Title
Variation in Unconscious Lexical Processing: Education an d Experience Make a Difference

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Authors
Libben, Gary
Sveinson, Leone

Publication Date
1994

Peer reviewed

Variation in Unconscious Lexical Processing: Education and Experience Make a Difference

Gary Libben
Department of Linguistics
4-36 Assiniboia Hall
University of Alberta
Edmonton, Alberta, Canada T6G 2E7
glibben@vm.ualberta.ca

Leone Sveinson
Languages Department
Mount Royal College
4825 Richard Road SW
Calgary, Alberta, Canada T3E 6K6

Abstract

Over the past twenty years numerous studies have investigated the extent to which morphological constituents of words are activated during the process of word recognition. In the vast majority of these studies it has been assumed that a correspondence exists between the formal linguistic analysis of a word and its representation in the minds of native speakers. This paper investigates the the extent to which this correspondence can be affected by individual variation that is associated with education, exposure and training. We investigated students who had recently completed a course in medical terminology. These students, and matched control subjects, responded to medical and non-medical multimorphemic stimuli in a lexical decision task. The results indicate that the medical terminology students' training affected their performance on novel medical words as well as their performance on very common medical words (e.g., psychiatry) that would have been part of their vocabulary prior to taking the course. The results therefore support the view that automatic unconscious lexical processing can indeed be modified by explicit training and specialized exposure. This finding has consequences for the generalizability of studies conducted on university students to the general population of native speakers.

Introduction

Linguistic studies of English morphology (e.g., Lieber 1980; Selkirk 1982) have shown that lexical forms can be represented as organizations of discrete morphemic units. Thus, words such as greenhouse, knocked, and remit would be assigned the following formal representations: $[[\text{green}]_{Adj}[\text{house}]_N]_N$, $[[\text{knock}]_V[\text{ed}]_{Af}]_V$, $[[\text{re}]_{Af}[\text{mit}]_V]_V$

Over the past two decades, there has been considerable discussion in the psycholinguistic literature on the extent to which the psychological processes involved in word recognition correspond to these linguistic characterizations of multimorphemic words. A broad array of studies have revealed considerable variation in the extent to which the properties of constituent morphemes play a role in word recognition. In general, it has been found that constituent morpheme effects are more pronounced for prefixed stimuli than for suffixed stimuli (e.g., Henderson Wallis and Knight 1984; Mandelis and Tharp 1977; Stanners et. al. 1979) and more pronounced for inflected words than for derived words (e.g., Smith & Sterling 1982; Stanners et al. 1979; but see Fowler Napps & Feldman 1985). It also has been suggested in a number of studies that morphological decomposition is a special strategy that is often induced in lexical decision tasks with non-word stimuli (e.g., Butterworth 1983) or induced as a result of the presence of particular real-word types in the stimulus set (see Andrews 1986).

This paper investigates the possibility that at least some of the variation reported in the literature on morphological processing is due to the fact that not all forms that linguists identify as multimorphemic are actually composed of discrete morphemes in the minds of native speakers of a language. In the case of the word remit, for example, it is questionable whether the root mit has any semantic representation for native speakers. In that sense, it may not qualify as a morpheme at all. This has also been a point of discussion in the theoretical morphological literature where, for example, Aronoff (1976, p. 11-14) makes the case for the status of mit as a morpheme on the basis of distributional and phonological regularities, but Sproat (1992, pp. 66) rejects its morphemic status on semantic and syntactic grounds. Is remit multimorphemic or monomorphemic? Is it perhaps multimorphemic for some subjects, but monomorphemic for others?

We investigated these possibilities by examining lexical representations in the domain of technical vocabulary, where differences in morphemic knowledge can be linked to differences in subjects' training and exposure to particular morphological constituents. In the technical vocabulary of medicine, for example, there is extensive use of morphemes derived from Latin and Greek. The resulting multimorphemic forms (e.g., pneumonia) may show very distinct internal representations depending on one's education, exposure, and training. The question we have addressed in this paper is: What role does this individual variation play in lexical processing as measured by a task such as lexical decision?

The focus of our study was a group of students who had recently completed a course in medical terminology at a Junior College. The course, which was designed for students who planned to pursue careers as health professionals and assistants, dealt almost exclusively with the acquisition of morphemes used in multimorphemic medical terms. Students were taught the meanings and forms of morphemes and were taught to isolate morphemes in medical terms as well as to create terms for novel medical conditions. Thus, the students differed from control subjects in that they had been taught the meanings of affixes and roots that we could presume the controls did not know, and they had been trained in the strategy of morphological decomposition.

