

CONSTRAINTS ON PROBABILITY DISTRIBUTIONS OF GRAMMATICAL FORMS¹

Aleksandar Kostić² and Milena Božić

Laboratory of Experimental Psychology, University of Belgrade, Serbia

In this study we investigate the constraints on probability distribution of grammatical forms within morphological paradigms of Serbian language, where paradigm is specified as a coherent set of elements with defined criteria for inclusion. Thus, for example, in Serbian all feminine nouns that end with the suffix "a" in their nominative singular form belong to the third declension, the declension being a paradigm. The notion of a paradigm could be extended to other criteria as well, hence, we can think of noun cases, irrespective of grammatical number and gender, or noun gender, irrespective of case and grammatical number, also as paradigms. We took the relative entropy as a measure of homogeneity of probability distribution within paradigms. The analysis was performed on 116 morphological paradigms of typical Serbian and for each paradigm the relative entropy has been calculated. The obtained results indicate that for most paradigms the relative entropy values fall within a range of 0.75 – 0.9. Nonhomogeneous distribution of relative entropy values allows for estimating the relative entropy of the morphological system as a whole. This value is 0.69 and can tentatively be taken as an index of stability of the morphological system.

Key words: *inflected morphology, inflectional paradigms, relative entropy.*

In inflected languages grammatical forms of nouns, adjectives, verbs and some pronouns are marked by inflectional suffixes, each word belonging to a particular

¹ This work was supported by the Ministry of Science and Environmental protection of the Republic of Serbia, grant No. D-149039.

² Autor's address: akostic@f.bg.ac.yu

declension or conjugation. In the present study we refer to declensions and conjugations as *paradigms*, where by paradigm we mean *a coherent set of elements with defined criteria for inclusion*. Thus, for example, in Serbian all feminine nouns that end with the suffix "a" in their nominative singular form belong to the third declension, feminine nouns that end with a consonant belong to the fourth declension etc. The notion of a paradigm could be extended to other criteria as well, hence, we can think of noun cases, irrespective of grammatical number and gender, or noun gender, irrespective of case and grammatical number, also as paradigms. Thus, for example, in Serbian, the case paradigm includes six elements (cases), the gender paradigm includes three elements (masculine, feminine and neuter), the grammatical number paradigm two elements (singular, plural) etc. In other words, paradigms can include both individual grammatical forms as well as elements of various grammatical categories.

Elements of different paradigms appear with unequal probabilities. The nominative case is more frequent than the dative, feminine gender is more frequent than neuter gender, singular is more frequent than plural etc. In table 1 we give an example of the unequal probability distribution of cases, singular and plural, of masculine nouns in Serbian.

Table 1. An example of the unequal probability distribution of cases

	Singular	Plural
Case	p	p
Nominative	0.289	0.0749
Genitive	0.1928	0.0892
Dative	0.0195	0.0063
Accusative	0.1236	0.0498
instrumental	0.0427	0.0134
Locative	0.0849	0.0136

We may ask whether unequal probability distributions within paradigms are unsystematic or is there some systematic factor that guides the observed distributions. Specifically, we ask whether probability distributions within paradigms vary freely, with no obvious regularity, or are there some constraints that allow probability variation within limited margin only. In order to answer this question we need to specify the metric in which different probability distributions can be described. The obvious metric derives from Information Theory. Specifically, it is the *entropy* of a paradigm, defined in Information Theoretic terms. In the forthcoming paragraphs we discuss the notion of entropy in more detail.

THE AMOUNT OF INFORMATION AND ENTROPY

In Information Theory information is defined in terms of probability. The amount of information carried by an event is inversely related to its probability: higher probability is paralleled by smaller amount of information and vice versa. The unit in which we express the amount of information is the *bit* (equation 1).

$$h_i = -\log_2 p_i \quad (1)$$

In equation 1 h_i refers to the amount of information (bits) carried by an event x within a system y , where p_i is the probability (proportion) of an event x , transformed by $\log_2$ and multiplied by -1. The obtained value is the amount of information (bits) carried by a particular event within some system. Note that probability is defined in terms of proportion relative to proportions of other events within a given system. Let us now apply equation 1 to cases singular and plural of Serbian masculine nouns (table 2: probabilities, expressed as proportions, and the amount of information carried by Serbian cases singular and plural masculine nouns).

Table 2. Masculine nouns

	Singular		Plural	
case	p	bit	p	bit
nominative	0.289	1.791	0.0749	3.739
genitive	0.1928	2.375	0.0892	3.487
dative	0.0195	5.680	0.0063	7.310
accusative	0.1236	3.016	0.0498	4.328
instrumental	0.0427	4.549	0.0134	6.222
locative	0.0849	3.558	0.0136	6.200

Inspection of table 2 indicates conspicuous variation of the amount of information carried by singular and plural cases of Serbian masculine nouns. The data presented in table 2 can be interpreted as the amount of information carried by each element of a system, where elements are grammatical forms and the system being the paradigm of masculine nouns. We can now express the amount of information carried by the system (i.e. paradigm) in terms of *entropy*, standardly defined as the average amount of information carried by a system /equation 2/.

$$H = -\sum p_i \log_2 p_i \quad (2)$$

In equation 2 H is the entropy of a system x , p_i is the probability of an event i within the system x and $\log_2 p_i$ is the amount of information carried by the event i . For each element of the system we calculate the product of its probability and the amount of information. The sum of products is the entropy of the system, also expressed in bits. Let us now return to the example from table 2. The entropy of the paradigm of masculine nouns that includes all cases singular and plural is 2.976 bits.

Assume that all singular and plural cases of masculine nouns appear with equal probability, where probability of each case is 0.083. What will be the of entropy of such paradigm? Applying the equation 2 the entropy is 3.585 bits. For the system in which all elements appear with equal probabilities we say that it is in the state of *maximum entropy*. In other words, maximum entropy equals $\log_2$ of number of elements within a system (equation 3).

$$H_{max} = \log_2 N \quad (3)$$

Equation 3 implies that systems with greater number of elements will have higher values of maximum entropy. This implies that entropies of two systems with unequal number of elements which appear with unequal probabilities will also generally differ. This feature may have considerable consequences for our study because it prevents us from direct comparison of paradigms with unequal number of elements. If entropies of two paradigms with unequal number of elements (e.g. case paradigm /six elements/ vs. gender paradigm /three elements/) differ, it remains unclear whether this difference is due to probability distributions within paradigms or due to the difference in the number of elements. With this in mind it becomes clear that entropy standardly defined (equation 2) may not be the proper descriptor that allows for comparisons among paradigms.

The Information Theoretic descriptor that is not sensitive to the number of elements of a paradigm is *the relative entropy* (equation 4).

$$Hr = H/H_{max} \quad (4)$$

The values of relative entropy range between 1 (maximum entropy) and asymptotically approach 0. If differences in probabilities of elements within a system are greater, the value of relative entropy will be smaller and vice versa. In other words, the more homogeneous the probability distribution, the greater the value of relative entropy. In other words, the value of relative entropy can be taken as an index of homogeneity of the probability distribution of elements within a system or, put differently, how far is the system from the state of maximum entropy.

The residual between relative entropy and maximum entropy is the *redundancy* (equation 5), which can also be used as an index of homogeneity of probability distribution of elements within a given system.

$$C = 1 - Hr \quad (5)$$

Due to the fact that relative entropy is a ratio, it is not sensitive to the number of elements of a system which, on the other hand, makes it a proper descriptor that allows for comparisons among paradigms with different number of elements.

With relative entropy being used as an index of homogeneity of probability distributions, paradigms can now be directly compared with respect to probability distributions of their grammatical forms or categories. Our initial question can now be rephrased. We ask whether paradigms vary unsystematically with respect to their relative entropy values or is there some preferred range of relative entropy values. In order to answer this question, the probabilities of grammatical forms should be estimated first, which will enable us to calculate the relative entropy values for different paradigms.

PROBABILITY ESTIMATES AND CRITERIA FOR PARADIGM SELECTION

Probabilities of grammatical forms were estimated for Serbian, which is highly inflected and to a great extent free word order language. The probabilities were estimated from samples of daily press and poetry and then averaged for the two registers (Kostić, Đ. 1965a). Each sample consisted of about one million words. The two registers are part of the Corpus of Serbian Language that consists of 11 million words (Kostić, Đ. 2001). The Corpus is diachronic and spans Serbian language from the 12th century to the contemporary language. The sample of contemporary language consists of about 7 million words and encompasses five registers (daily press, poetry, prose, scientific texts and political texts). Each word in the Corpus is manually annotated for its grammatical status with a system of annotation that distinguishes about 2000 grammatical forms in Serbian. The Corpus was compiled and annotated in the mid fifties and transferred into the electronic format in the late nineties.

