PROCESSING OF INFLECTED NOUNS AND LEVELS OF COGNITIVE SENSITIVITY

Aleksandar Kostić[1] and Jelena Mirković

Laboratory for Experimental Psychology
University of Belgrade, Serbia, Yugoslavia

In a series of experiments with lexical decision task it was demonstrated that processing of inflected Serbian noun forms is determined by the amount of information (bits) carried by those forms. The amount of information is derived from form's probability and number of syntactic functions/meanings carried by a form. Form's probability, on the other hand, is specified within gender paradigm (e. g. what is the probability of suffix x to be attached to a feminine noun?) by summing up probabilities of cases shared by a given inflected form. Within a paradigm of feminine nouns, however, there are number of subparadigms that differ in case distribution of their inflected forms and, by the same token, in distribution of the amount of information distribution. Previous studies have shown that the amount of information derived from probabilies of inflected forms derived from the dominant pardigm account for almost all processing variability. In this study we investigate whether processing of inflected forms from the non-dominant paradigm is affected by its probability distribution or by probability distribution of the dominant paradigm. The outcome of the experiment indicated that processing latencies to inflected forms are determined by probabilies derived from the dominant subparadigm.

Key words*: psycholinguistics, inflectional morphology, nouns, lexical decision*

One of the principal problems of research dealing with the inflected morphology is to account for variability in processing time to different forms of the *same* affixed word. There are two major approaches to this problem. One is based on the assumption that affixed words are represenetd in the lexicon as a whole and that processing variability is related to frequency of an affixed word (Manelis & Tharp, 1977; Rubin, Becker & Freeman, 1979; Kempley & Morton, 1982; Cutler, 1983; Butterworth, 1983; Henderson, 1985). The other approach assumes that

---

[1] Author's address: akostic@f.bg.ac.yu

morphologically complex words are represented through their constituents, therefore, processing of an affixed word should imply its decomposition into a base form and affix. The lexical search for the base form and affix is also frequency biased (Taft & Forster, 1975, 1976; Mackay, 1978; Taft, 1979a, 1979b; 1981; Jarvella & Meijers, 1983; Allen & Badecker, 1999; Badecker & Allen, 2002). In other words, both approaches assume that processing variability is due to affix frequency. Applied to processing of inflected morphology, time to processes an inflected word should therefore depend on inflectional suffix frequency.

In the present study we investigate some aspects of the processing of inflected forms of Serbian feminine nouns. Before we address this issue in more detail, we give a brief outline of Serbian noun system.

AN OUTLINE OF THE SERBIAN NOUN SYSTEM

Serbian nouns appear in three genders and seven cases singular and plural,[2] marked by inflectional suffixes. While noun can cross case and grammatical number, it can not cross grammatical gender, i.e. individual noun can be of one gender only. Morphological transformations of Serbian nouns are standardly classified into four declensions. The first declension includes regular masculine nouns that end with a consonant. It also includes neuter nouns whose base form ends with a consonant, with the vowels *o* and *e* being attached (e.g. *sel-o* /*villag*e/ viz. *sel-a*, *sel-u*, *sel-om*, etc.; *polj-e* - /*field*/ viz. *polj-a, polj-u, polj-em* etc.). The second declension includes neuter nouns whose base form ends with a vowel, while their inflectional suffix contains consonant and vowel (e. g. *ime* /*name*/ viz. *ime-na, ime-nu, ime-nom* etc). The third declension refers to regular feminine nouns that end with the vowel *a* in the nominative form and irregular masculine nouns that end with the same vowel (e.g. *žen-a* (F) - /*woman*/; *sudij-a* (M) - /*judge*/). Finally, the fourth declension includes irregular feminine nouns that end with a consonant (e.g. *strast* - /*passion*/).

