2013

Grammatical Typology and Frequency Analysis: Number Availability and Number Use

Dunstan Brown
University of York, UK

Greville G. Corbett
University of Surrey, UK

Sebastian Fedden
University of Surrey, UK

Andrew R. Hippisley
University of Kentucky, andrew.hippisley@uky.edu

Paul Marriott
University of Waterloo, Canada

Repository Citation
Brown, Dunstan; Corbett, Greville G.; Fedden, Sebastian; Hippisley, Andrew R.; and Marriott, Paul, "Grammatical Typology and Frequency Analysis: Number Availability and Number Use" (2013). *Linguistics Faculty Publications*. 64.
https://uknowledge.uky.edu/lin_facpub/64

Grammatical Typology and Frequency Analysis: Number Availability and Number Use

Digital Object Identifier (DOI)
http://dx.doi.org/10.15398/jlm.v1i2.69

Notes/Citation Information
Published in *Journal of Language Modelling*, v. 1, no. 2, p. 227-241.

Grammatical typology and frequency analysis: number availability and number use

Dunstan Brown[1], *Greville G. Corbett*[2], *Sebastian Fedden*[2], *Andrew Hippisley*[3], *and Paul Marriott*[4]
[1] University of York, UK
[2] University of Surrey, UK
[3] University of Kentucky, USA
[4] University of Waterloo, Canada

ABSTRACT

The Smith-Stark hierarchy, a version of the Animacy Hierarchy, offers a typology of the cross-linguistic availability of number. The hierarchy predicts that the availability of number is not arbitrary. For any language, if the expression of plural is available to a noun, it is available to any noun of a semantic category further to the left of the hierarchy. In this article we move one step further by showing that the structure of the hierarchy can be observed in a statistical model of number use in Russian. We also investigate three co-variates: plural preference, pluralia tantum and irregularity effects; these account for an item's behaviour being different than that solely expected from its animacy position.

Keywords: animacy hierarchy, frequency, number, Russian

1 INTRODUCTION

The morphosyntactic feature of number is found in many languages; it has the values singular and plural, and often others too, such as dual. Number distinctions and the availability of number have been generally well-studied cross-linguistically. One of the most important contributions in this area was the Smith-Stark hierarchy (Smith-Stark 1974), discussed in Corbett (2000). This hierarchy, often also called the Animacy Hierarchy, offers a typology of the availability of number in languages. In this article we move one step further by demonstrating that the structure of the Smith-Stark hierarchy can be observed in the use of the number feature in Russian[1]. The hierarchy we use in this paper, which is adapted from Smith-Stark (1974) is given in (1):

Speaker > Addressee > Kin > Non-human rational > (1)
Human rational > Human non-rational > Animate >
Concrete inanimate > Abstract inanimate

The labels 'speaker' and 'addressee' are used for the first and second person pronouns. The other positions of the Smith-Stark hierarchy in (1) are universally applicable lexical categories. We also refer to them as the animacy category of a noun. Nouns of the non-human rational category denote supernatural beings. Human rationals include humans except children, which belong in the Human non-rational category. Corbett (2000) points out that the rational/non-rational distinction has limited justification. However, given the typological importance of the Smith-Stark hierarchy, we took the decision only to extend distinctions within the hierarchy rather than eliminate any. We therefore maintained the human rational/non-rational distinction, and we also added a distinction of concrete and abstract within inanimates, which meant that the original structure of the hierarchy is recoverable. The hierarchy predicts that the availability of number is not arbitrarily distributed. For any language, if the expression of plural is available to a noun it is likewise available to any noun of a semantic category towards the left of the hierarchy. For example, if a language has a singular-plural contrast in animate nouns, it will also have such a contrast in human non-rational, human rational, and non-human rational nouns, kin nouns and the second and first person pronouns. In other words, there is a cut-off point somewhere along the hierarchy. Left of this point, plural is available; further down the hierarchy to the right of this point, plural is not available.