Probabilities were specified at the level of grammatical form and at the levels of different grammatical categories. The most detailed estimate was at the level of grammatical form (e.g. if a word is a noun, what is the probability for it to be a masculine noun in genitive singular). The coarser specifications were performed at the level of grammatical categories (e.g. what is the probability of a noun to be in the nominative case, or what is the probability of a noun to be in singular etc.).

As noted earlier, we define a paradigm as a coherent set of elements with specified criteria for inclusion in the set. According to this definition, each declension and conjugation could be treated as a paradigm. Likewise, a set of noun cases, irrespective of grammatical gender and number can also be considered a paradigm because the members of a set belong to the same grammatical category (case) within a defined class (nouns). This is also true for other categories like, for

example, noun gender, verb person, adjectival number etc. In order to get better understanding of the relative entropy variation of different paradigms, we need an exhaustive list of possible paradigms of Serbian nouns, adjectives and verbs. In addition to the very basic level of grammatical forms, paradigms will be specified at the level of grammatical categories as well.

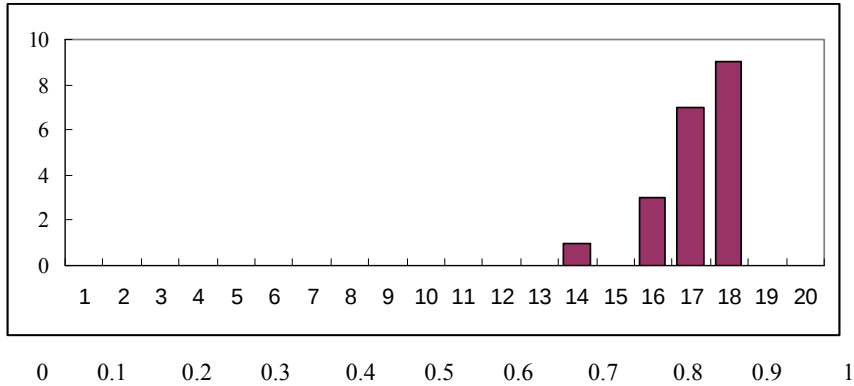
PROBABILITY DISTRIBUTIONS WITHIN NOUN PARADIGMS

Taking into consideration case, grammatical number and gender and their combinations, following paradigms are possible for Serbian nouns:

1. Case, irrespective of gender and grammatical number
2. Gender, irrespective of case and grammatical number
3. Grammatical number, irrespective of case and gender,
4. Case, masculine, irrespective of grammatical number
5. Case, feminine, irrespective of grammatical number
6. Case, neuter, irrespective of grammatical number,
7. Grammatical number, masculine, irrespective of case
8. Grammatical number, feminine, irrespective of case
9. Grammatical number, neuter, irrespective of case
10. Case, masculine singular
11. Case, feminine singular
12. Case, neuter singular
13. Case, masculine plural
14. Case, feminine plural
15. Case, neuter plural
16. Case, singular, irrespective of gender
17. Case, plural, irrespective of gender
18. Case singular and plural, masculine
19. Case singular and plural, feminine
20. Case singular and plural, neuter

Relative entropies for the above paradigms are presented in tables 1-17 in Appendix. The distribution of relative entropy values for noun paradigms is presented in Graph 1 (X-axis: value of relative entropy, Y-axis: number of paradigms that fall into defined value range (0.05)).

Graph 1. Distribution of relative entropy values for noun paradigms



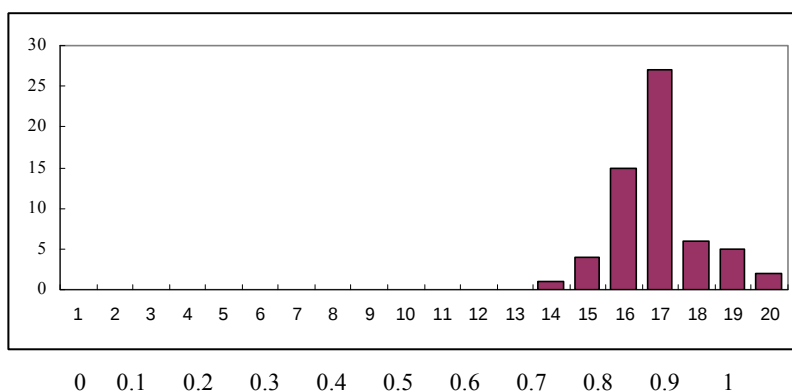
Inspection of Graph 1 indicates that most values of relative entropy range between 0.75 and 0.9.

PROBABILITY DISTRIBUTIONS WITHIN ADJECTIVAL PARADIGMS

Due to the fact that adjectives have the same grammatical characteristics as nouns (i.e. case, number and gender), paradigms of adjectives parallel those of nouns. However, unlike nouns, adjectives are characterized by three levels of comparison: positive, comparative and superlative, each level being characterized with the same grammatical properties (case, number and gender). Thus the number of paradigms should be multiplied by three (for each level of comparison). Likewise, there is a distinct paradigm for comparison (probability of an adjective to be in positive, comparative and superlative).

Values of relative entropy for adjectival paradigms are presented in tables 21-80 in Appendix. The Distribution of relative entropy values for adjectival paradigms is presented in Graph 2 (X-axis: value of relative entropy, Y-axis: number of paradigms that fall into defined value range (0.05)).

Graph 2. Distribution of relative entropy values for adjectival paradigms



Inspection of Graph 2 indicates that most values of relative entropy cumulate within a range of 0.75 and 0.95.

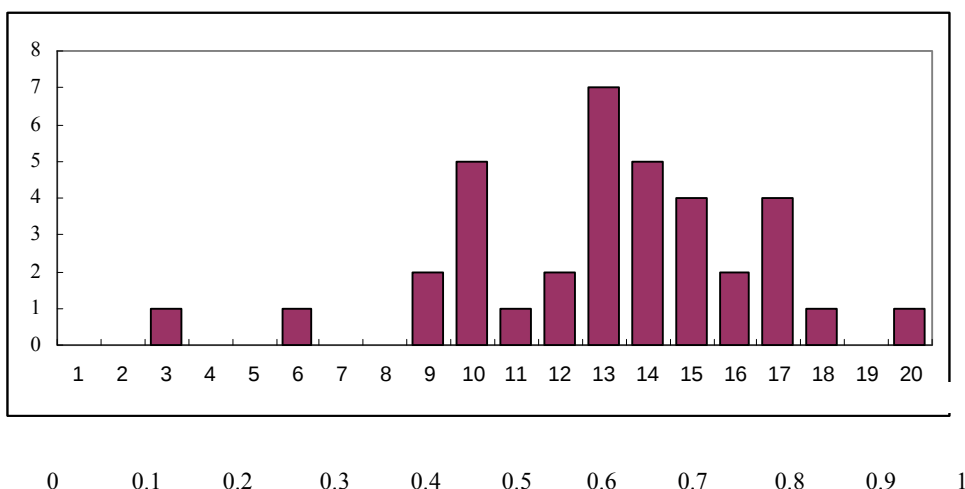
PROBABILITY DISTRIBUTIONS WITHIN VERB PARADIGMS

Serbian verbs are characterized by person, grammatical number, tense, aspect and sometimes gender. Aspect will be omitted from the current analysis due to the fact that it does not generate systematic paradigms. On the other hand, grammatical gender is marked only for past tense and plusquamperfect. Therefore, only within those tenses it is possible to specify paradigms with respect to gender. With this in mind, the following paradigms can be generated for verbs:

1. Verb person, irrespective of grammatical number, tense and gender
2. Grammatical number, irrespective of person, tense and gender,
3. Tense, irrespective of person, grammatical number and gender
4. Present tense, verb person, irrespective of number
5. Present tense, grammatical number, irrespective of person
6. Present tense, person singular
7. Present tense, person plural
8. Future tense, verb person, irrespective of number
9. Future tense, grammatical number, irrespective of person
10. Future tense, person singular
11. Future tense, person plural
12. Imperfect, verb person, irrespective of number
13. Imperfect, grammatical number, irrespective of person
14. Imperfect, person singular
15. Imperfect, person plural

