Differences between written and spoken input in learning new words

Jessica R. Nelson, Michal Balass, and Charles A. Perfetti
University of Pittsburgh, Learning Research & Development Center

We trained adult learners the meanings of rare words to test hypotheses about modality effects in learning word forms. These hypotheses are that (1) written (orthographic) training leads to a better representation of word form than phonological training, that (2) recognition memory for a word is partly dependent upon congruence between training and testing modality (written vs. spoken) but that (3) skilled learners are less dependent on the episodic context of training than are less skilled readers. These hypotheses were confirmed by results of a word recognition test following form-meaning training. We discuss these results in terms of an episodic account of word learning (Reichle & Perfetti, 2003) and variations in lexical quality (Perfetti & Hart, 2001) that can arise through differences in code generation during learning.

Differences between print and speech in learning new words

Learning new word forms and word meanings arises from a variety of experiences with words. Much word learning involves mapping new word forms onto existing spoken word representations; other learning involves “translating” a second language form to a first language form; still other word learning is more or less de-novo, when a previously unfamiliar word is encountered for the first time in a context that provides a meaning. In this case, the learner establishes a new word representation of variable specification of form and meaning, and one of the factors that may influence this variability is the episodic context of the word encounter, including whether it is read or spoken. This is the problem we address here: How are word forms learned from experiences with unfamiliar words, and how this is affected by the modality (written or spoken) of the word learning event?

Written Language & Literacy 8:2 (2005), **25–44**.
ISSN 1387–6732 / E-ISSN 1570–6001

When an adult encounters a new word in a meaningful context, several aspects of the encounter are established in memory, and an episodic trace of the experience is formed. Such a trace is detectible in ERPs, which can differentiate between a recently learned word and a familiar or unfamiliar word as early as 140 ms, with an additional later component between 400 and 600 ms that is also sensitive to comprehension skill. The later component is more pronounced in skilled comprehenders, indicating that skilled readers are more effective in establishing strong episodic traces for new words (Perfetti, Wlotko, & Hart, 2005). With repeated exposures to a new word, and multiple episodic traces, a reader should be able to establish a new lexical entry, more independent from the individual episodic traces, complete with semantic, phonological, and orthographic information. Reichle and Perfetti (2003) demonstrated that a lexicon that is built with such repeated episodic traces of co-occurring form and meaning information has the potential to account for not only basic frequency and regularity effects, but also to capture the emergence of inferential and derivational morphology.

While the model outlined by Reichle and Perfetti (2003) includes repeated traces of the *integrated* orthography, phonology, and meaning of words in the lexicon, readers often encounter a new word *either* visually *or* auditorily in a meaning context. Thus, an episodic trace of such an encounter is likely to include context-specific information such as visual or acoustic input features in addition to more context-independent information such as orthography or phonology. The orthographic and phonological traces are strengthened as they are repeated over many encounters, eventually creating the kind of unified traces required for an abstracted lexical entry, while more context-specific aspects of individual traces will not be strengthened with variable encounters with the word.

To the extent that future encounters with a recently learned word match the type of knowledge that has already been established in memory for that word, a reader should be able to more readily retrieve the previous memory traces or abstracted lexical knowledge. Thus, if a reader learns a new word via a visual encounter, that reader should more readily access the established memory for that word if they see the word again than if they hear it. Likewise, if they learn a new word via an auditory encounter, hearing the word again should facilitate recognition more than seeing the word. However, the ease with which the memory for the word can be accessed will also be dependent on the strength and quality of the initial traces, and this may be dependent on the input modality.

There are four main factors to consider concerning the encoding of a word in a written or speech modality: for a written word, the quality of the orthographic traces; for a spoken word; the quality of the phonological traces; and for both written and spoken words, the quality of cross-modal traces (if any) that might be generated as part of the encoding; i..e., phonological codes generated when reading and orthographic codes generated when hearing a word. Asymmetries in these factors would provide the basis for representational quality differences given different input modalities. We will describe evidence showing that (1) orthographic inputs have a memory advantage over phonological inputs, and (2) phonological recoding of orthography happens more readily than orthographic recoding of a phonological stimulus. Because of this, we expect an advantage for orthographic encounters, in terms of both strength and flexibility (across modality) of memory traces.

The hypothesis that orthographic inputs lead to better retention than phonological inputs has support from other research (Dean, Yekovich, and Gray, 1988; Gallo, McDermott, Percer, and Roediger, 2001). Dean et al (19998) found an overall superiority for visually trained words over auditorily trained words in later recognition, regardless of testing modality. Gallo et al (2001) found a modality effect for false memories: A false memory for a probe that was related semantically to a list of words was greater for a visual probe to an auditory list than for an auditory probe to a visual list. Gallo et al. discuss several possibilities for this modality effect, including, for example, that visual recollections may be more distinct than phonological ones. These studies suggest that words presented in the visual inputs establish more accessible traces in memory than words presented auditorily.

