading and spelling if the children were taught to sort pictures on the basis of shared sounds, but even more progress if they were also explicitly taught letter-sound correspondences. Bradley and Bryant's research revealed that phonological awareness is important for reading acquisition in English.

In non-alphabetic languages such as Chinese, phonological awareness has also been shown to make a significant contribution to successful reading acquisition (Ho 1997; Ho & Bryant 1997a; Ho & Bryant 1997b; Hu & Catts 1998; Li et al 2002; McBride-Chang & Ho 2000). This might be due to the fact that most Chinese characters are composed of two parts: a radical, which indicates the character's meaning, and a phonetic, which provides clues as to how the character should be pronounced. Very often, the sound sequence of the phonetic needs to be broken down into smaller components in order to obtain the pronunciation of the character (Wang 1994: 135-144). Knowing how to segment a sound sequence, therefore, would lead to facilitated reading in Chinese.

Ho and Bryant's (1997a) study tested the effects of phonological awareness on Chinese children's reading acquisition. The children's phonological awareness was measured by tasks testing children's detection of rhymes and tones. The children's scores in the two phonological awareness tasks were found to be correlated significantly with their later performance in character reading and pseudocharacter reading tasks.

Li et al's (2002) recent study also confirmed the effects of phonological awareness on Chinese children's reading acquisition. The children in the experiments were first and fourth graders (about 7 and 11 years old respectively) in a primary school in China. The phonological awareness test consisted of syllable reversing, onset deletion, and tone discrimination tasks. The syllable reversing task required the children to reverse a three- or four-syllable word, e.g. reverse *qiao·ke·li* (chocolate) into *li·ke·qiao*. The onset deletion task tested the children's ability to remove the first consonant of a syllable. In the tone discrimination task, the children were given several syllables and were asked to decide whether the syllables had the same tone or not. The results showed significant correlations between phonological awareness and reading proficiency.

Most research has revealed that phonological awareness is a predictor of successful reading acquisition in virtually every orthography that has been studied, including various kinds of alphabetic and non-alphabetic languages (Goswami 1999: 134). Since phonological awareness plays such an important

role in children's early reading, we assume that knowing two languages at a young age might facilitate reading acquisition if pre-literate bilingual children are found to be more sensitive than their monolingual peers to the phonological units of language. Although still scant in quantity, recent studies have shown that English-French, English-Italian, and Russian-Hebrew pre-literate bilingual children have advantages over their monolingual counterparts in doing certain phonological awareness tasks.

Campbell & Sais (1995) investigated the phonological awareness of 15 English monolingual children and 15 Italian-English bilingual children with a mean age of 4.8 years. One of the tasks given to the children was an oddity task in which the odd word differed from the other two words in the initial sound. They found that bilingual children were more successful in performing this task.

Bruck and Genesee (1995) administered a series of phonological awareness tasks to 72 English monolingual kindergarten children and 91 English-French bilingual kindergarten children. The bilingual children outperformed their monolingual peers in onset deletion, cluster onset detection, single onset detection and rime detection tasks. In the onset deletion task, a CVC non-word, e.g. [væp], was given to the child and the child was told to repeat the word without the first sound e.g. [æp]. In the cluster onset detection task, two CCVC non-words, e.g. [plam] and [plæk], were presented to the child and the child was required to say whether the two non-words shared the same two beginning sounds. The single onset detection task and the rime detection task were similar to the cluster onset detection task except that the two non-words had a CVC structure, e.g. [foik] and [feiv], and a CVCC structure, e.g. [pisk] and [nisk], respectively.

Eviatar and Ibrahim (2000) compared the metalinguistic awareness of 36 Russian-Hebrew bilingual kindergarteners and 40 Literary-Spoken Arabic bilingual kindergarteners with 40 Hebrew monolingual kindergarteners. (Literary Arabic is the basis of the writing system used in the Arab world. Spoken Arabic is the local dialect. The two forms of Arabic are phonologically, morphologically, and syntactically different.) The two groups of bilinguals outperformed the monolinguals in initial phoneme detection, final phoneme detection, and phoneme/syllable deletion tasks.

It is worth noting that the languages involved in previous studies of the relationship of bilingualism to phonological awareness are non-tonal languages. In contrast with non-tonal languages, tonal languages distinguish words not only on the basis of phonological segments but also on the basis of tones. Tones, the relative pitch levels or contours with which words are realised, play an important lexical role in the phonology of a tonal language, and so tone awareness is a potentially important component of phonological awareness in such a language. Tones occur simultaneously with vowel and consonant segments and can extend over multiple segments. They are therefore different in nature from the vowel and consonant segments, and so tone awareness may be quite different from other types of phonological awareness. In addition, the existence of tones in a tonal language might also influence the tonal language bilingual's awareness of segments. Research on the relationship of bilingualism and phonological awareness will therefore be

incomplete if it does not include the study of tone and segment awareness in tonal languages.

The aim of the present study was to further research on the relationship between bilingualism and phonological awareness by answering the following two research questions:

A. Will tonal language bilinguals show the same advantage in segmental awareness as previously found in studies with non-tonal language bilinguals?

B. Does the existence of two systems of tones promote tone awareness in children who are bilingual in tone languages?

It was hypothesized that the bilingual children would have across-the-board advantages over their monolingual peers in tone awareness as well as in segmental awareness.

Method

Subjects

The bilingual subjects were recruited from two different classes in a kindergarten in Guangzhou (Canton), China. Both Cantonese and Mandarin are used in the media and in most work places in this city. The children's age ranged from 5 to 6 years old. Thirty-five children from the two classes participated in the experiment. The language backgrounds of the children were obtained through interviews with the children's teachers and through questionnaires completed by the children's parents. Among the 35 children, one child spoke the Min dialect of Mandarin, one had newly moved to Guangzhou and spoke no Cantonese, and two could not speak Mandarin well. Since these four children were not bilingual in Cantonese and Mandarin, their responses were not used in the analysis below. The remaining 31 children were all simultaneous childhood bilinguals, who either came from a Cantonese-Mandarin bilingual family or had begun to acquire Mandarin or Cantonese at age two. They were all normal, healthy children.

The monolingual subjects were recruited from one class in a kindergarten in Chengde, China, a city close to Beijing. Their age ranged from 5 to 6 years old. Thirty children participated in the experiment. The language backgrounds of the children were obtained through interviews with the children's teachers and through questionnaires completed by the children's parents. All of the 30 children were monolingual Mandarin native speakers. They were all normal, healthy children.

Both groups of children had Chinese and English classes, which were two required courses in the kindergarten curriculum. In their English class, they were taught simple English words. They were not taught how to read and write in English. In their Chinese class, the children were taught some classic ancient poems and nursery rhymes. The children in Chengde were also taught a little *Pinyin* to pronounce a few simple Chinese characters. This was required by the city's kindergarten curriculum. However, by the time when the experiment was conducted, the children had had classes for *Pinyin* for only two months. I assumed that their experience with *Pinyin* should not affect the

results of the experiment to a significantly great extent. The children's knowledge of Chinese characters was limited. None of them could write any characters other than their own names. All of the children's parents were city residents whose social standing matches that of middle-class people in Western countries.

Materials

Four phonological awareness tests were administered to the children. The tests measured the children's awareness of Mandarin tones, initials (i.e. the first consonant of a Mandarin syllable), finals (i.e. the VVC sound sequence in a Mandarin CVVC syllable, and the VC sound sequence in a Mandarin CVC syllable), and rimes (i.e. the VC sound sequence in a Mandarin CVVC syllable, and the second V sound in a Mandarin CVV syllable). Each child responded to ten sets of materials for each of these four types of awareness. Initial, final and rime awareness were chosen for the tests of segmental awareness because previous studies have shown that awareness of these three predicts aspects of Chinese children's literacy acquisition. All four types of awareness were tested using an ABX design. In this design, the children listen to a stimulus set which includes a target stimulus, X, and two further stimuli, A and B, and they are asked to indicate whether X sounds like A or B (see Appendix I).

The stimuli used in the present study were monosyllabic non-words that obeyed the phonotactic constraints of Mandarin and so were potential and pronounceable but non-existent words in Mandarin. The non-words were coined by the investigator using a Mandarin syllabary. The purpose of using non-words was to avoid any interference from lexical information. Ten stimulus sets were used for each type of phonological awareness. The stimuli were listened to and checked by two native speakers of Mandarin and two native speakers of Cantonese to ensure that the non-words were not real words in either of the two languages. Since differences between the four Mandarin tones could be a confounding factor in the "segmental" awareness tests (i.e. tests for initial, final and rime awareness), these tests used non-words with the same tone. Tone 1 was chosen as the tone for these materials because it is easier to perceive than other tones. Tone 2 and 3 are easily confused and Tone 4 is relatively brief and difficult to distinguish (Gandour, 1978, 1983).

The non-word stimuli were recorded directly into a computer using the Microsoft Windows audio recording function with a microphone in a quiet room. A native speaker of Mandarin served as the reader. The Praat phonetics software programme (Boersma & Weenink 2001) was used to arrange stimuli in sets. The time interval between two neighbouring stimuli within a stimulus set was 1.5s, and the interval between two neighbouring stimulus sets was 2s to allow time for the experimenter to pause the tape so that the children could choose their answers. The order of the stimulus sets and the order of the A and B stimuli inside each stimulus set were randomised.

Procedure

Written consents were obtained from the principals of the kindergartens and the children's parents before the experiment started. The author administered all the tests. The experiment involved four sessions conducted on four consecutive days, and was carried out in the children's regular classrooms. The

children were divided into two groups and received the tests separately during each session. The stimuli were played from a TEAC cdx I Type I cassette tape, using a Philips AQ5211MK2 player in Guangzhou and a Sanyo M7024 player in Chengde that were supplied by the kindergartens. The children reported that the recording was clear and good enough for performing the task. Two trial stimulus sets were given to the children at the beginning of the first day's test. The formal test did not start until every child understood the instructions. The children were each given a small gift after the experiment.

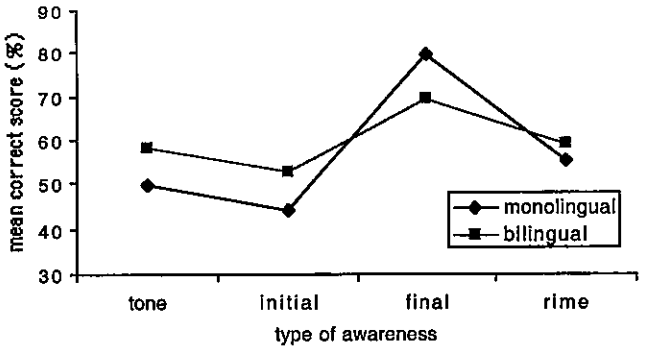
The children were told that they were going to play a game involving paper cut-outs of cartoon characters. One set of cartoon characters was used for each ABX stimulus set. Each set contained two identical cut-outs, and one bigger cut-out with the same picture as the other two (see Appendix II). The two identical cut-outs were shown to the children first, and then the bigger cut-out. The children were told that the bigger cut-out was an elder sibling (brother or sister) of one of the two identical cut-outs. They were asked to indicate which of these two identical cut-outs was the younger sibling of the bigger one. They were also told that the only way to find the younger sibling was to listen to the funny Mandarin names of the three characters, because the names of two siblings sounded similar. The recording of the three non-words in the set was played to the children, and the children were told which non-word was the name for which character.

Since the tasks were not administered to the children individually, the children had to write their responses on an answer sheet. To make it possible and interesting for the pre-literate children to do this, their answer sheets showed the three cartoon pictures of the cut-outs, reduced to a much smaller size. The children were asked to circle the cartoon character which they thought was the right answer.

