

Doctoral Dissertation

Title: From n-grams to n-sets: A Fuzzy-Logic-Based Approach to Shakespearian
Authorship Attribution.

A doctoral thesis submitted to the Faculty of Arts, Design and Humanities of De Montfort
University in partial fulfilment for the degree of Doctor of Philosophy.

By Foukis Dimitrios,
October 2019

Leicester, United Kingdom.
Supervisor: Professor Gabriel Egan.

This page intentionally left blank.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Egan for his full support and scientific criticism throughout the three years of this PhD thesis. I would like also to thank De Montfort University for the full bursary it offered me during my studies. Furthermore, I would like to acknowledge the valuable help of Professor Hugh Craig, the director of the Centre for Literary and Linguistic Computing (CLLC) of the University of Newcastle in Australia, for the authorisation to use texts (versions of plays) that were regularised by him and his collaborators. Finally, I dedicate this thesis to my father, Apostolos Foukis, whose chronic illness made me understand two things: the importance of caring for others and that scientific reasoning is the only tool for resolving complex problems and minimising human errors that are unavoidable however hard we try.

Declaration of Authorship

I hereby declare that this thesis is the product of my own work and has been generated as the result of my own original research. Wherever contributions of others are involved, every effort is made to indicate this explicitly, with due reference to the literature, and acknowledgement of the relevant resource. In addition, no part of this thesis has been published anywhere.

Abstract

This thesis surveys the principles of Fuzzy Logic as they have been applied in the last three decades in the micro-electronic field and, in the context of resolving problems of authorship verification and attribution shows how these principles can assist with the detection of stylistic similarities or dissimilarities of an anonymous, disputed play to an author's general or patterns-based known style. The main stylistic markers are the counts of semantic sets of 100 individual words-tokens and an index of counts of these words' frequencies (a cosine index), as found in the first extract of approximately 10,000 words of each of 27 well-attributed Shakespearian plays. Based on these markers, their geometrical representation, fuzzy modelling and on the ground of Set Theory and Boolean Algebra, in the core part of this thesis three Mamdani (Type-1) genre-based Fuzzy Expert Systems were built for the detection of degrees (measured on a scale from 0 to 1) of Shakesperianness of disputed and, probably, co-authored plays of the early modern English period.

Each of these three expert systems is composed of seven input and two output variables that are associated through a set of approximately 30 to 40 rules. There is a detailed description of the properties of the three expert systems' inference mechanisms and the various experimentation phases. There is also an indicative graphical analysis of the phases of the experimentation and a thorough explanation of terms, such as partial truths membership, approximate reasoning and output centroids on an X -axis of a two-dimensional space. Throughout the thesis there is an extensive demonstration of various Fuzzy Logic techniques, including Sugeno-ANFIS (adaptive neuro-fuzzy inference system), with which the style of Shakespeare can be modelled in order to compare it with well-attributed plays of other authors or plays that are not included in the strict Shakespearian canon of the selected 27 well-attributed, sole-authored plays. In addition, other relevant issues of stylometric concern are discussed, such as the investigation and classification of known 'problem' and disputed plays through holistic classifiers (irrespective of genre).

The results of the experimentation advocate the use of this novel, automated and computer simulation-based method of classification in the stylometric field for various purposes. In fact, the three models have succeeded in detecting the low Shakesperianness of non-Shakespearian plays and the results they provided for anonymous, disputed plays are in

conformance with the general evidence of historical scholarship. Therefore, the original contribution of this thesis is to define fully functional automated fuzzy classifiers of Shakespearianness. The result of this discovery is that we now know that the principles of fuzzy modelling can be applied for the creation of Fuzzy Expert Stylistic Classifiers and the concomitant detection of degrees of similarity of a play under scrutiny with the general or patterns-based known style of a specific author (in our case, Shakespeare). Furthermore, this thesis shows that, given certain premises, counts of words' frequencies and counts of semantic sets of words can be employed satisfactorily for stylistic discrimination.

Keywords: Fuzzy Logic, Set Theory and Boolean Algebra, Fuzzy Expert Stylistic Classifiers, Fuzzy Model, Authorship Attribution and Verification, Cosine Index of Similarity, Average-‘Ideal’ Document (-Vector of Counts of Words’ Frequencies), ‘Problem Plays’, Sugeno-ANFIS (Adaptive Neuro-Fuzzy Inference System)

Abbreviation

ANFIS

COA

FIS

LDA

Max

Min

Mf/mf

PCA

RSD

SD

SiS

SVM

WAN

Full Form

Adaptive Neuro-Fuzzy Inference System

Centroid of Area

Fuzzy Inference System

Linear Discriminant Analysis

Maximum Value or Union

Minimum Value or Intersection

Membership Function

Principal Components Analysis

Relative Standard Deviation

(Sample) Standard Deviation

Stylistic index of Similarity

Support Vector Machines

Word Adjacency Networks

Note regarding the length of the thesis

The length of the thesis without the Appendices (Section 9) is approximately 95,000 words. The section of Appendices is about 20,000 words. The necessity for exceeding the general limit of 80,000 words for the core part of the thesis was due to the nature of the experimentation, the variety of technical terms the application of Fuzzy-Logic entailed and the need to follow a repetitive structure for the generation of rules and the building of an inference mechanism. A shorter version would not permit me to illustrate adequately how by analogy to the engineering field I applied in concreto in digital humanities the principles of Fuzzy-Logic through a (Matlab-based) simulated computerised environment.

Note regarding numbers

Numbers from 0 to 10 are written in words, unless they are used in formulae or as determiners (for example Validation Stage 2 or Rule 9). Numbers more than 10 are written in the numerical form.

Supplementary files

The Matlab scripts and other relevant files are stored and can be accessed from the repository: <https://github.com/dimitrios-lab/fuzzy.git>

Table of Contents

Abstract	3
1 Chapter 1: Introduction & Literature Review	19
1.1 Introduction	19
1.1.1 Review of the Literature	22
1.1.2 Definitional Context of and Scope of the Topic	23
1.1.3 Outline of the Current Situation--Novel Method of Fuzzy Logic in Stylometry 25	
1.1.4 Bridging the Gap--Importance and Advantages of Novel Methodology.....	26
1.1.5 Research Problem, Objectives and Aim	27
1.1.6 An Outline of the Order of Information in the Thesis.	28
1.2 Authorship Attribution and Style Discrimination: The 18th, 19th, and 20th Centuries.	30
1.2.1 Discriminators of Verses, Lines Ending and ‘Peculiar Inaccuracies’: Edmond Malone and James Boswell, William Spalding. (Circa 1750-1850)	30
1.2.2 Middle and Late nineteenth Century (circa 1850-1895): Proportionality in Co- authorship Attribution & The Debates of the New Shakspere Society: James Spedding, Frederick James Furnivall.	33
1.2.3 Emergence of a Mechanic Stylometric Approach: Mendenhall and his Curves- Based Discrimination of Authorial Styles. (Circa 1895-1905)	34
1.2.4 Theoretical Foundations of the Beginning of the Twentieth Century (circa 1905-1931) and External (Henslowe’s Diary) and Internal Stylometric Evidence: W.W.Greg, H. Dugdale Sykes and Charles A. Langworthy.	35
1.2.5 Zipf’s First Systematic Rank-Correlation/Frequency Law (1932) and Its Modern Application.....	37
1.2.6 Collaboration and Further Theoretical Stylometric Clues by Muriel St. Claire Byrne (1932).....	39
1.2.7 Yule’s Theorems (1944): Words at Risk, Ratios of Vocabularies and Size of Texts. 40	

1.2.8	Probabilities-Based Emerging Stylometric Technicalities (1963): Investigating the Federalist Essays through Bayesian Inference (Frederick Mosteller and David L. Wallace).....	42
1.2.9	The New Computational Stylistic Research Environment (1969-1985): Bryan Drake, Ronald E. Bee, John R. Allen, Ching Suen.	45
1.2.10	Novel Stylometric Methods (1985-2000): Machine Learning Techniques.	45
1.2.10.1	Emergence of Neural Networks in Stylometry: Robert Matthews and Thomas Merriam (1993-1994).....	45
1.2.10.2	A Revisit of The Federalist Authorship problem (D.I. Holmes and R. S. Forsyth, 1995), (J. Tweedie, S. Singh and David I. Holmes, 1996) and Mechanic Content Analysis (Colin Martindale and Dean McKenzie, 1995).	47
1.2.11	Scope and Aim of Literature Review the 18 th , 19 th and 20 th Centuries.	49
1.3	Authorship Attribution and Style Detection: The twenty-first Century.....	49
1.3.1	New Emerging Stylometric Techniques (2002-2008) at the Beginning of the twenty-first Century.....	50
1.3.1.1	John Burrows's Delta Procedure and Z-scores (2002).	50
1.3.1.2	Clustering Classification, Centroid Analysis, Vectorisation of Documents and Cosine Similarity.....	53
1.3.1.3	Linear Discriminant Analysis (LDA) as an Alternative to Principal Components Analysis (PCA) and Support Vector Machines.	56
1.3.1.4	Beyond the Delta Method: John Burrows's Zeta and Iota Measures.	60
1.3.2	Recent Perspectives of Stylometric Approaches (2009-2017).	62
1.3.2.1	Hugh Craig and Arthur F. Kinney: <i>Shakespeare, Computers, and the Mystery of Authorship</i> (2009).	62
1.3.2.2	<i>The New Oxford Shakespeare: Authorship Companion</i> (2017).	63
1.3.3	Scope and Aim of Literature Review of the Stylometric Methods of the twenty-first Century.....	66
2	Chapter Two: Methodology and Objectives	67
2.1	Research Paradigm--Post-Positivism and Functionalism	67
2.2	A Mixed Research Method	69
2.3	Type of Research: Primary Resources & Quantitative Analysis	70
2.4	Research Design and Corpus Building: Techniques of Data Collection and Tools of Analysis.....	71

2.5	Null and Alternative Hypothesis	85
2.6	Methodological Validity	85
2.7	Data Analysis: Advantages and Limitations.	86
2.8	Objectives of Research Design	87
3	Chapter Three: From Micro-Engineering to Stylometry, A paradigm of Primary Experimentation.	88
3.1	A Functionalist Approach of a Fuzzy Fan Controller.	88
3.2	Design of Membership Functions and Manual Clustering of Data Points.....	97
3.3	Design of Inference Mechanism-Rules and Output Classes of the Fuzzy Expert Systems-Stylistic Classifiers.	103
4	Chapter Four: Three Genre-Based Fuzzy Stylistic-Classifiers of Shakesperianness--Modelling and Validation-Testing Stage.	108
4.1	The Comedies-Based Fuzzy Classifier of Shakesperianness: Design of Membership Functions-Classes.....	108
4.2	Rules and Inference Mechanism of the Comedies-Based Fuzzy Classifier of Shakesperianness.....	123
4.3	Validation Stage 1 or Validation of Inference Mechanism.....	128
4.4	The Three Additional Input-Variables of the Two-Layered Complete Fuzzy-Logic Stylistic Classifier of Comedies.....	139
4.5	Algorithm of Producing SiS2 Score.....	143
4.6	Design of the Validation Stage 2/Validation of Performance: Mapping A Non-Shakespearian Well-Attributed Play and Further Development of the Inference Mechanism of the Full-Fuzzy Stylistic Classifier.....	146
4.7	Design of the Testing Stage and Further Development of the Inference Mechanism of the Full-Fuzzy Stylistic Classifier.	152
4.8	Building of the Second Layer of the Fuzzy Classifier.	155
4.9	Complete Fuzzy-System Output: Validation Stage 2 and Testing Stage.....	157
4.9.1	Experimentation with Well-Attributed, Sole-authored Non-Shakespearian Comedies.	157
4.9.2	Experimentation with Anonymous or Disputed Plays: Testing Stage.....	163
4.9.3	Some more plays for Validation Stage 2 and Experimentation with a New Anonymous, Disputed Play.	165

4.10	Validity of Performance and Testing Mechanism.....	175
4.11	Histories-Based Fuzzy-Stylistic Classifier.....	177
4.11.1	The Histories-Based Fuzzy Classifier of Shakespearianness: Design of Membership Functions-Classes.....	177
4.11.2	Inference Mechanism of the Histories-Based Fuzzy Classifier of Shakespearianness.	191
4.11.3	Validation Stage 1: Validation of Inference Mechanism.....	194
4.11.4	The Three Additional Input-Variables of the Two-Layered Complete Fuzzy-Logic Stylistic Classifier of Histories.....	198
4.11.5	Design of the Validation Stage 2, Validation of Performance: Mapping A Non-Shakespearian Well-attributed Shakespearian Play and Further Development of the Inference Mechanism of the Histories-Based Fuzzy Stylistic Classifier.	201
4.11.6	Design of the Testing Stage and Further Development of the Inference Mechanism of the Histories-Based Fuzzy Stylistic Classifier.....	204
4.11.7	Building of the Second Layer of the Fuzzy Classifier.	206
4.11.8	Complete Fuzzy-System's Output: Validation Stage 2 and Testing Stage.....	207
4.11.9	Experimentation with Three More Disputed Histories (Testing Stage).	209
4.11.10	Validity of Performance and Testing Mechanism.....	217
4.12	Tragedies-Based Fuzzy-Stylistic Classifier.	222
4.12.1	The Tragedies-Based Fuzzy Classifier of Shakespearianness: Design of Membership Functions-Classes.....	222
4.12.2	Inference Mechanism of the Tragedies-Based Fuzzy Classifier of Shakespearianness.	227
4.12.3	Validation Stage 1: Validation of Inference Mechanism.....	229
4.12.4	The Three Additional Input-Variables of the Two-Layered Complete Fuzzy-Logic Stylistic Classifier of Tragedies.	234
4.12.5	Experimentation with two Well-attributed, Sole-authored Non-Shakespearian Tragedies (Validation Stage 2).	236
4.12.6	Experimentation with Two Disputed Tragedies (Testing Stage).....	240

5	Chapter Five: Comparison with an Automated Sugeno-ANFIS (Adaptive Neuro-Fuzzy Inference System).	246
5.1	Application of Sugeno-ANFIS Adaptive Neuro-Fuzzy Inference System in Stylometry.	246
5.2	Evaluation of Sugeno ANFIS.	251
6	Chapter Six: Conclusions.	255
7	Appendices: List of Plays' Versions, Textual Addendum and Technical Appendix.	272
7.1	List Addendum: Plays (and their Date Versions) of the Core Experimentation (Chapter Four).	272
7.2	Textual Addendum: Reaching Stylometric Conclusions Quickly and Efficiently by Using Fuzzy Logic.	273
7.2.1	Corpus Building and Sets of Words.	274
7.2.2	Rules and Inference Mechanism for Decision Making in the Stylometric Area.	284
7.2.3	Classes of Output Variable and Experimentation with the Three Plays of the Testing Corpus.	303
7.2.4	Results.	303
7.2.5	Conclusions of the Primary Experimentation.	309
7.3	Technical Appendix	310
7.3.1	Open and Closed Intervals.	310
7.3.2	'Ant' Statistical Analyser.	310
7.3.3	Absolute and Relative Humidity.	311
7.3.4	Centroid of Area (COA).	312
7.3.5	Trapezoidal and Triangular Functions.	313
7.3.6	Calculation of memberships (Y-axis) of values (u) on X-axis.	313
7.3.7	'None' Subclasses of an Aggregated 'None' Class.	316
7.3.8	Technicalities of Adaptation of 'None' Classes (Subclasses).	317
7.3.9	An Informal Validation of the Fuzzy Model and the Sets-based Approach--Cosine Similarity.	317

7.3.10	Script of Fan Controller (3.1).....	321
7.3.11	Script of Fuzzy Stylistic Classifier in the Primary Experimentation.	322
7.3.12	Script of Comedies-Based Fuzzy Stylistic Classifier.	324
7.3.13	Pseudo-code of the Algorithm of One-Level-Fall from SiS1-High.....	327
7.3.14	Design of Bottom Corners of Actual and ‘Not/None’ Subclasses.....	328
7.3.15	The Concepts of ‘Evidential Intervals’ of Belief and ‘Evidential Intervals’ of Degrees of Shakespearianness Expressed in Linguistic terms.	330
7.3.16	Method of Forming Compact SiS Intervals if SiS Intervals are Larger than 0.25.	331
7.3.17	Algebraic Product and Algebraic Sum.	332
7.3.18	Epoch and Hybrid Learning Algorithm (in ANFIS experimentation).....	333
7.3.19	Explanation of the Effect of ‘Overfitting’	333
7.3.20	Script of Histories-Based Fuzzy Stylistic Classifier.	335
7.3.21	Script of Tragedies-Based Stylistic Classifier.	338
7.3.22	Script of a Trained Sugeno-ANFIS Classifier.	340
	Works cited	348

Figure 1:	The optimal hyperplane separating theorem.	59
Figure 2:	Membership function of temperature.	89
Figure 3:	Membership function of humidity.....	90
Figure 4:	Membership function of output variable, fan’s speed.	91
Figure 5:	Two inputs, one rule.	93
Figure 6:	One input, two rules.	95
Figure 7:	Two inputs and two rules.	96
Figure 8:	Triangular and trapezoidal (triplets, duplets) patterns.....	99
Figure 9:	SiS and four classes of output variable.	105
Figure 10:	Membership functions of Set One to Four.	115
Figure 11:	Membership functions of Set Five.	115
Figure 12:	Structure of the comedies-based Fuzzy Stylistic Classifier.	118

Figure 13: Y-axis expresses Mean and StDev in numbers. RSDs are measured in percentages. The numbers on X-axis correspond to the five sets.	119
Figure 14: ‘Not/none’ subclasses of Set Three in the comedies-based fuzzy classifier of Shakespearianness.	123
Figure 15: Graph of counts of frequencies of 100 individual words in 12 Shakespearian comedies.	137
Figure 16: Counts of words’ frequencies of Set One.....	138
Figure 17: Structure of the complete comedies-based fuzzy classifier.....	139
Figure 18: Memberships of Set Five.....	140
Figure 19: Output SiS as input: SiS1.	141
Figure 20: Membership functions of Cosine Similarity.	143
Figure 21: Script for classes of Four Sets.	148
Figure 22: Membership functions of Set One and Two.....	149
Figure 23: Script for classes of Four Sets.	153
Figure 24: <i>The Staple of News</i> , SiS1 (R22, R23, R24, R25).	160
Figure 25: <i>The Staple of News</i> , SiS1 (R22, R23, R24, R25).	160
Figure 26: Input variables of <i>The Staple of News</i> for the production of SiS1.	161
Figure 27: <i>The Staple of News</i> , SiS2 (R32, R33).....	161
Figure 28: <i>The Staple of News</i> , SiS2 (R32, R33).....	162
Figure 29: Input data points for the production of SiS2 of <i>The Staple of News</i>	162
Figure 30: Input variables of <i>All’s Well That Ends Well</i> for the production of SiS1.	164
Figure 31: Script of four sets’ membership functions-classes.	167
Figure 32: Classes of Set Five.	174
Figure 33: The scatter plot of the SiS interval of the data points of five comedies under scrutiny.	176
Figure 34: Basic membership functions for Set Two of the six Shakespearian histories.....	181
Figure 35: Membership functions including the ‘not/none’ classes (see black triangles) inside the classes ‘2a’ (first blue triangle) and ‘2b’ (red triangle) of Set Two of the six Shakespearian histories.....	182
Figure 36: Zooming in the black triangles of the Figure 35.	182
Figure 37: Y-axis expresses Mean and StDev in numbers. RSDs on Y-axis are measured in percentages. The numbers on X-axis correspond to the five sets. The left figure shows the tendencies for histories and the right for the previously analysed comedies-based model.	185

Figure 38: Comparison of the relative standard deviations (RSDs) of Comedies and Histories.	185
Figure 39: Histograms of data points of Set Two in Comedies (X-axis) and Histories (Y-axis).	186
Figure 40: Script for actual and ‘not/none’ classes of the four sets’ data points. With particular focus on Set One [Input1].	191
Figure 41: <i>Henry IV Part I</i> ’s SiS1-score (0.794) produced by rules R1 and R8. SiS2 defaults to 0.5.	195
Figure 42: Memberships and data points of Set Five in the six histories.	199
Figure 43: Membership functions of Cosine Similarity.	200
Figure 44: Script for classes of Four Sets.	205
Figure 45: Data points of input variables of <i>Henry VI Part I</i> for the production of SiS2.	209
Figure 46: Script for the classes of Four Sets.	210
Figure 47: Membership functions of Set Five.	213
Figure 48: Counts of words’ frequencies of <i>Edward III</i> and the average-ideal document.	215
Figure 49: The scatter plot of the SiS interval of the data points of four histories under scrutiny.	220
Figure 50: Y-axis expresses Mean and StDev in numbers. RSDs on Y-axis are measured in percentages. The numbers on X-axis correspond to the five sets.	227
Figure 51: Memberships and data points of Set Five in the nine Tragedies.	234
Figure 52: Membership functions of Cosine Similarity.	235
Figure 53: Script displaying the membership functions of five of the seven inputs of the tragedies-based model.	237
Figure 54: The scatter plot of the SiS interval of the data points of four tragedies under scrutiny.	244
Figure 55: Distribution of the 34 plays and their target SiS score.	249
Figure 56: Selection of the number and type of membership functions for each of the five sets of Sugeno-ANFIS model.	250
Figure 57: Membership functions of Set Two’s data points in the Sugeno-ANFIS model.	250
Figure 58: Sugeno-ANFIS training of five sets’ data associated with their SiS1 scores of the core experimentation.	251
Figure 59: The structure of the adaptive fuzzy neuro-inference mechanism (Sugeno-ANFIS).	252
Figure 60: A Π -Shaped membership function.	266

Figure 61: An example of a probabilistic co-authorship estimator.....	268
Figure 62: Twelve plays and six sets of words.	277
Figure 63: Membership functions of six sets.	280
Figure 64: Eight combinations of Shakespearian style.	281
Figure 65: An example of ‘Not’ Class.	284
Figure 66: SiS and four classes of output variable.	303
Figure 67: Trigger of Rule 8 for the SiS of <i>All’s Well That Ends Well</i>	305
Figure 68: Minimum of memberships of data points of <i>All’s Well That Ends Well</i>	305
Figure 69: Minimum of memberships of data points of <i>Henry VI Part 1</i>	307
Figure 70: Minimum of memberships of data points of <i>The Jew of Malta</i>	308
Figure 71: ‘Ant’ Statistical Analyser.	311
Figure 72: Membership functions-classes of humidity.....	312
Figure 73: Premise parameters of a trapezoidal class and calculation of the degree of membership (y-value) and x-value.	314
Figure 74: Premise parameters of a triangular class and calculation of the degree of membership (y-value) and x-value.	315
Figure 75: An aggregated ‘none’ class.	317
Figure 76: An on-line tool of calculation of the cosine similarity between two documents- vectors.....	320
Figure 77: An example of calculation of cosine index.	321
Figure 78: An example of calculation of the cosine index from actual experimentation.	321
Figure 79: Membership functions of Set One for the comedies-based model.....	329
Figure 80: Membership functions of Set Two for the comedies-based model.	329
Figure 81: Membership functions of Set Three for the comedies-based model.	329
Figure 82: Membership functions of Set Four for the comedies-based model.	330
Figure 83: Membership functions of Set Five for the comedies-based model.	330
Figure 84: ‘Overfitting’ in the form of a complex polynomial equation.	334
Figure 85: ‘Overfitting’ in terms of a linear relation.	334
Table 1: Length of texts of the corpus of 27 well-attributed, sole-authored Shakespearian plays.....	74
Table 2: Data points of 12 well-attributed Shakespearian comedies.	110
Table 3: Five Sets’ RSD.	119

Table 4: Data points of 12 well-attributed Shakespearian comedies and their contribution to actual classes. Similarly coloured cells represent the distinct classes.	120
Table 5: Combinations that are found in more than one Shakespearean comedy.	122
Table 6: Classification of the 12 well-attributed Shakespearian comedies.	128
Table 7: Validation SiS1 scores of the 12 well-attributed Shakespearian comedies.	132
Table 8: Indices of cosine similarity of the angle formed by each of the 12 Shakespearian comedies-vectors with the vector of the average-ideal document.	135
Table 9: Descriptive statistics of the indices of Cosine Similarity of the 12 comedies.	142
Table 10: Five Sets' data points of <i>The Staple of News</i>	148
Table 11: Five Sets' data points of <i>All's Well That Ends Well</i>	152
Table 12: SiS1 and SiS2 results of <i>The Staple of News</i>	162
Table 13: SiS1 result of <i>All's Well That Ends Well</i>	163
Table 14: SiS1 and SiS2 of <i>All's Well That Ends Well</i>	164
Table 15: Sets' data points of three new comedies under scrutiny.	166
Table 16: SiS1 result of <i>A Mad World, My Masters</i>	168
Table 17: SiS1 result of <i>The Wild Goose Chase</i>	169
Table 18: SiS1 result for <i>Measure For Measure</i>	169
Table 19: Index of cosine similarity of the three new plays with the average-ideal document.	170
Table 20: Data points of three new plays under scrutiny.	172
Table 21: SiS1 and SiS2 of <i>A Mad World, My Masters</i>	173
Table 22: SiS1 and SiS2 of <i>The Wild Goose Chase</i>	173
Table 23: SiS1 and SiS2 of <i>Measure For Measure</i>	175
Table 24: Data points of the five sets of the six Shakespearian histories.	178
Table 25: Mean, Standard Deviation and RSD of the five sets in the six histories.	183
Table 26: Mean, Standard Deviation and RSD of the five sets in the 12 comedies.	183
Table 27: Data points of six well-attributed Shakespearian histories and their contribution to actual classes. Similarly-coloured cells represent the distinct classes.	189
Table 28: Validation SiS1 scores of the six Shakespearian histories.	195
Table 29: Cosine index of similarity of the six well-attributed histories with the average-ideal history.	197
Table 30: Descriptive statistics of the index of Cosine Similarity of the six histories.	199
Table 31: Five Sets' data points of <i>Edward II</i>	202
Table 32: Five Sets' data points of <i>Henry VI Part 1</i>	204

Table 33: SiS1 score of <i>Edward II</i>	208
Table 34: SiS1 and SiS2 of <i>Edward II</i>	208
Table 35: SiS1 score of <i>Henry VI Part 1</i>	208
Table 36: SiS1 and SiS2 of <i>Henry VI Part 1</i>	209
Table 37: Data points of Sets of <i>Edward III</i>	210
Table 38: Data points of Sets of <i>Henry VI Part 2</i>	210
Table 39: Data points of Sets of <i>Henry VI Part 3</i>	210
Table 40: SiS1 and SiS2 results of <i>Edward III</i>	212
Table 41: SiS1 and SiS2 results of <i>Henry VI Part 2</i>	213
Table 42: SiS1 and SiS2 results of <i>Henry VI Part 3</i>	213
Table 43: Cosine indices of <i>Edward III</i> , <i>Henry VI Part 2</i> and <i>Henry VI Part 3</i> in comparison with the average-ideal document.....	214
Table 44: SiS1 and SiS2 results of <i>Edward III</i>	216
Table 45: SiS1 and SiS2 results of <i>Henry VI Part 2</i>	216
Table 46: SiS1 and SiS2 results of <i>Henry VI Part 3</i>	216
Table 47: Counts of words' frequencies of six well-attributed Shakespearian histories (Rows 1-6) and <i>Edward III</i> (last row).....	219
Table 48: Data points of the five sets of the nine Shakespearian tragedies.....	224
Table 49: Data points of nine well-attributed Shakespearian tragedies and their contribution to actual classes. Similarly coloured cells represent the distinct classes.....	225
Table 50: Mean, Standard Deviation and RSDs of the five sets in the three genres.....	226
Table 51: Combinations that are found in more than one tragedy.....	227
Table 52: Validation SiS1 scores of the nine Shakespearian tragedies.....	231
Table 53: Cosine index of similarity of the well-attributed Shakespearian tragedies with the average-ideal tragedy.....	233
Table 54: Descriptive statistics of the indices of Cosine Similarity of nine tragedies.....	235
Table 55: Sets' data points and cosine indices of two well-attributed, non-Shakespearian tragedies. <i>The Jew of Malta</i> and <i>The Spanish Tragedy</i>	237
Table 56: SiS1 and SiS2 of <i>The Spanish Tragedy</i>	239
Table 57: SiS1 and SiS2 of <i>The Jew of Malta</i>	239
Table 58: Sets' data points and cosine indices of two disputed tragedies. <i>Timon of Athens</i> and <i>Titus Andronicus</i>	241
Table 59: SiS1 (R28, R29) and SiS2 (R30, R31) of <i>Timon of Athens</i>	242
Table 60: SiS1 (R1, R32) and SiS2 (R26, R34, R36) of <i>Titus Andronicus</i>	242

Table 61: Training corpus of Sugeno-ANFIS composed of 24 well-attributed Shakespearian plays and 10 non-Shakespearian or disputed plays (see the 10 plays in grey background). In the first five columns with data there are the inputs of the five sets' counts and the last column is the target-assigned SiS1.	249
Table 62: Size of extracts in primary experimentation.	275
Table 63: Data points of nine well-attributed Shakespearian plays (C= Comedies, T= Tragedies, H= Histories).....	278
Table 64: Data points of the three plays' six sets in the Testing Stage.	279

1 Chapter 1: Introduction & Literature Review

The first section of this Chapter (1.1) introduces the principles of Set theory and Fuzzy Logic with emphasis on the fuzzy methodology of control systems, known as Mamdani-Fuzzy inference and there is also a description of the definitional context and the scope of the thesis. Section Two of Chapter One (1.2) reviews the literature and the methods of authorship attribution and style discrimination in the period 1700-2000. The primary aim of this review is to survey the strengths and weaknesses of the different theorisations in stylometry from the primary, rather unsystematic, efforts of the dawn of the eighteenth century to almost the end of the twentieth century. Section Three (1.3) of the same Chapter explores the sophisticated, mainly machine learning based, techniques of authorship attribution and stylistic detection in the last two decades, during the period of 2002-2018.

1.1 Introduction

Authorship attribution as ‘the science of inferring characteristics of the author from the characteristics of documents written by that author’ is a challenging and complex field in which a variety of methodological tools can be applied. The scope of this thesis relates to the application in the area of digital humanities of a computational technique employed widely in micro-electronics known as Fuzzy Logic. Though the term fuzzy is usually synonymous with the adjectives ‘vague’ and ‘unclear’, in science fuzzy as a determinant of the word ‘logic’ is associated with the mathematical formalisation of Set Theory and it is employed for modelling attributes, phenomena, real world events or situations that are characterised by degrees of certainty. Fuzzy Logic is applied in this thesis in the context of textual criticism and uncertain authorship as ‘a basic concept which, though fuzzy rather than precise in nature, may eventually prove to be of use’ when there are ‘decision processes involving incomplete or uncertain data’ (Zadeh 1968, 12). Undoubtedly, as stylistic changes over time characterise even the way of writing of the same author, what has been called as stylochronometry (Forsyth 1999, Burrows 2002, 279), the general task of authorship attribution of anonymous, strongly disputed or co-authored plays is a complex decision process involving uncertainty that is derived from the fact that every written text contains in

general differentiated textual data, which can be processed for the identification of stylistic patterns that are found in other texts written in approximately the same period.

In many scientific disciplines and in real life there is sometimes a preference for modelling uncertainty and searching for solutions by using binary logic and functions analysis. Binary logic defines the boundaries of a problem domain through the use of two states (of one bit), zero and one in electronic circuitry, on and off in electronics, true or false in logic science, ill or healthy in medical science, existing and not existing in philosophy and the relevant theories of phenomenology and existentialism. At the same time, it has been a common understanding that in complex systems the delineation of the features of entities, systemic relations and the mapping of input to output data cannot be accomplished without the application of mathematical functions, especially if there is a need for precise solutions. A long series of multiple calculations, the function composition and the formation of complex rules require great processing effort, which the human being must expend when confronted with non-linear problems or multi-component systems that necessitate prediction and a classification of truth values. In the field of stylometry, an example of such problematic situations or complex problems is the authorship attribution of anonymous and co-authored plays. The difficulty of combining binary--that is, simply, numerical--logic with functions derives from the fact that this kind of logic entails the danger of over-generalisation with loss of information, and that functions need sometimes to be complex in order to deal with any possible combination of input data in a systemic environment with many uncertain parameters. (This numerical or quantifiable logic is sometimes described as crisp logic, with numerical logic having the quality of crispness – it is used in this sense throughout this thesis.) These two elements, loss of information and perhaps non-avoidable complexity, are not the only burdens of this traditional binary logic-based modelling and problem-solving approach. The requirement of the extensive use of multivariate statistics in a multi-level research design, the need for a massive amount of input data, the lack of methodological flexibility and the crispness of the known variables, are all additional factors that make the traditional tools of conventional analysis time consuming and inefficient in certain cases, as in the area of computational stylistics. This sort of unnecessary complexity and lack of flexibility can be confirmed by exploring the impact in stylometry of the Bayesian Inference of Frederick Mosteller and David L. Wallace in their book *Inference in an Authorship Problem* (Mosteller and Wallace 1963), where they use a plethora of delicate statistical methods and complex formulae. Despite their usefulness for a specific problem, in this case

the authorship attribution of the Federalist essays (1787), these complex multivariate methods cannot be and, in fact, have not been employed widely for attributing other works.

In this thesis, there is a novel stylometric methodology that aims at applying an engineering technique, Fuzzy Logic, which has been adopted successfully in the design of micro-controllers and in the area of robotics for navigation modelling. Micro-controllers are programmable device-embedded micro-chips that process an input of numerical information and execute a specific task for the device. Such a machine is called a controller because through its programmed memory it controls the functions of the device after processing the input data. An example of the use of fuzzy-microcontrollers is in automatic vacuum cleaners (robotic vacuums). Fuzzy Logic is a branch of natural maths developed in 1965 (Zadeh 1965; 1968). With the term ‘natural maths’ I refer to the development of non-complex and practical cognitive strategies, thus not necessitating the use of complex formulae, in order to solve problems that require some mathematical reasoning. Practically, Fuzzy Logic has been developed in the last three decades and contributed to the progress of micro-electronics and industrial engineering. The term ‘fuzzy’ signifies the indefinite, but this lack of clarity is identified in its application in the real world and science with a deterministic sense providing an optimised and accurate mode of describing attributes of objects’ classes and physical quantities as input and output data.

The investigation of the similarity of anonymous and co-authored texts to the works of Shakespeare (what I call their Shakespeariness) is carried out through a fuzzy controller that is based on the stylistic features of 27 well-attributed, sole-authored Shakespeare plays (see Table 1 in Section 2.4 and Section 7.1, which refers also to the versions of the plays). By the term well-attributed is implied that the corpus contains Shakespearian plays for which historical scholarship has not found internal evidence of contributions by other authors and hence the *prima facie* evidence (typically title-page attribution to Shakespeare) should be accepted. Despite any subjectivity in this claim, a canon of plays of such a distant era (sixteenth century) has to be formed on a number of acceptances, even though participation of other writers is not 100% rejected. Naturally, such acceptances may change over time and indeed suspicion is currently being cast on the sole-authorship of Shakespeare's *King John* (Merriam 2016; 2017; 2018) that this thesis assumes to be sole-authored. In this context and similarly to the fact that ‘style is a probabilistic concept’ (Dolezel and Bailey 1969, 328), it can be argued that the process of the formation of a ‘corpus of well-attributed Shakespearian plays’ in any experimentation is set on an approximate (but high) probabilistic threshold. The useful notion of single-author, well attributed plays is found in the book *Shakespeare*,

Computers, and the Mystery of Authorship by Hugh Craig and Arthur F. Kinney and all the 27 plays of the corpus of this thesis (including '*King John*') are listed as such (well-attributed Shakespearian plays) in Appendix A (Craig and Kinney 2009, 217–18).

In the investigation of this thesis, there are three fuzzy controllers, one for each of the genres of tragedy, comedy and history. Each of the three controllers in this thesis has been programmed using the counts of four sets of a total of 100 words and an index of the counts of the words' frequencies in the 27 known Shakespeare plays (see Table 1 on page 74). These counts are represented in the fuzzy controller through appropriate membership functions, recorded as geometrical shapes with certain mathematical properties. Once it has been programmed with the values for well-attributed Shakespeare plays, the Fuzzy-Logic-based stylistic controller is given as its input the counts, in the play being tested, of the frequencies of occurrence of the words in the four sets, recorded as overall counts for each set. The Fuzzy-Logic controller's task is to evaluate this play's Shakespearianness by comparing it with the known data points of the Shakespearian plays of each genre, using what is called its inference engine to output a 'Stylistic index of Similarity' or 'SiS' (or a SiS interval), a term coined by the present investigator. The evaluation of the quantification and its comparison with the known, already modelled Shakespeare's stylistic data are plotted using a Cartesian coordinate system, thus an X axis and a Y axis. This kind of experimentation with the fuzzy controllers is called also fuzzy simulation, since the effort is to 'emulate', or simulate in a fuzzy mode, 'what the human response would be and apply the most intelligent fit to the data' (Omega 2018). In our case and for problems of stylometric nature, I attempt to find for anonymous or co-authored plays the degrees of similarity with the actual Shakespearian style of the known corpus and indirectly resolve problems of authorship attribution by deducing how Shakespearian the anonymous or co-authored plays are. It has to be emphasised that Fuzzy Logic as an intuitionist tool, named as Mamdani-Fuzzy Inference after its inventor E. H. Mamdani (Mamdani and Assilian 1975), has never before been employed for problems of authorship verification and attribution.

1.1.1 Review of the Literature

The survey of past literature on this topic commences in with the exploration of foundations and the stylometric approaches of various scholars before the boom of the New Shakspeare Society in the late nineteenth century. In the primary stages of stylometric period (circa 1750-1850), the main discriminators were phraseological peculiarities, and rhyme and

other elements of metrical analysis. There is also a discussion of authors' styles and the debates of the scholars of the New Shakspeare Society (1873). Then, there is a description and assessment of the emergence of a novel mechanic stylometric method just before the beginning of the twentieth century, which is the use of histograms and graphical curves by Thomas Corwin Mendenhall (1841-1924), who devised a statistical methodology that looked methodically for length-based classes of words. In the middle of Section 1.2, there is a reference to George Zipf's First Systematic Rank-Correlation/Frequency Law (1932) and the theorems of G. Udny Yule (1944) about the nature of frequency distribution, the author's vocabulary 'at risk' and other aspects of stylistic differentiation. Among others, in Section 1.2 efforts were made by the present investigator to provide a comparative description of the aspects of practical experimentation of authorship verification and attribution whether the focus is on the Shakespearian and works of the Elizabethan/Jacobean era (1558-1625) or the collection of essays and documents of political/constitutional nature written in 1787 known as the Federalist essays. Near the end of the twentieth century, emphasis was laid on machine learning techniques, such as neural networks, and various classification and profiling methodologies. The Section Two of Chapter One (1.2) tells the story of this work up to the end of the twentieth century.

In 1.3, there is a description and theorisation of the novel methods of computational stylistics in the twenty-first century. John Burrows's Delta procedure (Burrows 2002) along with Zeta and Iota methods (Burrows 2007) are also reviewed in this section.

In general, in this thesis, Yule's and Zipf's theorems and also some of John Burrows's findings, such as the importance of the detection of the non-systematic use of specific words-terms by an author, have been considered regarding the building of the textual corpus and the selection of the words and their grouping into specific sets with different features, such as having generally high, medium and low absolute and relative-frequency counts.

1.1.2 Definitional Context of and Scope of the Topic

Two of the major theoretical foundations of this thesis are propositional calculus with the use of the Boolean algebraic operations, and Set Theory (Stoll and Enderton 2018), the latter of which was developed at the end of the nineteenth century by Georg Cantor (1845-1917), a German mathematician (Encyclopaedia Britannica 2018). Cantor researched the subset relations of sets of numbers and formed a relevant theory about the cardinality, continuity and infinity of sets of numbers that can be associated or paired. The practical

usefulness of Set Theory in this thesis lies in the fact that by grouping objects and numbers into sets we can specify numerically the relations between objects' properties and, in a way, form their topology of symmetries or asymmetries. This kind of topology in our context of stylometric research is parallel to the stylistic features of an author's writing, whereas the objects' properties can be rephrased as the counts of sets of words and counts of words' frequencies.