16. Aorist, verb person, irrespective of number
17. Aorist, grammatical number, irrespective of person
18. Aorist, person singular
19. Aorist, person plural
20. Past tense, person, irrespective of grammatical number and gender
21. Past tense, gender, irrespective of person and grammatical number
22. Past tense, person singular, irrespective of gender
23. Past tense, person plural, irrespective of gender
24. Past tense, person masculine, irrespective of grammatical number
25. Past tense, person feminine, irrespective of grammatical number
26. Past tense, person neuter, irrespective of grammatical number
27. Past tense, person masculine singular
28. Past tense, person feminine singular
29. Past tense, person neuter singular
30. Past tense, person masculine plural
31. Past tense, person feminine plural
32. Past tense, person neuter plural
20. Plusquamperfect, person, irrespective of grammatical number and gender
21. Plusquamperfect, gender, irrespective of person and grammatical number
22. Plusquamperfect, person singular, irrespective of gender
23. Plusquamperfect, person plural, irrespective of gender
24. Plusquamperfect, person masculine, irrespective of grammatical number
25. Plusquamperfect, person feminine, irrespective of grammatical number
26. Plusquamperfect, person neuter, irrespective of grammatical number
27. Plusquamperfect, person masculine singular
28. Plusquamperfect, person feminine singular
29. Plusquamperfect, person neuter singular
30. Plusquamperfect, person masculine plural
31. Plusquamperfect, person feminine plural
32. Plusquamperfect, person neuter plural

Values of relative entropy for verb paradigms are presented in tables 81-116 in Appendix. The Distribution of relative entropy values for verb paradigms is presented in Graph 3. (X-axis: value of relative entropy, Y-axis: number of paradigms that fall into defined value range (0.05)).

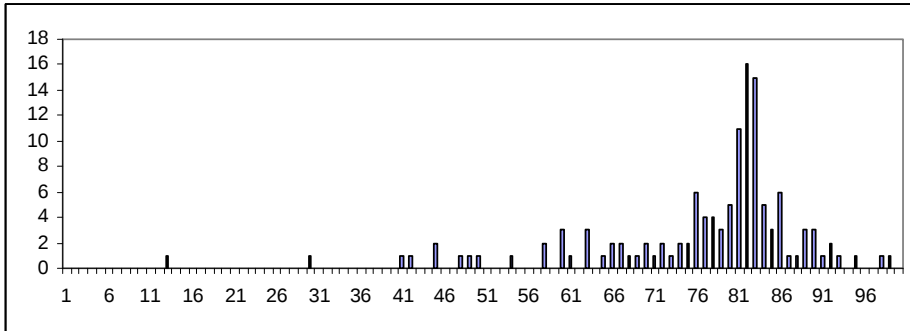
Graph 3. Distribution of relative entropy values for verb paradigms

Unlike values for noun and adjectival paradigms, relative entropy values for verb paradigms range mainly between 0.4 and 0.9. Some paradigms of verbs are characterized with extremely high redundancy (i.e. low relative entropy). A conspicuously low relative entropy is observed for the plusquamperfect gender (table 109) due to the fact that the proportion of masculine gender is prevalent relative to feminine and neuter gender ($H_r=0.129$). Likewise, an extremely low relative entropy has also been observed for verb person in the past tense (table 104) due to the extremely high probability of the third person singular ($H_r=0.296$).

RELATIVE ENTROPY DISTRIBUTION IRRESPECTIVE OF WORD TYPE

For the three word types analyzed in this study 116 paradigms had been specified and their relative entropies calculated. We may ask what is the distribution of relative entropy values irrespective of word type. In order to do this, we sum up the number of paradigms with a given entropy value /Graph 4 (X-axis: value of relative entropy, Y-axis: number of paradigms with a particular relative entropy value) and Appendix/.

Graph 4. Distribution of relative entropy values for paradigms of all three word-types



Inspection of Appendix 3 and Graph 5 indicates that almost two thirds (64%) of relative entropy value fall within a range of 0.75 – 0.9, while one third (33%) of the values fall within a range of 0.8 – 0.85.

RESTRICTION TO PRINCIPAL PARADIGMS

The cumulation of relative entropy values within the observed range should be taken with caution. It may not be surprising that values cumulated within some margins because there are many instances where the same grammatical categories appear in different paradigms. Take, for example, case. Case is a property of nouns, adjectives and some pronouns. Each noun and adjective can cross six cases singular and plural. Nouns, like adjectives, appear in three genders, while adjectives also appear in three levels of comparison. Cases cross gender, and number, while in the case of adjectives they also cross three levels of comparison. With this in mind, 9 paradigms can be specified for nouns and 27 for adjectives. Is there a reason to assume that probability distributions of cases in singular should differ from those in plural, or that the distribution of cases for masculine gender should differ from the distribution for feminine or neuter gender? Likewise, is there a reason to assume that case probability distributions differ for adjectives in positive as opposed to adjectives in comparative and superlative? Finally, there is strong reason to assume that probability distributions of case, grammatical number and gender for adjectives should parallel those of nouns because adjectives and nouns have to agree in case, number and gender. These assumptions were empirically confirmed. Additional analyses revealed extremely high correlations for case probability distributions for nouns of different gender and number, as well as distributions for adjectives and nouns. The same is true for grammatical number and gender. Note that each of those

instances was treated as a distinct paradigm with its relative entropy value. With this in mind, we may ask whether the observed cumulation of relative entropy values may be an artifact of the way paradigms were specified.

This problem is two-fold. On the one hand, it is true that the same category (for example, case) is multiplied in a number of paradigms. On the other hand, it is also true that this multiplication is a fact of language.

So, is there any distinct range of the relative entropy distribution that may not be an artifact of paradigm selection criteria? In order to answer this question, we have to restrict our analysis to probability distributions within paradigms that are unique. Inspection of grammatical properties of nouns reveals that case, grammatical number and gender are their principal grammatical properties, for adjectives those are case, grammatical number, gender and comparison, and person, grammatical number and tense for verbs. In table 3 we give the relative entropy values for those categories, with probability distributions within paradigms being calculated at the most coarse level. For example, for noun cases it means probability distributions irrespective of grammatical number and gender, for noun gender it means probability distributions irrespective of case and grammatical number etc.

Table 3. Relative entropy values for six principal grammatical categories

Category	Hr
nouns (case)	0,85
nouns (gender)	0,81
nouns (number)	0,77
adjectives (comparison)	0,21
Verbs (person)	0,71
Verbs (number)	0,87
Verbs (tense)	0,70

Inspection of table 3 indicates that values of relative entropy for six out of seven principal grammatical categories across three word types range between 0.7 and 0.87, the exception being adjectival comparison (0,21). This range is not substantially different from the one observed for all paradigms as presented in picture 4 where two thirds of the relative entropy values range between 0.75 and 0.9. The average value of relative entropy is 0.73.

HOMOGENEITY OF RELATIVE ENTROPY DISTRIBUTIONS

Inspection of the distribution of relative entropy values, as depicted in Graph 4 and Appendix 3, indicates cumulation of relative entropy values. Thus, for example, there are 16 paradigms with relative entropy value of 0.82, 6 with the value of 0.86,

4 with the value 0.78 etc. In other words, there is a nonhomogeneous distribution of number of paradigms with a given relative entropy value. At this point we may ask what is the degree of the observed nonhomogeneity. Again, the metric will be the relative entropy. However, in this metric needs to be elaborated in more detail. In Graph 4 relative entropy values are presented on the X-axis and the number of paradigms with a given relative entropy value on the Y-axis. Now we may rephrase the Y-axis in an alternative metric. Instead of asking how many paradigms take a particular entropy value, we may ask what is the probability that a given paradigm has a particular entropy value. To do this, we need to transform the raw values (i.e. number of paradigms with a particular entropy value) into proportions. By doing this, we can now treat the distribution of all 116 paradigms as a single paradigm with unequal probabilities of events, the events being a given value of relative entropy. (Probabilities of entropy values are given in Appendix 3). The relative entropy of the whole system is 0.69.³

What may be the interpretation of the relative entropy value for the whole system? At this point we take it as a tentative index of the overall stability of a system of inflected morphology. This assumption will be elaborated in more detail in the forthcoming paragraphs.

EXTENSION TO OTHER ASPECTS OF LANGUAGE

The observed range of relative entropy variation was obtained for the exhaustive set of morphological paradigms. We may ask whether this range is specific for morphology, or is it more general. Will the same range of entropy variation be observed for paradigms derived from other aspects of language like, for example, probability distributions of individual words? In order to answer this question the following paradigms were investigated:

- a. Probability distribution of lemmas.
- b. Probability distribution of forms of lemmas
- c. Probability distribution of lemmas within a given word type

Likewise, we can treat probability distributions of word types and probability distributions of all grammatical forms as a distinct paradigm.

