Word Order in Adjective-Noun Pairs
Inside and Outside the Prepositional Phrase
A contribution to the 15th International Congress of Slavists (Minsk, 2013)

Philip R. Minlos
Institute for Slavic Studies of the Russian Academy of Sciences, Moscow

Abstract
This paper discusses a statistical correlation between possessive prenominal placement and the presence of a preposition. The data from mediaeval Russian, Czech, and Croatian are validated against standard statistical measures (chi-square ($X^2$) test and phi ($\phi$) coefficient). Different explanations for the correlation are proposed; the most natural and simple one links the syntactic feature with the phonetic chunking of preposition and adjacent possessive, strengthened by their frequent co-occurrence.

Keywords
syntax, statistical analysis, noun phrase, word order
В статье обсуждается статистическая корреляция между препозицией притяжательных местоимений и наличием предлога. Данные средневековых русских, чешских и хорватских памятников оцениваются по стандартным статистическим процедурам (критерий $X^2$ и ф-коэффициент). В статье предложено несколько различных объяснений для этой корреляции; самое естественное и простое объяснение связывает эту синтаксическую особенность с фонетической: предлог со смежным притяжательным местоимением образуют единое фонетическое слово, причем это единство усиливается частым повторением этой пары в текстах.

Ю. Синтаксис, статистический анализ, именная группа, порядок слов

0. Introductory notes. The present contribution discusses a syntactic feature found in mediaeval Slavic languages: the presence of a preposition mostly triggers (or, to put it another way, statistically correlates with) the prenominal placement of a single adjective attribute (as in Old Russian в его соборе 'in (his) own town' in the Smolensk treaty of 1229), whereas if no preposition is present, the adjective is free to be postposed (e.g., дом своего in the same treaty). Old Russian data laid out here have already been presented in a series of articles, most notably [MINLOS 2008, MINLOS 2011, MINLOS 2012]; here, some Old Russian data are repeated, coupled with standard statistical analysis (chi-square test and phi coefficient). The new data from mediaeval Czech and Croatian texts introduced in the present article provide evidence that the same correlation is in effect in other Slavic languages. Furthermore, after broadening the empirical base it becomes more evident that it is the correlation that should be accounted for, not the specific manifestations of the correlation (e.g., “in prepositional phrases, adjectives follow the noun very rarely”). The discussion section will offer a clearer understanding of the problem as a whole. A variety of interpretations are offered for this correlation. My preferred explanation states that the preposition and the following adjective/possessive, although unrelated structurally, merge into a unit, as if some force of gravity binds them together. What is the reason for this? First, they clearly form a phonetic unit, and second, the phonetic chunking is most likely strengthened by frequent co-occurrence.1

1 Statistical analysis of Old Russian data. The correlation between adjective placement and prepositional phrase was observed by Dean Worth in [WORTH 1985], where the analyzed data set consisted of phrases with the adjective новгородский 'Novgorodian' in a selection of Novgorodian parchment docu-

I am much obliged to those who read the versions of the paper and commented on it: Peter Arkadjev, Claudia Jensen, Viktor Zhivov, and especially the anonymous reviewer.
ments. His findings may be summarized as follows: Suppose we have picked up a set of noun-adjective pairs (clear results are obtained from “simple” noun phrases containing no elements other than these two). Presumably, the sample is defined on some independent and solid grounds, that is, it consists of some structurally similar pairs\(^2\) from a single text or from a group of similar texts. There are two independent two-value parameters which may be applied to the pairs: 1) the relative ordering of noun and adjective: [adjective noun] and [noun adjective] (here abbreviated as AN ~ NA), and 2) whether the noun phrase is governed (or accompanied, depending on the model) by a preposition (hereafter –prep ~ +prep). Thus, the frequency distribution of two variables forms a 2×2 contingency table such as Table 1:

![Table 1. Contingency table for two variables (+-prep and AN-NA order variation)](image)

The four types may be illustrated by four examples with the collocation novgorod'skaja volostь ‘Novgorodian region’ in Worth’s material:

- AN, –prep Acc. Sg. novgorod'skaja volostь (№ 8);
- NA, –prep Acc. Pl. volosti novgorod'skij (№ 18);
- AN, +prep Loc. Sg. po novgorod'skoj volosti (№ 63);
- NA, +prep Loc. Pl. po volostemь po novgorod'kimь (№ 23).

(A careful reader might notice the differences in the examples: postnominal placement with plural nouns, prenominal placement with singular nouns; this is not a mere chance, as discussed in [MINLOS 2011: 55].)