Other tools of computational intelligence have also been employed for building parts of our fuzzy inference model, such as heuristic search and an adapted, rather hybrid, iterative-deepening search process. As this process of iteration during the building of a rules database is explained later in Chapter Three and Four, the focus here is only on the propositional calculus and Set Theory. The classical propositional calculus, known also as binary or Boolean logic, defines whether an element-object called x of a universe called X belongs to a set of elements-objects called S . In other words, according to the traditional logic we say that an element-object (x) belongs (denoted by the symbol \exists) to a set of objects (S) or does not belong (denoted by \nexists), and consequently the proposition-statement about the element x as a member of S can be only true (1) or false (0). Thus, the membership of x in S is defined inside a closed interval of only two values $[0, 1]$, which are also the endpoints of the interval. In Fuzzy Logic there are more than two truth values and the concept of membership $\mu_S(x)$ to a collection of objects is defined in the range of 0 to 1. (For practical reasons, I formalise this range of values as decimals.) As the grade of membership approaches the unity (1), the degree of membership rises. Therefore, in Fuzzy Logic the interval of truth values about the proposition of x belonging into a set S has the same endpoints $[0$ and $1]$ with the interval in classical (Aristotelian) logic, but it also includes the values between them, that is in decimals 0.1, 0.2... 0.9. Therefore, in Fuzzy Logic, the proposition-statement about the element x as a member of S can be true (1) or false (0) or has a membership degree of truth. The propositional logic and Set Theory jointly provide us with the mathematical formalisation and systematisation of the relations of various objects of different sets during decision-making processes. The fact that in Fuzzy Logic we are not confined to the strictly defined options of only 0 or 1 assists us in producing evaluations and classifications of objects-elements based on approximate reasoning. By avoiding the crispness of absolute truths (only 0 or 1) during the process of decision making, we can systematise our thinking when we cope with problematic situations, systems or phenomena that are characterised by high uncertainty and non-linearity. This thesis shows how Fuzzy Logic can be applied to the problem of determining the authorship of writings for which the

author is unknown or disputed and in which I have counted particular features such as the frequency of use of certain words and the counts of sets of words. Another element of Fuzzy Logic is the use of linguistic variables and modifiers for the representation of similarity or dissimilarity of objects' properties. Modifiers and linguistic variables allow the characteristic human quality of imprecise classification--as in 'Friday is *kind of* the weekend', 'John is very tall', and '5 feet and 1.4 inches (156 cm) is an extremely short height'--to be embodied in computer systems to enable human-like distinctions. In these examples, the words 'kind', 'tall' and 'short' are linguistic variables, whereas the adjectives 'very' and 'extremely' are modifiers. Linguistic variables and modifiers enable us to group into single sets (**S**) certain elements-objects ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_n$) of a universe **X**, considering their (approximate) similarities. In this context, if Jane is also 'very tall' we can group her with John in the class of 'very tall' human beings, and so we can discriminate Jane's and John's attribute from David whose height is '5 feet and 1.4 inches' and whose category might be that of 'extremely short' human beings. Of course, the exact representation of the linguistic variables, as will be discussed in Chapters Three and Four, needs careful application of a systematic design methodology. Linguistic variables are defined numerically based on the principles of fuzzy-approximate reasoning and they are represented by an interval of values. In other words, Jane might be a bit taller, at six feet and four inches than John who is six feet and three inches, and though they belong in the same class of 'very tall' human beings, Jane has a higher membership degree to their common class. This process of objects' properties' systematisation through linguistic variables, partial truths and classes in the field of Fuzzy Logic assists us in dealing with the uncertainties of a complex phenomenon and helps in deriving more precise conclusions and results than if we selected the monolithic, all-or-nothing, binary approach. With the use of modifiers, linguistic variables, partial truths and logical connectives, I can apply Fuzzy Logic to the counts I have made of the various features in the writing in order to assess numerically the stylistic similarities of a disputed or anonymous text to an author's known style. Based on the corpus of 27 well-attributed, sole-authored Shakespeare plays, this thesis's ultimate goal is to detect the degrees of Shakespearianness, thus to produce a quantified index of stylistic features of Shakespearian writing in disputed or anonymous plays.

1.1.3 Outline of the Current Situation--Novel Method of Fuzzy Logic in Stylometry

The application of fuzzy methods has been a challenge in many, mainly practical, areas, such as engineering, robotics, image processing and computer science. Concerning theoretical fields, there have been fuzzy approaches in digital forensics analysis, for the detection of authorship of technical documents and in decision making by combining heuristics, models of childhood conceptions and intuitionism (Reyna 2012). In addition, Fuzzy Logic has been applied in the area of textual sentiment (Haque and Rahman 2014) and ontological and semantic component analysis (Amira and Amel 2015). But all these applications relate to the domain of neurolinguistics and to the fuzzy-tracing of cognitive representations of memory's mechanisms and information retrieval during the processes of reasoning, judgment and decision making. Thus, until now the researchers have studied only the role of semantics and the effect of intuition on cognitive mechanisms and tried to build fuzzy models of cognitive schemes using overlapping hierarchical taxonomies, as produced in various experimentation trials with words, sentences and images. The stylometric field has seen the application of almost all the major advanced methods of computational intelligence, such as neural networks and the classifiers of Support Vector Machines (SVM). But, as already mentioned, Fuzzy Logic has never before been employed for problems of authorship verification and attribution.

1.1.4 Bridging the Gap--Importance and Advantages of Novel Methodology

The Mamdani-Fuzzy system includes an inference mechanism and the production of its output is based on the logical connectives applied on the inputs, the memberships' truths of propositions regarding the classification (specified as degrees of membership) of a number of elements into discrete sets of objects **S**. An advantage of Mamdani-Fuzzy Inference is that its technical features have already been modelled in the mathematics software called Matlab, and have been extensively used for building controllers for complex control situations, such as the combination of steam engines with boilers and automatic temperature control.

In this thesis I employed the principles of Fuzzy Logic in order to build three genre-based Stylistic Expert systems/Classifiers for the direct production of authorship verdicts and the detection of the degrees of Shakespeareanness in disputed or anonymous, possibly co-authored, plays. The aim is to show the large potential that such an advanced computational method can have for resolving problems of authorship verification and authorship attribution. Whether it concerns stylistics or profiling, the (Matlab-based) computerised environment of Fuzzy Logic can assist in the modelling of the known style of an author. So, based on the

Fuzzy-Logic-based quantification and modelling of the stylistic features of the well-attributed, sole-authored plays of an author, it is possible by employing natural approximation methods to proceed to the detection of this author's style in disputed or anonymous, possibly co-authored, plays. From a general theoretical standpoint, the investigation in this thesis deals with the development of a new computational method that can be employed in stylistics and, therefore, the principal goal is not to study Shakespeare. Nevertheless, the fact that Shakespeare has written a plethora of plays in three genres (which is not the case with many other of his contemporaries or authors of other periods) and the fact that his works are the most intensively studied in this field make him an ideal candidate for applying the principles of this new Fuzzy-Logic based computational methodology. At the same time, by building robust discriminating stylistic mechanisms it is also feasible, though as mentioned this is not the main concern in this thesis, to draw some conclusions about the degree of Shakespeareanness in plays that are disputed and, possibly, co-authored.

Overall, with this thesis I propose a new (automated) control framework, a *sui generis* expert system for detecting automatically the degree of Shakespearianness of disputed or anonymous plays. To be precise, in this thesis I am dealing with the challenges of stylometry in early modern English by applying the Mamdani-based Fuzzy Logic of Type-1. (In rough terms Type-1 is a unidimensional approach based on a single membership degree of an object in a specific class-set, whereas the so-called Type-2 is two-dimensional and generates a second membership degree of uncertainty for the first membership degree; for our purposes this additional dimension is unnecessary).

1.1.5 Research Problem, Objectives and Aim

Based on the principles of Fuzzy Logic, Set Theory and descriptive statistics, the present research focuses on the relationship between specific input variables as stylistic markers and aims at estimating an output variable which I call in our experimentation Stylistic index of Similarity (SiS). These input variables are the counts of several sets of words (six sets in the primary and four sets in the core experimentation) and an index of counts of (respectively 102 and 100) words' frequencies between the disputed or anonymous play and the average or ideal document of a group of well-attributed Shakespearian plays of the same genre as the play under scrutiny. The data points of the experimentation of the thesis are extracted from 27 well-attributed, sole-authored Shakespearian plays and the plays being tested. After the gathering of the data points of the 27 plays (see Table 1 on page 74), three

models of Shakespearian style are built, one each for comedy, history and tragedy, through the design of mathematical functions. These are represented by geometrical shapes (trapezoids and triangles) that model the relationships between the various features of writing that are counted and the membership (on a scale of 0 to 1) of the class of Shakespearianness. Using an algorithm to form a series of rules through an iterative process, an inference mechanism is built and the production of the output (SiS) is controlled, depending of course on the input of the data points of the new play.

The current thesis has various and deliberately overlapping objectives. A primary objective is to explore the statistical similarities occurring from sets of related words from the Shakespearian works. So, there is a tendency for selecting words which can be grouped together and constitute a set with a semantic attribute. From this point of view, it can be claimed that the objective is to explore internally ‘movable’ n-sems instead of static n-grams. With the term n-sems it is denoted that I am interested in the counts of semantic units-sets grouped in a complex of adjacent units and, for instance, in the core experimentation with the numerical analysis of four sets, I am looking for the counts of semantic tetra-sets by analogy to the tetra-grams. The interesting element of that approach is that it is possible to detect if some of the counts of the semantic sets of the well-attributed Shakespearian plays converge quantitatively whereas a, rather small, number of their contained words have very dissimilar frequencies. (This, in general, is something that applies mainly for words that are personal pronouns). The potentiality of internal counts-based permutation in each case-play (the ranking of words’ frequencies inside a set) diversifies the elements of our analysis, the sets, and that is the reason I call them internally ‘movable’ n-sems. So, I measure the counts of sets each of which has words situated on a common semantic space, like space-time for Set One, with one, two or even all of its three axes, thus semantic, functional and thematic (Osgood, Suci, and Tannebaum 1957). The second objective is to apply the techniques of Fuzzy Logic and the principles of Set Theory in order to develop a novel method for assessing stylistic similarities between an author’s known writing and anonymous, disputed and co-authored plays. The third and generic objective is to propose a framework for detecting automatically the degree of anonymous or disputed plays’ Shakespearianness and indirectly the authorship share(s) in cases of co-authorship.

1.1.6 An Outline of the Order of Information in the Thesis.

Section One of Chapter One (1.1) introduces the principles of Set theory and the fuzzy methodology of control systems, known as Mamdani-Fuzzy inference, and there is also a description of the scope of the thesis. Section Two of Chapter One (1.2) reviews the literature and the methods of authorship attribution and style discrimination in the period 1700-2000. Section Three (1.3) of the same Chapter probes into the computational, mainly machine learning based, techniques of authorship attribution and stylistic detection developed during the period of 2002-2018. Chapter Two presents this thesis's new method and describes the methodological steps of the pre-processing and processing stage regarding the criteria and method of the selection of the stylistic markers (words, sets of words) and the building of a corpus of 27 well-attributed, sole-authored Shakespearian plays. In addition, this chapter deals with some preliminary parametric concepts of the inference mechanism that acts as an expert system, whose database and rules are derived by applying a series of formulated steps (algorithms). In Chapter Three there is a presentation of the method of the descriptive representation of the collected data and a justification of its usefulness for the design of the parameters and classes of the Fuzzy Model. Furthermore, Chapter Three includes the Component Design Specification (CDS) of the built model, the Shakespearian stylistic classifier, and explains the design and building of the Fuzzy Model, the stylistic classifier of Shakespearianness. In the same chapter, apart from the descriptive statistical analysis of the extracted textual data, there is an explanation of the parameters of the experimentation and of the stages of the design of the components of the fuzzy system stating at the same time the specific objectives that these stages collectively or individually serve. Overall, Chapter Three gives an account of the delicate features of fuzzy modelling such as the fuzzy numerical representation of objects' attributes, the representation of counts of sets of words and words' frequencies, what I call an index of cosine similarity, and the defuzzification process which provides an output of a degree of stylistic similarity. The next two chapters (Chapters Four and Five) constitute two different methods of Fuzzy-Logic-based experimentation. In Chapter Four, there is the core experimentation (Mamdani-Type-1 inference) aiming at assessing the technical validity of the building of the three fuzzy models of the corpus of the well-attributed plays of Shakespeare. In addition, there is experimentation with disputed or co-authored plays and the culled results are combined with the general evidence of historical scholarship. The results from the primary experimentation (Chapter Three and Section 7.2) and, particularly, Chapter Four provide an evaluation of the overall functionality of the fuzzy model-classifiers, the algorithms applied for the formation of the rules and the production of

the measurement of Shakespearianness of new plays based on the modelled data points of the corpus of the well-attributed, sole-authored Shakespearian plays.

These two chapters, Three and Four, are the basis for the building in Chapter Five of a Sugeno-ANFIS (Adaptive Neuro-Fuzzy Inference System) and the assessment of the robustness and limitations of the application of the automated approximate reasoning-based fuzzy techniques in the traditional field of literature and early modern English plays. The Chapter of Conclusions (Chapter Six) explains how this thesis manages to bridge the current gap in the stylometric area exploiting technicalities of expert automation and fuzzy reasoning. Moreover, some implications are made in that concluding chapter about how researchers can build upon such automated stylistic assessors and how we might strive toward an automated universal stylistic classifier for any kind of written text. There are also two Textual Addenda, the first of which contains the list of the versions of the plays (Section 7.1) of the core experimentation, as well as a detailed Technical Appendix (Section 7.3) that clarifies technical terms and concepts of Fuzzy Logic.

1.2 Authorship Attribution and Style Discrimination: The 18th, 19th, and 20th Centuries.

This review tells the story of authorship attribution from the late eighteenth century to the end of the twentieth century and chronologically assesses the stylometric methods that were used. There is a theoretical apparatus to be assessed alongside the practical applications, too. As for texts, our focus is on efforts to attribute the plays of Shakespeare and his period, but space is also afforded to the Federalist Papers, essays written by Alexander Hamilton, John Jay, and James Maddison in defence of the constitution of the United States of America.

1.2.1 Discriminators of Verses, Lines Ending and ‘Peculiar Inaccuracies’: Edmond Malone and James Boswell, William Spalding. (Circa 1750-1850)

At the end of eighteenth century, Edmond Malone (1704-1774) discussed the quantitative metrical discrimination-diversification of verses in Shakespeare’s authorial style. Assessing internal evidence, he examined the provenance of the three plays of *Henry VI* in the First Folio of 1623. In fact, he argued that *Part 1* of *Henry VI*, which does not exist in quarto, was based on a previous, non-extant, play by an unknown dramatist that Shakespeare took over and adapted. He also came to similar conclusions about *Parts 2* and *3* of *Henry VI*.

In addition, Malone investigated critically the Shakespearian thematic preferences and the metrical system in the three plays of *Henry VI* in comparison with similar elements of the undisputed works of Shakespeare. His comparative analysis in his dissertation *on Henry VI* (Malone 1787) tended to indicate, if not prove, that a large portion, but not the whole, of *Parts 2 and 3* of *Henry VI* were written by Shakespeare. At the same time Malone brought up many differences in the way verses were formed or phraseology was composed. In this context, there was reference to the ‘peculiar inaccuracies’ and parallel ‘peculiarities’ of the Shakespearian ‘phraseology’, plus irregular transpositions and repetitions, in *Parts 2 and 3* in the Folio. These characteristics suggest that we cannot credit Shakespeare with the full authorship for *Parts 2 and 3* of *Henry VI*. In addition, there were unexpected metrical discrepancies in *Part 1*, thus the metrical system is dissimilar to the Shakespearian poetic style. Such discrepancies detected in *Part 1* of *Henry VI* were the ‘lack of alternate rhymes’ and the ‘very few dispersed rhymes’ (Malone 1787, 16) in comparison with the undisputed plays of Shakespeare. According to Malone, this was a strong indication that the first of these three parts of *Henry VI* should not be attributed to Shakespeare alone. Malone also observed that in the first part of *Henry VI* the lines end usually with a pause, which was not a Shakespearian feature. He noticed in *Part One* a pattern of classical allusions that was not congruent with Shakespeare’s style, and which does not exist in the other two parts of the play (Malone 1821a, XVIII:4–5).

In the preface to another book by Malone (Malone 1821b, I:v–xlix), James Boswell (1778–1822) referred to the general terms of stylistic analysis, such as the ‘versification of the poet’. Boswell dealt with versification in relation to the reality of corrupted texts, presumably in the form of copies of the original manuscript(s), and the tendency of commentators to follow Alexander Pope’s (1688–1744) rigid criteria of intervention. In that sense, Boswell’s concern, similar to that of George Steevens (1736–1800), was to discern the original style of a play’s versification from the inappropriate interventions, thus the versification that could have been technically ‘lengthened or curtailed’ altering the original traits of of an author’s play. Furthermore, Boswell did not propose systematic stylometric techniques, but he offered general guidelines and highlighted the difficulty of applying objective methods for the discrimination of authorship. In this context, he criticised those who were investigating authorship and style simply by exploring unilaterally the metrical system of verse lines or single peculiarities, such as the plural substantive of singular verbs. In fact, Boswell implied, narrow definitions of style lead to wrong conclusions since, as

Malone showed, in certain cases, deviations of the phraseology and of the known metrical system of Shakespeare might be simply circumstantial mishaps, printing errors or products of carelessness.

A few years after the work of Malone, was published the book titled *A Letter on Shakespeare's Authorship of 'The Two Noble Kinsmen': A Drama Commonly Ascribed to John Fletcher* (Spalding 1833). In this book the analysis of William Spalding epitomised in a few sentences the authorial portraits of Shakespeare and Fletcher (Spalding 1833, 11–12). Since, as the title page of the first edition in 1634 tells us, the play's authors were Shakespeare and Fletcher, the detection of Shakespeare's peculiarities in any part would imply the lack of Fletcher's pen and vice versa. According to Spalding, 'Shakespeare's versification is broken and full of pauses' (Spalding 1833, 11). This statement is an implication for the frequent use of punctuation by Shakespeare, for instance a comma or full stop, in the middle of the verse line, avoiding the strict monotonic identification of the end of sentences with the end of lines. It seems that compared to Fletcher Shakespeare had no strong preference for double or 'feminine' line-endings (Spalding 1833, 11-12). This kind of line-endings involves the addition of an eleventh unstressed syllable to a regular line of iambic pentameter. (Iamb or else Iambus constitutes in metrical terms a foot in which an unstressed syllable is followed by a stressed syllable. Consequently, an iambic pentameter consists of ten syllables or else five iambs.) Furthermore, Shakespeare's writing style is characterised in general by abruptness and the sudden ending of scenes and speeches. On the other hand, Fletcher's style, according to Spalding, is smoother, his lines are more complete in having no sudden pauses, and he employed double endings more frequently than Shakespeare (Spalding 1833, 12). Spalding in the rest of the book reversed his initial disadvantageous description and explained that Shakespeare's style and abruptness is precisely what ensured passion and dramatic vehemence, two elements essential for *The Two Noble Kinsmen*. Consequently, starting from a quantitative approach, Spalding posed a qualitative hyper-conclusion, emphasising the need to study the lack of metrical identities in connection with the parallel existence of stylistically dramatic and creative features in the sense that a metrical deviation can be the result of the intention of an author to change (momentarily) his style for reasons of dramatisation.

1.2.2 Middle and Late nineteenth Century (circa 1850-1895): Proportionality in Co-authorship Attribution & The Debates of the New Shakspeare Society: James Spedding, Frederick James Furnivall.

In the middle of nineteenth century there was a boom of authors investigating in a more systematic way Shakespearian works and authorship issues with certain particularities, starting with *The Two Noble Kinsmen* and *Henry VIII or All is True*. *The Two Noble Kinsmen* was likely written around 1613-14 and was printed as a quarto in 1634 though it does not appear in the First Folio (1623). *Henry VIII* is assumed to have been written in 1613 and was printed later in the First Folio. The acceptance that *Henry VIII* was co-authored started to trigger similar thoughts for other works, such as *The Taming of the Shrew* and *Pericles*, the second of which was printed in quarto in 1609 and omitted from the Folio. The most important researchers of this period were James Spedding (1808-1881), William Spalding (1809-1859), Frederick James Furnivall (1825-1910) --a co-creator of *Oxford English Dictionary*--Samuel Hickson (no available date), Edwin A. Abbot (1828-1916), Frederick Gard Fleay (1831-1909), John Kells Ingram (1823-1907), and Walter William Skeat (1835-1912).

Furnivall was the founder of the New Shakspeare Society (1873-1894), a union of scholars interested in dealing in a systematic way with the correct chronological classification of the theatrical works of the early modern English drama and complex issues of authorship attribution and verification (Britannica, n.d.; Furnivall et al. 1874).

Spedding's article on the authorship of *Henry VIII* appeared in August 1850 in *The Gentleman's Magazine* and was reprinted later in the *New Shakspeare Society* transactions. Spedding's article paid attention to the so-called metrical test of the masculine and feminine endings, of the extra stressed syllable at the end of the verse line and the use of 'stopped' or 'unstopped' (that is, enjambed) lines in order to assign different parts of *Henry VIII* to Shakespeare and Fletcher. Spedding, exploring the metrical similarities between *Cymbeline* and *The Winter's Tale*, intended to find the average rate of use of redundant syllables in verse lines (Rolfe 1884, Appendix, 14) and to compare them to the rates found in *Henry VIII*. Spedding considered the possible participation of additional writers apart from Shakespeare and Fletcher, stressing the fact that 'at least two different hands had been employed in the composition of *Henry VIII*; if not three' (Spedding 1874, 6-7). Spedding's conviction was based mainly on intuitive, nuance-based and emotional criteria. According to his arguments, though an attribution of exact portions to a third hand could not be verified, the whole work

seemed to have been written by different hands without any trace of parallel collaboration and is characterised by a plethoric inconsistency of style, which can be conceived mainly through ‘the general effect produced on the mind, [and] the ear...’ (Spedding 1874, 6–7).

Hickson discussed the co-authorship attribution with two articles (1847, 1850), which were later reprinted in the *New Shakspeare Society* transactions: the first (Hickson 1847) about *The Two Noble Kinsmen* and the second Shakespeare’s *Henry VIII*. Regarding *Henry VIII*, Hickson assessed the previous contribution of Spedding, who did not accept the idea of its sole authorship by Shakespeare, and offered Fletcher as the most probable co-author. Fleay later made clear that he independently arrived at the same conclusions and likewise attributed the works to Fletcher and Shakespeare. In 1847 Hickson in his review of William Spalding’s aforementioned letter on *The Two Noble Kinsmen* (Hickson 1847), reprinted in the First Part of *Shakspeare Society’s Transactions* for 1874, agreed that this was a work composed collaboratively by Fletcher and Shakespeare. Hickson and Fleay focused on the frequencies of the elements of the authors’ metrical system, such as the ‘stopped’ or ‘unstopped’ lines, feminine endings, and the presence or absence of rhymes.

1.2.3 Emergence of a Mechanic Stylometric Approach: Mendenhall and his Curves-Based Discrimination of Authorial Styles. (Circa 1895-1905)

Thomas Corwin Mendenhall (1841-1924) devised a ground-breaking statistical stylometric approach that looked methodically for length-based classes of words and thus presented his results using histograms and graphical curves. In his two articles, ‘The Characteristic Curves of Composition’ (Mendenhall 1897) and ‘A Mechanical Solution of a Literary Problem’ (1901), Mendenhall supposed that the lengths of words, measured in letters, and the frequencies of the occurrences of the resulting classes, that is the class of all one-letter words, all two-letter words, all three-letter words and so forth, can be indicative of writing style. The discrete points in a histogram showing the frequency of each class of words can be connected to form graphical curves and the differences in the styles of authors can be represented by these curves.

Mendenhall explored mainly the works of Charles Dickens and William Thackeray and the first histograms were from passages of 1,000 words from *Oliver Twist* and *Vanity Fair*. Averaged across their works, Mendenhall found consistently common curves for each author and certain similarities in their writing style, as expressed by these curves. Mendenhall considered two passages of 5,000 words from each of *Political Economy* and *Essay on*

Liberty by the economist John Stuart Mill and the result was provocative. Contrary to Mendenhall's general findings of the predominance of three-letter words in Dickens's and Thackeray's works, in Mill's works the most frequently used were the two-letter words (Mendenhall 1897). Despite this, Mendenhall found that there was also an actual predominance of frequencies of lengthier words in the political documents, and he attributed this difference of frequencies to the nature of the political documents and the liberal way prepositions are employed and sentences are formed in that genre of more verbose writing (Mendenhall 1897). That is, genre explained the unusual result that two-letter words and much longer words were more frequent in Mill's writing than in Dickens's and Thackeray's.

In his second study titled 'A Mechanical Solution of a Literary Problem' (Mendenhall 1901), Mendenhall again explored the length of words and the resulting word-class frequency distributions, this time in *Vanity Fair* by William Thackeray. Now, he paid more attention to the analysis of the two consecutive passages of 1,000 words looking for specific similarities in the curves of one author (Dickens) to see how consistently the authorial style emerged across the writing. Mendenhall proceeded to analyse two groups of 200,000 words, 400,000 in total, using the works of Shakespeare and irrespective of the genre. Mendenhall compared Shakespeare's style with that of Francis Bacon (using another two sets of 200,000 words each). In the same manner with the analysis of the two groups of words of the Shakespearian and Baconian works, Mendenhall also analysed two groups of about 75,000 words each from Ben Jonson's works (poetry and prose), and from other writers of the Elizabethan period, among whom were Christopher Marlowe. Mendenhall aimed to extract general results about the predominance of the class of three-letter words. He found that Bacon's curve was dissimilar to the Shakespearian and that Shakespearian style was characterised by the predominance of four-letter words in both selected groups (this stood in contrast to the predominance of three-letter words in other writers' canons). Mendenhall detected strong similarities between Marlovian and Shakespearian writing styles regarding the features he measured. In critical terms, Mendenhall's methods constitute historically the first systematic stylometric effort, though a more focused analysis could now be applied to the same evidence by re-sampling and cross-validation.

1.2.4 Theoretical Foundations of the Beginning of the Twentieth Century (circa 1905-1931) and External (Henslowe's Diary) and Internal Stylometric Evidence: W.W.Greg, H. Dugdale Sykes and Charles A. Langworthy.

The fact that early modern plays were written for actors to perform requires investigators of writing style to pay attention to the conditions and practices of the theatre industry, and at the start of the twentieth century two books by W. W. Greg (1907, 1908) made widely available the evidence held in the cache of documents that the theatre impresario Philip Henslowe left us. A practical application of this knowledge appears in the essays he edited on the subject of *Shakespeare's Hand in the Play of 'Sir Thomas More'* (Greg 1923). The main objective of the book was to explore the manuscript (British Library Harley 7368) of the play of *Sir Thomas More*, about that Catholic martyr and saint (1478-1535), and the claim that it contained three pages of Shakespearian dramatic writing in his own hand (designated 'Hand D'). The handwriting of those pages was compared to the handwriting of the six surviving signatures of Shakespeare and found to be the same. The book's contributors (A. W. Pollard, W. W. Greg, J. Dover Wilson, E. Maunde Thompson and R. W. Chambers) shed light on the Shakespearian authorship in relation to the play's literary-dramatic style, too, using persuasive, albeit subjective, illustrations of their thematic and verbal likenesses.

In the same year, H. Dugdale Sykes started his book titled *Sidelights on Elizabethan Drama: A Series of Studies Dealing with the Authorship of Sixteenth and Seventeenth Century Plays* by investigating *Timon of Athens* (Sykes 1923, B-48), exploring its various authorship contingencies with the two basic alternative assumptions that parts of the First Folio version are by 'an inferior hand' writing at the same time (Sykes 1923, B-48) and that Shakespeare took over and revised parts of a previously existent play. (Here, 'hand' of course means 'writer' rather than 'handwriting' as it does in relation to an extant manuscript.) Sykes laid emphasis on common doubts about the penultimate scene of *Timon of Athens* and its non-Shakespearian character. Sykes developed his stylometric criticism based on features such as verbal quibbles, which according to Sykes were used by Shakespeare in a concrete and more substantial way (Sykes 1923, B-48) than is found in the work of other Elizabethan writers, who rather used them as an ornament or static element of dialogue. For instance, in *Timon of Athens* in a dialogue between Apemantus and First Lord the use of the word 'time': 'F. L: What time ... is it?/ Ap.: Time to be honest. / FL.: That time serves still' (Sykes 1923, B-48). In other cases, Sykes sought words' frequencies and specific phrases, as for instance 'much ado' or 'such ado' found in *Famous Victories of Henry V* and *The Taming of a Shrew*. By detecting common phraseological patterns and classical allusions Sykes claimed that the

anonymous play of *The Taming of a Shrew* has stylistic similarities with *Famous Victories*. Sykes concluded that the first play was of dual authorship and that one of these two authors had also written *The Taming of a Shrew*. Sykes's survey led to the dramatist Samuel Rowley. But the gain of Sykes's investigation is mainly his methodological systematicity, as he insisted on the importance of comparing stylistic features between the plays under scrutiny only under the same circumstances. That is, he was searching for common lexical or phrasal patterns assessing also the similarities-generality of their linguistic and theatrical, parts of scenes or acts, contexts. In this sense, Sykes claimed that simple imitations of an author's style or random or sporadic uses of patterns by another author should not be counted as consistent elements of an authorial style.

At the same period, Charles A. Langworthy wrote an analytical research paper titled 'A Verse-Sentence Analysis of Shakespeare's Plays' (Langworthy 1931) which carried out a methodological investigation in Shakespearian plays of the properties of verse-sentence correlation, meaning the alignment of the end of a sentence with the end of a verse line. According to Langworthy, it was important to evaluate if this poetic habit was randomly and unintentionally applied or whether it was intentional and systematic.

The connection between verse lines and sentences, argued Langworthy, can be investigated in a broader way concerning the full triptych of grammatical units plus clauses and sentences. His results indicated that in early Shakespearian works there was a statistically significant correlation between the endings of verse lines and sentences--that is, they ended together--whereas in the late works independent clauses were more likely to start from a new line. In late Shakespeare sentences tended to obtain autonomy in relation to verse lines. In general, this kind of causality-based stylometric analysis, argued Langworthy, could be a useful tool of clarifying authorship issues in connection with the Shakespearian canon, too.

1.2.5 Zipf's First Systematic Rank-Correlation/Frequency Law (1932) and Its Modern Application.

From the mid-eighteenth century until Langworthy (1931), stylometric efforts were mainly concentrated on resolving authorship attribution by evaluating historical evidence and detecting quantitative internal evidence. This quantitative detection of discriminators had until that period an unsophisticated character, as investigations of authorial styles and similarities were performed based on a series of very basic mathematical operations applied without any kind of systematic classification of collections of infrequent terms, such as

lexical words, and frequent terms, such as function words. The 1930s mark the transition from the period of simple mechanic stylometric studies of early modern drama to stylometric initiatives that used mathematical and statistical analysis for the systematic study of linguistic traits. For the first time, words' frequencies began to be theorised in a complex and correlational way.

The change came with George K. Zipf's *Selected Studies of the Principle of Relative Frequency in Language* (Zipf 1932), a short book (just 51 pages) divided in two parts. The first part was concerned with the phonetics of Indo-European languages and the second with the frequencies of words of works in Latin, English and Chinese language and included a short reflection on semantics. Zipf introduced his 'Principle of Relative Frequency' and he laid emphasis on the term of conspicuousness, that is to say the fact that 'The accent or degree of conspicuousness of any word, syllable, or sound is inversely proportionate to the relative frequency of that word, syllable, or sound' (Zipf 1932, 1).

Most important was the formation of his Zipf rank-correlation/frequency law, in which the frequency of the most frequent word corresponds roughly to double the frequency of the second most frequent word, to triple the frequency of the third most frequent word and so on. That is, when words are ranked by their frequencies, the frequency of each word after the first (relative to the frequency of the first) is just 1 divided by its rank position. This remarkable property holds true for any sizeable document in any language. Zipf postulated an inverse frequency distribution correlation with an exponential potential. It is interesting, as it shows Zipf's theorem's impact and diachronic practical usefulness, that Zipf's primary correlative ranking law of frequency has been further widely developed with logarithmic representation, the so-called Riemman adjustments and constants additions.

G. Udny Yule (Yule 1944), whose theory of ratios will be explained below (1.2.7), and Zipf attempted to relate practically the most frequent word of a text with the next most frequent word and then the next most frequent, in a ranking order. Though Yule probes into the numerical correlation of the relative frequencies with the total number of the words in the text, Zipf's contribution to statistical analysis of language in general is more complex as along with his principle of least effort (Zipf 1949), whose interpretation is that 'people try to find an equilibrium between uniformity and diversity in the use of words', he investigated the dynamics of philology and concerned himself with the psychological, biological and social exploration of the statistics of language (Zipf 1935). Zipf's law applies not only to words but has been accepted as an empirical truth in economics and other social sciences, whether it concerns the populations of cities--the largest compared to the second largest and the third

largest and so on--or economic growth distribution (Gabaix 1999). It is not clear, nor can it really be defined, what constitutes the minimum amount of data to which Zipf's law applies. Would a single sonnet or a paragraph of an essay be liable to this kind of ranking order? What should be the required minimum amount of words or, in social science, the sample of cities whose populations obey this law (Gabaix and Ioannides 2003, 7)? Exceptions to Zipf's law have been found, as for example is the case of Japan's cities' populations. On the other hand, this is the first integral method of evaluating the internal quantitative correlation of all words' occurrences inside a text with emphasis on nouns. Summarizing Zipf's theory, it is built on an exponent-based inverse ranking classification principle that conveys that each word of a text is ranked proportionally to the total of its own and the other words' frequencies.

1.2.6 Collaboration and Further Theoretical Stylometric Clues by Muriel St. Claire Byrne (1932).

Almost simultaneously with Zipf, Muriel Byrne with her article 'Bibliographical Clues in Collaborate Plays' (Byrne 1932) dealt also with attribution of collaborative and joint authorship. Byrne's arguments revolved around two anonymously published plays, namely *The Downfall of Robert Earl of Huntington* and *The Death of Robert Earl of Huntington*. Byrne assessed the information of Henslowe's diary about the date and authorship of these works, associating Antony Munday with the former (*The Downfall*) and Henry Chettle with the latter (*The Death*). Byrne also referred to the authorship of the play *Robin Hood*, an assumed mixture of original work, revisions and alterations. *Robin Hood* along with *The Death of Robert Earl of Huntington* as works of joint authorship, argued Byrne, cannot be strictly theorised as sources of stylometric evidence of any single author. Byrne's major principle was that stylometric investigation should begin with the known sole-authored works and then move to the anonymous or collaborative works. At the same time, when investigation of authorship takes place, 'a negative check' should be applied to make sure that the features found to be in common between the text to be attributed and the body of writing by a candidate author are not also frequently found in other writers' work. That is, to establish authorship it is not enough to find features in common: one must find features that are common to the work to be attributed and to the candidate's works and that are highly rare in, or absent from, everybody else's works.

1.2.7 Yule's Theorems (1944): Words at Risk, Ratios of Vocabularies and Size of Texts.

In a different spirit to Zipf, G. Udny Yule (1871-1951) in his book *The Statistical Study of Literary Vocabulary* (Yule 1944) investigated first the nature of frequency distribution, and, building a conceptual framework of principle factors affecting stylistic differentiation, he analysed the 42,000 words of *De Imitatione Christi* compiled from the concordance of Rayner Storr (1835-1917) and widely believed to be by Thomas a Kempis (1380-1471). Yule criticised the fallacy that if there is a total of 2,000 nouns in a sample of writing and the most frequent occurs 50 times, then in a sample containing 4,000 nouns we will find around 100 occurrences of that noun. In practice, the increase of a text's size is not always followed by a proportional increase of entry rate of the new lexical words. There is not a linear correlation between the size of the text and the total numbers of occurrences of the nouns in it. In fact, there are critical points where this correlation is broken, especially as the text grows and reaches a specific length, and other critical points where again this correlation re-emerges and consequently there is a decreasing rate of new token entries. Of course, various theories and models have been adopted by linguists to account for this. For instance, according to Ye-Sho Chen and Ferdinand F. Leimkuhler textual elements such as text generation, sentence structure, vocabulary size, rank frequency relations and type-token relations can be analysed through stochastic modelling (Chen and Leimkuhler 1989, 45–46). This kind of modelling makes a prediction (for example of the rate of vocabulary's growth) based on the current state and the variables that affect the transition to a specific next state. In this category of modelling fall also the so-called Yule and Simon-based models that support a nearly linear theorisation of the way the text and vocabulary are simultaneously produced (Chen and Leimkuhler 1989, 47–50). Chen and Leimkuhler exploited the previous work of Gustav Herdan who first formulated statistical linguistic laws and showed that, in relation to the size of text, the growth rate of the vocabulary is a poly-parametric problem and is dependent on various factors such as the specific size of the text, its already accumulated vocabulary, and even the conditions of writing, such as genre and content (Herdan 1960).

Critical terms in Yule's approach are 'tokens' and 'types': the former relates to words counted once for each time they occur and the latter refers to the singular occurrence of distinct words. Thus, in the sentence 'there are two types and two tokens', the word 'two' is a type with two tokens because it appears twice; all the other types are represented by one token. Yule discussed the significance of the ratio of tokens and types.

Yule decided to analyse only counts of nouns' occurrences and he made comparative analysis of vocabularies of other works of Thomas a Kempis along with other less plausible candidates for *Imitatio*'s authors, such as Jean Gerson (1363-1429). Analysing the number of nouns in *De Imitatione Christi*, Yule drew a table of the counts of nouns' frequencies in relation to the types of the total vocabulary and tokens in comparison with the total number of tokens (Yule 1944, 10–11). In a multi-column table, he recorded the proportions of types and tokens on a scale, starting with the 520 nouns that occur only once to the single noun occurring most often, with 418 occurrences. In general, his token-type ratio analysis brought up the way the words' frequencies were distributed, and Yule claimed that 'the most frequent noun constitutes by itself less than one-tenth of 1 percent' of the total vocabulary as a type 'but contributes to over 5 percent of all occurrences' of a text (Yule 1944, 11). According to that claim of Yule, it can be deduced that the numerical evaluation of the tokens used by an author show us the habit of employing tokens of specific nouns in a more consistent and frequent manner than other nouns. At the same time the proportion of a type to the total vocabulary differs from the respective proportions of the occurrences of its tokens to those total occurrences in the text. Yule showed that this differentiation follows a certain orbit starting from the least frequent to the most frequent types, and as the size of the text-vocabulary is reduced there is a gradual deterioration of the correlation between the proportion of types-to-total-vocabulary and the proportion to tokens-to-total-text-size. Thus, the most frequent words occupy as tokens a far larger part of the total occurrences-tokens of a text than the less frequent words, though as types they constitute a much smaller part of the total vocabulary. In effect, it seems that authors have certain preferences towards a specific pool of words from the vocabulary of the language. This familiarisation with a certain pool of words makes it easier for the writer to use them more frequently and embodies Zipf's principle of the least effort. Nevertheless, as the sample gets larger, the increase in the use of the words that previously were most frequent does not keep pace with the increase in the size of the sample. As the sample grows, writers tend to find new words from their available--brain registered--vocabulary, which Yule idiomatically called the 'words at risk'.

Another interesting point according to Yule's findings and limit theorem is that as the small initial textual sample gets larger, there is a significant increase of the counts of the most frequent words of his vocabulary but not proportionally. In fact, after various critical points, and at the end of the full work, there is a significant reduction of this proportionality. In practice, Yule's experimentation initiated a theory of local maxima and minima of words' frequencies.

Yule also suggested that the full vocabulary of the writer--the total of nouns of all the writers' works--should be the base of the analysis. He inclined more towards nouns, even though he stated that other parts of speech, such as adjectives or adverbs, could be similarly analysed. Yule advised that function words, such as determiners should be excluded as non-significant.

In conclusion, G. Udny Yule drew special attention to the role of the size of texts in relation to the richness of vocabulary. He propounded that the formation of literary works is carried out based on the author's pool of 'words at risk', namely the available personal vocabulary. Thus, Yule was not simply counting words: he was using his findings to propose the outlines of a theory of creativity.