These additional paradigms may not be an exhaustive set of possible paradigms for aspects other than morphology, but they may be sufficient to get some preliminary insights about the generality of the range observed with morphological paradigms.

Table 4. The obtained relative entropy values

³ The same kind of calculation for the seven principal paradigms is not reliable due to small number of events (i.e. relative entropy values) (table 3).

Paradigm	ΣF	N	H	$H_{\max}$	H_r
lemmas					0,706
forms of lemmas					0,833
lemmas (nouns)	659 399	93 792	14,695	16,512	0,888
lemmas (adjectives)	223 058	68 972	14,303	16,074	0,890
lemmas (numerals)	26 111	1 417	7,731	10,468	0,739
lemmas pronouns)	208 957	2 461	7,672	11,266	0,681
lemmas (verbs)	356 230	60 449	12,316	15,887	0,775
lemmas (adverbs)	94 180	3 973	8,585	11,954	0,718
lemmas (prepositions)	202 986	196	3,979	7,615	0,523
lemmas (conjunctions)	181 902	241	3,248	7,913	0,41
lemmas (exclamations)	4 417	189	3,408	7,563	0,451
lemmas (particles)	35 333	287	4,586	8,167	0,562
grammatical forms					0,729
word types					0,76

Probabilities for the additional paradigms, like those for morphological paradigms, were derived from the "Frequency Dictionary of Contemporary Serbian Language" (Kostić, Đ., 1999). The Dictionary was compiled from samples of daily press and poetry, each containing about one million words. The Dictionary contains 65 000 lemmas and their frequencies and about 240 000 forms of lemmas (and their frequencies). The obtained relative entropy values for the respective paradigms are given in table 4.

Inspection of table 4 indicates that variation of relative entropy values of paradigms for other aspects of language also ranges mainly between 0.7 and 0.9

The observed outcomes suggest that the observed range is not morphology specific, but more general and may apply to all aspects of language. The obtained results also indicate that relative entropy variation is more generic than initially assumed. This, on the other hand, suggests some global constraints imposed over permissible probability variations of language events.

GENERAL DISCUSSION

In the present study we investigated probability distributions for various paradigms of Serbian inflected morphology. Specifically, we investigated probability distributions within noun, adjective and verb paradigms. The aim of the study was to evaluate whether there is some systematic patterning in probability distributions of grammatical forms belonging to different paradigms, or whether this distribution is nonsystematic and arbitrary. The first step was to estimate the probabilities of grammatical forms for different paradigms. This estimate was derived from a study in which probabilities of all grammatical forms in Serbian were given at the levels of individual grammatical forms and at the various levels of

grammatical categories (Kostić, 1965a). The next step was to specify the metric in which to express probability distributions in a way that will allow for comparison among paradigms with various number of elements. This metric proved to be the relative entropy, which can be treated as a measure of homogeneity of probabilities within a given paradigm or, alternatively, how far is a paradigm from the state of maximum entropy.

It is reasonable to make an a priori assumption that probabilities of grammatical forms may not be homogeneous, nor should they, by the same token, be close to the state of maximum entropy. It is less clear, however, whether probability distributions of grammatical forms should vary freely, or should their variation be constrained within relatively narrow margins. Put differently, there is no obvious reason why it would not be possible for a number of paradigms to take any value of relative entropy between, say, 0.1 and 0.9. Assume some paradigms having one grammatical form with extremely high probability, while other forms appear with small probabilities. In such a case the paradigm is characterized by high redundancy, i.e. low relative entropy. Likewise, it is conceivable to have a paradigm with minimal probability differences among grammatical forms, in which case the paradigm would have high relative entropy and low redundancy. If there are no constraints on probability distributions within paradigms, theoretically it could be expected that paradigms distribute homogeneously with respect to their relative entropy values.

Although theoretically plausible, the above assumptions were not empirically confirmed. Two thirds of relative entropy values for 116 paradigms ranged between 0.75 and 0.9. The same is true for the reduced number of paradigms, where six out of seven paradigms ranged between 0.7 and 0.9. The observed concentration suggests that probability distributions of grammatical forms within paradigms are not arbitrary nor is, by the same token, the variation of relative entropy values among paradigms. Most probability distributions within paradigms are restricted to a relatively narrow margin of ± 0.1 of relative entropy values, ranging between 0.7 and 0.9.

This conclusion, as noted earlier, has to be taken with caution. The caveat is related to the criterion of paradigm selection. Namely, some grammatical categories like, for example, case and grammatical number, cross many paradigms with minimal variation in probability of grammatical forms or subordinate grammatical categories. Consequently, cumulation of relative entropy values within a narrow margin may not be surprising. In order to eliminate this we introduced a somewhat restrictive criterion where only principal grammatical categories, specified at the most molar level (e.g. case of a noun, irrespective of gender and grammatical number) are taken into consideration. The obtained range of variation did not differ substantially from the one observed for 116 paradigms. Such an outcome suggests that the observed range may indicate the preferred relative entropy values of paradigms or, put differently, the preferred probability distributions of grammatical forms within paradigms.

Before we address the problem of "preference" for a given range of relative entropy variation, we need to elaborate the statement that the observed cumulation may be an artifact of paradigm selection. Although the subsequent analyses on the restricted number of paradigms indicated similar ranges of entropy variation, it is true that the observed cumulation obtained for 116 paradigms is also due to the fact that the same categories cross a number of paradigms. At this point we have to distinguish two aspects of the problem. The one, related to the *range* of relative entropy variation and the other, related to *cumulation* of relative entropy values within a given range. The first aspect is not an artifact of paradigm selection because similar ranges have been observed for both selection criteria. The latter aspect, however, requires additional elaboration. We defined a paradigm as a coherent set of elements with specified criteria for inclusion. Applied to inflectional morphology this allows for creation of a finite number of paradigms with clear selection criteria – whatever can be specified in terms of a coherent class with n grammatical elements can be treated as a morphological paradigm. This specification implies a continuum ranging from the molecular level of a given set of grammatical forms (e.g. nominative singular, masculine nouns) up to the molar level of grammatical categories (e.g. cases, irrespective of word type, grammatical number and gender). Once such a criterion has been adopted, we inevitably face the multiplication of the same set of elements across number of paradigms. The question is whether this is a matter of description or the fact of language.

It is a fact of language that case is a property of three distinct word types (nouns, adjectives and some pronouns). Likewise, it is a fact of language that grammatical number is a property of nouns, adjectives, verbs and pronouns. The same is true to some extent for grammatical gender. Criteria for paradigm selection should map onto these properties within and across word types if their purpose is to encompass exhaustive sets of possible paradigms at different grain size of description. If so, multiplication of paradigms with equivalent probability distribution and, by the same token, cumulation of paradigms with the same relative entropy value is natural consequence of language properties. With this in mind it can be stated that the observed cumulation is a matter of language rather than criteria by which paradigms were specified. This does not refute the fact that probabilities for entropy values to take a given entropy value are not inflated to some extent.

The fact that distribution of relative entropy values is not homogeneous and that there is a preferred margin within which those values cumulate may suggest that distribution of relative entropies is constrained in a way to conserve some overall divergence of relative entropy values. This divergence can also be expressed in terms of relative entropy, this time as relative entropy of the whole system. As demonstrated, the value of this entropy is around 0.7 and can tentatively be taken as an index of the system's overall stability. Any conspicuous change of relative entropy for a given paradigm (or paradigms) may change the relative entropy of the morphological system and push it to the state of potential instability. If such change happens over wider time span, the compensatory changes of probabilities have to

happen in order to conserve the overall probability divergence (or coherence) expressed in terms of the overall relative entropy value.

Let us elaborate this assumption in more detail. Probability fluctuation is an inherent property of language and can be observed in diachronic studies of different linguistic phenomena. Take, for example, the hypothetical decrease of probability of one case in case paradigm in some period of time. Such decrease will cause the decrease of relative entropy across number of paradigms. Consequently, this will cause the relative entropy of the system to decrease as well, thus shifting the system towards the state of instability. In order to conserve the stable state (i.e. the relative entropy value of the system) other changes have to occur as well (not necessarily in the case system) which will compensate for the case probability decrease. In other words, any change in probability within one paradigm has to be compensated by some proportional probability change (or changes) in other paradigm (or paradigms). This assumption, however, needs empirical evaluation in diachronic studies of language.

