
О ВАРИАТИВНОСТИ НЕКОТОРЫХ БУКВЕННЫХ ЧАСТОТ В СУПРАСЛСКОМ СБОРНИКЕ

Й. Табов, Св. Христова

Абстракт. В настоящей работе предлагаем начальные шаги исследования вариативности частот некоторых букв в текстах разной длины из Супраслского сборника. Получены интервалы, в которых варьируют частоты букв М, Н, Е и Ъ в текстах длиной одной до пяти страниц сборника.

Keywords: буквенные частоты, старославянские тексты, Супраслский сборник

Введение

Среди важнейших задач при изучении болгарского письменного наследия можно отметить атрибуцию и датирование старых болгарских текстов — определение автора текста, школы, к которой его причисляют, и время его создания. Подобной проблемой является установление однородности текста (наличие вставок другого автора или переписчика). Эту традиционную для палеографов задачу не всегда можно успешно решить только их усилиями — сравнением специфических характеристик текста. Желательно найти формальный, количественно-статистический подход, основанный на анализе буквенных частот, который давал бы дополнительные доводы, дающие предпочтение одной палеографической гипотезы перед остальными.

Для создания такого подхода нужны масштабные исследования старых болгарских текстов, одним из важнейших среди которых является Супраслский сборник. Он подходит для экспериментов с буквенными частотами еще и потому, что результаты от них можно сравнивать с уже имеющимися многочисленными палеографическими анализами.

В настоящей работе предлагаем начальные шаги таких исследований: сравнения частот некоторых букв в текстах разной длины из Супраслского сборника по его факсимильному изданию [Супрасълски сборник, 1983].

1 О вариативности буквы М

Диаграмма на **рис. 1** показывает наглядно распределение частот буквы М для 37 страниц рукописи Супраслского сборника: по абсциссе частоты в процентах, разделенные на интервалы (от 1% до 1,5 %, от 1,5 % до 2 % и т.д.), а по ординате – число страниц, на которых частоты на буквы М попадают в соответствующий интервал, отмеченный на абсциссе.

Диаграмма на **рис. 2** показывает наглядно распределение частот буквы М для 36 пар страниц (36 текстов, занимающих 2 страницы) в рукописи Супраслского сборника: по абсциссе частоты в процентах, разделенные на интервалы (от 1% до 1,5 %, от 1,5 % до 2 % и т.д.), а по ординате – число страниц, для которых частоты буквы М попадают в соответствующий интервал, отмеченный на абсциссе.

Аналогичная диаграмма для 35 троек страниц представлена на **рис. 3**.

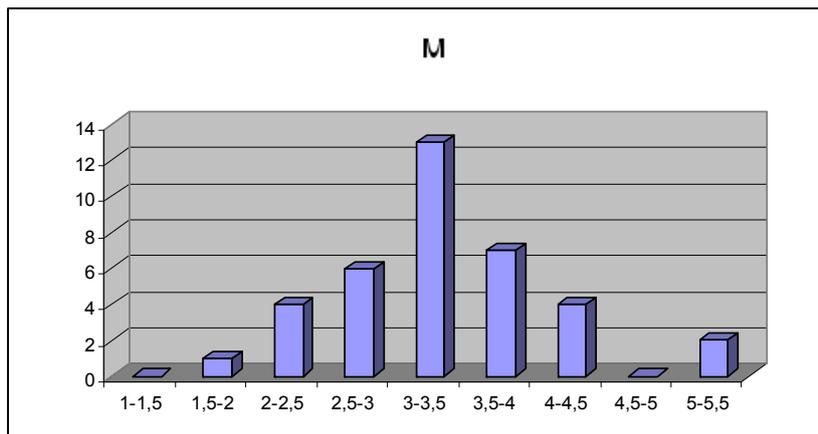


Рис. 1. Распределение частот буквы М для 37 страниц рукописи Супраслского сборника: по абсциссе частоты в процентах, разделенные на интервалы (от 1% до 1,5 %, от 1,5 % до 2 % и т.д.), а по ординате – число страниц, на которых частоты на буквы М попадают в соответствующий интервал, отмеченный на абсциссе.