The main result presented in [WORTH 1985] was that the number N\(_{NA, +prep}\) is the smallest one. The exact statements made in that article were more complicated, as the author introduced an additional parameter, that is, animacy. This additional parameter is rather controversial; the relevant discussion can be found in [MINLOS 2010b] and [MINLOS 2011].

As I carried on the work Worth had begun, I focused primarily on a single group of frequent phrases, that is, nouns with agreeing possessives and with no extra modifiers (“simple NPs”), such as moi smol'nyane ‘my Smolensk people’ [MINLOS 2011, MINLOS 2012]. The data were mostly taken from Old Russian texts (birchbark letters, chronicles, legal texts, testaments, treaties, and hagiographic texts).

\(^2\) That is, pairs including only a specific group of adjectives or even also specific pairs (e.g., [MINLOS 2012] deals at length with kn'az' velikij ‘grand duke’).
The present article repeats some tables from [MINLOS 2012], providing a basic statistical analysis for them. Table 2 summarizes the data from the Smolensk agreements\(^3\). Table 3 summarizes the data from the Old Russian legal code *Pravda ruskaja* according to the Troitsky manuscript (15th century).

**Table 2. Placement of agreeing possessive in Smolensk agreements (13–14th centuries)**

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−prep</td>
<td>12</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>+prep</td>
<td>17</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Totals</td>
<td>29</td>
<td>8</td>
<td>37</td>
</tr>
</tbody>
</table>

**Table 3. Placement of agreeing possessives in Pravda ruskaja with totals**

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>−prep</td>
<td>16</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>+prep</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Totals</td>
<td>21</td>
<td>5</td>
<td>26</td>
</tr>
</tbody>
</table>

The most visible feature of Tables 2 and 3 is the almost complete absence of examples with postposed possessives in prepositional phrases (the relevant cells are shaded gray). Though the numbers, informally speaking, do not seem to be convincing (see below for a more formal perspective), this tiny data set clearly illustrates an interesting distribution pattern. What is evident from this table is that the overall number of prepositional phrases is smaller than that of phrases without prepositions ($N_{−\text{prep}} > N_{+\text{prep}}$, and that is typical at least for Slavic) and that noun phrases with preposed possessive outnumber those with postposed possessive ($N_{AN} > N_{NA}$, and that is typical for Slavic secular legal texts). So, intuitively, we do not expect the $N_{NA, +\text{prep}}$ to be a big number.

As some Slavists may not have sufficient background in statistics, the following section is devoted to recapitulating the basic statistical notions of *null hypothesis*, *expected values*, *chi-square ($X^2$) test*, and *phi ($\phi$) coefficient*.

\(^3\) Agreements of Smolensk princes with their Western counterparts are analyzed on the basis of the [SUMNIKOVA, LOPATIN 1963] publication, and include: 1) the agreement of an unknown Smolensk prince with Riga and Gotland (1220s) (pp. 11–13); and 2) the agreement of Smolensk with Riga and Gotland (with six texts grouped in two versions and dated between the 13th and the 14th centuries) (pp. 18–52). The Smolensk-German agreements are a collection of nearly identical documents; the numbers presented in the paper refer to the union (set-theoretical sum) of the texts. This means that each fragment that recurs in identical texts is counted only once.
The essence of our study lies in the idea that one categorical variable (adjective placement) is determined by another categorical variable (presence / absence of preposition). The only way of testing this hypothesis (technically, an alternative hypothesis) is through constructing and testing the null hypothesis. In our case, the null hypothesis predicts that the variables (adjective placement and presence of preposition) are independent.

The data from Table 2 will serve as a model example. The respective sample numbers 37 outcomes. We measure the probability of the specific values (for example, NA word order) with the formula:

\[
\text{probability of the event} = \frac{\text{number of outcomes giving this event}}{\text{total number of outcomes}}
\]

Thus, the probability of AN word order is estimated by dividing the total of the respective column by the overall total:

\[
p(AN) = \frac{29}{37} \approx 0.78
\]

Each cell of the table relates to a joint event (for example, NA word order + prep). Assuming null hypothesis (= that presence of the preposition has no effect), joint probability is obtained simply by multiplying two respective probabilities:

\[
p(NA, +prep) = \frac{8}{37} \times \frac{18}{37}
\]

The expected value of the event is its probability multiplied by the total number of outcomes:

\[
p(NA, +prep) = \frac{8}{37} \times \frac{18}{37} \times 37 = 3.89
\]

So, the actual value 1 is somewhat less than the value expected according to the null hypothesis, which indicates that a deviation from the null hypothesis does exist. Table 4 presents the expected values.

