

Mental Lexicon and Derivational Rules

Marijan Palmović and Antonija Maričić

Laboratory for Psycholinguistic Research, Department for Speech and Language Pathology, Faculty of Education and Rehabilitation Sciences, University of Zagreb, Zagreb, Croatia

ABSTRACT

Lexical decision task in an event-related potential experiment was used in order to determine the organization of mental lexicon regarding the polymorphemic words: are they stored as unanalyzable items or as separate morphemes? The results indicate the later: while monomorphemic words elicit N400 component, usually related to lexical-semantic processing, prefixed words and prefixed pseudo-words elicit left anterior negativity (LAN), usually related to grammatical (morphosyntactic) processes. These components indicate that the speakers apply grammatical (i.e. word-formation) rules and combine morphemes in order to obtain lexical meaning of the prefixed word.

Key words: mental lexicon, lexical decision task, prefixation, event-related potentials, N400, LAN

Introduction

Native speaker usually understands more than 20,000 words of his language. Vocabulary size is important for a foreign language learner, as well. It is claimed, for example, that for a good score on TOEFL (Test of English as a Foreign Language) a candidate has to understand around 10,000 English words. The size of this body of knowledge requires some structure and this structure allows for quick recognition or retrieval of words. Indeed, the speed of word recognition and retrieval is regarded as sufficient evidence for the existence of such structure and internal organization of mental lexicon, storage for words in our minds, has been a major topic of psycholinguistics for decades. The first model that dealt with word recognition process¹ differentiated between word forms (*logogens*) stored in the *logogen system* and meanings stored in the *cognitive system*. The model is in a way parallel to the Wernicke-Geschwind model in which sound images and their meanings were considered to be separate (and the neural basis for the two aspects of word knowledge was considered to be different – Wernicke's area translates auditory input into phonological form which can then access semantic networks that are distributed in the brain). Other models focus on form-to-meaning mapping, for example, *cohort model*², *Neighborhood Activation Model*³ or connectionist models⁴. In short, these models explain word recognition in terms of similarity and frequency. The two terms play different roles in different models: while similarity can be purely

phonological (or orthographic) in some models, semantic similarity counts in others; while word frequency determines the speed or easiness of accessibility, the frequency of the similar words can interfere.

Morphologically complex words were usually studied in psycholinguistics in order to provide evidence for a decision between dual-route⁵ and connectionist models. In dual-route models it is assumed that the roots and affixes are stored separately. The application of a rule that combines them overrides frequency: a rule is applied equally to frequent and non-frequent items. On the other hand, as there are no rules in connectionist models, it is the frequency that determines the activation of a complex word if only a part of it is available, with the recall of the whole word-form stored in the memory as a consequence.

The method of event-related potentials (ERP) was used to investigate the organization of mental lexicon in two major paradigms: lexical decision tasks and priming paradigms. In this paper the lexical decision task will be employed. In mental lexicon studies this task has been employed mainly for studying inflected forms such as irregular vs. regular past tense in English⁶. Only few studies address derivational morphology. In McKinnon et al. (2003)⁷ prefixed English pseudo-words formed by prefixing a non-productive verbal stem elicited reduced N400 component in comparison to non-prefixed pseudo-words. As confirmed in numerous studies, the N400 component

can be associated with lexical/semantic processing⁸, therefore, the reduced N400 amplitude was interpreted as evidence in favour of morphological decomposition of complex words in the mental lexicon, i.e. as a consequence of successful access to the stored morphemes, not words. A more complicated pattern was obtained in a study by Janssen et al. (2006)⁹ in which German suffix *-ung* was used for words that violated a morphosyntactic rule, while suffixes *-keit* and *-heit* having complementary distribution regarding to the prosodic features of the stem violated only prosodic constraints on the suffixes. A »N400-like« response was obtained, but with different latencies and different distribution for the two types of violation.

Derivational morphology is a very productive part of Croatian grammar: from a given root it is usually possible to derive a number of words that are usually translated as derivationally unrelated words in other languages. To take an example: from the noun *bol* »pain« many words are derived: *bol-nica* »hospital«, *bol-est* »disease«, *bol-ničarka* »nurse«, *bol-estan* »ill«, etc. Croatian derivational morphology has been studied psycholinguistically in a number of studies, most explicitly (and relevant for this paper) in a study that addressed prefixation in typical and impaired language development¹⁰. The results of this study indicated differences between the two groups of children: while children with typical language development committed mainly morphological errors that usually involved the wrong choice of prefixes, children with language impairment committed mainly lexical errors (in general, their error pattern was more dispersed over all language components).