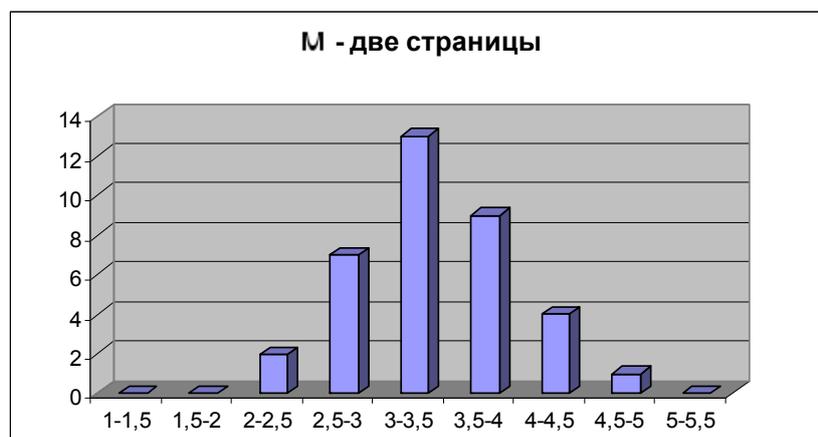


Рис. 2. Распределение частот буквы М для 36 пар страниц (36 текстов, занимающих 2 страницы) в рукописи Супраслского сборника: по абсциссе частоты в процентах, разделенные на интервалы (от 1% до 1,5 %, от 1,5 % до 2 % и т.д.), а по ординате – число пар страниц, для которых частоты буквы М попадают в соответствующий интервал, отмеченный на абсциссе.

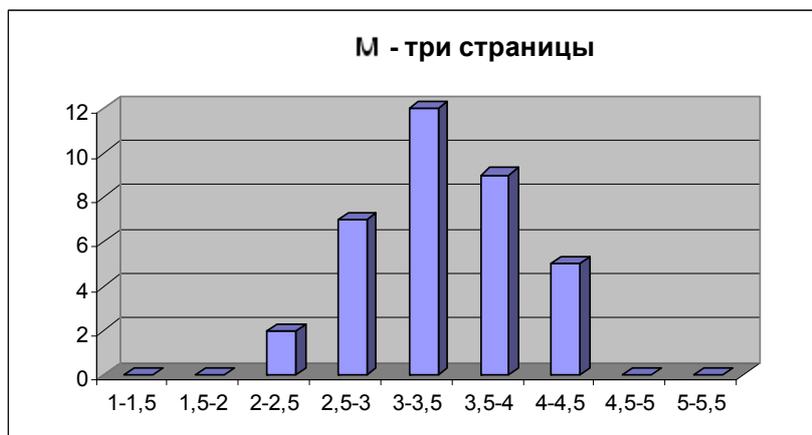


Рис. 3. Распределение частот буквы М для 35 троек страниц (35 текстов, занимающих 3 страницы) в рукописи Супраслского сборника: по абсциссе частоты в процентах, разделенные на интервалы (от 1% до 1,5 %, от 1,5 % до 2 % и т.д.), а по ординате – число троек страниц, для которых частоты буквы М попадают в соответствующий интервал, отмеченный на абсциссе.

В охваченных настоящим исследованием 37 страниц Супраслского сборника частота буквы М немного выше средней: около 3,343 %. Как видно из диаграммы на **рис. 1**, для этих страниц она изменяется от 1,5 % до 5,5 %, т.е. – в процентах – в интервале (1,5; 5,5). Для пар и троек страниц – судя по диаграммам на **рис. 2** и **рис. 3** – этот интервал сужается соответственно до (2 ; 5) и (2 ; 4,5).

Эти результаты дают возможность сравнить вариативность частот при уменьшении длины текстов: с троек к парам страниц и с троек к отдельным страницам (**рис. 4**)

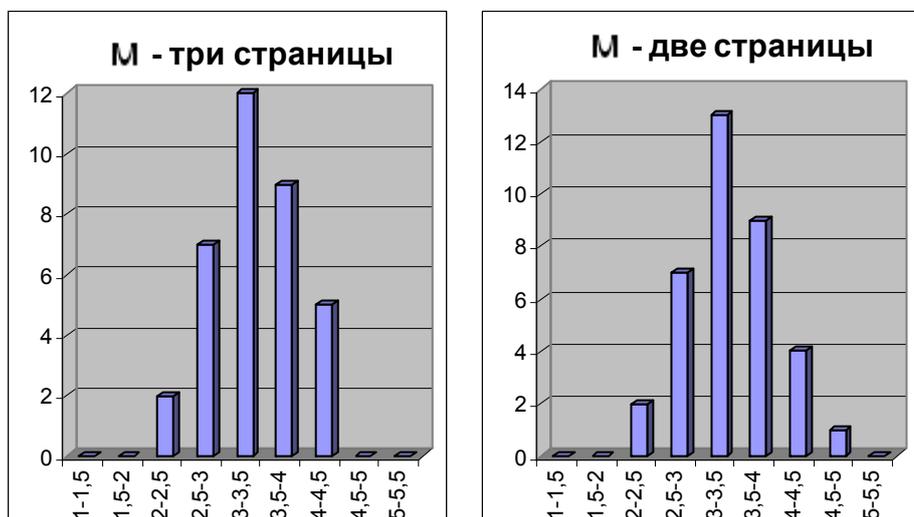


Рис. 4. Сравнение вариативности частот буквы М при уменьшении длины текстов: с троек (слева) к парам (справа) страниц в рукописи Супраслского сборника. Частоты М варьируют соответственно в интервалах (2 ; 4,5) и (2 ; 5).