Although in most cases nouns of a particular declension share morphological transformations, suffixes for nouns belonging to the same declension are not always identical. There are some suffix differences between masculine and neuter nouns (the nominative and the accusative plural), in spite of the fact that they belong to the same declension. There are also morphological differences within gender. Thus, for example, the genitive plural for the feminine nouns that have two consonants prior to the final vowel *a* (e.g. *tabla* /*board*/) ends with a suffix *i* (*tabl-i*), while for the feminine nouns that have only one consonant prior to the final vowel it is identical to the nominative singular (e.g. *žaba- frog*) (see Appendix 1). Likewise, there are morphological differences between animate and inanimate masculine nouns.

As indicated in Table 1 and Appendix 1, several cases of a particular gender can be comprised within a single noun form. Thus, for example, feminine noun form *žabe* comprises the genitive singular, the nominative plural and the accusative

---

[2] Most linguists agree that vocative should be attached to the nominative case.

plural. Similar examples can be found for other forms in all three genders. This property of Serbian nouns requires a distinction between ***inflected form*** and ***case***, since a form need not be always morphologically transparent for case (see Table 1 and Appendix 1). Thus, when visually presented in isolation, a Serbian noun is equivocal for case and grammatical number. While noun form of a particular gender can be equivocal for case and number, inflectional suffix per se can be also equivocal for gender and word type. Take, for example, the suffix ***i***. If attached to the feminine nouns it specifies the dative and the locative singular, if attached to the masculine nouns it specifies the nominative plural. The same suffix attached to adjectives and possessive pronouns specifies masculine nominative plural, but if attached to verbs it specifies the third person singular present tense. Hence, the accounts based on the assumption that processing variability of inflected word forms is due to suffix frequency have to specify the grammatical domain in which the suffix frequency is to be estimated.

Among the three attributes of nouns, case is the most complex one. The principal property of case is that it encapsulates a number of potential syntactic functions and meanings, which are realized in the sentence context. Take, for example, the nominative case. It can modify subject and predicate roles as in the following sentences: ***Prijatelj*** *je došao* (***The friend*** *has come*) and *Petar je* ***učitelj*** (*Peter is a* ***teacher***). Or take the Serbian accusative, which, in addition to its most common object role, can encompass a vast number of meanings like *time* (***Zoru*** *je proveo čekajući ga /He spent the* ***morning*** *waiting for him/*), *place* (*Popeo se na* ***planinu*** */He climbed the* ***mountain***/), *cause* (*On je odgovoran za njihovu* ***nesreću*** */He is responsible for their* ***tragedy***/), etc. Case thus serves as a syntactic nucleus with variety of potential syntactic functions and meanings.

Serbian noun cases differ both in the number of functions and meanings they encompass and probability of occurrence (Kostić, Đ., 1965a; 1965b) (Appendix 1). It should be noted that notions of case function and meaning are a matter of controversy. While linguists agree that these notions are among the principal properties of cases, it is far from clear what taxonomy should be accepted as the standard for a particular language. Therefore, the absolute number of case functions and meanings as reported in Appendix 1 should be taken as tentative and of marginal importance: what matters is *the proportion of functions and meanings modified by a particular case, relative to other cases*.

THE INFORMATION-THEORETIC APPROACH TO PROCESSING OF INFLECTED NOUNS

Generally, models dealing with processing of inflected morphology emphasize the relevance of suffix frequency, not taking into consideration the fact that inflected words contain encapsulated syntactic information. As noted earlier, Serbian cases differ in number of functions and meanings they encompass and so do the forms that comprise several cases. In other words, inflected noun forms differ not only in their frequency of occurrence, but also in number of syntactic

functions/meanings they encompass. While it is generally acknowledged that frequency is inversely related to processing latency, it could be assumed that the greater number of syntactic functions/meanings is paralleled by greater complexity of a form and, as a consequence, processing time increase. If so, the two parameters have inverse processing effects that can be expressed in terms of a frequency by number of syntactic functions/meanings ratio. The obtained measure is *the average frequency per syntactic function/meaning*. If frequency per syntactic function/meaning ratio for a particular form is expressed as proportion relative to ratios of other noun forms, and transformed by the log transform, the obtained unit is now *the amount of information (I) derived from the average frequency per syntactic function/meaning carried by a particular noun form* (Equation. 1).