In order to obtain more detailed data on lexical storage in Croatian an electrophysiological study that involved prefixed words and pseudo-words was conducted. The study is designed to provide insight into the organization of mental lexicon in Croatian: are the words stored as unanalyzed items in many of their forms or are they store separately, as single morphemes?

Materials and Methods

The ERP experiment consisted of four experimental conditions: (1) *words*, which consisted of one hundred two syllabic nouns. All nouns were high-frequency words with relative frequency (in one million words corpus) between 0.15 and 0.0035, according to Moguš et al. (1999)¹¹. (2) *pseudo-words* condition consisted of one hundred two syllabic pseudo-words phonologically similar to the words in the *words* condition. (3) *prefixed-words* condition consisted of one hundred two-syllabic nouns derived by prefixation of the verbal root as in *po-gled* »a look« or *pro-rez* »a slit«. (3) *prefixed pseudo-words* condition consisted of one hundred two-syllabic pseudo-words also derived by prefixation of the verbal root, but with the wrong choice of the prefix, as in *na-gled* or *o-rez*.

The logic of the experiment was rather simple: words and pseudo-words differ in the N400 component with

pseudo-words eliciting strong N400 effect. If morphologically complex (prefixed) words are stored as unanalyzed items, in *prefixed pseudo-words* condition similar strong N400 effect should be obtained while prefixed words should elicit no such effect. If prefixes and roots are stored separately, smaller N400 effect should be expected in the *prefixed pseudo-words* condition due to the separate lexical retrieval of »legitimate« Croatian prefixes and roots, i.e. the results should be similar to the *prefixed words* condition. The same logic was used in the mentioned ERP experiments with derivational morphology.

Twelve volunteers took part in the experiment, 5 males and 7 females. They were all students of the University of Zagreb, age between 21 and 27 (mean age=24, SD=1.9). All of them were right-handed, healthy and of normal or corrected-to-normal vision. They reported no neurological problems and none of them was on any medication. They received no credit for taking part in the experiments.

The participants were seated in a darkened and electrically shielded room (Holland Shielding System, MIL STD-285). The stimuli were presented visually, in the middle of the 19" computer screen. The stimulus words were presented in random order. Each of them was presented for 500 ms with the inter-stimulus interval of 1200 ms. After each 50–60 stimuli a black screen was presented to reduce fatigue. The experiment lasted for 12 minutes. The participants had to make a lexical decision by pressing a button (this information was, however, not used because only automatic processes were of interest here – overt action of the participant merely served to keep his/her attention during the experiment).

Continuous EEG was recorded on a NeuroScan Nu-Amps 40-channel amplifier. The signal was recorded from 36 sites using NeuroScan QuickCaps with standard electrode positions (32 recording channels and 4 channels for ocular artifact reduction). The recording was referenced to the average reference and filtered with high-pass filter set up on 0.01 Hz (minimal value for the system) and low-pass filter on 70 Hz. The signal was digitalized with the sampling rate of 1000 Hz and 22 bit resolution. Off-line averaging and analysis was done using NeuroScan Edit software. The average was obtained in the interval –100 – 1000 ms around the stimulus with the pre-stimulus interval taken as a baseline. For statistical analysis the data were exported into SPSS.

Results and Discussion

The results were somewhat surprising and did not follow the simple pattern of reduced N400 for prefixed pseudo-words, as in the mentioned study by McKinnon et al.⁷. The results show two distinct components: N400 for *words* and *pseudo-words* and Left Anterior Negativity (LAN) for *prefixed words* and *prefixed pseudo-words*. The N400 effect is stronger for the pseudo-words, as expected (Figure 1).

LAN is visible only on the left frontal and central electrodes F7, F3, Fz (Figure 2a). It cannot be observed on

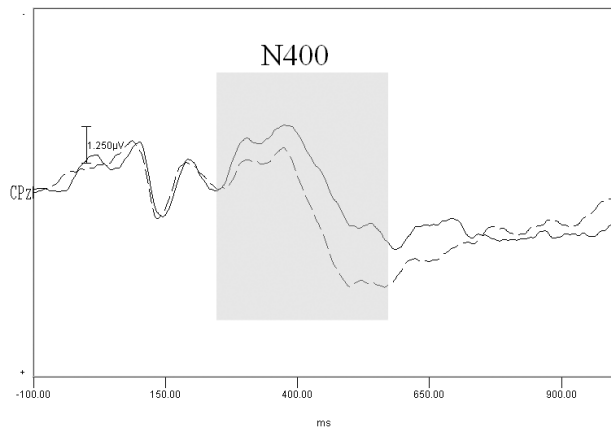


Fig. 1. N400 obtained in the words (dotted line) and pseudo-words condition.