Table 4. The expected values for Table 2

<table>
<thead>
<tr>
<th></th>
<th>(N_{AN})</th>
<th>(N_{NA})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_{-prep})</td>
<td>14.89</td>
<td>4.11</td>
</tr>
<tr>
<td>(N_{+prep})</td>
<td>14.11</td>
<td>3.89</td>
</tr>
</tbody>
</table>

The amount of deviation, or error, is obtained by subtracting the expected values (E) from the observed ones (O). This amount of deviation for the cells in our table takes two values: 2.89 and –2.89. The next step is to square the difference, which results (with approximation) in \((O-E)^2 = 8.35\). Then, for each cell the squared difference is divided by the expected cell frequency, which results in Table 5.

Table 5. \((O-E)^2/E\) values, based on Tables 2 and 4

<table>
<thead>
<tr>
<th></th>
<th>(N_{AN})</th>
<th>(N_{NA})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_{-prep})</td>
<td>0.56</td>
<td>2.03</td>
</tr>
<tr>
<td>(N_{+prep})</td>
<td>0.59</td>
<td>2.15</td>
</tr>
</tbody>
</table>
The sum of the four cells is $0.5 + 0.59 + 2.03 + 2.15 = 5.33$, which is the **chi-square value**.

First, the chi-square value is used in **Pearson’s test** to calculate the **p-value**, which is the probability of obtaining the observed results assuming that the null hypothesis is true. The p-value is calculated from the chi-square and the **degrees of freedom** in our case, 1); usually special software such as R or look-up tables are used to determine the value. In our case, the p-value is 0.0104 (when “one-tailed” calculation is used, which means that we conjecture the direction of the correlation before the test). The null hypothesis is rejected if the p-value is less than the **significance level** 0.05 or 0.01 (the significance level 0.05 means that we are willing to take a 5% chance of making a mistake by rejecting the null hypothesis). As the p-value is indeed less than the significance level 0.05 (but not 0.01), the association between the variables in Table 2 (the Smolensk agreements) is considered to be statistically significant (but not extremely significant).

Second, the chi-square value is used to calculate the phi coefficient $\varphi$, which is used to compare the measure of association of two binary variables. By definition,

$$\varphi^2 = \frac{X^2}{N} \quad (5)$$

From this, we obtain $\varphi = 0.38$ for Table 2 (the Smolensk agreements). As a rule of thumb, a phi coefficient of more than 0.30 is usually interpreted as indicating significant evidence; a coefficient of less than 0.30 is usually interpreted as evidence that the association is too weak. It confirms that the data from the Smolensk agreements are significant.

Now we are in a position to revisit the Old Russian data from [MINLOS 2010A] and [MINLOS 2012] (the exact description of the samples can be found in the cited articles) and estimate their statistical reliability.

**Table 6. Statistical estimates for the placement of agreeing possessives in Old Russian sources**

<table>
<thead>
<tr>
<th></th>
<th>chi square value</th>
<th>p value</th>
<th>Total number of examples</th>
<th>phi coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smolensk agreements</td>
<td>5.33</td>
<td>0.01</td>
<td>37</td>
<td>0.38</td>
</tr>
<tr>
<td>Pravda ruskaja</td>
<td>1.47</td>
<td>0.11</td>
<td>26</td>
<td>0.24</td>
</tr>
<tr>
<td>birchbark letters</td>
<td>13.68</td>
<td>0.001</td>
<td>121</td>
<td>0.34</td>
</tr>
<tr>
<td>Life of St. Andrey the Fool</td>
<td>6.94</td>
<td>0.004</td>
<td>100</td>
<td>0.26</td>
</tr>
<tr>
<td>First Novgorod Chronicle</td>
<td>13.07</td>
<td>0.002</td>
<td>161</td>
<td>0.28</td>
</tr>
<tr>
<td>Kiev Chronicle (a sample)</td>
<td>3.9</td>
<td>0.02</td>
<td>211</td>
<td>0.14</td>
</tr>
<tr>
<td>Pskov Court Charter</td>
<td>2.49</td>
<td>0.06</td>
<td>61</td>
<td>0.20</td>
</tr>
<tr>
<td>Novgorod parchment documents</td>
<td>23.75</td>
<td>&lt; 0.0001</td>
<td>267</td>
<td>0.30</td>
</tr>
</tbody>
</table>

