Learning Words amidst Phonemic Variability

Conor Frye and Sarah Creel

Learning language is one of the most complex and intriguing feats of learning that humans achieve; but how we do it remains obscured behind many layers of questions we have yet to answer. Each small advance, however, moves us one step closer to understanding this incredible process. One aspect of this problem that has been explored substantially is word learning. In principle, understanding how a small number of words are learned in the lab could be scaled up to an entire language, giving scientists a more manageable task. If we can understand how different populations learn a small number of words under highly controlled conditions, understanding language acquisition in the real world becomes a task of exploring all of the conditions that may occur in everyday language use.

In order to approach the problem of vocabulary acquisition, many researchers have classically treated linguistic elements (speech sounds, words) as discrete symbols. If each word is made up of specific, unchanging building blocks, then word learning can be broken down into learning these building blocks and then using that small number of building blocks to store representations of words. In this account, new words are learned when new combinations of symbols are encountered. In English, for example, the sounds that differentiate the word “cat” from the word “cut” are considered different phonemes. That is, exchanging the “a” for the “uh” sound is a clue to word learners that they are hearing a new word. These symbol-like sound categories, known as phonemes, indicate differences in meaning between words (Liberman et al., 1957). Once discrete phonemes are learned, it is argued, speakers can efficiently code learned words and identify when they are hearing a new word.

Most language acquisition models describe phonemes as crystallizing into fixed categories by early childhood (Kuhl, 2000), making word learning as simple as identifying new words by their component speech sounds. However, other evidence suggests that phoneme representations are more flexible. Children can recognize words even if a constituent phoneme is violated (“fish” pronounced as “fesh”; Creel, 2012), and adults can learn labels for the same object differing minimally in vowel phonemes (e.g. ziv/zev) as readily as they learn a single label (Muench & Creel, 2013). Here we explore the role of the phoneme as the building block of word form identity, and seek to deliver a more

* Conor Frye, University of California, San Diego: cifrye@ucsd.edu; Sarah Creel, University of California, San Diego: creel@cogsci.ucsd.edu

nuanced view of phoneme perception in word learning than currently prevails in the literature.

1. The phoneme as the atom of word form identity

Many models of language conceptualize the phoneme as elemental to word identity. Best’s perceptual assimilation model (PAM; Best, 1994; Best, McRoberts, & Goddell, 2001) suggests that second language phoneme perception is driven by articulatory similarity to the learner’s first-language phonemes. If a phoneme in a second language is not found in the native language, perceptual accuracy will diminish. This approach strongly implies that the phoneme as a discrete symbolic unit of word identity. Throughout the literature, there is an assumption that phoneme category boundaries are fairly fixed in adulthood (though see Flege, 1998, for a lifelong-learning perspective). These theoretical accounts imply that learners should no longer be able to learn phonemically-identical words as different—for instance, “ba” and nasalized “ba”—which differ phonemically in French, but are not phonemically different in English. They also imply the inverse—that learners should have difficulty learning phonemically-differing words as the same word, for example, learning that puff and baff are equally-acceptable labels for a novel concept. This second implication has been much less explored than the first one, and is the subject under consideration in this work.

2. Evidence for the rigidity of phoneme categories

As the current understanding goes, when novices learn specific phonemes in a language, they are acquiring the tools to encode words and to distinguish between known and new words in that language. The language learner’s most basic and important task is then to learn the phonemes themselves—find the phonemes, find the words. Thus, a great deal of research has focused on the perceptual problem of distinguishing between highly similar phonemes. Perhaps the paradigm case of phoneme category learning is the stop consonant voicing distinction. For example, the sounds “b” as in “beach” and “p” as in “peach” are highly similar, differing mainly in subtle timing parameters. One of these parameters, called “voice onset time” or VOT, is a major contributor to the category distinction. When asked to identify English speech sounds with a range of different VOTs, English speaking adult listeners show sharply different identification near the category boundary, with VOTs up to 25ms being classified consistently as /b/, and longer VOTs being classified consistently as /p/ (Pisoni & Tash, 1974). This does not mirror the continuous range of VOTs present in the sounds they are hearing. Some have argued for developmental acquisition of such categorical boundaries (though see Eimas et al., 1974, for an
account of early availability of the VOT boundary). With a creation of a boundary, two distinct categories can be used to encode words.