Results

Each response was scored as correct or incorrect, and percentage correct response scores were obtained for each child for each type of awareness. The responses of 31 bilingual children and 30 monolingual children were used in the statistical analysis. The average scores across all participants are given in Figure 1. Each child's percentage score in each condition was subjected to an arcsine transformation ($2*arcsine\sqrt{p}$) before being entered as the dependent variable in a two-way ANOVA. The independent variables were *awareness type* (tone, initial, final and rime) and *group* (monolingual vs. bilingual).

Figure 1. Mean correct scores in ABX task



The results indicate that there is no main effect of group, $F(1, 52) < 1$, but a main effect of awareness type, $F(3, 52) = 15.363$, $p<0.001$, i.e., there are differences between the means of the four phonological awareness types. However, *awareness type* also interacts significantly with *group*, $F(3, 52) = 5.587$, $p=0.001$, and so the awareness type effect was further analysed in two separate one-way ANOVAs (with *awareness type* as the independent variable) for each group of children. The results reveal that there are significant differences among the four awareness types in both the bilingual group, $F(3, 26) = 6.042$, $p<0.001$, and the monolingual group, $F(3, 27) = 24.741$, $p<0.001$. Post-hoc analysis (Fisher's protected LSD) further indicate that both groups of children performed significantly better in the final awareness test than in the other three awareness tests, $p<0.01$. The difference between initial and rime awareness is also significant in the monolingual data, $p<0.05$.

Another way of looking at the *awareness type*group* interaction is to argue that differences between the bilingual and monolingual groups vary across the different awareness types. Four one-way ANOVAs were therefore carried out to investigate whether there were significant differences between the bilingual and monolingual groups in each awareness type. The results show that the bilingual children outperformed the monolingual children in tone, $F(1, 59) = 4.355$, $p<0.05$, and initial awareness, $F(1, 59) = 4.648$, $p<0.05$; that the monolinguals outperformed the bilinguals in final awareness, $F(1, 59) = 6.799$, $p<0.05$; and that there was no significant difference between the groups in rime awareness, $F(1, 59) < 1$.

In summary, the present study indicates that the bilingual children performed significantly better than the monolingual children in tone and initial awareness. The monolingual children performed significantly better than the bilingual children in final awareness. There was no significant difference between the two groups of children in rime awareness.

Dicsussion

The present study indicates that the bilingual children performed significantly better than their monolingual peers in tone awareness. This result is the first

documented demonstration of an advantage in tone awareness for bilingual children. The bilinguals' better awareness of tones may point to at least two benefits of tonal language bilingualism: one is that young tonal language bilinguals are better than their monolingual peers in understanding the lexical role of tones, and the other is that young tonal language bilinguals are better than their monolingual peers in the ability to separate tones from the segments on which the tones are realised.

The results of the present study reveal that the bilingual children also outperformed their monolingual peers in initial awareness. The bilingual children's greater awareness of initials, i.e. the first consonant of a potential Mandarin word, is consistent with previous studies conducted with non-tonal language bilinguals (Bruck & Genesee 1995; Campbell & Sais 1995; Eviatar & Ibrahim 2000). In these three studies, bilingual children outperformed their monolingual peers in tasks measuring the awareness of the initial phoneme of a CVC word or non-word.

The results also show that the monolingual children, contrary to prediction, outperformed the bilingual children in final awareness; and that there was no significant difference between the bilingual and the monolingual children in rime awareness. Why is it that the exposure to the phonological systems of Mandarin and Cantonese did not promote the bilingual children's awareness of Mandarin finals and rimes?

Since the monolinguals had learned some *Pinyin* prior to the time of this study, one could possibly argue that their better performance in final awareness was a reflection of that knowledge. However, such an argument would also predict a superior result from the monolingual children in the initial awareness test, since Mandarin initials and Mandarin finals are almost always treated equally in instruction of Mandarin *Pinyin* knowledge. The effect for initials was the opposite to this, with bilinguals significantly outperforming monolinguals. It is therefore unlikely that the monolinguals' better performance in final awareness is a result of their familiarity with *Pinyin*.

Consider again the bilinguals' superior performance in initial awareness. Since initials are phonemes in Mandarin and Cantonese, one might assume that Cantonese-Mandarin bilingualism promotes awareness of phonemes instead of awareness of units which are larger than phonemes such as rimes and finals. That is, when responding to the stimuli used in the final awareness test, the bilingual children may have segmented the non-words into phonemes and compared each phoneme in order to get the correct answer, which turned out to be a poor strategy for that task. The monolingual children, however, might have segmented the same non-word only into an initial and a final, which provided a better basis for comparison and resulted in a higher correct score in the final awareness test. This approach to interpreting the data, however, cannot explain why the bilingual children performed significantly better in final awareness than in initial awareness. If the bilinguals made use of their phoneme awareness advantage in initial and final awareness and segmented the stimuli into phonemes and then compared every phoneme, their performance in final awareness should be worse than that in initial awareness because they would need to compare just one phoneme in the initial awareness test, but two or more phonemes in the final awareness test. The results of the present study, however, proved exactly the opposite finding, i.e.

that the bilinguals performed significantly better in final awareness than in initial awareness.

Considering the three segmental awareness types (initial, final and rime) together, we can see that the scores for the two subject groups scores across these tests seem to follow the same pattern. Both groups of children performed significantly better in final awareness than in the other two tests of segmental awareness. One might argue that the children's performance pattern in the present study is a result of the difference in the number of overlapping phonemes in Stimulus A and Stimulus X (see Appendix I). Because Stimulus A and Stimulus X had more common phonemes in the final awareness test than in the initial awareness test, the children's attention might be more attracted by Stimulus A in the final awareness test; and, consequently, the children performed better in this test. This approach, however, is still unable to explain why the bilingual children performed better than the monolingual children in initial awareness, but worse in final awareness. If one overlapping phoneme in the initial awareness test caused the bilingual children to pay more attention to Stimulus A than the monolingual children, there is no reason why several overlapping phonemes in the final awareness test caused the bilingual children to pay less attention to Stimulus A than the monolingual children.

The better performance by both groups of children in the final awareness test than in the other two segmental awareness tests may be a result of the fact that the Mandarin final contains all the vowels of a word. It has been shown that native listeners' perception of a Mandarin word depends heavily on the vowels of the word. If the vowels are removed from the word with the initial remaining intact, native listeners find it hard to figure out what the word is. When the situation is reversed, i.e. initial removed and vowels intact, the task becomes very easy for the native listeners (Gottfried & Suiter 1997). Because vowels play such an important role in Mandarin word perception, and are also the major loci of the tones of Mandarin words, it would not be surprising to find that they are more salient than initials for young Mandarin-speaking children. Since in turn the Mandarin final contains all the vowels of a word, it is a highly salient unit in a Mandarin word, and therefore attracts most of young Mandarin-speaking children's attention. In other words, young Mandarin-speaking children pay most of their attention to the final when perceiving a Mandarin word. Since the rime in the present study is somewhat different from the final in that it is only a component of the final, the rime should become less salient than the final in Mandarin listeners' perception of a Mandarin word. However, because rimes also contain vowels (though not all of the vowels of a final), they should still be more salient than initials, which contain only consonants.

If different amounts of attention are devoted to these different segment types depending on their relative saliency for Mandarin-speaking children, then the phonological awareness results of the present study may have an explanation in terms of attention. Since the bilinguals and the monolinguals achieved the same scores in the rime awareness test, we might assume that both groups of children assign the same amount of attention to the rime unit. Why do young Mandarin-speaking bilingual and monolingual children pay the same amount of attention to rimes? This is probably related to the children's experience with nursery rhymes and poems. In order to perceive a rime in Mandarin, children need to segment between the two vowels of a final.

This segmentation is unnecessary except in the case of perceiving nursery rhymes and poems. Since preliterate Chinese children are taught relatively the same nursery rhymes and poems in kindergarten, the amounts of attention that both groups of children pay to Mandarin rimes are also relatively equal.

The bilinguals and the monolinguals in the present study demonstrated different degrees of awareness in Mandarin finals, with the monolinguals outperforming the bilinguals. How might an attention-based account explain this rather unexpected result? The answer may lie in the fact that Cantonese-Mandarin bilinguals need to pay more attention to individual phoneme-sized segments than larger segments such as finals and rimes in order to differentiate Cantonese and Mandarin words. Since monolingual Mandarin children do not need to differentiate two linguistic systems, they may focus much of their attention on the finals, which contain mostly vowels.

Bilinguals' awareness of subsyllabic units such as rime and final has seldom been investigated in earlier research. An exception is a study by Bruck and Genesee (1995). In their research, bilinguals outperformed monolinguals in a rime awareness task. This finding is not consistent with the results of the present study, in which the monolinguals and bilinguals did not show a significant difference in the rime awareness task. Since Bruck and Genesee (1995) is the only study of this kind involving non-tonal language bilinguals, it is difficult to conclude whether the discrepancy between these two studies is due to differences in language type, e.g. tonal vs. non-tonal, or in the amount of linguistic input, e.g. more vs. less nursery rhymes or poems, or due to other reasons, e.g. age of acquisition.

The attention-based account sketched above suggests that bilinguals pay more attention to phoneme-sized segments in their daily language use, and thereby develop greater phonological awareness of units at this level. This appears also to be supported in the research of Eviatar and Ibrahim (2000), who found that Russian-Hebrew bilinguals outperformed their Hebrew peers in all the phoneme awareness tasks. How, then, can we account for the results of Bruck and Genesee (1995), who tested phoneme awareness in English-French bilingual children and English monolingual children, and found no difference between the two groups of subjects? One crucial factor could be that the English-French bilinguals employed in Bruck and Genesee's study were consecutive bilinguals, while the bilinguals involved in Eviatar and Ibrahim's research and in the present study were simultaneous bilinguals. It is possible that consecutive bilinguals and simultaneous bilinguals develop different patterns of phonological awareness, possibly because of the differing requirements with respect to distinguishing between the two languages being learned. It should be borne in mind, however, that the languages, tasks and ways of administering the tasks employed in the present study are very different from those of previous studies. More research is needed to obtain a cross-linguistic pattern of bilingual children's performance in phoneme awareness.

In summary, the findings of the present study suggest several conclusions concerning phonological awareness among Cantonese-Mandarin bilingual children and Mandarin monolingual children. Firstly, bilingualism does not provide an across-the-board advantage in phonological awareness. Both bilinguals and monolinguals show advantages and disadvantages in different

aspects of phonological awareness. Studies are still too scant for us to be able to predict accurately what types of awareness will be observed for each group. Secondly, the present research suggests that both Cantonese-Mandarin bilingual children and Mandarin monolingual children have stronger awareness of Mandarin segmental elements involving vowels than of those involving only consonants because both groups of children performed better in awareness of rimes and finals, which mainly contain vowels, than initials, which only contain consonants. Thirdly, the attention-based account suggests that the Cantonese-Mandarin bilingual children might have a better awareness of Mandarin phonemes. Finally, since the bilingual children in the present study outperformed the monolingual children in tone awareness, this indicates that tonal language bilinguals have an advantage in suprasegmental awareness.

References

Adams, Marilyn Jager 1990. *Beginning to Read: Thinking and Learning about Print.* Cambridge: MIT Press.

Boersma, Paul & David Weenink 2001. *Praat 4.0 – A software package for doing phonetics by computer* (http://www.praat.org)

Bradley, Lynette & Peter Bryant 1983. Categorizing sounds and learning to read -- A causal connection. *Nature* 301: 419-421.

Bruck, Maggie & Fred Genesee 1995. Phonological awareness in young second language learners. *Journal of Child Language* 22: 307-324.

Campbell, Ruth & Efisia Sais 1995. Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology* 13: 61-68.

Eviatar, Zohar & Raphiq Ibrahim 2000. Bilingual is as bilingual does: Metalinguistic abilities of Arabic-speaking children. *Applied Psycholinguistics* 21: 451-471.

Gandour, Jack 1978. The perception of tone. In Victoria A. Fromkin (ed), *Tone: A Linguistic Survey.* New York: Academic Press, 41-76.

