

Exploring the Applicability of Digital Humanities Tools to Ottoman and Turkish Languages

Exploring the Applicability of Digital Humanities Tools to Ottoman and Turkish Languages

By

Ayşe Tarhan

**Cambridge
Scholars
Publishing**



Exploring the Applicability of Digital Humanities Tools to Ottoman and Turkish Languages

By Ayse Tarhan

This book first published 2026

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2026 by Ayse Tarhan

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 978-1-0364-6409-7

ISBN (Ebook): 978-1-0364-6410-3

For my husband, my greatest inspiration.

TABLE OF CONTENTS

List of Tables.....	ix
List of Figures.....	x
Abstract	xvii
Acknowledgments	xviii
List of Abbreviations	xix
Chapter 1	1
Introduction	
Chapter 2	4
Research Methodology	
2.1 Data Collection Methods	4
2.2 The Tools	5
2.3 Data Analysis Methods	5
2.4 Data Features	6
Chapter 3	10
Literature Review	
3.1 Digital Studies on Ottoman and Turkish Languages.....	10
3.2 Overview of Language and Literature Analysis Tools and Algorithms.....	18
Chapter 4	43
Results and Discussions	
4.1 Identifying Gaps in the Digital Humanities Tools	43
4.2 Discussion and Analysis of Findings	213

Chapter 5	222
Conclusions	
References	224
Glossary.....	231

LIST OF TABLES

Table 2.1: Word Frequency Numbers Obtained in the Work Titled Dânişmend-nâme with Digital Humanities Tools.....	7
Table 2.2: Word Frequency Numbers Obtained in the Laîfî Tezkeres Kayseri Râşid Efendi Manuscript and the Hâlet Efendi Copy Using Digital Humanities Tools.....	8
Table 2.3: Word Frequency Numbers Obtained in the Dîvân of Âşık Sıdkî Baba with Digital Humanities Tools.....	9
Table 2.4: Word Frequency Numbers Obtained in the Novel Kuyucaklı Yusuf with Digital Humanities Tools	9
Table 3.1: 12 Digital Humanities Tools and Their Features.....	42
Table 4.1: List of Tools Included in the Research and Their Advantages and Disadvantages	214
Table 4.2: The Table for Compatibility of 12 Tools for Turkish and Ottoman Languages	220

LIST OF FIGURES

Figure 3.1: Akis Algorithm	14
Figure 3.2: An Image from the Akis's Demo Video	15
Figure 3.3: Conversion of the Mihri and Zeyneb Items of Latifi Tezkeres into Metatext with Transcription in Both Arabic and Latin Letters via Recyclable Katib/Scribe Tool	16
Figure 3.4: Exclude Select Vocabulary Application Produced Within the Scope of The Latifi Project	17
Figure 3.5: Word Categories Visual of Docuscope Corpus Based Rhetorical Analysis Tool	22
Figure 3.6: DocuScope Corpus Based Rhetorical Analysis Tool Bootstrap Image.....	24
Figure 3.7: Emotional Analysis Report of LIWC on Transcribed Text....	25
Figure 3.8: Voyant Tools Interface.....	26
Figure 3.9: AntGram Test Image.....	28
Figure 3.10: AntCorGen Test Image	29
Figure 3.11: AntConc 3.5.0 Interface Dated 2017.....	30
Figure 3.12: Wordcloud Image of Antconc 4.2 Version Dated 2023	31
Figure 3.13: {L}exos Interface.....	33
Figure 3.14: SketchEngine Turkish Corpora Dashboard.....	35
Figure 3.15: SketchEngine Shortcuts	36
Figure 3.16: Gephi 0.10.1 Version Image	37
Figure 3.17: Gephi Plug-ins Option.....	38
Figure 3.18: Gephi 0.10.1 Version Word Cloud Plugin	39
Figure 4.1: Processing Stage Visual of Orange3-Textable 3.13.0 Version.....	44
Figure 4.2: Word Cloud Image of Orange3-Textable 3.13.0 Version	44
Figure 4.3: Emotional Analysis Report of Linguistic Inquiry and Word Count on the Transcribed Text.....	45
Figure 4.4: Emotional Analysis Report of Linguistic Inquiry and Word Count on Arabic Letters.....	46
Figure 4.5: Testing Latin Transcribed Ottoman Text on Voyant Tool	47
Figure 4.6: Testing Arabic Letter Ottoman Text on Voyant Tools	47
Figure 4.7: Voyant Tools Analysis Interface.....	48
Figure 4.8: Voyant Tools Phrases Algorithm	49
Figure 4.9: Voyant Tools Trends Algorithm	50