1.2.8 Probabilities-Based Emerging Stylometric Technicalities (1963): Investigating the Federalist Essays through Bayesian Inference (Frederick Mosteller and David L. Wallace).

The Federalists papers are a run of anonymous newspaper essays published in late eighteenth-century America with the intention of persuading readers that the newly written Constitution of the United States should be ratified by the states. The identities of the papers' writers--John Jay, Alexander Hamilton, and James Madison--were guessed at during publication, but the attribution of each to its correct author was incomplete even after contemporary records of authorship were disclosed in the early nineteenth century, and the puzzle was not solved until the twentieth century and using analysis of internal evidence of literary style. Jacobs E. Cooke (Cooke 1961) adopted mainly a theoretical approach investigating mostly the intention and character of the writers and describing the Federalists' essays as a venture that stood out as the major contribution of Americans to political philosophy. On the other hand, in the first part of the book titled as *Inference in an Authorship Problem* (Mosteller and Wallace 1963), there was a more practical stance. Mathematician Thomas Bayes's theorem was applied to the problem of distinguishing Hamilton's writing from Madison's, since the attribution of papers to Jay had already been solved by external evidence.

Key terms in Bayes's theorem are 'prior' and 'posterior' probabilities. Bayes's theorem as an inference method is useful because it assists in making predictions, when we have only partial data and lack necessary information that would assist us in predicting an event. The partial data can be previous observations and measurements, or in some cases even empirical or subjective beliefs, that can be considered. The combination of initial

beliefs, and their corresponding ‘prior’ probabilities, with the findings from the theorem leads to the ‘posterior’ probability and the updating of our initial subjective or empirical beliefs. In other words, it is feasible to exploit certain known parameters and avoid a blind or subjective prediction by assessing statistics from what is already known, and in the case of authorship attribution this can be the stylistic markers of the authorial styles. Posterior probability is formed by the product of the prior evidence and the likelihood ratios, being divided by an index of constant probabilities. So, posterior probability equals the revision of the prior belief by exploiting new informational events. Of course, in a series of predictions each piece of posterior evidence can be used at the next stage of prediction as a prior belief for the next decision. Various more and less complex formulae can be applied in the framework of Bayesian inference, but the main point is that it is a methodology that builds on the belief of previous available evidence before the new event takes place and then tells us how much to revise our beliefs in the light of that new event.

Mosteller and Wallace started from the initial probabilities for each candidate’s authorship of a work, where were typically assigned arbitrary values such as 0.5 for Hamilton and 0.5 for Madison. Bayes’s theorem is primarily concerned with the degree to which we should update such probabilities in the light of new evidence: it gives an updating function to be multiplied by the prior probabilities arising as the posterior detected frequencies of specific words in a preceding round of counting. In the logic of Mosteller and Wallace, a large deviation of the findings of the disputed text from the average frequencies, expected ‘usages’, would be penalised accordingly and occurrences very similar to the previous observed ‘usages’ of the undisputed texts would be enhanced probabilistically. The kind of sequential probabilistic updating for a pool of certain words led to the final probabilities respectively of the authorship of Hamilton and Madison for the 12 disputed political essays.

Thus, in a worked example (Mosteller and Wallace 1963, 289–90), if the initial authorship belief in favour of Hamilton’s authorship of a paper was 0.999 and was 0.001 for Madison’s authorship, this initial probability for Hamilton should be updated by being multiplied by the division of the likelihood ratios of Hamilton by Madison’s for the found frequency, say four usages of the specific word. The probabilities of Madison would be the residual from the subsequent subtraction. Accordingly, the initial belief of 0.999 for the authorship of Hamilton could be weakened if the found frequency, ‘usage’, of a specific word was rarely observed in his undisputed works. In the event, the initial probability of his authorship might be updated to less than 40% on account of our repeatedly finding words he

rarely used in his undisputed works. The same applied for the sequential processing for the rest of selected words.

Whereas in the previous approach of Paul L. Ford and Edward G. Bourne (Ford and Bourne 1897), regarding the 12 disputed essays, each one from 900 to 3,500 words in length, it seemed that Hamilton was the author, Mosteller and Wallace believed that Madison wrote most of the disputed essays. Mosteller and Wallace provided an analysis of the results in the form of what are called t- and p-values and confidence limits for logarithmic metrics. The ultimate goals of Mosteller and Wallace were to proceed to a simultaneous bi-direction analysis, namely to find the probability of Madison having written each essay and of Hamilton having written it, and to compare the results identifying the difference of these descriptive data of the two candidate authors. T-values express the difference of the variation of the mean values of two groups of data. In our case these groups are formed by the two respective groups containing the number of occurrences of specific words of the disputed texts in comparison with the expected number arising from actual occurrences of those words in the undisputed corpus. P-values indicate how often the evidence that we found for each candidate's authorship would be found in any case, purely by chance, if they were not the author. In conclusion, Mosteller and Wallace claimed Madison's authorship for 10 out of the 12 disputed essays (numbers 49-58, 61, 62) from a total of the initial 77, not including the eight essays that were later added, of which 65 were undisputed.

Mosteller and Wallace attempted to find correlations between the frequencies of occurrences of the words in the known works of the candidates, Hamilton and Madison, and other features common to their known writing and the disputed essays. Mosteller and Wallace searched for words with stable frequency behaviour, namely consistency of frequencies in different works by the same authors. They looked for words (such as 'may', 'his', 'any') whose counts had statistical significance and eliminated words whose frequencies were about the same across works by different writers, or were depended on contextuality (such as 'war', 'executive' and 'law') or were close to zero. (The researchers describe their methodology and the corpora they employed but not the specific words they eliminated). Their aim was to explore the standardised difference of the means of these words' frequencies. The measured words were put into six classes (Mosteller and Wallace 1963, 289–90) according 'to the apparent degree of contextuality danger' (Mosteller and Wallace 1963, 289–90) by rather intuitive criteria, such as their stable and medium contextuality, their grammatical categorisation, or their high and low frequency occurrence rate. For instance, there were

selected for classification verbs and pronouns categorised as auxiliary, and prepositions such as ‘upon’ which is neither completely a function word devoid of lexical value nor a word whose meaning is affected significantly by its close context. This aspect of Mosteller and Wallace’s work was not as ground-breaking as their demonstration that the frequencies of occurrence of function words are indeed a powerful discriminator of authorial style.

1.2.9 The New Computational Stylistic Research Environment (1969-1985): Bryan Drake, Ronald E. Bee, John R. Allen, Ching Suen.

By the end of the 1960s, the advent of computerised processing emerged gradually as an opportunity for more efficient and dynamic textual analysis. In fact, it was already realised that computers could be used equally ‘in the technical disciplines and in the humanities’ with various goals from ‘machine translation to the determination of authorship’ (Peterson 1969, 9). Along with the advent of the computational textual tools, concerns of equal significance were raised about the ‘Unanswered Questions in Computerised Literary Analysis’. Ronald E. Bee ‘The Use of Statistical Methods in Old Testament Studies’ (Bee 1973) compares passages of Biblical texts with extracts of a long English poem, Edmund Spenser’s *Faerie Queene* Book 1 and Book 10. The following year, John R. Allen adapted Claude Shannon’s notion of textual entropy to assist the detection of the similarities between a sample piece of writing and the known writing style of an author (Allen 1974). As described by Claude Shannon’s article ‘A Mathematical Theory of Communication’ (Shannon 1948), entropy is the average information rate of a piece of writing treated as a sequence of symbols each of which appears at a different frequency rate.

1.2.10 Novel Stylometric Methods (1985-2000): Machine Learning Techniques.

Here I attempt a description and analysis of the most advanced computational features of novel stylometric methods, requiring in certain cases, such as for the Support Vector Machines, a high degree of mathematic knowledge and reasoning, and geometrical representation. There must also be a paradigmatic review of the so-called supervised and unsupervised machine learning techniques, as applied in the stylometric field.

1.2.10.1 Emergence of Neural Networks in Stylometry: Robert Matthews and Thomas Merriam (1993-1994).

In their article ‘Neural Computation in Stylometry II: An application to the works of Shakespeare and Marlowe’ (Matthews and Merriam 1994), Robert Matthews and Thomas Merriam simulated a neural network examining the canons of Shakespeare and Christopher Marlowe. A year before, the authors had written the article titled: ‘Neural Computation in Stylometry I: An Application to the Works of Shakespeare and Fletcher’ (Matthews and Merriam 1993). In both articles, the authors applied neural computation techniques in the stylometric area.

According to Robert Callan, neural networks are ‘parallel computing devices consisting of many interconnected simple processors’ (Callan 1998, 1). These processors work as receivers and senders of signals, which can be viewed as electrical currents of very specific informational content. Processors can be arranged in different ways, depending on the research design and the degree of non-linearity of the problems to be solved, namely the patterns that have to be classified. The interconnection of these processors looks similar to biological networks of neurons but they are not aimed at simulating biological functions: the aim is finding mechanically efficient solutions to classifying patterns. The designer of the neural network selects the topology-structure (the connections of the neurons/processors) and the type of algorithms to be applied to the learning parameters in the initial training phase. During training, each processor’s pattern of behaviour--its responses to its inputs and just what it outputs--is automatically self-adjusted until the entire network of them is making the discriminations that the investigator wants. After the training stage, there is a validating phase with the new input data, and in that way the performance of the training is cross-validated on new input-data not included in the training phase. Finally, there is a testing phase, where new input data are tested.

Neural networks are composed of the processing elements-units that are called nodes, which are interconnected to form the neuron-like structures. Like the nervous systems and the function of biological neural synapses, these artificial neurons are interconnected forming the topology of the neural network. In the biological field, the input layer of a neuron is called a dendrite and the output is an axon. In a similar manner, the topology of a neural network consists of at least three layers: the input layer, where the initial external input is received, at least one hidden layer, and the output layer.

Each layer includes multiple units comprising variables, and each of these variables stores a value. The learning process uses the stored memory of each value, which represents for example the magnitude that the input signal must reach (say, as the total of two

simultaneously received inputs) for that node to respond by presenting a certain output value. The value in each node--the input-threshold to which it responds--is itself modifiable as the entire network 'learns' the desired behaviour. As well as different nodes having different thresholds for action, the connections between them can have different 'weights' so that the signal coming from one node has a greater effect on the action of the next node than does the signal of a lower 'weighting' coming from elsewhere. The values held in the initial input layer are processed through adaptation or learning in conjunction with the type, supervised or unsupervised, of the output layers. A basic element of the training is the neural connection weights of the links of the layered units. These weights in the range of -1 to 1 are adjusted according to the adopted algorithm and the topology of the network. In a 'feedforward' topology each node's output is passed on to nodes further into the network, towards those nearer the output layer. In a 'recurrent' topology, by contrast, the output of a node can be routed backwards to other nodes nearer than itself to the input layer, thereby forming loops of information transmission across the network. In a 'feedforward' topology, information about the output being in error (say, a misclassification of authorship) can be fed into the nodes of the network by a process called back-propagation that calculates an adjustment to each node's behaviour (altering, say, its threshold for action) that will contribute to reducing the error in the overall output. Matthews and Merriam used a feedforward multi-layer network and applied the backward propagation algorithm as an 'error feedback' procedure (Hecht-Nielsen 1989, I-593).

The emphasis should be laid on the results of the second article (Matthews and Merriam 1994) since it directly concerns the classification of the works of the Shakespearian and Marlovian canons and the classification of three anonymous works, verifying at the same time the claim of Tucker Brooke (Brooker 1912) that *The True Tragedy of Richard Duke of York* (1595) is by Marlowe and *Part 3 of Henry VI* is Shakespeare's revision of this Marlovian original. In addition, doubts about the assumed undisputed authorship were justified in certain cases, as in Act Five of *Henry VIII*.

1.2.10.2 A Revisit of The Federalist Authorship problem (D.I. Holmes and R. S. Forsyth, 1995), (J. Tweedie, S. Singh and David I. Holmes, 1996) and Mechanic Content Analysis (Colin Martindale and Dean McKenzie, 1995).

In 1995 a new research paper was published dealing innovatively with the Federalist papers, an issue which has been already considered in this chronological review. David I.

Holmes and Richard S. Forsyth applied three new stylometric techniques to the political essays of the Federalists.

The investigators' approach was concentrated on three major stylometric methodologies: first a multivariate analysis of the vocabulary and its richness, secondly a clustering and Principal Component Analysis (PCA), and thirdly a genetic algorithmic process identifying relational expressions of authorial styles. Holmes and Forsyth included the novel elements of Mosteller and Wallace's (1984) findings in their new approach.

PCA explores large datasets and finds the most significant linear correlations between the input variables of the data sets. The limitations of this method derive from the fact that it is proper to use it only with a large number of variables and, as it is a reductionist approach, the relations formed by the large number of variables is examined through the lens, usually, of only the two most important variables. In that sense, PCA detects the most important linear correlations but it cannot detect non-linear relations and does not provide us with adequate information about all the variables. These features make this method inappropriate when, as in this thesis for instance, an investigator wants to model an author's style taking into account all the available information of all the variables that contribute to the modelling of the style.

Another drawback is that with PCA slight changes in the data of the input variables might give a completely different result and, consequently, the theorisation of the principal components might also differ dramatically. (In the Chapter of Conclusions, Chapter Six, there will be a further discussion of the limitations of PCA in comparison with the fuzzification-defuzzification process of the fuzzy stylistic expert systems-classifiers built in this thesis.)

Apart from their theoretical approach concerning the Federalist Papers, Holmes and Forsyth built upon previous work, including their own (Holmes 1991; Holmes 1992; Brunet 1978, Sichel 1975), in order to adapt an existing formula for exploring vocabulary richness by adding a new variable.

As well as the use of neural networks in the works of Matthews and Merriam (1993; 1994) and the genetic approach of Holmes and Forsyth (1995), in the same period Colin Martindale and Dean McKenzie initiated what is called a content analysis (Martindale and McKenzie 1995). In their article titled 'On the Utility of Content Analysis in Author Attribution: *The Federalist*', the authors compare content analysis in comparison with discriminant analysis based on function words and on lexical statistics, as first performed in studies of the length of certain units, such as the average length of words (Mendenhall 1901) and of sentences (Yule 1938, 1944). Martindale and McKenzie applied their analysis to the Federalist Papers problem, taking as granted the overwhelming authorial evidence of

Mosteller and Wallace (1964; 1984) in favour of Madison as author of the previously disputed works. The investigators aimed to check the Madisonian predominance and thus verify the prevailing theory and assess the value of content analysis for the same kind of work.

One year later than Martindale and McKenzie, F.J. Tweedie, S. Singh and David I. Holmes revisited also the Federalist Papers problem using a neural network (Tweedie, Singh, and Holmes 1996). The basic traits and suitability of neural networks derive from the fact that neural networks can learn from their own data, can generalize without the need of a large set of input data, can assess non-linear relationships, and are relatively immune to being distorted by a few faulty data in the training set. Most of all, neural networks cope well with the non-linearity of stylometric problems and benefit from their error-tolerance. In this context, non-linearity means that stylistic markers may indicate authorial styles based on the parallel existence of high and low, or even negatively or positively correlated, counts of words' frequencies.

1.2.11 Scope and Aim of Literature Review the 18th, 19th and 20th Centuries.

In general, this literature review presents in a critical mode the various concepts and trends of qualitative, statistical and computational processing, as they were progressively developed from the last quarter of eighteenth century till the end of the twentieth century.

1.3 Authorship Attribution and Style Detection: The twenty-first Century.

In the previous chapter, and regarding the style discrimination of the eighteenth, nineteenth and twentieth centuries, we saw that in a sequential manner the investigators were devising systematic stylometric methods adding up new techniques of quantification of stylistic features. Near the end of the twentieth century, it seems that the emerging stylometric tendencies tent towards the delineation and condensation of the area of computational stylistics with emphasis on the specialisation of specific machine learning techniques. Classification techniques, various metrics and profiling techniques are some of the methods analysed in that section.

1.3.1 New Emerging Stylometric Techniques (2002-2008) at the Beginning of the twenty-first Century.

The twenty-first century started with the quantification of the features of authorial styles through the use of advanced comparative descriptive statistics and profiling techniques that focussed on the exploration of semantic and socio-psychological/cognitive schemes.

Twenty-first century stylometric analysis is diverse in its methods and in its findings: many techniques are deployed and many styles of authorship are discovered.

1.3.1.1 John Burrows's Delta Procedure and Z-scores (2002).

A computational stylometric methodology that contributed fruitfully to the research of authorship attribution was described in John Burrows's article 'Delta: a Measure of Stylistic Difference and a Guide to Likely Authorship' (Burrows, 2002). This Delta procedure is appropriate especially for texts of about 1,500 words and can be applied in small textual segments as short as 100 words. As has been already mentioned (in Section 1.2.10.1) regarding neural networks, there are the open and closed types of authorship problems (Burrows 2002) and the supervised and unsupervised techniques of pattern recognition. Burrows argues that authorship candidates can be investigated in pairs and that descriptive statistics, such as means and standard deviations, are more powerful than simple frequencies of stylistic markers.

Before delving into the specific research and Burrows's Delta concept, it is critical at this point to refer to the investigator's experimentation and corpus building. Initially, a corpus of 540,244 words was created from texts of 25 English poets of the Restoration (1660-1685). Based on this corpus of texts a 'frequency-hierarchy of the most common words' was created (Burrows 2002, 268). Then, texts were divided into subsets by author. In parallel, Burrows built certain subsets of the most frequent words. A major pool of 150 common words was used for Burrows's various experiments. For each experiment, depending on the selective process, the 30 most frequent words were tabulated. Burrows's approach built on the descriptive statistics of the most common words from all selected texts of the corpus. In that corpus analysis, at first the individual scores-words' frequencies of each poet were calculated and represented in percentages of the total text. The mean of all the scores-frequencies of the selected words and the standard deviation were then calculated. Standard deviation is found

here by calculating the square root of the mean of the squared differences between the individual works' scores for each word and the corpus's means for each word.

At the same time, in order to identify the statistical descriptiveness of the authorial style of the poets, Burrows extracted data from other additional and known works of the poets. Burrows built a main set of a database of verse-lines from these 25 poets and provided correlatively, in the form of two columns, data of descriptive statistics. Thus, he tabulated the mean and standard deviation of the frequencies of the 30 most common words of the corpus. These were named as the figures of the main set. Several additional columns were then built including the descriptive statistics and the z-scores of these words' frequencies in the subset of the textual matter of each one of the English poets. Descriptive statistics and particularly z-scores and Delta-divergent difference were calculated displaying in comparative terms the difference of z-scores of these 30 frequent words, as produced by the works of the selected author and sets of textual matter from works of other poets of the corpus.

Burrows's Delta method is hinged upon a standardised type of score-frequencies, called the z-scores. This type of z-score assesses the difference of a score from the mean of all the scores of the general corpus, and divides this by the standard deviation. This numerical difference can be positive or negative denoting respectively how much higher or lower this score-frequency is in relation to the mean, and comparatively to the other scores of the distribution. In that manner, a holistic statistical analysis was performed displaying the positive or negative z-scores and the difference of the standardised mean of the counts of words' frequencies in the texts of a specific author in comparison to the poetic works under examination. Burrows found positive and negative signed z-scores can shed light on the role of stylistic markers. Probably, they can be viewed as a way to assess the positive or negative contribution of the frequencies of certain words to the standardised final divergent mean, the Delta distance.

In other words, the critical measure is the standardised divergent mean from all the z-scores of the frequencies of the subsets of selected words of the candidate authors under scrutiny. The critical clue of authorship attribution for the disputed texts derived from the difference between the Delta measures of each of 25 poets. The best, namely the lowest (least different) Delta measure among those of the 25 poets indicates which authorial style of the candidate writers is closest to the work under scrutiny.

As a matter of fact, z-scores in comparison to percentages provide more analytical features of a score. Thus, it is a way to represent individual scores/frequencies in a comparative fashion after taking into account the grouping(s) of frequencies and the total of

the individual scores along with various indices of descriptive statistics, such as the mean, variance, standard deviation, and the maximum and minimum scores. The z-scores, as defined above, standardise the individual scores in conjunction with the mean and the dispersion of the values. The means of z-scores of the words' frequencies of the selected works of the multi-authored corpus are then compared with the frequencies of the same words in the works to be attributed and indirectly to the known works, included in the multi-corpus, of all the candidate writers. Thus, initially are produced the z-scores of the differences of these words' frequencies between the multi-corpus and the works under scrutiny and then they are divided by the standard deviation of the respective words' frequencies of the multi-corpus. These differences are converted finally into their absolute form by removing any negative signs. So, rephrasing and shortening Burrows's exact definition, the Delta 'procedure' is the mean of the absolute differences between these z-scores of a set of specific words-variables of a reference and a target text (Burrows 2002, 269). In general, the critical measure is this standardised divergent mean from all the z-scores of the frequencies of the subsets of selected words of the candidate authors under scrutiny. Burrows' Delta method was efficiently applied to a large pool of poets and the processing was performed for each of the pool of 25 English poets. The critical clue of authorship attribution for the disputed texts derived from the differences between the Delta measures of each of 25 poets. The best, namely the lowest, Delta measure among the poets indicates which authorial style of the candidate writers is closest to the work under scrutiny. High deviation from the standardised mean of the scores of 150 words of the general multi-authored corpus would imply large divergence from the style of other authors. On the other hand, a low Delta score would imply little divergence in style and thus assist the detection of the most likely authors from a large pool of candidate poets.

Regarding his Delta procedure, Burrows referred briefly to the side effects of the genre and the 'stylochronometry' (Burrows 2002, 279), a term borrowed by Richard S. Forsyth (Forsyth 1999) to characterise changes in style over time. Burrows accepted that when authors write in a different genre, their authorial style changes, too. In parallel, Burrows added that the writer's style changes over time, but he stated that the existent 'aberrations are rational'.

In conclusion, Burrows devised a computational method overcoming, with the use of z-scores, the simplicities of percentage of words' frequencies and the constraints of the probability theorems that are usually applied only for a small number of candidate writers.

1.3.1.2 Clustering Classification, Centroid Analysis, Vectorisation of Documents and Cosine Similarity.

This subsection describes techniques of (density-based) data points' classification and representation through vectors.

1.3.1.2.1 Clustering and K-Means Partitioning.

In the book titled *Text Mining Application Programming* (Konchady 2006, 275), Manu Konchady described thoroughly machine learning techniques of textual analysis and presented methods of classifying useful and measurable data that are produced by text mining and a tokens-based layered model of information extraction. These measurable data in the stylometric field are usually the counts of tokens' frequencies, which may, for instance, represent the style of different texts of a limited number of authors. In Konchady's analysis, clustering of documents relies on the importance of their distinguishing words, whose co-occurrences in a set of documents can be valued and assessed by the degree of similarity of weighted frequency rates. The usefulness of the method of clustering lies in the fact that a set of random samples of uncategorised subjects can be efficiently and quickly divided into discrete subsets sharing common characteristics. Documents in the same subset have a degree of stylistic similarity, and consistent similarity in stylometry frequently denotes a specific authorial style or in other types of search, as in the analysis of web pages, shared subject matter.

1.3.1.2.2 Cosine Similarity Measure

When classifying authorial styles by unsupervised methods, a major element of measuring the stylistic similarity of documents is the cosine measure. The similarity of the stylistic features of the compared documents is assessed based on the cosine measure of the angle of the counts of the word-tokens' occurrences in these documents. These counts are represented as values in the coordinate space. As reported above (1.3.1.2), in supervised learning the inputs are pre-associated with a specific output category, whereas in unsupervised learning from blind data an algorithm itself makes the categorisation of unlabelled input data. By using the cosine measure the stylistic difference of the documents can be measured in degrees of an angle. From the computational aspect of stylometry, this

measurement counts the angle that the vectors of the counts of the tokens' frequencies of two documents form. Through the counts of frequencies of types' tokens documents are represented as vectors, dimensions by the different words and observations refer to the numerical values of these words.

As a paradigmatic example of the way we count the cosine angle, assume now we explore the counts of the occurrences of three words ['be', 'was', 'maybe'] in two documents (B and C) from a corpus of three documents (A, B, C). Assume that the three dimensional, three words-based vectors of documents B and C are represented respectively as follows:

$$B = [10, 5, 4] \text{ and } C = [4, 8, 7].$$

In other words, the word 'be' is found ten times in document B and four times in document C, the word 'was' is found five times in document B and eight times in document C, and the word 'maybe' is found four times in document B and seven times in document C. Assume that these three words are also found in document A, so that there is not any need of extra processing. In our case, the cosine similarity--of document B and document C--based on their counts for these three words, would be:

$$\begin{aligned} & (10*4) + (5*8) + (4*7) / \sqrt{(10^2+5^2+4^2)} * \sqrt{(4^2+8^2+7^2)} \\ & = 0.80 \text{ (approximately).} \end{aligned}$$

If we think of the three counts--for the word 'be', the word 'was', and the word 'maybe'--as Cartesian coordinates in three-dimensional space, then a line drawn from the origin (0,0,0) to the point (10,5,4) for document B is a vector pointing in a particular direction. Another line drawn from the origin to the point (4, 8, 7), likewise representing the counts for document C's occurrences of these three words, is another vector pointing in a different direction. The cosine similarity we have calculated is the cosine of the angle between these two vectors, measured at their intersection at the origin, and a cosine of 0.80 corresponds to an angle of about 36 degrees. The maximum of the cosine function is 1, corresponding to an angle of 0 degrees in which the two vectors are pointing in the same direction in three-dimensional space, and hence that that their proportions (although not their absolute counts) of the three words 'be', 'was', and 'maybe' are identical.

The minimum of the cosine function is -1, corresponding to an angle of 180 degrees in which the two vectors are pointing in opposite directions on account of one of them having negative counts for the words in question and the other having positive counts. In practice it is impossible for the counts of the occurrences of words to be negative (the least possible count is zero), so the practical minimum value of the cosine function is zero, corresponding to an angle of 90 degrees in which the two vectors are perpendicular to one another (also

known as being orthogonal) because one document has extremely large numbers of the words being counted and the other has extremely small numbers of them. Thus cosine similarity is a measure, on a scale of 0 (for least alike) to 1 (for most alike), of likeness of two documents in their frequencies of occurrence of the words we are counting.

In practice, there are various formulae for calculating the difference between two documents. The use of different formulae affects the criteria of overlapping when numerical values derived from texts are treated as points in multi-dimensional Cartesian space but all of them assess also the asymmetry of the length of documents, short and long, researching words' frequencies in conjunction to what is widely known as Term Frequency and Inverse Document Frequency.

Term Frequency refers to the counts of specific words' occurrences in a single document as a proportion of its total words. Inverse Document Frequency refers to the counts of these words' occurrences after allowance is made for their occurrence in a wider corpus, most often to diminish the significance of words found very often in the wider corpus--as 'the', 'of', and 'and' are in English writing--so that less frequent words are given greater weight.

In practice, the formulae of Term Frequency are expressed by the number of the occurrences of the target words in a document divided by the total number of all the words-tokens of this document. The Inverse Document Frequency of a word is defined as the common logarithm, with base 10, of the number of documents in the corpus divided by the number of documents containing this word. Thus, the Inverse Document Frequency of a word is low when this word is observed in all the documents in a corpus--hence it has low discriminating power--and extremely high when the word is found many times in just a few short documents. Typically, for each target word the Term Frequency will be multiplied by the Inverse Document Frequency to produce a combined measure that incorporates both. Based on these initial formulae, other variants of more complex calculations can be produced considering the counts of the frequencies of specific words and the total of words' frequencies in each and all the documents of a corpus.

In all, combining cosine similarity, Term Frequency, and Inverse Document Frequency, any bias in favour of lengthier documents is avoided, and therefore normalised measurements of documents' stylistic similarities can be efficiently produced. This kind of normalisation is satisfied in the case of the measurement of cosine similarity using the inner product, which contributes to assessing words' frequencies not simply as single numbers, but inside the normalised space formed by the angle of the two documents. The inner product is

defined as a series of multiplications, of the magnitude-length of a vector-document X , which could include numerical frequencies of words, by vector Y corresponding to another document. Cosine similarity is the product of the multiplication of this inner product by the cosine of the θ , which is the angle that is formed between these two vectors-documents. Formally, cosine similarity is defined as the sum of the product of two documents' common words divided by the products of the magnitudes-lengths of these two vectors-documents (Konchady 2006, 275). It must be noted that usually inner product is used interchangeably with the term dot product, though there is a subtle difference in terms of dimensionality particularities.

In conclusion, various metrics have emerged in the area of computational stylistics and can be efficiently applied in order to determine the degree of stylistic similarity between documents of different size inside a corpus of numerous documents. The significant feature of these various distance metrics is that the counts of frequencies of specific words can be analysed in a correlational but also systematic and weights-based parametric mode.

1.3.1.3 Linear Discriminant Analysis (LDA) as an Alternative to Principal Components Analysis (PCA) and Support Vector Machines.

Linear discriminant analysis (LDA) is a similar method of analysis to Principal Component Analysis (PCA) though with certain subtle differences. Both techniques aim at the reduction of the dimensionality of the dataset and the elimination of certain less-distinguishing variables. In PCA, as mentioned above (1.2.10.2), each data point representing a set of readings for a set of variables is treated as a point in multidimensional Cartesian space (with one dimension per variable). These points are projected onto new axes of fewer dimensions than the initial dataset. This technique is applied in order to maximise the variability of the projected points. In fact, with PCA a relatively large number of linearly correlated variables is reduced to a smaller set of variables that maintain as much as possible of the information of the original variables. In LDA the same projection occurs, but the new axes are chosen to maximise the linear separation between the categories of the projected data points. This is possible because each data point is first tagged to show the category it comes from, such as all the works by one author or all the degrees of expression of one gene. Because of this distinction between uncategorised and categorised data, PCA is known as an unsupervised method and LDA a supervised one. In comparison with LDA, PCA as an unsupervised method does not entail any pre-categorisation and it is blind to the categories

the data fall into. Of course, the common element of LDA and PCA is that they both project the data onto fewer dimensions than it originally possessed.

Support Vector Machines (SVM) were invented as a general categorisation technique by V. Vapnik in 1974 (Vapnik 2000; Information Resources Management 2017, 1256). Due to the power of their mathematical complexity SVMs have started recently to be used in fields such as optical recognition of characters and image analysis (Juola 2006, 282). An element of this complexity is for instance the need for the creation of multiple vectors-based hyperplanes, thus a number of associated vectors forming a vector space based on the data-observation points. As we have seen, a vector is a quantity having a magnitude and a direction in the sense of 'pointing' in a direction within multi-dimensional Cartesian space. The results of any data-gathering operation can be treated as a vector since it is simply a list of numbers. A vector space is identified with a set of operations (addition, multiplication) applied to a set of vectors. In three-dimensional Cartesian space, any three points that are not in a straight line define a plane that extends to infinity in all directions and divides the entire space into two regions so that any point in that space falls on one or other side of the plane or else lies on the plane itself. Such a plane has two dimensions, which is one less than the number of dimensions of the ambient space. Equally, two points in a two-dimensional space define a line of infinite length that divides that space in two, which line has only one dimension. This principle can be extended in the other direction into multi-dimensional Cartesian space. Each such space may contain what is called a hyperplane that is defined by a set of points--using as many points as there are dimensions in that space--which divide the space, and which has one fewer dimensions than the ambient space.

It is formally argued (Krisnamourti, n.d.) that two parallel lines or three non-collinear points, or even two intersected lines, define a specific plane and that a single line or two collinear points can form only a line and thus in essence define an infinite number of planes-flat surfaces. In addition, it is usually asserted that hyperplanes represent $n-1$ subspaces of n -dimensions. (Subspaces are vector spaces inside other subspaces). This description of collinearity and planes is not completely accurate. A hyperplane for a line can be a point separating the remaining of its left and right sets of data points (Kowalzyk 2014). Between two parallel lines, the term hyperplane is identified with any single separating line of their contained space, but the most valid separating line-hyperplane is the one that satisfies the best separation in terms of maximum distance to the points-margins of both the line's sides, half-spaces. As the number of dimensions of the (hyper-) topological structure increase, the definitional context becomes more complex. For instance, in 3-dimensional

space two hyperplanes are identified with the lines of the two axes (X and Y axis) and therefore in a three-dimensional space exist three planes (XY, YZ, ZY), thus three flat surfaces.

Support vector methods usually apply a linear classifying function in contradistinction with the non-linear learning mechanism of neural networks. That is, they find the straight line (in two dimensions) or the plane (in three dimensions) or the hyperplane (in higher dimensions) that separates into two classes the points in space that the data correspond to. (In fact, non-linear SVM also exist and employ a more complex structure, the kernel mechanism). SVM modelling employs existent data for training by deriving the plane that best separates the data into the classifications specified by the investigator. After validation, the new data--in our case new observations of counts of words-tokens' frequencies, and hence new texts from which these observations derive--can be tested by projecting these new data as new points into the divided space and seeing which side of the plane they fall, and hence to which class they belong. In general, the use of hyperplanes may be viewed in various fields, even in liquefaction and the prediction of earthquake parameters, as a sophisticated 'pattern recognition tool' aiming at classifying observations into classes (Siddhartha 2015, 126–28).

The hyperplane that this method derives from the data should be one that maximises the 'gap' between the hyperplane and the nearest data points on either side. The principle of maximisation of the margin between the classes of data and the creation of the optimal hyperplane to achieve this facilitate the efficient classification and separation of data, assisting the avoidance of any extra processing burden of pre- or post-selection, since neither any prior standardisation of training data nor all the data apart the margin points are needed. In effect, in that way the remaining of the data, thus those observations distinguished from the markers-'maximisers', do not need to be assessed. Due to the discriminative power of the 'maximisers', an optimal distance is found between the margin markers of the classes of data, as depicted in the graph below with points C1 and B. Consider the simple linear model $y = wx + b$, where w , symbolising the weights of the training patterns-data, is a value for calculating different planes and b , the bias, a line is formed through x and y points. (More commonly and for general use, the formula is known as $y = mx + c$). Such linear functions are employed in maths to measure in a system the constant rate of changes between two numbers. In a two-dimensional space, the point b is where a line intercepts the Y -axis, w is the gradient-slope of the change of one quantity in relation to another, and x , y are the two dimensional points representing the two quantities. The linear correlation of the changes of

the values of these two quantities is described by that line intercepting the Y -axis at point b . In the area of SVM and hyperplanes theorem, three lines parallel to the X -axis can be defined by a linear function where y is greater than zero, for instance plus one, y less than zero, for instance minus one, and y equals zero. The third line can be viewed as the decision boundary line standing equally distant in the middle of the other two lines (and so the three lines intercept y respectively at $+1, -1, 0$). In a slightly different viewing in the simplified graph below a separating intermediate line stands in parallel to other two lines that pass each one through the extreme points-observations (points B, C1) of the two respective classes corresponding to the left and right side. There can be many parallel lines, but the best, or most valid, represents the largest distance between the distinct closest points, already measured observations, of the data of the two different classes on opposite sides. At last, these closest points, known as ‘maximisers’, constitute the Support Vector Constituents. The hyperplanes-based separation can be illustrated inside two parallel lines forming two classes of data, where each line is coinciding with the extreme(s) point(s)-data observation(s) respectively of each of the two classes:

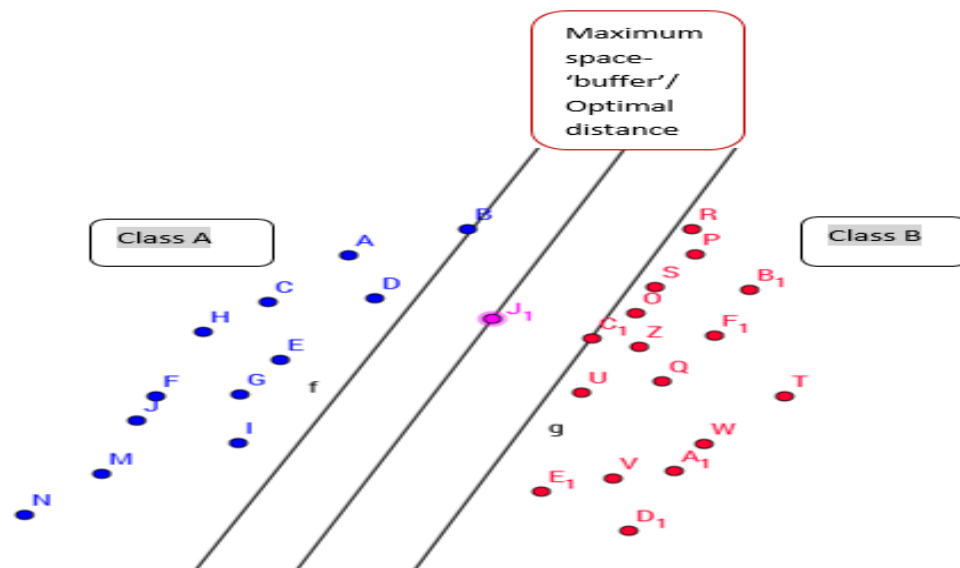


Figure 1: The optimal hyperplane separating theorem.

J1 point has a pink colour. It is not red neither blue, it does not belong to either of the two classes. If this was the case (pink turned to red), then we should say an observation has not been classified (imperfect hyperplane) or a new hyperplane should be searched in a higher dimensional space. J1 is simply a point of the intermediate line and has been selected to visualise the correlation of lines and points. This intermediate line that passes through the pink J1 point is parallel to the lines that pass respectively through blue B and red C1 point.

The width of the gap between the lines passing through B and C1 is maximised and the line that passes through J1 creates two half-planes which are equal in width by definition.

Points-observations in the area of textual analysis and authorship attribution could be viewed as counts of words' or n-grams' frequencies.

In all, despite the burden of its mathematical complexity, the method of SVM has achieved impressive results in various experiments. Such a case is the innovative experimentation of authorship categorisation of 'English and Arabic extremist group forum messages' exploring the advantages of SVM to deal with noisy data and to offer robust results through the principle of 'structural risk minimisation' (Abbasi and Chen 2005). SVM through its advanced mathematical formalisation evaluates numerical observations of words' frequencies maximising those with the most discriminating power.

1.3.1.4 Beyond the Delta Method: John Burrows's Zeta and Iota Measures.

John Burrows in 'All the Way Through: Testing for Authorship in Different Frequency Strata' (Burrows 2007) developed two new stylometric methods of textual discrimination, similar to his previous Delta procedure methodology (Burrows 2002), as described above (1.3.1.4). His new theorisation, built also upon his previous own (Burrows 2005) and others' similar investigations (Hoover 2007), revolves around three different strata of words' frequencies. The first stratum concerns the frequencies of the most frequent terms, function words typically. As already seen in this review, this stratum constitutes a basic layer of analysis in computational stylistics and it is used for stylistic discrimination by a plethora of researchers. But the real innovation of Burrows is the two new strata of analysis. The first one contains the rare words used consistently by a specific writer--in most of their works--and 'sporadically' by others. The second new stratum contains the words employed sporadically by a specific writer and not at all by the majority of the other candidate writers. Burrows intended to demonstrate that this three-stage analysis can lead to consistent authorship attribution, especially if the length of the disputed texts is a bit more than 2,000 words.

Burrows accepted that specific parameters, such as the genre, may cause the deviation of a writer from their usual style. On the other hand, argued Burrows, it is always possible to explore these deviations, the specific idiosyncratic vocabulary of authors, as there are almost always certain indices of an authorial style that pertain to the mentality of a specific writer.

These indices, aiming, according to Burrows, at ‘trends’ and not ‘universals’, were researched through the analysis of the second and third strata of words.

The analysis of the three strata in Burrows’ novel experimentation was made on a set of eight poems, three of which were of known authorship. In fact, two of them, namely, *On the Danger his Majesty (being Prince) Escaped* and *Instructions to a Painter*, are attributed to Edmund Waller, and one, *The First Anniversary of the Government under O.C.*, to Andrew Marvell. The remaining five poems were published anonymously, and their authorship is still uncertain. Burrows used two other sets of a total of 41 poems of the Restoration period. As with his Delta analysis (Burrows 2002), he built a database of 25 candidate authors, including Waller and Marvell. The analysis of the first stratum was performed according to the Delta methodology, which was explained above (1.3.1.1). In the new investigation of Burrows, the first stratum analysis clearly indicated as the most probable writers of the three unquestionable works their real writers, Waller and Marvell. Regarding the remaining of the disputed works, Waller and Marvell ranked low in the list of the 25 candidate writers.