REFERENCES

Kostić, Đ. (1965a). *Struktura upotrebne vrednosti gramatičkih oblika u srpskohrvatskom jeziku*. Beograd, Institut za eksperimentalnu fonetiku i patologiju govora.

Kostić, Đ. (1965b). *Sintaktičke funkcije padežih oblika u srpskohrvatskom jeziku*. Beograd, Institut za eksperimentalnu fonetiku i patologiju govora.

Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika*. Tomovi I – VII (7-560). Beograd, Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju.

Kostić, Đ. (2001). *Elektronski korpus srpskog jezika*. Beograd, Institut za eksperimentalnu fonetiku i Laboratorija za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu.

APPENDIX 1

**Probability distribution within paradigms of nouns,
Verbs and adjectives**

n – nominative
g – genitive
d – dative
a – accusative
i – instrumental
l – locative

s – singular
pl – plural
m – masculine
f – feminine
nt – neuter

1. p – first person
2. p. – second person
3. p. – third person

Nouns

Proportion of cases, irrespective of gender and number (1), proportion of number, irrespective of case and gender (2), proportion of gender, irrespective of case and number (3)

case	
1	p
n.	0.3292
g.	0.2836
d.	0.0209
a.	0.1967
i.	0.0631
l.	0.1067

gramm. number	
2	p
s.	0.7477
pl.	0.2525

gramm. gender	
3	p
m.	0.4546
f.	0.4011
nt.	0.1443

Proportion of cases of masculine (4), feminine (5) and neuter (6) nouns, irrespective of grammatical number

M	
4	p
n. m.	0.3639
g. m.	0.2821
d. m.	0.0259
a. m.	0.1734
i. m.	0.0561
l. m.	0.0986

F	
5	p
n. f.	0.3178
g. f.	0.284
d. f.	0.0136
a. f.	0.2107
i. f.	0.0684
l. f.	0.1079

NT	
6	p
n. nt.	0.2536
g. nt.	0.2893
d. nt.	0.0249
a. nt.	0.2325
i. nt.	0.0707
l. nt.	0.1289

Proportion of grammatical number for masculine (7), feminine (8) and neuter (9) nouns, irrespective of case

M	
7	p
s. m.	0.7525
pl. m.	0.2475

F	
8	p
s. f.	0.7131
pl. f.	0.2869

NT	
9	p
s. nt.	0.8283
pl. nt.	0.1717

Proportion of cases in singular of masculine (1), feminine (2) and neuter (3) nouns

M		F		NT	
10	p	11	p	12	p
n. m. s.	0.3841	n. f. s.	0.3165	n. nt. s	0.2467
g. m. s.	0.2562	g. f. s.	0.2819	g. nt. s	0.2967
d. m. s.	0.0259	d. f. s.	0.0135	d. nt. s	0.0267
a. m. s.	0.1642	a. f. s.	0.1962	a. nt. s.	0.2184
i. m. s.	0.0567	i. f. s.	0.0694	i. nt. s.	0.0738
l. m. s.	0.1128	l. f. s.	0.1224	l. nt. s.	0.1376

Proportion of cases in plural of masculine (4), feminine (5) and neuter (6) nouns

M		F		NT	
13	p	14	p	15	p
n. m. pl.	0.3027	n. f. pl.	0.3183	n. nt. pl	0.1724
g. m. pl.	0.3605	g. f. pl.	0.2866	g. nt. pl	0.1526
d. m. pl.	0.0255	d. f. pl.	0.0139	d. nt. pl.	0.0096
a. m. pl.	0.2013	a. f. pl.	0.2447	a. nt. pl.	0.1807
i. m. pl.	0.0543	i. f. pl.	0.0653	i. nt. pl.	0.0333
l. m. pl.	0.0556	l. f. pl.	0.0711	l. nt. pl.	0.4514

Proportion of cases singular and plural, irrespective of grammatical gender

S		PL	
16	p	17	p
n. s.	0.3367	n. pl.	0.3083
g. s.	0.2731	g. pl.	0.3163
d. s.	0.0193	d. pl.	0.0193
a. s.	0.1855	a. pl.	0.2308
i. s.	0.0644	i. pl.	0.0594
l. s.	0.1207	l. pl.	0.0669

Proportion of cases singular and plural for masculine (18), feminine (19) and neuter (20) nouns

18	p	19	p	20	p
n. m. s.	0.289	n. f. s	0.2263	n. nt. s	0.2044
g. m. s.	0.1928	g. f. s.	0.2016	g. nt. s	0.2458
d. m. s.	0.0195	d. f. s.	0.0096	d. nt. s	0.0221
a. m. s.	0.1236	a. f. s.	0.1403	a. nt. s	0.1809
i. m. s.	0.0427	i. f. s.	0.0496	i. ž. s.	0.0611
l. m. s.	0.0849	l. f. s.	0.0875	l. ž. s.	0.1139
n. m. pl.	0.0749	n. f. pl.	0.0916	n. nt. pl.	0.0445
g. m. pl.	0.0892	g. f. pl.	0.0824	g. nt. pl.	0.0436

d. m. pl.	0.0063	d. f. pl.	0.004	d. nt. pl.	0.0027
a. m. pl.	0.0498	a. f. pl.	0.0704	a. nt. pl.	0.0516
i. m. pl.	0.0134	i. f. pl.	0.0188	i. nt. pl.	0.0095
l. m. pl.	0.0136	l. f. pl.	0.0204	l. nt. pl.	0.015

ADJECTIVES: POSITIVE

Proportion of cases, irrespective of gender and number (21), proportion of number, irrespective of case and gender (22), proportion of gender, irrespective of case and number (23)

case		gramm. number		gramm. Gender	
21	p	22	p	23	P
n.	0.4039	s.	0.6814	m.	0.431
g.	0.2391	pl.	0.3176	f.	0.4226
d.	0.0192			nt.	0.1459
a.	0.200				
i.	0.0585				
l.	0.0788				

Proportion of cases of masculine (24), feminine (25) and neuter (26) nouns, irrespective of grammatical number

M		F		NT	
24	p	25	p	26	P
n. m.	0.4299	n. f.	0.3899	n. nt.	0.3669
g. m.	0.23	g. f.	0.2495	g. nt.	0.2369
d. m.	0.0215	d. f.	0.017	d. nt.	0.0188
a. m.	0.1855	a. f.	0.2061	a. nt.	0.2262
i. m.	0.0571	i. f.	0.0591	i. nt.	0.0619
l. m.	0.0759	l. f.	0.0784	l. nt.	0.0892

Proportion of grammatical number for masculine (27), feminine (28) and neuter (29) adjectives, irrespective of case

M		F		NT	
27	p	28	p	29	P
s. m.	0.6854	s. f.	0.6576	s. nt.	0.7396
Pl. m.	0.3145	pl. f.	0.3424	pl. nt.	0.2604

Proportion of cases in singular of masculine (30), feminine (31) and neuter (32) adjectives

M		F		NT	
30	p	31	p	32	p
n. m. s.	0.4622	n. f. s.	0.4072	n. nt. s	0.3814
g. m. s.	0.1962	g. f. s.	0.238	g. nt. s	0.2302
d. m. s.	0.0217	d. f. s.	0.0191	d. nt. s	0.0205

Constraints on Probability Distributions of Grammatical Forms

a. m. s.	0.1819	a. f. s.	0.1889	a. nt. s.	0.2065
i. m. s.	0.0588	i. f. s.	0.0595	i. nt. s.	0.0604
l. m. s.	0.0792	l. f. s.	0.0871	l. nt. s.	0.0974

Proportion of cases in plural of masculine (33), feminine (34) and neuter (35) adjectives

M		F		NT	
33	p	34	p	35	p
n. m. pl.	0.3597	n. f. pl.	0.3566	n. nt. pl.	0.3257
g. m. pl.	0.3037	g. f. pl.	0.2716	g. nt. pl.	0.256
d. m. pl.	0.0216	d. f. pl.	0.0129	d. nt. pl.	0.0142
a. m. pl.	0.1932	a. f. pl.	0.2389	a. nt. pl.	0.282
i. m. pl.	0.0534	i. f. pl.	0.0582	i. nt. pl.	0.0559
l. m. pl.	0.0688	l. f. pl.	0.0617	l. nt. pl.	0.0659