Figure 4.10: Voyant Tools Export Algorithm	51
Figure 4.11: Voyant Tools Terms Algorithm	52
Figure 4.12: Voyant Tools Bubblelines Algorithm	53
Figure 4.13: Voyant Tools Bubbles Algorithm	54
Figure 4.14: Voyant Tools Cirrus Algorithm	55
Figure 4.15: Voyant Tools Links Algorithm	56
Figure 4.16: Voyant Tools Collocates Algorithm	57
Figure 4.17: Voyant Tools Contexts Algorithm	58
Figure 4.18: Voyant Tools Correlations Algorithm.....	59
Figure 4.19: Voyant Tools Document Terms Algorithm.....	60
Figure 4.20: Voyant Tools Terms Algorithm	61
Figure 4.21: Voyant Tools Trend Algorithm.....	62
Figure 4.22: Voyant Tools Documents' Words/Sentence Algorithm	63
Figure 4.23: Voyant Tools Knots Algorithm.....	64
Figure 4.24: Voyant Tools Mandala Algorithm	65
Figure 4.25: Image of the Word Sultân in Voyant Tools Mandala Algorithm.....	66
Figure 4.26: Voyant Tools MicroSearch Algorithm.....	67
Figure 4.27: Voyant Tools Phrases Algorithm	68
Figure 4.28: Voyant Tools ScatterPlot Algorithm.....	69
Figure 4.29: Voyant Tools StreamGraph Algorithm	70
Figure 4.30: Voyant Tools TermsRadio Algorithm.....	71
Figure 4.31: Voyant Tools TextualArc Algorithm	72
Figure 4.32: Voyant Tools Topics Algorithm	73
Figure 4.33: Voyant Tools Trends Algorithm	74
Figure 4.34: Voyant Tools Veliza Algorithm	75
Figure 4.35: Voyant Tools WordTree Algorithm	76
Figure 4.36: Voyant Tools DreamScape Algorithm.....	78
Figure 4.37: Voyant Tools Summary Algorithm.....	79
Figure 4.38: Interface of AntConc 4.2 version	80
Figure 4.39: File Algorithm of AntConc 4.2 Version.....	81
Figure 4.40: Word Algorithm of AntConc 4.2 Version.....	82
Figure 4.41: KWIC Algorithm of AntConc 4.2 Version	83
Figure 4.42: KWIC Algorithm Test Result in AntConc 3.5 Version.....	84
Figure 4.43: KWIC's Search Quarry List Algorithm in AntConc 4.2 version.....	85
Figure 4.44: Test Result of the Words shu'arâ, kaşide, beyt, maṭla', neşr, şî'r and şâ'ir on 7 Items of the Râşid Efendi Copy of Latîfî Tezkere in KWIC's Search Query List Algorithm in AntConc 4.2 Version	86
Figure 4.45: Plot Algorithm and Test Result in AntConc 4.2 Version	87
Figure 4.46: File View Algorithm in AntConc 4.2 Version	88

Figure 4.47: Test Result of the Turkish Suffix ki in the Cluster Algorithm in AntConc 4.2 Version.....	89
Figure 4.48: Test Result of N-Grams Algorithm ki Structures in AntConc 4.2 Version	90
Figure 4.49: Test Result of the Word “ve” in the Collocate Algorithm in AntConc 4.2 Version	91
Figure 4.50: Test Result of the Word “neşr” of the Collocate Algorithm in AntConc 3.5 Version	92
Figure 4.51: Test Result of the Persian Word “ü” in the Collocate Algorithm in AntConc 3.5 Version.....	93
Figure 4.52: Keyword Algorithm in AntConc 4.2 Version	94
Figure 4.53: Negative Keywords Algorithm in Tool Settings-Keyword in AntConc 4.2 Version	95
Figure 4.54: Keywords Algorithm in AntConc 3.5 Version.....	96
Figure 4.55: Test Result of the 8 Items in the Râşid Efendi Copy of Latîfî's Tezkere on the WordCloud Algorithm in AntConc 3.5 Version.....	97
Figure 4.56: Clone Results Algorithm in AntConc 3.5 Version.....	98
Figure 4.57: Save Current Tab Results and Save Current Tab Database Tables Algorithm in AntConc 4.2 Version	99
Figure 4.58: Saving the txt File in UTF-8 Code.....	100
Figure 4.59: Stylo Set Parameters	101
Figure 4.60: Stylo Analysis Parameters.....	102
Figure 4.61: Image of Stylo Hâlet Efendi and Râşid Efendi Corpora	102
Figure 4.62: {L}exos Interface.....	105
Figure 4.63: Corpora Loaded on the {L}exos Tool.....	106
Figure 4.64: {L}exos Manage Algorithm.....	107
Figure 4.65: {L}exos Prepare Algorithm	109
Figure 4.66: {L}exos Cut Algorithm.....	110
Figure 4.67: {L}exos Tokenize Algorithm.....	111
Figure 4.68: {L}exos Visualize Algorithm	112
Figure 4.69: {L}exos Word Cloud Algorithm Colour Arrangement.....	113
Figure 4.70: {L}exos Export Image	113
Figure 4.71: {L}exos MultiCloud Algorithm	114
Figure 4.72: {L}exos Bubbleviz Algorithm	115
Figure 4.73: {L}exos Rolling Window Algorithm.....	116
Figure 4.74: {L}exos Analyze Algorithm	117
Figure 4.75: {L}exos Similarity Query Algorithm.....	118
Figure 4.76: {L}exos Content Analysis Algorithm.....	119
Figure 4.77: SketchEngine Language and Corpora Selection List	120
Figure 4.78: SketchEngine Turkish Corpora Dashboard.....	121