$$I_m = \left[-\log_2 \left(\frac{\dfrac{F_m}{R_m}}{\displaystyle\sum_{m=1}^{M} \dfrac{F_m}{R_m}} \right) \right]$$

(1)

In Equation 1 **I** stands for the amount of information carried by a noun form (**m**), **F** refers to frequency of a form, while **R** stands for the number of funtions/meanings modified by a form. The obtained descriptor refers to relative complexity of a noun form: the higher the **I** value, the higher the complexity of a form. If so, it could be assumed that increase in the amount of information should be paralleled by processing time increase.

The preliminary evaluation of Equation 1 was performed on reaction times reported in several studies with Serbian nouns. When regressed on response latencies for the four forms of feminine nouns reported by Todorović (1988), values derived from Equation 1 accounted for 98% of processing variability. For values obtained on six forms of masculine nouns, reported by Kostić & Katz, (1987), 92% of processing variability was accounted for, while all the variability was explained for the replicated versions of experiments reported by Lukatela and his associates (Lukatela, Mandić, Gligorijević, Kostić, A. & Turvey, 1980; Lukatela, Carello & Turvey, 1987) (Kostić, A. 1991; 1995). The above experiments were characterized by factorial design with only few inflected forms of a particular gender being presented. In order to evaluate the generality of the Information-theoretic Approach in a series of experiments all inflected forms of masculine, feminine regular and irregular and neuter nouns were presented (Kostić, A., 2003 submitted). The outcome of these experiments demonstrated that processing time for Serbian inflected noun forms highly correlates with the amount of information (bits – Equation 1) carried by those forms. Specifically, 88% of variability of inflected forms of masculine nouns was accounted for by values derived from Equation 1, 98% of processing variability of feminine nouns, 99% of variability of irregular

feminine nouns and 99% of variability of inflected forms of neuter nouns. These outcomes strongly suggest that time to process inflected nouns is determined by the amount of information derived from the average frequency per syntactic function/meaning modified by a particular noun form.

CRITERIA FOR UNCERTAINTY SPECIFICATION

In the summarized experiments probabilities of inflected forms were specified relative to the paradigm of a particular gender. Thus, for example, for masculine nouns we asked what is the probability of suffix x to be attached to a masculine noun or what is the probability of a suffix y to be attached to a feminine noun. Probability of a suffix (and, by the same token, the amount of information - Equation 1), was defined as *sum of case frequencies encompassed by a given inflected form within a given gender paradigm* (see Table 1). The same procedure was applied for irregular feminine nouns as well. The outcome of the summarized experiments (Kostić, A. 2003, submitted) indicated that proper specification of probability estimate is tied to paradigm of a particular gender and not to probability of suffix per se, irrespective of gender. This indicates that suffix probability for a given paradigm is case and gender dependent.

If the above statement is correct, the implication is that the same should hold not only for irregular feminine nouns but also for the subparadigms of regular feminine nouns. Within the paradigm of Serbian feminine nouns ten subparadigms could be distinguished. The respective subparadigms differ in their probabilities, the most frequent being the one of type "žaba" that encompasses about 78% of all feminine nouns (see Appendix 2) (Kostić, Đ. 1999). Inspection of Appendix 2 indicates two subparadigms that differ in case repertoire of inflected forms from the dominant subparadigm (type "žaba"). Those two subparadigms encompass nouns of type "tabla" and type "bajka", both of them having suffix "i" in the genitive plural. However, nouns of type "bajka" undergo sibilization in the dative and locative singular, thus creating an additional inflected form (seven instead of six distinct forms). Therefore, clear contrast could be obtained only for the subparadigms of type "žaba" and type "tabla" – they both appear in six distinct inflected forms that differ in probability distribution due to the fact that forms encompass different cases. In addition, the probability of the two subparadigms also differs (78% vs. 13% - see Appendix 2).