Gandour, Jack 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11: 149-175.

Goswami, Usha 1999. The relationship between phonological awareness and orthographic representation in different orthographies. In Margaret Harris & Giyoo Hatano (eds), *Learning to Read and Write: A Cross-Linguistic Perspective.* Cambridge: Cambridge University Press, 134-156.

Goswami, Usha & Peter Bryant 1990. *Phonological Skills and Learning to Read.* London: Erlbaum.

Gottfried, Terry & Traci L. Suiter 1997. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics* 25: 207-231.

Ho, Connie Suk-Han 1997. The importance of phonological awareness and verbal short-term memory to children's success in learning to read Chinese. *Psychologia* 40: 211-219.

Ho, Connie Suk-Han & Peter Bryant 1997a. Phonological skills are important in learning to read Chinese. *Developmental Psychology* 33: 946-951.

Ho, Connie Suk-Han & Peter Bryant 1997b. Learning to read Chinese beyond the logographic phase. *Reading Research Quarterly* 32: 276-289.

Hu, Chieh-Fang & Hugh W. Catts 1998. The role of phonological processing in early reading ability: What we can learn from Chinese. *Scientific Studies of Reading* 2: 55-79.

Juel, Connie, Priscilla L. Griffith, & Philip B. Gough 1986. Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology* 78: 243-255.

Li, Wenling, Richard C. Anderson, William Nagy, & Houcan Zhang 2002. Facets of metalinguistic awareness that contribute to Chinese literacy. In Wenling Li, Janet S.

Graffney & Jerome Lee Pachard (eds) *Chinese Children's Reading Acquisition: Theoretical and Pedagogical Issues*. Boston: Kluwer Academic Publishers.

McBride-Chang, Catherine & Connie Suk-Han Ho 2000. Developmental issues in Chinese children's character acquisition. *Journal of Educational Psychology* 92: 50-55.

Wang, Bingyu 1994. *Chinese Character: Its Correct Form and Pronunciation*. Beijing: Chinese Language Press.

Appendix I: Non-word Stimuli

Initial (All Tone 1)

A	B	X
bia	fia	be
hi	ki	hueng
fai	rai	fou
müan	füan	muan
lua	tua	lüan
nia	tia	nui
piu	tiu	püan
kü	mü	kiao
dia	mia	duai
gin	fin	güan

Final (All Tone 1)

A	B	X
büan	biang	tüan
gia	güan	mia
küe	kiu	güe
hiao	hüan	kiao
tiu	tuai	giu
fiong	füan	riong
gie	gü	fie
bua	buai	nua
füe	fao	büe
nuan	niong	buan

Rime (All Tone 1)

A	B	X
fai	fao	luai
tia	tüe	fua
biang	biong	muang
nua	nüan	nia
ria	rian	tua
buang	bü	fiang
fiou	fin	mou
piang	pong	muang
hiao	hiu	fao
muan	miang	ran

Tone

A	B	X
fiāng	fiáng	pīu
nuǎ	nuà	gǐn
giào	giāo	füè
lǔn	lǔn	tuāi
hiáo	hiāo	tuái
pìu	píu	luà
tǔn	tǔn	liōng
gié	giè	büé
buǎn	buān	miǎ
rài	rāi	sìu

Appendix II: Sample Stimulus Arrangement

1 2

The asymmetrical change in progress of NEAR and SQUARE vowels in NZE: psycholinguistic evidence

Megan Rae and Paul Warren

Abstract

This paper presents results from a recent psycholinguistic investigation of the merger of the NEAR and SQUARE diphthongs in New Zealand English (see also Rae and Warren 2002). The pattern of lexical decision response times in a semantic priming experiment is found to reflect the asymmetry in this merger reported in production studies, providing some confirmation from another quarter that the change is towards the closer NEAR variant.

Introduction

The research project[1] that provides the data reported in this paper addresses the issue of how the speech recognition and comprehension system copes with ongoing sound change. In addition, it provides some psycholinguistic evidence that relates to the progress of one particular change that has been studied extensively in recent sociolinguistic and phonetic research, namely the merger of the NEAR and SQUARE diphthongs in New Zealand English (NZE).[2]

In part, our interest in this topic has been stimulated by popular opinion expressed about changes in the pronunciation (and other aspects) of language, especially a view that maintains that standards are slipping, or that "children can't speak or write properly any more" (Milroy 1998). In the case of the merger of the NEAR and SQUARE diphthongs in NZE, we read this from a 1994 opinion column in the Evening Post:

> To use the [...] examples of "beer" and "bare" and "here" and "hair", I go into this bar and say, "Beer, please" and the barmaid, being an obliging girl, takes off her top and bra. Because I am devoutly decent, I say indignantly, "Here! Here!" and the barmaid who knows when enough's enough whacks me with a jug of Old Dark which starts a bloody brawl.
>
> You see, the potential for misunderstanding is substantial and the consequences may be horrendous.
>
> Alex Veysey - Opinion column, *Evening Post*, October 29th 1994

and more recently from a correspondent to the New Zealand Listener:

---

[1] "Sound change and spoken word recognition in New Zealand English", funded in part by Victoria University Research Fund (URF) grant 1/11 to the second author. The experiment reported in this paper was carried out as an Honours research paper by the first author under the supervision of the second author, as part of this URF project.

[2] For clarity, we use NEAR and SQUARE (Wells 1982) to represent the lexical sets which in non-merging dialects contain the /iə/ and /eə/ vowels respectively.

> Regional accents are wonderfully enriching contributions to the English language – but an accent becomes degenerate when the spoken word cannot differentiate between totally different meanings. Too often we hear of people crossing on the Cook Strait "fairy" and flying "Ear" New Zealand.
>
> L. Bravery, letter to the editor, *Listener* June 9th 2001.

As psycholinguists, we find ourselves compelled to point out that situations in which such sound change results in miscommunication are probably unlikely. This is because information from the contexts in which NEAR or SQUARE words are encountered will make the intended meaning clear. There is considerable psycholinguistic evidence that suggests that the human language processing system makes rapid and efficient use of such contextual information to resolve ambiguity resulting from homophony. Some of this evidence is reviewed below. It is certainly unlikely that Veysey's barmaid will reach the interpretation illustrated above, or that, given the context in which an ambiguous word is used, the listener will not be able to "differentiate between totally different meanings." However, the fact that the NEAR/SQUARE merger is a change in progress, meaning that both NEAR and SQUARE forms are likely to be encountered, suggests that we should perhaps not assume that NEAR/SQUARE words are treated just the same as more established homophones.

As an aside, it should be noted that there are also productive consequences of the merger, as in headlines ("Spear a Thought" on a cookery item at the beginning of the asparagus season) and in the many entertaining names for hairdressing establishments such as "Hair Today, Gone Tomorrow", "Hair, There and Anywhere", "Always Hair", "Hair We Go" and so on. And of course these names are not so very different in character from names exploiting 'established' homophony such as "Thyme for Herbs" and "The Wright Builders", to take just two from the Wellington phone book.

Progress of the NEAR/SQUARE merger in NZE

Some of the most recently published work on the NEAR/SQUARE merger is that of Gordon and Maclagan (1999; 2001). These authors report a series of production studies conducted with fourth-form students (14-15 years old) in the same four Christchurch schools in 1983, 1988, 1993 and 1998. The students were recorded reading sentences and word lists containing word pairs with the NEAR and SQUARE vowels. Table 1 presents data from Gordon and Maclagan (1999), showing that while the early samples show a high degree of variability (i.e. more than 60% of productions showed no significant preference for either form), by 1993 there is a clear tendency for a merger on NEAR, although some speakers still keep the forms distinct. An additional study (Maclagan and Gordon 1996) compared two age groups recorded in 1994 and confirmed that a merger on EAR was more complete for younger speakers (20 – 30 years) than for older speakers (45 – 60 years).

YEAR	Merge towards NEAR variant	Merge towards SQUARE variant	NEAR and SQUARE kept distinct	No significant preference
1983	10	15	13	62
1988	27	5	3	65
1993	81	0	16	3
1998	85	0	11	4

Table 1. *Realisation patterns (%) for NEAR and SQUARE words (data from Gordon and Maclagan, 1999).*

Other studies have confirmed that the NEAR and SQUARE vowels are undergoing merger elsewhere in New Zealand (for a recent review see Batterham 2000), but they have not all agreed on the direction that the merger is taking. In one study, surveying speakers from Porirua near Wellington (Holmes and Bell 1992), the move amongst the younger speakers sampled appeared to be towards the SQUARE variant. It has been noted that Holmes and Bell's younger speakers would be near contemporaries of the first cohort of Gordon and Maclagan's fourth-formers, who also showed a slight preference for merger on SQUARE (see Table 1). The latter authors suggest that these SQUARE preferences may be due to *hypercorrection* (Maclagan and Gordon 1996), i.e. certain speakers became conscious of a merger towards NEAR, which they perceived to be a change towards a stigmatised form, and attempted to redress this by using the alternative form. Since, however, the merger was already well established, these speakers were no longer able to use the NEAR and SQUARE forms in a way that reflected the distribution found earlier in NZE and elsewhere in the English-speaking world, and ended up using the SQUARE form for NEAR words as well as for SQUARE words. Other reports of an increase in SQUARE forms come from Auckland. One survey (Starks et al. 1998) showed an apparent-time increase in SQUARE realisations of the word "ear", progressing from 5% in the 55-64 group through to 28% in the 15-24 group. Batterham (2000) also reports more SQUARE realisations of NEAR words amongst younger Auckland speakers, but additionally notes that there is a stronger trend for change in words with SQUARE words, with working-class females in particular realising these with NEAR vowels. Most other research supports a merger on the NEAR form, but with varying degrees of completion depending on a number of demographic variables, in particular speaker age but also including speaker sex and ethnicity.

Psycholinguistic studies of ambiguity

If words with NEAR and SQUARE vowels (e.g. *cheer, chair*) have become indistinct for some speaker groups then we are likely to be dealing, for those speakers, with homophony. That is, the pronunciation of pairs of lexically different and phonetically formerly distinct words have become indistinguishable. Recent psycholinguistic research using a lexical decision task (where the participants must decide whether each stimulus is a real word) has shown that homophones take longer to recognise than unambiguous words (Rodd et al. 2002), and so it might be supposed that there is some truth in the claim that the merger is resulting in less clarity. However, it should be stressed that despite this relative disadvantage for homophonous words, they are still recognised very rapidly as real words in the lexical

decision task. Responses were made on average less than a second after the beginning of the word, or just over a third of a second after the end of the word, and less than 50msec later than responses to unambiguous words.

It is interesting to note that while homophones are recognised more slowly than unambiguous words, the opposite is observed for polysemous words, i.e. words with different but related senses, which are recognised more rapidly than unambiguous words. Rodd and her colleagues explain this difference between homophones and polysemes in terms – effectively – of an accumulation of activation of a single lexical item on the basis of a range of senses in the case of polysemy but competition between different lexical candidates in the case of homophony.

Rodd et al's research used isolated words, and simply required that a word be recognised as a real word, without asking which sense(s) had been recognised. Other studies have shown that when ambiguous words are presented in context, multiple senses may be available, but context is quickly used to select the appropriate meaning. Swinney (1979), for example, used a semantic priming task, in which he presented subjects with spoken sentences containing ambiguous words such as in (1)

(1) ... the man was not surprised when he found spiders, roaches, and several other bugs$_{(A)}$ in the corner$_{(B)}$ of his room

At either point (A), the offset of the ambiguous word, or point (B), a few syllables later, subjects were presented with a probe word on a visual display, about which they had to make a lexical decision. These words were either related to one of the meanings of the homophone, such as *ant*, or *spy*, or they were unrelated to either meaning, such as *sew*. Swinney found equal facilitation of both *ant* and *spy*, compared with *sew*, at point (A), even though the context biases the sentence towards the insect meaning of *bugs*. However, at point (B), facilitation was only found for the word related to the appropriate meaning, *ant*, with both *spy* and *sew* receiving slow responses in comparison. This demonstrates that on encountering a homophone, regardless of context, both meanings of the homophones are accessed. Importantly, though, context then helps to quickly select the appropriate meaning from those available.