Although most languages have sounds like “b” and “p”, VOT cues (and related cues) differ from language to language, meaning that it must be learned specifically for one’s native language. Infants are thought to be born with the neural capacity to discriminate nearly all phonemes in a language-universal manner (Werker & Tees, 1984), but during their first twelve months, linguistic exposure causes a pruning of perceptual abilities so that only native language phoneme categories remain (Kuhl et al., 2006). The older the learner is when a language is learned, the poorer the learner’s ability to perceive and produce distinctions between these categories (Flege, 1991). This means that non-native speakers may often confuse (for example) a “b” sound with a “p” sound, because of the inflexibility of comparison with their native categorical boundary perception.

The specific alteration of speech sound categories across languages is also well-documented. In one of the best-studied cases of difficulty learning a second-language phoneme boundary, native Japanese speakers find it notoriously difficult to distinguish an English “r” sound from an English “l” sound (Miyawaki et al., 1975). This fits with a symbol-like account of the phoneme in demonstrating that adult learners possess rigid categories for their native contrasts. This may present special difficulty when one’s first language has fewer categories than one’s second language. Spanish, for example, is considered to have five vowels (Maddieson, 1984), whereas standard American English has been noted to have 11 (Clopper et al., 2005). For native bilingual speakers and those learning a second language, this difference in category number might make moving between languages difficult, and could cause problems in new word learning if the phoneme categories do not perfectly match up. For example, if a native Spanish speaker wanted to learn an English word with a vowel that fell between two known Spanish vowels, what is the best course of action? Is it easier to treat perception of the new vowel category as a poor instance of the closest Spanish category, or to learn an entirely new set of boundaries for a new category? A symbolic account is unable to answer these questions.

3. Evidence for flexibility of phoneme categories

In contrast to studies outlined above, other evidence suggests that phonemic representation is not so rigidly defined. There is evidence for flexibility even in the most prototypical case of categorical speech perception, VOT differences. In a study asking participants to locate a picture whose label differed from another picture’s label in VOT (e.g. beach vs. peach), participants sometimes categorized long-VOT “b” sounds as /p/ and short-VOT “p” sounds as /b/—that is, their word recognition behavior showed effects of gradient VOT.
similarity (McMurray, Tanenhaus, & Aslin, 2002). This indicates that the categorical boundary that seems quite stark when tested in a phoneme identification task is actually much less categorical in practice when given natural words (McMurray et al., 2002).

In another task, adults were purposefully given multiple labels for a single object (e.g. “diy” and “dev”; Muench & Creel, 2013). Adults learned these variable-vowel labels as readily as they learned a single label (“dev”), suggesting highly-flexible phoneme boundaries, or at least graded speech sound similarity across phoneme boundaries (Muench & Creel, 2013). This means that when learning at least some labels that contain phoneme variability, adults have no more difficulty than with learning a single label. But why and when such variable-label learning might be easy is not entirely understood. And notably, in accounts where phonemes must remain categorically separate from their neighbors, this result is difficult to explain.

Even studies with non-words sharing close neighbors work to demonstrate this phonemic flexibility. Adults can be primed by similar-but-not-identical non-words; recognition of the word RED in a lexical decision task is primed equally well by the word “tomato” and the close nonword “pomato” (Marslen-Williams, Moss, & van Halen, 1996). The flexibility to phoneme variation reviewed here lends more evidence to the argument that understanding word learning through a categorical or symbolic lens is incomplete.