Figure 4.79: SketchEngine Ottoman Corpora Dashboard	122
Figure 4.80: SketchEngine Text Type Analysis Algorithm. Contents Graph of Turkish Corpora.....	123
Figure 4.81: Collection List of Ottoman Corpora in SketchEngine Text Type Analysis Algorithm.....	124
Figure 4.82: SketchEngine Word Sketch Icons	125
Figure 4.83: SketchEngine Export Example.....	126
Figure 4.84: SketchEngine Parallel Concordance	127
Figure 4.85: SketchEngine Parallel Concordance Algorithm.....	128
Figure 4.86: SketchEngine Trends Algorithm	129
Figure 4.87: SketchEngine Trends' Year of the Document and Word Algorithm.....	130
Figure 4.88: Frequency of Use of the Word Predictability in SketchEngine Trends' Algorithm from 1989 to 2015	131
Figure 4.89: SketchEngine Bilingual Terms Algorithm.....	132
Figure 4.90: SketchEngine Trends' Word Sketch Algorithm	133
Figure 4.91: Akıl and Mantık Combinations in SketchEngine Word Sketch Algorithm.....	134
Figure 4.92: Akıl and Mantık Collocations Lists in SketchEngine Word Sketch Algorithm.....	134
Figure 4.93: KWIC List Created by the Word Akıl and Mantık in the SketchEngine Word Sketch Algorithm.....	135
Figure 4.94: List of Synonyms of the Word Akıl in SketchEngine Thesaurus Algorithm	136
Figure 4.95: Concordance List of the Word Human Detected as a Synonym of the Word Mind in the SketchEngine Thesaurus Algorithm.....	137
Figure 4.96: KWIC List of the Word Neâr in the SketchEngine Concordance Algorithm.....	138
Figure 4.97: Distribution of Hits in the Corpus Graph of the Word Neâr in the SketchEngine Concordance Algorithm.....	139
Figure 4.98: Keywords' N-Grams List of Ottoman Corpora in SketchEngine Keywords Algorithm	140
Figure 4.99: Word Frequency List of Ottoman Corpora in SketchEngine Wordlist Algorithm.....	141
Figure 4.100: Converting Excel File to CSV File	142
Figure 4.101: Nodes and Edges.....	143
Figure 4.102: Sample Nodes and Edges File Model for Use in Gephi...	144
Figure 4.103: New Project Creation Interface in Gephi	145
Figure 4.104: Uploading Nodes CSV File to Gephi.....	146

Figure 4.105: General CSV Options Stage in Uploading Nodes CSV File to Gephi 147

Figure 4.106: Import Settings Stage in Uploading Nodes CSV File to Gephi..... 148

Figure 4.107: Append to Existing Workspace Stage in Uploading Nodes CSV File to Gephi..... 149

Figure 4.108: Viewing the Nodes CSV File in Gephi in the Data Laboratory Workspace..... 150

Figure 4.109: Uploading Edges CSV File to Gephi 151

Figure 4.110: General CSV Options Step of Uploading Edges CSV File to Gephi 152

Figure 4.111: Import Settings Stage of Uploading Edges CSV File to Gephi..... 153

Figure 4.112: Append to Existing Workspace Stage in Uploading Edges CSV File to Gephi..... 154

Figure 4.113: Viewing the Edges CSV File in Gephi in the Data Laboratory Workspace..... 155