In the more specific case of reading, even children who do not know how to read are able to take advantage of visual cues in print (e.g. word length, initial letters) to temporarily learn associations between a printed word and its meaning (Gough, Juel, & Griffith, 1992). Attention to just a few beginning letters might also characterize adult learning of unfamiliar words.

An additional perspective on the visual-orthographic advantage comes from the possibility that phonological representations have variable quality. Specific problems in representing phonology have been attributed to dyslexics (Elbro et al., 1994; Elbro & Jensen, 2005; Snowling, et al, 1986). Elbro & Jensen (2005) summarize the argument that dyslexics tend to have underspecified phonological representations, which, for example, leads to the production of non-distinct vowels. Although this idea of underspecified phonology has been applied only to dyslexics, it might capture a general factor that applies to even skilled readers when they are required to encode rare and phonologically complex. Their representations may be less stable or less well-specified than they would be for well known words.

In alphabetic reading, an important factor is that the mapping between orthography and phonology allows a memory representation that links visual-graphic with auditory-phonological information. A reading event activates both of these and results in a memory trace that includes both. Indeed, phonological information is activated routinely during the visual presentation of words and pseudowords (e.g. Naish, 1980; Perfetti, Bell, & Delaney, 1988; Perfetti & Bell, 1991; Van Orden, Johnston, & Hale, 1998; Lukatela & Turvy, 1994 a&b; Lukatela, Frost, and Turvey, 1999; Unsworth & Pexman, 2003). If a visual encounter with an unknown word also activates pronunciations, the memory trace of this encounter may include both orthographic and phonological information. This seems likely, given that phonology is activated early, automatically, and pre-lexically for known words.

A similar cross modal code generation can occur for spoken words, as orthographic information is activated by phonological inputs (e.g. Seidenberg & Tanenhaus, 1979; Ziegler & Ferrand, 1998; Slowiaczek, Soltano, Wieting, & Bishop, 2003; Castles, Holmes, Neath, and Kinoshita, 2003). However, it likely the generation of orthography from phonology is less reliable, less automatic, and perhaps more lexically-dependent than is the generation of phonology from orthography. Thus the orthography-to-phonology coding may occur more generally than the phonology-to-orthography coding (Borowsky, Owen, & Fonos, 1999).

Applied to new word learning (for words without spoken lexical entries), these cross-modal codings are, in effect, pre-lexical. Johnston, McKague, and Pratt (2004) showed that when orthographic representations were formed from auditory presentations of new words, the orthographic representations were underspecified compared to known words. If we relate the generation of phonology and the generation of orthography to reading and spelling, respectively, we would assume that the latter would be more difficult because spelling is more difficult, requiring retrieval of a completely specified and accurate orthographic form (Perfetti, 1997). Alternatively, reading requires only a recognition/discrimination process that is sufficient to select the read word from other words. Studies show better performance on reading tasks than on spelling tasks across languages. Bosman & Van Orden (1997) argue this results from more consistent letter-phoneme relations than phoneme-letter relations, and certainly English has notoriously inconsistent letter-phoneme mappings. However, this pattern of behavior holds even in Italian, which has highly consistent phonology-spelling mappings (Cossu et al., 1995). Whereas reading builds on spoken language from its beginning, speech perception is established pre-literately and, in principle, does not strictly require orthography even after literacy is established. Additionally, competent readers are likely to have had far more practice reading than spelling.

Evidence for such a lack of symmetry between decoding and orthographic recoding of phonological input comes from an implicit memory study in which readers were primed either visually or auditorily for stem completion tasks, which were also given in both modalities (Berry, Banbury, & Henry, 1997). Visual presentation primed both visual and auditory stem completion tasks, whereas auditory presentation primed only the auditory stem completion task. Berry et al. attributed this to phonological recoding of the visual stimuli and supported the claim with a second experiment that used articulatory suppression to eliminate the visual priming of auditory stem completion without affecting visual priming of visual stem completion.

The asymmetry between visual and auditory word processing appears to have two closely related components relevant for the quality of word representations. First, learning visually presented words may be easier than learning spoken words. Second, the visual presentation of a word is more likely to lead to a compound memory trace that contains both orthographic and phonological components than is a spoken presentation. And the second may explain the first. Baddeley, Papagno, and Vallar (1988) found that subjects were faster to learn word-nonword pairs visually than auditorily. They pointed out that "it is possible that memory for nonwords in our study might be enhanced by visual presentation because this allows both a phonological and a visual encoding of the unfamiliar material" (p. 591).