[1] The research reported here was originally funded by the ESRC (UK) under grant R000222419. For the time for recent updating, Brown and Corbett are indebted to the European Research Council under grant ERC-2008-AdG-230268 MORPHOLOGY. The support of both funding bodies is gratefully acknowledged. We thank Alexander Krasovitsky for helpful discussion of specific Russian examples.

The Smith-Stark hierarchy is a typological generalization and as such should be valid cross-linguistically. Our hypothesis is that the use of the grammatical category *number* can be predicted from a typology which in turn makes predictions about the availability of number. A necessary way of testing this generalization is to apply it to a test language. Russian was selected since number is (generally) available to nominals, and the rich morphology of Russian typically makes the expression of number clear, as can be shown by the items in (2) which exemplify each of the different points on the hierarchy.

ja 'I'	vs.	*my* 'we'	[speaker]	(2)
ty 'you (singular)'	vs.	*vy* 'you (plural)'	[addressee]	
otec 'father'	vs.	*otcy* 'fathers'	[kin]	
bog 'god'	vs.	*bogi* 'gods'	[non-human rational]	
podruga 'girlfriend'	vs.	*podrugi* 'girlfriends'	[human rational]	
rebenok 'child'	vs.	*deti* 'children'	[human non-rational]	
lošad' 'horse'	vs.	*lošadi* 'horses'	[animate]	
stol 'table'	vs.	*stoly* 'tables'	[inanimate]	
sistema 'system'	vs.	*sistemy* 'systems'	[abstract inanimate]	

This article has four sections. In section 2 we give a summary of our methods and the statistical model we used in our study. In section 3 we present the results of our study. We show that there is a relationship between the points in the availability hierarchy and number use, but that other co-variates can come into play that result in a much higher plural proportion than expected from the position on the hierarchy. This is for example the case for nouns whose referents typically come in pairs (*glaz* 'eye') or in multitudes (*gramm* 'gramme'), and for pluralia tantum, such as *rebjatiški* 'kids', i.e., nouns which have only plural forms. Finally, we give our conclusions.

2 METHODS AND STATISTICAL MODEL

In this section we outline the methods used for data preparation and data analysis. We also sketch the statistical model used in this research.

2.1 *Data preparation*

To test our hypotheses, we used the corpus of contemporary Russian texts prepared at Uppsala University, Lönngren (1993), which contains about one million tokens. At the time the research was carried out this was the most suitable corpus of Russian as far as scope and design were concerned, as it covered a range of texts within a 25-year time period (1960–1985).[2]

We prepared the data as follows. Nouns were taken from the corpus and marked for semantic, morphosyntactic, and frequency information. The dataset contains 5,450 noun and pronoun lexemes occurring five or more times, with morphosyntactic and frequency information about their 243,466 word forms. This includes first and second person pronouns, but excludes third person pronouns. The third person deserves a separate study; there are around 29,000 examples of third person pronouns in the corpus. We used the concordance tool 'WordSmith' (Oxford University Press) to extract the nouns from the corpus and we indexed them according to position on the Smith-Stark hierarchy, and recorded number information, i.e., the distribution of singulars and plurals. This information was formatted in Microsoft Excel and encoded in such a way so as to facilitate statistical analysis. In particular we noted for each lexeme the proportion of plural forms being used. Numerical values were given for all information on animacy category, i.e., position on the Smith-Stark hierarchy, case and number. The statistical software package used for data analysis was S-PLUS.

The dataset resulting from our study has been made available on our web site.[3]

2.2 *Statistical model*

A number of differing modelling approaches were used for the analysis. The non-parametric bootstrap (Efron and Tibshirani 1993) was

[2] The offline version of the Russian National Corpus is a similar size (see `http://ruscorpora.ru/corpora-usage.html`), while the online version is much bigger. The semantic categories available for searching the online version should map straightforwardly onto the Smith-Stark hierarchy, but currently it is not possible to download the full results of a search. Replicating our results using the RNC would, of course, be a useful future piece of research. For more on the RNC and its history see Grišina and Plungian (2005). See Maier (1994) for more information on the Uppsala corpus.

