LEXICAL DECISION FOR INFLECTED NOUNS*

G. LUKATELA, Z. MANDIĆ, B. GLIGORIJEVIĆ, A. KOSTIĆ, M. SAVIĆ
University of Belgrade

and M.T. TURVEY
University of Connecticut and Haskins Laboratories

Lexical decision times were measured for three grammatical cases of inflected Serbo-Croatian nouns. The grammatical cases occur with different frequencies. Decision times were not related by a unique constant multiplier to the logarithms of the respective case frequencies. The result suggests that a principle of organization in addition to frequency of occurrence is involved in the lexical memory of inflected nouns.

INTRODUCTION

Several investigators have suggested that during reading, the recovery of word information involves a relatively extensive search of lexical memory (e.g., Rubenstein, Garfield and Milliken, 1970; Stanners and Forbach, 1973; Forster and Bednall, 1976). Individual words are said to be represented as lexical entries, with the lexical entries ordered by frequency of occurrence. A search of the lexicon, then, might be construed as beginning at the most frequent entry and searching serially through the list of lexical entries, in accordance with the frequency ordering, until the target entry is determined. If there is no entry, then the search is exhaustive (see Forster and Bednall, 1976).

The focus of this paper is the structure of lexical memory for the Serbo-Croatian language in which inflection is a major grammatical device, perhaps the principal one. Thus for nouns, all grammatical cases in Serbo-Croatian are formed by adding to the root form an inflectional element, namely, a suffix consisting of one syllable of the vowel or vowel-consonant type.

For any given noun the grammatical cases produced by inflection are not equal in their frequency of occurrence. Table 1 is taken from data collected by Dj. Kostić (1965a); it gives the case frequencies for nouns in the singular that are more frequent than nouns in the plural (74% to 26%). We see, in short, that for any given noun of frequency of occurrence (f) in the language, the singular nominative form will appear with a frequency of approximately 0.25f, the singular genitive form will appear with a frequency of approximately 0.20f, and so on.

How might the nouns of an inflected language such as Serbo-Croatian be organized in lexical memory? One hypothesis is that each grammatical case for each noun receives a

This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade, and in part by NICHD Grant HD-01994 to Haskins Laboratories. The authors wish to thank Leonard Katz for his helpful advice on data analysis.

TABLE 1

Case	Symbol	Frequency (per cent)
Nominative	$(CF)_{nom}$	24.55
Genitive	$(CF)_{gen}$	19.90
Dative	$(CF)_{dat}$	1.86
Accusative	$(CF)_{acc}$	13.52
Instrumental	$(CF)_{ins}$	4.70
Locative	$(CF)_{loc}$	8.79

lexical entry and these lexical entries are ordered according to frequency of occurrence. An alternative hypothesis concurs that each grammatical case for each noun receives a lexical entry, but stresses that frequency is not the sole principle of organization. For any given noun the nominative singular is the most frequently occurring grammatical case and it is that which is learned first. The alternative hypothesis might take the form that nominative singulars are ordered in the lexicon according to frequency of occurrence, but that the other grammatical cases for any given noun are subentries to the noun's nominative singular, and these subentries are organized by some principle other than frequency. A simple prediction follows. If the first hypothesis is correct, then the lexical decision ("Is this a word?") latencies for the different grammatical cases of a noun should be determined by frequency of occurrence. However, the lexical decision latencies need not be so determined if the second hypothesis holds.

The present experiment examines Serbo-Croatian nouns from the mid-range of word frequencies (Kostić, 1965b). For each noun, three singular cases were considered: nominative, locative and instrumental. If the noun occurs with frequency (f), then by the first hypothesis, decision time should be related by a unique constant multiplier to the corresponding logarithms of the proportional frequencies, 0.25f, 0.09f, 0.05f (corresponding to the nominative, locative and instrumental, respectively). By the second hypothesis, decision time to the nominative singular should be fastest, but the relation among the decision times should not be accountable for by the proportional frequencies.

METHOD

Subjects

Thirty-nine students from the Psychology Department of the University of Belgrade participated in the experiment. They were experienced with reaction-time procedures.

Materials

The nouns were selected according the the following criteria: (1) easy to read aloud; (2) easily imagined (concrete nouns); (3) only one meaning that was invariant for all grammatical cases; (4) written as alternations of single consonants and vowels. One hundred twenty words were selected for the experiment: 57 masculine nouns, 52 feminine and 11 neuter, corresponding to the proportion of genders in the Serbo-Croatian language.