Proportion of cases singular and plural, irrespective of grammatical gender

S	
36	p
n. s.	0.4269
g. s.	0.2186
d. s.	0.0205
a. s.	0.1887
i. s.	0.0599
l. s.	0.0853

PL	
37	p
n. pl.	0.3543
g. pl.	0.2834
d. pl.	0.0166
a. pl.	0.2246
i. pl.	0.0559
l. pl.	0.0652

Proportion of cases singular and plural for masculine (38), feminine (39) and neuter (40) adjectives.

imenice	
38	p
n. m. s.	0.3168
g. m. s.	0.1345
d. m. s.	0.0149
a. m. s.	0.1247
i. m. s.	0.0403
l. m. s.	0.0543
n. m. pl.	0.1132
g. m. pl.	0.0955
d. m. pl.	0.0067
a. m. pl.	0.0608
i. m. pl.	0.0168
l. m. pl.	0.0216

imenice	
39	p
n. f. s.	0.2678
g. f. s.	0.1565
d. f. s.	0.0126
a. f. s.	0.1243
i. f. s.	0.0391
l. f. s.	0.0573
n. f. pl.	0.1221
g. f. pl.	0.093
d. f. pl.	0.0044
a. f. pl.	0.0818
i. f. pl.	0.0199
l. f. pl.	0.0211

imenice	
40	p
n. nt. s.	0.2821
g. nt. s.	0.1702
d. nt. s.	0.0151
a. nt. s.	0.1527
i. ž. s.	0.0474
l. ž. s.	0.0721
n. nt. pl.	0.0849
g. nt. pl.	0.0667
d. nt. pl.	0.0037
a. nt. pl.	0.0734
i. nt. pl.	0.0145
l. nt. pl.	0.0172

ADJECTIVES: COMPARATIVE

Proportion of cases, irrespective of gender and number (41), proportion of number, irrespective of case and gender (42), proportion of gender, irrespective of case and number (43)

case		gramm. number		gramm. gender	
41	p	42	p	43	p
n.	0.4931	s.	0.7294	m.	0.471
g.	0.1416	pl.	0.2706	f.	0.3841
d.	0.0211			nt.	0.1449
a.	0.2364				
i.	0.0564				
l.	0.0514				

Proportion of cases of masculine (44), feminine (45) and neuter (46) nouns, irrespective of grammatical number

M		F		NT	
44	p	45	p	46	p
n. m.	0.5223	n. f.	0.4846	n. nt.	0.4208
g. m.	0.1378	g. f.	0.138	g. nt.	0.1636
d. m.	0.0186	d. f.	0.0169	d. nt.	0.0382
a. m.	0.235	a. f.	0.2334	a. nt.	0.2493
i. m.	0.0425	i. f.	0.0715	i. nt.	0.062
l. m.	0.0437	l. f.	0.0556	l. nt.	0.0659

Proportion of grammatical number for masculine (47), feminine (48) and neuter (49) adjectives, irrespective of case

M		F		NT	
47	p	48	p	49	p
s. m.	0.7152	s. f.	0.7234	s. nt.	0.7929
pl. m.	0.2848	pl. f.	0.2765	pl. nt.	0.2071

Proportion of cases in singular of masculine (50), feminine (51) and neuter (52) adjectives

M		F		NT	
50	p	51	p	52	p
n. m. s.	0.5524	n. f. s.	0.4996	n. nt. s.	0.4343
g. m. s.	0.1105	g. f. s.	0.1228	g. nt. s.	0.1564
d. m. s.	0.0153	d. f. s.	0.0192	d. nt. s.	0.0483
a. m. s.	0.2334	a. f. s.	0.2285	a. nt. s.	0.2413
i. m. s.	0.0431	i. f. s.	0.0734	i. nt. s.	0.0665
l. m. s.	0.0453	l. f. s.	0.0563	l. nt. s.	0.0532

Proportion of cases in plural of masculine (53), feminine (54) and neuter (55) adjectives

M		F		NT	
53	p	54	p	55	p

Constraints on Probability Distributions of Grammatical Forms

n. m. pl.	0.4467	n. f. pl.	0.4452	n. nt. pl	0.3694
g. m. pl.	0.2063	g. f. pl.	0.1777	g. nt. pl	0.191
d. m. pl.	0.027	d. f. pl.	0.0108	d. nt. pl.	0.0003
a. m. pl.	0.2389	a. f. pl.	0.2549	a. nt. pl.	0.2802
i. m. pl.	0.0413	i. f. pl.	0.0664	i. nt. pl.	0.0445
l. m. pl.	0.0398	l. f. pl.	0.0539	l. nt. pl.	0.1146

Proportion of cases singular and plural, irrespective of grammatical gender

S		PL	
56	p	57	p
n. s.	0.5137	n. pl.	0.4375
g. s.	0.1224	g. pl.	0.1934
d. s.	0.0219	d. pl.	0.0176
a. s.	0.2328	a. pl.	0.2463
i. s.	0.0583	i. pl.	0.0515
l. s.	0.0507	l. pl.	0.0536

Proportion of cases singular and plural for masculine (58), feminine (59) and neuter (60) adjectives.

58	p	59	p	60	p
n. m. s.	0.395	n. f. s	0.3615	n. nt. s	0.3443
g. m. s.	0.079	g. f. s.	0.0889	g. nt. s	0.124
d. m. s.	0.0109	d. f. s.	0.0139	d. nt. s	0.0382
a. m. s.	0.1669	a. f. s.	0.1653	a. nt. s	0.1913
i. m. s.	0.0308	i. f. s.	0.0531	i. ž. s.	0.0527
l. m. s.	0.0324	l. f. s.	0.0407	l. ž. s.	0.0422
n. m. pl.	0.1272	n. f. pl.	0.1231	n. nt. pl.	0.0765
g. m. pl.	0.0587	g. f. pl.	0.0491	g. nt. pl.	0.0395
d. m. pl.	0.0077	d. f. pl.	0.0029	d. nt. pl.	0.0004
a. m. pl.	0.0681	a. f. pl.	0.068	a. nt. pl.	0.058
i. m. pl.	0.0117	i. f. pl.	0.0184	i. nt. pl.	0.0092
l. m. pl.	0.0113	l. f. pl.	0.0149	l. nt. pl.	0.0237

ADJECTIVES: SUPERLATIVE

Proportion of cases, irrespective of gender and number (61), proportion of number, irrespective of case and gender (62), proportion of gender, irrespective of case and number (63)

case		gramm. number		gramm. gender	
61	p	62	p	63	p
n.	0.4339	s.	0.6271	m.	0.4957
g.	0.2052	pl.	0.3727	f.	0.3699

d.	0.0208	nt.	0.1342
a.	0.2151		
i.	0.0535		
l.	0.0713		

Proportion of cases of masculine (64), feminine (65) and neuter (66) nouns, irrespective of grammatical number

M		F		NT	
64	p	65	p	66	p
n. m.	0.4419	n. f.	0.4321	n. nt.	0.4092
g. m.	0.2144	g. f.	0.1978	g. nt.	0.1924
d. m.	0.0279	d. f.	0.0148	d. nt.	0.0108
a. m.	0.1857	a. f.	0.2372	a. nt.	0.2629
i. m.	0.0573	i. f.	0.0482	i. nt.	0.0542
l. m.	0.0729	l. f.	0.0699	l. nt.	0.0705

Proportion of grammatical number for masculine (67), feminine (68) and neuter (69) adjectives, irrespective of case

M		F		NT	
67	p	68	p	69	p
s. m.	0.5778	s. f.	0.6673	s. nt.	0.6992
pl. m.	0.4222	pl. f.	0.3327	pl. nt.	0.3008

Proportion of cases in singular of masculine (70), feminine (71) and neuter (72) adjectives

M		F		NT	
70	p	71	p	72	p
n. m. s.	0.5464	n. f. s.	0.5	n. nt. s.	0.4845
g. m. s.	0.1105	g. f. s.	0.1386	g. nt. s.	0.1473
d. m. s.	0.0241	d. f. s.	0.0133	d. nt. s.	0.0116
a. m. s.	0.1525	a. f. s.	0.2271	a. nt. s.	0.2442
i. m. s.	0.0609	i. f. s.	0.0428	i. nt. s.	0.0504
l. m. s.	0.1055	l. f. s.	0.0826	l. nt. s.	0.062

Proportion of cases in plural of masculine (73), feminine (74) and neuter (75) adjectives