4. Segment type and phoneme category flexibility

A consideration relevant to questions of phoneme category flexibility is that different phoneme types (consonants vs. vowels) may differ in how malleable they are. Consonants and vowels have been described phonologically as two separate types of categories (Maddieson, 1984; Ladefoged, 2001). For instance, in a study asking participants to compare a continuum of stimuli to a standard, both consonants and vowels showed flexibility based on which type of standard was presented, but vowels demonstrated a much larger range of acceptable alteration (Macmillan, Goldberg, Braida, 1988). In a study designed to teach new vocabulary, English-speaking adults were more likely to confuse words that differed only in their vowels (such as pibo and pabu) than to confuse words that differed only in their syllable-onset consonants (such as pibo and zigo; Creel, Aslin, and Tanenhaus, 2006)—though vowel-differing words were less confusable than words differing only in coda consonants. Toro, Nespor, Mehler, and Bonatti (2008) argued on the basis of several statistical segmentation studies that segmentation is driven by consonants, whereas the structural (grammatical) information is carried by vowels. Cross-linguistic experimental evidence suggests that vowels and consonants provide speakers with different information about words as well. Testing Spanish and Dutch speakers (languages which differ greatly in their vowel to consonant ratios),
Cutler, Sebastian-Galles, Soler-Vilageliu, and Van Ooijen (2000) found that learners look to consonants for word recognition more reliably than vowels. Taken together, these data imply that, across multiple languages, onset consonants provide more robust lexical information than vowels, giving vowels more flexibility within a word, but that consonants’ advantage may depend partly on being in an early position within a word. Importantly, this also suggests that perhaps sounds that distinguish words less well (vowels) may be easier to merge, that is, they might be more flexible to input variability in a study like that in Muench and Creel (2013).

5. Comparisons with Multilingual Speakers

Another consideration when assessing phoneme category flexibility is the role of language knowledge: will knowing two or more languages give the listener greater category flexibility? According to a 2006 European Commission survey, the majority of the world (~56%) speaks more than one language. Early bilingualism has been associated with facilitation of phonological processing, possibly leading to better word learning (Bialystok et al., 2003). This improved phonological ability has been linked to phonological relationships between languages (see Bialystok et al., 2003; Bialystok et al., 2005). In word learning tasks, bilinguals were asked to learn paired associations between a novel word and a known word, and outperformed monolinguals (Papagno & Vallar, 1995; Van Hell & Mahn, 1997). Another study found that bilinguals excelled in word learning tasks when taught words with sounds novel to any language they knew (Kaushanskaya & Marian, 2009). Exploring how bilinguals will respond when presented with phoneme variability will provide insight into how phonemes are learned in general, as well as how new words and languages are perceived when they are presented to learners with different language backgrounds.

6. Open Questions

To sum up, current research leaves questions about the extent to which phonemes are the atoms of word form identity; to what extent flexibility in phonemes is dependent on the type and position of the phoneme, and whether differences between languages change these relationships.

We know there is some flexibility in phonemes for word form identity, but the extent to which this is the case is unclear. It remains to be seen if all types of phonemes are flexible in the same manner. To what extent are specific types of phonemes (vowels or consonants) flexible within word recognition?

It is also clear that different languages divide up the acoustic-phonetic space differently when giving meaning to word forms. However, it is unknown if this difference is driven by basic properties of the language, such as the number of vowels and consonants in a particular language.
The current study tests the hypothesis that categorical use of phonemes is not rigidly defined in adult brains and behavior, but can be modified with minimal training. Our flexible lexical probability hypothesis predicts that phonemic categories are much more plastic and can be modified and merged: exposure to different probability distributions should alter how phonemes are used in word recognition. To explore the idea that word identity is probabilistically defined, we teach participants novel vocabularies with various “collapsed” phonemic differences in order to gauge the malleability of word-form representations at a behavioral level.

7. Method
7.1. Participants

We tested 80 normal-hearing adults recruited from the UC San Diego SONA subject pool. Forty were native English speakers (avg. age: 21.1 years, sd: 1.9) who grew up hearing no other languages and reporting fluency in no other language but English. The remaining 40 were bilingual Spanish-English speakers (avg. age: 21.1, sd=2.9) who grew up hearing Spanish from birth, and English either at birth or sometime later. Age of English acquisition ranged from 0-14 years, and the average age was 5.0 years, sd=3.0. All bilingual speakers considered themselves fluent in both languages.