Non-words were generated as follows. The selected 120 words were listed according to frequency of occurrence. Every other three words in the list were converted into non-words. For nominatives and locatives this was done by changing the first letter. For example, the noun in nominative KIŠA 'rain' was transformed into the nonsense letter string GIŠA. In the locative this noun is KIŠI; the nonsense form was LIŠI. For instrumentals, half of the non-words were produced by changing the first letter and half by changing the last letter or the last two letters. This was done to minimize the influence of the idiosyncratic instrumental endings. For purposes of subsequent analysis it should be noted that the dative and locative for all genders have identical codings and are indistinguishable in the absence of sentential context. Similarly, in the singular, nominative and accusative for masculine and neuter gender are of identical form, whereas in the singular of the feminine gender, the nominative and the accusative are different. In Serbo-Croatian, for all genders, the instrumental is the only unequivocal grammatical case in either the singular or the plural.

The words and non-words were presented as lower case, printed Roman letters (Helvetica Light, 12 point), horizontally arranged at the center of 35 mm. slides.

Procedure

Each of the 120 letter strings was exposed for 1500 msec. in one channel of a three-channel tachistoscope (Scientific Prototype, Model GM) illuminated at 10.3 cd/m^2. Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches further away. The closer button was depressed for a "No" response (the string of letters was not a word), and the further button was depressed for a "Yes" response (the string of letters was a word).

Latency was measured from stimulus onset. The total session lasted for half an hour with a short pause after every eighteen slides.

Design

One hundred twenty stimuli were presented to each subject. Twelve stimuli were used for practice, but were not taken into the final analysis. The subjects were divided into three groups in order to exclude the possibility that the same word, though in different grammatical cases, could be presented to the same subject. Hence, a subject saw one-third of the words and non-words in nominative, one-third in dative, and one-third in instrumental.

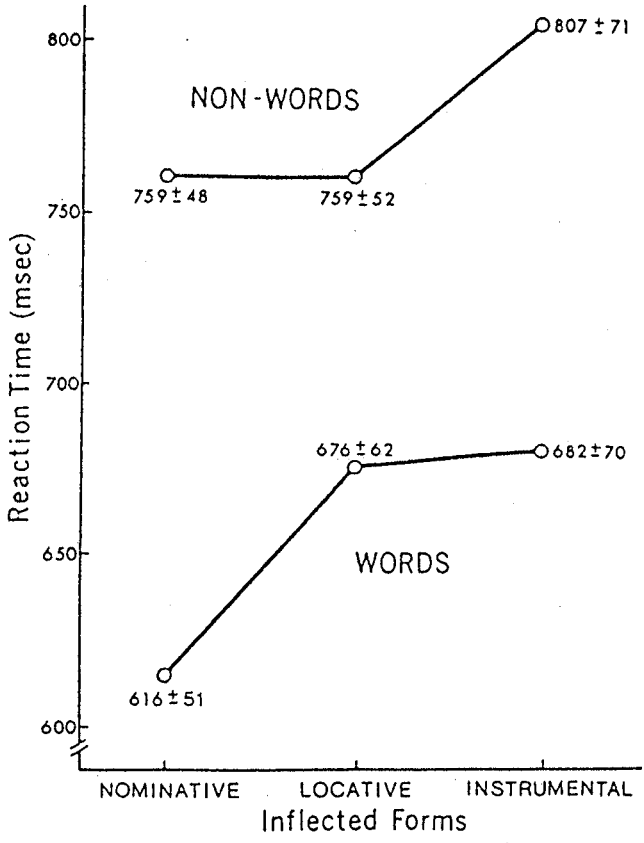

Fig. 1. Lexical decision latency as a function of grammatical case.

RESULTS

The reaction time of each subject to each stimulus was the basic datum for the analysis. If the subject gave an incorrect answer, his average latency for the given class of stimuli replaced the missing data. The number of incorrect decisions was relatively small (2.4%); those responses that were either too fast (less than 300 msec.) or too slow (more than 1500 msec.) were also considered as errors. The data are summarized in Fig. 1.

The reaction times for three inflected forms within each word were compared. A given word in a particular grammatical case was seen by a third of the total number of subjects and, therefore, for the purpose of analysis, the words were divided into three groups of eighteen words each.

The analysis of variance included the three factors: fixed factor – grammatical case,

random factor – subjects and random factor – words. A group of thirteen subjects was nested under a particular grammatical case, while the same eighteen words appeared in three inflected forms under the respective treatments.

The differences between the nominative on one side and the instrumental and locative on the other are statistically significant (see Clark, 1973) F' (1,32) = 5.4, $p < 0.05$ and F' 1,35) = 4.05, $p < 0.05$ respectively, whereas the difference between locative and instrumental was not significant.

DISCUSSION

The results of the experiment demonstrate that, in lexical decision, the latency to nouns in the nominative case is shorter than to nouns in the locative and instrumental cases, and that non-words take longer to classify than words.