If the formation of memory traces from word encounters reflects modality asymmetries, we can make some predictions about how vocabulary is established when new, rare words are encountered in a single modality. First, literate adults should be faster to learn the meaning of the word if the word is learned though print rather than through speech. Although there are several possible reasons for such a prediction — greater distinctiveness for visual stimuli, less specificity in phonological representations — we assume that one important factor is the modality asymmetry: that a memory representation for a written word is more likely to contain cross-modality information. Second, recognition of learned words should be better if the word is tested in the same modality in which it was learned. This is because there is more overlap between the memory traces laid down during learning and the test item, making a match resonate from memory more strongly. Finally, the speed of learning, as well as the ability to extract more abstract word representations that are less dependent on episodic factors, will likely vary according to individual differences. The ability to form strong memory traces, the ability to extract higher level orthographic and phonological representations from graphic and auditory input, and the ability to decode probably all play a role in an individual's ability to learn and later recognize new words.

To test these hypotheses, we trained participants on the word meanings of a collection of very rare words. The words were presented either visually or auditorily and were learned to criterion. We then tested recognition for the items when they were presented either in the same modality as training or in the opposite modality. In this way, we were able to measure learning speed, in terms of trials to criterion, as well as recognition accuracy.

Methods

From a pool of approximately 500 undergraduate students from the University of Pittsburgh who completed a battery of tests that assess comprehension ability and vocabulary knowledge for course credit, 35 were recruited to participate for compensation. They were chosen to represent a range of comprehension skill, as assessed by the Nelson-Denny comprehension test (Nelson & Denny, 1973). Each subject was paid $21 for participating in the three hour study. Subjects knew that they would receive the same amount, even if they completed the training portion of the study in less than the expected 2.5 hours. This acted as incentive to work quickly to achieve the goal. The study was composed of two portions: a training portion, in which participants learned rare words, and a testing portion, in which participants were probed for their memory of the words.

Materials

A set of 105 rare words were used for training and 35 additional words were used as rare word foils for testing. These words were all very rare, low frequency words. Only 45 of the 141 words appeared in the English Lexicon Project database (Balota et al., 2002), and 15 of those were not in the Kucera-Francis database (Kucera, & Francis, 1967). The average Kucera-Francis frequency for the remaining 30 words was 1.3, with the highest frequency word being "bivouac" with 5 occurances per million. The mean length of trained words was 7.0 letters (range 4–9) and 2.5 syllables (range 2–3). The rare word foils had a mean word length of 7.4 characters and 2.6 syllables, although they were more variable, ranging from 4 to 11 characters and 1 to 5 syllables. Stimuli were displayed and responses recorded using E-Prime software from Psychological Software Tools.

Training phase

Participants were trained to a criterion of 100% correct. Training ended when this criterion was reached or after 2.5 hours if the criterion had not been reached by then. All participants proceeded to the assessment and testing session after their training ended. The participants were trained on 35 different rare words in each of three training conditions, for a total of 105 words. All participants were in all three conditions, making condition a within-participants variable. The order of the training conditions was counterbalanced across subjects and remained constant throughout the experiment. Words within training conditions were randomized to eliminate order effects. The training was self-paced so that the participant could spend as much time on any word as needed. The training conditions were as follows:

orthography/meaning: In this condition subjects were required to learn the meaning of each word given its written form. They saw the word presented along with its definition and an example sentence during training.

phonology/meaning: In this condition participants were required to learn the meaning of each word given its pronunciation. They heard the word presented along with its written definition and a written example sentence with the target word replaced by a blank space. They were able to hear the pronunciation of the word as many times as needed by pressing a key.

orthography/phonology: In this condition participants were required to learn the pronunciation of each word given its written form. They saw and heard the word during training. They were able to hear the pronunciation of the word as many times as needed by pressing a key. This condition was included for the purposes of another study, and data from this portion will not be discussed here.

Initial Familiarity Check

Although the words were very rare, to assess prior familiarity, the training session began with a familiarity check. Participants proceeded once through each block of 35 words within each condition, responding for each word whether it was familiar by pressing 1 (familiar) or 2 (unfamiliar). If the familiarity response was made within 4 seconds, feedback appropriate for the condition was given: the word's definition (in the orthography/meaning and phonology/meaning conditions) or pronunciation (in the orthography/phonology condition), and the participants were able to study each word as long as they wanted. If 4 seconds passed without a response, the next item was given, and the participant did not have a chance to study the word. Participants were reminded of the instructions at the beginning of each training condition block and proceeded through the entire list once at their own pace.

Reaching criterion

After the familiarity check was completed, participants proceeded through the list of words, which remained blocked by training condition, with the experimenter. Words within each training condition were presented in random order. Subjects either saw or heard the word, and they were required to give the appropriate definition or pronunciation, depending on the condition, aloud. The experimenter judged whether the definition or pronunciation was correct and keyed in a response. This prompted a feedback screen telling the participant whether they were correct or incorrect along with the correct answer. Regardless of whether the participant was correct, they could study the correct answer again for as long as they needed before proceeding to the next word.

The participant and experimenter proceeded through the list of words in this manner as many times as needed for the subjects to learn 100% of the words. When a participant gave the correct definition or pronunciation for a given word twice in a row, the word was considered learned and was removed from the list. That word did not appear again in the remainder of the training portion. Criterion was reached when participants correctly answered 100% of the words twice in a row so that no words remained in the list. Thus, each word was encountered at least three times: once during the familiarity check, and at least twice subsequent to reach criterion.

