Lexical Access via an Orthographic Code: The Basic Orthographic Syllabic Structure (BOSS)

MARCUS TAFT

Massachusetts Institute of Technology

Five experiments are described which detail both the structure of lexical representations accessed in visual word recognition and the method of parsing words in order to access these representations. All experiments employ the lexical decision task. The first three experiments provide support for the view that the lexical representation of a word is a representation termed here the Basic Orthographic Syllabic Structure (BOSS) of that word rather than any articulatorily defined structure. The principle for determining the BOSS is based on both orthographic and morphological factors. Experiments 1 and 2 demonstrate that words are disrupted more when a division is made in the word according to its pronunciation than when it is made according to its BOSS. It is shown in Experiment 3 that final silent E's are not a part of the lexical representations of words and this is predicted by the BOSS principle. The last two experiments provide evidence for a parsing procedure for retrieval that is a reiterative left-to-right technique whereby lexical searches are made for combinations of letters that begin with the initial letter of the word. The results, in all, support a system of lexical access that is not mediated by phonological encoding.

In order to recognize and comprehend a written word, the orthographic representation of that word must be encoded in some way and matched with a memory, or lexical representation, in the reader's long-term word store, the lexicon. The type of encoding that is undertaken by the reader has been the subject of considerable interest in recent years. Several types of encoding have been suggested (not necessarily mutually exclusive):

1. **Phonological encoding.** The claim has been made (e.g., Rubenstein, Lewis, & Rubenstein, 1971; Meyer, Schvaneveldt, & Ruddy, 1974) that the orthographic representation of a word is converted by grapheme-phoneme correspondence rules into a phonological representation, since it is supposed that all lexical representations of words are stored in a phonological form (though they may be stored in an orthographic form as well).

2. **Morphological encoding.** It has been proposed (e.g., Gibson & Guinet, 1971; Murrell & Morton, 1974; Taft & Forster, 1975) that polymorphic words are stripped of their prefixes and inflections in the recognition process since such words are stored in the lexicon in a morphologically decomposed state. For example, the lexical representation of REHEATING would be HEAT, with information about RE and ING being stored within this lexical entry.

3. **Syllabic encoding.** Hansen and Rodgers (1968) and Spoehr and Smith (1973) have suggested that words are parsed into syllables (or, at least, syllable-like units) since it is supposed that polysyllabic words have their syllable structure represented in the lexicon. Taft and Forster (1976) also propose that...
items because there were more letters provided before the gap. Thus LANT conveys more information about LANtern than does LAN. In order to counter this criticism, it must be shown that O-Group words are more quickly recognized than words divided up at a point after their BOSS. For example, the BOSS of BOYCOTT would be BOY (since YC is an orthographic violation) and thus BOY COTT should be quicker to recognize than BOY COTT despite the fact that BOY has more letters than BOY. Morphological principles also provide items of interest. For example, the BOSS of FULL will be FUL on morphological grounds, and thus FULL should be more rapidly recognized than FULL ILL.

Twenty of each item were included in Experiment 2 to ensure that the effect obtained was not a result of the length of the first word chunk. It was found that the mean reaction time to the words broken up according to their BOSS (BOY COTT) was shorter than that to words broken up at a point after their BOSS (BOY COTT); 614 vs 656 mses. This difference was significant on both the subject and item analyses but not on minF (1.48, n = 377, p > 05). However, the failure to reach significance is not crucial for the argument against the notion that the advantage of LANT E RN over LAN TERN was a result of there simply being more letters in LANT than LAN. Had the number of letters in the first chunk been the essential factor, then items like BOY COTT should have been more quickly responded to than items like BOY COTT. But, if anything, the reverse was true.

Since the results of this experiment have such important implications for theories of lexical storage, and further, since the finding is hardly an intuitively obvious one, it was considered advisable to try to replicate the O-Group/P-Group difference with a different set of items and subjects. In this second experiment, however, rather than separating the units of a word by means of a gap, the boundary between units was marked by a case transition. A word either shifted from upper case to lower case (e.g., MUSTard) or from lower case to upper case (e.g., zero) would be unconvincing if the effect observed in the first experiment was dependent upon the word being divided by a gap and not by any other means. If, on the other hand, the O/P difference was still observed using the case alternation method of division then the conclusions drawn from Experiment 1 would be strengthened.