The aim of the present study is to investigate whether the cognitive system is sensitive to the amount of information derived from probability of inflected form *within a given subparadigm* or to the amount of information derived from probability defined relative to a *dominant subparadigm*. In the experiment with all six inflected forms of feminine nouns, referred to earlier (Kostić, A. 2003, submitted), only nouns from the dominant subaradigm (type "žaba") were presented. The amount of information was specified in the following way. Inflected forms, i .e. cases encompassed by inflected forms, were specified relative to the dominant subparadigm (type "škola"). Case frequencies, on the other hand, were derived from

the paradigm of feminine nouns and not from the dominant subparadigm. The fact that almost all variance has been accounted for by values derived from Equation 1 may suggest that cognitive system is sensitive to probabilities derived from a given subparadigm, in spite of the fact that case frequencies were derived from the global paradigm of feminine nouns. However, the fact that stimulus materials consisted of nouns from the dominant subparadigm may obscure the relevant level of cognitive sensitivity. It remains unclear whether the cognitive system is sensitive to the subparadigm that encompasses 78% of feminine nouns or to the paradigm of feminine nouns that encompasses all feminine nouns.

In order to find out which of the two possible probability counts is cognitively relevant, in the present study we investigate processing of inflected noun forms from the non-dominant subparadigm. Specifically, we investigate processing of inflected forms of "tabla" subparadigm. There are two distinct probability counts and, therefore, two distinct predictors: the one, based on case distribution of the dominant subparadigm ("žaba") and the other based on case distribution of the subparadigm of nouns of type "tabla" (Table 1). Inspection of Table 1 indicates that the genitive plural for nouns that belong to the dominant paradigm (*žab-a*) is morphologically identical to the nominative singular. In contrast, the genitive plural for nouns whose base form ends with two consonants (*tabl-a*) is morphologically identical to the dative and locative singular. As a consequence, informational values for the two types of nouns differ. The question is whether this difference has cognitive consequences, i.e. whether patterning of response latencies for inflected forms of the two subparadigms will also differ. The amount of information carried by inflected forms of the two subparadigms is presented in Table 1 (see also Appendix 1). Note that informational values (Equation 1) are derived from F and R values, presented in Appendix 1.

Table 1: Distinct forms of the two types of feminine nouns and the amount of information (I) carried by these forms

Form	case	I
žab-a	*nom. sing+gen. pl.*	1.464
žab-e	*gen. sing.+nom. pl.+acc. pl*	2.280
žab-i	*dat. sing.+loc. sing.*	2.803
žab-u	*acc. sing.*	2.705
žab-om	*ins. sing.*	3.346
žab-ama	*dat. pl.+loc. pl.+ins. pl.*	4.77
tabl-a	*nom. sing.*	0.175
tabl-e	*gen. sing.+nom. pl.+acc.pl*	4.713
tabl-i	*dat. sing.+loc.sing.+gen. pl*	5.478
tabl-u	*acc. sing.*	5.138
tabl-om	*ins. sing.*	5.778
tabl-ama	*dat. pl.+loc. pl.+ins. pl.*	7.206

EXPERIMENT

All six forms of nouns of type "tabla" were presented in the experiment.

Method

Participants: 60 first-year undergraduates from the Department of Psychology, University of Belgrade participated in the experiment as part of their academic requirements.

Stimuli and procedure: Six groups of participants were presented with 48 feminine nouns of type *tabla* and 48 pseudo-nouns in six forms. All stimuli were equalized for length (4 letters in the base form). Nouns and pseudo-nouns were presented on a computer screen (AppleII-e) with 1500 ms exposure duration. The subject's task was to answer as quickly and as accurately as possible (by pressing yes/no keys) whether the presented string of letters is a word or not.

Results

The mean response latencies are presented in Table 3.