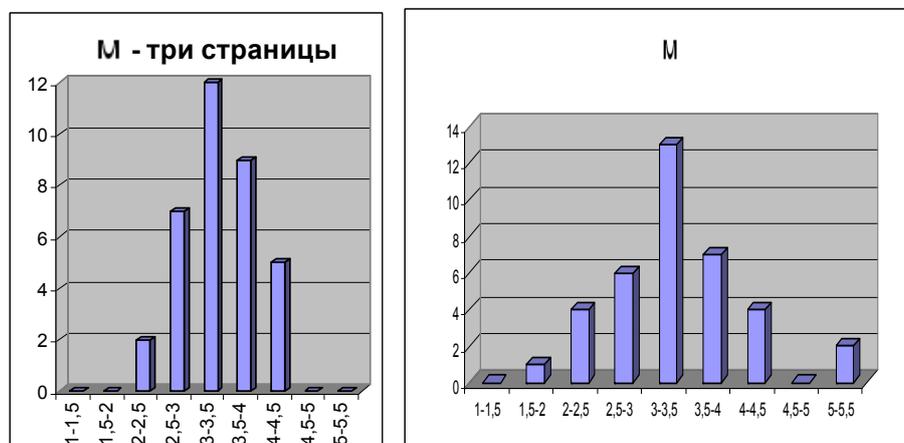


Рис. 5. Сравнение вариативности частот буквы М при уменьшении длины текстов: с троек (слева) к отдельным страницам (справа) в рукописи Супраслского сборника. Частоты М варьируют соответственно в интервалах (2 ; 4,5) и (2 ; 5).

Диаграммы на **рис. 4** и **рис. 5** иллюстрируют расширение интервала вариативности частот буквы М при уменьшении объема соответствующего текста – ожидаемое изменение. В процентном соотношении оно означает расширение интервала соответственно на 20 % и 60 %.

2 О вариативности буквы Н

В охваченных нашим исследованием 37 страниц Супраслского сборника частота буквы Н большая, выше средней: около 9,052 %. Как видно из диаграммы на **рис. 6**, для этих страниц она изменяется от

5,7 % до 12 %, т.е. – в процентах – в интервале (5,7 ; 12). Для троек и пятерок страниц – судя по диаграммам на **рис. 7** и **рис. 8** - этот интервал сужается соответственно до (7,8 ; 11,3) и (7,8 ; 10,8).

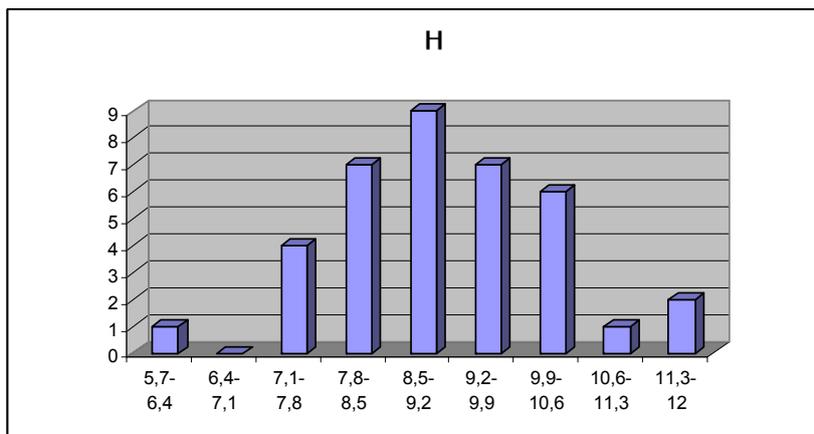


Рис. 6. Разпределение на честотите на буквата Н за 37 страници от ръкописа на Супраслския сборник: по абсцисата са честотите в проценти, а по ординатата – броят на страниците, в които честотите на буквата Н са в съответния интервал.