Testing phase

In the testing phase, participants were presented with a series of words. One third were the words from the orthography/meaning condition, one third were the words from the phonology/meaning condition, and one third of were rare words that had not been presented previously (rare foils). Half of the words from each condition were presented in spoken form, and half were presented visually. This created four conditions for learned words: O-O (words learned orthographically and tested orthographically), P-P (words learned phonologically and tested phonologically), O-P (words learned orthographically and tested phonologically), and P-O (words learned phonologically and tested orthographically). Rare foils were tested half orthographically and half phonologically. The words were presented randomly during the testing portion, intermixing those presented auditorily and those presented visually.

The task was to recall whether each word had been seen, heard, or not presented (in the case of the untrained foils) during the training phase. Participants were asked to rate their certainty using a 7 point scale, anchored by the visual modality at the low end and the auditory modality at the high end: 1, "I definitely saw this word", and 7 "I definitely heard this word". Intermediate scale positions were associated with less certainty: 2 "I think that I saw this word," and 3, "I may have seen this word," 6, "I think that I heard this word," and 5, "I may have heard this word." The mid-point of the scale, 4, was the judgment that the word has not been presented in training: "I neither saw nor heard this word". It was emphasized that they should judge based on the modality in which they were given the word in the training phase of the experiment ignoring its form at test. Although the test list was presented a second time in a new random order with each word given in the other test modality, only the data from the first list are considered to be sure we have memory judgments that are uncontaminated with the modality of a recent probe experience. The testing portion was untimed, but was generally completed quite rapidly.

Results

Word learning

The main result for word learning was faster learning for orthographic than phonological presentations. Thirty of the thirty-five participants required fewer trials to reach criterion in the orthography-meaning condition than in the phonology-meaning condition. A paired t-test of the mean number of trials to criterion showed that reliably fewer trials were required to learn words in the orthography-meaning condition ($M = 3.44$, $SD = 1.15$) than in the phonology-meaning condition ($M = 4.04$, $SD = 1.48$), $t(34) = -4.24$, $p < .01$ (two-tailed).

Recognition accuracy

The scale used in the testing phase of the experiment was broken down into three measures. First, old/new recognition accuracy was assessed by scoring whether the response was "4" on the scale ("I have neither seen nor heard this word") or any other point on the scale. For "4" the word was considered marked "new," and for all other responses, the word was considered marked "old." Our second measure was modality accuracy. For those words correctly identified as "old," we determined whether the participant thought they had seen or heard the word in training. If the answer was "1," "2," or "3" on the scale, we marked the answer as "saw." If the answer was "5," "6," or "7" on the scale, the answer was marked as "heard." This was then compared against the actual modality in which the word was learned to assess accuracy. Our final measure was confidence. For this measure, we transformed the scale so that the highest confidence rating for either modality (each end of the scale) was given a "1" and the lowest was assigned a "3." Given that the word was correctly identified as old, and that the modality was correctly identified, the confidence in the memory was determined using this 3-point scale.

Old/new recognition accuracy

To test memory for word form, we calculated *A'* as a measure of memory strength controlled for response bias, as indicated by false recognition of new items given in the same test modality.[1] This analysis disregarded the modality judgment, and was instead based only the participant's judgment that the word had occurred in the training phase of the study.[2] Table 1 shows the hit rates and false alarm rates, and Figure 1 shows the results of the *A'* analysis. The main

Table 1. Old/New accuracy

Training	**Orthography**		**Phonology**		**Untrained (false alarms)**	
Testing	See	Hear	See	Hear	See	Hear
Identified "Old"	98.99%	79.37%	85.87%	98.32%	18.89%	14.76%

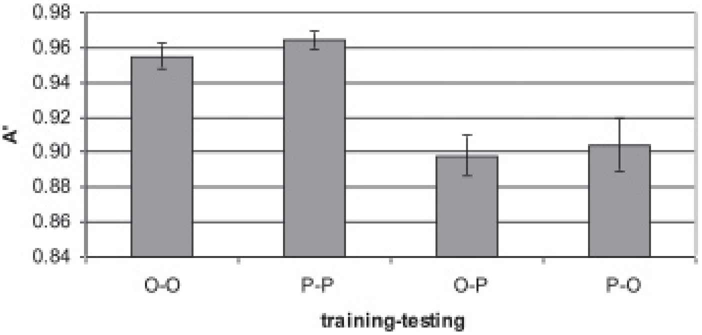

Figure 1. Mean *A'* ± one standard error for words trained in the orthography-meaning and phonology-meaning training conditions, tested either in the same modality or different modality.

result is that memory that the word had been presented in the training phase was more accurate when the modality at test matched the modality at training. This conclusion is supported by a two-way repeated ANOVA of the *A*' s, with Training Modality (O and P) and Test Modality Congruence (Same Modality and Different Modality) as the two factors. Words tested in the incongruent modality showed lower *A*' s ($A'=.901$, $SE=.01$) than words tested in the congruent modality ($A'=.96$, $SE=.01$), $F(1,33)=51.05$, $MSE=.002$, $p<.001$. There was no effect of training modality, and no reliable interaction between training modality and testing congruency.