These medical terminology students (M-Group subjects) were tested in a lexical decision task and were compared to age-matched controls (C-Group subjects) using four types of stimuli: (a) novel medical words that only those trained in medical terminology would be able to interpret (e.g, hyponucleal); (b) existing medical words that both the M-Group and C-Group would know (e.g., hypodermic); (c) multimorphemic real words that were unrelated to medical terminology (e.g., disclaimer); (d) multimorphemic nonsense words that were also unrelated to medical terminology (e.g., disarranger). The use of such stimuli promised the opportunity to address two sets of important questions in the processing of multimorphemic words:

(1) Is morphological decomposition a strategy that can be acquired? Can strategic training affect the unconscious processes that are assumed to underlie the lexical decision task?

(2) Does the extent to which subjects have semantic representations for constituent morphemes affect lexical decision latencies?

Method

Subjects

Forty-eight subjects participated as volunteers in the experiment. Twenty-four of the subjects had, within the previous two weeks, completed a first-year Medical Terminology course at a junior college. The other 24 subjects were recruited from unrelated first-year courses at the same college.

Materials

The stimuli consisted of eighty trimorphemic strings. This stimulus set contained four stimulus groups of 20 words each. Within each group, 10 stimuli were composed of a prefix, a root and a suffix. The other 10 were composed of a root plus two suffixes. All stimuli possessed a frequency of 10 or less in the Kucera and Francis corpus (Kucera and Francis 1967) and ranged in length from 6 to 13 characters. The real non-medical words were selected through a random search blocked for morphological structure, using a developing morphological database system (Libben & Vanderweide 1992). The set of non-medical nonwords was constructed through the recombination of morphemes in the real-word set. The real medical words were matched on frequency and length to the set of non-medical words. The 20 words in this set were selected following a pen and paper pretest with 50 items. In this pretest, 30 first-year university students were asked to indicate for each word on the list, whether they knew its meanings. All stimuli in the final set of 20 items showed an accuracy rate of over 95% on the pretest.

The medical nonwords were formed by recombining morphemes from the final set of real medical words. Here again a pretest was conducted, in this case to ensure that the novel medical forms were plausible. In the pretest, the 20 nonword stimuli were compared to 20 low frequency real trimorphemic medical terms (which we assumed normal subjects would not know). The forty stimuli were printed in random order on a sheet of paper. Thirty-seven first year university students were each presented with the sheet and were asked to decide which words were real medical terms and which were phony. An analysis of the number of responses of 'real' showed no difference between the low frequency real medical words and the words in the novel stimulus set ($\bar{X}1= 10.97$, $\bar{X}2= 10.70$; $t(36)=.667$, two-tailed)

Design

As has been discussed above, the design tested the role of a Group factor that represented the medical terminology student and control groups, and four levels of the within-subjects factor Word Type. An additional within-subjects factor, Presentation Condition, played a crucial role in the experiment.

The Presentation Condition factor was included to increase the sensitivity of the experiment. Each stimulus word was shown under three presentation conditions, which may be conceptually described as morphologically transparent, morphologically opaque, and morphologically neutral. The morphologically neutral condition simply involved the presentation of the stimulus word in normal 24 point typeface (e.g., disclaimer). The morphologically transparent condition was created by presenting the stimulus word such that the second constituent morpheme was in bolded typeface (e.g., dis**claim**er). Because all stimuli were trimorphemic, this manipulation ensured that each morpheme boundary was marked by a change of typeface. Finally, the morphologically opaque condition was also created by presenting medial letters in boldface type. The bolded substring was equal in length to the second morpheme, but in this case began either to the left or right of the actual morpheme boundary (e.g., di**scla**imer or discl**aime**r).

The use of the three presentation conditions made it possible to investigate lexical decision latencies using subject groups as their own controls. The morphologically opaque condition served as a control for whether changes in response latency were simply a result of typeface changes within the string or were actually related to the highlighting of constituent morphemes. We reasoned that if subjects in the M-Group were indeed activating constituent representations, they should show elevated response latencies for novel multimorphemic strings in the morphologically transparent condition as compared to the morphologically opaque condition. The C-Group should also show elevated response times for non-medical nonwords in the transparent condition, but not for medical nonwords. Here we reasoned that if the subjects did not have lexical representations for the constituent morphemes, then the transparent condition would to them be no different from the opaque condition. In both cases, a substring would be bolded, but in neither case would the bolded substring correspond to a lexical representation.

To summarize, the experiment tested three factors: subject group, word type, and presentation condition. These were analyzed in the 2X4X3 design with repeated measures on the last two factors.