Ети резултати дават възможност сравнить вариативность частот при уменьшении длины текстов: с пятерок к тройкам страниц и с пятерок к отдельным страницам (**рис. 7** и **8**).

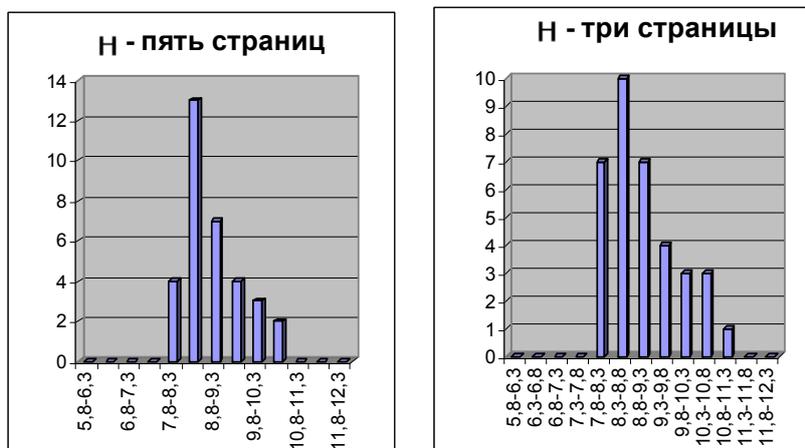


Рис. 7. Сравнение вариативности частот буквы Н при уменьшении длины текстов: с пятерок (слева) к тройкам страниц (справа) в рукописи Супраслского сборника. Частоты Н варьируют соответственно в интервалах (7,8 ; 10,8) и (7,8 ; 11,3).

Диаграммы на **рис. 7** и **рис. 8** иллюстрируют расширение интервала вариативности частот буквы Н при уменьшении объема соответствующего текста – ожидаемое изменение. В процентном соотношении оно означает расширение интервала соответственно на 17 % и 110 %.

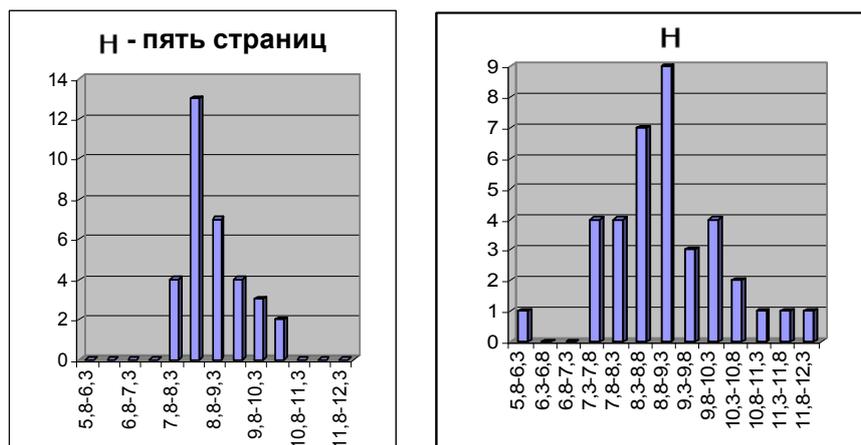


Рис. 8. Сравнение вариативности частот буквы Н при уменьшении длины текстов: с пятерок (слева) к отдельным страницам (справа) в рукописи Супраслского сборника. Частоты Н варьируют соответственно в интервалах (7,8 ; 10,8) и (5,8 ; 12,3).