Modality accuracy

To assess the accuracy of participants' modality judgments, we scored each item that was correctly identified as "old" according to whether the training modality was correctly identified. A two-way repeated measures ANOVA parallel to the one for word form accuracy showed a main effect of Training Modality: the modality of orthographically trained words ($M=.95$, $SE=.01$) was more accurately recalled than the modality of phonologically trained words ($M=.90$, $SE=.01$), $F(1, 33)=11.253$, $p<.01$. In addition, those words tested in the congruent modality had a reliably higher modality accuracy ($M=.95$, $SE=.01$) than those tested in the incongruent modality ($M=.91$, $SE=.02$), $F(1,33)=5.306$, $p<.05$. There was a marginally significant interaction between training modality and testing congruency, suggesting that orthographically trained words were more affected by a change in modality at test, $F(1,33)=3.667$, $p=.064$. The means (with standard errors in parentheses) of the orthography-same (O-O), orthography-different (O-P), phonology-same (P-P), and phonology-different (P-O) conditions were, respectively, .98(.01), .92(.02), .91(.01), and .90(.02).

Confidence ratings

To test participants' confidence in their recognition using the full 7-point rating scale, we transformed the scale so that the highest confidence for either modality was given a "1" and the lowest was assigned a "3". A two-way repeated measures ANOVA for words that were correctly recognized in the correct modality showed a significant effect of Test Congruency: words presented in the same modality at test were more confidently rated ($M=1.11$, $SE=.04$) than words presented in the opposite modality ($M=1.3$, $SE=.05$), $F(1,33)=42.59$, $MSE=.03$, $p<.001$. There was also a trend toward orthographically trained words ($M = 1.17$, $SE = .04$) being rated more confidently than phonologically trained words ($M = 1.23$, $SE = .04$), $F(1,33) = 3.22$, $p = .08$.

New words

For new words, we can define a false alarm rate as the percentage of participants' judgments that fell into some category other than "neither saw nor heard". As can be seen in Figure 2, most participants made fewer than 4 false alarms, with 18.9% false alarms in the visual modality and 14.8% false alarms in the auditory modality. Test Modality did not affect the rate of false alarms: there was no reliable difference in a paired-samples t-test. However, when false alarms did occur, Test Modality reliably affected the modality rating. New words tested visually were rated as being learned auditorily (88.1% marked as learned auditorily), and words presented auditorily were rated as being learned visually (61.3% marked as learned visually). For the 22 subjects who made false recognition errors to at least 1 visually tested and 1 auditorily tested new item, we carried out a paired samples *t*-test on the mean ratings. This test verified that the tendency to mark falsely recognized new words as having been learned in the modality opposite the test modality is significant, $t(21) = -3.29$, $p < .01$ (two-tailed). Thus, when participants judged a new word to have been presented earlier — presumably a weaker memory trace (noise rather than signal) than for a word actually presented — they inferred that they must have learned the word in the alternative modality. The assumption that a false memory is subjectively a weaker memory was confirmed by a paired samples *t*-test on the 3-point confidence scale showing that the untrained words ($M = 2.13$, $SE = .12$)

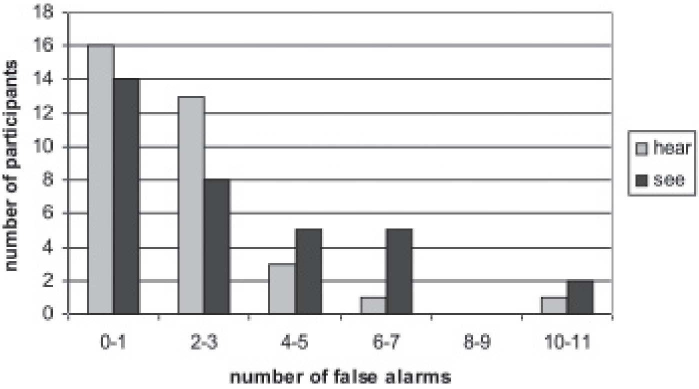

Figure 2. Histogram of the number of false alarms made when untrained words were presented at test either visually or auditorily.

were rated with far less confidence than the trained words, ($M = 1.19$, $SE = .04$), $t(33) = -8.9$, $p < .001$.

When a new word was heard at test and mistakenly identified as old, participants were about one and a half times more likely to judge that they had previous seen it than heard it. When a new word was seen at test and mistakenly identified as old, participants were nearly seven and a half times more likely to say they had heard the word in training than that they had seen the word in training.