Subsequent research has extended and refined Swinney's initial findings. For instance, it has been shown that the different meanings of an ambiguous word such as rose ("flower", "stood up") are initially accessed regardless of the fact that one of them may be syntactically as well as semantically inappropriate to the sentence context (Seidenberg et al. 1982). A more recent experiment (Simpson 1994) measured reading times for ambiguous words in visually-presented sentences, and compared words where the frequencies of the different meanings are relatively equal (*balanced* words) and words where one meaning is much more likely than the other (*unbalanced* words). This study found that if there was no disambiguating prior context, then the balanced words required more processing (as reflected in longer reading times) than the unbalanced words. If there was a preceding disambiguating context, then the longest reading times were for an unbalanced word when the less frequent meaning was signalled by the context. Reading a balanced word was no different from reading an unbalanced word when the dominant

meaning was signalled. This suggests that the resolution of lexical ambiguity can involved a rather complex interaction of properties of the ambiguous word and aspects of the context.

Considerations of context such as those outlined in the preceding paragraphs are clearly important in our continuing deliberation of the processing of newly ambiguous forms that result from sound merger. However, they were not pursued in the study reported in this paper, which started our experimental investigation of the consequences for word recognition of the NEAR/SQUARE merger by considering words in isolation. In particular, we wished to determine whether words with NEAR or SQUARE vowels are treated as though they are homophones, giving access to the meanings of both words in a NEAR/SQUARE pair. If there is evidence of homophony, then is this true of both pronunciations of the pair, and if so, is it *equally* true of both pronunciations?

To use a concrete example – when listeners hear [tʃiə], do they access meanings associated both with *cheer* and with *chair*? Is the same true when they hear [tʃeə]? If evidence of homophony is equally present in both cases, then this suggests that both pronunciations are equally good realisations of a merged form. This merged form may be phonetically intermediate between [tʃiə] and [tʃeə], with each of these being in some sense a deviation from the merged form. Alternatively, the merged form may have multiple realisations including [tʃiə] and [tʃeə] in free variation. If homophony is evident in both cases, but the relevant meanings are more strongly accessed, say, in the [tʃiə] case, then this might suggest that merger on [tʃiə] is complete, leaving the [tʃeə] form as a less good phonetic variant. Finally, if homophony is evidenced only in the case of one of the forms, with the other form giving access to just one set of meanings, then this could reflect an asymmetric merger-in-progress, with the merger proceeding towards one form and with the other form still available to listeners as an unambiguous word, a "relic" as it were of the old system. Given that the predominant evidence is that the merger is towards NEAR, we would predict that such an asymmetric system for NZE would have [tʃiə] as a homophonous form for *cheer* and *chair*, but [tʃeə] as a form for just *chair*. Alternatively, of course, if the evidence from Holmes and Bell's (1992) Porirua survey reflects a merger in the Wellington region on the SQUARE form, then we would expect this last prediction to be reversed.

To investigate these possibilities, we needed to make use of an experimental paradigm that explores access to the meanings of ambiguous words in isolation from context. One such paradigm involves semantic priming, and requires subjects to respond to pairs or series of items which bear some meaning relationship to one another, as well as to unrelated items. Semantic priming obtains when a word is responded to more rapidly or more accurately when it has been preceded by a related word. For instance, *nurse* would be expected to be recognised more rapidly (as reflected in lexical decision times) when it has been preceded by *doctor* than when it has been preceded by *table*, because of the semantic relationship between *doctor* and *nurse*. In such a study, *nurse* would be the probe word of interest, and *doctor* and *table* would be primes (related and control primes respectively). This semantic priming paradigm was used in an early study (Meyer and

Schvaneveldt 1971) which revealed significant facilitation of probe words in a lexical decision task when these words were preceded by related primes.

An important issue in priming studies is the type of semantic relationship between prime and probe. A recent cross-modal priming study (Hino et al. 1997) used auditorily-presented ambiguous words as primes followed by visually-presented words as lexical decision probes. Hino et al explored both meaning dominance (the more frequent meaning of *habit* is something you do often rather than the nuns' outfit) and possible differences between associatively related words (e.g. *habit* with *smoke* or *nun*) and words related through their semantic features (e.g. *habit* with *frequent* or *dress*)[3]. They found that that if the probe was presented immediately after the prime then they obtained priming of associates related to both meanings of the ambiguous word but no priming of words related through semantic features. Later probes (700msec after offset) resulted in priming of words related to the dominant meanings of the ambiguous word (*habit* primed *smoke* and *frequent*) but not of words related to subordinate meanings (*habit* failed to prime *nun* or *dress*), regardless of whether or not they are associatively related. This was taken to suggest that lexical (associative) relationships have an immediate effect while non-linguistic (cognitive-semantic) relationships take longer to build up. Other research using auditory-auditory priming (Moss et al. 1995) with an interval of 1000msec between items showed significantly stronger priming for associates than for non-associatively semantically related items. These researchers also found (Moss and Marslen-Wilson 1993) that if two words are related by association, then the probe word is primed regardless of context, but if they are related via some semantic feature, then context can affect priming. These studies suggest that the best basis for testing whether [tʃiə] and [tʃeə] are homophones is to present each of them as auditory primes followed by probes that are *associatively* related to each of *cheer* and *chair*.

A further important issue is whether to use a single- or cross-modal version of the semantic priming task. Evidence relevant to this question comes from studies that have compared paired and list presentation of items. Paired presentation is when items are presented as isolated pairs of stimuli, and subjects are required either to respond to each item in a pair, or to ignore the first and respond only to the second, or to respond to the pair (e.g. by only making a positive response in a lexical decision task if both items are real words). In list presentation, word pairs of interest (such as *doctor-nurse*) are embedded in a longer list of real and nonsense words, with every item in the list requiring an "independent" response. Research comparing these presentation formats has argued that paired presentation can encourage both expectancy generation and post-lexical checking (McNamara and Altarriba 1988; Shelton and Martin 1992). Expectancy generation is when subjects generate a set of possible targets on the basis of the prime, and then check these off against the probe. This would result in facilitation of the probe when

[3] Associative relationships are the kind reflected in word association tasks ("what is the first word that you think of when I say *habit*?") and are largely to do with the relationships between words (and may also be referred to as lexical semantic relationships). Feature-based relationships reflect the fact that the meanings of words may have properties in common, or the properties of one word may be involved in defining the other. This type of relationship is cognitive rather than lexical.

it is in the set, but inhibition when it is not (for instance after unrelated primes). Post-lexical checking includes assessing the relation between the prime and target before making a decision about the target. Decisions might then be based on whether such a relationship exists (which must indicate a word response). These decisions can also be affected by the strength of the relationship. Single item presentation is claimed to eliminate relatedness-checking as a post-lexical strategy (McNamara and Altarriba 1988), since it increases the number of unrelated real word items occurring next to one another and so reduces the efficacy of the strategy. McNamara and Altarriba also found that list presentation produced mediated priming (e.g. the prime word *lion* facilitates the probe word *stripes* via the relationships of *lion* with *tiger* and of *tiger* with *stripes*) regardless of whether the lists include directly related pairs (*tiger-stripes*). In paired-item tasks, however, mediated priming is reduced when the experiment also includes directly related pairs, since the presence of these directly related pairs encourages participants to check the relatedness between pair members. It would thus seem that list presentation is less likely than paired presentation to result in strategic approaches such as post-lexical relatedness checking. This encouraged us to use list presentation in our experiment, which means also that we need to use auditory-auditory rather than cross-modal presentation of prime and probe items.

Experiment

Materials

The list of items to be presented in our experiment had to include both members of a NEAR/SQUARE pair (e.g. [tʃiə] and [tʃeə]) as primes, followed by probe words semantically associated to each of the members of the pair (i.e. associates of *cheer* and of *chair*). In addition, in order to obtain a baseline response time against which the priming of these associates could be measured, we needed conditions in which the probe words were preceded by unrelated primes (Jonides and Mack 1984). Overall, we selected 12 pairs of words containing the NEAR/SQUARE vowels and 12 pairs of matching semantic associates, giving 24 primes and 24 probes. All items were monosyllabic; the real words were approximately matched for frequency. In addition to these 12 target word pairs and their 12 pairs of associates, we obtained 24 control stimuli that were unrelated in meaning to the probes.

A number of factors had to be considered in selecting the test words (primes and probes). For some NEAR/SQUARE pairs there are NEAR or SQUARE words which are themselves homophones (e.g., *deer* and *dear; bear* and *bare; sheer* and *shear*). To make the following description clearer, we will refer to these as *homophonous members* (distinguishing them from the potential homophone pairs *beer* and *bear*, etc.) As the stimuli are presented auditorily, choosing between the meanings of these homophonous members does not affect the form of the prime, but it does affect the selection of the probe word. Note that in the list presentation format we need to leave sufficient time between items (including primes and probes) for a lexical decision to be made to each item, and that it was reported above that Hino et al (1997) found in a cross-modal priming experiment that dominant meanings persevered over an interval of 700msec, but not subordinate meanings. Other research also suggests that the more frequent meaning is accessed faster and retained

longer (Simpson 1981). However, it has also been pointed out that associates exhibit more robust priming in single-modal presentation than in cross-modal experiments (Moss et al. 1992). Nevertheless, wherever possible we used probe words associated with the most frequent meaning of a homophonous member.[4] However, it transpired that in some cases the most frequent meaning had less clear associates than another meaning. For example, the frequency corpora gave *sheer* a higher frequency than *shear*, but intuitively, *shear* followed by *sheep* would provide greater priming effects than *sheer* followed by *cliff* or *tights*. Since associative priming appears to be fairly robust (Moss et al. 1995), we felt it important to be sure that there were strong associative relationships between the primes and probes used. Therefore, on occasion we used a lower frequency prime instead of the most dominant meaning of a homophonous member. However, it is worth emphasising that the frequency differences between the meanings of these homophonous members was never great.

Probe words were selected on the basis of three sets of published association norms[5]. Polysyllabic probe words were avoided, and where possible concrete and natural category words were used (Moss et al. 1995). In some instances these constraints meant that the probes were not the most frequent response in the association data. Some of the prime words were not included in the association norms, and in these cases probe words were chosen based on our own intuitions.

Control words and filler words were selected from a computerised list of monosyllables available in our speech laboratory. Control words were unrelated to the probes and were phonologically matched with the prime words in terms of being open syllables. Fillers were 20 real words, which were included in the test lists to disguise the relationships in the test data and reduce the risk of strategies such as relatedness checking. Finally, 44 non-words were included, which were constructed by altering one phoneme in existing words – e.g., *jatt* from *cat*.

All stimuli were spoken by a native NZ English speaker who was asked to be careful to distinguish the NEAR and SQUARE vowels. Her own spontaneous speech does distinguish the two, but not consistently for all words. Stimuli were recorded to digital tape in a sound-treated room, and then transferred to Macintosh computer for use in the experiment.

Design

The overall design consisted of 4 test conditions and 2 control conditions:

[4] Frequencies were taken from the Wellington Corpora of New Zealand English.

[5] EAT (Edinburgh University); Florida (University of Southern Florida) (see reference list for URLs) and the Birkbeck Association Norms (Moss and Older 1996).

