Word Segmentation in Written Text: an Argument for a Multiple Subunit System

Ruth Ann Atchley, Ovid Tzeng, and Melanie Cass
Department of Psychology: University of California, Riverside
Riverside, California 92521
ruthann@cassandra.ucr.edu

Abstract

Two types of word stimuli, easily syllabified (eg. *balcony*) and ambisyllabic English words (eg. *balance*), were used in a reading task designed to determine if words are processed using strictly syllables as the unit of segmentation or if multiple units of segmentation are used. We could not replicate work done by Prinzmetal, Treiman, and Rho (1986) who found that, with text, more illusory conjunction occurred within syllables than across syllable boundaries. In contrast, our work supports the hypothesis that English is too complicated of a language to use only one segmenting unit. Thus, the pattern of results was dependent on the structure of the words themselves, with the ambisyllabic words being processed using phonemes and not syllables as the unit of segmentation.

Introduction

Some form of word segmentation, the process of finding simple units by which either speech or text can be processed and accessed in semantic memory, is generally believed to be a necessity given the complexity of our language. Particularly in speech processing, where the signal is continuous (Cutler & Norris, 1988), the number of individual patterns which must be processed by the perceiver is infinite. In written language, despite the fact that the structure of the stimuli is better segmented, the number of potential inputs that the system could possibly be called upon to process is no less staggering. This general type of reasoning is well accepted in the literature (Henderson, 1982; Spoehr, 1981; Cutler, Mehler, Norris & Sequi, 1986). However, what these units of segmentation are is a matter of some controversy. Furthermore, with a few exceptions, this body of research tends to ignore the question of whether or not the same type of processing or segmentation that is used by audition is used in written text. Therefore, we must address the issue of whether empirical results from one modality can be generalized to the other. Lastly, the complexity of the English language must be taken into account when generating stimuli. For example, a study of the processing of easily syllabified word, for example *pardon*, in which a CVCCVC structure leads to segmentation between the central, nonidentical, adjacent consonants may not yield the same results as a study including ambisyllabic words which have a CVCVC type of structure, with a central consonant that "properly belongs to both the first and second syllables" (Cutler, Mehler, Norris & Sequi, 1986) . As a result, carefully chosen stimuli which generalize to the multiple-structured and complex nature of the English language must be used in order to truly test any potential subunit's validity.

One unit of segmentation which is often suggested as the key primary unit for processing has been the syllable. Savin and Bever (1970) showed that prespecified, or primed, syllables were detected faster than prespecified phonemes. Thus, they argued that the syllable is the first unit of processing followed by a decomposition to phonemes. Sequi, Fruenfelder and Mehler (1981) also found that in monitoring tasks syllable sized targets were easier for listeners to detect than phoneme sized targets. Liberman, Shankweiler, Fisher, and Carter (1974) used work with young children in order to see which unit of process is more readily used. They found that children can easily identify a syllable but that a phoneme is harder to identify.

In addition, Spoehr and Smith (1973; Spoehr, 1981) showed that subjects could process a one syllabled, tachistoscopically presented, five-letter word more accurately than a two syllabled word of the same length. They argued that this difference was due to the fact that words are processed one syllable at a time, thus a one syllabled word is chunked and stored quickly as a whole, while a two syllable word requires twice the storage time (see Spoehr, 1981 for review). A more recent study (Prinzmetal, Treiman & Rho, 1986) uses an interesting research methodology to again establish that the syllable is the important subunit. Prinzmetal, et al., used a grouping paradigm, usually used in attention research, to show that more illusory conjunctions occur within a syllable then across syllable boundaries. Therefore, they argued that the syllable is the most strongly grouped unit in a word, much like a line of similar figures would be strongly grouped in a Gestalt display.

The second school of thought on segmentation has been voiced by Cutler, Mehler, Norris, and Sequi (1986, 1992). This approach argues that word segmentation is very much affected by the language being processed and, within English, by the structure of the words themselves. Therefore, for the English language, the complexity of the language due to its "stress-based" linguistic structure, requires a processor to either use a phoneme based segmentation strategy or to use multiple subunits. This new approach will investigate whether the earlier results that showed that the syllable is the most important and first accessed subunit, was not partly due to the use of uncomplicated, easily syllabified words as stimuli. In this research, however, Cutler et al., has used only auditory stimulus. Will the processes underlying auditory segmentation generalize to segmentation of written text?

These two theoretical questions lead us to take advantage of the the paradigm proposed by Prinzmetal et al. (1986). Our study included stimuli, both from that paper and from research done by Cutler et al. (1986). The hypothesis being tested proposes that the processing of text will mimic speech, and the subunit of segmentation will be dependent on the word being processed. In other words, a syllable is one of the possible subunits that may be used, but that the language system will not solely segment words into syllables.