It is noteworthy that whereas the Delta procedure is concerned with the processing of a single text and the candidacies of many writers (25), and it searches for the most probable writer (the closest Δz , Delta z-scores), these new Zeta and Iota tests can be also employed to assign more than one text to the most probable candidate writer. This is achievable as various selections of large sets of words-types can be employed as frequency lists of types with various consistency usages, and as sets of stylistic markers of a particular style in comparison with other authors’ works. So, whereas in the Delta method the emphasis was placed on the style of a specific play, on its descriptive statistics on the basis of a small pool of words, and its attribution to one among many candidate authors, in the Zeta and Iota tests the analysis revolves in a multivariate mode around a number of texts that can or cannot be attributed to a specific author.

For Zeta, the investigator first forms two sets of texts, each being the known work of one author or group of authors. These sets are broken into evenly sized segments and the word-tokens to be used in the experiment are derived from the segments themselves, being those that are used more frequently and the words used less frequently in one or other set. Specifically, what is counted is the number of segments from each set that contain or lack each word-token. The selection of words-tokens was based on the criteria of less commonality of frequency and in the case of Zeta tests, the second stratum included words-tokens found in more than two of five segments of works of the base set and less than three in the counter set. The criteria were more stringent for the Iota tests, as the selected

words should appear in less than three of the five segments of base set and not at all in the counter set.

Overall with Zeta and Iota tests Burrows makes a transition from the lexical stratum of the common words employed in the Delta procedure to two more specialised lexical strata where words of medium and low frequency can play under specific premises a role of stylistic discriminator.

1.3.2 Recent Perspectives of Stylometric Approaches (2009-2017).

This final section of the literature review covers the modern period of delicate theoretical and practical stylometric research, as well as considering the work at its margins. From the entropic experimentation of the Word Adjacency Networks to Pascucci's compression algorithm, the general aim of the investigators of this period is to find novel ways of stylistic analysis departing from the monolithic analysis of the counts of words' frequencies and initiating elaborate profiling methods that reveal and quantify some of the structural properties of whole textual parts.

1.3.2.1 Hugh Craig and Arthur F. Kinney: *Shakespeare, Computers, and the Mystery of Authorship* (2009).

Hugh Craig and Arthur F. Kinney (Craig and Kinney 2009) in collaboration with others dealt with Shakespearian authorship attribution and stylistic discrimination by applying novel tools of statistical analysis, such as the Burrows's Zeta methodology and the centroid metrics, both of which were described above (1.3.1.2.2). Craig and Kinney also applied multivariate statistical methodologies, such as the t -test. This is a technique for measuring the difference of the means of data from different groups in conjunction with, in fact divided by, the respective group's data variability/dispersion index. Dispersion measures the extension-variability from central values, such as the mean. T -tests can be performed in the stylometric area for the comparison of the counts of the occurrences of the same words-types in the texts of two different authors. (A t -test in that case starts with the assessment of the difference of the counts of each word between two different authors, then continues with the calculation of the mean, the standard deviation and the standard error of the differences, which occurs from the division of the standard deviation by the square root of the number of the samples (n). The division of the mean difference by the standard error

produces what is called *t*-statistics, which provide the likelihood of the occurred-actual results under the acceptance-expectations of the null hypothesis, for instance implying that there is no significant substantial stylistic difference between two compared authorial patterns.) Furthermore, Craig and Kinney made mention of Principal Component Analysis, analysed above (1.2.10.2), as a way of combining, for instance at the simplest level, two variables and forming a single new representative variable (Craig and Kinney 2009, 28–30). In addition, the investigators of this book adopted Brian Vickers's approach (Vickers 2007) of selecting as stylistic markers not simply rare words but rare 'phrases and combination of words'. Craig and Kinney proposed that rarity of stylistic features, terms that are found in the passages of a specific writer and not or rarely found in other authors' passages, can assist stylistic discrimination. In addition, they commented that the rarity of terms, words or phrases, can be quantitatively determined with approximate fixed criteria as for example terms that cannot be detected in more than 50 or even a larger collection of 500 books (Craig and Kinney 2009, 57–58).

In conclusion, the book of *Shakespeare, Computers, and the Mystery of Authorship* contains practical applications by various researchers of most of the novel computational stylometric methodologies that were devised in the beginning of the twenty-first century. New evidence was also presented through the various segments' analysis indicating the Shakespearian co-authorship of *Arden of Faversham*, *Edward III*, *Sir Thomas More*, and *The Spanish Tragedy*. In fact, the researchers isolated Shakespeare's contribution to these plays. They found Shakespeare's hand in the middle of the section of the play of *Arden of Faversham*, for instance (Kinney 2009a, 78–99). At the same time, significant clues were provided for the revision by Shakespeare of *Sir Thomas More* and the quarto text of *King Lear* (Kinney 2009b).

1.3.2.2 *The New Oxford Shakespeare: Authorship Companion* (2017).

Mark Eisen, Santiago Segarra, Gabriel Egan and Alejandro Ribeiro with their article titled 'Stylometric Analysis of Early Modern Period English Plays' (Eisen et al. 2016) devised an innovative method of stylometric analysis by building function word adjacency networks (WANs). By analysing the serial-selective combination of function words in plays and building their adjacency networks (WANs.) Eisen et al. in practice created aptly linear, and indirectly probabilistic, classifiers of authorial styles. But the major breakthrough was made a year later with a recent book that was composed of the research efforts of many

scholars, that is *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan (Taylor and Egan 2017). As the title indicates, this book aims to give a comprehensive account of various scholars' critical perspectives regarding authorship attribution and its validation. Gabriel Egan highlighted in the chapters titled 'A History of Shakespearian Authorship Attribution' (Egan 2017a) and 'The Limitations of Vickers's Trigram Tests' (Egan 2017b) that authorship attribution and textual similarities have to be validated on the basis of internal evidence in a broad perspective and by employing scientifically robust methodologies.

From the whole book, two specific methods deserve to be mentioned for their novelty in comparison to the stylometric methods that already have been discussed in this literature review. The first is a compression algorithm of Giuliano Pascucci (Pascucci 2017), and the second a multi-variate statistical analysis of Jack Elliott and Brett Greatley-Hirsch (Elliot and Greatley-Hirsch 2017). The latter essay combines Burrows's Delta, Zeta and Iota test with the Nearest Shrunken Centroid and Random Forests techniques.

Giuliano Pascucci in the essay 'Using Compressibility as a Proxy for Shannon Entropy in the Analysis of *Double Falsehood*' (Pascucci 2017, 409) employed generally accepted principles of engineering analysis and after dividing the work of *Double Falsehood* into scenes, he created a list of plays divided into sections of 32 kilobytes. He employed a differentiated version of the LZ77 algorithm, which was named BCL, in order to compare the compressibility of certain Elizabethan and Jacobean writings after the algorithm had built its dictionary of strings to be tokenised by compressing a work of known authorship.

Jack Elliott and Brett Greatley-Hirsch in their stylometric exploration of '*Arden of Faversham*, Shakespearian Authorship, and "The Print of Many"' (Elliot and Greatley-Hirsch 2017, 149) applied a range of stylometric methods: Delta, Zeta, Nearest Shrunken Centroid and Random Forests. The Nearest Shrunken Centroid is a way of extracting an average value from a set of data points, each of which represents the frequency of a word in a text. Treating these frequencies as Cartesian coordinates, the occurrences of all the marker function words in the different segments were represented by single points (one per segment) in multi-dimensional space. The term 'multi' in this case corresponded to the number of variables-dimensions, thus function words, to be explored. The crucial point in this technique was the measuring of the distances of these points from one another and, where they cluster together, finding the centre of the cluster, called the centroid. The centroid has for its coordinate in each dimension the arithmetic mean of the values for that dimension held by the set of points in the cluster. The 'nearest sunken' refinement of this calculation discounts the

contribution to the mean made by those words for which an author's rates of usage (across different segments) are highly inconsistent. Finally, another method that was considered in *The New Oxford Shakespeare: Authorship Companion* by Elliot and Greatley-Hirsch was Random Forests, a technique that is based on the decision trees. In such a tree, every node except the leaf-nodes at the bottom represents a question with a 'yes'/'no' answer, such as 'is the frequency of the word *and* > 0.012 ?' or 'is the frequency of *an+a* divided by the frequency of *a* < 0.2 ?'. Two lines emerging from the node, one for when the answer is 'yes' and one for when it is 'no', lead to further nodes and further questions. Each leaf-node at the bottom of the tree represents a conclusion (such as 'this text is by Middleton') reached because of particular set of 'yes'/'no' answers that lead to that node. In principle, the questions in decision trees need not to have only 'yes'/'no' answers: three or more lines could emerge from each node to represent three or more possible answers to the question being posed. Elliott and Greatley-Hirsch did not make clear whether their trees were binary.

The process of creating a decision tree is algorithmic. A training dataset of texts for which the different authors' names are known--which will correspond to the leaf-node values--is repeatedly decomposed into smaller subsets based on the results of the various tests that were applied to it (such as 'frequency of the word *and*' and 'frequency of *an+a* divided by the frequency of *a*') until a set is found to be 'pure', in the sense of all its texts being by the same author. Each subdivision into new subsets adds a new node to the tree, representing a rule based on one of the tests that were applied. When a resulting subset is 'pure', we have found a leaf-node and can label it with a decision (such as 'the text is by Middleton'). The process is repeated until every subset is 'pure' and has been labelled as a decision about authorship. The paths to some decisions will be short (passing through only a few questions) and the paths to others will be long (traversing many questions), but every path through the tree leads to a decision node. When making an authorship attribution, each tree in the set of trees (known as a forest) is given the text to be attributed and its 'decision' is noted; the final decision is reached by treating these multiple, competing decisions as votes. In Elliott and Greatley-Hirsch's method, every decision tree contributes to the final result, for instance a positive or negative attribution outcome for a specific candidate writer. To sum up, the importance of such a methodology in the computational stylometry is that, similarly to Burrows's Zeta and Iota method, it avoids the monolithic stylistic exploration of words of high frequencies, but instead it evaluates the correlational value of words with low frequencies.

1.3.3 Scope and Aim of Literature Review of the Stylometric Methods of the twenty-first Century.

Making a critical evaluation of the computational methodologies of the twenty-first century, Burrows's claim that styles can be recognised despite the changes of authors' styles over time, what is known as 'stylochronometry', and genre's particularities, have not been extensively discussed or tackled in the experimentation field of stylometric investigation. Regarding the cosine similarity and the representation of documents as vectors of counts of frequencies of words in computational stylistics, there is not any kind of reference in the existing literature or technical resources about the proper number of words in relation to the length of the compared documents. Furthermore, there is not any specific evidence about the features of the words-types included in the two vectors, for instance if they are function or lexical words or if a combined set can be used. In addition, the result of the cosine similarity and the stylistic similarity of two compared documents will be differentiated if low, medium or high frequently words are used as stylistic discriminators. In that sense, innovative computational methods, such as the index of cosine similarity, have also to be further described and developed taking into account the type of the texts and the goal of the experimentation since this index is mostly employed for the stylometric Web-based exploration of documents.

At this point, it is worth mentioning that almost all of the problems of authorship attribution that have been tackled with all prior methods described in the literature review can be dealt--with the proper adaptation--by employing the principles of the Fuzzy logic and the Matlab-based computerised environment (specific examples are mentioned in this thesis's conclusions in Chapter Six). That is, the method described here can use as its inputs the results of almost the existing methods described above and so it offers a general-purpose tool for rapidly synthesising a broad range of approaches.

2 Chapter Two: Methodology and Objectives

In this chapter there is an analytical description of the research paradigm and type of research adopted. The focus is on the pre-processing stage, the criteria of selection of samples and the building of the two corpora. Then, there is a discussion of the selection of the stylistic markers-words (100 in total) and their grouping into four distinct semantic sets. In addition, the structure and functional role of the three stages of the experimentation are discussed. Finally, the general rationale, advantages, limitations and objectives of research design are reviewed in the context of delivering a practical and automated Fuzzy stylistic Simulator-Classifier.

2.1 Research Paradigm--Post-Positivism and Functionalism

It has been already highlighted in the Introduction (1.1) and through the literature review in Section Two (1.2) and Three (1.3) of Chapter One that the determination of authorship of anonymous and co-authored plays in the stylometric area is a complex issue. This risky venture necessitates a range of theoretical evidence of historical scholarship and a variety of measurement methods. As the goal of this kind of determination is associated with the assessment of quantifiable stylistic traits, the computational stylometry by nature falls into the field of the positivist research paradigm. Traditionally, the term positivism in scientific research denotes the effort to measure, gather and assess numerical data in order to make inferences and draw objective conclusions about a phenomenon, an event or a system (Ayer 1978). In the positivist area, the extraction, measurement, assessment and classification of numerical data is construed as an effort to explore rationally and model a problematic situation or an event without the influence of subjective evidence or preferences. And indeed there are two main reasons for building models, that is the understanding of the problem and the desire ‘to communicate important aspects of the solution system’, whereas at the same time there can be use of a ‘sophisticated visual modelling tool’ allowing adaptations and changes after the reception of feedback during its operation. As the determination of authorship is not always a straightforward process and due to the fact that it falls into the field of Digital Humanities with many indistinct conceptions and tendencies of phenomenological theorisations, which entails subjective evidence, it is more appropriate for the investigation in this thesis to add the prefix ‘post-’ in our positivism research paradigm. By the term indistinct

conceptions I imply the difficulty in defining what constitutes an extract in terms of a large or a short text or what is the ideal size of analysis for a specific author. In addition, this term expresses the unclear theorisation of what principally constitutes a style (words, n-grams or words' adjacency networks) and the ambiguity arising from the term style that, for instance, may or not include the stylistic 'peculiarities' of an author, as 'peculiarities' do not conform always to the general style of an author and, therefore, their use for stylistic discrimination may be dependable on a number of parameters or complexities.

The research paradigm in this investigation has also a functionalist perspective. Though the term functionalism has been primarily used in the political field (Ernst Bernard Haas, 1924-2003), sociological domain (Emile Durkheim, 1857-1917) and in linguistics (Prague school, 1930s), it can be equally argued that the research in this thesis targets computationally at the disclosure of hidden truths about the latent functional and structural layers of texts and latent individual patterns of an author's style by employing as stylistic markers not only counts of individual words but also the measurement of counts of semantic sets of words. This kind of functionality is enhanced technically by the building of an automated Fuzzy-Logic-based expert system and should be interpreted in our case as an attempt to devise an efficient method, which can be largely employed for discriminating writing styles semi-blindly and classifying efficiently common and uncommon features of an author's writing style. (I mean 'common' in the sense that these features-stylistic markers are detected very frequently in an author's various plays and I use 'uncommon' if these features are not frequently detected.) The term 'semi-blindly' expresses the functional-discriminating role of the sets of words and not only of the counts of words' frequencies.

In the study of political science the terms 'functionalism' (Mitrany 1975) and 'neo-functionalism' (Risse 2005) express respectively the top-down and bottom-up process of gradual integration of different parts-agents in a *sui generis* pseudo-organism that performs a number of functions after the completion of integration, despite the fact that its constituent parts do not form a new unique entity. An example in political science is the European Union as an intergovernmental organisation. Similarly, in the domain of authorship attribution there exists a functionalist perspective. In effect, this thesis through approximate reasoning and the application of fuzzy methodologies aims to evaluate the functional role of counts of partial textual data as constituent markers of the style of an author who over time has produced distinct integral plays. The term integral denotes here Aristotle's definition about what constitutes a whole (ὅλον), a coherent play (Aristotle's definition lays emphasis on tragedy), that is, a play of some size, a series of 'actions of mimetics', with a beginning (ἀρχήν) and

middle (μέσον) and end (τελευτήν) (Aristotle est. 335-323 AD, 1450b; Fyfe 1932). With the use of the term ‘partial textual data’ it is implied that only a small amount of the whole text, just a pool of words, is each time quantified and assessed stylistically.

2.2 A Mixed Research Method

The method of the research in this thesis is of exploratory nature and it is based on inductive reasoning. The exploratory character of the method is determined by the research goal, namely to establish ‘What is the Shakespearian canon?’. In contrast to deductive reasoning which is a top-down approach and uses a parallel reasoning from case to case, the research method applied here clearly follows a reverse, bottom-up, direction. Let us examine this direction stage by stage. First, the intention is to discern if new conclusions emerge by extracting not only data points of words but also of the sets of the selected marker-words. Then, in the second stage during the processing of the data and the calculation of descriptive statistics the task is to find ways of modelling the data points in a fuzzy mode (which term will shortly be explained) and to proceed to the formulation of some tentative hypotheses. In the third stage, the purpose is to generate a new theory with practical orientation: the building of a fuzzy classifier that assesses numerically the stylistic similarities of a disputed or anonymous text to an author’s known style (in our case, the style of Shakespeare).

With the description of these three stages the inductive technique prevails and there is a bottom-up approach, which constitutes a transition from particular facts (the modelled-patterned observations and differently rewarded rules) to a general conclusion. This process from specific to general and the postulation of a new theory is produced after the validation and testing of the fuzzy classifier through computer simulation. The core computer simulation takes place at the third stage. The simulation in the experimentation arises from the fact that with the design of a system with multiple inputs (the counts of various features in the text), with the output of a single numerical index of similarity to a given writer’s style in one case (in the primary experimentation) or of two indices constituting a SiS interval (in the core experimentation) and the associated inference mechanism, a computer program emulates human reasoning and decision making based on prior knowledge in the form of modelled data. The fuzzy representation of the multiple inputs and one output together with the inference engine constitute the Fuzzy Knowledge Base. But whereas the applied research method is characterised by induction, the computer program simulates human thinking in a deductive mode. The modelling of the fuzzy multiple-input-one output classifier system, the

construction of the rule-based mechanism and its rewarded rules, and in general the simulation of human thinking, all hinge not upon any kind of training using data but upon prior knowledge.

In technical fields, computer simulation is employed as an autonomous research tool; in the current set-up of the experiment it constitutes--along with the exploratory approach and the inductive reasoning--the constituent part of a mixed, hybrid research method.

2.3 Type of Research: Primary Resources & Quantitative Analysis

The corpus of the 27 well-attributed, sole-authored Shakespearian plays provides the textual data which are processed, represented numerically and modelled fuzzily. These textual data are combined with a second corpus of well-attributed non-Shakespearian plays and disputed (and sometimes co-authored) plays. This makes for authoritative evidence for writing styles in the early modern period. These are considered primary sources for this study. Historical scholarship regarding the formation of the Shakespearian canon and the provenance of the plays under scrutiny in the testing stage is considered secondary resource: only counts of textual data can be objectively measured and these constitute the expression of an author's actual style. Though the theoretical investigations about the problems of authorship attribution and what in general is considered as secondary resource requires qualitative analysis, the major type of the research in this thesis is quantitative. The results in the last stage of the experimentation, which is described in the next subsection, entail also qualitative analysis through the general evidence of historical scholarship. Consequently, the typology of research investigation has again mixed elements, but the quantitative analysis prevails.

The quantitative analysis in the experimentation of this thesis follows multiple paths and though these are analysed in detail in the next two chapters, it is useful to mention them briefly in this subsection. The quantitative analysis starts from the primary processing of the texts with the quantification of certain stylistic features and the calculation of descriptive statistics, continues with the fuzzification, which is the estimation of the membership degrees and the delimitation of intervals of input classes, and finally ends up with the procedure of defuzzification, which produces (through differential calculus) a central value that expresses the relationship of the inputs and the output of the system. Defuzzification is the process of turning fuzzy input values into a crisp, single numerical value. I analyse these concepts ('fuzzification', 'defuzzification') below in Chapters Three (especially section 3.1) and Four.

2.4 Research Design and Corpus Building: Techniques of Data Collection and Tools of Analysis

The first step in the research design is the stage of pre-processing: the building of the corpora of the texts that are the subject of experimentation. The first corpus consists of 27 plays (see Table 1 on page 74) that were written by Shakespeare alone. The rationale for the selection of the well-attributed Shakespearian, well-attributed non-Shakespearian and disputed plays was explained in the Section of Introduction (1.1). The second corpus (of 14 plays) contains only well-attributed, sole-authored plays of other authors and anonymous or co-authored plays. The same logic-rationale of the selection of the known 27 Shakespearian plays applies also in the selection of the 14 plays used in the Validation Stage 2 and the Testing Stage, as these plays are listed either as well-attributed non-Shakespearian (Validation Stage 2) or disputed plays in the Appendix A of the book titled *Shakespeare, Computers, and the Mystery of Authorship* (Craig and Kinney 2009, 213–18). Consequently the ratio of the Shakespearian to non-Shakespearian plus disputed plays is almost 2 to 1 (2:1). In fact, for the building of the corpus of the Shakespearian, non-Shakespearian and disputed plays apart from Appendix A of that book (Craig and Kinney 2009, 213–18) has been also consulted by the present investigator a relevant article submitted on the ‘5th International Conference on Corpus Linguistics’. In that article a corpus of 27 known texts (written in Spanish) has been built and its conclusions advocate the use of texts of around 10,000 words against shorter texts (500 words) regarding the discrimination of the general style of an author (López-Escobedo et al. 2013).

The first corpus is used in order to build the Fuzzy Knowledge Base, which comprises the membership functions of the input-variables, the inference engine and the database of rules. Typically, the initial stage is called the training stage but due to the fact that the model is built based on prior knowledge and not what is normally meant by ‘training’ in Machine Learning, it is more appropriate to call that stage the modelling stage. The second corpus is employed both for the second stage, validation, and the third stage, testing. The term ‘testing’ should be viewed with caution because the success of the Testing Stage is evaluated through the general evidence of historical scholarship, too. Taking into consideration this evaluation stage, we could say that the experimentation includes three stages, and the last of them is followed by a fourth meta-stage providing feedback that could be exploited for the optimisation of the model’s parameters.

During research design and the pre-processing stage the first concern is to avoid bias. The present investigator tried to ensure the avoidance of two main types of bias, which are selection bias and what is known as the Observer Expectancy Effect. The minimising of biases was attempted by selecting all the texts from the same academically valid resource (identified below), the categorisation of the plays into the three widely accepted genres of comedy, history and tragedy, and, particularly regarding the Fuzzy-Logic-based modelled Shakespearian style, by the selection of plays that are well-attributed and sole-authored by Shakespeare. To fall into this category, plays had to be without general suspicion of revision or adaptation by another author. Even more important is the fact that I respected Yule's theorems and Zipf's frequency-ranking law. As discussed in the literature review (1.2.10.1), Yule proved that there is not a linear correlation between the size of the text and the total numbers of occurrences of words, particularly of the nouns, it contains. There are, he showed, critical points where this correlation is broken, especially as the text grows, and so we could deduce that extracting the first part of the integral text of a play, for instance the first 10,000 words, is not exactly the same as extracting another part, say the next 10,000 words (words 10,001-20,000). Yule showed that authors tend to become familiar with a certain pool of words and this makes it easier for them to use these words more frequently, especially in the beginning of the text.

This insight of Yule's embodies Zipf's principle of the least effort (Zipf 1949), which states that as the sample of a text gets larger and larger, the increase in the use of the words that previously were most frequent is not commensurate with the increase in the size of the text. In other words, as the text grows, authors tend to retrieve 'words at risk', which are new words from their wider available vocabulary. To treat all texts equally despite their different sizes, for this study only approximately the first 10,000 words of each play were extracted during the building of our two corpora. The texts of these plays are encoded using eXtensible Markup Language (XML) that conformed to the specifications of the Text Encoding Initiative (TEI). These were modernised, regularised versions by researchers at the Centre for Literary and Linguistic Computing (CLLC) of the University of Newcastle in Australia under the guidance of Hugh Craig, who is expert in sixteenth and seventeenth century literature. The texts (see versions in List Addendum of Section 7.1) were regularised by the researchers with the use of specialised software (see Section 7.3.2 of Technical Appendix). An example of such regularisation is that the early modern words 'my selfe' were turned into their modern equivalent 'myself' in order not to miss an occurrence of the pronoun 'myself' or not to count

one word as two. This textual intervention is particularly important for the chopping of texts and the proper calculations of counts of words.

The dataset of the regularised plays is composed of two folders, one of them is of 275 and the second of 64 plays. This division has been made by the researcher. The first folder contains 275 unique plays and the second, named Versions and Selections, repeats some of the plays in the larger set but also includes folio and quarto versions and splits some plays by author. In fact, the second, much smaller, folder contains plays that have been also the primary object of Craig's research (Taylor and Egan 2017). So, the collection of these 343 plays included quarto and folio versions of the majority of plays of the early modern drama. For the building of the first corpus of the 27 well-attributed Shakespearian plays (see Table 1 on page 74), only versions of the First Folio (1623) were selected. This collected plays' edition was favoured because it represents the output of a single printshop within a fairly short period (the years 1622–23) by a single team of compositors and proofreaders and for the works in question no one has seriously questioned the authority of the Folio versions. The same dataset was used also for the selection of plays for the validation and testing stages. Thus, all the plays were regularised using the same criteria and by the same researchers. After the selection of the plays that constitute the first corpus, the texts were then processed by the present investigator and all speaker or other non-textual tags were deleted along with the descriptions of scenes (for instance, textual about information who enters or exits), and care was taken to filter any mishaps or typographical errors that could affect our counts, such as a punctuation mark appearing between two spaces.

The texts were made roughly of equal length by deleting all but the first 10,000 words, as explained above in relation to the findings of Yule and Zipf. The texts were not chopped blindly at the threshold of 10,000 words but instead I set intentionally some constraints. In the range of minimum around 9,950 up to maximum 10,150 words I chopped the text of each play at the end of an act, when this was possible. When there was not an end of act in the range of 9,950 to 10,150 words, I used the end of a scene. In cases where there was not a scene in that range, I chopped the text at the end of the speech which was nearest the limit of 10,000 words. In addition, if a series of alternate rhymes (stichomythia) were found in that range of 9,950 to 10,150, I cut the texts at a speech just before or after them, keeping always the approximate minimum and maximum thresholds of 9,950 (in fact 9,945 in *Richard II* because an act ends at that limit) and 10,150. In practice, the average length of the selections from the 27 texts is 10,044 words and the standard deviation close is 52.4 words, as can be

seen by the table below. Furthermore, for 22 of the plays the word count of the extract falls within one standard deviation of the average.

Plays (27)	Statistics	Number of words
<i>As You Like It</i>		10,004
<i>The Comedy of Errors</i>		10,145
<i>Love's Labour's Lost</i>		10,002
<i>The Merchant of Venice</i>		10,012
<i>The Merry Wives of Windsor</i>		10,037
<i>A Midsummer Night's Dream</i>		10,073
<i>Much Ado about Nothing</i>		10,008
<i>The Taming of the Shrew</i>		10,136
<i>The Tempest</i>		10,013
<i>Twelfth Night</i>		9,966
<i>The Two Gentlemen of Verona</i>		9,949
<i>The Winter's Tale</i>		10,069
<i>Henry IV Part 1</i>		10,042
<i>Henry IV Part 2</i>		10,077
<i>Henry V</i>		9,995
<i>King John</i>		10,094
<i>Richard II</i>		9,945
<i>Richard III</i>		10,058
<i>Antony and Cleopatra</i>		10,094
<i>Coriolanus</i>		10,043
<i>Cymbeline</i>		10,057
<i>Hamlet</i>		10,013
<i>Julius Caesar</i>		10,017
<i>King Lear</i>		10,092
<i>Othello</i>		10,082
<i>Romeo and Juliet</i>		10,116
<i>Troilus and Cressida</i>		10,057
	MAX IS:	10145
	MIN IS:	9,945
	MEAN IS:	10,044
	STANDARD DEVIATION IS:	52.4 (Rounded to one decimal place)

Table 1: Length of texts of the corpus of 27 well-attributed, sole-authored Shakespearian plays.

(For a full list of the plays and their versions, including also the 14 texts of the Validation Stage and Testing Stage, see List Addendum of Section 7.1.)

The justification for this small flexibility around the 10,000-word boundary is that words and their counts should be viewed in the context of integral textual entities: plays, acts,

scenes, and speeches. We should not mechanically cut segments from a whole play but let the author finish what he has to say at specific integral parts of. Using a fixed word-count could cause a minor bias by breaking abruptly the exponential sequence of use of words from the writer's available dictionary as understood from Zipf's and Yule's principles. Shakespeare might at one point employ a repetitive structure with interrogative pronouns in a stichomythia but very probably he would not intend to finish a segment like that but rather would have the characters return to a more natural style of speaking. The use of stichomythia, we might say, implies the use of something else after it before a segment ends. Likewise, with the other segmental considerations of dramatic writing the minor difference of the exact length of texts, averaging 1% of the total length, does not significantly affect the calculations and the extraction of descriptive statistics. This kind of chopping of texts is congruent to the Fuzzy Logic and the so-called theorem of (weak or strong) alpha-cut α of a fuzzy set S . Roughly speaking, this theorem defines, instead of zero or one, an intermediate membership degree as a threshold value (weak if this value is not inclusive and strong if it is). Along similar lines, it can be argued that the segmenting of the texts (and their inclusion in each corpus) was made on the basis of two attributes-variables and two alpha-cuts, one for the class of the minimum variable (9,950) and one for the class of maximum variable (10,150).

Let us see now the plays that compose the first corpus of the 27 well-attributed, sole-authored Shakespearian plays.

COMEDIES (12 plays)

As You Like It

The Comedy of Errors

Love's Labour's Lost

The Merchant of Venice

The Merry Wives of Windsor

A Midsummer Night's Dream

Much Ado about Nothing

The Taming of the Shrew

The Tempest

Twelfth Night

The Two Gentlemen of Verona

The Winter's Tale

HISTORIES (6 plays)

Henry IV, Part 1

Henry IV, Part 2

Henry V

King John

Richard II

Richard III

TRAGEDIES (9 plays)

Antony and Cleopatra

Coriolanus

Cymbeline

Hamlet

Julius Caesar

King Lear

Othello

Romeo and Juliet

Troilus and Cressida

Regarding the second stage of validation, this uses well-attributed, sole-authored plays of other authors (non-Shakespearian plays), whereas the third, Testing Stage contains plays that are disputed and probably co-authored (see full list of plays for Validation Stage and Testing Stage in Section 7.1). For comedies the validation plays are: *The Staple of News*, a play by Ben Jonson (Craig and Kinney 2009, 215), *A Mad World, My Masters* well-attributed by Thomas Middleton (Craig and Kinney 2009, 216) and *The Wild Goose Chase* well-attributed by John Fletcher. For histories the validation play is *Edward II* by Christopher Marlowe (Boas 1960; Parks 1999; Craig and Kinney 2009, 215). For tragedies the validation plays are: *The Jew of Malta* by Christopher Marlowe (Craig and Kinney 2009, 215) and *The Spanish Tragedy* by Thomas Kyd (the version without the additions attributed to Shakespeare and perhaps others) (Craig and Kinney 2009, 217–18).

In this second, validation stage, I experimented with a separate fuzzy model for each genre. I selected non-Shakespearian plays for validating each of our three genre-based models in order to examine the behaviour of the fuzzy Shakespearian classifier for non-Shakespearian plays and to assess stylistic norms without the possible distortions caused

by the differentiations of genre. If the model asserts that the Shakespeariness of non-Shakespearian plays is high, I would have to optimise the parameters and/or make changes to the chosen stylistic markers (words or sets of words), or even add new stylistic markers to the input system of the fuzzy classifier of Shakespeariness.

In order to see how the model of the corpus of the well-attributed Shakespearian plays behaves in the real world of the disputed and anonymous plays, and to test the three models' plausibility as classifiers of the degree of Shakespeariness for disputed cases, I selected for the testing stage in each genre the following plays:

Comedies: *All's Well That Ends Well*, *Measure For Measure*.

Histories: *Henry VI Part 1*, *Edward III*, *Henry VI Part 2*, and *Henry VI Part 3*.

Tragedies: *Timon of Athens* and *Titus Andronicus*.

The goal in this third stage is to attribute the degree of Shakespeariness of anonymous plays and to form concrete stylometric conclusions such as if *All's Well That Ends Well* is almost entirely a sole-authored Shakespearian play--the *New Oxford Shakespeare* claims that perhaps a few dozen lines are Thomas Middleton's (Taylor 2017, 322–336; Loughnane 2017, 287–302)--whereas *Henry VI Part 1* is probably a play of at least two authors, one of them being Shakespeare who might well have had only a minor role in its creation (Holdsworth 2017, 338–365; Taylor and Loughnane 2017, 416–518).

Regarding the claim of *New Oxford Shakespeare* for *All's Well That Ends Well*, in fact, it is the (King's) speech of 2.3.109–36 that is identified, with a high probability, with Middleton's style (Nance 2017, 329). This short passage is included in the extract of 10,000 words that is examined in this comedy-based experimentation. With the proper adaptation a mini-Fuzzy-Logic-based classifier could be built using John V. Nance's markers (heroic couplets and specific n-grams), as detected in well-attributed passages of the two candidate authors, Shakespeare and Middleton, giving us an authorship verdict for that part of the play. Perhaps through a rolling procedure another option could be to explore set-wise the whole Scene 3 (of Act 2) of *All's Well That Ends Well*. Of course, this would require modelling Shakespeare's and Middleton style. In any case, if the SiS scores differed in case the play was examined without the speech of 2.3.109–36, it would be an additional indication that this speech should not be attributed to Shakespeare.

The method of style analysis adopted here is to count the frequencies of certain words that the investigator has, subjectively, but based on semantic and functional criteria, put into

certain categories so that as well as counts for each word I can tally counts for each set of words. A total of 100 words have been split into four sets. One of the sets, Set Three, is, in fact, composed of three further subsets that play the role of independent sets in the primary experimentation (see Section 7.2). The rationale for the subdivision of Set Three is that it could enable me to increase the number of the basic four sets or change the words they contain if our initial results from the validation stage were not satisfactory or the discrimination power of the initial distinction of the four sets was low and mirrored general instead of an individual style. The three unified sets of Subset One, Subset Two and Subset Three that constitute Set Three were proven in the primary experimentation of Chapter Three to have on average the smallest counts of all the six sets: 0.2 % (Set Three), 0.4% (Set Six) and 1.1% (Set Five). With the unification of these three subsets in the core experimentation of the thesis (Chapter Four), there is the option to avoid counts less than 1%. Furthermore, with the use of fewer sets-input variables, for instance four instead of six, it is feasible to control the fuzzy rule-base more efficiently and with less complexity. In authorship attribution we count various things in the text at the same time, each forming a separate input to our method of analysis. In the investigation of this thesis, the four inputs are sets of words though, as explained in Chapter Four and Chapter Seven, these words were also employed for the extraction of an additional input, an index of counts of words' frequencies from play to play. This index is used to generate a metric known as cosine similarity, as previously described (1.3.1.2).

Let us describe now the process of the selection of the 100 words that are contained in the four sets. The main decision was to apply a kind of deterministic randomness in building sets of words on the basis of various semantic and grammatical criteria. Almost half of the 100 words are function-grammatical words and the other half of the words have a lexical meaning. I also tried to avoid employing words that require extensive lemmatisation or homographs categorisation (such as 'that') and very high frequency words that have a strict functional role, such as the determiners 'the', 'a', and 'an'. (Function words can be categorised at a differential scale. For instance, the pronoun 'she' arguably has more information than the determiner 'the' since it nominates its referent's gender whereas 'the' does not. These considerations shaped the choices of certain function words, but otherwise such arguments were set aside and all function words were treated as essentially equivalent.) Effort was also put to avoid using words that would produce zero counts in all plays of the corpus. Furthermore, I took into account that less favoured by Shakespeare (Merriam 2016, 405) were function words such as 'and', 'in', 'of', 'that', 'the', and 'to' and so I avoided

including them in any of the four sets. The N-gram viewer provided by Google was also occasionally consulted for theorising the general use of the selected words during the period of the early modern English period. As for the grammatical variety, the aim was to include in the sets words of various grammatical categories, such as nouns in singular or plural (such as ‘night’ and ‘nights’), pronouns and adjectives. Effort was also made, as Mosteller and Wallace advice, not to include words that are contextual (Mosteller and Wallace 1963, 277–78).

The logic of the decision making was to form four main semantic sets that would meet the general principles of semantic, functional and thematic axes (Osgood, Suci, and Tannebaum 1957) and the principles of the Russian Formalism (a movement that flourished around 1916-1929), which lays emphasis on the structural elements of narration and plots’ formation through the so-called ‘devices’ of time and space, the mechanism of metaphors (Steiner 1984), the ‘representational elements of action and event in their natural chronological and causal order’, and ‘the textual presentation created by artistic compositional patterns’ as well (Margolin 2005, 3). By forming these four semantic sets, the orientation is towards detecting these interrelational, compositional patterns. The sets were formed with the subjective and intuitive selection of words that responded to the semantic requirements of each set, as described below. (Of course various other combinations could be made). So, Set One includes words that signify mainly space and time.

Set One is composed of 14 words: ‘if’, ‘then’, ‘else’, ‘here’, ‘there’, ‘more’, ‘less’, ‘now’, ‘night’, ‘nights’, ‘day’, ‘days’, ‘morning’, and ‘tonight’. In Set One, I included also conditional conjunctions since hypothetical events are situated always on a time axis. Most of these words have been gathered from the spreadsheet of the General Inquirer Categories (GIC) (“Harvard-IV-4 Psychosocial Dictionary,” n.d.) and, for instance, were selected words such as ‘now’, ‘night’, ‘tonight’, ‘morning’. Some adaptation has been made in the pre-processing stage in order for the sets to have the appropriate discriminative between them relation in quantitative terms (thus I wanted to avoid having Set One much larger than Set Four), and so the word ‘where’ has been put in Set Four (as it has also ‘relational’ character as most of the words-relative pronouns of Set Four do) rather than Set One.

Set Two is composed of the following 20 words: ‘you’, ‘me’, ‘her’, ‘thou’, ‘this’, ‘he’, ‘thee’, ‘him’, ‘she’, ‘we’, ‘they’, ‘them’, ‘us’, ‘these’, ‘none’, ‘those’, ‘himself’, ‘herself’, ‘themselves’, and ‘nothing’. The second set is composed of personal, demonstrative and reflexive pronouns or words that show an index (e.g. ‘nothing’, ‘none’). So, with the production of the second set we can theorise the actors, agents situated on the axis of Set One (space-time). Set Two’s

words have also been widely used as stylistic markers by many other researchers (Mosteller and Wallace 1963, 280–81; Elliot and Greatley-Hirsch 2017, 145; Burrows 2002, 272), they are high frequency words and they assist with the gender discrimination, identity information and narrative perspectives of the protagonists or heros of the play.

Set Three is composed of 53 words and divided into three subsets. Set Three refers again, as Set One, to the content analysis words included in the General Inquirer Categories or to words that are congruent mainly with the three Osgood's axes since they show the wider lexical context of the play-text and bring out the active or contrasting nuances as represented, for example, by the antonyms of 'strong' and 'weak' or 'black' and 'white', or 'life' and 'death'. In fact, Set Three is composed of three subsets. Subset One is of 'physical properties' including colour, such as 'black' or 'white' (extracted from the existing category of colours in the GIC), but also includes words that relate to some of the human senses, such as 'touch'. Subset Two is composed of hyponyms that denote body parts (and are also selectively extracted from the category of body parts in the GIC), such as 'hand', 'hands', 'brain', 'face', or members of a 'family' (this is an invention of the present investigator), such as 'son', 'daughter', 'wife', 'man'. As for Subset Three, this includes words which indicate mainly environmental setting or Heraclitus' elements, such as 'fire', 'water', 'sea' and 'earth' or metaphorically 'life' and 'death'. (This adaptation calls for the application of the metaphorical machinery of meaning according to the Russian Formalists (Steiner 1984, 39–137).) Heraclitus was a Pre-Socratic Greek philosopher living around 500 BC. He was the first philosopher that discussed both physical-cosmological and metaphysical concepts. Heraclitus developed theoretical claims about opposite concepts, such as 'earth' and 'sea' (and metaphorically 'life' and 'death'), and their balanced relation in cosmological terms and regarding the symbolism of the transformation of material and the implied features of a unique divinity that controls the cosmos (Graham, n.d.; Barnes 1982). The last (third) subset of Set Three constitutes the building, for reasons of stylometric experimentation, of a set of non-function words with a philosophical connotation. (The search for the counts of words' frequencies was based on strings and so there was not any kind of adaptation--lemmatisation for cases where the word 'present' denoted the gift and not the time, and that's why this word--together with 'past'--was included in the unified third set indicating a more general meaning, such as the environmental setting.) Subset1 is composed of 16 words: 'white', 'black', 'red', 'yellow', 'colour', 'colours'/colors', 'smell', 'odours'/'odors', 'speed', 'touch', 'liquid', 'matter', 'weak', 'weakness', 'strong', and 'strength'. Subset2 is composed of 23 words: 'eye', 'eyes', 'eyed', 'ear', 'ears', 'face', 'faces', 'hand', 'hands', 'throat', 'throats', 'brain', 'bone',

'bones', 'heart', 'hearts', 'part', 'world', 'wife', 'man', 'married', 'daughter', and 'son'. Subset3 is composed of 14 words: 'life', 'death', 'fire', 'earth', 'water', 'sea', 'bright', 'dark', 'present', 'past', 'moon', 'sun', 'shines', and 'shine'. In the extended theorisation (as I built also Set Three) of sets' interrelations, Set Three relates to the actors, agents of Set Two as situated on the space-time theatrical continuum (Set One) of the text.