Figure 4.114: Gephi's Overview Interface..... 157

Figure 4.115: Gephi's Overview Layout Algorithm 158

Figure 4.116: Gephi's Overview Icons 159

Figure 4.117: Gephi's Overview Layout Yifan Hu Algorithm 160

Figure 4.118: Gephi's Overview Layout Fruchterman Reingold Algorithm..... 161

Figure 4.119: Gephi's Preview Interface 162

Figure 4.120: CARTO Workspace Interface 163

Figure 4.121: CARTO Workflow Example..... 164

Figure 4.122: CARTO Workspace Menu 165

Figure 4.123: An Example of a File for the CARTO 166

Figure 4.124: Steps of Uploading Files to CARTO..... 167

Figure 4.125: CARTO Builder Layer Map..... 168

Figure 4.126: CARTO Layer Algorithm 169

Figure 4.127: CARTO Layer Type, Color and Radius Algorithms..... 170

Figure 4.128: CARTO Widget Algorithm..... 171

Figure 4.129: CARTO Interactions Algorithm..... 172

Figure 4.130: CARTO Interactions Click and Hover Color Arrangement 173

Figure 4.131: CARTO Legend Algorithm..... 174

Figure 4.132: CARTO Basemap Algorithm..... 175

Figure 4.133: Adding a Layer in the CARTO Legend Algorithm..... 176

Figure 4.134: CARTO Data Table..... 177

Figure 4.135: CARTO Turkish and Ottoman Place Names Labelling View.....	178
Figure 4.136: CARTO Export Data Algorithm	179
Figure 4.137: CARTO Maps Algorithm.....	180
Figure 4.138: CARTO Sharing Algorithm	181
Figure 4.139: Recogito Interface	182
Figure 4.140: Recogito Account Settings.....	183
Figure 4.141: Recogito New File Upload Algorithm	183
Figure 4.142: Recogito Workspace View.....	184
Figure 4.143: Recogito Named Entity Recognition Algorithm	185
Figure 4.144: Recogito Named Entity Recognition Engines.....	186
Figure 4.145: Recogito Named Entity Recognition Authorities Files	187
Figure 4.146: Recogito Document Views Labelling Option Line with English Language Option	187
Figure 4.147: Project Page Created After Recogito NER.....	188
Figure 4.148: Recogito Project Page Icons.....	188
Figure 4.149: Recogito Person Tagging	190
Figure 4.150: Recogito Location Labelling.....	191
Figure 4.151: Recogito Event Labelling.....	192
Figure 4.152: Recogito Fast Algorithm	193
Figure 4.153: Recogito Relations Labelling	194
Figure 4.154: Recogito Relations Labelling Test	194
Figure 4.155: Recogito Location Labelling Image	195
Figure 4.156: Pop-up Algorithm in Recogito Place Names Tagging	196
Figure 4.157: Recogito Activity Reporting Algorithm.....	197
Figure 4.158: Recogito Entities Reporting Algorithm.....	198
Figure 4.159: Recogito Tags Reporting Algorithm	199
Figure 4.160: Recogito Download Options Algorithm.....	200
Figure 4.161: Recogito Document Settings Metadata Algorithm.....	201
Figure 4.162: Recogito Document Settings Annotation Preferences Algorithm.....	202
Figure 4.163: Recogito Document Settings, Annotation Preferences, Tagging Vocabulary Algorithm	203
Figure 4.164: Recogito Sharing and Participant Adding Algorithm	204
Figure 4.165: Recogito Labelling on the Image Workspace View.....	205
Figure 4.166: Recogito English Image Labelling Option Line.....	206
Figure 4.167: Recogito Labelling Options Line	206
Figure 4.168: Recogito Labelling Option Line and Point Algorithm with Turkish Language Options.....	207
Figure 4.169: Recogito Point Algorithm Application.....	208
Figure 4.170: Recogito Rectangle Algorithm.....	209

Figure 4.171: Recogito Tilted Box Algorithm.....	210
Figure 4.172: Recogito Linked Box Algorithm.....	211
Figure 4.173: Recogito Image Labelling Test Image	212
Figure 4.174: Recogito Image Tagging Export Algorithms	213

ABSTRACT

The first known computerised language and literature analysis was carried out by Father Roberto Busa in 1949. The number of such studies accelerated significantly after 1980 and today the tools are diversified with artificial intelligence. Unfortunately, except for a few minor studies, this development cannot be seen in the fields of Ottoman and Turkish language. Therefore, as the first step in producing computer tools for Ottoman and Turkish, it is necessary to understand the algorithms of existing English-based tools and determine the applicability of these algorithms to the Ottoman and Turkish languages.

The research conducted for this book is the first step in the Latifi Project, the aim of which is to develop a stylistic analysis computer tool suitable for the Ottoman and Turkish languages. In this study 12 tools based on text and network analysis will be examined in terms of their suitability for Ottoman and Turkish. Tests on the tools will be evaluated under the headings of *Corpus*, *Visualisation*, *Word Count*, *N-Grams*, *Structure Analysis*, *Network Analysis*, and *Labelling*. The data obtained will be evaluated in terms of the suitability of the algorithms of the tools, and suggestions will be made for the algorithms that seem suitable for the Ottoman and Turkish language and literature analysis tool or tools to be developed in the future.