---

4 For our current purposes, the very informal definition from [GRIESE 2009: 47] suffices: “Degrees of freedom are the way in which sample sizes and the amount of information you squeeze out of a sample are integrated into the significance test.” The more parameters we take into account, the higher the degrees of freedom.
Pearson's chi-square test depends heavily on the size of the data set; for that reason the First Novgorod Chronicle and the Life of St. Andrey the Fool show rather good p values, whereas the phi coefficient is below the 0.3 threshold. The Kiev Chronicle, Pravda ruskaja, and Pskov Court Charter show rather poor results for both measures. Thus, two Old Russian texts providing clear evidence that the possessive placement depends on the presence of the preposition are the corpus of birchbark letters and the Smolensk agreements. The Novgorod parchment documents are also an important source.

2. Data from outside Old Russian. From the descriptive perspective, additional data are always welcomed; it is important that the correlation can now be extrapolated from Old Russian to Slavic languages in general. But what is more, the data from Czech and Croatian texts presented here are crucial for the choice of interpretation. In Old Russian texts, the correlation mostly surfaces as a small number of NA phrases after a preposition. In Tables 7, 9, and 10, the correlation is most visible in the small numbers of AN phrases without preposition.

Here we present the data from some mediaeval Czech and Croatian texts in tabular form. The Chronicle of the Priest of Duklja (we used the edition provided in [KUKULEVIČ 1851]) dates from the 16th century, but the original text may well be much older. The notarial records signed by the notary public Mikula Krstinić were written in glagolitic script on the island of Lošinj at the end of the 16th century and published in [KOŠUTA 1988] (texts 1–170). The chronicle of the clerk Bartoš [ERBEN 1851] describes events that took place in Prague at the beginning of the 16th century (and of which the author was a witness), and was presumably composed in the first half of the century. The present paper uses the data from chapters 1–20 of Book 1. The huge historical work Historie církevní (Church History) is actually by no means restricted to ecclesiastical events. It was written by Pavel Skála ze Žhoře in the first half of the 17th century. The examined sample is published in [JANÁČEK 1984: 25–48]. The archive of Olřich (Ulrich) from Rosenberg (Rosenberg is now a district of the city of Brno) dates from the first half of the 15th century. The archive consists of the correspondence of Olřich with various members of the nobility, including Elizabeth (Alžbeta) of Luxemburg, queen of Hungary and Bohemia. The examined sample is limited to the text numbers 1–42 (from the period 1420–1445) from the edition [PALACKÝ 1844]. Two frequently recurring phrases were removed from the data as they are used invariably: služba naše and TM (standing for Tvá Milost).
Table 7. Agreeing possessive placement in the *Chronicle of the Priest of Duklja* (Croatia)

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>− prep</td>
<td>3</td>
<td>52</td>
</tr>
<tr>
<td>+ prep</td>
<td>16</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 8. Agreeing possessive placement in the records of Mikula Krstinić (Croatia)

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>− prep</td>
<td>63</td>
<td>53</td>
</tr>
<tr>
<td>+ prep</td>
<td>69</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 9. Agreeing possessive placement in a sample from the *Chronicle of the clerk Bartoš* (Bohemia)

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>− prep</td>
<td>12</td>
<td>71</td>
</tr>
<tr>
<td>+ prep</td>
<td>42</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 10. Agreeing possessive pronouns in a sample from *Historie církevní* by Pavel Skála ze Zhoře (17th century)

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>− prep</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td>+ prep</td>
<td>15</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 11. Agreeing possessive pronouns in a sample from the archive of Oľich (Ulrich) from Rosenberg (Bohemia)

<table>
<thead>
<tr>
<th></th>
<th>AN</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>− prep</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td>+ prep</td>
<td>40</td>
<td>7</td>
</tr>
</tbody>
</table>
The quantitative distribution in the historical works (Tables 7, 9, and 10) differs from that in the legal documents and official communications: in the chronicles, the fraction $N_{AN}/N$ is rather small. As already noted above, the data with a small fraction of $N_{AN}/N$ is crucial for the general discussion.

The statistical evaluation of the data is presented in Table 12.