Set Four is composed of the following 13 words: 'with', 'what', 'when', 'since', 'which', 'where', 'who', 'within', 'whom', 'whose', 'wherein', 'things', 'people'. Set Four is composed mostly of relative pronouns that can be viewed in the form of the interrelationships of the actors, agents of Set Two, as situated on the space-time continuum (Set One) and with the characteristics ('environmental setting') or attributes of Set Three. Relative pronouns introduce relative clauses and they provide a specific kind of information for the actors, agents of Set One and they can be, generally, employed in order to clarify information relating to a person, thing, situation or micro-chronological period. The four sets' words connect the events or actors, agents in relation to the setting or the place where these events take place. Furthermore, most words of the Set Four, as those of Set One, have been widely employed for stylistic discrimination by other researchers (Mosteller and Wallace 1963, 280–81; Elliot and Greatley-Hirsch 2017, 145; Burrows 2002, 272).

Therefore, the selection of the four sets' words is generally based on the effort to formulate a mixed theoretical approach that treats all contained words as a unified grid of 'lexico-semantic' interrelated markers of the style of a playwright and writer. Set Two and Set Four are composed mainly of grammatical, function words and Sets One and Three are mostly of words that have some lexical meaning or indicate an attribute. Since the words of Set Three are words that relate to content analysis and they constitute the investigator's choice according to some criteria (weakness or strength of meaning, antonymic relation, meronyms-hyponyms in the context of a mereological interrelation or metaphorical theorisation of Heraclitus' concepts), let us go back mainly at the grammatical-function words and check if these words of our sets have also been markers of other researchers in the stylometric field (Mosteller and Wallace 1963, 280–81; Elliot and Greatley-Hirsch 2017, 145; Burrows 2002, 272).

Mosteller and Wallace (Mosteller and Wallace 1963, 280–81) use the following words, most of them high frequency function words:

more, our, **their**, then, **thing**, this, what, when, which, with, who, your, where, **matter**, they, us, himself, his, we.

The following words are used by Elliott and Greatley-Hirsch (Elliott and Greatley-Hirsch 2017, 145) as markers, though with their homographing categorisation:

he, her, here, him, himself, his, if, most, none, nothing, now, past, she, since, than, them, themselves, then, their, them, themselves, there, these, they, this, those, thou, you, us, we, what, when, where, which, who, whom, whose, with, within.

Moreover, Burrows in the development of the Delta procedure (Burrows 2002, 272), as can be viewed in the only existing list of the 35 of the total 150 words-markers, has used for discrimination:

his, with, he, their, her, you, they, we, this, when.

Only the first 35 words are tabulated in the Burrows's article but it is likely that his full list of 150 words has many in common with our Sets Two and Four, as pronouns generally play a key role in the discrimination process of the Delta method.

There also exist other alternatives for the selection of the words. It is possible, for example, to neglect lexical words or words of philosophical connotation and, instead, build the four sets from previously used lists of function words or from lists of words that could be created after measuring the counts of all the Shakespearian corpus and forming sets based on various strata criteria as formed by the examined and modelled author's style. And as it is true that alternative ways of making the sets exist, two other lists of words for the formation of sets of words could be employed, thus one third of the 221 function words used in Elliott and Hirsch (Elliott and Greatley-Hirsch 2017, 145) or a large group of the 165 function and lexical words employed by Mosteller and Wallace (Mosteller and Wallace 1963, 280–81) or, preferably, a combination of these two (anyway, both lists have words in common) in order to avoid the limitation of using only function words and ignoring completely the words that an author is employing consciously. (Why should we completely ignore the conscious use by an author of lexical words since author's intentionality can assist in discrimination?)

The sets of words have been generated according to generic criteria avoiding a monolithic tendency (as would be the case if someone selected a number of the most frequent corpus' words in order to form Set Two, and then formed other three Sets with words whose counts for instance range on average from 1 to 5 counts, from 6 to 10 and from 11 to 20

counts). The only concern for other researchers should be to follow the sets' quantitative distribution adopted in this thesis, for example putting in Set Two words that on average would almost equal or exceed slightly the total counts of all three remaining sets in order for Set Two to be treated as a special case. The present investigator is claiming that the approach in this thesis is a holistic approach and takes into account a poly-strata criterion (though not setting clear limits of ranges of counts as Burrows does), since the selection of sets entails a mixture of elements-criteria and the Russian Formalism-based devices (played by the four sets) adds validity to our method of words' selection. Therefore, the implication is that the selection of words for general Fuzzy-Logic-based modelling of style should not include only grammar-function words but should be generated based on some generic criteria avoiding being strictly corpus orientated.

Overall, the main goal was to construct sets of words with different attributes, whether these concern grammatical categories or the classification of the words as very frequent, moderately frequent and rare (Burrows 2007). By counting set-wise--so that an occurrence of any word in a set is tallied as a hit for that set—I am measuring the appearance of whole semantic categories at once, allowing types within each category to stand for each another. Thus, the occurrence in a play of either 'day' or 'night' will have the same effect of increasing by one our count for words in Set One.

Each word type in the six sets occurs multiple times in each of our extracts of Shakespearian writing; for example, 'if' occurs 42 times in the extract from *A Midsummer Night's Dream*. Each of these 42 occurrences is called a 'token' so that the type 'then' has a token count of 42 for that 10,000-token extract. In the event, the total token count for all the occurrences of all the types in the four sets did not exceed 17% of the token count of each extract (that is, it was under 1,700 tokens).

The selected words-types were all of them in their uninflected form, and few of them in both the singular and plural form, for example, 'night' and 'nights'. The words were counted without considering their grammatical role, for instance, 'face' or 'faces' were counted both as nouns or verbs because they were searched for as alphabetical strings and not words in the linguistic sense. In a simple cross-check based on a manual search and investigator judgement, in almost all 27 Shakespearian plays more than 90% of the occurrences of 'face' are nouns. It might be argued that in plays of the sixteenth and seventeenth century the words that function as nouns and verbs, as in our case with the lexemes of 'face' and 'hand', would probably be employed by authors mainly as nouns probably due to their descriptive power in the context of scene. Though there has not been

much research yet carried out for this kind of lexemes-words, and such a specialised issue exceeds the purpose of our studies, in principle it can be argued that these words were most frequently used in a literal body-based sense. Generally, as also George J. M. Lamont implies in ‘The Historical Rise of the English Phrasal Verb’ (Lamont 2005), the production of phrasal verbs in English language was influenced by the North German language and though their use started to rise in the sixteenth century, this was the case for only a specific pool of verbs.

I count for each set the number of tokens in the extract that are types in that set, so for Set One I sum the counts of the occurrences of the tokens ‘then’, ‘if’, ‘here’, and so on, in an extract, and I express this as a percentage of all tokens in the extract. So, for example, if I find in a 10,000-token extract that there are 200 tokens corresponding to the types in Set One, I record Set One contributing 2% of the tokens for that extract. If for another 10,000-token extract I find Set One contributing 300 tokens that is recorded as 3% for Set One for that extract. By measuring the sets’ counts and adopting a set-wise approach it is feasible to detect the different strata of sets’ counts (rather than individual words’ counts), the magnitude of the Set Two, which is treated as a special case (with the major contribution), and the ranking of the other three Sets whose counts are fluctuating around approximately 2% to 3%. In that manner, the focus is not merely on the words’ counts but on a set of words’ counts, and, therefore, even if the words’ counts differ from play to play they can be identified with the Shakespearian style. Consequently, this method lays emphasis on the set-based Shakespearian style and the compensation of certain words’ counts by others in the various modelled well-attributed plays of Shakespeare.

When the corpora of texts were built, a specific text analyser program, AntWordProfiler (see Section 7.3.2 of Technical Appendix), was employed for the grouping of words into sets and the measurement of the counts of the percentage of each set and the words’ frequencies. AntWordProfiler, developed at the Faculty of Science and Engineering at Waseda University in Japan (Laurence 2014), is a useful tool that aims at profiling statistically the vocabulary level and evaluating the complexity of texts. After the data points of the words and sets of words of plays of the corpora were gathered, the AntWordProfiler software provides the results in a default customised format for the whole text that was entered in a spreadsheet in order to derive the measures of central tendency and spread (standard deviations). These measures provide important information about the data points of counts of words grouped into sets. The extracted data and their representation through the central tendency and spread constitute the prior knowledge upon which the decision-making hinges. The decision making in this experimentation can be phrased as a compound question,

namely, ‘Is the new play (under scrutiny in Validation Stage 2 and Testing stage) of Shakespearian style and to what degree?’.

2.5 Null and Alternative Hypothesis

The null hypothesis (H_0) in this research is that there is no causal relation between the dependent and independent variable. The dependent variable in this computer simulation is an output variable that expresses in comparison with the modelled data of known Shakespearian plays of each genre the degree of Stylistic index of Similarity (SiS) of new, disputed, co-authored or of other authors’ plays to be tested. The independent variables are the counts of the sets of words-input variables. The alternative hypothesis (H_1) is that there are similarities regarding the writing style of a specific author and these can be uncovered by exploring through an automated Fuzzy-Logic program not only the counts of distinct words but also the counts of sets of related words. Even if the selected words are not all distinguishing stylistic markers, each Shakespearian play’s correlations of counts of words grouped into sets might indicate an authorial style. An unexpected large increase or decrease of counts of one set’s words in one play can be compensated with a proportional increase or decrease of other words of the same set. So, the research explores the possibility that many differentiated factors of authorial modelling exist, and instead of relying on the predominance of particular words as markers I test whether I can reliably infer authorship from the perhaps unconscious, more or less consistent rate of use of sets of related words.

2.6 Methodological Validity

The building of two distinct corpora, one composed of the well-attributed Shakespearian plays and one of anonymous, strongly disputed or non-Shakespearian plays (such as the well-attributed Marlovian play *The Jew of Malta*), based on the same mode of regularisation performed by the same academic researchers, provides us with the first positive index, and a test of the practicality of using pre-processing of the textual corpora in a way that avoids bias. Instead of selecting plays from different digital libraries and with the modernisation made by various people, I selected balanced, regularised texts with the same criteria and from the same source. Effort has been also put regarding the stylistic markers to avoid the selection of words which are associated with extensive lemmatisation or spelling

variations, although of course the methods used here could easily be adapted for use with different word-sets if the present ones are considered less than ideal.

Overall, the research methodology was articulated from the beginning based on the need for primary quantification of the counts of selected words and the fuzzification of these counts. Moreover, the multi-level character of the research design facilitates the straightforward assessment of the plausibility of the SiS results (carried out in Chapter Four and Seven), and it is feasible to check if these SiS results are consistent with the theoretical claims of the historical scholarship.

2.7 Data Analysis: Advantages and Limitations.

The research design and the integration of all the components of the fuzzy classifier is an effort to provide a tool of natural mathematics and approximate reasoning for decision-making of a stylometric nature. The fact that such an engineering method has not previously been applied to the problems arising from the plays of Shakespeare and his contemporaries is part of this thesis's original contribution to knowledge. By employing Fuzzy-Logic, Set Theory and propositional calculus this thesis manages to extract authorship verdicts by evaluating in a fully automated environment and in the form of an expert system the descriptive statistics and correlational evidence of sets' counts and counts of words' frequencies inside the frame of uncertainty and ambiguity, and this in a creative way. This creativeness arises from the fact that with the use of partial truths membership the effect of 'ambiguity' and 'uncertainty'--generally, identified with negative signs--can be estimated and, therefore, taken into consideration when dealing with complex problems, such as authorship attribution of a play written in a distant era, and concepts characterised by a degree of definitional relativity, such as style. In other words, by applying the principles of Fuzzy-Logic it is attainable to avoid the high risk that entail the precise mathematical models in case they fail. Another contribution of this thesis is the plausibility of the argument that the use of sets of words in combination with the counts of words' frequencies can assist more efficiently stylistic discrimination than simply the use of words as markers.

An important advantage of building of such a fuzzy and universal stylometric model is that it can be employed in many ways and in many cognate fields. In psycho/neuro-linguistics, input variables of sets of words indicating neuroticism might indicate from their writing style if someone is neurotic ((Moshe, Schler, and Argamon 2008). In addition, in other fields, such as forensic science, gathering data points (sets of words

expressing anger or aggressiveness) from social networking (posts on Facebook, Twitter messaging, email messages, short texts) can help experts classify individuals as potential threats of committing crimes. At the same time, with the use of different stylistic markers as input variables, such as the ratio of lexical density, it is possible to assess the degree of stylistic variety of a student's essays and her vocabulary at risk (Yule 1944), which thus assess some aspects of cognitive abilities. Instead of building an ad hoc stylometric tool, the present method might enable creation of a new automated universal stylistic classifier of any kind of written text and for various purposes.

2.8 Objectives of Research Design

The primary objective of the research design is to integrate as components of an approximate-reasoning-based automated stylistic classifier the descriptive statistics of the collected data points, the principles of Fuzzy Logic and the algebra of Set Theory in order to investigate the relationship between specific stylistic markers used as inputs and to estimate the output variable which I already mentioned, that is, the Stylistic index of Similarity (SiS). The second objective is to design the investigation in such a way that after the completion of the structure of the model to make it possible when feedback is available from the validation stage to go back and optimise the technical parameters such as the design of memberships functions of inputs, and adjust the rule-base in assessing high-Shakespearianness. The third objective of the research design is to ensure that with the validation and testing process--by experimenting with new plays with various characteristics (anonymous, co-authored, well-attributed plays of other authors)--it is possible to assess the robustness and rationality of our Fuzzy-Logic-based authorship attribution model.

3 Chapter Three: From Micro-Engineering to Stylometry, A paradigm of Primary Experimentation.

In this chapter, there is a description of the design and the function of a fuzzy fan controller, intended to convey the principles of Fuzzy Logic. The sensors of the fan receive input information about the room's temperature and humidity and the microchip of the fan is programmed in a Fuzzy-Logic-based mode in order to automatically control the atmosphere of the room and to provide comfort for human occupants of the room in the most economical way. In the second (3.2) and third subsection (3.3) of this chapter and, in detail, in the Textual Addendum (7.2), by analogy to the design of the fuzzy fan controller there is an analysis of the steps for modelling a Fuzzy-Logic-based Stylistic Simulator-Classifer of Shakesperianness for disputed, anonymous or other authors' well-attributed plays. This primary experimentation is not included in the main part of the thesis since it is not our proposed model and for reasons of space. On the other hand, it assists the reader, before diving into complex technical concepts, to understand through a primary example the application of Mamdani-Type-1 Fuzzy-Logic reasoning for resolving problems of authorship attribution.

3.1 A Functionalist Approach of a Fuzzy Fan Controller.

As was discussed in the Introduction, lots of complex problematic situations in life cannot be defined according to the binary logic simply as 'black or white', 'on or off', 'tall or short', 'ill or healthy'. We think in fuzzy concepts that treat some phenomena as partially belonging to particular sets or classifications. We say that a comedian is somewhat funny but also a bit coarse, that Friday night is sort of the weekend but strictly a week-night, that a cup of coffee is rather too black for our taste but basically acceptable, and so on. Perhaps surprisingly, there is a branch of computer science--called 'Fuzzy' Logic--that manages to systematise our thinking in this way and apply it to practical decision making. Fuzzy Logic can be applied to the problem of determining the authorship of writings for which the author is unknown or disputed and in which we have counted features such as the frequency of use of certain words.

To illustrate the ideas, let us start with the how we think about temperature. In Fuzzy Logic, instead of making a sharp determination that some temperatures are ‘high’, we instead say that a certain temperature’s membership of the ‘high temperature’ class takes a value between 0 (not at all a member) through 0.5 (half-ways a member) to 1 (fully a member). So, a temperature of 35 degrees Celsius has a membership of 0.6 to the class of ‘high temperature’, being on the low side of this class: a somewhat higher temperature would be more fully a member (see Figure 3). I can add the class of ‘medium temperature’, and naturally enough the same temperature of 35 Celsius has a different membership for this new class: 0.2 membership, being on the high side of this class: a much lower temperature would be more fully a member. I could go on adding classes, such as ‘low temperature’, but for this demonstration we will use just these two. It is obvious that individuals will disagree about what is hot or warm, cold or warm, as these are overlapping judgements including objective and subjective elements. A Fuzzy Logic system is a way to embody a set of such judgements and use them for a practical purpose.

Suppose that the system for which our Fuzzy Logic system is being developed is to keep a room comfortable by controlling a fan set in the ceiling. If we are measuring just one relevant value, here the temperature, it is necessary to decide which membership function value to use for the temperature 35 degrees Celsius: the 0.2 membership of the ‘medium temperature’ class or the 0.6 membership of the ‘high temperature’ class. Naturally, from these numbers the temperature of 35 degrees Celsius can be said to be more fully a ‘high temperature’ (at 0.6 membership of that class) than a ‘medium temperature’ (at 0.2 membership), so it is reasonable to apply the rule ‘take the maximum’ in such a case of overlapping membership functions.

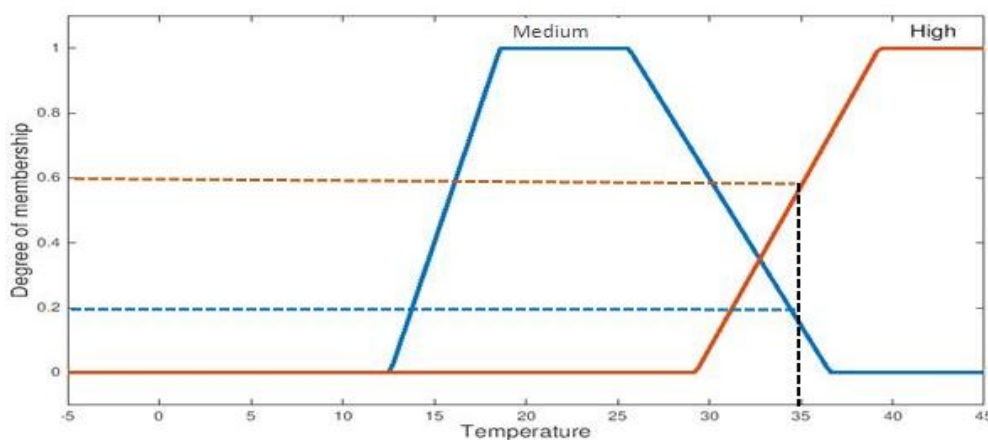


Figure 2: Membership function of temperature.

Of course, I might have more information to work with than just the temperature: I may also, for example, measure the humidity of the air in the room. For this measurement I may again have overlapping membership functions for ‘medium humidity’ and ‘high humidity’ and again I could apply the ‘take the maximum’ rule and, in this case, say that the humidity of 18 grams of water per cubic meter of air belongs more to the ‘medium humidity’ class (at a membership of 0.8) than it belongs to the class of ‘high humidity’ class (at a membership of 0.4).

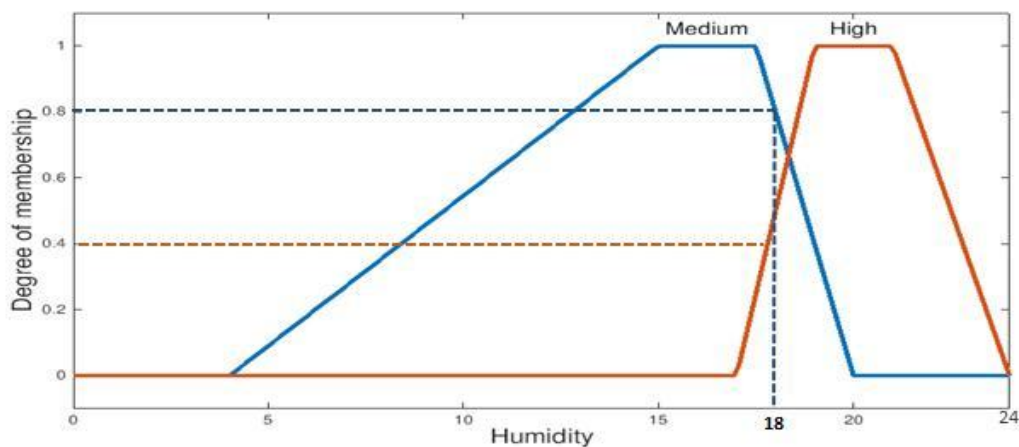


Figure 3: Membership function of humidity.

So far, then, I have derived from our input temperature of 35 degrees Celsius a value of 0.6 as the membership of the class of ‘high temperature’ and from our input humidity of 18 g/m³ a value of 0.8 as the membership of the class of ‘medium humidity’. The next question is: ‘how do we combine these numbers to control the fan?’ Our instinct might be to take the higher value again, but a little reflection using worst-case-scenario thinking shows why this would be a mistake. If some temperature value gave us a membership of 1.0 of the class of ‘high temperature’ and some humidity value gave us a membership of 0.6 class of ‘high humidity’, the fan should not be running at its highest speed because things could still get worse: the humidity could still go up. Likewise, if some temperature value gave us a membership of the class ‘high temperature’ of 0.7 and some humidity value gave us a membership of the class ‘high humidity’ of 1.0 this again is not the worst case: the temperature could still go up. (The term humidity and the issue of recommended maximum indoor humidity are much more complex than presented here (see Section 7.3.3 of Technical Appendix).) The worst case would be both inputs being at their peaks, and it is necessary to

reserve some fan power for that event. To leave ourselves some power to for a worse situation, we might reasonably combine such inputs using a ‘take the minimum’ rule.

Before I get to the rules by which a Fuzzy-Logic system might combine the inputs of temperature and humidity to control the speed of a fan to keep a room comfortable, it is necessary to apply one more concept. The notion of a ‘membership function’ applies quite intuitively to the classifications (‘medium’ and ‘high’) that I use for the inputs to our system (the temperature, the humidity). And I can also apply this notion to the output, in this case the speed (measured in revolutions per minute, RPM) of the fan. Let us suppose that I have two membership functions for the fan’s speed, one I call ‘medium speed’ and one I call ‘high speed’ and that, as shown, they do not overlap.

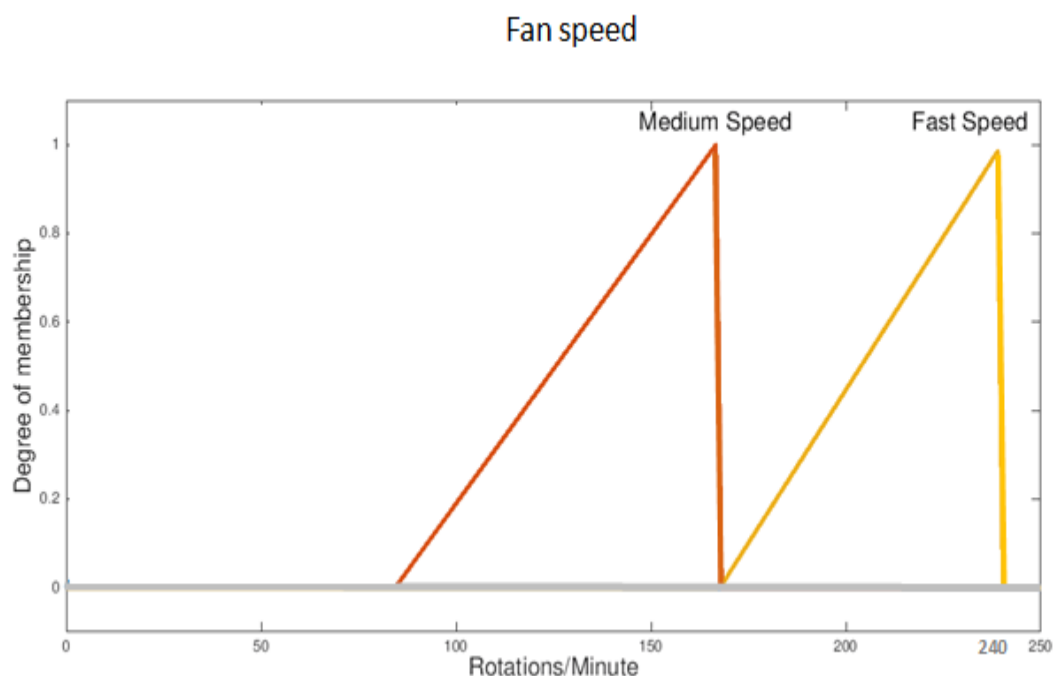


Figure 4: Membership function of output variable, fan’s speed.

To use membership functions for the output (the fan speed) it is necessary to think in reverse. Instead of starting with a value on the X-axis (temperature or humidity) and deriving from it a y-value, here I want to derive a value on the X-axis from (the speed at which I should run the fan) from known values on the Y-axis. How can I combine the y-value inputs to derive an x-value output for the fan speed? We have seen two ways that we might combine y-values. we can ‘take the maximum’, which may be a reasonable thing to do when combining y-values taken from overlapping membership functions because the corresponding

x -value most fully ‘belongs’ to the class for which it has the highest membership function. And we can ‘take the minimum’, which may be a reasonable thing to do when combining y -values from different inputs since, by the worst-case-scenario principle, we want to leave ourselves some headroom to raise the fan speed whenever there is at least one input below 1.0.

For this illustration I am only going to use the latter of these two methods, the ‘take the minimum’ method applied to two inputs. The ‘take the maximum’ method is mentioned only to show that in more complex situations it might be useful. Instead of selecting from the values given by overlapping membership functions using a ‘take the maximum’ principle, I am instead going to allow the rules to select which membership function supplies the y -value I use to control the output. I will define one or more rules that govern the relationship between the inputs and the output. To see how the inputs operate with the rules, I am going to consider separately three distinct scenarios. In the first, there are 2 inputs and 1 rule; this entails the application of a single rule to our present case of 2 inputs (temperature and humidity). Then we will see how multiple rules may operate at once but simplify the situation by having only 1 input to consider: this is the 1 input and 2 rules scenario. Then multiple inputs can be combined with multiple rules: the 2 rules and 2 inputs scenario.

Scenario: Two inputs, one rule.

Remember that our temperature is 35 degrees Celsius and our humidity is 18 g/m^3 . These inputs give us two membership values for temperature (0.2 membership of the class of ‘medium temperature’ and 0.6 membership of the class of ‘high temperature’) and two membership values for the humidity (0.8 membership of the class of ‘medium humidity’ and 0.4 membership of the class of ‘high humidity’). The single rule is:

If temperature is ‘high’ and humidity is ‘medium’ then fan speed is ‘high’

In the rule, the word ‘is’ means ‘has a greater-than-zero membership function value for that class’. Both parts of the rule are satisfied: the membership of the ‘high temperature’ class is indeed above zero (it is 0.6) and the membership of the ‘medium humidity’ class is also above zero (it is 0.8). Because the rule is satisfied, the fan speed class selected is ‘high speed’.

Looking at its antecedents, the rule specifies the ‘high temperature’ class and the ‘medium humidity’ class, so these are the classes for which our membership functions provide the inputs: I will be using 0.6 membership of the ‘high temperature’ class and the 0.8 members of the ‘medium humidity’ class. As it happens, these would also be the values I would use if I was applying the ‘take the maximum’ procedure to the values produced by overlapping membership functions, since 0.6 is the larger of 0.2 and 0.6, and 0.8 is the larger of 0.4 and 0.8. But that is not how we are selecting our inputs: the rule tells us that we use the memberships of the ‘temperature high’ and ‘humidity medium’ classes and we never need to apply a ‘take the maximum’ procedure.

How do I get from this selection of a fan-speed membership function to an x -value on the fan speed scale? First, I apply the ‘take the minimum’ principle that I justified above using the worst-case-scenario thinking. So, of 0.6 and 0.8 I take the minimum, which is 0.6. I draw a horizontal line on the fan speed graph at the point $y=0.6$ and use this line to chop the top off the triangle representing the ‘high’ fan speed.

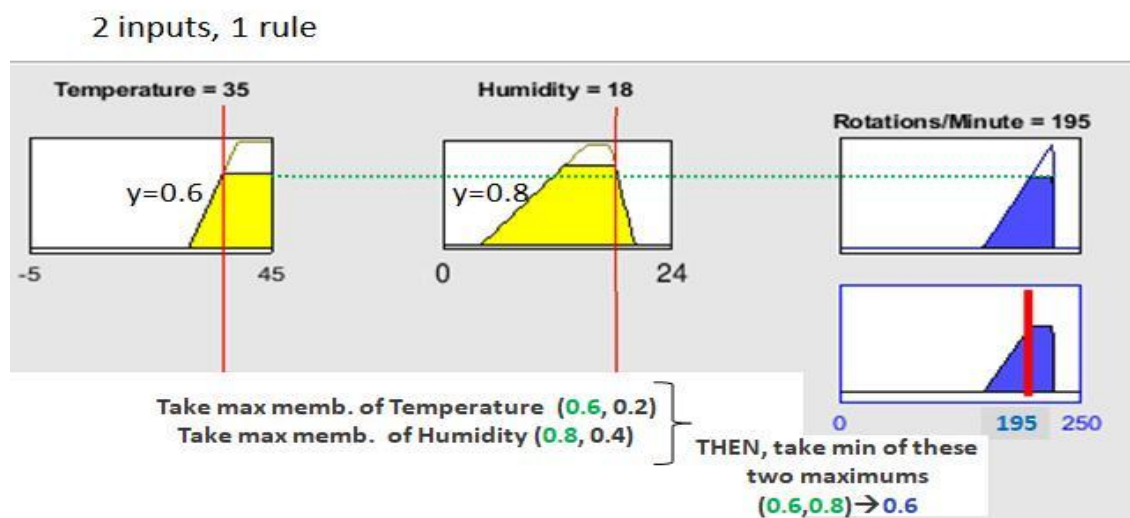


Figure 5: Two inputs, one rule.

The resulting trapezoid has a centroid, equivalent to its centre of mass if the trapezoid were made of a thin sheet of material of uniform density. (Informally, this is the point where we could hold such a sheet between pinched fingers and the sheet would have no inclination to rotate because no ‘side’ would be heavier than any other.) This centroid has an x -value, so a line dropped from the centroid to the X -axis gives us the optimal fan speed corresponding to these two inputs and this one rule. Behind the calculation of the centroid there is the complex

operation of differential calculus (see Section 7.3.4 of Technical Appendix. (In practise, for trapezoids and triangles the use of two respective mathematical formulae is adequate for the production of the centroid. Calculus is used for complex, curve, functions but since Matlab employs a generalised built-in tool for the production of the centroid for all (11) pre-built functions, I stick with that terminology).

Scenario: One input, two rules.

As before, the temperature is 35 degrees Celsius but, in this scenario, I omit the humidity reading. This 35 degrees Celsius reading gives two membership values for temperature: 0.2 membership of the class of ‘medium temperature’ and 0.6 membership of the class of ‘high temperature’. The two rules are:

If temperature is ‘high’ then fan speed is ‘high’

If temperature is ‘medium’ then fan speed is ‘medium’

As before, a rule selects a fan speed membership function. But whereas before there was only one rule and hence only one fan speed membership function selected, here there are two rules and both are satisfied: the membership of the ‘high temperature’ class is indeed above zero (it is 0.6) and the membership of the ‘medium temperature’ class is also above zero (it is 0.2). Thus, both fan speed membership functions are selected, but, as we shall see, they are active to different degrees.

1 input, 2 rules

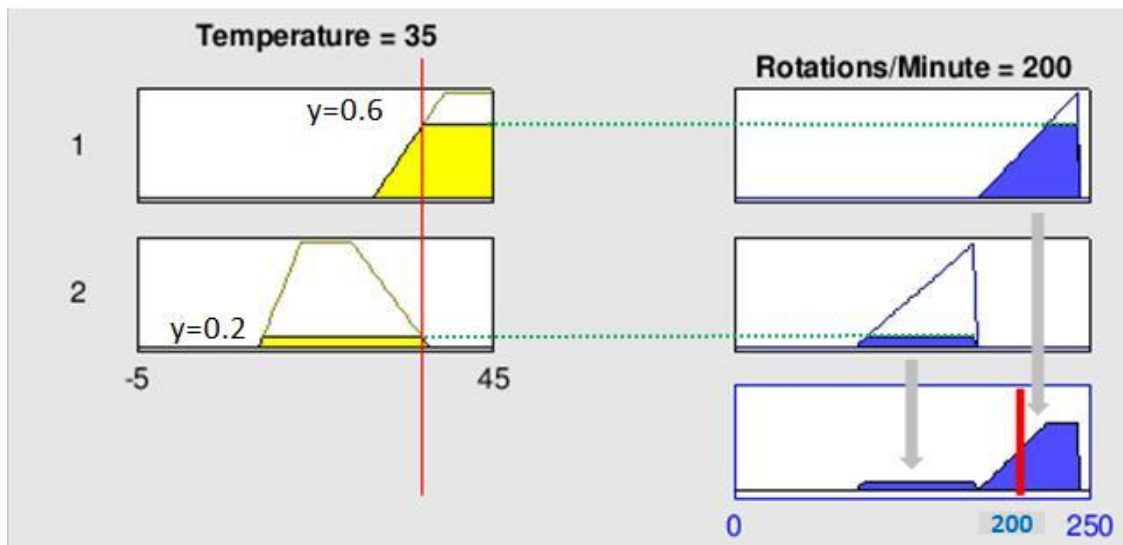


Figure 6: One input, two rules.

I apply to each fan speed membership function a truncation derived from the membership function specified in the antecedent(s) named in the rule that selected that fan speed membership function. Thus, the first rule selects the ‘high speed’ membership function for the fan, and this membership function is chopped off by a line drawn at $y=0.6$ because that is the extent to which 35 degrees Celsius belongs to the ‘high temperature’ class named in the rule’s antecedent.

I apply the same logic to the second rule, which selects the ‘medium speed’ membership function for the fan, and this membership function is chopped off by a line drawn at $y=0.2$ because that is the extent to which 35 degrees Celsius belongs to the ‘medium temperature’ class named in the rule’s antecedent. I now have two trapezoids (truncated triangles) in the fan speed graph and I combine them so that the entire area shaded in blue in the above figure is treated as a thin piece of material of uniform density for which I take the centre of mass, or centroid. From this centroid I drop a line to the X-axis to find the optimal fan speed for this single input using this one input and these two rules.

Scenario: Two inputs, two rules.

As before, the temperature is 35 degrees Celsius and the humidity is 18 g/m^3 . These inputs give two membership values for temperature (0.2 membership of the class of ‘medium temperature’ and 0.6 membership of the class of ‘high temperature’) and two membership

values for the humidity (0.8 membership of the class of ‘medium humidity’ and 0.4 membership of the class of ‘high humidity’). The two rules are:

If temperature is ‘high’ and humidity is ‘high’ then fan speed is ‘high’

If temperature is ‘medium’ and humidity is ‘medium’ then fan speed is ‘medium’

Each rule specifies a fan speed membership function and since both rules are satisfied, both fan speed membership functions are selected. But since each rule contains two antecedents, because I am considering two inputs, it is necessary to decide which of them gets to truncate the fan speed membership function specified by the rule. The solution is that I ‘take the minimum’ of the inputs, on the worst-case-scenario principle.

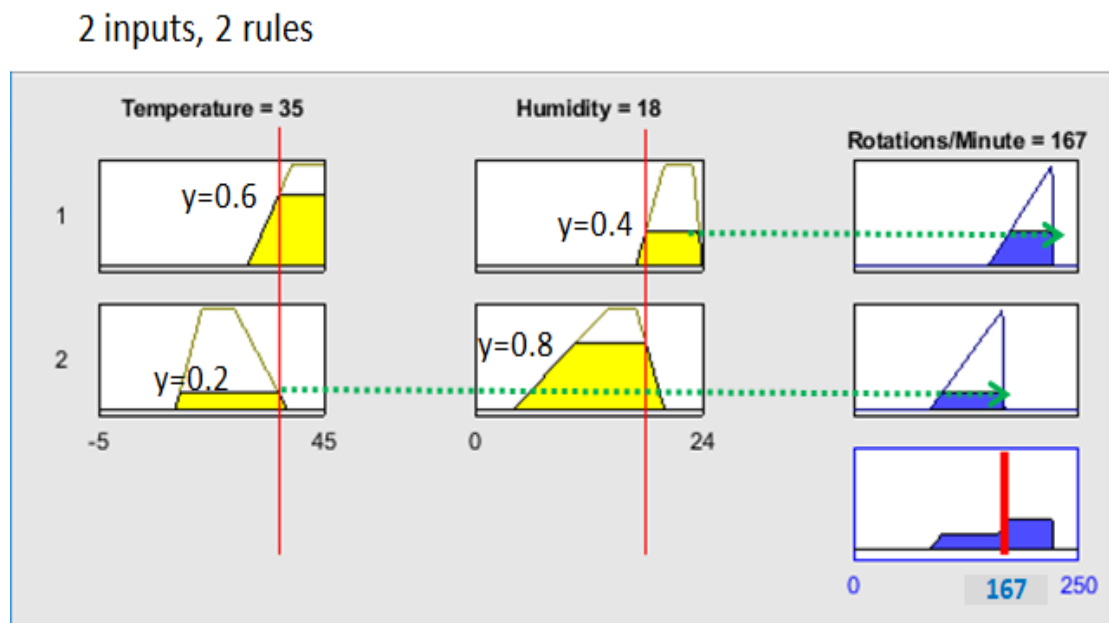


Figure 7: Two inputs and two rules.

Thus, for the first rule I take the minimum of 0.6 (the degree to which 35 Celsius is a ‘high temperature’) and 0.4 (the degree to which 18 g/m³ is a ‘high humidity’), which is 0.4. So, I truncate the fan speed ‘high’ membership function at y=0.4 to produce a trapezoid. Then I do the same for the second rule. I take the minimum of 0.2 (the degree to which 35 Celsius is a ‘medium temperature’) and 0.8 (the degree to which 18 g/m³ is ‘medium humidity’), which is 0.2. So, I truncate the fan speed ‘medium’ membership function at y=0.2 to produce a trapezoid. As before, I combine these two trapezoids to make a single shape, take its

centroid, and drop a line from the centroid to the *X*-axis and read off from the *X*-axis the optimal fan speed for these two inputs and these two rules. Through the explanation of the design of a fuzzy-fan controller I describe by analogy in the Section 7.2 how can be built a holistic (irrelevant of genres) stylistic Fuzzy-Logic-based classifier of Shakesperianness. The simplified example used here is to make for better comprehension of the genre-based fuzzy stylistic classifiers in the next chapter.

3.2 Design of Membership Functions and Manual Clustering of Data Points.

To illustrate how I apply the method used for the fuzzy fan controller with two variables and two membership classes, let us imagine the data of six sets' counts as detected in 15 plays instead of nine in the primary experimentation (7.2) and 12 (for comedies), six (for histories) and nine (for tragedies) in the core experimentation of Chapter Four. That would give us for Set One data like this:

Set One: 7.1, 6.9, 7.5, 6.6, 6.6, 7.4, 7.4, 6.3, 6.1, 6.9, 7.3, 6.9, 7.0, 8.2, 8.6

Data are gathered for the other five sets in a similar way. Regarding Set One, note that each data point refers to one play's proportion of words belonging to that set. Note also that I round to one decimal place. I need now to generate membership classes from these numbers, and there are two steps to the process: the first creates trapezoidal classes based on triplets of three data points or/and duplets of two data points in a set being close to one another, and the second uses the remaining data (if there are any) to create a triangular class below the lowest trapezoid and, if there are data left, another triangular class above the highest trapezoid. (If there are more than three data points that are within 0.1 of each other [as in Set Six: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7], then these data points are not considered as triplets or duplets but they are represented through a triangular membership function-class with the exception of the constraints described below (see Section 7.3.5 of Technical Appendix). The algorithmic steps can be described in terms of two sequential steps, as follows.