3 О вариативности буквы Е

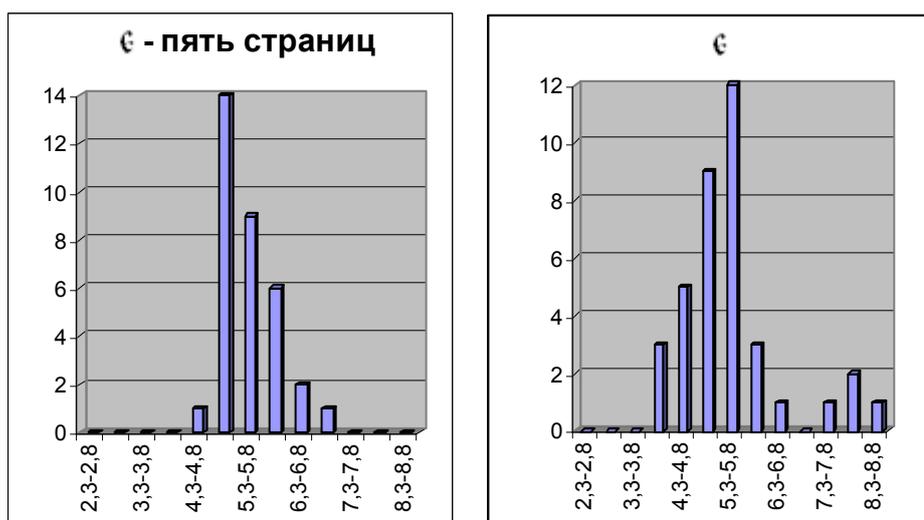


Рис. 9. Сравнение вариативности частот буквы Е при уменьшении длины текстов: с пятерок (слева) к отдельным страницам (справа) в рукописи Супраслского сборника. Частоты Е варьируют соответственно в интервалах (4,3 ; 7,3) и (3,8 ; 8,8).

Диаграмма на **рис. 9** иллюстрирует расширение интервала вариативности частот буквы Е при уменьшении объема соответствующего текста: расширение интервала на 67 %.

4 О вариативности буквы Ё

Диаграмма на **рис. 10** иллюстрирует расширение интервала вариативности частот буквы Ё при уменьшении объема соответствующего текста: расширение интервала на 160 %.

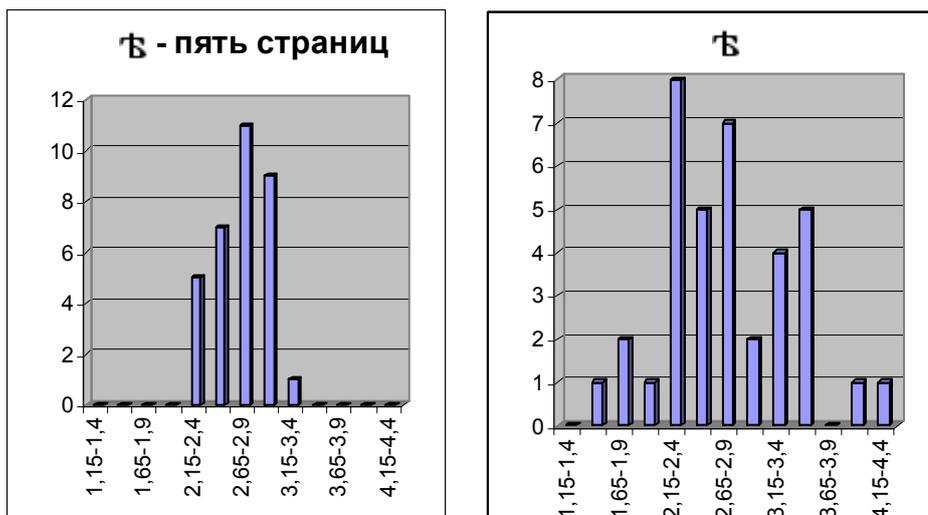


Рис. 10. Сравнение вариативности частот буквы **Ѣ** при уменьшении длины текстов: с пятерок (слева) к отдельным страницам (справа) в рукописи Супраслского сборника. Частоты варьируют соответственно в интервалах (2,15 ; 3,4) и (1,4 ; 4,4).

5 Выводы и перспективы

Число текстов фиксированной длины, для которых в нашем исследовании вычисляются частоты букв - от 33 до 37, что представляет собой выборку, достаточную для того, чтобы играть роль первоначального ориентира для выводов, относящихся к всему сборнику.

Судя по полученным нами данным, для текстов длиной в одну страницу частота данной буквы может варьировать в интервале длиной до около 5 % (т.е. плюс-минус 2,5 %). Для текстов длиной в 5 страниц этот интервал значительно уже – около 3 %.

Предположим, что нужно сравнить два длинных текста на старославянском языке (длиной больше 5 страниц по масштабам Супраслского сборника) X и Y, и что частоты некоторой буквы (хотя бы одной) в X и в Y отличаются больше, чем на 1,5 %; это отличие указывало бы на разное происхождение этих текстов – что они писаны не одним и тем же человеком («одной и той же рукой»). Более существенные отличия буквенных частот указывали бы на то, что X и Y являются творением разных школ.

Для того, чтобы установить точные параметры вариативности отдельных букв и использовать их на практике, нужны дальнейшие исследования

5 Библиография

[Супрасълски сборник, 1983] Супрасълски или Ретков сборник. Издат. на БАН, София, 1983.

Информация об авторах



Йордан Табов – Институт математики и информатики БАН, ул. Акад. Г. Бончев блок 8, 1113 София, Болгария; e-mail: tabov@math.bas.bg

Основные области научных исследований: Применения математики и информатики в гуманитарных науках, дидактика математики и информатики