7.2. Procedure

The study consisted of vocabulary training with feedback, and vocabulary testing (no feedback). Participants learned labels for 16 novel objects over 256 trials using a two alternative forced choice task (2AFC). On each trial, participants saw two novel objects on a screen, and heard a recorded word. They chose one of the objects by clicking, and received accuracy feedback before advancing to the next learning trial. In the vowel merge condition, labels for specific objects varied in their vowel phonemes (e.g. the same object was labeled both “teev” and “tiv”). In the onset merge condition, the onset consonant was variable (e.g. “div” and “tiv”)—varying in exactly the voicing contrast that is the paradigm case of phoneme categorization. In the dissimilar (difficult) condition, all phonemes except the coda varied between the two labels for a single object (e.g. “buv” and “div”). In two control conditions (matched to the vowel-merge and consonant-merge respectively), each picture received only one label.

Each participant was tested (128 trials) on the trained word-meaning pairings in the same 2AFC task as during training, except that there was no feedback. Accuracy and reaction times were measured. While training trials only depicted dissimilar-sounding words on a given trial, the test trials contained four different levels of difficulty. In close-competitor trials, the target and the
competitor share two of the three phonemes in the word (target: /bIʃ/-/biʃ/ competitor: /pIʃ/-/piʃ/; note that for the consonant-merge and consonant-control conditions, target and competitor shared onsets and differed in vowels, e.g., target: /bIʃ/-/pIʃ/ competitor: /biʃ/-/piʃ/). In same-vowel and same-consonant trials, the target and competitor shared only the vowel or the onset consonant, respectively (target: /bIʃ/-/biʃ/ competitor: /dIv/-/div/; target: /bIʃ/-/biʃ/ competitor: /buv/-/bUv/). Finally, in the far trials, only the coda is shared (target: /bIʃ/-/biʃ/ competitor: /fʌʃ/-/faʃ/).

8. Results

An ANOVA was conducted on test accuracy, with Merge Condition (Onset Merge, Vowel Merge, control conditions and dissimilar label conditions) and Language Background (Spanish-English Bilingual and English Monolingual) as between-subjects factors, and Test Trial Type (close competitor, same onset competitor, same vowel competitor, and far competitor) as a within-subjects factor. A main effect of language background (F(1,78)=8.13, p<.006) resulted from English monolinguals outperforming Spanish-English bilinguals overall. Although this runs counter to findings of bilingual advantages in word learning, it is similar to Muench and Creel’s (2013) finding that greater English dominance predicted better learning, perhaps because our materials, like Muench and Creel’s, contain English phonology/subphonetic cues. However, Language Background did not interact with any other factors, so for all additional analyses we collapsed across this factor.

There was a main effect of Merge Condition (F(4,78)=17.61, p<.001). Specifically, all conditions showed higher test trial accuracy than the dissimilar-labels condition (t(30)=6.00, p<.001 (Onset Change condition to dissimilar-labels condition); t(30)=9.53, p<.001 (Onset Control condition to dissimilar-labels condition); t(30)=7.04, p<.001 (Vowel Change condition to dissimilar-labels condition); t(30)=9.55, p<.001 (Vowel Control condition to dissimilar-labels condition). The relationships between merge conditions are shown in Figure 1. To explore patterns of test trial type and to compare each merge condition to its corresponding control condition, we dropped the dissimilar-labels condition from the data set and computed a different ANOVA, with Merge Condition (merged, control), Segment Type (consonant, vowel), and Trial Type as factors.
In this second ANOVA, there was an effect of Merge Condition, such that merged participants were overall less accurate than control participants ($F(1,78)=5.90$, $p=.02$). There was also a main effect of Trial Type ($F(3,78)=77.55$, $p<.001$). This effect resulted from close competitor trials (where the two pictures’ labels shared two out of three segments) being significantly less accurate than any other trial type (same onset competitors: $t(63)=3.78, p<.001$; same vowel competitors: $t(63)=6.79, p<.001$; far competitors: $t(63)=8.31, p<.001$). Same consonant and same vowel trial accuracies differed as well ($t(63)=, p=.02$). See Figure 2 for this relationship between trial types.