As depicted in Fig. 1, the two reaction time plots – one for words and the other for nonwords – display two different patterns. Let us first address the less significant issue of why the latencies to the "instrumental" non-words were longer than those to the "nominative" and "locative" non-words. The relative difficulty with the "instrumental" non-words most probably stems from the fact that the "instrumental" nonsense letter strings were, on average, one letter longer than the other nonsense letter strings. We recall that an "instrumental" non-word was produced by changing one letter in a noun that was grammatically inflected in the instrumental case. We recall, also, that the characteristic ending of the instrumental case in Serbo-Croatian consists of two letters (of the vowel-consonant type) and that the characteristic endings of other cases in the singular consist of a single letter (a vowel). As a result of the transformation rules, all non-words in the experiment were orthographically legal. The "nominative" non-words were mono- or bisyllables. The "locative" as well as the "instrumental" non-words were bi- or trisyllables, but each "instrumental" non-word had one letter more than its "dative" mate. These facts and the data in Fig. 1 suggest that the effect of number of syllables on lexical decision time for non-words was not significant. On the other hand, the effect of number of letters in a nonsense string proved to be significant. This finding is in agreement with the results of Forster and Chambers (1973) and Fredericksen and Kroll (1976).

Further comment on the non-word data is unnecessary. Let us focus on the main issue of why the latencies to the inflected words did not follow the general pattern that was predicted by the word-frequency effect. In the lexical decision task, the reaction time (RT) is inversely proportional to the word frequency, (f). In the first approximation there is a linear regression of a general form:

$$(1)\quad RT = -A \ln f + B$$

where A and B are the regression coefficients that depend on the number of letters in the word. For English five-letter words, given their frequency of occurrence (Kučera and Francis, 1967), it has been found[1] experimentally that the appropriate numerical values

[1] *L. Katz, personal communication.*

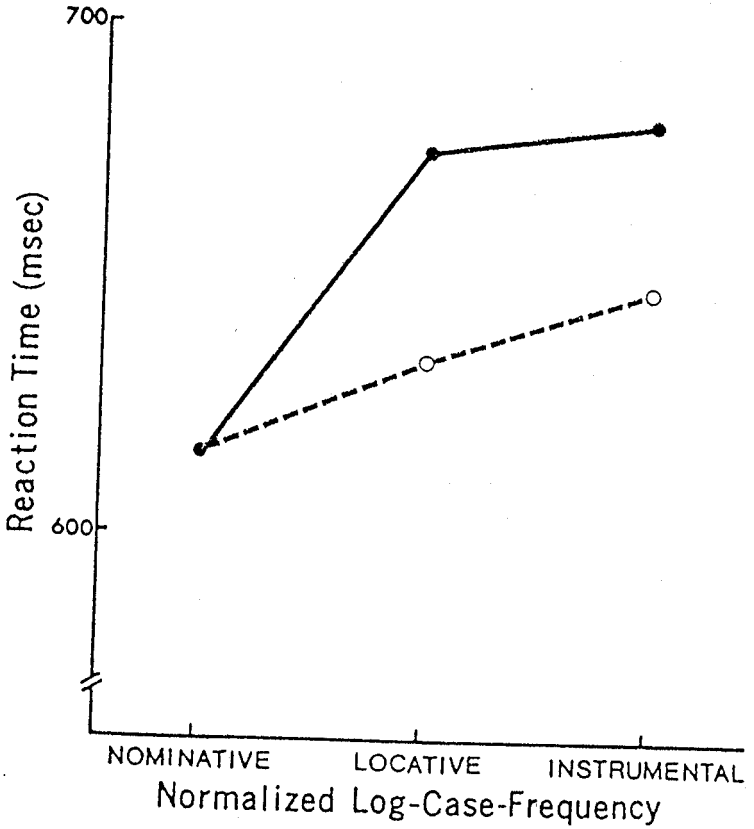

Fig. 2. Contrast between obtained (solid line) and predicted (broken line) latencies as a function of normalized log case frequency.

of the regression coefficients are: A = 17.78 and B = 644.

In the present experiment the average number of letters (when averaged across all nouns in all inflected forms) was about five per word. Therefore, if the reaction time for inflected forms was governed uniquely by the case frequency CF, then the slope of the function relating RT to log CF should be about 17.78, as shown by the dashed line in Fig. 2. The zero-intercept of the dashed line, in agreement with out data, was set at B = 616 msec.

The experimental data in Fig. 2 are represented by black dots and, for convenience, are connected by solid lines. The suggestion is that the solid curve differs from the dashed-line curve not only quantitatively, but also qualitatively. Hence, the plots in Fig. 2 suggest that the word-frequency effect cannot explain the experimental results.

There is, of course, some theoretical possibility that the numerical value of the slope coefficient A, as plotted in Fig. 2, is not appropriate for Serbo-Croatian words. What we

need, therefore, is a stronger proof that the experimental data and the data predicted uniquely by the word-frequency effect are significantly different for any arbitrary value of the regression coefficients.