Procedure

Subjects were individually. shown stimuli in the center of a Macintosh computer screen in 24 point font. For each stimulus, subjects were required to press a key labelled 'yes' if the stimulus was a word that they had seen before and to press a key labelled 'no' if they had never seen the stimulus. Each testing session consisted of 18 practice trials and 240 experimental trials. In each trial cycle, the subject was first presented with a 'prepare for stimulus' prompt for 500 milliseconds. The screen was then cleared for 500 milliseconds after which the stimulus appeared centered on the screen. The stimulus remained on the screen until the subject pressed a response key. The screen was then cleared for 500 milliseconds before the onset of the next trial cycle. The 240 stimulus presentations were given in a different random order for each subject. In the morphologically opaque presentation condition, half the bolded substrings were offset two characters to the left of the actual morpheme boundary and half were offset two characters to the right. For stimuli which contained two-character prefixes or suffixes, this offset was reduced to single character. The experiment was controlled by a Pascal program which utilized a software millisecond timer.

Results and Conclusions

The dependent measure for all analyses was lexical decision latency, which was calculated from stimulus onset to response.

In all analyses, separate ANOVA's were conducted treating subjects (F_S) and items (F_i) as random factors. In the subjects analysis, Word Type and Condition were treated as within-subjects factors. In the items analysis, Word Type was a between-items factor. Group and Condition were within-items factors. Only correct responses were analyzed and response times below 300 milliseconds and above 1500 milliseconds were removed from the dataset. The response patterns for the C-Group and the M-Group are shown in Figures 1 and 2 respectively.

Overall ANOVA

Both the subjects and items analyses yielded main effects for Presentation Condition ($F_S(2,92)=7.43$, $p<.001$; $F_i(2,152)=3.16$, $p<.05$) and Word Type ($F_S(3,138)=150.6$, $p<.001$; $F_i(3,76)=54.85$, $p<.001$). There was no main effect for Subject Group in either analysis.

Thus on the whole, the M-group and C-group did not differ in response times to the 240 test stimuli (the relatively high average response times could result from the fact that all real word stimuli were multimorphemic and low frequency). The significant effect for Presentation Condition reflects an average lower response time for the neutral condition (906 ms) as compared to the transparent (923 ms) and the opaque conditions (925 ms). Finally, as is apparent in Figures 1 and 2, the main effect of word type reflects higher latencies for 'no' responses as compared to 'yes' responses. All interactions except the Condition X Group interaction were significant in both subjects and items analyses.

Control Group Subjects

The analysis of simples effect and pairwise Tukey comparisons for the the C-Group subjects showed no simple effects of Presentation Condition for any of the word types (all $p>.15$). The Tukey comparisons yielded no differences in 'yes' latencies between the medical words and the nonmedical words (all $p>.05$). However, the C-Group subjects rejected medical nonwords significantly more quickly than non-medical nonwords for all levels of stimulus presentation.

Thus, as Figure 1 displays, the C-Group subjects seemed insensitive to the typeface manipulations across the three presentation conditions. They also showed no sensitivity to the difference between medical words and nonmedical words However, their consistently elevated response times for nonmedical nonwords may reflect the activation of constituent morphemes for those stimuli but not for the constituent morphemes of the medical nonwords.

Medical Group Subjects

As can be seen in Figure 2, the response patterns for the M-Group subjects differed considerably from those of the C-Groups subjects. The M-group subjects showed a significant effect of presentation condition for all word types (all $p< .01$) except the nonmedical nonwords ($p=.47$). For the real word stimuli, no significant differences were found between medical and nonmedical words in either the neutral or opaque conditions. However, for the medical and nonmedical words, the M-Group subjects showed significant differences between the neutral condition and the opaque condition ($p<.05$) and between the transparent condition and both of the other conditions ($p<.01$) for the medical words. For these stimuli, the opaque condition created an interference effect whereas the transparent condition generated a facilitation effect. These findings are consistent with the view that the M-Group's preferred processing mode was through the activation of constituent morphemes. The opaque presentation condition interfered with this mode of processing and therefore resulted in elevated response times. The transparent condition, on the other, hand presented the trimorphemic stimuli as 'partially processed' and thus facilitated the recognition process.