Individual differences

To determine whether individual differences in the rate of learning were related to the memory results, we assigned each participant a learning speed measure, the average number of trials to criterion in the meaning training conditions (orthography-meaning and phonology-meaning).[3] This trials-to-criterion measure was reliably correlated with old/new recognition performance (in terms of *A*') for both different-modality conditions: for O-P, $r = -.526$, $p = .001$; for P-O, $r = -.490$, $p = .003$. Trials-to-criterion were not reliably correlated with either of the same-modality conditions.

We can calculate the dependency of recognition on modality by subtracting accuracy for different-modality conditions from the same-modality conditions. A large difference means a large dependence on retrieval of the initial modality information. Our measure of learning speed was significantly correlated with the difference in *A*' between old/new judgments in the same-modality and the different-modality testing conditions, $r = .62$, $p < .001$.

Learning speed was also significantly correlated with the difference between modality accuracy for same-modality and different-modality conditions. Again, this indicates that there are individual differences in the dependence on encountering a word in the learned modality for accurate recognition, $r = .50$, $p < .01$. Learning speed was also significantly correlated with confidence: fast learners were more confident in their recognition. This was true in all conditions: O-O, $r = .539$, $p = .001$; P-P, $r = .583$, $p < .001$; O-P, $r = .655$, $p < .001$, and P-O, $r = .529$, $p = .001$. Finally, learning speed was also significantly correlated with the dependence of confidence on modality, meaning that slower learners showed a bigger difference in confidence between same-modality conditions and different-modality conditions. Dependence on modality was calculated by subtracting confidence ratings for the same-modality test items from the different-modality test items. $r = .371$, $p < .05$. For a summary of all reliable correlations with trials-to-criterion, see Table 2.

Table 2. Individual differences: correlations with trials-to-criterion

measure	*r*	*p*
old/new recognition (A')		
O-P	−.526	.001
P-O	−.490	.003
confidence		
O-O	.539	.001
P-P	.538	.001
O-P	.655	.001
P-O	.529	.001
Modality Dependence (same-modality — different-modality)	*r*	*p*
old/new recognition A'	.620	< .001
modality accuracy	.500	< .01
confidence	.371	< .05

Discussion

The main findings of this study confirmed hypotheses concerning modality effects in learning words. First, learners were faster to learn printed words than spoken words. Second, learners were more accurate and confident in identifying words tested in the same modality in which they were trained. Third, learners were more accurate and probably more confident in judging the training modality of words they had learned through print than words they had learned through speech. Fourth, individual differences in learning rate were associated with differences in dependence on modality congruency for all measures and in confidence. We discuss each of these findings in turn.

The finding that experienced readers are faster to learn new words through print than through speech is consistent with previous research showing visual inputs led to faster paired associate learning than spoken inputs (Baddeley et al, 1988; Beery, 1968) as well as better recall (Papineau and Lohr, 1981). Baddeley et al. (1988) and Atkins and Baddeley (1998) explain the visual advantage by suggesting that learners add information to a visually presented word by recoding to phonology.

The process of "adding" a phonological code is a central idea in the self-teaching hypothesis (Jorm & Share, 1983; Share, 1995, 1999). Indeed, the phonological coding is critical for acquiring an orthographic representation of a word that a child already knows and for strengthening a newly formed orthographic representation. We think there is a natural extension of this idea to the learning of rare words by skilled adult readers. Because the words are novel, they have no lexical representation phonologically or orthographically prior to training. Decoding the written word is possible because of readers' knowledge of letter-phoneme correspondences. This decoding can add a phonological code, that in turn, can stabilize the new orthographic representation. If readers are exposed to only the phonology, however, the new lexical representation may be of low quality. The phonological segments may be underspecified in the initial representation and orthographic recoding will not readily take place. This difference between novel spoken and written inputs would arise if lexicality plays a different role in the two cases, and it may. An encounter with a previously unknown word does not produce lexical access because there is no lexical entry. Not having a lexical entry may be a larger problem for a spoken word than a written word for a literate person. A novel written word can be decoded pre-lexically. The analogous process for speech, orthographic coding, may be more dependent on a post-lexical process that retrieves the spelling of a word, rather than generating the spelling of a phoneme sequence.

Turning to the more general result of modality dependence between encoding and testing, this effect is consistent with the assumption that word learning occurs when episodic traces of word encounters are laid down in memory. The episodic context of the word event (e.g. its modality) is part of this memory. The degree to which words are later recognized accurately and confidently depends on two things: the amount of overlap between the memory trace and the memory probe and the quality of the memory trace. Learners encountering a word in the same modality in which the word was learned can recognize the word more accurately and with more confidence than words encountered in a new modality, because the probe overlaps with the memory trace to a greater degree.