In addition to altering the method of word division in the second experiment, an attempt was made to improve upon the method of item selection. Items were chosen for Experiment 1 purely on the basis of whether they possessed a medial consonantal cluster or a long first vowel. Other than this, there were no systematic criteria for deciding which words would be included and which would not. The second experiment, on the other hand, employed a totally automatic and thus objective method of item selection. Highly rigorous criteria were adopted for item selection and all words meeting these criteria were used in the experiment.

**Experiment 2**

Experiment 2 differed from Experiment 1 in five ways:

1. Word items were selected systematically.
2. Rather than splitting each item into units by means of a gap, the units were demarcated by a transition from upper case to lower case or from lower case to upper case (e.g., MUSTard vs MUSTard, zero vs zero).
3. A third condition was included where the item was left intact, either completely in upper case or completely in lower case.
4. Nonword items were constructed so that they could be analyzed in the same way as word items.
5. Stimulus display was controlled by a PDP 11/10 computer.

The changes outlined in Points 1, 2, and 5 were methodological changes that should not have altered the basic findings. If O-Group words were still found to be recognized faster than P-Group words in Experiment 2 then the generalizability of the effect would be well established.

The predictions made for nonwords (Point 4) is that O-Group items and P-Group items should not differ since there is no stored structure for nonwords and thus an exhaustive search would be undertaken in both cases trying out all possible encodings. Thus for a nonword like SILPONTH, both a SILPONTH parse and a SILPONTH parse must be found to be unsuccessful before it can be decided that SILPONTH is not a word.

The non-disrupted condition was included (Point 3) to examine whether the O-Group items were not only more easily recognized than P-Group items but also facilitated recognition compared to an intact normally presented item. It would be very interesting if the results did demonstrate such a facilitation since the O-Group items are physically disrupted and this may well delay the letter identification stage of processing. Words written with case alternating from letter to letter to do indeed increase recognition difficulty (Coltheart & Freeman, 1974; Fisher, 1975).

Of course, the present experiment involves only one shift in case type and is thus less of a disrupted stimulus than when case type alternates from letter to letter.

**Method**

**Materials.** The items that were selected were all those words falling in a particular frequency range (20–30 according to Kučera & Francis, 1967) that met the following criteria:

1. The word was to be polysyllabic and 4-7 letters long.
2. The word was to either have a medial pair of nonidentical consonants (e.g., the ST of MUSTard) or a long first vowel (e.g., the E of zero).
3. A medial consonant pair could not be NG or NK since the N could then be pronounced as /ŋ/ and not /n/’.
4. A P-Group division would not represent the correct pronunciation in this case. For example, BLANKet would lead to the pronunciation /blænk kat/ and not /blæk kat/.
5. The word was not to be polyphormic (e.g., NERVous, PROTEST, RACES). Every word that met the above criteria was used in the experiment. There were 44 such words and these are presented in the Appendix.

Words were presented either as P-Group items or as O-Group items. Grouping of letters was achieved by presenting the letters of the grouping in a different case from the rest of the word (i.e., either upper or lower case). Half of the items were randomly assigned to an upper case followed by lower case set (e.g., MUSTard) while the other half were assigned to a lower case followed by a lower case set (e.g., zero). P-Group items were words where the case transition was made at the phonetic syllable boundary as defined in Experiment 1. That is, where there was a medial consonant pair the division was made between the two consonants (e.g., MUSTard) and where the first vowel was long the division was made immediately after this vowel (e.g., zero). Grouping of letters in the O-Group items conformed to the BOSS principle regardless of pronunciation (e.g., MUSTard, zero).

The items were presented to two groups of subjects and were selected randomly to appear either as P-Group words for the first set of subjects while being O-Group words for the second set, or as O-Group words for the first set of subjects while being P-Group words for the second set. Thus one group of subjects would have been presented with MUSTard while the other group was presented with MUSTard and P-Group words and 22 O-Group words.
supported, then the argument for the BOSS principle will be further strengthened.

**Experiment 3**

The design of this experiment follows the design of several interference experiments reported by Taft and Forster (1975, 1976). If pairs of words are presented to subjects and found to have relatively slow classification times as nonwords or words, then it is concluded that these word parts are stored in the lexicon. Thus, it was concluded by Taft and Forster (1975) that INFECT is stored as FECT and PREVENT is stored as VENT, since FECT and VENT took longer to classify than control items. Similarly, it was concluded (Taft & Forster, 1976) that ATHLETE is stored as ATH and NEIGHBOR stored as NEIGH, since ATH and NEIGH took longer to classify than control items. The fact that FAW and SHREW did not take longer to classify than control items indicated that FAWN is not stored as FAW and SHREW is not stored as SHREW.