[3] `http://www.surrey.ac.uk/englishandlanguages/research/smg/files/rusnoms.xls`

used to test if there was a significant difference between the median values of plural usage between groups, while the two sample Kolmogorov-Smirnov test (Conover 1971) was used to test for differences in distributions of the plural usage, again across pairs of groups defined by the hierarchy. The results from non-parametric approaches were checked using a parametric approach using the log-likelihood for inference. The S-PLUS code for this model and explanatory text has been made available at the Surrey Morphology Group website.[4]

Since the results for the parametric method were qualitatively the same as the non-parametric, only the non-parametric results are reported here.

In order to test the differences between the median values of two groups, the bootstrap, a form of randomisation, was used. We extract a subset of lexemes S from the corpus C according to animacy category. We calculate the median frequency of the distribution of the required frequency. Denote this to be $m(S)$ in the subset S and $m(C)$ in the full corpus, C. We need to see if $m(S)$ is significantly different from $m(C)$ assuming the null hypothesis that there is no relationship between the extraction criterion (animacy category) and the measure quantity (frequency). Under this assumption we can evaluate the distribution of $m(S)$ by randomly selecting (with replacement) samples of equal size to S from C, and calculating their median. This procedure is repeated many times and an estimate of the underlying distribution of the median is constructed. This will be the bootstrap distribution of the median under the assumed hypothesis. The actual value of $m(S)$ can then be compared to this bootstrapped distribution to see if it is extreme. A p-value can then be directly calculated from the bootstrap distribution. For details of this procedure see Efron and Tibshirani (1993), Chapter 13.

Initially, informal graphical methods were used to explore the data before any modelling or formal testing was done. The exploratory data analysis showed observed proportions varying continuously in the range from 0 to 1, but also with appreciable finite atoms of probability at exactly 0 or 1. Hence a mixture model was selected using

[4] `http://www.surrey.ac.uk/englishandlanguages/research/smg/files/statisticalmodel.pdf`

a beta distribution as a continuous model for the interval $(0,1)$ and with the discrete atoms modelled separately. The model was fitted using maximum likelihood and showed very good agreement with the data.

3 RESULTS AND DISCUSSION

In this section we present the details of the results of our investigation into number use in Russian and discuss those cases in which the proportion of plural forms was much higher than we would expect from the position on the hierarchy.

3.1 *The relation between plural marking and hierarchy position*

We analysed 5,450 Russian noun and pronoun lexemes from the Uppsala corpus according to the methodology outlined in Section 2.1, which were represented by 243,466 word forms. We recorded lexemes for their distribution of singular and plural forms, as well as for their animacy category. The sample details are given in Table 1.

The p-value in the second rightmost column in Table 2 represents the probability that the observed median was due to chance variation computed via the bootstrap. The p-value in the last column is from the Kolmogorov-Smirnov test. There is very strong evidence that there is structure in most of the categories. (A value less than 0.05 is

Table 1: Details of the sample of Russian nouns

Animacy category	Lexeme frequency	Word-form frequency	Word-form proportion of sample (%)
Speaker	1	9,610	3.9
Addressee	2	2,805	1.2
Kin	45	4,155	1.7
Non-human rational	5	267	0.1
Human rational	498	17,127	7.0
Human non-rational	28	2,054	0.8
Animate	102	2,826	1.2
Concrete inanimate	2,437	93,442	38.4
Abstract inanimate	2,332	111,180	45.7
TOTALS	5,450	243,466	100

strong evidence that the group is significantly different from the corpus.) From Table 2 we see that the evidence is less strong for Speaker, Addressee, and Non-human rational. The group Kin was significant using the Kolmogorov-Smirnov comparing distributions.

Table 3 gives the *p*-values for pairwise tests of equality of distribution across the groups in the hierarchy.