Methods

Subjects

The subjects were 30 English speaking undergraduates at the University of California, Riverside. All reported having normal eyesight and all were right handed. Subjects participated in order to earn research credit as required by an introductory psychology course.

Apparatus and Materials

The stimulus were presented on a IBM PC with color monitor using MEL software. The computer was housed in a blackened room, and all subjects participated alone. A researcher, however, was present in the room at all times to monitor the subjects' concentration.

Stimuli consisted of a white target letter presented on a black screen for 150 msec, and a dual-colored stimulus word which was made up of two of four possible colors: blue, green, pink, or yellow. The word was centrally presented for 225 msec. Words ranged from five to eight letters in length. Subjects were given 2500 msec to respond to each trial.

The design of the experiment was a 2X4 factorial design. With the first factor, colored letter position, having two levels, second letter and fourth letter in the word. The position of the colored letter should interact with the syllable structure in such a way as to cause more illusory conjunctions with the target letter (more color naming mistakes) if the colored letter falls in the same syllable as the target. For example, if the stimulus word is *fancy,* having *a* be of a different color then the rest of the word should cause more errors then if the odd colored letter is in the fourth letter position, such as *c* in this example.

The second independent variable corresponded to words in one of four possible categories (syllable types): two letter syllables, three letter syllables (stimuli for these two conditions were taken from the Prinzmetal et al. study, 1986), strong first syllabled words and weak first syllabled words (these two were modeled after Cutler et al., 1986). Strong first syllable words had CVC structure for its initial syllable, usually followed by another consonant in the fourth letter position. A weak first syllable word contained an ambisyllabic third letter (eg. *balcony* vs. *balance*). An example of each type of trial is as follows: two letter syllable *armor,* three letter syllable *larva,* strong syllabled *napkin,* weak syllabled *helium.* Frequency of words was not controlled for.

Of the total number of trials ten percent were catch trials. These catch trials were ones in which the target letter was not present in the dual-colored word. If the target letter was present, it was the third letter of the word. If it was a target absent trail, the target letter was a consonant chosen from among the other possible target letters. This insured that every target letter used was sometimes responded to as present and sometimes as absent.

Procedure

Each subject signed up for one hour of experimental credit. At the beginning of the session, each subject was given a short interview to determine language experience and handedness. Subjects then moved to the computer and were presented instructions on the screen. They were asked to read these instructions aloud.

The instructions told the subject to: first, memorize a target letter, wait for a rectangular mask of asterisk to pass, and then read a word in which one letter would be a different color then the rest. They were then told to respond by pressing one of five buttons on the keyboard that corresponded to the four colors or the space bar, if the target was not present.

After the instructions, subjects were given approximately 20-25 practice trials. If a trial was responded to correctly then no feedback was given and the next trial begins. However, if the practice trial was responded to incorrectly then a tone sounded, which indicated an incorrect response and then the trial was rerun with the same target and word screens. This processes of reshowing incorrect trials functionally gave a subject who was having trouble understanding some aspect of the instructions more practice time. Once all of the practice trials were completed correctly, subjects were presented two blocks of 150 trials each. The subject was able to complete all of these steps in the allotted hour. Afterwards, a debriefing occurred.

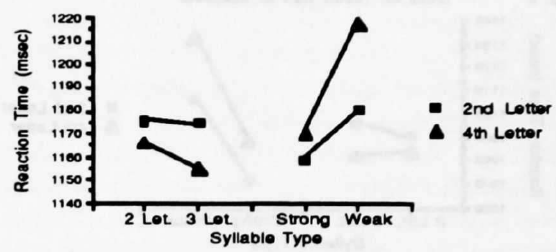
Results

Means for the subjects, across the 2x4 design (color position x syllable type) are listed in Table 1. There was an overall main effect using reaction time as the dependent variable ($F(7,232) = 4.48$, $p<.0001$). However, there was no significant main effect using the dependent variable, accuracy.

Table 1. Data for all subjects: Color Position X Syllable Type (msec)

Syllable Type	2nd Letter	4th Letter
2 letter syllables	1175	1166
3 letter syllables	1174	1155
Strong syllables	1158	1170
Weak syllables	1180	1218

Fig. 1 Data for all subjects



There was a main effect for the variable, syllable type, $F(3,232)=6.48$, $p<.001$ and for the interaction term, syllable type by colored letter position, ($F(3,232)= 3.73$, $p<.05$). A Tukey test of these results indicates that the main effect for syllable type is due to weak syllables being significantly slower than each of the other three syllable types (2-lettered syllables, 3-lettered syllables and weak syllables). Post-hoc paired comparisons of individual cell means indicate that the significant interaction is due, primarily, to the slowness of subjects when responding to weak syllabled words with the fourth letter being of a different color ($F(1,58)= 7.55$, $p<.01$) (See Figure 1.)