M		F		NT	
73	p	74	p	75	p
n. m. pl.	0.2991	n. f. pl.	0.2959	n. nt. pl.	0.2342
g. m. pl.	0.3565	g. f. pl.	0.3166	g. nt. pl.	0.2973
d. m. pl.	0.033	d. f. pl.	0.0177	d. nt. pl.	0.009
a. m. pl.	0.2313	a. f. pl.	0.2663	a. nt. pl.	0.3063
i. m. pl.	0.0522	i. f. pl.	0.0592	i. nt. pl.	0.0631

Constraints on Probability Distributions of Grammatical Forms

l. m. pl .	0.0278	l. f. pl.	0.0443	l. nt. pl.	0.0901
------------	--------	-----------	--------	------------	--------

Proportion of cases singular and plural, irrespective of grammatical gender

S		PL	
76	p	77	p
n. s.	0.5189	n. pl.	0.291
g. s.	0.1271	g. pl.	0.3369
d. s.	0.0179	d. pl.	0.0254
a. s.	0.1938	a. pl.	0.2509
i. s.	0.0522	i. pl.	0.0557
l. s.	0.0899	l. pl.	0.04

Proportion of cases singular and plural for masculine (78), feminine (79) and neuter (80) adjectives.

imenice		imenice		imenice	
78	p	79	p	80	p
n. m. s.	0.3157	n. f. s	0.3337	n. nt. s	0.3387
g. m. s.	0.0639	g. f. s.	0.0925	g. nt. s	0.1029
d. m. s.	0.0139	d. f. s.	0.0089	d. nt. s	0.0081
a. m. s.	0.0881	a. f. s.	0.1486	a. nt. s	0.1707
i. m. s.	0.0352	i. f. s.	0.0285	i. ž. s.	0.0352
l.m.s.	0.0609	l. f. s.	0.0551	l. ž. s.	0.0434
n. m. pl.	0.1263	n. f. pl.	0.0984	n. nt. pl.	0.0705
g. m. pl.	0.1505	g. f. pl.	0.1053	g. nt. pl.	0.0894
d. m. pl.	0.0139	d. f. pl.	0.0059	d. nt. pl.	0.0027
a. m. pl.	0.0976	a. f. pl.	0.0886	a. nt. pl.	0.0921
i. m. pl.	0.022	i. f. pl.	0.0197	i. nt. pl.	0.0189
l. m. pl.	0.0117	l. f. pl.	0.0148	l. nt. pl.	0.0271

VERBS: PRESENT TENSE

Proportion of verb person singular and plural (81), proportion of grammatical number (82)

person		gramm. number	
81	p	82	p
1. p. s.	0.1533	s.	0.7117
2. p. s.	0.0487	pl.	0.2883
3. p. s.	0.5097		
1. p. pl.	0.0291		
2. p. pl.	0.0088		
3. p. pl.	0.2504		

Proportion of verb person singular (83) and plural (84)

83	p	84	p
1. p. s.	0.2154	1. p. pl.	0.1010
2. p. s.	0.0684	2. p. pl.	0.0304
3. p. s.	0.7162	3. p. pl.	0.8685

VERBS: FUTURE TENSE*Proportion of verb person singular and plural (85), proportion of grammatical number (86)*

person		gramm. number	
85	p	86	p
1. p. s.	0.1703	s.	0.7474
2. p. s.	0.0683	pl.	0.2526
3. p. s.	0.5087		
1. p. pl.	0.0534		
2. p. pl.	0.0380		
3. p. pl.	0.1613		

Proportion of verb person singular (87) and plural (88)

87	p	88	p
1. p. s.	0.2279	1. p. pl.	0.2113
2. p. s.	0.0914	2. p. pl.	0.1503
3. p. s.	0.6807	3. p. pl.	0.6384

VERBS: AORIST*Proportion of verb person singular and plural (89), proportion of grammatical number (90)*

person		gramm. number	
89	p	90	p
1. p. s.	0.2105	s.	0.8263
2. p. s.	0.0245	pl.	0.1737
3. p. s.	0.5914		
1. p. pl.	0.0278		
2. p. pl.	0.0040		
3. p. pl.	0.1419		

Proportion of verb person singular (91) and plural (92)

91	p	92	p
1. p. s.	0.2547	1. p. pl.	0.1601
2. p. s.	0.0296	2. p. pl.	0.0229
3. p. s.	0.7157	3. p. pl.	0.8171

VERBS: IMPERFECT*Proportion of verb person singular and plural (93), proportion of grammatical number (94)*

person		gramm. number	
93	p	94	p
1. p. s.	0.2911	s.	0.7440
2. p. s.	0.0121	pl.	0.2560
3. p. s.	0.4408		
1. p. pl.	0.0254		

2. p. pl.	0.0145
3. p. pl	0.2162

Proportion of verb person singular (95) and plural (96)

95	p	96	p
1. p. s.	0.3912	1. p. pl.	0.0991
2. p. s.	0.0162	2. p. pl.	0.0566
3. p. s.	0.5925	3. p. pl	0.8443

VERBS: PAST TENSE (PERFECT)

Proportion of verb person singular and plural (97), proportion of grammatical number (98) and proportion of grammatical gender (99)

person	
97	p
1. p. s.	0.1137
2. p. s.	0.0403
3. p. s.	0.5756
1. p. pl.	0.0344
2. p. pl.	0.0071
3. p. pl	0.2288

gramm. number	
98	p
s.	0.7296
pl.	0.2704

gramm. gender	
99	p
m.	0.6345
f.	0.2747
nt.	0.0908

Proportion of verb person singular (100) and plural (101)

100	p	101	p
1. p. s.	0.1558	1. p. pl.	0.1274
2. p. s.	0.0552	2. p. pl.	0.0264
3. p. s.	0.7890	3. p. pl	0.8463

Proportion of verb person singular and plural masculine (102), feminine (103) and neuter (104) gender

102	p	103	p	104	p
1. p. s. nt.	0.1553	1. p. s. f.	0.0544	1. p. s. nt.	0.0023
2. p. s. nt.	0.0324	2. p. s. f.	0.0687	2. p. s. nt.	0.0090
3. p. s. nt.	0.5076	3. p. s. f.	0.6536	3. p. s. nt.	0.8151
1. p. pl. nt.	0.0536	1. p. pl. f.	0.0015	1. p. pl. nt.	0.0002
2. p. pl. nt.	0.0103	2. p. pl. f.	0.0022	2. p. pl. nt.	0.0001
3. p. pl. nt.	0.2408	3. p. pl. f.	0.2195	3. p. pl. nt.	0.1733

VERBS: PLUSQUAMPERFECT

Proportion of verb person singular and plural (105), proportion of grammatical number (106) and proportion of grammatical gender (107)

person	
105	p
1. p. s.	0.0690
2. p. s.	0.0402
3. p. s.	0.3985

gramm. number	
106	p
s.	0.5077
pl.	0.4923

gramm. gender	
107	p
m.	0.7625
f.	0.1667
nt.	0.0709

1. p. pl.	0.0096
2. p. pl.	0.0038
3. p. pl.	0.4789

Plusquamperfect. Proportion of verb person singular (108) and plural (109)

108	p	109	p
1. p. s.	0.1358	1. p. pl.	0.0195
2. p. s.	0.0792	2. p. pl.	0.0078
3. p. s.	0.7849	3. p. pl.	0.9728

Plusquamperfect. Proportion of verb person singular and plural masculine (110), feminine (111) and neuter (112) gender

110	p	111	p	112	p
1. p. s. nt.	0.0653	1. p. s. f.	0.0115	1. p. s. nt.	0.2432
2. p. s. nt.	0.0452	2. p. s. f.	0.0345	2. p. s. nt.	0.0001
3. p. s. nt.	0.3015	3. p. s. f.	0.7356	3. p. s. nt.	0.6486
1. p. pl. nt.	0.0126	1. p. pl. f.	0.0001	1. p. pl. nt.	0.0000
2. p. pl. nt.	0.0050	2. p. pl. f.	0.0001	2. p. pl. nt.	0.0001
3. p. pl. nt.	0.5704	3. p. pl. f.	0.2184	3. p. pl. nt.	0.1081

PROBABILITIES OF VERB PERSON AND NUMBER**IRRESPECTIVE OF TENSE AND GENDER*****Proportion of verb person singular and plural (113), proportion of grammatical number (114)***

person	
113	p
1. p. s.	0.1480
2. p. s.	0.0478
3. p. s.	0.5671
1. p. pl.	0.0279
2. p. pl.	0.0072
3. p. pl.	0.2020

gramm. number	
114	p
s.	0.7209
pl.	0.2791

Proportion of verb person singular (115) and plural (116)