These results give more structure to the patterns shown later in Figure 1. Thus, for example, we see that while the Human non-rational and Animate groups are significantly different from the corpus as a whole (Table 2), they are not different from each other (Table 3). On the other hand, groups at the lower end of the hierarchy are both different from the corpus and different from each other. These results show how the structure of the hierarchy is reflected in the observed distribution of number use. It is clear that the position that a lexeme takes in the Smith-Stark hierarchy can have a strong effect on the proportion of one number (plural) being used over another. We can compare the hierarchy for number availability with the broad picture

Table 2: Details of the sample of Russian nouns

Animacy category	Singular forms	Plural forms	Singular + plural forms	Mean plural proportion	Median plural proportion	*p*-value Bootstrap	*p*-value K-S test
Speaker	6197	3413	9610	35.5%	35.5%	0.83	0.75
Addressee	2600	205	2805	8.7%	8.7%	0.43	0.71
Kin	3733	422	4155	14.7%	5%	0.07	<0.001
Non-human rational	248	19	267	5.8%	5.5%	0.46	0.12
Human rational	9392	7735	17127	45.1%	45.5%	< 0.001	< 0.001
Human non-rational	854	1200	2054	58.4%	61.8%	< 0.001	< 0.001
Animate	1599	1227	2826	43.4%	48.1%	< 0.001	< 0.001
Concrete inanimate	65427	28015	93442	30%	23.1%	< 0.001	< 0.001
Abstract inanimate	84698	26482	111180	23.8%	0.5%	< 0.001	< 0.001
TOTALS	174,748	68,718	243,466	28.2%	16.7%		

of the results of our investigation into number use. The Smith-Stark hierarchy is given in (3), repeated from (1) above.

Speaker > Addressee > Kin > Non-human rational > (3)
Human rational > Human non-rational > Animate >
Concrete inanimate > Abstract inanimate

We have made explicit the distinction between human rational and human non-rational (children), and extended the hierarchy to distinguish inanimates that are concrete from inanimates that are abstract. The classes which distinguish singular and plural occupy the upper segments of the hierarchy, and languages make the split between items distinguishing number and those failing to do so at different points of the hierarchy.

Our investigation into number use yielded statistically significant results. We can compare the version of Smith-Stark's hierarchy for number availability in (3) with the picture of number use in Figure 1.

The data are structured with each animacy position having its own median point. The median is represented by the line in the middle of the box; the box itself represents a range of proportions covering the middle 50% of the lexemes in the category; the whiskers cover the

Table 3: Comparison of pairs of groups in the hierarchy

Animacy category	Addressee	Kin	Non-human rational	Human rational	Human non-rational	Animate	Concrete inanimate	Abstract inanimate
Speaker	0.667	0.422	0.375	0.856	0.820	0.858	0.815	0.567
Addressee	–	1.000	0.867	0.196	0.080	0.179	0.519	0.977
Kin		–	0.906	< **0.001**	< **0.001**	< **0.001**	< **0.001**	0.083
Non-human rational			–	0.003	< **0.001**	**0.005**	< **0.042**	0.538
Human rational				–	0.416	0.960	< **0.001**	< **0.001**
Human non-rational					–	0.258	< **0.001**	< **0.001**
Animate						–	< **0.001**	< **0.001**
Concrete inanimate							–	< **0.001**

remaining 50%, except potential outliers which are indicated separately with circles (Daly *et al.* 1995). This demonstrates that there is a relationship between the positions in the availability hierarchy and number use.

On the one hand, we might have hoped for a correlation between the positions on the hierarchy and number, and clearly this is not found. This means that the hierarchy which accounts well for number availability across languages does not apply straightforwardly to number use, since Russian appears to be a counterexample. On the other hand, when we compare the medians of the proportion of plural forms for the different animacy categories of Smith-Stark, we see that each lexical category has its own median point (Figure 1). This strongly indicates that at a general level, the hierarchy position to which a lexeme belongs has an impact on the way it will distribute its forms. There is a dramatic difference between groups of nominals. Nouns denoting humans and other animates show the highest proportion of plural use, with concrete and abstract inanimates lower. Moreover, for all positions below non-human rationals the *p*-values are highly significant (Table 2 rightmost column). For the kin and non-human rational cate-