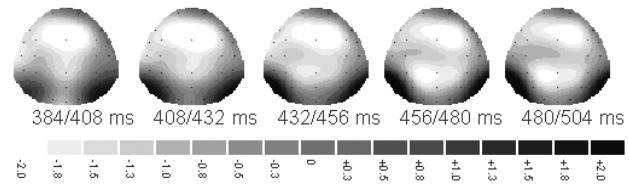


Fig. 3. The scalp distribution for N400 in the pseudo-words condition.

central and parietal electrode sites, typically for N400. On the other hand, LAN is distributed more frontally, with the visible difference between the left and right side, with the left side being more negative (lighter) than the right side (darker), as shown on Figure 4.

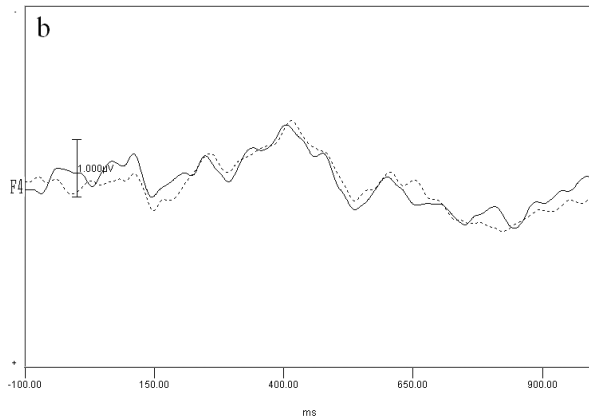
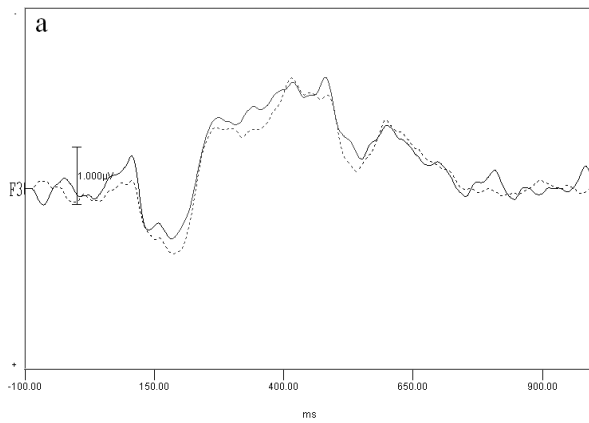


Fig. 2. Results for prefixed words (dotted line) and prefixed pseudo-words condition on left (a) and right (b) frontal electrodes.

the right frontal electrodes (Figure 2b). The difference between *prefixed-words* and *prefixed pseudo-words* condition is rather small, i.e. the LAN effect is rather small.

However, the distribution of the potentials clearly shows the difference between the obtained components. Figure 3 shows broad negative wave over the frontal,

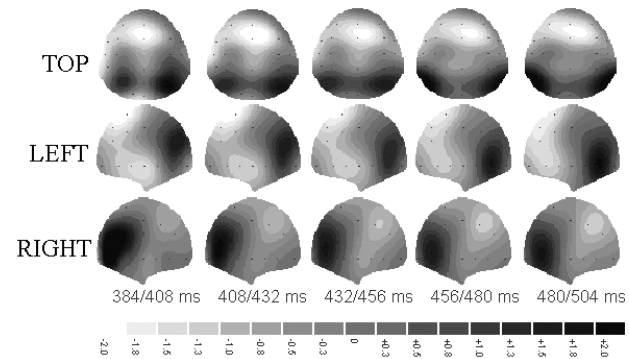


Fig. 4. The scalp distribution for the prefixed pseudo-words condition.