Step One: Find a triplet (= three data points) in the set data that are within 0.1 of each other, as in [6.9, 7.0, 7.1] and [7.3, 7.4, 7.5] above. For each, the lowest value produces the top-left corner of the trapezoid (at $x=6.9, y=1.0$) and the highest value produces the top-right corner of the trapezoid (at $x=7.1, y=1.0$). To get the bottom left and bottom right corners (where

y=0) I gather all the data that fall within the range of the triplets, so [6.9, 6.9, 7.0, 7.1]--notice that there are four data points within the range of this triplet--and calculate their mean and Standard Deviation (SD). The bottom left corner of the trapezoid has an x -value that is one SD below the lowest value in the triplet (= the x -value of the top left corner) and a y -value of zero and the bottom right corner of the trapezoid has an x -value that is one SD above the highest value in the triplet (= the x -value of the top right corner) and a y -value of zero. I repeat this step, crossing out all data values I have used in the calculation of the SD, until I can find no more triplets in the data. In the above example I would find another triplet at [7.3, 7.4, 7.5] and use it to create a trapezoid and cross out the data points [7.3, 7.4, 7.4, 7.5]. I label the two trapezoids 'Shakespeare Medium 1' and 'Shakespeare Medium 2'. When I have found all the triplets (and there may be none if the data points are widely spaced), I repeat the process searching for duplets: pairs of data points also no more than 0.1 apart. In a duplet, the lower of the two values forms the x coordinate of top-left corner of the trapezoid (whereas in a triplet this was formed by the lowest of the three values) and the higher of the duplet's two values forms the x coordinate of the top-right corner of the trapezoid (whereas in a triplet this was formed by the highest of the three values).

Step Two: The remaining data points, if any, must fall below or above the trapezoids I have created in Step One, since I crossed out all the data points that fall within the range of each triplet or duplet. Starting with the points below the lowest trapezoid, I pool them to create in this case the set [6.6, 6.6, 6.3, 6.1]. I calculate the mean and SD of this set and the mean gives the x -value of the apex of an isosceles triangle of which the y -value is one. One SD below this mean gives the x -value of the bottom left corner of this triangle (at $y = 0$) and one SD above this mean gives the x -value of the bottom right corner of this triangle (at $y = 0$). This deals with all the data points below the lowest trapezoid, and I label the result 'Shakespeare Low'. Then I repeat the process for all the remaining data points, which will fall above the highest trapezoid, and I label this last triangle 'Shakespeare High'. So, in the end, I have one or more trapezoids and possibly a triangle below them and possibly a triangle above them, as represented in the next figure.

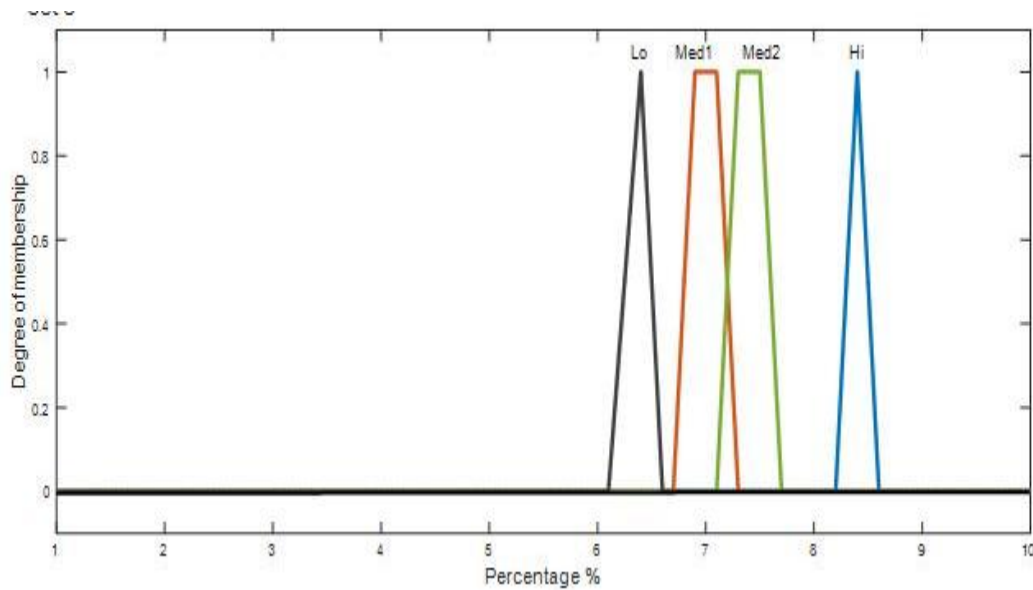


Figure 8: Triangular and trapezoidal (triplets, duplets) patterns.

Notice also that the classes may overlap as in Figure 8, but need not: it depends again on how spread out the data are. In general, this algorithm for deriving classes from data points is the present investigator's invention based on descriptive statistics and the principles of Fuzzy Logic. That is, the algorithm is an aspect of this thesis's originality.

It is possible for a set's data points to yield no triplets or duplets, in which case the above algorithm produces no central trapezoids and it becomes meaningless to speak of triangles being formed above or below the central trapezoid(s). If there are no triplets and duplets, then, with the exception of the above constraints of single, isolated x -values (a), all data points, for the set are represented by a single isosceles triangular class. In this case (sets 4, 5, 6), all data points are summed up and their mean gives the x value of the apex of an isosceles triangle of which the y -value is one. One SD below this mean gives the x -value of the bottom left corner of this triangle (at $y = 0$) and one SD above this mean gives the x -value of the bottom right corner of this triangle (at $y = 0$).

Standard Deviation is employed here because I am attempting to use mean values that represent an author's actual word usage (derived from the author's known works) to derive realistic expectations of that author's likely word usage in other works I have not examined. Standard Deviation captures the 'spread' of data around a mean value, and hence is a measure of the author's willingness to depart from the habit that is represented by that mean value. In all the experimentations in this thesis I employed the sample Standard Deviation instead of the population Standard Deviation as the former estimates the Standard Deviation for the

whole population including an estimator of bias since for the creation of each membership function I explore a small sample from the whole population (that is, data points of all well-attributed Shakespearian plays). Furthermore, it seems that the population Standard Deviation would more often than the sample standard deviation be smaller than 0.1 and might be less helpful for the differentiation of classification and the creation of membership functions, since I look for decimal patterns. The sets' counts are measured by the AntWordProfiler Software in the first decimal point at the first level.

Below there are a few more basic constraints--mainly activated in the experimentation of Chapter Four--for the selection and design of membership functions. The methodology of the design of membership functions exploits the properties of values-data points' neighbourhood connectivity and defines neighbourhood constraints in the context of a kind of constraint-based local search technique and combinatorics (Hentenryck and Michel 2005, 3–10), where combinatorics here denote the combinations of values-data points in finite sets-classes.

These constraints are activated if the spread of the data allows it:

a) If an x -value, a set's data point, is the only point more than 0.1 below or above any of the other data points and these 'other' remaining data points do not constitute duplets or triplets but form an open continuous interval of decimal values, then instead of a unique triangular class for all data points, one triangular function is formed for this single isolated value and a second triangle for the other remaining data points. (This interval includes real numbers as integers or numbers in first decimal, since the AntWord Profiler Software produces at first level the output in first decimal. See also the definition of Open and Closed intervals in Section 7.3.1 of Technical Appendix.) In that case, the single x -value plays the role of mean of two hypothetical values--one below ($- 0.1$) and one above ($+ 0.1$) the mean--within the distance of 0.1 and constitute the apex of a separate membership function/class, an isosceles triangle of which the y -value is one and whose bottom left and right corners are one SD above and one SD below this hypothetical mean. (Note regarding the rule 'if an x -value, a set's data point, is the only point more than 0.1 below or above any of the other data points' that there might be, depending on the spread of data, a continuous interval of decimal points and one 'isolated' value before that interval plus one 'isolated' value after that, in which case I have to form two hypothetical single x -value-based triangles.). If the distance of a single x -value to any other sets' data point

above or below is more than 0.5 and there is more than one value before or after that 'isolated' value (so that these 'more than one' values can form a duplet, triplet or a non-hypothetical triangular function), again a separate single-point triangular class is created for the single, isolated x -value that plays the role of mean of two hypothetical values--one below $[-0.1]$ and one above $[+0.1]$ this actual value-hypothetical mean and constitutes the apex of a separate membership function, an isosceles triangle of which the y -value is one and whose bottom left and right corners are one SD above and one SD below this hypothetical mean. If there exist only two consecutive 'isolated' data points, then a triangular class is formed based on these data points--their mean constitutes the apex--with the differentiation again of the additional use of internal single-point-based 'not/none' classes in the core experimentation (see Section 4.11.1 on 'not/none' classes). This differentiation will be discussed in Chapters Six and Seven with the comparative assessment of the genres-based and the primary holistic model. (In the rare case of widely spaced data, such as with three or more consecutive data points distancing from themselves more than 0.5, a proper adaptation has been made based on the principles of the constraints described here. Only one such case was met, and it is extensively analysed in the design of the membership functions of the comedies-based fuzzy stylistic classifier). In general, the use of hypothetical triangles has been adopted in order to deal efficiently with the outlier data point(s).

The necessity of such a design derives from the fact that the continuous sequence of values (real numbers in first decimal) or the maintenance of an upper threshold (0.5) of their proximity is more congruent with the building of Shakespearian patterns using multiple classes than the expression of the central Shakespearian stylistic tendency by a single triangular class would be.

b) When there is a validation stage (and this is the case in the core experimentation) if the SD above and below the mean of the data points of any triangular class (since this cannot happen for trapezoidal classes) does not cover the data points of the modelled plays then the second SD (instead of the first SD) below and above the mean is employed for the creation of this class. This extension from the first SD to the second SD for the design of the bottom left and right corners of a class can be called the principle of full coverage of the modelled data points. (If a known play's data point is not covered by the two SD then no further action is taken, though this is not common). When a set-input variable contains classes that are made using only on the first SD below and above the mean of the data

points that contributed to the classes' formation, then for reasons of uniformity of 'notness/noneness' the bottom corners of all 'not/none' subclasses essentially start and end where the x -values of the first SD of the neighbouring actual class, if any, has a membership of zero in the actual class and a membership of one in the 'not/none' class. (This has been the case in primary experimentation). On the other hand, if at least one of the classes of a set-input variable has been designed based on the two SD below and above the mean of the data points that contributed to the formation of the class (the principle of full coverage), then all 'none' subclasses of this set-input variable start 0.01 before the first SD bottom right corner of the left neighbouring actual class and end 0.01 after the first SD bottom left corner of the right neighbouring actual class. (This has been the main case in the core experimentation.) With this adaptation of 0.01, actual classes overlap with 'none' subclasses, even if the first SD is adequate for covering all data points, and the system manages to form the memberships of the x -values in the fuzzy 'none' subclasses based on the second SD or first SD of the data points of any left or right neighbouring actual class. This depends on which SD was selected for the formation of the actual class's bottom corners (as said, this is the case if at least one of the classes of a set-input variable has been designed based on the two SD below and above the mean of the data points). With this adaptation of fuzzy 'none' subclasses, the system compensates with the use of second SD, instead of the first SD, and the creation of larger 'none' areas in each variable. (See also the discussion about the avoidance of fuzzy singletons in Section 7.3.14 of Technical Appendix.)

Regarding the above constraints (a) and (b) in connection to the formation of triangular functions treated as special cases and the constraint about the selection of the first or second SD, the rationale is that the design of the membership functions should be associated with the way each sets' data are spread, and this distribution may be influenced by the number of plays in the corpus and the counts of the sets. With sets of small counts, as in the primary experimentation sets with on average counts around 0.5 % (Sets Three and Six) or around 1% (Set Five), it might be easier for the data to form a continuous interval of (more than three) values (that is, real numbers with a precision of first decimal), and this denotes that there is an ordered continuous variation in the data, and this ordered variation allows them to be represented in terms of the central stylistic tendency rather than in terms of stylistic patterns. When data points of the plays of the corpus form a continuous interval, it is logical to expect either the formation of duplets or triplets (Set One and Three in primary

experimentation) or the formation of a first SD-based single triangular function (Set Four, Five and Six in primary experimentation) that expresses the central (Shakespearian) stylistic tendency of a set's data points. The central stylistic tendency (set-input variable with first SD-based single actual class) is expected to be represented when the data are spread out and each data point is far from the others but with a distance from 0.1 to 0.5 (inclusive).

So, regarding the design of the membership functions-classes I apply a kind of manual dense, mean and first or second SD-based clustering.

Based on the fact that in the primary experimentation (7.2) the data points that contribute to the formation of classes Set One and Two are either covered by the first SD (actual duplet and triplet points, last triangular function of Set Two) or do not include any x -values meeting all the relevant constraints, and the fact that Set Three to Six are represented by one actual class (and so one 'not' actual class), there is no need to employ in the primary experimentation two SD for the design of the membership functions and so the design is less complex than that of the core experimentation in Chapter Four.

3.3 Design of Inference Mechanism-Rules and Output Classes of the Fuzzy Expert Systems-Stylistic Classifiers.

Creating the inference mechanism and building the database of rules of the Fuzzy Expert Systems-Stylistic Classifiers is a complex process that is described extensively in the primary experimentation of Section 7.2.2 and in each of the three genre-based fuzzy stylistic classifiers of Chapter Four but in general there are three basic steps. The first step is to map the combinations of the input variables' membership functions that represent each time the counts of every play's data points. I build one rule for each combination, and if any of the current combinations coincide, duplication is avoided. If a new play that I test provides data that, for each set, falls within the shapes for that set and in doing so 'selects' shapes to form one of these actual combinations then we should say that it belongs to the membership class of highest Shakespearianness. As such, it has what I will call the highest possible Stylistic index of Similarity (SiS). That is what the rules must achieve: they must select the SiS-High shape in the output membership function for any play that matches the known Shakespeare plays in this way. The second step is through a search and harvesting procedure to add rules from a much larger set of possible combinations with actual and 'not/none' actual classes as it is theoretically possible for a new play to mostly match the scores from the known Shakespeare plays, but perhaps for just one or some of the input variables (counts of sets of

words) to fall outside the shapes for that set. In that case a new play is not quite in the class of most Shakespeare-like ('SiS-High') but falls instead into a lower output class (say, 'SiS-Medium') according to an algorithm that assesses how many, how much and which input variables-sets' counts are Shakespearian or not. (There is one algorithm for the core experimentation of Chapter Four (4.5) and one for the primary experimentation described in the second addendum of the Technical Appendix (7.2.2)). The second step has been initiated in order to add naturalness to the inference mechanism and to ensure that this mechanism is fully functional and does not trigger rules that a new play's data points should not. The third step is to check the data points of each new play of the experimentation and see if its data points are mapped by the current set of rules, as produced in Steps One and Two. If the current rules map them (which is very rare), then there is no need to further extend the database of rules. If it does not, it is necessary to map the new plays' data points according to the membership functions of the modelled known Shakespearian plays and the algorithm that assesses the data points' actual Shakesperianness or/and 'notness/noneness'. (I say 'or/and' because, as will be discussed in the various parts of the core experimentation, new plays' data points may have as a result the production of more than one rule that might play off another.) And this process (Step Three) goes on with every play that I add during the experimentation (Validation Stage and Testing Stage) with new plays.

As for the output variable, this expresses the degree of the (Shakespearian) Stylistic index of Similarity (SiS) of the play to be tested and it is measured in the range of 0.00 to 1. The output variable of the fuzzy model, as shown in Figure 9, is represented by four trapezoids, the classes of the Low, Medium1, Medium2 and High Stylistic index of Similarity (SiS).

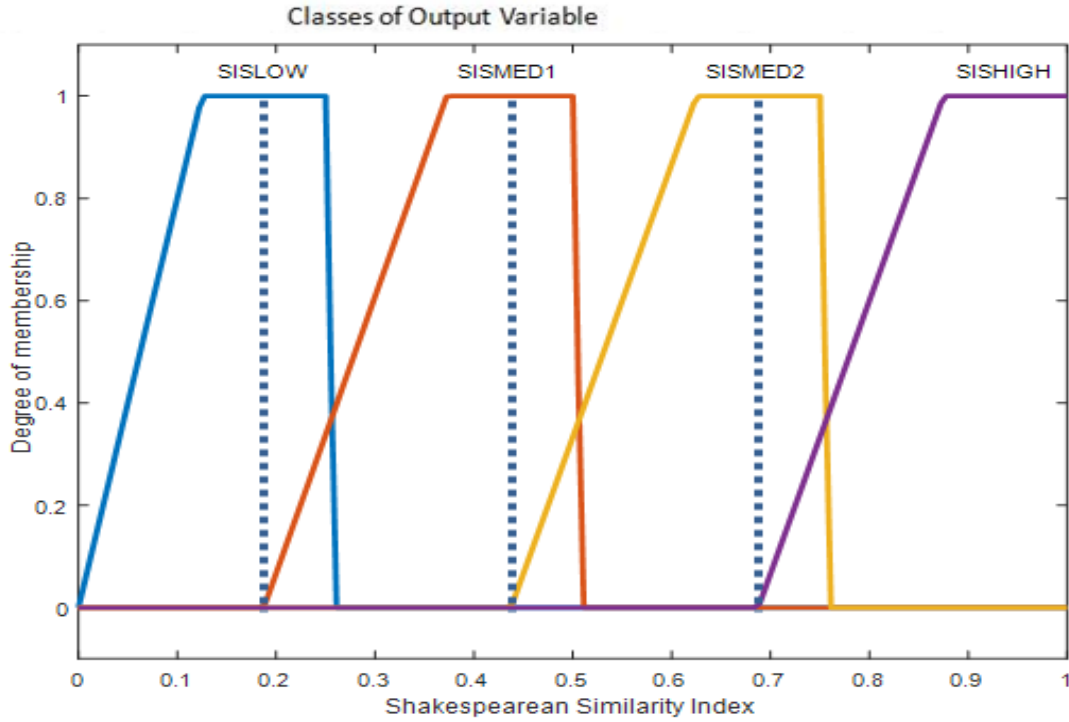


Figure 9: SiS and four classes of output variable.

These four terms are called fuzzy quantifiers. Each of the four classes corresponds on the X-axis roughly respectively to a quartile of the number of 1. The design is based on four class intervals, which overlap, in terms of x -values: $[0-0.25]$, $[0.1875-0.50]$, $[0.4375-0.75]$, $[0.6875-1]$. The top-right corners ($y=1$) of the four trapezoids correspond on X-axis to the values of **0.25**, **0.50**, **0.75** and **1**. These four values define the four respective quartiles of the number one. As can be viewed in the figure above, the right lateral sides are not vertical but rather slope a little because I have added 0.01 to the x -value of their bottom-right corners in comparison to their top-right corners to give 0.26, 0.51, and 0.76. This adjustment with the formation of the little slope has a minor effect on the production of results and assist us in keeping the naturalness of a (non-right) trapezoid avoiding the sharpness of the vertical line which breaks sharply the distribution of membership degrees for the x -values to the upper-right corner ($y=1$). The top-left corner of each trapezoid has a y -value of 1 and an x -value that is the midpoint of the quartile it represents, so 0.125 (halfway between 0 and 0.25), 0.375 (halfway between 0.25 and 0.5), 0.625 (halfway between 0.5 and 0.75), and 0.875 (halfway between 0.75 and 1).

The bottom-left corner of the first trapezoid starts at $x=y=0$, and for the other three trapezoids the y -value is of course 0 and the x -value is derived by dividing in half again the

quartile represented by the preceding trapezoid. So, for the second trapezoid I look to the first trapezoid, whose upper-left corner has an x -value of 0.125 and whose upper-right corner has an x -value of 0.25. I take the mid-point of this range and set the bottom-left corner of the second trapezoid at $x=0.1875$. For the third trapezoid, I look to the second trapezoid, whose upper-left corner has an x -value of 0.375 and whose upper-right corner has an x -value of 0.5. I take the mid-point of this range and set the bottom-left corner of the third trapezoid at $x=0.4375$. For the fourth trapezoid, I look to the third trapezoid, whose upper-left corner has an x -value of 0.625 and whose upper-right corner has an x -value of 0.75. I take the mid-point of this range and set the bottom-left corner of the fourth trapezoid at $x=0.6875$. This method of deriving the bottom-left corner of each trapezoid by reference to the mid-point of the upper edge of the preceding trapezoid implements a version of what is called the ‘mean of maxima’ principle. In the fuzzy fan controller, the design of the two classes of the output variable was more straightforward and there were not any overlapping areas between them. In the stylometric experimentation, the mean of maxima of the upper side of the first, second and third trapezoid is the mid-point in the upper side of each trapezoid. In other words, the mean of maxima is a point whose distance to the upper-left corner equals the distance to the upper-right corner of the same trapezoid. For instance, the mean of maxima for the first trapezoid is 0.1875 and it is the central point between 0.125, the upper-left edge point, and 0.25, the upper-right edge point. The means of maxima for the second and third trapezoid are respectively 0.4375 and 0.6875.

Technically, I employed the quartiles in order to divide at a first stage the range of SiS-values into four classes, each of which contains 25% of the total SiS-values. Each bottom-right endpoint of the first three quartiles is also the median of all values between two consecutive closed intervals (see Section 7.3.1 of Technical Appendix). In fact, at a second stage I formed three of our four trapezoids (the second, third and fourth) through an interquartile division by dividing each of the four initial quartiles into two sub-quartiles. In conclusion, by employing natural mathematics it was feasible to construct systematically and symmetrically four overlapping classes of SiS. Much dictates the design of membership functions. Their overlapping depends on many factors too and there is not any *a priori* optimum solution. Here, for our interquartile divisions, I ascertained that there is a ‘sufficient overlap’ which approaches ‘25 to 50 percent of the classes’ bases’ (Negnevitsky 2011, 124–25, Cox 1999). In other words, around 25-50% of each output class-set is overlapped by another neighbouring set-class.

The reason I selected the divisions of quartiles can be illustrated by analogy with what is known as the coffee cup size problem (Asquith 2018). The three categories, small, medium and large cup of coffee have come to dominate the industry because historically it proved difficult to treat all kind of coffees in exact patterns. Any standardisation must take into account many variables, such as the kind of coffee, their combination with milk, the strength of the coffee and, finally, the concept of coffee from the scope of what constitutes a large, medium and small cup of coffee in the mind of different individuals doing the coffee in their own house (and in different geographical locations of the world). There were also other problems of definitional nature, such as the fact that ‘the original cappuccino was five to six ounces’ but a ‘cup with one ounce of espresso, and four to five ounces of milk, measures five to six ounces – that’s roughly 150 mL to 180 mL’ (Asquith 2018).

In the same logic by using four output classes of which three overlap it is feasible to apply natural maths and the approximation logic, whereas the technique of overlapping exploits the symmetrical properties of the interquartile division creating indirectly new additional output classes. The objective from the beginning was to form three-four basic output classes in order to detect high Shakespearianess and low Shakespearianess with high certainty and to create two classes of intermediate Shakespearianess. Furthermore, data points in textual analysis and digital stylometry are not characterised by the preciseness of environmental indicators, such as temperature, and, therefore, trapezoids were selected for the design of the output variable because their upper side conforms to the detection on x -axis of areas-group of values (and not a single value) that share a high membership degree (y -axis) in the respective trapezoidal class.

Of course, the experimentation in Chapter Four is more complex than that of the primary experimentation, but the explanation is analytical, too. Reading a full example of the primary experimentation in Section 7.2 will help the comprehension of how the principles of Fuzzy Logic are applied.

4 Chapter Four: Three Genre-Based Fuzzy Stylistic-Classifiers of Shakespearianness-- Modelling and Validation-Testing Stage.

This chapter contains the core experimentation and the modelling of three Fuzzy Stylistic Classifiers of Shakespearianness. These three Fuzzy-Logic-based classifiers are based on the well-attributed, sole-authored comedies, tragedies and histories of Shakespeare. In principle, I adopted the methodology that was described in the previous chapter. Simultaneously, I make a few slight technical adjustments based on feedback from the primary experimentation (7.2). Each of the three models is composed of three basic stages. First, there is a design of possibly overlapping class-membership functions of each input, the data points of each set of words, secondly, there is the formation of actual combinations of these classes in the well-attributed plays of each genre, and thirdly the generation of rules by associating the class-membership functions of inputs' sets with the output SiS classes. It needs to be highlighted that the different sequence or proximity of the data points of each set in each genre has as a result the different parametrisation of the design of the respective models.

4.1 The Comedies-Based Fuzzy Classifier of Shakespearianness: Design of Membership Functions-Classes.

Let us now start building the first, comedies-based, Fuzzy Stylistic Classifier of Shakespearianness. This model is based on the 12 well-attributed, sole-authored comedies written by Shakespeare.

COMEDIES (12 plays)

As You Like It

The Comedy of Errors

Love's Labour's Lost

The Merchant of Venice

The Merry Wives of Windsor

A Midsummer Night's Dream

Much Ado about Nothing

The Taming of the Shrew

The Tempest

Twelfth Night

The Two Gentlemen of Verona

The Winter's Tale

For the building of this model, I employed as stylistic discriminators the four sets of words that are described in Sections 2.4 and 7.2.1, plus one more expressing the total of four sets' counts, and I represented them in triangular or trapezoidal functions by following the same steps as in the primary experimentation. As before, the shapes of the membership classes are derived from the numbers of the data points of each set and the way these data are spread. First I create trapezoidal classes based on triplets of three data points and/or duplets of two data points in a set being close to one another, and then I use the remaining data (if there are any) to create a triangular class below the lowest trapezoid and, if there are data left, another triangular class above the highest trapezoid with the exception of the constraints a) of Section 3.2. When I have found all the triplets (and there may be none if the data points are widely spaced), I repeat the process searching for duplets: pairs of data points also no more than 0.1 apart. In a duplet, the lower of the two values forms the x coordinate of top-left corner of the trapezoid (whereas in a triplet this was formed by the lowest of the three values) and the higher of the duplet's two values forms the x coordinate of the top-right corner of the trapezoid (whereas in a triplet this was formed by the highest of the three values).

So, the first concern in the examination of the data points of all the comedies' scores for each set is to search for triplets and duplets that are within 0.1 of each other. In other words, each of the 12 comedies provides only one data point for Set One, one data point for Set Two, and so on, so that to get the membership functions for each set it is necessary to combine the data for all the comedies. This is explained below in the formation of the combinations of the actual classes. When there are triplets and duplets, I apply trapezoidal membership functions for the modelling of the class of data points of the sets' counts. As before, if there are more than three data points that are within 0.1 of each other, as in Set One (2.1, 2.2, 2.3, 2.4, 2.5, 2.6), then these data points are not considered as triplets or duplets. We should also recall the constraints described in the design of the membership functions (Section 3.2, constraint a). Note in the table below that the *Love's Labour's Lost* data points of Set One, Two and Five are x -values that play the role of mean (of two hypothetical values--one below and one above--within distance 0.1) and constitute the apex of an isosceles triangle of which the y -value is one and whose bottom-left and bottom-right corners are one SD above and one SD below the mean. (See also Section 7.3.14 of the Technical Appendix

regarding the adaptation of 0.01 for the values of corners in order to avoid what are called fuzzy singletons.) Data points in cells that have the same colour and form duplets or triplets in columns of Set Two, Four and Five constitute trapezoidal classes. In fact, there is one triplet for Set Two, one duplet for Set Four and two triplets for Set Five. Data points of Set One and Set Three form respectively two and one triangular classes.

Comedy title	Set One %	Set Two %	Set Three %	Set Four %	Total percentage of the counts of the four sets %
<i>As You Like it</i>	2.6	9	1.8	2.5	15.9
<i>The Comedy of Errors</i>	2.3	9.4	1.7	2.6	16
<i>Love's Labour's Lost</i>	1.9	6.2	1.5	2.5	12.1
<i>The Merchant of Venice</i>	2.4	8	1.6	2.7	14.7
<i>The Merry Wives of Windsor</i>	2.2	9.2	1.5	2	14.9
<i>A Midsummer's Night</i>	2.5	8.3	2.1	2.6	15.5
<i>Much ado about Nothing</i>	2.1	9.9	1.5	2.1	15.7
<i>The Taming of The Shrew</i>	2.2	9.6	1.6	2.1	15.5
<i>The Tempest</i>	2.6	8.8	1.9	2.8	16.1
<i>Twelfth Night</i>	2.4	9.1	1.9	2.3	15.7
<i>The Two Gentlemen of Verona</i>	2.5	9.1	1.5	2.5	15.6
<i>The Winter's Tale</i>	2.3	8.3	1.8	2.8	15.2

Table 2: Data points of 12 well-attributed Shakespearian comedies.

As can be seen in the table above, for input Set One there are no unique triplets and duplets. Leave out for the moment the single value (1.9) in the cell of *Love's Labour's Lost's* row and of Set One's column. The other 11 data points in green cells are summed and the mean (2.37) and standard deviation (SD) are produced. The mean gives the x -value of the apex of an isosceles triangle of which the y -value is one. In the primary experimentation, the spread of the data did not call for the use of two SD and so in all classes of the six sets the first SD was employed and one SD below this mean gave us the x -value of the bottom-left corner of this triangle (at $y=0$) and one SD above this mean gave us the x -value of the bottom right corner of the triangular functions (at $y=0$). Here we should remind ourselves of the fact that in certain cases I extend the base of the triangle to two SD above and below the mean. As described above in (b), when the first SD above and below the mean does not cover all the data points that contributed to the formation of a class (and this can happen only for triangular classes) and this class is not the single class of the set-input variable then the second SD (instead of the first SD) below and above the mean is employed for the creation of this class.

Frequently I had to employ the second SD as all the sets-variables were represented by more than a single class and there were cases in which the first SD was not adequate for the creation of patterns. Clearly, the primary experimentation was proved to be a paradigm of central Shakespearian stylistic tendency (only first SD), whereas the spread of the data of the corpus of the 12 comedies calls for the creation of Shakespearian stylistic patterns (second SD if the first SD is not adequate for the principle of full coverage as described in Section 3.2, see constraint b). Such a case of creating stylistic patterns is particularly that of the membership functions of Set One (composed of words that denote space and time) and Two (composed of personal, demonstrative and reflexive pronouns) in the current experimentation. The use of the two SD is made for the design of the second triangular class of Set One in order not to leave out of this triangular class the data points of 2.1 and 2.6, which are more than one SD from the mean for the set. Sets One, Two and Five have many data points that form continuous sequences (pi-function) of values in first decimal but also a single, isolated x -value as I defined this concept in the constraints of the design of membership functions in the primary experimentation in Section 3.2, constraint a). In Set One by creating a separate single x -value based triangular class, I manage to have a second triangular class that represents the patterned style for the 11 plays and that will attribute a non-zero membership value for any actual Shakespearian count. I label that class '1b' because there is a data point (1.9) that precedes that class.

I likewise extend to two SD any triangular classes of the other sets with the same criterion-constraint of complete data points' coverage as explained in the primary experimentation. Therefore, for Set One I formed also a second triangle which graphically precedes the triangular function just described ('1b'). The data point of 1.9 gives the x -value of the apex of an isosceles triangle of which the y -value is one. Though there are no other actual data points, 1.8 and 2.0 are the hypothetical values that are within 0.1 of the left and right of 1.9. I used these two values in order to produce the bottom-left and bottom-right corners of a second triangle. In this single-point-based new triangle, the single x -value (as with 1.9) plays the role of a mean between two hypothetical continuous x -values in first decimal (one below and one above 1.9, that is, 1.8 and 2.0). This mean is the apex of an isosceles triangle of which the y -value is one.

In every similar case of the five sets with this kind of triangular classes, explained with the description of the two relevant constraints (a), the corners of this triangle are one SD below and one SD above the mean of the two hypothetical values. Let us explain the element of this fuzzy representativeness in Set One. In the first set, with the three data points (1.8, 1.9, 2), the actual 1.9 and the hypothetical 1.8 and 2.0, the mean is 1.9 and the first SD below 1.9 is 0.1 and the first SD above 1.9 is again 0.1. So, the two hypothetical values are 1.8 [$1.9 - \text{SD} (= 0.1)$] and 2 [$1.9 + \text{SD} (= 0.1)$] and so a separate single-point triangular class is created for the single, isolated x -value that plays the role of mean (of the two hypothetical values--one below [-0.1] and one above [$+0.1$]). This actual value-data point of *Love's Labour's Lost* constitutes essentially a distinct single-point-based pattern and it is the apex of a separate membership function, an isosceles triangle. I do not need to use here two SD as 1.9 is the data point of the only play that contributes to an actual triangular class and the principle of full coverage is self-fulfilled. The SD below this mean gives the x -value (1.8) of the bottom-left corner of this triangle (at $y=0$) and one SD above this mean give us the x -value (2.0) of the bottom-right corner of this triangle (at $y=0$).

This adaptation with the triangle of hypothetical values and the use of two SD for the class of '1b' has two major advantages. First, I eliminated the effect of the non-existent data point of 2.0, as none of the plays had that count for Set One. A second advantage is that by using two SD in the next triangle and so more spread corners, I have an overlapping area of the two triangles that constitute the classes '1a' and '1b' of Set One. Consequently, the data point of '2' does not need to be represented by a 'not/none' class but in fact shares a low membership in the first triangular class and a zero membership ('the elimination effect') in the second triangular class of Set One. (See also the slight adaptation of 0.01 for the corners

of the actual classes in Section 7.3.14 of the Technical Appendix). This non-exclusiveness is an interesting element of fuzziness and so the spread of data of the 12 Shakespearean comedies allow the differentiation of the current design with that of the experimentation of the previous chapter. As highlighted above in the description of the first constraint (in Section 3.2), the methodology of the design of membership functions lies in the properties of values-data points' neighbourhood connectivity and neighbourhood constraints and it resembles the constraint-based local search technique (Hentenryck and Michel 2005, 3–10) by employing a kind of combinatorics (that is, the combinations of value) for the creation of (finite) actual classes.

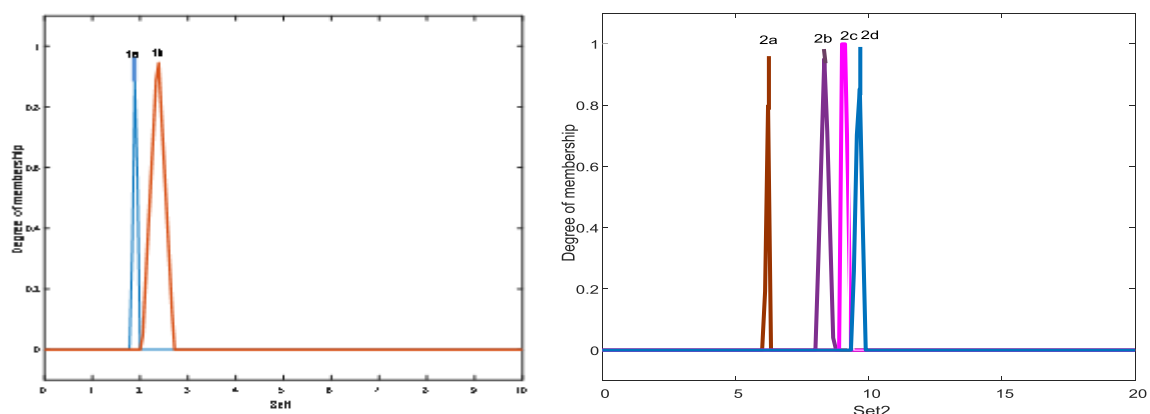
Let us now continue with the description of the design of the classes of the second input of the first fuzzy model. In Set Two, as can be viewed from the yellow cells of the above table, there is one triplet (9.0, 9.1, 9.2 for the values 9.0, 9.2, 9.1, 9.1). For each triplet, the lowest value produces the top-left corner of the trapezoid (for input of Set Two at $x=9.0$, $y=1.0$) and the highest value produces the top-right corner of the trapezoid (at $x=9.2$, $y=1.0$). To get the bottom-left and bottom-right corners (where $y=0$) I gather all the data that fall within the range of the triplets, so [9.0, 9.1, 9.1, 9.2]--notice that there are four data points within the range of this triplet--and calculate their mean and standard deviation. As described in (7.2.1), the bottom-left corner of the trapezoid has an x -value that is one SD below the lowest value in the triplet (that is, the x -value of the top-left corner) and a y -value of zero and the bottom-right corner of the trapezoid has an x -value that is one SD above the highest value in the triplet that is, the x -value of the top-right corner) and a y -value of zero. This is the class which I label '2c'.

As there are no other triplets or duplets for Set Two I formed two more triangular classes, the one with the single data point of 6.2 as the x -value of the apex of an isosceles triangle of which the y -value is one. I repeat the step I adopted for Set One because in Set Two the next close value to 6.2 is very distant (8.3 and thus more than 0.5), and 6.1 and 6.3 are the hypothetical values that are within 0.1 of the left and right of 6.2. I use these hypothetical x -values (6.1, 6.3) in order to produce the bottom-left and bottom-right corners of a triangle, which I label class '2a', as they precede all other data points of Set Two. Then, I gather the remaining values in order to produce two more triangular classes. The one is located after '2a' and just before the triplet-trapezoidal class of '2c' and the other just after the triplet. The mean of the data points 8.0, 8.3, 8.8, 8.3 is 8.35 and it gives the x -value of the apex of an isosceles triangle of which the y -value is one. Two SD below this mean give us the x -value of the bottom-left corner of this triangle (at $y=0$) and two SD above this mean give us

the x -value of the bottom-right corner of this triangle (at $y=0$). I label this triangle class '2b', as it follows class '2a' and it precedes the triplet, trapezoidal class of '2c'. Finally, a fourth class is formed with the data points 9.4, 9.6, 9.9. The mean of these data points is 9.63 and it gives the x -value of the apex of an isosceles triangle of which the y -value is one. The corners of this triangle are again formed by two SD below and two SD above the mean. I label this triangle class '2d', as it follows the triplet, trapezoidal class '2c' and there are not any other remaining data. So, the data points of Set Two form, in total, three triangular and one trapezoidal class.

I followed the same methodology for the design of classes of Set Three, Four and additional set called Set Five that holds the total counts of the preceding four sets. Set Three is composed of two triangular classes, Set Four of one trapezoidal and one triangular class and Set Five is of two triangular and two trapezoidal classes.

The additional fifth set input to the fuzzy system is critical for creating coherent and accurate patterns of Shakespearian style as far as it concerns his comedies, because there is a high consistency of the total counts of the four sets. In fact, two triplets (15.5, 15.6, 15.7 and 15.9, 16, 16.1) encompass the exact patterns of eight actual Shakespearian comedies, which represent two thirds of the 12 comedies in our corpus. In other words, it can be argued that Shakespeare, when writing comedies, consistently employed the 100 words listed above in the range of 15.6% and 16% of the time, or else in the range of 15.5% to 16.1%. Below, in Figure 10 and Figure 11 there is the graphical display of the membership functions-classes of the five sets.