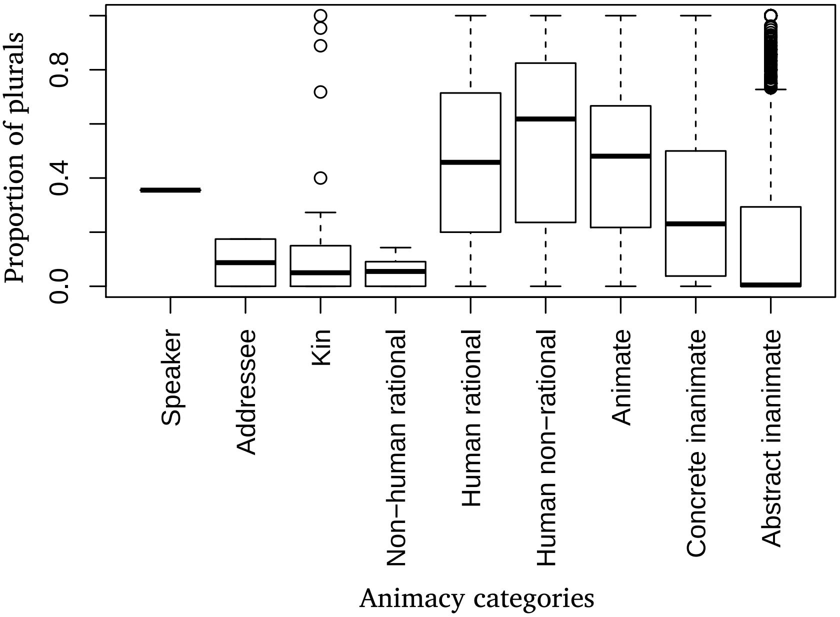

Figure 1: Box plot of proportion of plurals and animacy

gories there are plausible reasons why this might be so. One concerns standard use. As kin terms are often used for addressing individuals, it is reasonable to expect a high proportion of singular forms. Another contributing factor could be the uniqueness of the father and mother kin relations with respect to each individual. For non-human rationals (i.e., god, devil, angel) we expect a higher proportion of singular forms given that Russia's major religion is monotheistic. On the other hand, there is no obvious reason to assume that the pronouns for speech participants would differ in terms of number use.

Another possible explanation for the different structures of availability and use is based on the notion of individuation. When we compare number availability with number use, an interesting picture emerges. If the medians of the proportion of plurals are compared amongst the lexemes belonging to each slot in the hierarchy, as shown in Figure 1, we have a steep hill shape, peaking at the human non-rationals. In other words the left and right edges of the hierarchy have a smaller proportion of plurals, and the middle portion has a significantly higher proportion of plurals. An explanation for the steep hill shape may be based on individuation, running from most individuated (Speaker), to least individuated, to completely non-individuated items (abstract mass nouns). The small proportion of plurals at the bottom of the hierarchy is due to 'individual' plurals being largely unavailable, and only the (rarer) 'sort' and 'container' plurals being available. In this scenario the small proportion of plurals at the top segment of the hierarchy is due to the conceptual difficulty of pluralising highly individuated items. Describing a person using a kin term is individualising him/her further. Pluralising the same person would act to make him/her less individuated. This would explain the lack of plurals in this category.

In sum, the position of a lexeme on the hierarchy has a strong effect on number use. However, further co-variates come into play which account for an item's behaviour being different to that solely expected from its animacy position. We will discuss each of these co-variates, plural preference, pluralia tantum and irregularity effects in turn below.