ACKNOWLEDGMENTS

I would like to thank Prof Ahmet Rizaner for his advice on my research. I am most grateful to Prof Walter G. Andrews, who opened the world of Digital Humanities to me, and to the organisers of the many universities' online project meetings and education programmes, that contributed greatly to my development in this field. Finally, thank you to my husband and my family for their support.

LIST OF ABBREVIATIONS

DH	Digital Humanities
HTR	Handwritten Text Recognition
IT	Information Technology
n.d.	No Date
NELC	Near Eastern Language and Civilization
OCR	Optical Character Recognition
OTAP	Ottoman Text Archive Project
USA	United States of America

CHAPTER 1

INTRODUCTION

Analysis techniques are changing rapidly, in parallel with developments in computer technology. While the term Digital Humanities embraces all human sciences, it has begun to adapt its methodology to differing fields. The use of computer technology in literature and language studies has changed the filing methods that have been applied for years, and the data that used to be obtained only by way of extensive study is now provided to the researcher in a few minutes, thanks to these computer tools. Today, due to the development of many analysis tools, language and literature experts develop their work using computer technologies. In Europe and the USA, Digital Humanities has become a very productive discipline of study, from undergraduate to doctoral levels and beyond. The research dimensions of this field are now extending across the world.

Computer-assisted text analysis is a growing field that utilises computational techniques to examine language and literature (Mason, 2014). Many computer tools for stylistic analysis are produced in various languages, but especially in English. To develop tools that focus on the Turkish language and grammar, it is necessary to learn the algorithms of digital humanities tools developed on other languages and to determine which of these algorithms are suitable for the Ottoman and Turkish languages.

The research conducted for this book is the first step in the Latifi Project, the aim of which is to develop a stylistic analysis tool suitable for the Ottoman and Turkish languages¹.

The importance of this study is that it presents essential information to the scientific community and will serve as the primary source for the development of computer tools for the analysis of the Turkish and/or Ottoman languages.

¹ For further information about the Latifi Project, please see Chapter 3 and the Digital Ottoman Projects website: www.digitalottomanprojects.org

When Turkish scholars attempt to use currently available computerised language analysis tools to expedite their work and obtain objective data for their research projects, they may realise that these tools are not suitable for Ottoman and Turkish language structures. There is no analysis application created for the Ottoman and Turkish languages, and existing computer applications created in other languages are not suitable for Ottoman and Turkish. The numerical data obtained from the examinations in these applications are not correct. This will waste the researcher's time, and incorrect evaluations are likely to be reached. Therefore, there is a need to conduct research on this subject and address the question: "What algorithms are used for digital humanities applications, and how compatible are they for use in Ottoman and Turkish languages?"

This study may be of interest to scientists from Turkology departments, researchers who study language structures, and Digital Humanities professionals who develop computerised language analysis tools. The research involves the investigation of available algorithms for computer-aided language analysis tools, to determine which language tools use algorithms that are more suitable for the potential development of tools for the analysis of Ottoman and Turkish language structures.

The terms used throughout this study are explained below:

1. Turkish language structure

The Turkish language is thought to have origins dating back to before Christ, but the available data provide the first official records of this language as Kül Tigin stone writings in the 8th century (Akar, 2005). Turkish is an agglutinative language. Due to this structure, Turkish is thought to be from the Altaic language family. Among the Altaic languages are Turkic languages, Mongolian, Tungusic, Japanese, Korean and Ainu languages, which share similar linguistic structures with Turkish.

2. Ottoman Language Structure

Following the acceptance of Islam by the Turks, the Arabic language influenced the cultural language of Turkish, while the Persian language influenced the Turkish literary language. With these interactions, the Ottoman state created an Imperial language called Ottoman, as the language of the upper classes and the Palace. These Arabic and Persian influences entered the Turkish language structure and caused the Turkish language to become a predominantly Arabic and Persian language. However, today, the Ottoman Language is a dead language, like Latin. While the use of Arabic

and Persian languages' grammatical features in the Ottoman language disappeared after the establishment of the Republic of Turkey, many of the words from these languages remained in Turkish, their use gradually decreasing over time. The Ottoman language was not a spoken language; it was a written language, known only by Ottoman experts and the Palace.

3. Computerised Language Tools

This refers to the tools produced for the purpose of language analysis and learning.

The scope of this research covers computer tools used for language analysis, not those which have educational purposes.

4. Digital Humanities

This is a field of scientific activity that combines digital technologies and the humanities.