Now, if silent E's are not present in the lexical representations of words, then items like STONE (Condition 1) should take longer to classify as nonwords than items like SLOW (Condition 2), since the entry for STONE (namely STON) will be encountered during the lexical accessing process. Similarly, items like SHINE (Condition 3) should take longer to classify as words than items like SWAN (Condition 4) since the lexical accessing process will encounter the entry for SHINE (being of higher frequency than SHIN) prior to the entry for SHIN.

**Method**

**Materials.** Fifteen Condition 1 nonwords were constructed by removing the final silent E from 15 monosyllabic words (e.g., STON from STONE). These were matched for length and CVC structure with 15 Condition 2 nonwords which were not generated from words with silent E's (e.g., SLOW).

The word items of Conditions 3 and 4 were equally frequent, CVC structure, and length. There were 15 words in each condition. The words of Condition 3 were words which were of lower frequency than words of the same form but with a final silent E (e.g., SHIN, which is lower frequency than SHINE). Condition 4 words were not related by a final silent E to any other words (e.g., SWAN).

The 60 items were presented to subjects in different random orders following 10 practice items. The procedure was otherwise the same as that described in the first experiment.

**Results and Discussion**

Mean reaction times across items for the four conditions are presented in Table 3. A comparison of the two nonword conditions revealed a significant difference in the pronunciation of STONE (namely /ston/) as compared to the pronunciation of SLOW (namely /sλon/), whereas /slon/ is not similar to the pronunciation of any real word. This possible explanation for the result seems unfounded, however, when one considers some of the other items used in the experiment. For many of the item pairs, the Condition 2 nonword was just as similar in pronunciation to a real word as was its matching Condition 1 nonword, and in some cases, was even more similar. For example, /dek/ (the pronunciation of the Condition 1 item DANCE) is similar to /de:n/ (the pronunciation of the word DANCE), and /fpk/ (the pronunciation of the Condition 2 item RING) is just as similar to /n:ʃ/ (the pronunciation of the word RING). Similarly, while HOM (Condition 1) has a similar pronunciation to HOME, LOM (Condition 2) has just as similar a pronunciation to LOAM or even OOM. The pronunciation of SCENE (Condition 1) is not at all similar to the pronunciation of SCENE (i.e., /sken/ compared to / s:n/), and CRING (Condition 1) is less similar in pronunciation to CRINGE than BLING (Condition 2) is to BLINK (i.e., /brŋ/ compared to /brŋ/ versus /blŋ/ compared to /blŋ/). So it can be seen from these examples (and these are not the only ones) that it is highly unlikely that the result obtained in Experiment 3 can be accounted for by phonological similarity to words.

Experiments 1, 2, and 3 have provided evidence for the storage of words as representations of their BOSS, being that unit of the first morpheme of the word that includes as many consonants after the first vowel as is orthographically valid. So the word LANTERN would have LANT as its BOSS since this takes in all the consonants after the first vowel of the morpheme LANTERN; the word SLANT would have SLANT as its BOSS since SLANT is the first (and only) morpheme of the word and at the same time includes all the consonants after the first vowel; the word PEARLY would have PEARL as its BOSS since the first morpheme is PEARL; and the word NEARLY would
It is clear from the results that the presence of an inappropriate word at the beginning of a nonword or word interferes with the classification response. This finding, then, suggests that retrieval of words from the lexicon is achieved via a left-to-right parse technique, namely the LR hypothesis put forward by Taft and Forster (1976). If retrieval were carried out according to the BP hypothesis (initially applying the BOSS principle to the word and then re-parsing if this was unsuccessful) then no interference should be observed for W nonwords and W words since the word appearing at the beginning of the item should never be encountered.

So, to reiterate, while it is postulated that words are stored as representations based upon the BOSS principle it is proposed that words are retrieved by a reiterative left-to-right parse technique and not by application of the BOSS principle.