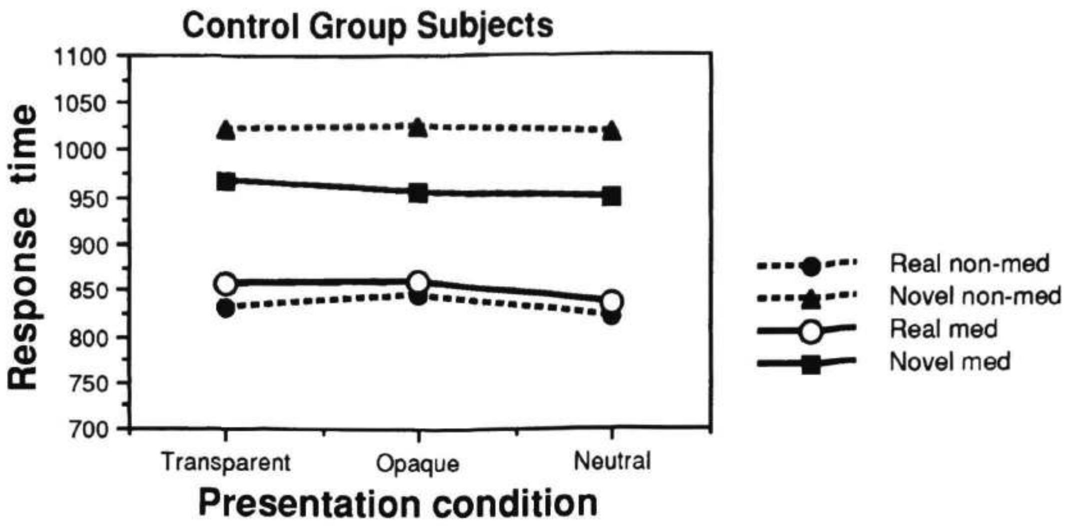

Figure 1: Response times of Control Group subjects for medical words, medical nonwords, non-medical words and nonmedical nonwords under three presentation conditions (transparent, opaque, neutral).

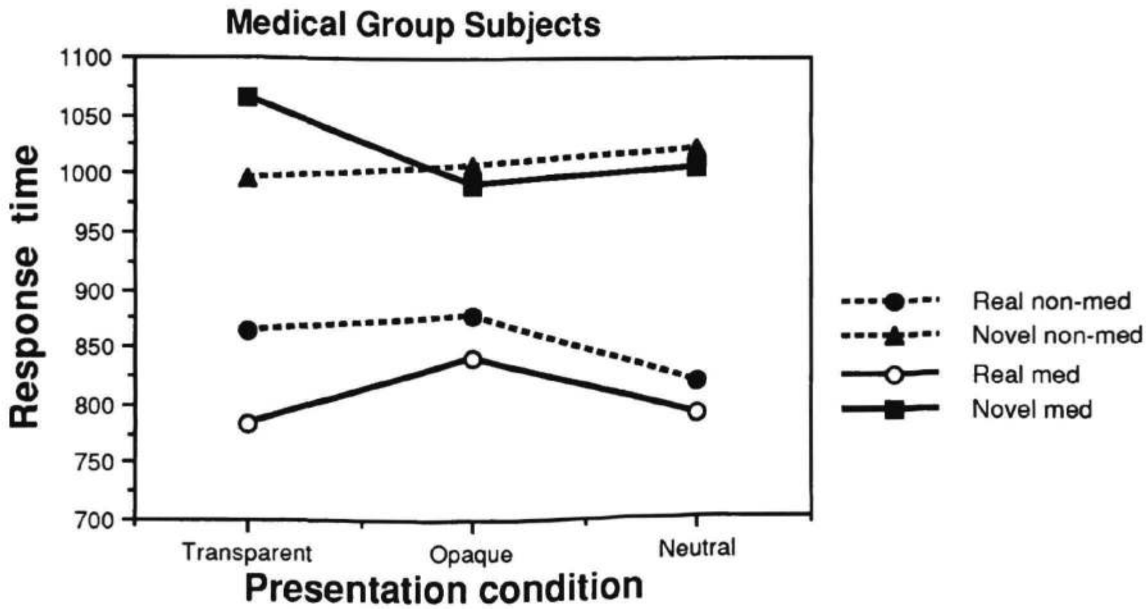

Figure 2: Response times of Medical Group subjects for medical words, medical nonwords, non-medical words and nonmedical nonwords under three presentation conditions (transparent, opaque, neutral).

An examination of the M-Group subjects' response patterns for the nonword stimuli supports this view. Unlike the C-Group subjects, the M-Group subjects showed no difference between the medical nonwords and the nonmedical nonwords in either the neutral or opaque conditions. We have suggested above that the observed significant difference for the C-Groups subjects could be explained by positing that their 'no' responses were not slowed down by the real morphemes in the medical nonwords because, for them, the constituents do not have lexical representations. This explanation predicts exactly the observed data for the M-Group. The Medical Terminology students were trained to acquire semantic representations for the constituent morphemes and thus should process medical nonwords in the manner that they process nonmedical nonwords. As can be seen in Figure 2, the M-Group's response latencies for the medical nonwords in the transparent condition are elevated with respect to both the nonmedical nonwords in the same presentation condition ($p<.01$) and the medical nonwords in the two other presentation conditions ($p<.01$). Here we conclude, as we did for the real medical words, that the transparent condition is offering the subjects partially preprocessed stimuli. In this case, however, the activation of real constituent morphemes makes the 'no' response slower.