Table 12. Statistical estimates for the placement of agreeing possessives in mediaeval Croatian and Czech texts

<table>
<thead>
<tr>
<th></th>
<th>chi square value</th>
<th>p value</th>
<th>Total number of examples</th>
<th>phi coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Chronicle of the Priest of Duklja</td>
<td>13.66</td>
<td>0.0001</td>
<td>102</td>
<td>0.37</td>
</tr>
<tr>
<td>The records of Mikula Krstinić</td>
<td>6.9</td>
<td>0.004</td>
<td>212</td>
<td>0.18</td>
</tr>
<tr>
<td>The chronicle of the clerk Bartoš</td>
<td>31.84</td>
<td>&lt;0.001</td>
<td>156</td>
<td>0.45</td>
</tr>
<tr>
<td>Church History</td>
<td>13.47</td>
<td>0.0001</td>
<td>80</td>
<td>0.41</td>
</tr>
<tr>
<td>The archive of Olřich</td>
<td>24.32</td>
<td>&lt;0.001</td>
<td>135</td>
<td>0.42</td>
</tr>
</tbody>
</table>

The most significant correlation between two variables is observed in Table 9 (in the chronicle of the clerk Bartoš).

3. Discussion. Two different ways of interpreting the distribution are considered:

(a) in the prepositional phrase, the adjective is preferably placed before the noun;

(b) in the noun phrases void of preposition, the noun is preferably placed before the adjective.

Theoretically, proposals (a) and (b) do not contradict each other; both forces may be in effect. The discussion of the choice of interpretation follows the presentation of these proposals.

Proposal (a) comes in three different varieties, a1–a3:

(a1) From a structural point of view, at least some postnominal placements are induced by sentence-level rules. The clearest case of such a rule is the NA order in sentence-initial thematic position (some considerations regarding this placement can be found in [Bonnot 2008], [Bonnot 2009], [Minlos 2012]). Noun phrases without preposition are free to be conditioned by some other rules, as they are not dominated by a prepositional phrase. Prepositional phrases may be simply inaccessible for such rules. The absence of a preposition, according to this version, does not force postnominal placement, but allows it to take place (for example, due to some information structure rules). To
summarize, the preposition blocks the external rules and retains the AN order (which is viewed as the basic order, in this case).

(a2) Perhaps the preposition forms a unit of some kind with the neighboring adjective, irrespective of the fact that they do not form a conventional syntactic constituent. Indeed, the preposition does form a phonological unit with the leftmost wordform of the noun phrase; in the case of the AN phrase, it is the adjective that occupies this position. The rest of the phrase may be separated from the preposition-adjective pair by a phrasal clitic or by some other external material, but the preposition cannot be left alone. The relevant examples are (1) and (2) below.

(1) Ukrainian *Ujake vin pojide misto*? [FÉRY, PASLAWSKA, FANSELOW 2007: 5]
   in which he will go town
   'To which town will he go?'

(2) Serbian *U koji je stigao grad?* [PENN 1999: 158]
   in which *cl*-3*sg* arrived city
   'In which city has he arrived?'

An additional hypothetical reason is more subtle: as many noun modifiers (first of all possessives, demonstratives, and indefinites) are much more frequent than common nouns, the specific preposition-modifier pairs (such as Old Russian [*na, svoi*]) are also more frequent than the specific preposition-noun pairs (such as [*na, podvorje*]). While mainstream linguistics is more concerned with structural dependencies, much research has been done in the frequently co-occurring wordforms which do not correspond to any structural constituent (called *chunks* in works by J. Bybee [BYBEE 2010]), and *n-grams* in natural language processing [JURAFSKY, MARTIN 2008]). Such sequences seem to matter a lot for speech production and parsing. At the extreme, such frequent pairs clearly merge into a new unit, such as the Italian preposition-article units *del, al*, etc.

(a3) The explanation proposed by D. Worth should be classified as a separate point of view. To quote: “inanimates are forced into the order pAN by the obligatory syntactic coherence of the prepositional phrase, the opening of which is signaled by prepositional phrase and the closure of which is signaled by the governed noun (note the tendency to mark the continuation of the phrase by prepositional repetition in the rare cases where inanimates occur in the order pNA)” [WORTH 1985: 543]. His interpretation (“preposition-adjective-noun” as a settled sequence of words) lies somewhere in between our (a1) and (a2).