**EXPERIMENT 5**

An essential feature of the left-to-right search model is that it begins at the extreme left of the word, that is, at the first letter of the word. It is necessary to postulate this in order to explain why the presence of a word in the last syllable position of a nonword (e.g., FLURBPARI) does not lead to interference effects (Taft & Forster, 1976). To further demonstrate this point Experiment 5 was designed using items where a single letter was added to the beginning of a word to form a nonword (e.g., GLIVE from LIVE) or a less frequent word (e.g., CLOVE from LOVE). If it were shown that the presence of a word at the end of a nonword or less frequent word did not cause any interference to lexical decision responses, then the left-to-right search procedure would be further supported since it was shown, in Experiment 4, that one does get interference when there is a word at the very beginning of a nonword or less frequent word.

**Method**

Fifteen W nonwords were constructed by adding a letter to the beginning of 15 common words (e.g., GLIVE, TWAIT). For each of these nonwords was another nonword that did not end with a word, but was similar in consonant-vowel structure to its matching W nonword (e.g., GLITE, TWILK). These nonwords constituted the N nonword condition. The 30 nonword items were presented along with 40 word items. Half of the word items were W words being constructed by the addition of a letter to the beginning of 20 common words (e.g., CLOVE, THIGH). Matched for frequency with each of these words was an N word. These words did not end with another word (e.g., CROAK, THUMB).

The apparatus and procedure were the same as in Experiment 3. There were 15 subjects.

**Results and Discussion**

Mean reaction times and percentage error rate are presented in Table 5.

There was found to be no difference in reaction time either between W nonwords and N nonwords (min F < 1) or between W words and N words (min F(1,33) = 2.14, p > .05), although the subject analysis was significant [F(1,14) = 5.21, p < .05], with reaction times actually being shorter for the W words than for the N words).

These results confirm the prediction that a word contained in a nonword or word item will only cause interference to the classification response if it is at the beginning of that item, that is, a lexical entry for the word that is included in the item will only be encountered if it is at the beginning. This finding, in conjunction with the results of Experiment 4, support the claim that the lexical representation of a word (namely, its BOSS) is retrieved via a parsing technique whereby a lexical search is made for all letter combinations within an item that begin with the first letter of that item. When there is a lexical entry that is congruent with one of these letter combinations, then this letter combination becomes a candidate for being the BOSS of the word.

**GENERAL DISCUSSION**

As stated in the introduction, a written word is recognized after the orthographic representation of that word is parsed in some way and then found to match with a lexical representation. The present experiment addresses the two issues of how the word is parsed and how the representation of the word is encoded in the lexicon. The results of the first three experiments, along with those obtained by Taft and Forster (1976), argue for the notion that words are encoded in the lexicon as representations of their first syllable defined orthographically and morphologically rather than phonetically. More precisely, the lexical representation of a word is a representation of that part of its first morpheme that includes after its first vowel all consonants that do not violate rules of orthographic co-occurrence. This is the BOSS.

The method of parsing is suggested by the last two experiments to be a reiterative left-to-right technique whereby lexical look-up is attempted for those letter groupings within a word that begin with the initial letter of that word. When a lexical representation is located by this technique the lexical entry must be examined further in order to ascertain whether or not it is the appropriate entry. For example, if the lexical entry for BEAR is accessed when the word BEARD is presented, the absence of any information within this entry saying that BEAR + D is a word will necessitate the continuation of the lexical search until the correct entry for BEARD is located.
### Orthographic Access Code

| flower | 512 | TRIBUTE | 589 |
| savage | 557 | PANIC | 520 |
| gather | 546 | CHAPEL | 565 |

**Nonwords**
- ampOW: 681, amPOw: 628
- BLUNdIn: 682, BLUNdIn: 651
- braikO: 627, braikO: 577
- padONY: 617, paDONy: 633
- THANDeR: 632, THAnDEr: 650
- gavIAL: 643, gaVIAl: 678
- spADOR: 639, spADOR: 674
- nooDATE: 634, nooDATE: 683
- lobEN: 615, lobEN: 600
- FLENdIn: 600, FLENdIn: 623
- omnPLE: 682, omnPLE: 700
- prustIN: 653, prustIn: 621
- ramONY: 660, raMONy: 683
- LAtire: 606, LAtire: 605
- CHIMBer: 718, CHIMber: 726
- BASTop: 628, BASTop: 676
- BLATer: 649, BLATer: 679
- zABLe: 591, zABLe: 590
- GRONDIN: 664, RIBEN: 613
- plandi: 615, serody: 643
- CATULE: 671, SHENKER: 632
- LASKIP: 599, toscarp: 632
- RITER: 647, VOLPENCH: 588

---

**References**


---

**Reference Note**


(Rceived February 21, 1978)