Finally, prefixed words and prefixed pseudo-words elicited late positive wave over the parietal electrode sites, P600 (Figure 5). P600 or »Syntactic Positive Shift«, as sometimes called¹², is usually regarded as a measure of integration costs or repair or reanalysis processes in sen-

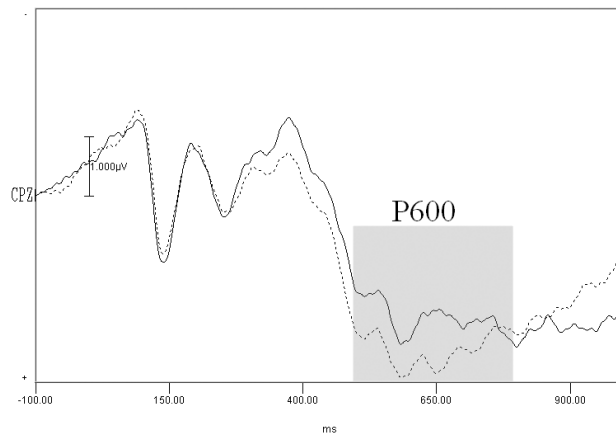


Fig. 5. P600 in prefixed words (dotted line) and prefixed pseudo-words condition.

tences, although morphological or even orthographic violations may elicit it, as well¹³. Therefore, it could be predicted that P600 would be larger in the *prefixed pseudo-words* condition. This is obviously not the case and calls for an explanation. It might be the case that the negative wave simply did not fade away at the latencies around 600 ms. Indeed, in the later interval, around 750 ms, as the potential falls back to the baseline, positivity is larger for the *prefixed pseudo-words* condition.

Statistical analysis (ANOVA) reveals statistically significant difference between conditions on the relevant electrodes (electrode \times condition) in the 350–450 ms window in which both N400 and LAN are present. The analysis was performed on Cz electrode for N400 and F3 electrode for LAN. Statistically significant difference was found on Cz electrode ($F(3,392)=26.17$ $p<0.01$) and on F3 ($F(3,392)=31.41$ $p<0.01$). Post hoc tests (Scheffe) were performed as well in order to show which conditions differ on the given electrodes. Post hoc tests proved to be informative: on the Cz electrode the *pseudo-words* condition significantly differs from all other conditions, while other conditions differ only below the significance level of $p=0.05$. On the F3 electrode the *prefixed pseudo-words* condition significantly differs from other conditions with the significance level $p=0.01$. On the other hand, *pseudo-words* condition does not significantly differ from *words* and *prefixed words* condition while *words* and *prefixed words* differ between themselves with the significance level $p=0.05$. This pattern of statistical results that relates prefixed pseudo-words with F3 and pseudo-words with Cz electrode makes a good ground for the claim that the obtained waveforms are not some variants of N400, but two distinct components. In this respect the small, but statistically significant difference between words and prefixed words on F3 suggests that some grammatical operation has left its trace as LAN in the *prefixed-words* condition, as well.

There is an important difference between the *words* and *pseudo-words* conditions on one side and *prefixed words* and *prefixed pseudo-words* condition on the other. While N400 effect is prominent, LAN effect is barely visible for the latter conditions. The interpretation is rather straightforward: while words and pseudo-words differ in terms of the existence of the items in the Croatian lexicon – and this is reflected in the lexical/semantic component such as N400, prefixes and roots used for stimuli in the later pair of experimental conditions are equal in both experimental conditions in this respect, i.e. they do not differ in terms of the existence in Croatian, in short, they exist. It is only the combination of these morphemes that does not exist in Croatian. It may be even stated that since the LAN effect is rather small here, this can be evidence for the prefixation as psycholinguistically very productive part of Croatian morphology. Croatian speakers, it seems, try to access the meaning on the basis of the meaning of the morphemes, regardless of the existence of the lexical item as a whole in the Croatian lexicon. The statistical difference between *words* and *prefixed words* conditions found on F3 speaks in favor of such an inter-

pretation, as stated above. Further on, a pseudo-word is only phonologically possible Croatian word, but prefixed pseudo-word, as used in this experiment, is more: it is morphologically and, to a certain degree even semantically possible, as well. It is less clear what rule prefixed pseudo-words violate: it may be semantic incoherence of the combination, but also some part of the aspectual meaning of the verbal root (or its *Aktionsart*) which makes the combination non-existent (e.g. ingressive meaning of the prefix with the root that has a punctual meaning). In that case it is not odd that LAN (and not N400) has been obtained with this sort of stimuli; the prefixed pseudo-words used in this experiment in fact violated their logical form, i.e. the features they enter syntactic relations with.