The strength of the memory trace, however, seems to vary with both training modality and individuals. Learners establish memory traces for new words through written forms more quickly than through spoken forms. Despite having fewer and less recent exposures to the visual words during training, learners were more accurate in modality judgments for visually trained words and tended to express more confidence in their judgments. Performance was especially high when words were both presented and tested visually. It is important to note that the analyses performed on modality accuracy did not take into account any bias in judging modality, because there were too few participants who made false alarms to both written and spoken foils to do a complete *A'* analysis for modality judgment. However, when participants did make false alarms, they more often reported learning the word through speech rather than through reading. Thus a general tendency to respond that a word had been heard would lead to better performance (hits) on spoken words. However, the result was the opposite — reliably better performance for visually learned words. This justifies the conclusion that not only are words learned more quickly in the visual modality, but the resulting representations are also more robust.

Similarly, although fast word learners have fewer encounters with the trained words (by definition), they show better performance than slower learners on both different-modality testing conditions. One line of explanation for a reduced modality dependence for fast learners is the bi-modal coding hypothesis, which assumes a tendency for fast learners to add the "missing" code to a word, i.e. phonologically coding a written word and orthographically coding a spoken word. This would establish a more complete, higher quality representation (Perfetti & Hart, 2001). However for this hypothesis to hold, fast learners would need to do what we said above is less likely in general, namely to generate an orthographic code pre-lexically for a novel spoken word. Clearly people can spell words they have never heard before, and our fast learners may have used this pre-lexical spelling strategy, producing both faster learning and less dependence on input modality.

A related but different line of explanation is based on the link between phonological memory and vocabulary acquisition (Atkins & Baddeley, 1998, Gathercole, Willis, Emslie, & Baddeley, 1991; Service & Kohonen, 1995). This explanation assumes that faster word learners are faster at learning phonology (than slower learners) because they are better able to acquire and store a well-specified phonological representation. To explain fast learners' reduced dependence on input-test modality matching, this phonological memory hypothesis would need to add an assumption about what happens at test: The participant decodes a written word and test and then compares the phonological output of this decoding with a memory of the phonological form of a word learned through speech. This explanation does not assume that faster word learners added an orthographic code to the spoken word, but rather that differences in phonological skills (decoding, acquisition, specification) form the basis for all recognition effects observed. Notice that bi-modality coding hypothesis is also consistent with this assumption concerning phonological skill. Such skill could be the basis of an ability to convert pre-lexical phonology into orthography. This experiment can not distinguish between these two similar hypotheses, which really differ on whether the reduction of modality dependence occurs in the encoding of the word at learning or the decoding of the word at test.

Finally, we consider how more abstract less modality-dependent knowledge of a word emerges. An initial memory for a novel word includes episodically specific traces. When subsequent experiences with the word do not vary its episodic context, this initial context is retained as a strong part of the memory trace. However, as exposure and context variability increases, the memory for the word becomes less dependent on any one context. Higher levels of word learning allow an increasingly more abstracted representation, one that is robust and only loosely tied to experience-specific information. These more robust representations are enabled by a reader's knowledge of orthography-phonology correspondences that, when applied to word encounters, can strengthen the representation of the word by adding orthographic and phonological information. Although experiencing words in written form rather than spoken forms is better for learning, theoretically a combination of written and spoken presentations should be better for retention. A learning process with variable context should reduce the learner's dependence on a single context.

Notes

* This research was supported by a grant from the Institute of Education Sciences to Charles Perfetti.

1. We used A' instead of *d'* to accommodate multiple cases in which the false alarm rate was 0 or the hit rate was 1. *A'* does not assume underlying normal distributions with equal variance and has been called a "nonparametric" test of discriminability. For an explanation of how to compute *A'*, see Snodgrass & Corwin (1988).

2. One participant was eliminated due to identifying all items, including all foils as "old". This created a *d'* of zero for every condition. The participant may have had difficulty understanding the instructions.

3. Nelson-Denny comprehension scores were not significantly correlated with the rate of vocabulary learning or with the dependent variables in this study.

References

Atkins, P.W.B. and A.D. Baddeley. 1998. Working memory and distributed vocabulary learning. *Applied Psycholinguistics* 19: 537–552.

Balota, D.A., MJ. Cortese, K.A. Hutchison, J.H. Neely, D. Nelson, G.B. Simpson and R. Treiman. 2002. The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. http://elexicon.wustl.edu/, Washington University.

Baddeley, A.D., C. Papagno and G. Vallar. 1988. When long-term learning depends on short-term storage. *Journal of Memory and Language* 27: 586–595.

Beery, K. E. 1968. Auditory vs. visual learning of words. *Perceptual & Motor Skills* 26(3): 862.

Berry, D.C., S. Banbury, L. Henry. 1997. Transfer across form and modality in implicit and explicit memory. *The Quarterly Journal of Experimental Psychology* 50A(1): 1–24.

Borowsky, R., W.J. Owen and N. Fonos. 1999. Reading speech and hearing print: Constraining models of visual word recognition by exploring connections with speech perception. *Canadian Journal of Experimental Psychology* 53: 294–305.