This study provides a comprehensive overview of existing computer-assisted text analysis tools and techniques and evaluates the tools' most suitable features appropriate for Ottoman and Turkish language and literature analysis, laying the foundation for further advancements in this field. The information given in this study will provide many ideas to experts in the Information Technology and Digital Humanities fields and will thus facilitate the creation of further new projects. Information will be given about the Ottoman and Turkish digital studies included in this study and about some of the applications that have been developed for other languages.

The study is organised as follows. The first chapter introduces the main issues and the importance of the study. The second chapter contains the research methodology used for the study. The third chapter consists of two subsections, in the first, existing digital studies for the Ottoman and Turkish languages are investigated, and in the second subsection, information is provided on the algorithms developed for other languages, particularly English-based tools, which were tested and evaluated for their suitability for the Ottoman and Turkish languages. In the fourth chapter, the results obtained in the study are presented and interpreted. The results are then discussed, and suggestions are made for the tool or tools to be created in the future. In the conclusions, the main findings are presented and possibilities for future studies are suggested.

CHAPTER 2

RESEARCH METHODOLOGY

This chapter describes the research methodology used in this study to explore the applicability of digital humanities tools to the Ottoman and Turkish languages.

Research is first made on the limited existing algorithms created for the Ottoman and Turkish languages, and then tools created for language and literature analysis in other languages and their algorithms are examined. Following this, the compatibility of these English language-based computer applications with the Ottoman and Turkish languages are tested. At this stage, corpora are created in both poetry and prose for the Ottoman and Turkish languages with both Latin and Arabic scripts. In the tests obtained, the applicability of the applications and algorithms for the Ottoman and Turkish languages are discussed. In the final section of the study, suggestions are given on how to evaluate the obtained data. This research will provide the main source for future technological developments in this field.

2.1 Data collection methods

Up-to-date information on digital humanities tools was compiled by way of searches on Google search engine and Google Scholar, Research Gate, and Academia Edu with keywords such as "language and literature analysis applications", "language and literature analysis tools", "style analysis tools", and "language and literature digital projects". To help determine the tools to be used in the study, participation was undertaken in seminars, meetings, and summer and winter schools of the institutes and research centres listed below:

- Compute Canada Federation, Humanities and Social Sciences Series
- Digital Humanities Leipzig (FDHL), Leipzig University
- Helsinki Centre for Digital Humanities, Aalto University HELDIG
DH

- Islamicate Digital Humanities Network (IDHN), Georgetown University
- Literary Studies & Digital Humanities, Leiden University
- Social ComQuant Project, Koç University
- The New York University Abu Dhabi Winter Institute in Digital Humanities, New York University Abu Dhabi Campus.

2.2 The tools

Six text mining tools, Orange3-Textable, Voyant Tools, AntConc Tool, Stylo, {L}exos, and SketchEngine were selected from the tools recommended in the training sessions and based on previous field experience. Three network and spatial analysis tools used in social sciences, Gephi, CARTO, and Recogito, were added to the study. DocuScope Project, LIWC, and The Metrical Tool Hylas tools, which do not have an infrastructure suitable for Ottoman and Turkish languages, but whose algorithm logic was found to be usable in producing a tool for the relevant languages, were also included in the study. Thus, a total of 12 tools were selected.

2.3 Data analysis methods

The basic algorithms in the tools were classified and tested under the headings *Corpus*, *Visualisation*, *Word Count*, *N-Grams*, *Structure Analysis*, *Network Analysis*, and *Labelling*. The algorithm features of all tools were evaluated in Table 3.1 in terms of these basic algorithm structures, and the results obtained through the tests were interpreted in Table 4.2.

During the selection phase of the texts to be tested on the tools, texts were selected from five different periods, ranging from the arrival of the Turks to Anatolia up to the establishment of modern Turkey. The first of the texts is Dânişmend-nâme, dated 1360, one of the first texts created by the Turks when they settled in Anatolia. The second and third are the Kayseri Râşid Efendi manuscript, dated 1546, and the Hâlet Efendi manuscript, dated 1576, which are the Tezkeres of Laîfî, one of the writers of the 16th century, when Ottoman literature was at its peak. The two manuscripts of the Laîfî Tezkeres were chosen because they show two different writing styles of the 16th century. The fourth work taken as a sample for the study was the Divan of Aşık Sıdkî Baba, a Bektashi poet who lived between 1865 and 1928 in the last periods of the Ottoman Empire. Finally, Sabahattin Ali's first work, Kuyucaklı Yusuf, written in 1931, was chosen to represent modern Turkish.

Since the first 4 texts were written with Arabic letters, they were all transcribed into the Latin alphabet.