Table 3: Mean reaction time to six forms of nouns of type tabla

Form	RT(ms)
tabl-a	658
tabl-e	669
tabl-i	666
tabl-u	667
tabl-om	683
tabl-ama	707

The analysis of variance, performed on subjects' mean reaction times, indicated a significant main effect of noun form: $F(5, 365) = 24.85$, $p<0.001$. When response latencies for forms of *tabla* type nouns were regressed on their informational values, no significant proportion of explained processing variability was obtained: $F(1,4)=5.982$, $r^2=0.599$, $p>0.05$. In spite of the fact that the amount of information was calculated with respect to cases shared by each form, the proportion of explained variability did not reach significance. On the other hand, when the response latencies were regressed on informational values derived from the dominant subparadigm (*žaba*), the proportion of explained variability did reach significance: $F(1,4)=36.097$, $r^2=.90$, $p<0.001$ (Figure 1).
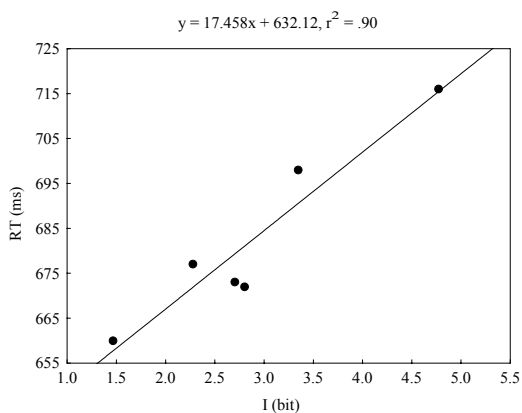
Figure 1: Relation between processing latencies for six forms of feminine nouns of tabla type and the amount of information derived from the dominant subparadigm.

Nouns of the non-dominant subparadigm (*tabla*) are processed as if they were of the dominant subparadigm (*žaba*). In spite of the fact that for the nouns of the non-dominant subparadigm the nominative singular is morphologically unique, the form is processed as if containing both the nominative singular and the genitive plural.

GENERAL DISCUSSION

The outcome of the present experiment gives further support to the claim that processing of inflected nouns is determined by the amount of information (I) derived from the average frequency per syntactic function/meaning for a given noun form. No processing differences were observed for the two subparadigms of feminine nouns (*žaba* vs. *tabla*), in spite of differences in distribution of informational values of their inflected forms. Feminine nouns of *tabla* type were processed as if they were of *žaba* type. This outcome may suggest that nouns of a particular gender are cognitively instantiated through their paradigmatic subparadigm which, in turn, determines cognitive distribution of cases within inflected forms for all nouns of the respective gender. For the feminine nouns it is the most frequent subparadigm whose base form ends with one consonant (e.g. *žaba*). This implies that the amount of information carried by forms of a particular gender is derived from cases comprised in forms of a dominant subparadigm. The criterion for specifying the dominant subparadigm is the proportion relative to proportion of other subparadigms within a defined paradigm.

REFERENCES

Allen, M., Badecker, W. (1999). Stem homograph inhibition and stem allomorphy; representing and processing inflected forms in multilevel lexical system. *Journal of Memory and Language, 41,* 105-123.

Badecker, W., M. Allen. (2002). Morphological parsing and the perception of lexical identity: A masked priming study of stem-homographs. *Journal of Memory and Language*, *47,* 125-144.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.) *Language Production, Vol. II: Development, writing and other language processes.* London: Academic Press.

Cutler, A. (1983). Lexical complexity and sentence processing. In G. B. Flores d'Arcais & R. Jarvella (Eds.) *The Process of Language Understanding*. New York: Wiley.

Henderson, L. (1985). Towards a psychology of morphemes. In A. W. Ellis (Ed*.) Progress in the psychology of language*. *Vol. I*, London: Lawrence Erlbaum Associates Ltd.

Jarvella, R., Meijers, G. (1983). Recognizing morphemes in spoken words: Some evidence for a stem-organized mental lexicon. In G. B. Flores d'Arcais & R. Jarvella (Eds.) *The Process of Language Understanding*. New York: Wiley.

Kempley, M., Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, *73*, 441- 454.

Kostić, A., Katz, L. (1987). Processing differences between nouns, adjectives and verbs. *Psychological Research*, *49,* 229-236.