NEAR prime followed by NEAR probe (NEAR_NEAR – e.g., *cheer, shout*)
SQUARE prime followed by SQUARE probe (SQUARE_SQUARE – e.g., *chair, sit*)
NEAR prime followed by SQUARE probe (NEAR_SQUARE – e.g., *cheer, sit*)
SQUARE prime followed by NEAR probe (SQUARE_NEAR – e.g., *chair, shout*)
Control prime followed by NEAR probe (control_NEAR – e.g., *bee, shout*)
Control prime followed by SQUARE probe (control_SQUARE – e.g., *toe, sit*)

The 24 test items occurred in each of the four test conditions. These conditions were rotated across the 24 items to produce four test blocks, each of which contained six items in each condition. Because our research goals included a comparison of priming effects in each subject's reactions in the lexical decision task with distinctions between NEAR and SQUARE in the subject's own productions of the word pairs involved, we tested each subject on each block. The entire experiment was run over two sessions (a week apart) with two of the test blocks in each session. Session one included block 1 and block 2 and session two included block 3 and block 4, but the order of blocks within sessions varied across subjects. Subjects were randomly allocated to one of two groups – group one heard block 1 before block 2 in session one and block 3 before block 4 in session two. This was reversed for the group two. For any item the two blocks in a single session did not include a repetition of either the prime or the probe word for that item. Between the test blocks in each session, subjects took part in filler tasks in order to separate the two blocks and reduce long-term priming of probes by related items occurring in the preceding block. However, the intervening tasks appeared not to fulfil this function, since the second block in each session had much reduced condition differences compared to the first block. For this reason, the data analysed for this paper are taken only from the first block in each session.

Each test list consisted of 12 primes (6 NEAR words and 6 SQUARE words) together with 12 probes, as well as 20 further real words as fillers and 44 nonwords The control primes were included with the probe words and filler items in a separate list that was administered as a further block at the end of the second session (and after a further filler task). As well as the control prime-probe combinations (48 items), the control list contained 26 real word fillers and 74 non-words. The word and non-word fillers reduced the number of sequences of unrelated words, reducing expectations that a word would be followed by another semantically related word.

Subjects and procedure

The 16 subjects were Linguistics undergraduates at Victoria University who volunteered as a part of their course of study, but who were naïve concerning the purpose of the experiment. They were all fluent native NZE speakers, aged 18-25 with no known hearing impairment.

Subjects were tested individually in a quiet room. They heard the sequence of real word and non word stimuli over closed-ear headphones and pressed one of two response buttons on a millisecond timer to indicate a real word or non word decision. The experiment was controlled by the PsyScope programme (Cohen et al. 1993). The inter-stimulus interval was set at 1400 milliseconds, which had been determined as long enough to allow subjects to respond, but short enough to maintain a fast response rate.

Predictions

Our main prediction for this experiment was based on the documented trend in NZE towards merger on the NEAR vowel, together with the observation that the merger is a change in progress. In other words, we predict that the NEAR forms of our test pairs will be treated as though they are homophones for NEAR and SQUARE words. This will result in semantic priming of probe words related to both NEAR and SQUARE words after hearing the NEAR form. In terms of an actual example, the form [tʃiə] will result in access of both *cheer* and *chair*, which will in turn prime *shout* and *sit*. This priming will be reflected in facilitation of lexical decision responses to *shout* and *sit* after [tʃiə] compared with responses to the same items after their unrelated control primes.

The SQUARE forms of our test pairs, in contrast, will be responded to as though they are not homophones, producing semantic priming only of words related to the SQUARE word. That is, on hearing [tʃeə] subjects will access only chair, which will prime *sit* but not *shout*. Therefore there should be facilitation of lexical decision responses to *sit* but not to *shout*, relative to their control conditions.

Results

Response times (RTs) for each item were recorded in milliseconds. One item set – *fear/fair* – was excluded due to an error in the experiment design. Missing data and incorrect responses were excluded from the analysis, affecting 3.69% of the test and control data. Facilitation times (FTs) were calculated by subtracting a subject's RT for a probe in a test condition from that subject's RT to the same probe in the control condition. Average FTs in each of the four test conditions were calculated for each subject, and the resulting averages were subjected to Analysis of Variance, with Facilitation Time as the dependent variable and Prime (NEAR word vs SQUARE word) and Probe (*appropriate* [e.g. word related to the NEAR word occurring after the NEAR word] or *inappropriate* [e.g. word related to the SQUARE word occurring after the NEAR word]) as independent variables. In addition, planned comparisons of probe types following each kind of prime were used to test the predictions outlined above. Overall averages for the four conditions are shown in Figure 1.

The statistical analysis confirmed that *appropriate* probes received greater priming than *inappropriate* probes ($F(1,15)=5.85$, $p<0.03$). In addition, there was a significant interaction of prime and probe type ($F(1,15)=4.90$, $p<0.05$). In the planned comparisons, SQUARE primes facilitated *appropriate* probes more than *inappropriate* probes ([tʃeə]-*sit* > [tʃeə]-*shout*; $F(1,30)=6.10$, $p<0.02$), while there was no difference in the priming of *appropriate* and *inappropriate* probes by NEAR primes ([tʃiə]-*shout* ≈ [tʃiə]-*sit*) ($F(1,30)=0.23$, $p>0.60$).

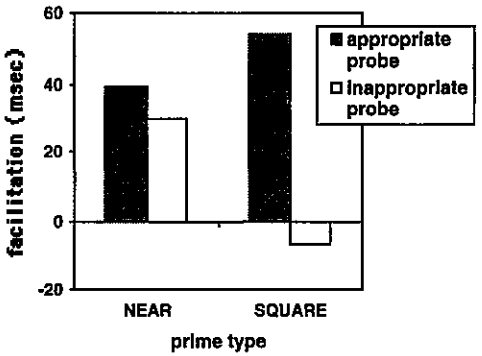


Figure 1. Average facilitation times (in milliseconds, relative to the control condition) for appropriate and inappropriate probes following *ear* and *air* primes (eg., from left to right in the figure, *cheer-shout; cheer-sit; chair-sit; chair-shout*).

General discussion

Our result for NEAR forms presented above supports our prediction made earlier, namely that these forms are ambiguous between NEAR words and SQUARE words, following the trend in NZE towards merger on NEAR. In other words, forms containing the [iə] vowel are treated as though they are homophones, resulting in semantic priming of both NEAR-related and SQUARE-related associates. The fact that there is facilitation of SQUARE-related probes after SQUARE forms shows that these forms are still recognised. However, the fact that there is no facilitation of NEAR-related words after the [eə] vowel shows that the SQUARE form is not homophonous between SQUARE and NEAR. So the SQUARE form appears not to be a phonetic variant of the NEAR form, but is still recognised as a separate form.

A further observation from the above results that might support our contention that NEAR forms are homophonous but SQUARE forms are not relates to the difference in facilitation levels after NEAR and SQUARE. Recall that recent work by Rodd and her colleagues (Rodd et al. 2002) showed that homophones are responded to more slowly than unambiguous words in a lexical decision task, reflecting competition between two different words with the same form. If this disadvantage for ambiguous words has a knock-on effect for associates presented subsequently, then we would predict that words associated with the homophones might be less strongly primed than words associated with unambiguous words. The data shown in Figure 1 would appear to support this position, with lower levels of facilitation for either word following NEAR primes than for the SQUARE word following the SQUARE prime. Note though that compared to the control condition there is still facilitation of probe words related to both meanings, so that any disadvantage is relative to priming by unambiguous words, rather than a claim that uncertainty about the lexical identity of NEAR words results in inhibition of lexical representations.

Overall the result of this experiment provides support from a spoken word recognition study for the view that the merger of NEAR and SQUARE vowels is proceeding towards the closer NEAR form. The asymmetry in the data would be difficult to explain if the merger was in the opposition direction, towards SQUARE, since this would predict that the SQUARE form would prime associates of the NEAR word, which is the condition in our experiment that shows no facilitation. It would seem that the presence of the SQUARE form (mainly in the speech of more conservative NZE speakers) is sufficiently strong to result in recognition of this form and significant facilitation of associates of the SQUARE words.

As noted earlier, these results are based on the first block of test items in each session. The second block produced much less clear results, which may be due to long-term priming across the session or possibly to fatigue effects. Since we have therefore discounted half of the data from the experiment, the current study should be regarded as a pilot study, and a replication of the experiment is needed to confirm the results. With such a confirmation in hand, we will be better able to relate results for individual participants and/or items in the priming experiment to production data for those participants and/or items, and also to discrimination data for pairs of NEAR/SQUARE items. This will allow us to explore more fully the relationship between production and perception/recognition in the process of sound change.

In addition, our future research will examine the effects of sentence context on the recognition of our NEAR/SQUARE tokens. For instance, will we find for NEAR words the same pattern as Swinney (1979) of immediate access to multiple meanings and rapid integration of contextual information in the process of selecting from those meanings? What will we find for SQUARE words – will there be any early access to multiple meanings that is not evidenced in the auditory-auditory priming task with its longer inter-stimulus interval?

Finally, we wish in our future research to examine the effects on perception and recognition of demographic information concerning the speaker. If the distinction between NEAR and SQUARE forms is indicative of a more conservative speaker, will participants who believe the speech to have come from an older speaker respond differently from those who believe the same tokens to have come from a young speaker?

References

Batterham, Margaret 2000. The apparent merger of the front centring diphthongs - EAR and AIR - in New Zeland English. *New Zealand English*. K. Kuiper. Amsterdam, John Benjamins: 111-145.

Cohen, Jonathan D., Brian MacWhinney, Michael Flatt and Jefferson Provost 1993. Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioural Research Methods, Instruments & Computers* 25(2): 257-271.

Gordon, Elizabeth and Margaret Maclagan 1999. "Are we NEAR the SQUARE?" A real time study of the merger of two centering diphthongs in New Zealand English. *Linguistic Society of New Zealand*, Palmerston North.

Gordon, Elizabeth and Margaret Maclagan 2001. "Capturing a sound change": A real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. *Australian Journal of Linguistics* 21(2): 215-238.

Hino, Yasushi, Stephen J. Lupker and Chris R. Sears 1997. The effects of word association and meaning frequency in a cross-modal lexical decision task: Is the priming due to "semantic" activation? *Canadian Journal of Experimental Psychology* 51(3): 195-210.

Holmes, Janet and Allan Bell 1992. On shear markets and sharing sheep: The merger of EAR and AIr diphthongs in New Zealand English. *Language Variation and Change* 4: 251-273.

Jonides, John and Robert Mack 1984. On the cost and benefit of cost and benefit. *Psychological Bulletin* 96(1): 29-44.

Maclagan, Margaret and Elizabeth Gordon 1996. Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change* 8: 125-147.

McNamara, Timothy P. and Jeannette Altarriba 1988. Depth of spreading activation revisited: semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language* 27: 545-559.

Meyer, David E. and Roger W. Schvaneveldt 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 21: 863-883.

Milroy, James 1998. Children can't speak or write properly any more. *Language Myths*. P. Trudgill. London, Penguin: 58-65.

Moss, Helen E. and William D. Marslen-Wilson 1993. Access to word meanings during spoken language comprehension: Effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 1254-1276.

Moss, Helen E. and Lianne Older 1996. *Birkbeck Word Association Norms*. Hove, UK: Lawrence Erlbaum.

Moss, Helen E., Ruth K. Ostrin, Lorraine K. Tyler and William D. Marslen-Wilson 1992. Semantic and associative priming within and across modalities. *25th International Congress of Psychology*, Brussels, Belgium.

Moss, Helen E., Ruth K. Ostrin, Lorraine K. Tyler and William D. Marslen-Wilson 1995. Accessing different types of lexical semantic information: evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(4): 863-883.

Rae, Megan and Paul Warren 2002. Goldilocks and the three beers: sound merger and word recognition in NZE. *New Zealand English Journal* 16: 33-41.

Rodd, Jennifer, M. Gareth Gaskell and William D. Marslen-Wilson 2002. Making sense of semantic ambiguity: semantic competition in lexical access. *Journal of Memory and Language* 46: 245-266.

Seidenberg, Mark S., Michael K. Tanenhaus, J.M. Leiman and M. Bienkowski 1982. Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. *Cognitive Psychology* 14: 489-537.