Discussion

The experiment reported above has investigated the extent to which lexical decision performance can be affected by differences in subjects' knowledge of constituent morphemes and their experience in the decomposition of multimorphemic words. Our results have indicated that when constituent structure is highlighted (in the transparent presentation condition) the Medical Group subjects show elevated response times in the rejection of novel multimorphemic medical nonwords — an effect that did not obtain for nonmedical nonwords. This result points to the view that the processes involved in word recognition are in fact highly permeable. Unconscious high-speed processing can be influenced by explicit lexical instruction and interacts with the type of presentation to which subjects are exposed.

For real medical words, we again found effects of morphological training and an interaction effect with presentation type. The M-Group subjects showed significant interference for the presentation condition in which morpheme boundaries were obscured. This effect, which was not found among the C-Group subjects suggest that the M-Group's manner of processing words with which they were previously familiar was altered by the morphological training they received. Thus, we conclude that prior to their instruction, the M-Group subjects were equivalent to the C-Group subjects and processed the multimorphemic medical forms as though they had no morphological structure. The result of their training may have been an alteration in how they represent multimorphemic medical terms, how they process them, or both.

We cautiously suggest that the dominant factor in this shift of processing manner is more fundamentally associated with the development of semantic representations for constituent morphemes than with the development of a morphological decomposition strategy for two reasons: (1) The M-Group's elevated response times for the rejection of medical nonwords seems solely attributable to differences in morphemic representations (under the assumption that decomposition occurs for all 'no' responses). (2) If subjects had developed a general strategy of morphological decomposition, the effect of this strategy should have been seen for the nonmedical as well as the medical words. This was not the case. The M-Group subjects showed no difference between the transparent and opaque presentation conditions for nonmedical words and, in general, their performance for nonmedical stimuli was comparable to the performance of the C-Group subjects.

In conclusion, we return to the considerations at the outset of this paper: To what extent do linguistic representations of morphological structure correspond to the psycholinguistic phenomena associated with word recognition? We propose no simple answer to this question and indeed suggest that there may never be one. In the investigation of the processing of multimorphemic words, it is important to keep in mind that words do not have constituent morphemes — people do. As we have found for medical terms, constituent representations may vary from subject to subject within a class of words. They may also differ between word classes. An extension of our findings to multimorphemic forms with bound roots suggest that these may be functionally monomorphemic for most native speakers of English. It is also possible that the differences between inflected and derived words and between prefixed and suffixed stimuli reported in the literature may reduce to the question of semantic representations for constituent morphemes. Inflectional suffixes such as -ing or -ed seem more likely to have independent semantic representations than derivational suffixes such as -al or -ity. Similarly, all of these suffixes seem to carry less semantic load than prefixes such as re- or un-, which native speakers can easily identify as meaning 'again' and 'not' respectively.

Finally, we would like to propose two cautions: If unconscious morphological processing can indeed be influenced by educational background, it may be the case that experiments conducted with University students as subjects yield data that are not representative of lexical processing in the general population. In addition, we may find significant differences across languages resulting from the interaction of a language's morphological structure and cultural educational practices that address it.

Acknowledgements

Preparation of this paper was supported by Social Sciences and Humanities Council of Canada Research Grant 410-90-1502 to G. Libben. The authors are grateful for the collaboration of David Wood, Mount Royal College, in the development of this study.

References

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Learning*, 25, 726-740.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge Mass.: MIT Press.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* New York: Academic Press.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition, 13*, 241-255.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma &

D. Bouwhuls (Eds.), *Attention and Performance X.* Hillsdale, New Jersey: Lawrence Erlbaum.

Kucera, H. & Francis, W. (1967). *Computational analysis of present day American English.* Providence R.I.: Brown University Press.

Libben, G. & Vanderweide, T. (1992, October). *MorphBase.* Paper presented at the Alberta Conference on Language. Banff, Alberta.

Mandelis, L. & Tharp, D. A. (1977). The processing of affixed words. *Memory and Cognition* 5(6), 690-695.

Smith, P. T., & Sterling, C. M. (1982). Factors affecting the perceived morphological structure of written words. *Journal of Verbal Learning and Verbal Behavior,* 21, 704-721.

Sproat, R. (1992). *Morphology and computation.* Cambridge Mass.: MIT Press.

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