Example	Example's animacy	Plural proportion	Plural proportion of example's animacy category (median)
roditel' 'parent'	Kin	95%	5%
bliznec 'twin'	Human rational	97%	45.5%
soavtor 'co-author'	Human rational	90%	45.5%
glaz 'eye'	Concrete inanimate	90%	23.1%
botinok 'boot'	Concrete inanimate	88%	23.1%
gramm 'gramme'	Abstract inanimate	81%	0.5%

Table 4: Nouns in the corpus locally unmarked for plural

3.2 *Plural preference*

Some items are naturally 'more plural' regardless of their lexical category. These can be viewed as locally unmarked for plural (Tiersma 1982), for instance items such as *glaz* 'eye' and *bliznec* 'twin' which would be expected to occur in the plural more frequently than the singular because singular contexts are unusual. Table 4 shows how the proportion of plurals for a locally unmarked item was found to be much greater than that expected from its animacy group.[5] Such nouns occur as outliers in our boxplots.

It might be asked why there is no similar section on singular preference. The basic answer is that for a noun to have singular preference is completely normal, as is evident from Table 2 (see column 'Mean plural proportion'), and from cross-linguistic data (see Corbett 2000, p. 281, for data on French, Latin, Sanskrit, Slovene and Upper Sorbian, as well as on Russian). In our count one third of the nouns (almost exactly) occur in the singular only. Note that this does not imply that they are singularia tantum; recall that for inclusion we require that the noun occurs five times or more. It is evident from the list that many nouns which occur five times only, all in the singular, are normal count nouns; they happen not to have occurred in the plural in the corpus.

[5] For further discussion of the semantics of number in Russian, see Ljaševskaja (2004) and references therein.

Table 5: Pluralia tantum in the corpus

Example	Example's animacy	Plural proportion	Plural proportion of example's animacy category (median)
rebjatiški 'kids'	Human non-rational	100%	61.8%
sani 'sledge(s)'	Concrete inanimate	100%	23.1%
brjuki 'trousers'	Concrete inanimate	100%	23.1%
xlopoty 'troubles'	Abstract inanimate	100%	0.5%
sutki '24 hours'	Abstract inanimate	100%	0.5%

3.3 *Pluralia tantum*

Some items lack a means of marking singular; in other words, for them singular is unavailable and they will always appear morphologically plural (even where there is a singular interpretation). Such pluralia tantum are given in Table 5. For example, the noun *sani* 'sledge' is morphologically marked for plural, but can have a singular and a plural reading.

Pluralia tantum are recognizable and are few in number in Russian. On the other hand, genuine singularia tantum are hard to identify; while many nouns normally occur in the singular, there are possibilities for recategorization: that is, they may be recategorized with unit reading or with instance reading (see Corbett 2000, pp 81–82, 84–87, for discussion). To illustrate the instance reading, we may take *mnogo raznyx vin* 'many different wines', where different types of wine are intended. The key point is that while such recategorizations are visible in the plural, the recategorization from mass to count gives a singular form too, hence *odno očen' xorošee vino* 'one very good wine'. This recategorized singular is not distinct from the normal singular.

3.4 *Irregularity effects*

There is a third important co-variate. In certain instances irregularity can affect the distribution of plurals. To appreciate this, it is important to distinguish absolute counting (the straightforward count of items in the corpus) from relative counting (the relation of forms within a lexeme; in our study this is plural versus singular). Irregularity in a lexeme is correlated with a high occurrence of plurals of that lexeme in the corpus.

Corbett *et al.* (2001) demonstrate for Russian that there is a relation between irregularity in noun lexemes and absolute plural anomaly, i.e., a high absolute number of plural forms in the corpus, and that there is a relation between non-prosodic irregularity (where irregularity is not confined to stress placement), and relative plural anomaly, i.e., a high proportion of plural forms compared to forms in the singular. This means that irregular Russian nouns in general have a high number of plural forms in the corpus. Prosodic irregularity means that there is also a high number of singular forms to match the plural ones (hence no relative plural anomaly), whereas nouns which display segmental irregularity have a higher proportion of plural forms in comparison with singular forms (hence high relative plural anomaly).

In sum, these three types of co-variate (plural preference, pluralia tantum, and irregularity effects) broadly account for the plural outliers in Figure 1.