Kostić, A. (1991). Informational approach to processing inflected morphology: Standard data reconsidered. *Psychological Research, 53,* 62-70.

Kostić, A. (1995). Informational load constraints on processing inflected morphology. In L. B. Feldman (Ed.) *Morphological Aspects of Language Processing*, Hillsdale, NJ: Erlbaum.

Kostić, A. (2003). The effects of the amount of information on processing of inflected morphology (submitted).

Kostić, Đ. (1965a). *Struktura upotrebne vrednosti gramatičkih oblika u srpskohrvatskom jeziku (Probability Structure of Grammatical Forms in Serbo-Croatian)*. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora.

Kostić, Đ. (1965b). *Sintaktičke funkcije padežnih oblika u srpskohrvatskom jeziku. (Syntactic Functions and Meanings of Serbo-Croatian cases)*. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora.

Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika*. Tom I–VII. (*Frequency Dictionary of Contemporary Serbian Language*. Vol. I–VII). Beograd: Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu.

Lukatela, G., Mandić, Z., Gligorijević, B., Kostić, A., Savić, M., Turvey, M. T. (1978). Lexical decision for inflected nouns. *Language and Speech*, *21,* 166-173.

Lukatela, G., Mandić, Z., Gligorijević, B., Kostić, A., Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. *Memory and Cognition, 8,* 415-423.

Lukatela, G., Carello, C., Turvey, M. T. (1987). Lexical representation of regular and i rregular inflected nouns.*Language and Cognitive Processes*, *2*, 1-17.

Mackay, D. G. (1978). Derivational rules and the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 17,* 61-71.

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

Rubin, G. S., Becker, C. A., Freeman, R. H. (1979). Morphological structure and its effect on visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 757-767.

Taft, M., Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638-647.

Taft, M., Forster, K. I. (1976). Lexical storage and retrieval of polymorphenic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, *15,* 607-620.

Taft, M. (1979a). Recognition of affixed words and the word frequency effect. *Memory and Cognition, 7*, 263-272.

Taft, M. (1979b). Lexical access via orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior, 18,* 21-39.

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior*, *20,* 289-297.

Todorović, D. (1988). Hemispheric differences in case processing. *Brain and Language*, *33,* 365-389.

APPENDIX 1

Frequency (F) expressed in percentages (%) and number of syntactic functions and meanings (R) for inflected cases of Serbian feminine nouns

Number	Case	Form A	Form B	F%	R
Singular	nominative	žaba	tabla	8.841	3
	genitive	žabe	table	7.876	51
	dative	žabi	tabli	0.377	22
	accusative	žabu	tablu	5.480	58
	instrumental	žabom	tablom	1.939	32
	locative	žabi	tabli	3.419	21
Plural	nominative	žabe	table	3.577	3
	genitive	žaba	tabli	3.220	51
	dative	žabama	tablama	0.157	22
	accusative	žabe	table	2.750	58
	instrumental	žabama	tablama	0.734	32
	locative	žabama	table	0.799	21

* Frequency values (F%) and number of thematic roles based on functions and meanings (R) adapted from Kostić, Đ. (1965a; 1965b).

* žaba (A) – frog; tabla (B) – board; strast (C) – passion

* Form (A) – nom. sing. = gen. pl.; Form (B) – dat. + loc. sing = gen. pl.; Form (C) – irregular feminine nouns

APPENDIX 2

Subparadigms of regular feminine gender nouns and their probabilities

Example	A	B	C	D	E	F	%
žaba	*						**77.929**
jabuka	*				*		5.035
zemlja	*					*	0.847
devojka	*				*	*	0.423
tabla		*					**12.659**
bajka		*			*		2.353
ruka			*				0.094
kretnja				*			0.282
tetka				*	*		0.282
olovka				*			0.094

(A) gen. pl – A; (B) gen. pl. I; (C) gen. pl – U; (D) gen. pl. A and I; (E) sibilization; (F) fleeting A.

Note: Probabilities of subparadigms were estimated from Kostić, Đ. (1999).