The alternative idea (b) states that in the noun phrases without preposition, the noun is preferably placed phrase-initially, that is, before the adjective. Essentially, this idea links the Slavic NP-syntax with the well-known facts from German grammar: in dependent clauses governed by a complementizer

---

5 The reasons Worth limited the interpretation to inanimates does not concern us here; see [MINLOS 2011] for details.
or a subordinating conjunction, the finite verb form is placed at the end (\textit{Maria glaubt, dass Hans den Mann sah}), otherwise, both in main and in dependent clauses, the finite verb form occupies the second position in the clause (\textit{Hans sah den Mann}; \textit{Marie glaubt, Hans sah den Mann}). These facts received some attention from specialists in transformational syntax (the second position was claimed to instantiate \textquotedblleft head movement\textquotedblright). Another approach is a constraint requiring that a prepositional / noun phrase begin with a head — either with a preposition or a noun, accordingly. I proceed from the assumptions stated in the research of J. Hawkins, for example in [HAWKINS 1994] and [HAWKINS 2004]: as the human processor parses the linear sequence of forms and properties, it works more efficiently as more properties are assigned earlier; the linguistic structures are evolving to meet the processing requirements. In our case, the heads are sufficient to detect the corresponding phrase (the preposition 	extit{constructs} the prepositional phrase, the noun 	extit{constructs} the noun phrase). Adjectives, however, may occur outside NP boundaries (first of all, in predicative constructions), and are therefore not sufficient to detect the NP.

Which of the proposed interpretations should we prefer? Proposal (b) does not fit very well with the data in which AN groups, even without preposition, are the predominant pattern (such as the data from \textit{Pravda ruskaja}, Table 3). Conversely, the proposals of the (a) group do not clearly explain the picture in the language varieties where NA groups, even with preposition, abound (such as the sample from the chronicle of the clerk Bartoš, table 9). In the most general terms, (a) and (b) are mirror images of each other and the difficulties accompanying the approaches are the same.

But upon closer scrutiny a difference does appear. Proposal (a1) says that NA order results from sentence-level rules, but when nested within the prepositional phrase, the AN phrase is inaccessible to these external rules. This proposal does presuppose that the AN order is the unmarked / basic one, which is problematic for many language varieties (for example, Tables 7, 9, and 10), if we judge from a quantitative point of view. But proposal (a2)—a preposition forms a unit with an adjective — does not rely on any vague presuppositions, as they clearly do form a phonetic unit.

Thus, the general conclusion is that the \textquotedblleft chunk\textquotedblright hypothesis gets the best score.


In this paper, we tacitly assumed that it is word order that can depend on the preposition, not the other way round. That is, in designing her utterance, the speaker chooses the construction first and then chooses the word order in the NPs in question that best fits the construction. But perhaps the speaker’s choice can also proceed the other way round. Why can a (preferable) position...
of the attribute affect the choice of the construction (that is, the presence or absence of the preposition)? Such dependence (word order → construction) obviously contradicts both the linear realization of speech (left to right) and the established linguistic levels. But such dependence fits well with the view that speech to some extent consists of surface sequences (chunks or n-grams, as discussed above).

The technical focus here on a single formal parameter determining word order in mediaeval Slavic noun phrases does not presuppose that this is the mainstream line of further research. On the contrary, I am confident that word order is determined by a host of factors, among which are clause-level constructions, semantics of the noun-adjective pair, and last but not least, grammatical number of the noun (some considerations on these parameters can be found in [MINLOS 2011] and [MINLOS 2012]). We cannot even completely rule out the possibility that the impact of the preposition on word order is only an epiphenomenon. But the goal of this technical report is to point to the texts which exhibit the correlation relatively clearly (luckily, there are huge texts of this kind, such as the chronicle of the clerk Bartoš) and thus can shed more light on the nature of the parameter.

References

Bonnot 2008

Bonnot 2009

Bybee 2010

Dryer 1992

Erbén 1851
Erbén K. J. (publ.), Bartošova Kronika pražská, od léta páně 1524 až do konce leta 1530, Praha, 1851.

Fény, Paslawska, Fanselow 2007

Greenberg 1963
Gries 2009

Hawkins 1994

Hawkins 2004

Janáček 1984

Jurafsky, Martin 2008

Košuta 1988

Kukulević 1851

Hawkins 1994

Hawkins 2004

Minlos 2008

Minlos 2010a

Minlos 2010b

Minlos 2011

Minlos 2012

Noha 1971

Palacký 1844
Palacký Fr. (publ.), “Rozličné dopisy pana Olřich z Rosenberka od r. 1420 do 1453”, *Archiv český čili staré písemné památky české a moravské*, 3/1, 1844, 3–64.

Penn 1999
Word order in adjective-noun pairs inside and outside the prepositional phrase

**Starčević 1852**


**Štefanić 1969**

Štefanić V. (publ.), *Hrvatska književnost srednjega vijeka*, Zagreb, 1969.

**Sumnikova, Lopatin 1963**


**Večerka 1989**


**Večerka 2006**


**Worth 1985**


**Zaliznjak 2004**