4 CONCLUSIONS

Typology is typically concerned with the availability of a feature in a language. The special interest of our contribution lies in juxtaposing questions of availability with those of actual use. One hypothesis about the relationship between number use in one language (here Russian) and its relationship with the hierarchy of number availability is that there should be a correlation, a strictly linear relationship where those categories furthest left in the hierarchy show the greatest median plural proportion, with this proportion decreasing as we move rightward along the hierarchy. However, this hypothesis must be rejected. The reality is perhaps more interesting: we have good evidence that the middle part of the hierarchy shows the highest plural proportions of usage, with a consistent decrease in plural proportions as we move rightward from the human rationals to the abstract inanimates. We are in a position to say that this is significant. For the top end of the hierarchy there is less that can be said with certainty, given the lack of significance for certain of the higher positions. If anything our results point to the difference between the pronoun proportion of the hierarchy (where the results are not significant) and the nominal proportion (where the results are significant). Something that is worthy of further investigation is the question of why the human (rational and non-rational) part of the hierarchy has the highest proportions, compared to animates and concrete inanimates. Further investigation would enable us to decide between two different theories about the way the hierarchy partitions the semantics of plural in use. In one theory, associative readings, 'normal' readings and recategorization effects partition the hierarchy, and the observation of high plural occurrence in the middle of the hierarchy is evidence for the high frequency of 'normal' readings associated with this part of the hierarchy. An alternative theory is that plural usage in the middle of the hierarchy is a reflection of the fact that it can have multiple plural semantics available to it (rather than just the 'normal' readings), and these multiple possibilities are reflected in greater use. While the first of these theories is the more plausible, we have no evidence yet to decide between them. Our research has therefore suggested a new programme of future research to investigate this matter in greater depth.

Our examination of the category of number in a language where nouns typically mark number has shown that the typology proposed by Smith-Stark for number availability has a partial analogue for number use. In other words, we have shown that answers to questions about availability can be reflected in use.

REFERENCES

W.J. CONOVER (1971), *Practical Nonparametric Statistics*, John Wiley & Sons, New York.

Greville G. CORBETT (2000), *Number*, Cambridge University Press, Cambridge.

Greville G. CORBETT, Andrew HIPPISLEY, Dunstan BROWN, and Paul MARRIOTT (2001), Frequency, regularity and the paradigm: a perspective from Russian on a complex relation, in Joan BYBEE and Paul HOPPER, editors, *Frequency and the Emergence of Linguistic Structure*, pp. 201–226, John Benjamins, Amsterdam.

Fergus DALY, David HAND, Chris JONES, Daniel LUNN, and Kevin MCCONWAY (1995), *Elements of statistics*, Addison-Wesley, London.

Bradley EFRON and Robert J. TIBSHIRANI (1993), *An introduction to the bootstrap*, Chapman and Hall, London.

Elena A. GRIŠINA and Vladimir A. PLUNGIAN (2005), Perspektivy razvitija Natsional’nogo korpusa russkogo jazyka [Prospects for developing a national corpus of Russian], *Indrik*, http://ruscorpora.ru/sbornik2005/19grishina.pdf.

Ol’ga N. LJAŠEVSKAJA (2004), *Semantika russkogo čisla [The semantics of Russian number]*, Jazyki Slavjanskoj Kul’tury, Moscow.

Lennart LÖNNGREN (1993), *Častotnyj slovar’ sovremennogo russkogo jazyka [A frequency dictionary of contemporary Russian] (Acta Universitatis Upsaliensis, Studia Slavica Upsaliensis 33)*, University of Uppsala, Uppsala.

Ingrid MAIER (1994), Review of Lennart Lönngren (ed.), Častotnyj slovar’ sovremennogo russkogo jazyka, *Rusistika Segodnja*, 1:130–136.

T. Cedric SMITH-STARK (1974), The plurality split, in Michael W. LA GALY, Robert A. FOX, and Anthony BRUCK, editors, *Papers from the Tenth Regional Meeting, Chicago Linguistic Society*, pp. 657–671, Chicago: Chicago Linguistic Society.

Peter M. TIERSMA (1982), Local and general markedness, *Language*, 58:832–849.