Bosman, A.M.T. and G.C. van Orden. 1997. Why spelling is more difficult than reading. In *Learning to Spell,* C.A. Perfetti, L. Rieben and M. Fayol (Eds.), 173–194. Hillsdale NJ: Lawrence Erlbaum Associates.

Castles, A., V.M. Holmes, J. Neath and S. Kinoshita. 2003. How does orthographic knowledge influence performance on phonological awareness tasks? *The Quarterly Journal of Experimental Psychology* 56A(3): 445–467.

Cossu, G., M. Gugliotta and J.C. Marshall. 1995. Acquisition of reading and written spelling in a transparent orthography: Two nonparallel processes? *Reading and Writing* 7: 9–22.

Dean, R.S., F.R.Yekovich and J.W. Gray. 1988. The effect of modality on long-term recognition memory. *Contemporary Educational Psychology* 13(2): 102–115.

Elbro, C. and M.N.Jensen. 2005. Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology* 46: 375–384.

Elbro, C., L. Nielsen and D.K. Petersen. 1994. Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia* 44: 205–226.

Gallo, D.A., K.B. McDermott, J. Percer and H.L. Roediger. 2001. Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(2): 339–353.

Gathercole, S. E., C. Willis, A. Emslie and A.D. Baddeley. 1991. The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics* 12: 349–367.

Gough, P.B., C. Juel, P.L. and P.L Griffith. 1992 Reading, spelling, and the orthographic cipher. In P.B. Gough, L.C. Ehri and R. Treiman (eds), *Reading Acquisition,* 35–48. Hillsdale NJ: Lawrence Erlbaum Associates.

Hogaboam, T. C. and C. Perfetti. 1987. Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology* 70(5): 717–729.

Johnston, M., M. McKague and C. Pratt. 2004. Evidence for an automatic orthographic code in the processing of visually novel word forms. *Language and Cognitive Processes* 19(2): 273–317.

Jorm, A.F. and D.L. Share. 1983. Phonological recoding and reading acquisition. *Applied Psycholinguistics* 4: 103–147.

Kucera, H. and W. Francis. 1967. *Computational Analysis of Present-Day American English.* Providence RI: Brown University Press.

Lukatela, G., S.J. Frost and M.T. Turvey. 1999. Identity priming in English is compromised by phonological ambiguity. *Journal of Experimental Psychology: Human Perception and Performance* 25(3): 775–790.

Lukatela, G. and M.T. Turvey. 1994a. Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General* 123(2): 107–128.

Lukatela, G. and M.T. Turvey. 1994b. Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General* 123(4): 331–353.

Naish, P. 1980. The effects of graphemic and phonemic similarity between targets and masks in a backward visual masking paradigm. *Quarterly Journal of Experimental Psychology* 32: 57–68.

Nelson, M.J. and E.C. Denny. 1973. *The Nelson-Denny Reading Test*. Houghton Mifflin Company.

Papineau, W. and J.M. Lohr. 1981. Effect of sensory modality for presentation and word imagery on paired-associate learning. *Perceptual & Motor Skills* 52(3): 747–750.

Perfetti, C. 1997. Psycholinguistics of spelling. In C.A. Perfetti, L. Rieben and M. Fayol (eds), *Learning to Spell,* 21–38. Hillsdale NJ: Lawrence Erlbaum Associates.

Perfetti, C. and L. Bell. 1991. Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language* 30: 473–485.

Perfetti, C., L. Bell and S. Delaney. 1988. Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language* 27(1): 59–70.

Perfetti, C.A. and L. Hart. 2001. The lexical bases of comprehension skill. In D. Gorfien (ed.), *On the Consequences of Meaning Selection,* 67–86. Washington DC: American Psychological Association.

Perfetti, C.A., E.W. Wlotko and L.A. Hart. 2005. Word learning and individual differences in word learning reflected in ERP's. Manuscript submitted for publication.

Reichle, E.D. and C.A. Perfetti. 2003. Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading* 7(3): 219–237.

Seidenberg, M. and M. Tanenhaus. 1979. Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory* 5(6): 546–554.

Service, E. and V. Kohonen. 1995. Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* 16: 155–172.

Share, D.L. 1995. Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition* 55: 151–218.

Share, D.L. 1999. Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology* 72: 95–129.

Slowiaczek, L., E. Soltano, S. Wieting and K. Bishop. 2003. An investigation of phonology and orthography in spoken-word recognition. *The Quarterly Journal of Experimental Psychology* 56A(2): 233–262.

Unsworth, S. and P. Pexman. 2003. The impact of reader skill on phonological processing in visual word recognition. *The Quarterly Journal of Experimental Psychology* 56A(1): 63–81.

Van Orden, G., J. Johnston and B. Hale. 1988. Word identification in reading proceeds: From spelling to sound to meaning. *Journal of Experimental Psychology* 14(3): 371–386.
Ziegler, J. C. and L. Ferrand. 1998. Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review* 5(4): 683–689.