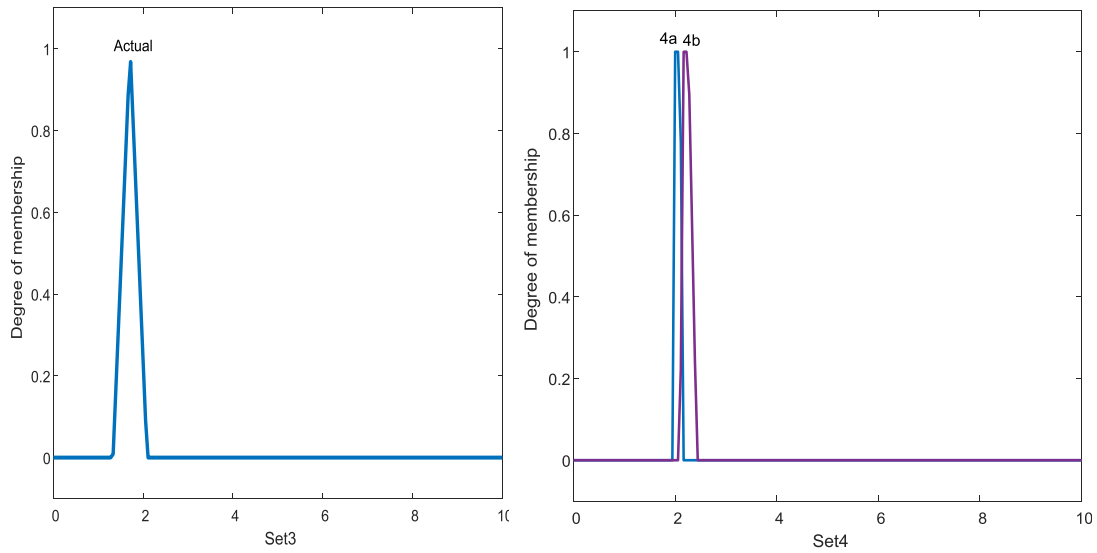


Figure 10: Membership functions of Set One to Four.

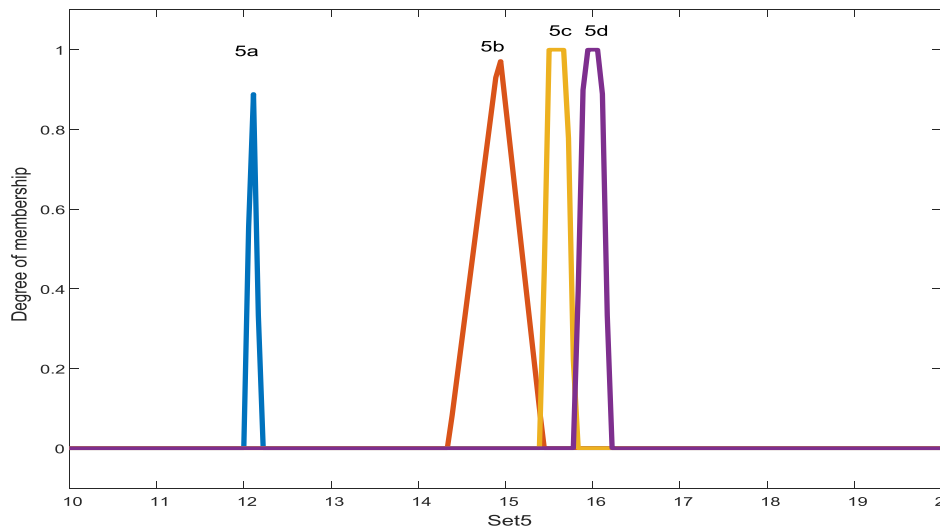


Figure 11: Membership functions of Set Five.

As have been derived the class-membership functions for each set--comprising triangles and trapezoids--from the data points provided by the 12 plays' occurrences of the words in each set from Set One to Set Five, the next step is to proceed to the derivation of the rules by which a new and previously untested comedy's frequency of occurrence of the words in each set is judged to show the degree of Shakespearianness of the new play. The formation of the rules is derived systematically from the same data points that provided the classes and the trapezoidal or triangular membership functions. Each of the 12 plays (across the X-axis of

the matrix) produced an actual data point for each of the five sets (across the *Y*-axis of the matrix).

In order to reduce uncertainty of stylometric conclusions in this experimentation I added a second output in our fuzzy system. This means that I could claim that without deviating from the criteria set in the primary experimentation a multi-input and two output fuzzy system is built for the exploration of Shakespeariness of comedy plays. The only difference is the addition of a second inference mechanism for the SiS-classification, which essentially does automatically what I have done manually in the validation process (Section 7.3.9) by exploring the plays' indices of cosine similarity referring to the 100 individual words-types' counts of frequencies. This second inference mechanism produces the SiS2 score, which, as said, evaluates the first SiS output. I will call the second output SiS2 in order to differentiate it from the initial, SiS1 output. Essentially, the aim was to build two fuzzy models, one for testing new comedies for their Shakespeariness based on the assessment of the counts of the four sets' data points, and so in that stage the initial SiS is produced, and one fuzzy model for validating this initial SiS by taking into account the total of 100 individual words' counts of frequencies and the number expressing the total of the four sets.

The same fuzzy principles of the primary experimentation apply both for the design of the membership functions of the two additional stylistic markers and the further evaluation of the initial SiS, thus the classification of the new plays' Shakespeariness. In fact, it was technically feasible through the inference engine to merge these two fuzzy systems into one. I split the five membership functions of the five sets by building a fuzzy model that has as input variables the counts of Set One, Two, Three, and Four. I added to the same fuzzy system three more inputs, stylistic markers, thus Set Five, SiS result of primary stage, and an index of cosine similarity as detected by the 12 plays (the ideal-average document discussed in validation of Chapter Seven, see Section 7.3.9 of Technical Appendix). These three inputs (Set Five, SiS as Input and index of cosine similarity) and the associated rules can be activated only after inputting to the system the initial produced SiS-result.

Why did I follow such a complex design with the use of a second output? To ensure that high Shakespeariness (SiS-High) does not contradict with other stylistic findings. By exploring general indices I do not deviate from our claim that counting set-wise makes more for a discriminating model. I try to reduce further the uncertainty of our conclusions by applying natural maths and employing as new inputs two additional fuzzy-modelled stylistic markers in situations where uncertainty cannot be eliminated by nature. In other words, by creating a second stage of SiS-evaluation with two additional inputs, I attempt to cope with

the ‘Mountain of Uncertainty’ by walking up the ‘Mountain of Logic’. Such a task is requisite when building a system ‘to act like a human in reaching valid conclusions given valid premises’ (Giarratano and Riley 2005, 343). Besides, the variation that occurs with the addition of the two inputs and a second output, the index of SiS2, is congruent with the principle that a new variation in modelling a problem and postulating a new theory means ‘new variation in a fractal mountain of uncertainty’ (Giarratano and Riley 2005, 343). And though the word ‘uncertainty’ in general causes fear, it is by nature a core principle in Computational Intelligence and Fuzzy Science/Expert Systems. In other words, ‘if there was a certain theory, there would be no uncertainty’ (Giarratano and Riley 2005, 343) and so we would not bother with problems of authorship attribution and ways of resolving them.

In addition, this second level fuzzy filtering, a constituent element of the same method which is based on the fuzzification of inputs and defuzzification of output, assists us in rejecting paradoxical results. For instance, imagine that a data point of a new play/artificial artefact is judged Shakespearian by the four sets-based fuzzy system, but this is only due to a couple of words that amount to the total set’s counts. The SiS2 can detect this completely coincidental event. The structure will be further analysed in the validation stage, the next subsection, but a first glance of the structure of the comedies-based fuzzy model can be viewed in Figure 12 below.

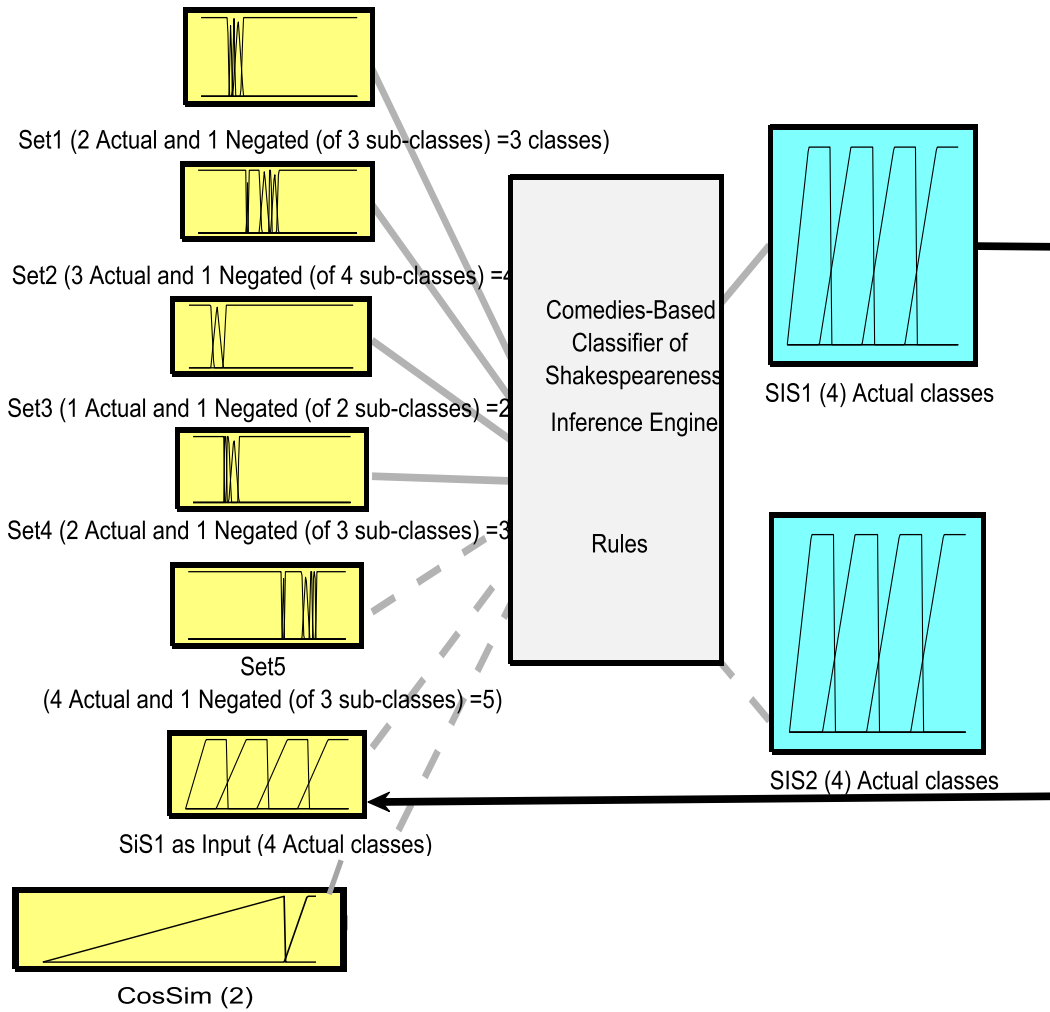


Figure 12: Structure of the comedies-based Fuzzy Stylistic Classifier.

In this core experimentation, the general aim is to work with small and semantically well-defined sets of words, as has been also the case in the holistic model and the primary experimentation, and then to follow a strict design and exploit data points of aggregated counts of multiple sets and multiple words. Another reason for employing Set Five in the second validation stage and without any combination to the classes of the other sets in the 12 plays, but as a general index, derives from the relative standard deviation (RSD) which measures in percentage the spread of the data around the mean. The RSD measures how far the SD is from the mean in terms of percentage of the mean value. The comparison of the RSD of different sets of words can also show which sets' SD of data points has on average in percentage the least distance from the mean. Despite the fact that Set Five does not correspond to a specific semantic set but sums up the total of the four sets' counts, it can play

the role of a general index-stylistic marker, because in statistical terms it has the lowest RSD, meaning it is the set whose data points cluster more tightly around the mean than those of the other four sets.

Statistics	Set One	Set Two	Set Three	Set Four	Set Five
Mean	2.33	8.74	1.70	2.46	15.24
STDEV	0.21	0.97	0.20	0.27	1.07
RSD %	9.01	11.15	11.76	11.17	7.05

Table 3: Five Sets' RSD.

Set Three has the largest RSD. After the end of the experimentation I can assess the importance of the RSD for the selection of the most appropriate sets as stylistic markers. The major principle of the selection of these five sets as stylistic markers is the fact that the connection of the tops of the bars of all RSD values of the five sets should form a symmetric function. In fact, here they form a combination of a (pimf) Π -shaped membership function and a Gaussian-like curve membership function with feet Set One and Five and shoulders Set Two and Four (see black curve in Figure 13). It is also feasible to investigate why such an approach is preferable to another option, namely the possible reshuffling of a set's words so as their data points form a curve approach a straight horizontal line (very similar RSDs). (By the term reshuffling is meant the possible extraction of a few words from one Set and their addition to another, for instance from Set One to Four or vice-versa.)

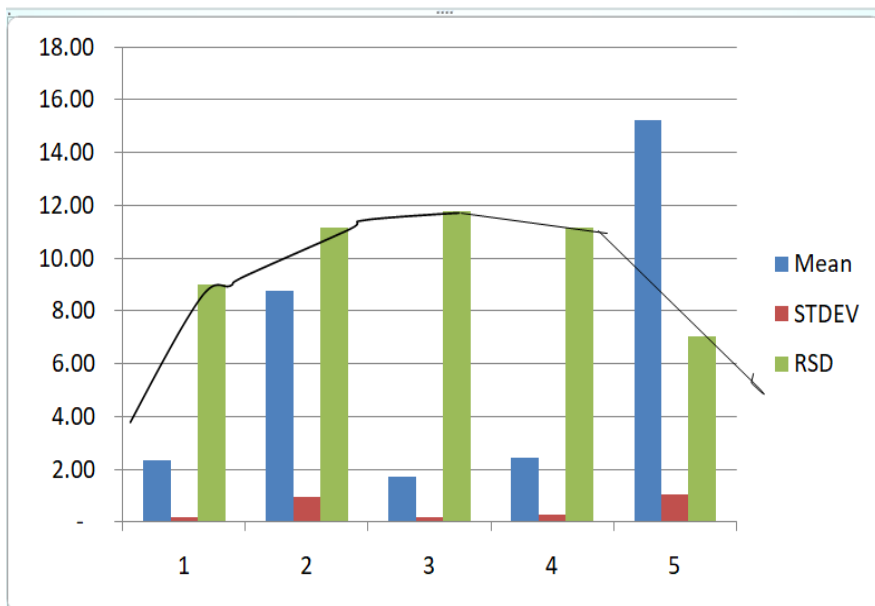


Figure 13: Y-axis expresses Mean and StDev in numbers. RSDs are measured in percentages. The numbers on X-axis correspond to the five sets.

Let us see now which classes each of the play's data fall within for each of the five sets, expressed as a matrix of colours. (Note: The colours in the matrices of the experimentation throughout the thesis are all assigned randomly).

Comedy title	Set One %	Set Two %	Set Three %	Set Four %	Total percentage of the counts of the four sets %
<i>As You Like it</i>	1b	2c	3	4b	5d
<i>The Comedy of Errors</i>	1b	2d	3	4b	5d
<i>Love's Labour's Lost</i>	1a	2a	3	4b	5a
<i>The Merchant of Venice</i>	1b	2b	3	4b	5b
<i>The Merry Wives of Windsor</i>	1b	2c	3	4a	5b
<i>A Midsummer's Night</i>	1b	2b	3	4b	5c
<i>Much ado about Nothing</i>	1b	2d	3	4a	5c
<i>The Taming of The Shrew</i>	1b	2d	3	4a	5c
<i>The Tempest</i>	1b	2b	3	4b	5d
<i>Twelfth Night</i>	1b	2c	3	4b	5c
<i>The Two Gentlemen of Verona</i>	1b	2c	3	4b	5c
<i>The Winter's Tale</i>	1b	2b	3	4b	5b

Table 4: Data points of 12 well-attributed Shakespearian comedies and their contribution to actual classes. Similarly coloured cells represent the distinct classes.

Let us recall the process of generating membership functions from sets' data points. Looking at the table above (Table 4), for Set One there are two shapes and hence in the above Table 4 the column for Set One has two background colours: a green background showing that all plays apart from *Love's Labour's Lost* Love contributed the data that formed the

second triangle ('1b') and a background in olive showing that the data point of play's *Love's Labour's Lost* together with the two hypothetical proximal values contributed to the formation of the first triangle ('1a'). I have discussed this constraint before: an isolated, x -value forms a separate triangular function together and its corners are a hypothetical value 0.1 before and 0.1 after it. Set Two has four background colours: a background in olive showing that data point of *Love's Labour's Lost* contributed (with the two hypothetical values) to a triangle ('2a') that is represented with the background in olive, a light blue background showing that *The Merchant of Venice*, *A Midsummer Night's Dream*, *The Tempest* and *The Winter's Tale* contributed to the formation of second triangle ('2b'), a yellow background showing that *As You Like It*, *The Merry Wives of Windsor*, *The Two Gentlemen of Verona* and *Twelfth Night* contributed to a trapezoidal class ('2c'), and a light orange background showing that *The Comedy of Errors*, *Much Ado About Nothing* and *The Taming of the Shrew* contributed to the fourth class, a triangle ('2d'), of Set Two. The same connotation applies with the background colours of the other three sets. In fact, Set Three has one background colour, Set Four two and Set Five four. So, each background colour represents a class, a membership function and the data points of each column of five sets that share the same background colour have contributed to the same class-shape. This association is also explained in the Textual Addendum of the primary experimentation (Section 7.2.2).

In the primary experimentation there are 12 theoretical ways to combine the shapes of the sets but only eight were selected based on the nine actual combinations found in the plays, as two of the plays had the same combination. In the current model, one pair, one triplet and one quadruplet of the 12 plays have a same combination and therefore there are six distinct combinations of 12 plays that generate six SiS-High rules. Looking at the number of actual shapes of each of four sets ($2 \times 4 \times 1 \times 2$)--I revisit the design of the classes of Set Five in Section 4.3-- there are two possibilities for Set One times the four possibilities for Set Two times the one of Set Three times the two possibilities for Set Four times (16 in total). But again, as in primary experimentation, I ignore the theoretically possible combinations because they were not combinations 'selected' by the 12 Shakespearian users here (see the complete explanation of this point in Section 7.2.2). So, six SiS-High rules are derived from the sets-based stylistic patterns of 12 plays. The following version of the Table 4 shows using colours the combinations that are found in more than one play:

<i>As You Like it</i>	1b	2c	3	4b
<i>The Comedy of Errors</i>	1b	2d	3	4b
<i>Love's Labour's Lost</i>	1a	2a	3	4b
<i>The Merchant of Venice</i>	1b	2b	3	4b
<i>The Merry Wives of Windsor</i>	1b	2c	3	4a
<i>A Midsummer Night's Dream</i>	1b	2b	3	4b
<i>Much Ado about Nothing</i>	1b	2d	3	4a
<i>The Taming of the Shrew</i>	1b	2d	3	4a
<i>The Tempest</i>	1b	2b	3	4b
<i>Twelfth Night</i>	1b	2c	3	4b
<i>The Two Gentlemen of Verona</i>	1b	2c	3	4b
<i>The Winter's Tale</i>	1b	2b	3	4b

Table 5: Combinations that are found in more than one Shakespearean comedy.

From the above list of ways of combining the shapes it is deduced that *Much Ado About Nothing* has the same combination with *The Taming of the Shrew*, *As You Like It* the same combination with *Twelfth Night* and *The Two Gentlemen of Verona*, and *The Merchant of Venice* the same combination with *A Midsummer Night's Dream*, *The Tempest* and *The Winter's Tale*. Therefore, there are six unique combinations and so the following six SiS-High rules (R1-R6) are formed with R1, R4 and R6 expressing the combination of more than a single play:

R1 IF 1b AND 2c AND 3 AND 4b THEN SiS1-High

R2 IF 1b AND 2d AND 3 AND 4b THEN SiS1-High

R3 IF 1a AND 2a AND 3 AND 4b THEN SiS1-High

R4 IF 1b AND 2b AND 3 AND 4b THEN SiS1-High

R5 IF 1b AND 2c AND 3 AND 4a THEN SiS1-High

R6 IF 1b AND 2d AND 3 AND 4a THEN SiS1-High

4.2 Rules and Inference Mechanism of the Comedies-Based Fuzzy Classifier of Shakespearianness.

Apart from the SiS-High rules, there is the necessity, as I highlighted in Section 3.3 and describe in primary experimentation Section 7.2.2, to define an algorithm adding rules that invoke the ‘not/none’ classes. Such a case is if data from a new play fall within some but not all the shapes specified in the above rules. As before, outside of the shapes that represent the data from the 12 Shakespearian plays are areas defined by ‘not/none’ functions. (The ‘not/none’ areas are, in certain cases, overlapping with an actual class adding further methodological fuzziness and natural mathematics-based accuracy in results. See also Sections of 7.3.7 and 7.3.8 of the Technical Appendix). As already stated in Section 3.3, a new play which is not quite in the class of most Shakespeare-like (SiS-High) falls instead in a lower level. Let us describe the algorithm for the formation of non-SiS-High rules and the generation of ‘not/none’ classes which contain the x -values outside of the range of x -values that fall within each (actual) class. (For the necessary technical adaptation of the corners of the ‘actual’ and ‘none’ subclasses, see Section 7.3.14 of the Technical Appendix.) The design of the ‘not/none’ classes of the data points of the 12 comedies meet the criteria set in Section 3.2 (see constraint 'b) and so there can be an overlapping of the actual with the left and right neighbouring ‘not/none’ subclasses whereas the ‘not/none’ classes contain degrees of membership (and the closest the x -values are to the left/right actual class, if any, the least is their membership in the ‘not/none’ class).

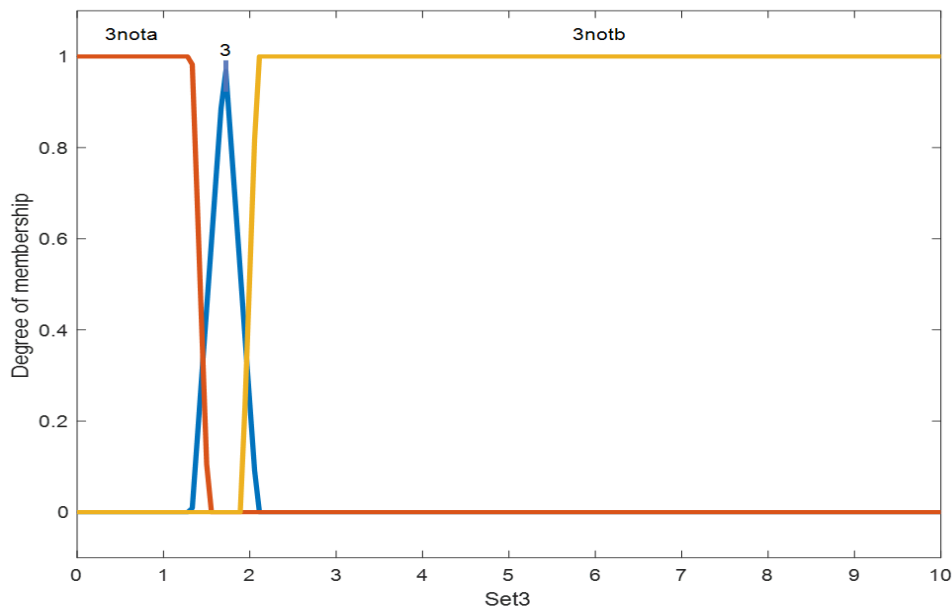


Figure 14: ‘Not/none’ subclasses of Set Three in the comedies-based fuzzy classifier of Shakespearianness.

Our algorithm for the derivation of rules from the scope of constraints follows the same approach that is used in Chapter Seven, and so the algorithm relies heavily on data points of Set Two. Of all the sets, Set Two has for each of the 12 well-attributed Shakespearean comedies the highest percentage of words' frequencies' counts, 8.7%. In fact, Set Two accounts for more than half of all four sets' data points. This set and Set Four contain pronouns and these are rich in stylistic information and have been employed by many other researchers (Mosteller and Wallace 1963, 280–81; Burrows 2002, 272; 2007; Elliot and Greatley-Hirsch 2017, 145). If the tested play is not detected as SiS-High, the system must detect the play's scores within the 'none' or/and 'not' classes and measure how far from the class of SiS-High its data should put it. In this experimentation with the selection of four sets and the unified third set the algorithm is more straightforward than in the primary experimentation and the detection of each 'not/none' class of the four sets in the new plays causes the fall of one level from the SiS-High class. To say that differently, if a new play triggers no 'not/none' classes, the SiS-High is selected. If it triggers one 'not/none' class the result is produced from SiS-Medium-2. If it triggers two 'not/none' classes the output is derived from SiS-Medium-1. And lastly if the new testing play triggers three or four 'not/none' classes the result is produced from SiS-Low. I have followed this ranking order at the medium level-classes since the classes of output variable start from a left SiS-Low class and then continue with SiS-Medium1 and SiS-Medium2 classes. The output classes could be viewed in a clockwise mode, thus 0 = SiS1-Low, 1 = SiS-Medium1, 2 = SiS-Medium2, 3 = SiS-High or conversely-anticlockwise as class 3, class 2, class 1, class 0. For the first output variable SiS1 the classes are SiS1-Low, SiS1-Medium1, SiS1-Medium2, SiS1-High and when there is a second output index, the classes of the second output variable are named SiS2-Low, SiS2-Medium1, SiS2-Medium2, SiS2-High. With the use of overlapping areas between 'actual' and 'none' classes, the result in some cases, depending on the interrelation of rules, can be produced by more than a single class of the output variable, the so-called SiS. (See Section 7.3.13 of the Technical Appendix, the description of the basic algorithm with Set Two treated as a special case.)

In general, the four sets of this experimentation are categorised, slightly differently than in the primary experimentation, into sets of words with very high (average 8.7% for Set Two), high (average 2.4 & 2.3% respectively for Set One and Set Four), and medium to low counts of frequencies (1.7% for Set Three), avoiding sets (and now subsets of Set Three) of very low counts of frequencies (less than 0.5%). This adaptation was made in order to avoid

any critique that the counts are close to zero and there is no much space for differentiation. There are 16 ways ($= 2 \times 4 \times 1 \times 2$) to match the actual shapes but if is added one negative to the existing actual shapes for each of the four sets there are 90 ways ($= 3 \times 5 \times 2 \times 3$) to combine the negated and actual shapes. But only six of the total 16 ways of combining actual shapes (not including negated sets) have been until now seen in the Shakespearian real-world of comedies (SiS1-High). Therefore, similarly to the process adopted in primary experimentation (see paragraph just before the subsection of 7.2.2) I discount the block of 16 SiS1-High rules and say that of the possible 90 combinations, 6 correspond to SiS1-High classification of real-world plays and 10 correspond to hypothetical SiS1-High matches that I have not (yet) seen in real-world plays. If data points of any new play in the validation or testing stage matches with possible actual combinations of the Shakespearian four sets' classes, then the real world of Shakespearian plays becomes larger and a new SiS1-High rule, apart from the current six SiS1-High rules, will be added.

For the time being, that leaves 80 ($90 - (16 - 6)$) possible ordered combinations, since I subtract from the total of 90 initial ordered combinations the ten hypothetical SiS1-High matches that have not yet been met and that it may not be necessary to form. In other words, based on the current evidence of the real world of the Shakespearian plays (as defined by the six actual SiS1-High matches) and the possible combinations of 'not/none' with actual classes, 80 (and this can be extended to 90) different stylistic patterns of comedies can be tested and assessed by the comedies-based Fuzzy Stylistic Classifier of Shakespearianness.

I will now proceed to the addition of 'not/none' class-based rules and leave for later the generation of new rules based on the new plays used for the validation of the model. The validation stage includes the assessment of the SiS of the well-attributed Shakespearian comedies, the experimentation with well-attributed comedies of other authors and the use of general indices of stylistic discrimination, such as the comparison of the counts of words' frequencies in the 12 Shakespearian comedies. Dealing further with the process of enlarging our database of rules, I can follow the methodology also adopted in the primary experimentation (Section 7.2.2) and form the possible combinations of data for the total of four sets as a tree structure in which the root node has four branches (technically known as 'edges') emerging from it, each one of which represents one of the following four possibilities:

- 1) That for one set the matched shape is 'not/none'
- 2) That for two sets the matched shapes are 'not/none'

- 3) That for three sets the matched shapes are 'not/none'
- 4) That for four sets the matched shapes are 'not/none'

As can be viewed also more analytically in Section 7.2.2, each of these selections represents a collection of ways in which the criterion it embodies can be met. The selection is made here taking into account that Set One has two actual shapes and one 'not/none' shape, Set Two has four actual shapes and one 'not/none', Set Three one actual and one 'not/none' shape and Set Four two actual and one 'not/none' (though the aggregation of 'not/none' class is technically complex and entails the notion of subclasses, see Section 7.3.14 of Technical Appendix). One of the variables in this process is the depth to which I descend before I start harvesting nodes, and another variable is how many antecedents of rules are generated before the process stops. On this occasion, the setting for the first variable is two and for the second variable it is six, as I intend to add six more rules. Let us enumerate the first two selections-children (this is the 'depth' variable) extracted in sequence from the existing pools-branches until the total of new rules reaches the number six (the 'how-many-rules' variable):

```

S1 matched as 'not/none'
S2 matched as 'not/none'
-----
S1+S2 matched as 'not/none'
S1+S3 matched as 'not/none'
-----
S1+S2+S3 matched as 'not/none'
S1+S2+S4 matched as 'not/none'
-----

```

The search and harvesting process stops here before the fourth pool whose only existing combination is

S1+S2+S3+S4 matched as 'not/none'.

Accounting for the first actual classes of the sets-inputs, the produced combinations of the rule's antecedents with 'not/none' classes are:

```

1n 2a 3 4a
1a 2n 3 4a
1n 2n 3 4a
1n 2a 3n 4a

```

1n 2n 3n 4a

1n 2n 3 4n

After turning these antecedents into rules by adding the appropriate consequents (SiS1-Low, SiS1-Medium1, SiS1-Medium2, SiS1-High) and by applying the algorithm of the special case of Set Two and one level fall from the SiS-High for the negated sets One, Three and Four, the following new rules are formed:

IF 1n AND 2a AND 3 AND 4a THEN SiS1-Medium2

IF 1a AND 2n AND 3 AND 4a THEN SiS1-Low

IF 1n AND 2n AND 3 AND 4a THEN SiS1-Low

IF 1n AND 2a AND 3n AND 4a THEN SiS1-Medium1

IF 1n AND 2n AND 3n AND 4a THEN SiS1-Low

IF 1n AND 2n AND 3 AND 4n THEN SiS1-Low

Therefore, for the time being, the full set of rules is:

R1 IF 1b AND 2c AND 3 AND 4b THEN SiS1-High

R2 IF 1b AND 2d AND 3 AND 4b THEN SiS1-High

R3 IF 1a AND 2a AND 3 AND 4b THEN SiS1-High

R4 IF 1b AND 2b AND 3 AND 4b THEN SiS1-High

R5 IF 1b AND 2c AND 3 AND 4a THEN SiS1-High

R6 IF 1b AND 2d AND 3 AND 4a THEN SiS1-High

R7 IF 1n AND 2a AND 3 AND 4a THEN SiS1-Medium2

R8 IF 1a AND 2n AND 3 AND 4a THEN SiS1-Low

R9 IF 1n AND 2n AND 3 AND 4a THEN SiS1-Low

R10 IF 1n AND 2a AND 3n AND 4a THEN SiS1-Medium1

R11 IF 1n AND 2n AND 3n AND 4a THEN SiS1-Low

R12 IF 1n AND 2n AND 3 AND 4n THEN SiS1-Low

As far as it concerns the output variable of the fuzzy model, the design is exactly the same with the primary experimentation (see Sections 3.3 and 7.2.3), and it is represented by the four trapezoidal membership functions of the classes of the Low, Medium1, Medium2

and High index of Shakespearian Similarity (SiS). The design of the output variable, as can be viewed also in the extensive analysis of Section 7.2.3, is based on the concept of the mean of maxima (MOM) and four overlapping classes-intervals ([0-0.25], [0.1875-0.50], [0.4375-0.75], [0.6875-1]).

4.3 Validation Stage 1 or Validation of Inference Mechanism.

Let us now see how the comedies-based model with the current 12 rules classifies the 12 well-attributed comedies, whose four sets' data points contributed to the modelling parameters of the first layer of the fuzzy classifier and the production of the unique SiS or else SiS1 (I will keep calling it SiS1 even if it is the unique output index in order to avoid confusion later when I will add the second output index of SiS2).

Comedy title	Which Rule fires	SiS1
<i>As You Like it</i>	1	0.866
<i>The Comedy of Errors</i>	2	0.87
<i>Love's Labour's Lost</i>	3	0.866
<i>The Merchant of Venice</i>	4	0.867
<i>The Merry Wives of Windsor</i>	5 and 6	0.866
<i>A Midsummer Night's Dream</i>	It loses SiS1-High for just 0.01 of the count of Set Three in Rule 4—New rule with one 'not/none' set/class is needed in order to have a precise result. (This will be later R17.)	SiS1-Medium2
<i>Much Ado about Nothing</i>	6	0.856
<i>The Taming of The Shrew</i>	6	0.868
<i>The Tempest</i>	4	0.861
<i>Twelfth Night</i>	1	0.848
<i>The Two Gentlemen of Verona</i>	1	0.866
<i>The Winter's Tale</i>	4	0.862

Table 6: Classification of the 12 well-attributed Shakespearian comedies.

These results are based on the SiS-High rules and the six added ‘not/none’ rules. As I explained above and in the design of constraints, when the classes are formed by the two SD below and above the mean, the ‘not/none’ class overlaps with the actual class from the end of the first to the end of the second SD below and above the mean. Consequently, if the data point of any of the 12 well-attributed plays has an x -value after the first SD, and the actual class is extended to the second SD, then this data point should be assessed also as a ‘not/none’ class. In other words, if any data point of the 12 well-attributed Shakespearian comedies is not before the first SD below and above the mean of an actual class, but falls on or after the second SD, it is at once somewhat actual and somewhat ‘not/none’. This point, therefore, should be assessed to see how much actual and how much ‘not/none’ it is. The rational expectation is that the majority, if not all, of the 12 well-attributed plays after the new mapping of their data points still have a value that falls into range of x -values of the class of SiS-High which, according to the design of the output classes, is [0.6875-1]. In order to continue with the validation, let us consider the plays one by one and see if any of their data points are also a bit ‘none’. If they are, then I need to proceed to the mapping of these points and update our rules mechanism--currently with 12 rules--and add new rules for each play. This would mean that the final result takes into account the centroid of more than one output variable’s class apart that of SiS-High.

1. *As You Like It*

Based on the ‘notness /noneness’ of the plays’ data points, two new combinations need to be added in the set of our rules:

1nc 2c 3 4b then SiS1-Medium2

1nc 2b 3 4b then SiS1-Medium2

(The reason is that the data point 9% of Set Two falls into two actual classes, thus that of ‘2c’ and ‘2b’.)

2. *The Comedy of Errors*—No change is needed; The play's data points do not fall into any ‘not/none’ actual class.

3. *Love’s Labour’s Lost*

One new combination needs to be added in the set of our rules:

1a 2a 3na 4b then SiS1-Medium2

4. *The Merchant of Venice*

One new combination needs to be added in the set of our rules:

1b 2nb 3 4b then SiS1-Low

5. *The Merry Wives of Windsor*- No change, the plays ' data points do not fall into any 'not/none' actual class.

6. *A Midsummer Night's Dream*

One new combination needs to be added in the set of our rules:

1b 2b 3nb 4b then SiS1-Medium2

7. *Much Ado about Nothing*

One new combination needs to be added in the set of our rules:

1nb 2nc 3 4a then SiS1-Low

8. *The Taming of The Shrew*- No change, the plays ' data points do not fall into any 'not/none' actual class.

9. *The Tempest*

One new combination needs to be added in the set of our rules:

1nc 2b 3 4nc then SiS-Medium1

10. *Twelfth Night*

One new combination needs to be added in the set of our rules:

1b 2c 3 4nb then SiS1-Medium2

11. *The Two Gentlemen of Verona*-No change, the plays ' data points do not fall into any 'not/none' actual class.

12. *The Winter's Tale*

One new combination needs to be added in the set of our rules:

1b 2b 3 4nc then SiS1-Medium2

Therefore, the full set of rules is now formed as the previous set of rules:

(Note: 'na' stands for 'not/none' a subclass, 'nb' for 'not/none' b subclass and so forth)

R1 IF 1b AND 2c AND 3 AND 4b THEN SiS1-High

R2 1b AND 2d AND 3 AND 4b THEN SiS1-High

R3 1a AND 2a AND 3 AND 4b THEN SiS1-High

R4 1b AND 2b AND 3 AND 4b THEN SiS1-High

R5 1b AND 2c AND 3 AND 4a THEN SiS1-High

R6 1b AND 2d AND 3 AND 4a THEN SiS1-High

And the new combinations of 'not/none' classes from the tree search:

R7 IF 1na AND 2a AND 3 AND 4a THEN SiS1-Medium2

R8 IF 1a AND 2na AND 3 AND 4a THEN SiS1-Low

R9 IF 1na AND 2na AND 3 AND 4a THEN SiS1-Low

R10 IF 1na AND 2a AND 3na AND 4a THEN SiS1-Medium1

R11 IF 1na AND 2na AND 3na AND 4a THEN SiS1-Low

R12 IF 1na AND 2na AND 3 AND 4na THEN SiS1-Low

And the new combinations of 'not/none' classes of the 12 comedies:

R13 IF 1nc AND 2c AND 3 AND 4b THEN SiS1-Medium2

R14 IF 1nc AND 2b AND 3 AND 4b THEN SiS1-Medium2

R15 IF 1a AND 2a AND 3na AND 4b THEN SiS1-Medium2

R16 IF 1b AND 2nb AND 3 AND 4b THEN SiS1-Low

R17 IF 1b AND 2b AND 3nb AND 4b THEN SiS1-Medium2

R18 IF 1nb AND 2nc AND 3 AND 4a THEN SiS1-Low

R19 IF 1nc AND 2b AND 3 AND 4nc THEN SiS-Medium1

R20 IF 1b AND 2c AND 3 AND 4nb THEN SiS1-Medium2

R21 IF 1b AND 2b AND 3 AND 4nc THEN SiS1-Medium2

Let us input again the data points of each comedy and see if the initial SiS scores of each of the 12 comedies remain still SiS-High. The results can be viewed by the table below:

Comedy title	Which Rule fires	SiS1
<i>As You Like it</i>	1, 13 and 14 (R14 with minimal effect)	0.732
<i>The Comedy of Errors</i>	2	0.87
<i>Love's Labour's Lost</i>	3 and 15	0.814
<i>The Merchant of Venice</i>	4 and 16	0.789
<i>The Merry Wives of Windsor</i>	5 and 16	0.866
<i>A Midsummer Night's Dream</i>	17	<u>0.623</u>
<i>Much Ado About Nothing</i>	6 and 18	<u>0.677</u>
<i>The Taming of the Shrew</i>	6	0.868
<i>The Tempest</i>	4, 14, 19 and 21	<u>0.612</u>
<i>Twelfth Night</i>	1 and 20	<u>0.666</u>
<i>The Two Gentlemen of Verona</i>	1	0.866
<i>The Winter's Tale</i>	4 and 21	0.742

Table 7: Validation SiS1 scores of the 12 well-attributed Shakespearian comedies.

In the above table and in this stage of inference validation (Validation Stage 1) I have underlined the scores of four plays that do not fall into the range of x -values of the class of SiS-High. The class of SiS-High is constituted by the interval which is defined by the lower bound of 0.6875 and upper bound the number 1. (See definition of the term interval in Section 7.3.1 of the Technical Appendix). But in fact, two of these four plays are very close to the left margin of the class of SiS-High, as the SiS scores are 0.666 for *Twelfth Night* and 0.677 for *Much Ado about Nothing*. Only the plays *A Midsummer Night's Dream* and the *Tempest* have scores of 0.623 and 0.612 and so they are clearly deviating from the range of x -values of the class of SiS-High but they are above the other three lower classes. As none of the actual 12 Shakespearian comedies have been judged to be of Low or Medium1 Shakespearianness, the verdict is that a robust and coherent model has been built. Ten of the plays can be described stylistically as plays indicative of High Shakespearianness and two of them as being of SiS1-Medium2.