Shelton, Jennifer R. and Randi C. Martin 1992. How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(6): 1191-1210.

Simpson, Greg B. 1981. Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior* 20: 120-136.

Simpson, Greg B. 1994. Context and the processing of ambiguous words. *Handbook of psycholinguistic research*. M. A. Gernsbacher. San Diego, Academic Press: 359-374.

Starks, Donna, Scott Allan and Cath Kitto 1998. Why vernacular speech? Speech samples from the taped Auckland rapid and anonymous survey. *New Zealand Language and Society Conference*, Victoria University of Wellington.

Swinney, David A. 1979. Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour* 18: 645-659.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

URLs

Edinburgh Associative Thesaurus: http://monkey.cis.rl.ac.uk/Eat/htdocs/eat_old.html

Florida Free Association Norms: http://w3.usf.edu/freeassociation

Experiments on /r/-intrusion

Jen Hay and Paul Warren

University of Canterbury and Victoria University of Wellington.

Introduction and Background

This paper reports the results of three simple experiments on the use of intrusive /r/ in New Zealand English. Most non-rhotic dialects of English exhibit /r/-sandhi (see e.g. Docherty and Foulkes 1999, Foulkes 1997a, b, Tollfree 1999, Bauer 1984, Trudgill 1974, Wells 1982, Williams and Kerswill 1999). /r/-sandhi can take two forms – linking /r/ and intrusive /r/. The term linking /r/ refers to cases in which the /r/ is orthographically present, and is produced across a morpheme or word boundary when followed by a vowel (e.g fearing, car alarm). Intrusive /r/ occurs in the same environments, but when there is no orthographic /r/ present at the point of relevance (e.g drawing, ma and pa).

In dialects which display /r/-sandhi, production of /r/ across word boundaries is variable (Jones 1964, Gimson 1980, Wells 1982). Linking /r/ appears to occur at higher rates than intrusive /r/, perhaps because the latter may be associated with a certain amount of stigmatisation (see, e.g. Lewis 1975; 1977, Pring 1976, Fox 1978; Brown 1988). There may also be an effect of spelling pronunciation involved. Word-internally, linking /r/ "almost invariably occurs", while there are "occasional" instances of word-internal intrusive /r/ (Wells 1982: 224). Gimson and Cruttenden (1994: §12.4.7) claim that in RP "the insertion of /r/ is obligatory before a suffix beginning with a vowel, where the /r/ is historically justified", whereas "the insertion of intrusive /r/ before a suffix is strongly resisted".

The existence of variability in the production of /r/-sandhi raises the question of the degree to which this variability is socially or linguistically conditioned.While there have been several attempts to model /r/ sandhi processes in formal theoretical models, empirical (data-oriented) work on this topic has remained fairly sparse. The most comprehensive corpus-oriented study was conducted by Foulkes (1997a, b), who studied /r/-sandhi in a stratified sample of speakers from both Newcastle upon Tyne and Derby. While social categories appeared to play no role in Derby, there were clear class and age differences in Newcastle. These results suggest that /r/-sandhi may be socially conditioned in some dialects.

Jones (1964) and Gimson (1980) claim that intrusive /r/ is more likely following schwa than other vowels. It is also claimed to be less stigmatised in this context - one explanation that has been put forward for this is that schwa is definitionally unstressed, so /r/ tends to be "less noticed" in this environment (Crystal 1984:43; Brown 1988:149). Wells claims intrusive /r/ after /ɔ/ is more stigmatised than after other vowels because it was a later innovation (1982: 225). Sudbury and Hay's (2002) study of 19th century New Zealand English shows the lowest rates of intrusive /r/ after /ɔ/, supporting Wells' claim.

No large-scale systematic study has investigated rates of /r/-intrusion following different vowels. And while it has been claimed that /r/-sandhi is more common across word boundaries than morpheme boundaries, no-one has investigated whether different types of suffixes may give rise to different rates of /r/-intrusion. Empirical data on the possible conditioners of intrusive /r/ are virtually non-existent.

Phonological theories of /r/-sandhi focus on modelling the alternation between pre-vocalic /r/, and non-prevocalic /r/-lessness, with much debate focussing on whether the /r/ is underlyingly present or not (see, e.g. McCarthy 1993, Harris 1994, Vennemann 1972, Johansson 1973, McMahon, Foulkes and Tollfree 1994, McMahon 2000). Virtually all of this literature assumes intrusive /r/ to be a categorical phenomenon, assuming implicitly that word external /r/-sandhi processes are obligatory in the dialects that contain them. This is in stark contrast with the descriptive literature outlined above, which seems to have struck a consensus that these processes are extremely variable.

Thus, there are important respects in which the descriptive and the theoretical literature are incompatible. In addition, the majority of the descriptive literature is itself based on informal observation rather than extensive corpus work.

The lack of empirical data on intrusive /r/ likely relates to the scarcity of appropriate environments for its occurrence. For example, in Sudbury and Hay's (2002) analysis of rhoticity, linking /r/ and intrusive /r/ in natural speech, they analysed 13760 potential environments for rhoticity, 2975 for linking /r/, and only 203 for intrusive /r/. That is, for every 68 potential environments for non-prevocalic /r/ and every 15 potential environments for linking /r/ there was just one potential environment for intrusive /r/.

In the studies reported here, then, we decided to sacrifice the advantages of natural speech in favour of reading and production studies, which ensured collection of sufficient data, and enabled us to systematically manipulate potentially relevant linguistic factors. Our goal was to conduct a series of simple experiments, in order to acquire some preliminary data on the linguistic conditioning of intrusive /r/.

Experiment 1

The first experiment involved the reading aloud of sentences containing potential environments for intrusive /r/. Sixteen University of Canterbury students received chocolate fish for their participation in a set of production and perception tasks on a range of linguistic topics. All students were native speakers of New Zealand English, and non-rhotic. The tasks included 48 sentences targeting intrusive /r/. A total of 768 tokens (16 speakers * 48 sentences) were analysed. The 48 sentences were randomised together with 86 "filler" sentences which were designed to test hypotheses unrelated to /r/-sandhi. All sentences were read from index cards, and recorded on Digital Audio Tape. Only the 48 sentences relating to intrusive /r/ are discussed here.

Base word	*expected final*	*Following environment*
Sofa	ə	-ing
Oprah	ə	-y
Ma	ɐ:	-ism
Bra	ɐ:	-ify
Claw	o:	-ish
Plough	æʉ	-ese
		-ize
		# in

Table 1 : Experiment 1 base words and environments

Table 1 shows the base words and environments tested. The base words were chosen to represent a range of different vocalic environments. The finals listed in the second column are the expected vowel qualities at the end of the base, although it should be noted that there is some variation in these qualities amongst speakers of NZ English (see Bauer and Warren, submitted). The word "plough" was included to assess the degree to which intrusive /r/ is used after /æʉ/ - which appears to be an emerging feature of New Zealand English, occurring concurrently with a change in the phonetics of /æʉ/. While intrusive /r/ cannot occur after [+high] vowels (see, e.g. Wells 1982: 226), /æʉ/ is undergoing change in NZ English. The second element of the diphthong is now tending towards COMMA rather than FOOT (see Maclagan, Gordon and Lewis 1999). /æʉ/ is therefore becoming a centering diphthong, rather than a closing one. This appears to make it eligible for intrusive /r/ for some speakers.

All base words were combined with all affixes listed in the final column of Table 1. A range of suffixes was tested, to test the effect of boundary strength on the likelihood of emergence of intrusive /r/.

Over all of the sentences recorded, 29% were produced with intrusive /r/. Individual subjects ranged from 0% to 77%, with a median of 21%. There was only one subject who produced no tokens of intrusive /r/. The overall rate of intrusive /r/ is likely to be much lower than would occur in natural speech - the sentences are clearly contrived, and the subjects are being recorded. However we still predict that the places in which intrusive /r/ is most likely under these circumstances reflect the environments in which it would be most likely to occur in a natural setting.

Note, also, that this rate of occurrence of intrusive /r/ is far from categorical, a fact which is perhaps not surprising, but is nonetheless difficult for the formal accounts of this phenomenon to explain. The use of intrusive /r/ in this setting also appears to reflect some social variation. The number of participants is too small for us to conduct any large-scale social analysis of the speakers' use. However we did record the Elley-Irving scores of the participants' parents. The Elley-Irving scale is a New Zealand-specific index of social stratification which assigns scores from one (highest) to six to New Zealand occupations (Elley and Irving 1985). We found the Elley-Irving score for the father to be a significant predictor of an individual's overall rate of intrusive /r/. Individuals with fathers with lower Elley-Irving scores (and so

higher socio-economic status) produced fewer instances of intrusive /r/ (spearman's rho = 0.56, $p<0.05$). The score for the mother approached significance (in the same direction), but did not reach it, and is rejected by stepwise models of the data.

Figure 1 shows the overall rate of intrusive /r/ for different base words. There is clearly a large amount of variation across the different words, with *claw* and *sofa*, for example, attracting more than three times the rate of intrusive /r/ as *Oprah* (chi-square = 34.3, $p<0.001$). Part of this variation is no doubt due to the identity of the preceding vowel (e.g. *claw* versus *plough*). There may also be some other factors at work, for example a potential avoidance of /r/ after proper names (c.f. *bra* vs *ma*, *sofa* vs *oprah*).

It is particularly interesting that the highest rates of /r/ occurred with *claw*, given previous claims (e.g. Wells 1982) that this vowel is in fact the most resistant to intrusive /r/, and Sudbury and Hay's (2002) finding that intrusive /r/ after THOUGHT was not prevalent among speakers of 19th century New Zealand English. The data collected here suggest that in fact the opposite may now be true. Future research will need to address whether this is due to a change in the /r/-intrusion rule, or a change in the quality of the vowel.

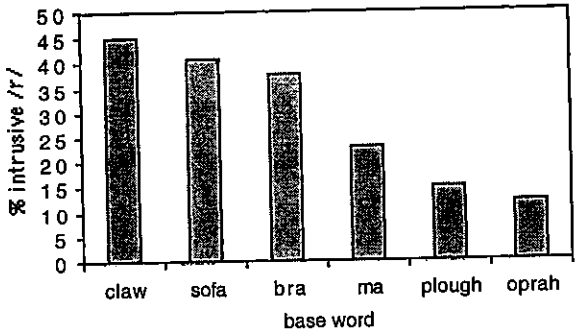


Figure 1: Rate of intrusive /r/, by base word.

Figure 2 shows the rate of intrusive /r/ broken down by the following linguistic environment. Here, too, we see a high degree of variation. The lowest rate of intrusive /r/ occurred across word boundaries (where there is scope for a pause to intervene), and the highest rate occurred with the inflectional affix *-ing*.

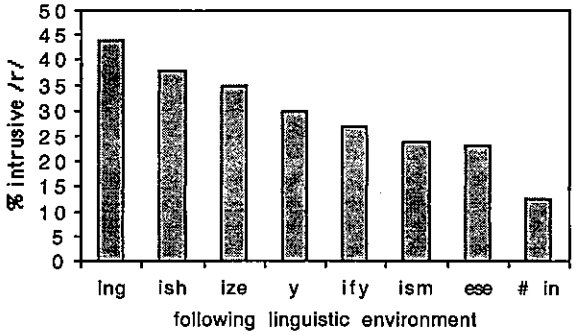


Figure 2: Rate of intrusive /r/, by following linguistic environment

In fact, if we consider the nature of the different affixes involved, the ordering shown in Figure 2 does not look completely random. *-ing* is the affix with the strongest (most word-like) boundary, and attracts the highest rate of intrusive /r/. *ish* is also a relatively productive, relatively separable affix. This contrasts with affixes like *-ese* and *-ify*, which are less frequent, less separable, and more likely to occur with bound roots (see, e.g. Hay and Baayen 2002)

Hay and Baayen (2002, forthcoming) investigate English derivational affixes, and identify a range of measures which are associated with different aspects of parsability and productivity. Figure 3 shows how the rates of intrusive /r/ for the 6 derivational affixes correlate with 4 of these measures. The top left panel shows the rate of intrusive /r/ as predicted by the summed token frequency of the affix. This figure represents the cumulative frequency of occurrence of all words which contain the affix. Affixes with high summed token frequencies tend to be represented by a large number of words, and/or high frequency words. The overall frequency of occurrence of the affix is high. The figure shows how high token frequency tends to correspond to higher rates of intrusive /r/.