Conclusion

Different patterns of electrophysiological response to words, pseudo-words (N400), prefixed words and prefixed pseudo-words (LAN), as obtained in this experiment, indicate that prefixes and roots are stored separately in the mental lexicon. This result is different from the results obtained in similar experiments in English and German and provides a psycholinguistic evidence for a typological trait of Croatian: that derivational morphology is a very productive part of Croatian grammar. If this result may be generalized and used for making a claim regarding the mental lexicon, it could be stated that morphemes (roots and affixes in general) are stored separately and that speakers employ rules to combine them in order to access a particular form of the word. This is consistent with the dual-route models. However, further evidence is needed to corroborate this claim. In particular, future work should, first, try to manipulate the P600 component in order to obtain more information about the integrative processes in, for example, compound words. Second, frequency should be taken into account: frequent and less frequent prefixed words might be processed differently; if frequent words are accessed faster, frequent prefixed words might be stored as wholes. Finally, developmental data and data from different populations can be particularly informative especially due to the mentioned difference in processing of prefixed word forms found in children with Specific Language Impairment and children with typical language development. Therefore, the results obtained in this experiment can be regarded as initial research into the organization of mental lexicon in Croatian using a brain imaging technique.

Acknowledgements

The study was financed by the project funded by Croatian Ministry of Science, Education and Sport No. 013-0131484-1488 (*»Higher cortical functions and language: Developmental and acquired disorders«*, Senior researcher Melita Kovačević, professor)

REFERENCES

1. MORTON J, Psych Rev, 76 (1969) 165. — 2. MARSLEN-WILSON WD, WELSH A, Cog Psych, 10 (1978) 29. — 3. LUCE PA, PISONI DB, Ear & Hearing, 19 (1998) 1. — 4. MCCLELLAND JL, ELMAN JL, Cog Psych, 18 (1986) 1. — 5. PINKER S, Science, 253 (1991) 530. — 6. MORRIS J, HOLCOMB PJ, Cog Brain Res, 25 (2005) 963. — 7. MCKINNON R, ALLEN M, OSTERHOUT L, NeuroReport, 14(6) (2003) 883. — 8. KUTAS M, HILLYARD SA, Nature, 307 (1984) 161. — 9. JANSSEN U, WIESE R, SCHLESEWSKY M, J. of Neuroling, 19 (2006) 466. — 10. KOVAČEVIĆ M, KUVAC J, CEPANEC M. [in Croatian] Je li glagolska prefiksacija morfološka ili leksička kategorija? In: 3rd Croatian Logopaedic Congress, Dubrovnik, Sept. 28 – Oct. 1, 2005. — 11. MOGUŠ M, BRATANIĆ, M, TADIĆ M, Hrvatski čestotni rječnik [in Croatian] (Zagreb: Školska knjiga, 1999). — 12. OSTERHOUT L, MCLAUGHLIN J, BERSICK M, Trends in Cog Sci, 1 (1997) 203. — 13. MÜNTE TF, HEINZE HJ, MATZKE M, WIERINGA BM, JOHANNES S, (1998) Neuropsychol, 36(3) (1998) 217.

M. Palmović

*Laboratory for Psycholinguistic Research, Department for Speech and Language Pathology, Faculty of Education and Rehabilitation Sciences, University of Zagreb, Zvonimirova 8, 10000 Zagreb, Croatia
e-mail: palmovic@erf.hr*

MENTALNI LEKSIKON I TVORBENA PRAVILA

SAŽETAK

Organizacija mentalnog leksikona ispitivala se metodom kognitivnih evociranih potencijala u zadatku leksičke odluke. Željelo se odgovoriti na pitanje jesu li riječi u mentalnom leksikonu pohranjene kao gotove cjeline ili su pojedini dijelovi, tj. morfemi pohranjeni posebno, s tim da govornik pojedinu riječ priziva uz primjenu gramatičkih pravila. U ovome eksperimentu zadatak leksičke odluke sastojao se od govornikova prepoznavanja riječi u četiri skupine podražaja. Prvu su činile dvosložne riječi, npr. *more*. Drugu su skupinu činile dvosložne pseudo-riječi, npr. *osmar*. Treću i četvrtu skupinu činile su prefingirane riječi i pseudo-riječi, npr. *po-gled* i *na-gled*. Dok se pseudo-riječima postigao efekt N400, tj. komponenta N400 veća je za pseudo-riječi, nego za riječi, prefingirane pseudo-riječi dale su slab efekt lijeve anteriorne negativnosti (*eng. left anterior negativity – LAN*). Budući da se N400 povezuje s obradom na leksičko-semantičkoj, a LAN s obradom na morfosintaktičkoj razini, zaključuje se da su riječi u mentalnome leksikonu pohranjene podijeljene na morfeme i da se prilikom njihova prizivanja primjenjuju gramatička pravila, u ovom slučaju, pravila tvorbe i da ona nisu produktivna u lingvističkome, nego i u psiholingvističkome smislu.