There were also two two-way interactions, one between segment type and trial type ($F(3, 78)=7.90, p<.001$) and another between merge condition and trial type ($F(3, 78)=3.34, p=.02$). To explore the first interaction, we compared trial types for each segment-merge type (consonants and vowels). Here, the only trial type that was significantly different was the close competitor trials, where competitors shared two out of three phonemes with the target. Performance on close competitors in the vowel condition was significantly more accurate than performance on close competitors in the consonant condition ($t(62)=2.57, p=.01$). Recalling the earlier discussion of the greater role of onset consonants in word identity than vowels, this difference likely stems from the fact that close competitors in the vowel conditions (both vowel-merge and vowel-control) differed in their onset consonants (div/div’s competitor was tlv/tlv), while in the onset-consonant conditions, the onset consonant was the same and the vowel differed (div/tiv’s competitor was dlv/tlv).

Figure 1. Accuracy in each merge condition, collapsed across language background. Dissimilar label condition was different from every other condition, $p<.001$. 
Figure 2. Close trials were significantly different from all other conditions (\(p<.001\)), same consonant trials were significantly different from same vowel trials (\(p=.02\)). The dissimilar-label condition is not included here.

We also explored the significant interaction between merge condition (merged vocabularies or control vocabularies) and trial type (\(F(3,78)=3.34, p=.02\)). The nature of the interaction was such that, while merge participants showed numerically lower accuracy than control participants across the board, only two test trial types showed a significant disadvantage for merge participants: same onset competitor trials (\(t(62)=3.70, p<.001\)) and far competitor trials where competitors shared coda consonants (\(t(62)=2.79, p=.007\)). The fact that trials with the least phonological competition (far trials) were affected suggests that difficulty may have stemmed from learning merged words themselves, rather than from phonological competition.

No other interactions reached significance, including the interaction between Merge Condition and Segment Type. This suggests that the magnitude of the effects in the vowel merge and the consonant merge conditions did not differ.

9. Discussion

Previous research showed that adults are capable of learning phonemically-variable vocabularies (Muench & Creel, 2013). Based on that work, we expected the dissimilar labels condition to be the most difficult, and the single-label control conditions to be the easiest, which was the case. The question was where the vowel merge and onset merge conditions would fall.
Although there was not a clear effect of segment type, the findings here suggest that learning words with variable onsets or variable vowels, while slightly more difficult than learning non-varying words, imposes limited difficulty relative to learning dissimilar labels. This implies there is a gradient of phoneme flexibility in word recognition and learning, with single-feature variation adding mild difficulty to the learning process.

The effect of language background on overall accuracy was inconsistent with some previous findings, but consistent with others. For instance, previous research indicated that knowing a second language might convey some benefits for word learning (Kaushanskaya & Marian, 2009). Were this the case, then in this word-learning task bilingual speakers should have performed better. However, Spanish-English bilinguals uniformly performed less accurately. This could be explained by cognitive factors unrelated to cross-linguistic phoneme perception, or possibly, as Muench and Creel (2013) suggested, lower familiarity with English phonology—the phonology of the to-be-learned words. However, it could also be related to greater sensitivity to within-category variation because Spanish has fewer vowels than English. That is, having vowel categories in Spanish that span more of the possible vowel space might make Spanish speakers more sensitive to within-category variability. If this is the case, having fewer vowels as legal phonemes in a language might provide greater focus on the ideal version of the phoneme, though we should not see this pattern for consonants. This would cause the Spanish-English bilinguals to be more practiced at rejecting phonemes that do not conform to the more specialized forms they have learned. Further experimentation is needed to elucidate the reason behind the data pattern.

Our findings contribute to research on word learning in two respects. First, results add nuance to the ongoing debate regarding bilingual advantages in word learning: not every learning task benefits from knowledge of multiple languages and their contrasting sound categories, and in fact some interactions between languages might make learning specific sounds more difficult. Second, results suggest impressive flexibility in adult word learning: not only variability in vowels—generally agreed to be flexibly perceived—but also, to some extent, variability in consonants differing categorically in voicing, can be learned. This hints at greater plasticity in L2 (and perhaps) L1 acquisition than previously thought.

References

Best, Catherine T., McRoberts, Gerald W., & Goodell, Elizabeth, (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s


