

Social and Linguistic Analysis of medieval Urban Data

Scientific Advisor : Prof. Olga Khotskina

Rishabh Tiwari

Novosibirsk State university

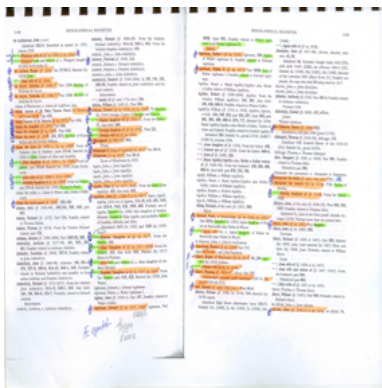
28 may, 2020

- ▶ This study shows development of social categorisation based on the social positions and relationships.
- ▶ To measure socio – economic indices such as education, income, wealth, occupation, ethnic heritage, and family dynamics to better understand the relationship between language and society.
- ▶ What can forenames tell us about English medieval society.
- ▶ The work demonstrates the next step in the development of anthroponymic research.

- ▶ It is considered as combination of linguistics and sociology.
- ▶ It takes language samples from sets of random population subjects and looks at variable that include such things as pronunciation, word choice, and colloquialisms.
- ▶ This field of study deals with anthropological linguistics, dialectology, discourse analysis, ethnography of speaking , geolinguistics, language contact study, secular linguistics, the social psychology of language, and the sociology of language.

- ▶ The source material for this thesis is drawn from the Survey of Medieval Winchester.
- ▶ It is the logical continuation of the volume Winchester in the Early Middle Ages.
- ▶ It includes various manuscripts of legal, religious, personal and other origins.
- ▶ It is the reviews of the legal structure, property distribution, town organization, town development, population size and development, the trading system.
- ▶ I selected a large sample from the whole material of the Survey of Medieval Winchester collection and described development of English society based on the first name material

- ▶ To read the data from multiple sources, PDF Documents Scanned Documents.
- ▶ I used optical Character Recognition Technique.
- ▶ I used Python for analysing the data but the data need not be in the required format always. In such cases, we convert that format (like PDF or JPG etc.) to the text format, in order to analyze the data in better way.



me Adderly, Edward (ff. 1604-29). From his ?relative Richard Adderl(e)y: 614-15, 640-1, 643. From his ?relative Stephen Adderly(e): 183.

Adderly, John, v. John oo

Nicholas (fl. 1590). :
 ce e ha ee oa Richard, - Richard Adderl(e)y.
 See 27/30-2. Married (by Adderly, Stephen, v. Stephen Adderly(e).

: Adderlye, John, v. John Adderl(e)y.

Adderly(e), Stephen (fl. 1590-1604). 4, 183, 184, 185, 589-90. Possibly related to John Adderl(e)y and Edward Adderly.

Haberdasher.

~, family of (fl. early 17th cent.) 184.

Adrian, Willi . 1502-3). ?Part 905.

le Ac(h)atour, John (cont.)

, Inhabited 622-3. Described as master by 1320,

citizen 1325.

- ▶ Regular Expressions
- ▶ Common Python Regex Methods are `re.findall()`, `re.split`, `re.sub`, `re.search`
- ▶ Word Tokenization
- ▶ Tokenization is the process of converting a text or corpus into tokens.
- ▶ Stemming.
- ▶ Lemmatization.
- ▶ Part of speech tagging (POS).
- ▶ Named entity recognition.
- ▶ Chunking

Questions

- ▶ How many males from the general number had name John in the 14th century?
- ▶ How many of these Johns had wives and how many of them were Maries?
- ▶ What was the most popular male name among merchants in the 15th century? In 1401-1449 or in 1450-1499?
- ▶ Were there any Johns having the surname Smith in the 14th century?
- ▶ How many Johns in the 15th century had sons bearing the same first name?
- ▶ Which crafts were practiced by women in 12th century? 13th century?
- ▶ Etc

- ▶ Master Thesis in social and Linguistic analysis by Prof. Olga Khotskina.
- ▶ [https://www.academia.edu/42696597/Ancient text recognition review](https://www.academia.edu/42696597/Ancient_text_recognition_review).
- ▶ [https://www.academia.edu/42984049/An approach to solve the problem of creating an automatic knowledge control system based on text analysis](https://www.academia.edu/42984049/An_approach_to_solve_the_problem_of_creating_an_automatic_knowledge_control_system_based_on_text_analysis).

THANK YOU