Hay and Baayen demonstrate that when base frequency is regressed on derived frequency, there is a fair amount of variation across affixes. In particular the location of the intercept of this line on the y axis (base frequency) varies considerably. A high intercept reflects a distribution in which bases tend to be frequent, relative to derived forms. This indicates that the affix has many forms with frequency characteristics which tend to make them highly decomposable. The top right panel of figure three shows how this value (the intercept) relates to /r/-intrusion. Affixes with high intercepts (and so many highly decomposable words) are more prone to /r/-intrusion than affixes with low intercepts.

The bottom left panel shows the type-parsing ratio. This is a second index of the decomposability of forms containing an affix. Affixes with high type-parsing ratios are represented by a high proportion of forms for which the base word is sufficiently frequent relative to the derived word to facilitate parsing. Affixes with high type-parsing ratios, then, tend to be more separable – represented by more words which are highly decomposable. The graph shows how rates of /r/-intrusion are related to this aspect of an affix – more separable affixes are more prone to /r/-intrusion.

Finally, the bottom right panel shows the summed type frequency – how many different words are attested which contain that affix. Affixes with high type frequencies have a high degree of use, affixes with low type frequencies occur in a strictly limited set of words. Again, we see a relationship with /r/-intrusion – affixes with higher type frequencies are more likely to attract intrusive /r/. Together, these four graphs indicate that rate of /r/-intrusion is a partial function of affix identity. Affixes which are highly parsable attract greater rates of intrusive /r/. In short, the more word-like the affix boundary is, the more likely intrusive /r/ is to occur.

Taken together, the results indicate that intrusive /r/ is affected by the individual speaker (as evidenced by the social class result), the preceding word (as evidenced by the large variation across words seen in Figure 1), and the following morpheme (Figures 2 and 3). Indeed, a stepwise binomial generalised linear model retains all three factors as significant predictors of the data ($p < 0.0001$).

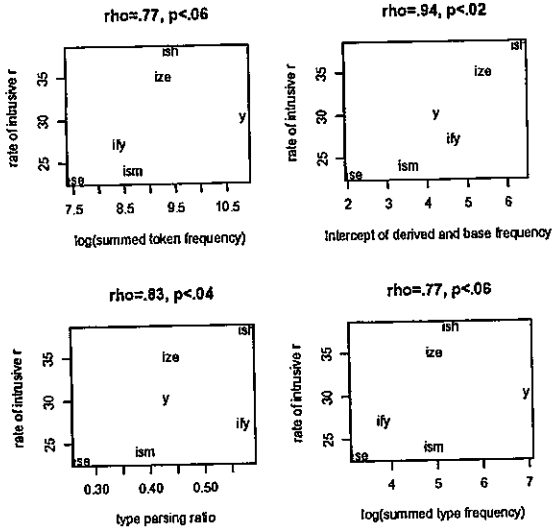


Figure 3. Rate of intrusive /r/ as it relates to affix token frequency (top left), intercept (top right) type-parsing ratio (bottom left), and type frequency (bottom right). High values of all of these measures are associated with high levels of parsability and productivity (see Hay and Baayen 2002, forthcoming, for details).

This set of results is contrary to formal phonological accounts, in that the phemonenon is both variable, and linguistically and socially conditioned. Intrusive /r/ is affected by social status, it occurs to different degrees following different words/vowels, and increases in likelihood with the increasing strength of the morphological boundary.

Experiment 2

Experiment 1 established that rates of intrusive /r/ appear to differ according to the specific affix produced. However some of the affixed words in experiment one will have been previously encountered by the participants (e.g *ploughing*), whereas others almost certainly have not (*Oprah-ese*). This may have lead to some of the variation observed. In experiment two, participants productively coined novel words based on novel base items – none of the words will have been previously encountered. This enabled us to examine the degree to which intrusive /r/ is used with nonsense words, and to further investigate possible variation across affixes.

17 students from Victoria University of Wellington were paid for their participation in a series of studies, which included the production of sentences, as in experiment 1. Each student was recorded twice, on two separate occasions – leading to 34 productions for each token. Due to a recording error, not all repetitions were captured on tape. We thus have 31 tokens per target word, with the exception of one target word (*tilpa*-ish), for which we have only 21. Tokens which were produced with an unintended affix were also omitted – though there were few such tokens.

Three trochaic base templates were created (X-*efta*, X-*ogga*, and X-*ilpa*), where "X" was represented by different initial consonants for each affix. These were cross-classified with five affixes - *-ify, -ing, -ish, y,* and *ese.*

Participants were given sentence frames, which they were asked to complete. E.g. the entire X-*efta* paradigm is given in (2). Note that stress marks are indicated on the test words, and participants were encouraged to pronounce the words with trochaic stress.

(2)

a. Helen wanted to cement and 'befta the driveway, until she discovered that her flatmate had already been cementing and _____ it all week.

b. The handbook said that you should only sleep in rooms which are humid and 'gefta, so Helen tried to humidify and _____ her room.

c. Helen usually liked talking about squirrels and about 'pefta, but today she just wanted her flatmate to be quiet. His speech was just too full of squirrel-ese and _____.

d. Helen had seen many yachts which are white and 'wefta, but this one she found unusually whiteish and _____.

e. The milkshake had the consistency of peaches and 'sefta. It was very peachy and _____.

Overall rates of /r/ intrusion for the five affixes are shown in Figure 4. The results of experiment 2 resemble experiment 1, in that there is variation across

different affixes, and that the highest rate of intrusive /r/ occurs with the inflectional affix *-ing*. However the ranking of the derivational affixes (with respect to one another) differs from experiment 1. The pattern for *–ing, -ish, -y* and *–ify* is essentially the same, but with much lower overall levels of intrusion. However *–ese* is the odd one out, showing much more intrusion in this experiment (relative to the other affixes) than in experiment 1. It is also striking that the rate of /r/-intrusion is sufficiently low that there is a complete absence of intrusive /r/ with *–ify*.

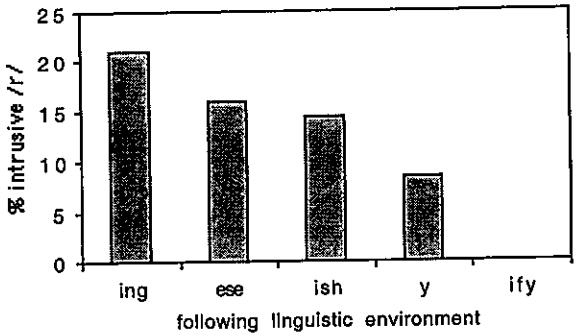


Figure 4: rates of intrusive /r/ with nonsense words, experiment 2

It seems relevant that the results of experiment one generalise over a range of different types of phonological bases. In this experiment, however, all of the bases were trochaic bases ending in a reduced centralised schwa-like vowel. In fact, if we go back and examine the results of experiment 1 more closely, we find that while *-ify* attracted intrusive /r/ at a rate of 25% overall, it only occurred at a rate of 12% when associated with *sofa* – the base in experiment one most similar phonologically to the nonsense bases used here. And while *-ify* was not the least likely affix to attract intrusive /r/ overall in experiment one, it was the least likely affix to attract intrusive /r/ when associated with *sofa*. This is because, rather than separate the two adjacent schwas with /r/, the most common strategy in this case is simply to merge them into a single vowel: e.g. *geftafy, sofafy*. *–ese*, on the other hand, was relatively more likely to occur after *sofa* (37.5% /r/ intrusion, as opposed to 23% over all the bases combined). When attached to multi-syllabic forms, *–ese* also seems to more often attract primary stress, which is likely to be a factor here. In their analysis of intrusive primary stress, which is likely to be a factor here. In their analysis of intrusive /r/ across word boundaries, Sudbury and Hay (2002) found that intrusive /r/ was significantly more likely if the following vowel was stressed.

These results regarding the differing likelihood of intrusive /r/ with different affixes across different base types highlights the fact that it is not just the phonology of the base and the nature of the following affix which condition the likelihood of intrusive /r/, but it is also the *interaction* between the base and the affix.

Experiment 3

Experiment 3 had two aims. The first was to test whether the finding (from experiment 1) that *claw* was more likely to attract intrusive /r/ than other vowels, could be definitively pinned to its phonology. Previous work has suggested that this vowel tends to be more resistant to intrusive /r/ than other vowels, and that intrusive /r/ is most common following schwa (see, e.g. Wells 1982). We therefore wanted to establish more firmly the degree to which the identity of the preceding vowel influenced the probability of intrusive /r/ in experiment 1. A second aim was to gauge the degree to which differing morphological productivity of a single phonological affix may affect the likelihood of intrusive /r/.

The affix *-able* can take two kinds of bases – nouns (e.g. *fashionable*) and verbs (e.g. *forgiveable*). It is more productive in the latter context than in the former (see Bauer and Warren 2002). Because there is some evidence that more separable affixes are more likely to attract intrusive /r/ (see experiment 1), and because separability and productivity are highly correlated (Hay and Baayen 2002, forthcoming) we hypothesized that intrusive /r/ may be more likely when *-able* was associated with a verb than with a noun. Six 'aw'-final nonsense words were chosen, together with 6 schwa-final nonsense words. Each nonsense word was paired together with both a deverbal and a denominal *-able* word, as shown in Table 3. Example sentences are shown in (3) and (4). Each participant produced both the (a) and (b) sentences, and these productions were separated by at least a week. 17 participants completed the experiment

Nonsense word	*Noun-base*	*Verb-base*
Clomfaw	Palatable	Polishable
Pontaw	Fashionable	Forgivable
Glintaw	Companionable	Interpretable
Duffaw	Seasonable	Fathomable
Brinlaw	Knowledgeable	Purchasable
Gockaw	Marriageable	Perishable
Driba	Sizeable	Washable
Dunsta	Peaceable	Quotable
Tringa	Impressionable	Extinguishable
Flockta	Miserable	Handleable
Stiga	Saleable	Eatable
Loppa	Personable	Governable

Table 3: Stimuli for experiment 3.

(3)

a. Mary refused the offer of marriage and 'gockaw. She considered herself neither marriageable nor _____.

b. Jerry was worried that the fruit would perish and 'gockaw. But luckily the fruit wasn't perishable or _____

(4)

a. The manager was a lovely person, and a friendly 'loppa. She was extremely personable and _____.

b. The country was difficult to govern and to 'loppa, because the people were not very governable or _____.

Figure 5 presents a summary of the results, showing overall rate of /r/ production for "clomfaw"-type and "loppa"-type nonsense words, primed by denominal and deverbal *-able* forms. The results provide strong support for the hypothesis that the phonology of the base word affects rate of intrusive /r/, with the *clomfaw*-type bases attracting much higher rates of intrusive /r/ than the bases ending in schwa. However the results provide no support for the hypothesis that intrusive /r/ rates may differ across the different *-ables*.