115	p	116	p
1. p. s.	0.1940	1. p. pl.	0.1182
2. p. s.	0.0626	2. p. pl.	0.0304
3. p. s.	0.7434	3. p. pl.	0.8514

APPENDIX 2**Values of entropy, maximum entropy and relative entropy for all paradigms**

tab.	H	Hmax	Hr	tab.	H	Hmax	Hr
1	2.2147	2.5845	0.8579	59	2.7204	3.5841	0.7590
2	0.8150	1.0000	0.8150	60	2.8391	3.5841	0.7921
3	0.1449	1.5850	0.9140	61	2.0822	2.5845	0.8057
4	2.1130	2.5845	0.8176	62	1.4215	1.5850	0.8968

5	2.1534	2.5845	0.8332	63	0.9529	1.0000	0.9529
6	2.0429	2.5845	0.7905	64	2.1039	2.5845	0.8140
7	0.8073	1.0000	0.8073	65	2.0470	2.5845	0.7290
8	0.8647	1.0000	0.8647	66	2.0600	2.5845	0.7970
9	0.6616	1.0000	0.6616	67	0.9825	1.0000	0.9825
10	2.1880	2.5845	0.8466	68	0.9177	1.0000	0.9177
11	2.2231	2.5845	0.8602	69	0.8823	1.0000	0.8823
12	2.3084	2.5845	0.8932	70	1.9591	2.5845	0.7580
13	2.1130	2.5845	0.8176	71	1.9555	2.5845	0.7566
14	2.1534	2.5845	0.8332	72	1.9507	2.5845	0.7548
15	2.0429	2.5845	0.7905	73	2.0683	2.5845	0.8003
16	2.2231	2.5845	0.8602	74	2.0971	2.5845	0.8114
17	2.3084	2.5845	0.8932	75	2.1591	2.5845	0.8354
18	2.9757	3.5841	0.8303	76	1.9669	2.5845	0.7610
19	3.0716	3.5841	0.857	77	2.1000	2.5845	0.8125
20	2.9322	3.5841	0.8181	78	2.9868	3.5841	0.8334
21	2.1424	2.5845	0.8219	79	2.9183	3.5841	0.8142
22	1.4536	1.5850	0.9171	80	2.8950	3.5841	0.8077
23	0.9026	1.0000	0.9026				
24	2.0994	2.5845	0.8123	101	1.8313	2.5845	0.7086
25	2.1282	2.5845	0.8234	102	0.8665	1.0000	0.8665
26	2.1752	2.5845	0.8416	103	1.0866	1.5850	0.6856
27	0.8984	1.0000	0.8984	104	0.6641	1.5850	0.4190
28	0.9271	1.0000	0.9271	105	2.0248	2.5845	0.7834
29	0.8273	1.0000	0.8273	106	0.8154	1.0000	0.8154
30	2.0729	2.5845	0.8020	107	1.1794	1.5850	0.7441
31	2.1329	2.5845	0.8252	108	1.2981	1.5850	0.8190
32	2.1749	2.5845	0.8415	109	1.6275	2.5845	0.6297
33	2.1219	2.5845	0.8210	110	0.6660	1.0000	0.6660
34	2.1024	2.5845	0.8135	111	0.9983	1.5850	0.6299
35	2.1234	2.5845	0.8217	112	0.7859	1.5850	0.4959
36	2.1189	2.5845	0.8199	113	1.8168	2.5845	0.7030
37	2.1174	2.5845	0.8193	114	0.8207	1.0000	0.8207
38	2.9865	3.5841	0.8333	115	1.0736	1.5850	0.6774
39	3.0494	3.5841	0.8508	116	0.7710	1.5850	0.4865
40	2.9957	3.5841	0.8358	117	1.2070	2.5845	0.6605
41	1.9657	2.5845	0.7606	118	0.8420	1.0000	0.8420
42	1.4456	1.5850	0.9121	119	0.9184	1.5850	0.5794
43	0.8423	1.0000	0.8423	120	0.7208	1.5850	0.4548
44	1.8723	2.5845	0.7245	121	1.2427	1.5850	0.7841
45	1.9941	2.5845	0.7716	122	1.8629	2.5845	0.7208
46	2.1396	2.5845	0.8279	123	1.409	2.5845	0.5452
47	0.8619	1.0000	0.8619	124	0.7641	2.5845	0.2956
48	0.8507	1.0000	0.8507	125	1.5852	2.5845	0.6133
49	0.7359	1.0000	0.7359	126	0.9998	1.0000	0.9998
50	1.8041	2.5845	0.6980	127	0.9998	1.5850	0.6308
51	1.9781	2.5845	0.7654	128	0.9553	1.5850	0.6028
52	2.1325	2.5845	0.8251	129	0.2039	1.5850	0.1286

53	1.9983	2.5845	0.7732	130	1.5604	2.5845	0.6038
54	2.0228	2.5845	0.7827	131	1.0494	2.5845	0.4061
55	2.0627	2.5845	0.7981	132	1.251	2.5845	0.484
56	1.9320	2.5845	0.7475	133	1.7432	2.5845	0.6745
57	2.0273	2.5845	0.7844	134	0.8542	1.0000	0.5842
58	2.7204	3.5841	0.7590	135	1.0274	1.5850	0.6482
				136	0.715	1.5850	0.4511

APPENDIX 3

Number of paradigms with a given relative entropy value

H_r	N	H_r	N	H_r	N	H_r	N
.01	0	.26	0	.51	0	.76	6
.02	0	.27	0	.52	0	.77	4
.03	0	.28	0	.53	0	.78	4
.04	0	.29	0	.54	1	.79	3
.05	0	.30	1	.55	0	.80	5
.06	0	.31	0	.56	0	.81	11
.07	0	.32	0	.57	0	.82	16
.08	0	.33	0	.58	2	.83	15
.09	0	.34	0	.59	0	.84	5
.10	0	.35	0	.60	3	.85	3
.11	0	.36	0	.61	1	.86	6
.12	0	.37	0	.62	0	.87	1
.13	1	.38	0	.63	3	.88	1
.14	0	.39	0	.64	0	.89	3
.15	0	.40	0	.65	1	.90	3
.16	0	.41	1	.66	2	.91	1
.17	0	.42	1	.67	2	.92	2
.18	0	.43	0	.68	1	.93	1
.19	0	.44	0	.69	1	.94	0
.20	0	.45	2	.70	2	.95	1
.21	0	.46	0	.71	1	.96	0
.22	0	.47	0	.72	2	.97	0
.23	0	.48	1	.73	1	.98	1
.24	0	.49	1	.74	2	.99	1
.25	0	.50	1	.75	2	1.00	0

REZIME

RESTRIKCIJE U OKVIRU DISTRIBUCIJA VEROVATNOĆA GRAMATIČKIH OBLIKA

Aleksandar Kostić i Milena Božić

Laboratorija za eksperimentalnu psihologiju, Filozofski fakultet u Beogradu

U ovom radu ispitane su restrikcije u okviru distribucija verovatnoća gramatičkih oblika za paradigme srpskog jezika, pri čemu je “paradigma” definisana kao koherentni skup elemenata sa jasno određenim kriterijumom njihovog uključivanja (npr. gramatički rod - muški, ženski, srednji; gramatički broj - jednina i množina; padež – nominativ, genitiv, dativ itd.). Pojam paradigme može da se proširi i na bilo koji drugi skup sa jasno definisanim kriterijumima (npr. padež s obzirom na gramatički rod i broj ili padež s obzirom na gramatički broj ali ne i na gramatički rod itd.). Svaki element ovako definisane paradigme javlja se sa određenom verovatnoćom. Postavlja se pitanje u kojoj meri je distribucija elemenata unutar paradigmi homogena. U ovoj studiji kao meru homogenosti uzeli smo relativnu entropiju koja je određena kao količnik dobijene i maksimalne entropije. Analiza relativnih entropija za 116 paradigmi srpskog jezika pokazuje da se njihove vrednosti u najvećem broju slučajeva kreću između 0.75 i 0.9. Razlike u vrednostima relativnih entropija dozvoljavaju da se izračuna i relativna entropija morfološkog sistema u celini. Ovakva analiza pokazuje da relativna entropija morfološkog sistema srpskog jezika iznosi 0.69 što se, uslovno, može shvatiti i kao indeks njegove stabilnosti.

Ključne reči: *morfologija, gramatička paradigma, relativna entropija*