In the study, corpora from both Arabic script and Latin script transcript formats were converted into a txt file with UTF-8 encoding formats for testing. Punctuation marks were deleted from the texts to avoid problems when the tools read the corpora. Also, capital letters were converted to lowercase letters.

2.4 Data features

The 12 tools were each tested individually on 5 corpora to determine the ability of the tools to read the corpora. CARTO, whose only analysis language was English, could not read Ottoman and Turkish letters. The tools calculated word frequency data incorrectly because they did not have Ottoman and Turkish language infrastructures. The total word counts of the tools used in the corpora in the study were close, but they did not provide the same results. The data generated by the AntConc tool, which is frequently used by literary researchers, is presented below as an example to show the lengths of the corpora taken as samples for the study. In that data, the correct word numbers obtained manually are also given after removing the suffixes and structures of the language that the AntConc tool miscalculated.

The first corpus taken as a sample in the study is *Dânişmend-nâme*, written in a simple style by Ârif Ali in 1360 during the reign of Murad II. The reason for choosing this text is that it is one of the first epics that the Turks transferred from oral tradition to writing in Anatolia and so is one of the first written works. In the study, the event in which *Dânişmend Gazi* led *Artuhtı* to Islam, which was narrated in the 15-line pages of *Dânişmend-nâme*, was selected, from the first line of 14a up to the first line of 15a.

Digital humanities tools produced based on English are not suitable for the language structure of the Ottoman or Turkish languages, as they also consider the phrase structures of the Arabic and Persian languages as words. According to the data given in Table 2.1, there is a total of 199-word types in the *Dânişmend-nâme* copy and the number of repetitions is calculated as 270. However, following sorting and corrections in this incorrectly calculated data, a total of 251 was reached.

Table 2.1: Word Frequency Numbers Obtained in the Work Titled *Dânişmend-nâme* with Digital Humanities Tools.

	Word Type	Word Total Frequency
<i>Dânişmend-nâme</i>	199	270

Laṭîfî, Abdüllatîf (Lutfullâh) b. Abdullah Kastamonulu is thought to have lived between 1491 and 1582. The first tezkere taken as a sample for the study is the copy registered at Kayseri Râşid Efendi Library, Manuscripts Department, Inventory Number 1160. The second tezkere taken as a sample is manuscript number 342 in the Süleymaniye Library, Hâlet Efendi Collection. The manuscripts of Laṭîfî Tezkeres dated 1546 and 1576 were chosen as samples for the study because they represent the plain and ornate writing traditions of the Ottoman Empire. The Ottoman language represents a period of the Turkish language, and the fact that it also contains the grammatical structures of both Arabic and Persian languages is the most important reason for choosing these tezkeres as samples. From both tezkeres, the items of poets named Ahmed Paşa, Dâ'î, Fânî, Fazlî, Mihrî, Şeyhi, Zeyneb and the Hâtîme section, which is the conclusion of the manuscript, were compiled to be used in the research.

When the word counts of the above-mentioned sections and items in the Laṭîfî Tezkere 1546 Kayseri Râşid Efendi manuscript and the word counts of the same section and items in the 1576 Hâlet Efendi copy were accessed with the tools, the data in Table 2.2 were obtained. This data showed that the tools counted structures in Ottoman and Turkish languages, such as Persian and Arabic phrases and Turkish suffixes, as words, and thus calculated the number of words incorrectly. When corrections are made on the data of the corpora, in Râşid Efendi's copy, there are a total of 629 words in the Ahmed Paşa item, a total of 229 in the Dâ'î item, a total of 373 in the Fânî item, a total of 210 in the Fazlî item, a total of 381 words are counted in the Mihrî item, a total of 700 in the Şeyhi item, a total of 284 in the Zeyneb item, and a total of 461 in the Hâtîme section.

When the incorrectly calculated Turkish, Persian, and Arabic language structures and suffixes in the Hâlet Efendi copy are removed, there are a total of 979 in the Ahmed Paşa item, a total of 241 in the Dâ'î item, a total of 1417 in the Fânî item, a total of 620 in the Fazlî item, a total of 548 in the Mihrî item, a total of 1914 in the Şeyhi item, a total of 369 in the Zeyneb item, and a total of 2516 words were calculated in the Hâtîme section.

Table 2.2: Word Frequency Numbers Obtained in the Laṭîfî Tezkeres Kayseri Râşid Efendi Manuscript and the Hâlet Efendi Copy Using Digital Humanities Tools.