The data of Table 1 show that the case frequencies of the nouns in nominative, locative and instrumental relate as follows:

$$(2)\quad (CF)_{nom} > (CF)_{loc} > (CF)_{ins}$$

In a lexical-decision task the case frequencies of nominative and accusative for masculine and neuter gender have to be compounded. In the experiment the number of nouns in masculine and neuter gender was 68, as compared with 52 nouns in feminine gender. The joint frequency of occurrence of the unequivocal and equivocal nominative forms, when averaged across all of the 120 nouns, results in the compounded nominative-accusative case frequency: $(CF)_1 = 31.31\%$. Similarly, we have also to compound the case frequencies of the locative and dative for all nouns. The compounded locative-dative case frequency is: $(CF)_2 = 10.65\%$.

If it were true that the mean reaction time and the case frequency were related by equation (1), then between the reaction time to the compounded nominative case $(\overline{RT})_1$ and the reaction time to the compounded locative case $(\overline{RT})_2$, the following hypothetical relation should hold:

$$(3)\quad (\overline{RT})_2 - (\overline{RT})_1 = A \ln \frac{(CF)_1}{(CF)_2}$$

where A is an arbitrary constant; $(CF)_1$ is the compounded case frequency for nominative and accusative, and $(CF)_2$ is the compounded case frequency for locative and dative.

Similarly, for the difference between the mean reaction time to the instrumental, $(\overline{RT})_{ins}$, and the mean reaction time to the compounded nominative $(\overline{RT})_1$, the predicted hypothetical relation would be:

$$(4)\quad (\overline{RT})_{ins} - (\overline{RT})_1 = A \ln \frac{(CF)_1}{(CF)_{ins}}$$

By dividing equation (3) by equation (4) we obtain:

$$(5)\quad \frac{(\overline{RT})_2 - (\overline{RT})_1}{(\overline{RT})_{ins} - (\overline{RT})_1} = \frac{\ln \dfrac{(CF)_1}{(CF)_2}}{\ln \dfrac{(CF)_1}{(CF)_{ins}}}$$

If we substitute the numerical values of the mean RTs from Fig. 1 into the left side of equation (5), we find that the ratio of the normalized $(\overline{RT})$ difference is:

$$\frac{(\overline{RT})_2 - (\overline{RT})_1}{(\overline{RT})_{ins} - (\overline{RT})_1} = \frac{676 - 616}{682 - 616} = 0.91$$

On the other hand, if we substitute the numerical values of the compounded case frequencies as well as the instrumental case frequency into the right side of equation (5) we find that:

$$\frac{\ln \frac{(CF)_1}{(CF)_2}}{\ln \frac{(CF)_1}{(CF)_{uns}}} = 0.56$$

Thus, we conclude that the hypothetical equation (5) is not correct: the left side is numerically about two times as large as the right side.

The preceding mathematical analysis supports the hypothesis that the longer latency to inflected words cannot be accounted for by the difference in the frequency of occurrence of the grammatical forms. We are led, therefore, to the tentative conclusion that frequency of occurrence is not sufficient to capture the lexical organization of the grammatical cases of inflected nouns. If the nominative singular is the base form, then the other grammatical cases of a noun may be represented in the lexicon as subentries to the noun's nominative singular, as suggested above. A different view, however, is that the nouns are not represented in the lexicon with their grammatical cases. Instead, root morphemes for all nouns, as well as the small set of inflectional morphemes, are stored in the lexicon. Appropriate combinations of a root and inflections are determined by separately stored syntactic rules. Subsequent research will be directed to these contrasting perspectives.

REFERENCES

CLARK, H.H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. verb. Learn. verb. Behav.*, 12, 335-59.

FORSTER, K.I. and CHAMBERS, S.M. (1973). Lexical access and naming time. *J. verb. Learn. verb. Behav.*, 12, 627-35.

FORSTER, K.I. and BEDNALL, E.S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition*, 4, 53-61.

FREDERIKSEN, J.R. and KROLL, J.F. (1976). Spelling and sound: approaches to the internal lexicon. *J. exp. Psychol.: Human Perception and Performance*, 2, 361-79.

KOSTIĆ, Dj. (1965a). Frequency of occurrence of words in Serbo-Croatian. Unpublished paper, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.

KOSTIĆ, Dj. (1965b). The structure of usage value of grammatical forms in Serbo-Croatian: nouns. Unpublished paper, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.

KUČERA, H. and FRANCIS, W.N. (1967). *Computational Analysis of Present-Day American English* (Providence, R.I.).

RUBENSTEIN, H., GARFIELD, L. and MILLIKAN, J.A. (1970). Homographic entries in the internal lexicon. *J. verb. Learn. verb. Behav.*, 9, 487-94.

STANNERS, R.F. and FORBACH, G.B. (1973). Analysis of letter strings in word recognition. *J. exp. Psychol.*, 98, 31-5.