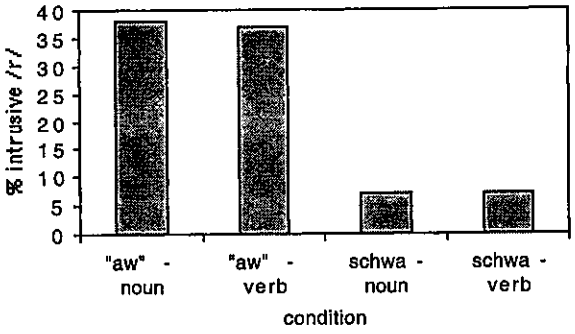


Figure 5: Experiment 3 results

One possible explanation for this is that the production of so many *–able* forms increased the salience of this affix, making all the forms in the experiment appear highly decomposed. Perhaps a large number of fillers may have facilitated the emergence of the desired effect. In addition, requiring the subjects to productively coin nonce forms with both *–able* forms may temporarily eliminate the true difference in productivity. While individuals may not regularly coin denominal *–able* forms, they are required to do so repeatedly in this experiment. Thus, if a difference in /r/-intrusion rates does indeed exist across the two forms of the affix, we may need to look to real words to document it.

A second possibility is that the correlations in Figure 3 are in fact due to chance, and that the affixal differences we have observed are due solely to differences in the phonology of the affix. Hence, in this experiment, there is no phonological difference, and so no difference in /r/-intrusion. Further research is clearly needed to follow up on this apparent correlation between affixal separability and intrusive /r/. However since the highest rates of intrusive /r/ tend to be across word boundaries, and our results show a clear pattern for inflectional *–ing* to have higher rates than derivational affixes, it seems an

intuitively reasonable result that the stronger the affixal boundary, the more likely a derivational affix would be to attract intrusive /r/.

Conclusion

The three experiments reported in this paper between them provide good evidence for both social and linguistic conditioning of intrusive /r/ in New Zealand English. Two particular points warrant more extensive research.

The first is the role of the preceding phoneme, which is clearly strongly influential (Experiments 1 and 3). It is particularly interesting that the phoneme which has previously been reported to least facilitate /r/ in fact occurs most frequently with /r/ in these data-sets (the THOUGHT vowel transcribed here as /o:/). More empirical work on both New Zealand English and other varieties will be required in order to establish the degree to which this trend is specific to New Zealand English, or whether it also extends to other varieties of English. Also worth watching is the emergence of intrusive /r/ after /æʉ/, which appears to be a relatively recent innovation in NZE.

The second result which requires more careful research is the role of the identity of the following suffix. We are not aware of any research which investigates the use of /r/ with different affixes, and no formal phonological theory predicts that different affixes should be associated with different degrees of /r/-intrusion. Yet we find clear evidence that some affixes are more likely to occur with /r/ than others. The results from our experiment 1 suggest that boundary-strength may be a relevant factor, with more separable/productive/frequent affixes leading to greater rates of /r/-intrusion. This requires substantially more data. If this is indeed the case, the theoretical implications will not be trivial.

References

Bauer, Laurie 1984. Linking /r/ in RP: Some Facts. *Journal of the International Phonetic Association* 14: 74-79.

Bauer, Laurie & Paul Warren 2002. Affix-able words: morphological productivity as the basis for (de)composition in processing. *Wellington Working Papers in Linguistics*, 14:

Bauer, Laurie & Paul Warren (submitted). New Zealand English: Phonology. In Bernd Kortmann, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider and Clive Upton (eds) *A Handbook of Varieties of English*, Vol. 1: Phonology. Berlin/New York: Mouton de Gruyter

Brown, Adam 1988. Linking, intrusive, and rhotic /r/ in pronunciation models. *Journal of the International Phonetic Association* 18(2): 144-151.

Crystal, David 1984. Should intruders keep out? In David Crystal (ed) *Who Cares about English Usage?* London, Penguin, 36-44.

Docherty, Gerard and Paul Foulkes 1999. Derby and Newcastle: instrumental phonetics and variationist studies. In Paul Foulkes & Gerard Docherty (eds) *Urban Voices: Accent Studies in the British Isles*. Arnold, London, 47-71.

Elley W.B and J.C Irving 1985. The Elley-Irving socio-economic index: 1981 census revision. *New Zealand Journal of Educational Studies* 20: 115-128.

Foulkes, Paul 1997a. Rule Inversion in a British English Dialect: A Sociolinguistic Investigation of [r]-sandhi in Newcastle upon Tyne. *University of Pennsylvania Working Papers in Linguistics*, 4.1: 259-270.

Foulkes, Paul 1997b. English [r]-sandhi: a sociolinguistic perspective. *Histoire Epistemologie, Langage* 19: 73-96

Fox , Anthony 1978. To 'r' is human? Intrusive remarks on a recent controversy. *Journal of the International Phonetic Association* 8: 72-4.

Gimson, A.C.. 1980. *An Introduction to the Pronunciation of English* (3rd Ed), London: Arnold.

Gimson, A.C. and Alan Cruttenden 1994. *An Introduction to the Pronunciation of English* (5th Ed), London: Arnold.

Harris, John 1994. *English Sound Structure*. Oxford: Blackwell.

Hay, Jennifer and Harald Baayen 2002. Parsing and Productivity. In G. Booij & van J. Marle (eds) *Yearbook of Morphology 2001*. Kluwer Academic Publishers, 203-235.

Hay, Jennifer and Harald Baayen Forthcoming. *Phonotactics, Parsing and Productivity*. Manuscript under review.

Johansson, Stig 1973. Linking and intrusive /r/ in English: a case for a more concrete phonology. *Studia Linguistica* 27: 53-68.

Jones, Daniel 1964. *An Outline of English Phonetics* (9th edition), Heffer and Sons Ltd, Cambridge.

Lewis, J. Windsor 1975. Linking /r/ in the General British pronunciation of English. *Journal of the International Phonetic Association*. 5, 37-42.

Lewis, J. Windsor 1977. The r-link business - a reply. *Journal of the International Phonetic Association* 7, 28-31.

Maclagan, Margaret, Elizabeth Gordon and Gillian Lewis 1999. Women and Sound Change: Conservative and Innovative Behaviour by the same speakers. *Language Variation and Change* 11: 19-41.

McCarthy, John 1993. A Case of Surface Constraint Violation. *Canadian Journal of Linguistics* 38: 169-195.

McMahon, April, Paul Foulkes and Laura Tollfree 1994. Gestural Representation and Lexical Phonology. *Phonology* 11: 227-316.

McMahon, April 2000. *Lexical Phonology and the History of English*. Cambridge University Press, Cambridge.

Pring, Julian T. 1976. More thoughts on the /r/-link business. *Journal of the International Phonetic Association*. 6: 92-5.

Sudbury, Andrea and Jennifer Hay 2002. The fall and rise of /r/: Rhoticity and /r/-sandhi in Early New Zealand English. In *Selected Papers from NWAV 30*, UPenn Working Papers in Linguistics 8.3, 281-295.

Tollfree, Laura 1999. South-east London English: discrete versus continuous modelling of consonantal reduction. In Paul Foulkes & Gerard Docherty (eds) *Urban Voices*. London: Arnold, 163-84.

Trudgill, Peter 1974. *The social differentiation of English in Norwich*. Cambridge University Press, Cambridge.

Vennemann, Theo 1972. Rule Inversion. *Lingua* 29: 209-242.

Wells , John C. 1982. *Accents of English* (3 Volumes). Cambridge. Cambridge University Press.

Williams, Ann & Paul Kerswill 1999. Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In Paul Foulkes & Gerard Docherty (eds) *Urban Voices*. London: Arnold, 141-162

PREVIOUS VOLUMES

CONTENTS: Volume 1, 1990
'Anahina 'Aipolo and Janet Holmes: The Wellington Tongan community: prospects for language maintenance
Nicola Daly: Sri Lankans and Sinhala language maintenance in New Zealand
Jane Pilkington: Mixed marriages and language maintenance: some Samoan data
Wendy Thompson: Attitudes to Maori and the use of Maori lexical items in English

CONTENTS: Number 2, 1990
Laurie Bauer: English vocabulary during the Twentieth century: an experimental approach
Lisa Matthewson: The negative clitic in Middle High German
Elizabeth Pearce: An analysis of negated infinitivals in Middle French
Xuezhong Zhang: The structure of Chinese dative/benefactive and double object sentences

CONTENTS: Number 3, 1991
Janet Holmes: Women's verbal contributions in public settings
Caroline McGhie: The jargon of Wellington taxi dispatchers
Catherine Neazor: Language maintenance and shift in the Wellington Polish community
Michael Williams: Post-vocalic (r) in the urban speech of the Isle of Wight

CONTENTS: Number 4, 1992
Laurie Bauer: Review of J. Kelly and J. Local *Doing Phonology*
Gavin Carr: On Heads in Morphology and Syntax
Lisa Matthewson: Prosodic Morphology in Alabama Negation
Matthew Scott: An Assimilatory Neutralization in New Zealand English

CONTENTS: Number 5, 1992
Chris Lane: Repetitive questioning in courtroom interaction and interlanguage communication
Lisa Matthewson: Talking to the TV. Conversational behaviour of television viewers
Jane Pilkington: 'Don't try and make out that I'm nice!' The different strategies women and men use when gossiping
Maria Stubbe: What's the score? Qualitative analysis in gender research.

CONTENTS: Number 6, 1994
Anita Easton: Talk and laughter in New Zealand women's and men's speech
Jen Hay: Jocular abuse in mixed gender interaction
Janet Holmes and Allan Bell: Consonant cluster reduction in New Zealand English
Camille Plimmer: Language maintenance and shift in the Italian community in Wellington

CONTENTS: Number 7, 1995
Laurie Bauer: Is the morpheme dead?
Jen Hay: The Adjacency Condition and the Atom Condition: compatible morphological constraints
Elizabeth Pearce: V-movement and optionality
Paul Warren: The sound of desert trains: delay strategies and constraints in spoken sentence processing

CONTENTS: Number 8, 1996
Jen Hay: No laughing matter: gender and humour support strategies
Gary Johnson: The management of interaction in the television interviews of Maggie Barry
Shelley Robertson: Maori English and the bus-driving listener: a study of ethnic identification And phonetic cues
Ben Taylor: Gay men, femininity and /t/ in New Zealand English. Report on Research in Progress
Janet Holmes and Chris Lane: Interactional Variation in Environmental Interviews

Book Review: Mary Crawford *Talking Difference.* Reviewed by Janet Holmes

CONTENTS: Number 9, 1997
Diane Massam: Three Faces of Niuean *aki**
Elizabeth Pearce: DP Licensing and Spec roles in Maori
Elizabeth Pearce: Genitive Case in the Maori DP
Shizuka Torii: Two Types of Evaluation Time and subject Marking in Japanese

CONTENTS: Number 10, 1998
Janet Holmes: The linguistic construction of gender identity
Jane Pilkington, Janet Holmes and Chris Lane: Gossip: its context and its boundaries
Sandra Shearn: Interviewing styles on New Zealand National Radio: is Mike Hosking really so outrageous?
Robert Sigley: Interpreting social variation using stylistic continua: the strange case of relativiser choice
Lynnette Sollitt-Morris: The paradox of silence in interaction: an indicator both of power and powerlessness

CONTENTS: Number 11, 2000
Sai Hui: Humour in a Cantonese family: an analysis in an asymmetrical small group setting
Jinnie Potter: Chartwell School: sixteen years of a Japanese-English bilingual venture
Maria Stubbe, Chris Lane, Jo Hilder, Elaine Vine, Bernadette Vine, Janet Holmes, Meredith Marra and Ann Weatherall: Multiple Discourse Analyses of a Workplace Interaction

CONTENTS: Volume 12, 2000
Angela Ford: A DP analysis of Iaai noun phrases
Elizabeth Pearce: Transitivity, Incorporation and Animacy in Iaai
Elizabeth Pearce and Samuel Ukewea Wadjeno: Uny me Cako: Iaai text
Shizuka Torii: What is tense?

CONTENTS: Volume 13, 2001
Laurie Bauer: Whither the Thesaurus
Laurie Bauer: How and Why the Phonological Word and the Morphosyntactic Word do not Coincide
Hannah Buchanan: Neutralisation of DRESS and TRAP before /l/ in New Zealand English
Yono Sukarno: vP-aspect: Aspectual Base Properties of Indonesian Predicates