	Râşid Efendi		Hâlet Efendi	
	<i>Word Type</i>	<i>Word Total Frequency</i>	<i>Word Type</i>	<i>Word Total Frequency</i>
<i>Ahmed Paşa</i>	508	813	711	1166
<i>Dâ'î</i>	172	261	202	273
<i>Fânî</i>	288	425	920	1711
<i>Fazlî</i>	185	246	459	667
<i>Mihri</i>	303	433	415	626
<i>Şeyhi</i>	516	792	1,280	2234
<i>Zeyneb</i>	238	326	314	442
<i>Hâtime</i>	345	525	1,498	2801
<i>Total</i>	2555	3821	5,799	9920

Another corpus analysed and evaluated in the study is Âşık Sıdkî Baba's Dîvân, from which 50 poems were selected, including those frequently shared on the internet and subject to many academic evaluations. Âşık Sıdkî Baba was a Bektashi poet who lived between 1865 and 1928. In selecting the poems sampled for the study, the first 44 poems of Âşık Sıdkî Baba's Dîvân and 6 further poems that were found to be most shared on social media were taken into consideration. These poems are numbered 1-44, 82, 198, 200, 299, 319 and 386 of Âşık Sıdkî Baba's Dîvân. These poems were taken as samples for the study to show the folk language representing the last period of the Ottoman Empire.

It can be seen in Table 2.3 that 2979 different word types are repeated 3673 times in the 50 poems of Âşık Sıdkî Baba, making a total of 6652 words. Of course, since computer tools see some Turkish, Persian, and Arabic structures and suffixes as words, these structures were also included in the word frequency data. When the incorrect data of the tools are removed, a total of 6388 words is reached in the corpus.

Table 2.3: Word Frequency Numbers Obtained in the Dîvân of Âşık Sıdkî Baba with Digital Humanities Tools.

	<i>Word Type</i>	<i>Word Total Frequency</i>
<i>Dîvân of Âşık Sıdkî Baba</i>	2979	6652

The last text taken as a sample for the study, representing Modern Turkish, is the novel *Kuyucaklı Yusuf*, written by Sabahattin Ali in 1931. Briefly, the novel tells the story of Yusuf, who, having lost his family in a disaster in Aydın's Kuyucak village, enters city life with his new family, but cannot get used to the artificial nature of life in the city. A total of 55445 words were used in Sabahattin Ali's novel *Kuyucaklı Yusuf*, with 14194-word types and repetitions of these words, as shown in Table 2.4. Since the tools were not compatible with Turkish, some Turkish suffixes were misread as words in this corpus, as in the first four corpora. When these errors were corrected manually, it was determined that a total of 53950 words were used.

Table 2.4: Word Frequency Numbers Obtained in the Novel *Kuyucaklı Yusuf* with Digital Humanities Tools.

	<i>Word Type</i>	<i>Word Total Frequency</i>
<i>Kuyucaklı Yusuf</i>	14194	55445

It should be noted that, because showing the visualisation and interpretations on all 5 works (which in fact showed similar results) would considerably increase the length of the study, 2 texts were chosen as examples for illustration purposes. Visualisations and interpretations, showing both the grammatical structure of the Ottoman period under the influence of Arabic and Persian and the structure of the Turkish language, were made on the Kayseri Râşid Efendi manuscript and the Hâlet Efendi manuscript of Laîfî's *tezkeres*.

CHAPTER 3

LITERATURE REVIEW

In this chapter, information is provided, firstly, about the Ottoman and Turkish languages digital studies included in this study and, secondly, about some of the applications that have been developed for other languages.

3.1 Digital studies on Ottoman and Turkish languages

Computerised studies on Ottoman and Turkish are very limited, but Turkology and Ottoman researchers have recently begun to embrace Digital Humanities and to apply its techniques to their studies. Some of these researchers have developed websites for their projects, and a platform providing information about the projects carried out in this field has been created.²

One of the studies on Ottoman texts is the digitisation of manuscripts. When we look at the world in general, digitisation started to become widespread in the 1980s, with the first studies on converting printed sources into digital file formats. In the following decade, digitisation efforts increased significantly, producing large-scale, ambitious projects with corporate funds. In the 2000s, the scanning of printed sources into digital media and transferring them from file formats such as JPEG, PNG, and TIFF to files such as txt, doc, pdf and xml with the use of Optical Character Recognition (OCR) technologies has accelerated (Terras, 2011). Over time, the use of OCR technology led to the improvement of Handwritten Text Recognition (HTR) technology.

With the digitisation of many Ottoman texts in the last 10 years, these texts have become accessible to researchers with OCR and HTR tools. These studies have been carried out in many countries of the world, such as the Republic of Turkey's yazlar.org. Apart from institutional studies, various projects based on transferring works to digital media have also been

² Tarhan, Ayse. "Digital Ottoman Projects". Accessed August 14, 2024. <https://www.digitalottomanstudies.com/>