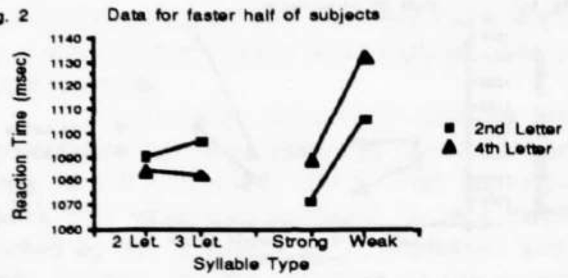
During analysis, it was noted that a many trials were given a maximum value of 2500 msec because the subject had run out of time when responding. A median split of the data was performed in order to avoid any bias this capping effect might have had on the data, particularly for the weak syllable condition, for the slower subjects. Furthermore, there is no theoretical reason why the faster subjects would be any different then the slower subjects.

For the faster subjects, the weak syllable condition was slower then the other three, $F(3,120) = 4.61$, $p<.01$ (see Table 2).

Table 2: Data for faster subjects

Syllable Type	2nd Letter	4th Letter
2 letter syllables	1089	1084
3 letter syllables	1096	1082
Strong syllables	1071	1089
Weak syllables	1105	1132

Fig. 2



However for the faster subjects the was no significant interaction F(3,120) = 1.47, p=NS (see Figure 2).

Discussion

The hypothesis that segmentation strategy is affected by the complexity of the English language, and that it is not just a simple syllable driven process, is supported by these results. This is particularly true given that we obtained a pattern of data that strongly resemble the results of the Cutler et al. (1986) paper. These results are equally interesting because of the trend in the data for the stimuli used in Prinzmetal et al.'s (1986) earlier study. This trend has a two fold importance to the study; first, because it supports the assumption that this methodological approach is valid, in that there is an effect of syllable structure on where illusory conjunctions occur. Theoretically, however, because the trend is not significant it shows that by adding more complicated stimuli to the word list the syllable effect found by Prinzmetal et al. is dissipated. The combination of these two effects indicate that a multiple subunit processing system is probably used to segment a complicated language like English. So when the syllable structure is pronounced and easily used a reader or listener will take advantage of the syllable to segment the word. However, when the word is ambisyllabic, such as *balance*, then the processing system must shift to a second subunit such as the phoneme. This shift from one unit to another, we believe, explains the significant difference in time it takes to process ambisllylabic versus strong syllabled words. However, there is no difference in accuracy or no more illusory conjunctions across the four conditions. This would argue that the language processor is equally comfortable with both subunits. In other words, the system first attempts to use the more frequent subunit, the syllable, but then, when this leads to ambiguity the system quickly switches to another subunit. One possible alternative subunit would be the phoneme, as argued by Cutler, et al (1986). One way to possibly get at the nature of this alternative subunit would be to use eye tracking technology to determine if in the weak syllable words a smaller unit is attended to.

Acknowledgment: We want to thank Curt Burgess for his help and criticism.

References

Cutler, A., Mehler,J., Norris, D. & Sequi, J. (1986) The syllable's differing role in the segmentation of French and English. Journal of Memory and Language, 25,385-400.

Cutler, A., Mehler,J., Norris, D. & Sequi, J. (1992) The monolingual nature of speech segmentation by bilinguals. Cognitive Psychology, 24, 381-410.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access Journal of Experimental Psychology: Human Perception and Performance, 14, 113-121.

Henderson, L. (1982). Orthography and Word Recognition in Reading. London: Academic Press.

Liberman, I., Shankweiler, D., Fisher, F., & Carter B. (1974). Reading and the awareness of linguistic segments. Journal of Experimental Child Psychology, 18, 201-227.

Prinzmetal, W., Treiman, R., & Rho, S. (1986). How to see a reading unit. Journal of Memory and Language, 25, 461-475.

Savin, H., & Bever, T. (1970). The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 9, 295-302.

Sequi, J., Frauenfelder, U. & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. British Journal of Psychology, 72, 471-477.

Spoehr, K. (1981). Word recognition in speech and reading: toward a single theory of language processing. In P. D. Eimas & J. L. Miller (Eds.), Perspectives on the Study of Speech. Hillsdale, NJ: Erlbaum.

Spoehr, K., & Smith, E. (1973). The role of syllables in perceptual processing. Cognitive Psychology, 5, 71-89.