Now let us look at the comparison of the counts of 100 words' frequencies of each of the comedy-vector with the (average) counts of the average-ideal Shakespearian document-vector of the 12 comedies. The angle formed between the two documents-vectors each time can be also viewed. The maximum of the cosine function is one (1), corresponding to an angle of zero (0) degrees in which the two vectors are pointing in the same direction in three-dimensional space (see subsection for the index of cosine similarity in 1.3.1.2.2). Regarding the counts of words' frequencies and the calculation of the indices of cosine similarity, the same methodology with the primary experimentation was followed (see also explanation of the term and technique of Inverse Document Frequency in Section 7.3.9 of the Technical Appendix). The term of cosine similarity was explained above (Section 1.3.1.2.2). This index counts the angle that the vectors of the counts of the tokens' frequencies of the two documents form. If a 90° angle is formed by two vectors-documents, the cosine of the angle ($\cos \theta$) formed by these two vectors-documents is 0 and the documents are completely dissimilar. If an angle of 0° is formed between these two vectors-documents, the cosine of the angle is one (1), and the documents are completely similar.

In addition, through the counts of frequencies of types' tokens documents are represented as vectors, dimensions by the different variables-words and observations refer to the numerical values of these variables-words. For the numerical analysis, vectors can be identified with a single row or column of the counts of the frequencies of the individual 100 words of the four sets.

In terms of the multi-dimensional space (though not possible to represent graphically), we should imagine that a line is drawn from the origin to a point with 100 observations (that is, dimensions) for each document-vector pointing in a particular direction. Here there are 12 vectors (with the counts of the entries of the 100 words' frequencies), 12 lines representing each one of the 12 comedies. In fact there are 12+1 lines-vectors as documents including the vector of the average-ideal document. With a small adaptation I went further than in the primary experimentation and for reasons of standardisation and in order to compensate for not applying the automatic cut of texts. Though the difference of the texts around 10,000 is minor, the counts of words in different plays were standardised as they were counts of equal texts (equal to the average of the 12 plays), thus if 71 tokens of 'if' were found in a comedy play of 10,004 and the average of the 12 comedies is 10,044 words, I proceed to the adaptation $((71/10,004) * 10,044)$ and the count of 'if' should be updated to 71.28. Therefore I multiplied the counts of each single word's frequencies by the number of the average length of texts (10,044 words) divided by the exact number of words in the play

from which this count is taken. For instance, 10,004 for the play *As You Like it*. (The exact size of the documents, which are all around 10,000 words, does not affect cosine similarity as anyway the cosine index is a measurement of orientation, but I wanted to add symmetry to the documents' vector space.) A separate, thirteenth, vector is that of the vector of the average-ideal document that contains the standardised average counts of the frequencies (tokens) of the 100 words-types as detected in the 12 well-attributed Shakespearian plays (comedies). By the term standardised I refer to the process of an adapted Inverse Document Frequency Standardisation, thus the tokens of each of the 100 words-types were summed, divided by 12 and finally were weighted negatively if they were not detected in all 12 comedies. (See the explanation of primary experimentation in Section 7.3.9 of the Technical Appendix.) If the cosine of the angle ($\cos \theta$) formed by the vector of the average-ideal document and the vector of a document under scrutiny is 0 the documents are completely dissimilar. These two vectors have a 90° angle between them. If the cosine of the angle is one (1), and an angle of 0° formed between the two vectors, the documents are completely similar. In the first case (90° angle), when there is orthogonality, the two documents represented as vectors are characterised by perpendicular direction. The indices of cosine similarity of each comedy-vector of the 12 plays in comparison with the index of the average-ideal Shakespearian document-vector are the following:

Comedy title	Cosine index of similarity with the average-ideal comedy
<i>As You Like it</i>	cosine similarity = 0.980 angle = 11.560 angle type = acute
<i>The Comedy of Errors</i>	cosine similarity = 0.981 angle = 11.225 angle type = acute
<i>Love's Labour's Lost</i>	cosine similarity = 0.975 angle = 12.820 angle type = acute

<i>The Merchant of Venice</i>	cosine similarity = 0.976 angle = 12.486 angle type = acute
<i>The Merry Wives of Windsor</i>	cosine similarity = 0.970 angle = 14.003 angle type = acute
<i>A Midsummer Night's Dream</i>	cosine similarity = 0.980 angle = 11.509 angle type = acute
<i>Much ado about Nothing</i>	cosine similarity = 0.95 angle = 17.68 angle type = acute
<i>The Taming of The Shrew</i>	cosine similarity = 0.970 angle = 14.04 angle type = acute
<i>The Tempest</i>	cosine similarity = 0.906 angle = 25.090 angle type = acute
<i>Twelfth Night</i>	cosine similarity = 0.984 angle = 10.258 angle type = acute
<i>The Two Gentlemen of Verona</i>	cosine similarity = 0.975 angle = 12.909 angle type = acute
<i>The Winter's Tale</i>	cosine similarity = 0.979 angle = 11.729 angle type = acute

Table 8: Indices of cosine similarity of the angle formed by each of the 12 Shakespearian comedies-vectors with the vector of the average-ideal document.

As indicated by the results above, all the plays as vectors of the entries of the counts of 100 individual words' frequencies in comparison with the vector of the respective entries (of the counts of words' frequencies) of the average-ideal document (Section 7.3.9) have a cosine similarity which is more than 0.9. In fact, the minimum of the 12 plays is the cosine similarity index of 0.906. This lowest cosine similarity is *The Tempest*; *Twelfth Night* has the maximum cosine similarity, 0.984. The second level of the fuzzy-system producing the SiS2 explores further the initial SiS1--produced from the four sets' counts--and assesses further the sets' data points from the perspective of three additional stylistic markers. The most important part of this further assessment is the comparison of the entries of the counts of words' frequencies of the new play-vector with those of the vector-ideal document, which in this core experimentation is represented by a range of x -values corresponding to the actual class of the input variable of cosine similarity. At the same time, a second general stylistic marker is included so that the total counts of four sets are treated as a fifth set. As stated before, in the primary experimentation was built a multi-input and single output fuzzy system. The output was named SiS standing for Stylistic index of Similarity. In this core experimentation I employed again this sets-based assessor output of SiS but I call it SiS1 as I added a second output in our fuzzy system, SiS2.

With the use of a second inference mechanism, the second score of SiS2 is produced and the first SiS1 output is further evaluated. SiS1 is measured based on the Shakespearianness of the data points of the four sets of the new input play. SiS2 is produced with the assessment of the Shakespearianness of the counts of the fifth set (that is, the total count of four sets), the SiS1 as input and the index of cosine similarity of the play (used as input), whose 100 words' counts of frequencies were compared with those of the average-ideal document (comedy). I call the second output SiS2 in order to differentiate it from the initial SiS, hence called SiS1, output. Essentially, the aim was to build two layers for the fuzzy classifier, one for testing new comedies for their Shakespearianness based on the assessment of the counts of the four sets' data points, and so in the first stage the SiS1 is produced, and one fuzzy layer for validating this initial SiS1 by assessing the total of 100 individual words' counts of frequencies and the data point expressing the total of the four sets. The same fuzzy principles of the primary experimentation apply both for the design of the membership functions of the two additional stylistic markers (total of sets and cosine similarity) and the further evaluation of the SiS1, thus the classification of the new play's Shakespearianness. If a new play's SiS1 is high and the data point of its fifth set is

Shakespearian, and its index of cosine similarity in comparison with the average-ideal document is also Shakespearian, then the SiS2 is also SiS-High.

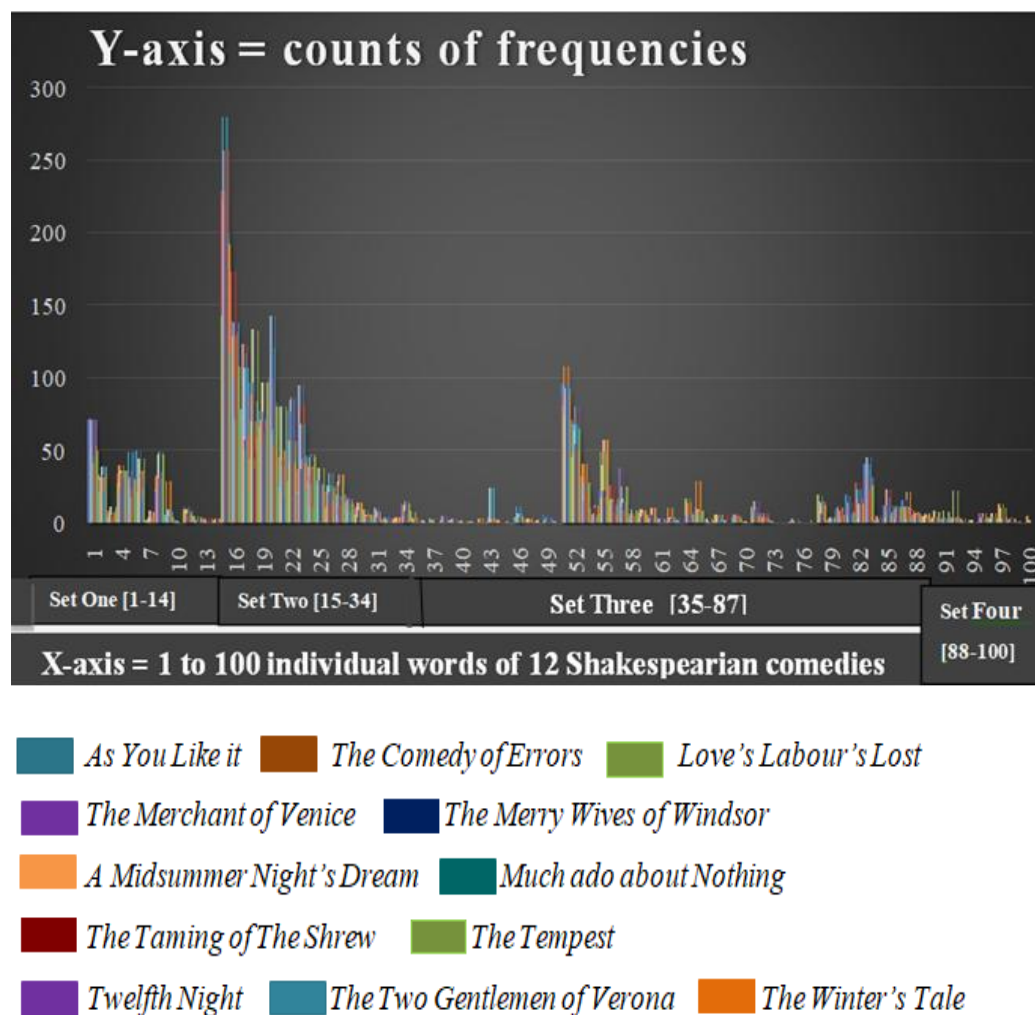


Figure 15: Graph of counts of frequencies of 100 individual words in 12 Shakespearian comedies.

In the graph above (Figure 15) are displayed the counts of 100 individual words' frequencies detected in the 12 well-attributed Shakespearian comedies. The Y-axis expresses the counts of the words' frequencies and the X-axis the four sets' individual words, which are numbered from 1 to 100. (The words in each set are arranged in a random order.) For each word in each of the 12 plays, there can be a maximum of 12 differently coloured bars with each colour representing a specific play. So, for 100 words in 12 comedies there are 1,200 counts, which could produce 1,200 vertical bars. A number of these counts are close or equal

to zero and so many fewer than 1,200 bars are visible. The important element of this graph is the representation of variety of the counts of words' frequencies. Set One is represented by word number 1 to 14 (inclusive), Set Two by word number 15 to 34 (inclusive), Set Three (composed of three subsets) by word number 35 to 87 (inclusive) and Set Four by word number 88 to 100. As can be seen, Set Two includes the words with the highest counts and Set Three has mainly words with low counts (fewer than 10). Sets One and Four contain words of various counts with the maximum exceeding slightly the 70 counts in Set One and the 100 counts in Set Four.

By zooming in to the first of the four sets becomes clearer the way the bars are formed representing the counts of every word's frequencies:

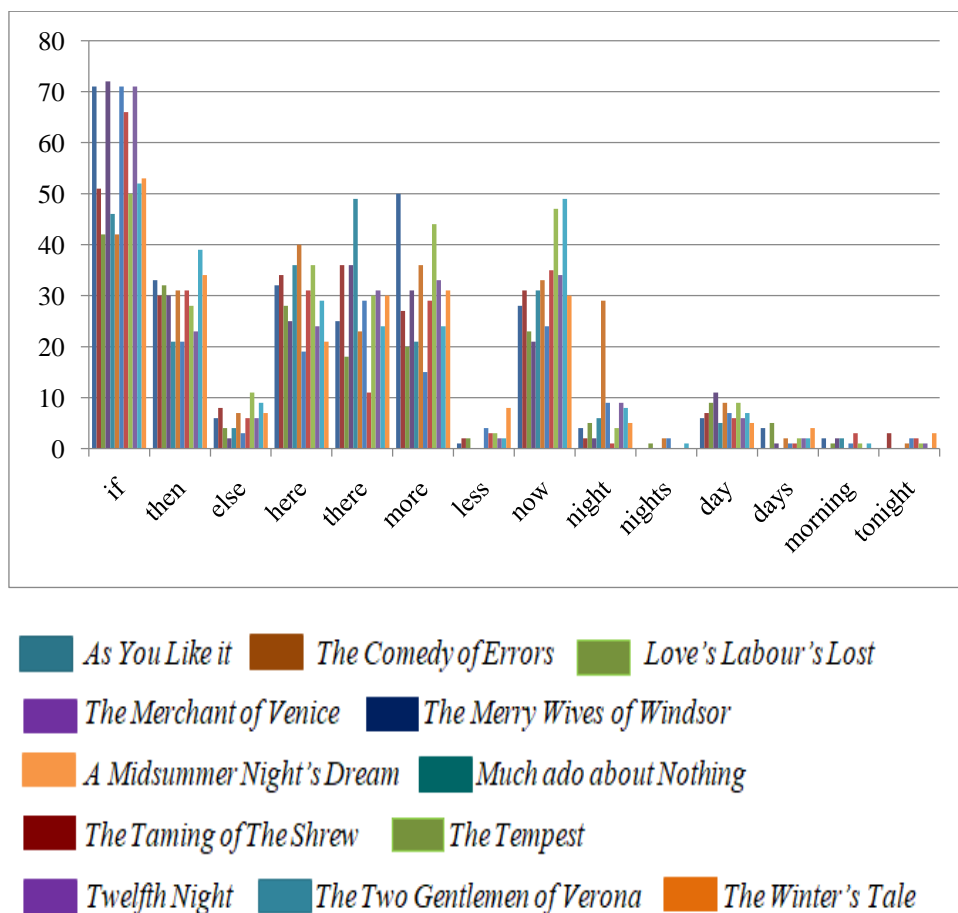


Figure 16: Counts of words' frequencies of Set One.

The two figures above (Figure 15, Figure 16) show the combination of, in general, very frequently employed words of Set Two with the other three Sets that contain words of medium and low counts. This diversity can assist the evaluation of the existence, rarity or non-existence of individual words in the Shakespearian style of the 12 well-attributed

comedies. In this subsection, after experimenting with each of the 12 well-attributed, sole-authored Shakespearian comedies it is possible to validate the inference mechanism of the first layer of the Fuzzy Stylistic Classifier based on the four sets' data points. In addition, all other necessary data were gathered, such as each play's cosine similarity index in comparison with the vector of the average-ideal document representing the average style of the 12 well-attributed plays. The next necessary step is to proceed to the description of the second layer of the fuzzy classifier (which I call SiS2) which uses three additional stylistic markers-input variables and one additional output-second score.

4.4 The Three Additional Input-Variables of the Two-Layered Complete Fuzzy-Logic Stylistic Classifier of Comedies.

The second layer of the fuzzy system contains as inputs the frequencies of occurrence of a fifth set of words (being the fifth variable in the whole fuzzy system), the SiS1 output from the SiS1 layer (as the sixth variable in the whole fuzzy system), and the index of cosine similarity to an average-ideal document (as the seventh variable in the whole fuzzy system). Let us recall graphically the full structure of the fuzzy simulator:

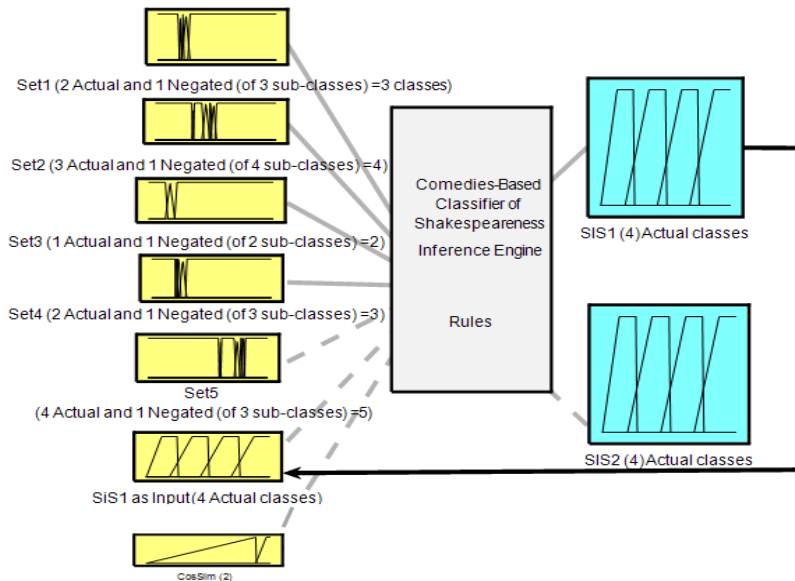


Figure 17: Structure of the complete comedies-based fuzzy classifier.

The fifth set of words forming the fifth input is simply the combined set of all the words in Set One to Set Four. A series of membership classes is formed for this input in the usual way using the actual data: the frequencies of occurrences of these words in the 12 Shakespearian comedies. Below in Figure 18 are the data points of the fifth set which is composed of two

trapezoidal and two triangular actual classes (Notice that there is a correspondence of the colours of the cells of the right column with the respective coloured membership class.)

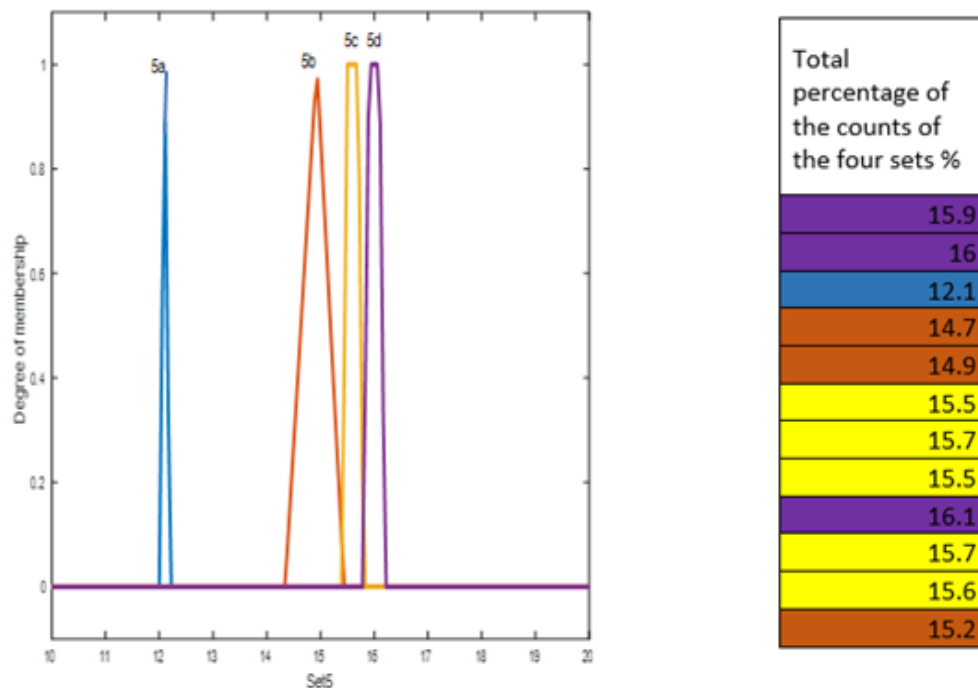


Figure 18: Memberships of Set Five.

The words in this fifth set do not form a semantically related group, but they constitute a general stylistic marker that aggregates the counts for Sets 1 to 4. By adding this marker, the criteria for SiS-High scoring become even more difficult than it was in the primary experimentation. There are various reasons that make it difficult to reach SiS-High with the addition of the fifth set. First, the evaluation of a new input adds a new constraint to the inference mechanism and the attribution of SiS-High for the plays under scrutiny. Secondly, Set Five does not constitute a semantic set, so it is a more general marker. Thirdly, and most importantly, the range of the decimal patterns of the 12 Shakespearian plays forms a much larger non-continuous interval than each of the first four sets. As you can see above in Figure 18, based on the counts of the data points of the 12 well-attributed plays, the non-continuous interval of Set Five ranges from 12.1 to 16.1, whereas that of Set One ranges from 1.9 to 2.6. Therefore, the first interval of Set Five covers a much larger area of ‘not/none’ areas (‘not/none’ Shakespearianness) than any of the four semantic Sets. Based on that correlational evidence, it seems rational to claim for a play under scrutiny that it is more

possible to obtain for its data point of Set One any of the decimal patterns that falls in the actual/Shakespearian plays than those in Set Five.

Regarding the output classes of the SiS1 which plays the role of an input (the sixth input) in the second layer of the fuzzy program and the production of SiS2, the classes are exactly the same with the classes of the output variable employed in the primary (Section 7.2.3) and current experimentation (Section 4.2). Let us remind ourselves of the four membership functions of the output variable that now plays the role of the sixth input of the fuzzy system, the input of SiS1, which is produced each time in the first layer of the fuzzy program after the evaluation of the data points of the four semantic sets of the new input play:

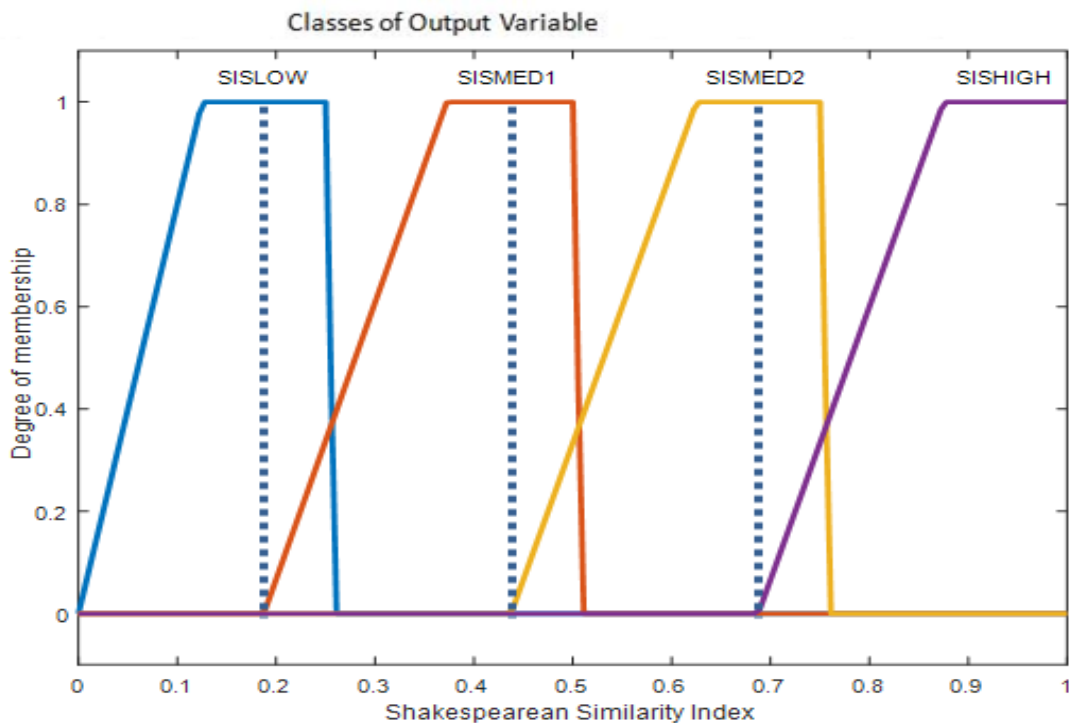


Figure 19: Output SiS as input: SiS1.

The only input variable that has not yet been designed is that of the cosine similarity. In the previous subsection I presented the results of the indices of cosine similarity. This similarity is the angle formed between two vectors in 100-dimensional space. The first vector is formed by the frequencies of occurrence of 100 individual words' frequencies in each of the 12 well-attributed Shakespearian plays. The second vector is formed by the frequencies of occurrence of those same 100 words in what I call the average-ideal document (Section 7.3.9). As postulated at the beginning of this thesis, the goal is to assess the discriminating

power of the counts of words' sets. At this point, in order to avoid paradoxical SiS-results, it is necessary to delimit the actual classes of Shakespearianness for the cosine similarity index. When I mention the term paradoxical results, I refer to the eventuality of almost 'completely' similar sets' counts (SiS-High) but 'excessively' dissimilar counts of words' frequencies (low cosine similarity) of the data points of a play under scrutiny with those data points of the 12 Shakespearian comedies.

In other words, one of the primary objectives of the second layer of the fuzzy system is the addition of the seventh variable (cosine similarity, henceforth called CosSim) in order to explore from a general perspective the counts of the frequencies of these individual 100 words grouped into the four semantic sets. As can be deduced in the table related to the comparison of the index of cosine similarity of each of the 12 comedies with that of the average-ideal document (Section 7.3.9), all 12 comedies have a cosine similarity roughly between 0.9 and 1, corresponding to an angle between 26° and 0° so that the two vectors point in roughly the same direction and one lies almost on top of the other. Thus, a class of actual Shakespearianness and at the same time class of high cosine similarity might have a bottom-left point at 0.9 and a bottom-right point at 1. Of course, it is necessary to follow a bit more precise design by evaluating the following descriptive statistics of the indices of the cosine similarity of each of the 12 comedies-vectors with the average-ideal document-vector:

Average index (mean) of the indices of Cosine Similarity of 12 comedies	0.968
Standard Deviation	0.021
Minimum	0.906
Left Corner (= Minimum – (minus) one Standard Deviation before the mean) of the actual triangular class of the Cosine Similarity	0.884

Table 9: Descriptive statistics of the indices of Cosine Similarity of the 12 comedies.

Based on the above evidence, I will design two classes for the input variable of the cosine similarity. One will be the class of high cosine similarity and a second class will model the range of x -values that have a membership in the class of low cosine similarity, which could be named also as 'not' the Shakespearian class. The class of high cosine similarity will be a right trapezoidal class with bottom-left corner the x -value of 0.884 (which is one SD below the mean) and has $y=0$ and bottom-right corner at $x=1$, $y=0$. The left shoulder point (with $y=1$) of the trapezoidal class on the X -axis equals the average of all

values of the 12 cosine indices. That is, 0.968. The right shoulder point (again with $y=1$) on the X -axis is the value 1 and coincides with the bottom-right corner of the trapezoid. The resulting High CosS(im) shape is shown on the right in Figure 20 below.

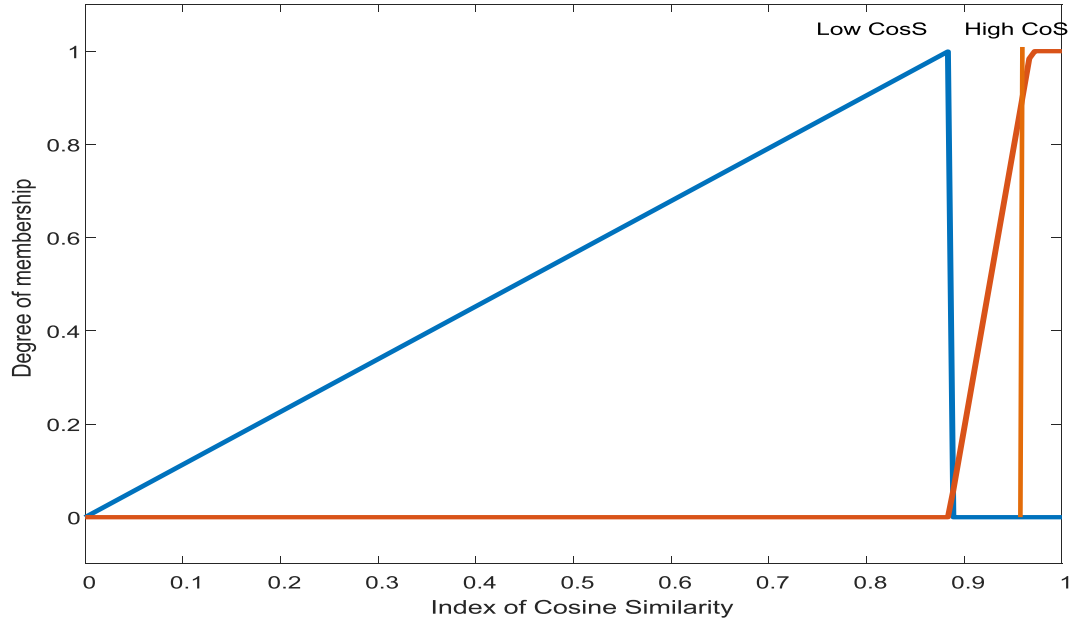


Figure 20: Membership functions of Cosine Similarity.

The second membership function that precedes the trapezoidal class is named class of low cosine similarity and it is represented by an almost orthogonal triangle with its bottom-left corner on the X -axis (and $y=0$). The peak ($y=1$) of this triangle gives on the X -axis the number 0.884. The bottom-right corner of this low-class triangle on the X -axis is 0.885, thus it is located on the X -axis 0.001 after the bottom-left corner of the triangular class of high cosine similarity.

4.5 Algorithm of Producing SiS2 Score.

Essentially, the algorithm in the second layer that completes the fuzzy classification of Shakesperianness of a newly tested play and produces the SiS2 score can be expressed with the following rules:

1. If for a new play the SiS1 is SiS1-High and its index of cosine similarity of the entries of counts of words' frequencies of this new play-vector compared with the entries of the average-ideal document-vector is actual, Shakespearian (that is, High-CosSim), and if the data point of the new play's fifth set-input is actual, Shakespearian (that is, it falls into an

actual class) then the SiS2 is also SiS1-High (though the exact score of SiS2-High might differ from SiS1 as different inputs now influence the truncation of area of the class of SiS2-High).

2. If for a new play the SiS1 is SiS1-High and its index of cosine similarity of the entries of counts of words' frequencies of this new play-vector compared with the entries of the average-ideal document-vector is actual, Shakespearian (that is, High-CosSim) and the data point of the new play's fifth set is not Shakespearian (that is, it falls in a 'not/none' class) then the SiS2 is produced from the centroid of the truncated area of SiS-High (SiS1) plus the class of SiS-Medium2, as with 'not/none' actual the data point of the fifth set-input, the SiS-High of the first layer falls one level.
3. If for a new play the SiS1 is SiS1-Medium2 and its index of cosine similarity of the entries of counts of words' frequencies of this new play-vector compared with the entries of the average-ideal document-vector is actual, Shakespearian (that is, High-CosSim) and the data point of the new play's fifth set is not Shakespearian (it falls in a 'not/none' class) then the SiS2 is produced from the centroid of the truncated area of SiS1-Medium2 (SiS1) plus the class of the SiS1-Medium1, as with 'not/none' the data point of the new plays' fifth set-input the SiS1-Medium2 of first layer falls one level.
4. If for a new play the SiS1 is SiS1-Medium1 and the data point of the new play's fifth set is not Shakespearian (Low-CosSim) then the SiS2 is produced from the centroid of the truncated area of the class of SiS1-Medium1 (SiS1) plus the class of SiS1-Low, as with 'not/none' actual the data point of the new plays' fifth set-input, the SiS1-Medium1 of first layer falls one level.
5. If for a new play the SiS1 is SiS1-Medium2 and its index of cosine similarity of the entries of counts of words' frequencies of this new play-vector compared with the entries of the average-ideal document-vector is actual, Shakespearian (that is, High-CosSim) and the data point of the new play's fifth set is also actual, Shakespearian then the SiS2 is produced from the centroid of the truncated area of SiS1-Medium2 (SiS1) plus the class of SiS1-High, as with actual the data point of the new play's fifth set-input, the SiS1-Medium2 of first layer rises a level.
6. If for a new play the SiS1 is SiS1-Low and its index of cosine similarity of the entries of counts of words' frequencies of this new play-vector compared with the entries of the average-ideal document-vector is actual, Shakespearian (that is, High-CosSim) and the data point of the new play's fifth set is also actual, Shakespearian then the SiS2 is produced from the centroid of the truncated area of SiS1-Low (SiS1) plus the class of

SiS1-Medium1, as with actual, Shakespearian the data point of the new play's fifth set the SiS1-Low of first layer rises a level.

7. If the index of cosine similarity of the entries of counts of the words' frequencies of a new play-vector compared with the entries of the average-ideal document-vector is not Shakespearian (that is, Low-CosSim) then, irrespective of the Shakespearianness of the fifth set, the SiS2 is produced from the centroid of the truncated area of the class of SiS1-Low.

Note: In practice, the truncation in the second layer is applied on the output classes of SiS2-Low, SiS2-Medium1, SiS2-Medium2 and SiS2-High but since the design of the output classes of the two SiS are exactly the same and the classification of SiS1 (thus, the output class in which SiS1 falls) is taken into account for producing SiS2, it was decided--in the description of the algorithm--to refer to the update mechanism ('rises or falls one level') in terms of SiS1 output-terminology.)

Notice that in the algorithm the index of cosine similarity does not influence positively the production of the SiS2 score if this index is Shakespearian. This is because the normal expectation is that it is Shakespearian. When this index is not Shakespearian it has a negative effect on the production of SiS2 score, as the centroid is produced from the lowest class (SiS1-Low) of the output variable. This happens because when the index of cosine similarity is 'not/none' Shakespearian or its data point falls into the class of Low Cosine Similarity this tells us that the counts of the individual words deviate much from the average counts of words' frequencies as detected in the 12 Shakespearian comedies. By the adverb 'much' I indicate the case where the index of cosine similarity of a new play under scrutiny in comparison with the average-ideal document has a zero membership in the class of the high, actual Shakespearian cosine similarity. As has been highlighted throughout, the aim is to investigate and show that sets can be more efficient stylistic markers than counts of words' frequencies. Steps have been taken, however, to avoid contradicting or violating the findings of the researchers about the stylistic information that an individual word may carry and therefore a safety threshold of the use of individual words by the same author, here Shakespeare, has been taken into account as a compensation mechanism so as not to disregard completely the role of words' counts.

After designing the membership functions-classes of the three additional variables, building the complete fuzzy classifier and forming the algorithmic steps of updating the initial SiS (SiS1) in order to produce SiS2 I can continue with the mapping of a

well-attributed, sole-authored non-Shakespearian play and therefore the design of the validation stage of the performance of the comedies-based Fuzzy Stylistic Classifier.

4.6 Design of the Validation Stage 2/Validation of Performance: Mapping A Non-Shakespearian Well-Attributed Play and Further Development of the Inference Mechanism of the Full-Fuzzy Stylistic Classifier.

In this stage I map data points of a well-attributed, sole-authored, non-Shakespearian play in order to validate later the performance of the automated Fuzzy Stylistic Classifier with non-Shakespearian comedies. The aim is to assess with well-attributed non-Shakespearian plays the discriminating power of the fuzzy classifier that has been based on the four sets' data points of the 12 well-attributed Shakespearean comedies.

The experimentation in the core experimentation is split into three stages, the validation stage of inference mechanism, the stage of the validation of the performance which contains the classification of the well-attributed, sole-authored non-Shakespearian plays and the third testing stage, which deals with disputed, possibly co-authored plays. In general, when a classifier is built, the validation options can be various. As a general principle in computational stylometry, viewed also in the *New Oxford Shakespeare*, the researchers usually form, similarly to the division in this thesis, three subcorpora (Shakespearian, non-Shakespearian and disputed plays) but instead of the validation stage of performance that I have employed in this Chapter, they applied a hold-out technique. In our case, this would mean developing the comedies-based model for 11 instead of 12 comedies and then test it by applying it to the twelfth. This hold-out method has not been adopted in this thesis for various reasons.

One basic reason is that this hold-out validation technique would require a complete re-application of the density-based clustering method for each genre-based model and the creation of a separate model with the possible extra-burden of redesigning at least a membership function in each variable. Furthermore, different or a reduced number of patterns (11 instead of 12 plays) can change the world of known plays and this would require to use the new rebuilt models for the three stages of each genre-based model, which would be computationally expensive. In addition, there is strong evidence that the 27 plays that are contained in the known corpus of this thesis are purely Shakespearian, apart from very restricted cases, such as *King John*, as a thorough literature review has been made in order to build a safety-threshold. As noted in Section 1.1, all the 27 plays (see Table 1 on page 74) of

the corpus of this thesis, including *King John*, are listed as such (well-attributed Shakespearian plays) in the Appendix A of the book *Shakespeare, Computers, and the Mystery of Authorship* (Craig and Kinney 2009, 217–18). And the well-attributed non-Shakespearian (Validation Stage 2) or disputed plays (Testing Stage) are also congruent with the categorisation and listing in the Appendix A of Craig and Kinney (Craig and Kinney 2009, 215–18).

Of course, a different version of validation could be initiated. By nature, the definitional context of the well-attributed, sole-authored Shakespearian corpus entails a (rather low) risk, and so the investigator can reasonably accept the risk, which arises also with other computational methods, such as the neural networks where the lack or insertion of specific training patterns can affect the final result-classification. In this context, it can be argued that the misclassification-validation rate of the developed method should be construed as follows. We should ask how many of the total of well-attributed Shakespearian plays are classified by each genre-based Fuzzy-Logic based Stylistic Classifier as SiS1-High (Validation Stage 1) and how many (of the total) well-attributed non-Shakespearian plays are classified as SiS-Low or SiS-Medium1 (Validation Stage 2).

As previously stated, an initial SiS1 will be produced based on the data points of the new plays' four inputs, the semantic sets, and then the new play's data points of three additional inputs (fifth set, SiS1, cosine similarity) will further evaluate SiS1 and derive a second score (SiS2) expressing the degree of stylistic similarity of the new play with the Fuzzy-Logic-based modelled Shakespearian comedy style. These two scores of SiS (the SiS1 produced in the first layer and the SiS2 in the second layer of the fuzzy program) form an evidential interval where one SiS can be the upper and the other SiS the lower bound. This evidential interval is based on the theoretical foundations of Dempster-Shafer Theory (Giarratano and Riley 2005, 284–90) and expresses 'a range of belief' in the evidence. Though Dempster-Shafer theory is related to probabilities and belief networks (Pearl 1988, 12–16, 50–51), its key principles are that it is better to express an attribute and the plausibility of the degree of this attribute in an interval of values rather than a single value. A slight difference with the application of these terms is that the belief is expressed in our experimentation in its general meaning (see Section 7.3.15 of the Technical Appendix) and that the evidential interval refers to the degrees of Shakespearianness. The play I experiment with in the validation of performance stage is the satirical comedy *The Staple of News* by Ben Jonson (Craig and Kinney 2009, 215). It was first published in 1631 and it is close to the period of the publication of the Shakespearian plays of the First Folio collection (1623).

A set of 21 rules has been already mapped. Let us map the four sets' data points of the new play to see if they trigger any of the existing 21 rules or it is necessary to form at least one new rule based on the algorithm's principle 'one-level-fall' (see Section 7.3.13 of the Technical Appendix).

Play for Validation Stage	Set One %	Set Two %	Set Three %	Set Four %	Set Five- Total of Four Sets %
<i>The Staple of News</i>	2.1	7.9	1.1	2	13.1

Table 10: Five Sets' data points of *The Staple of News*.

```
[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a':'trimf',[1.79 1.9 2.01]
MF2='1b':'trimf',[2.04 2.37 2.71]
MF3='1nota':'trapmf',[0 0 1.7 1.81]
MF4='1notb':'trapmf',[1.99 2.1 2.11 2.2]
MF5='1notc':'trapmf',[2.54 2.71 100 100]

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=7
MF1='2nota':'trapmf',[0 0 6 6.1]
MF2='2notb':'trapmf',[6.29 6.4 7.69 8.02]
MF3='2notc':'trapmf',[9.88 10.14 100 100]
MF4='2a':'trimf',[6.09 6.2 6.31]
MF5='2b':'trimf',[7.69 8.35 9.01]
MF6='2c':'trapmf',[8.94 9 9.2 9.26]
MF7='2d':'trimf',[9.13 9.63 10.14]

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=3
MF1='3':'trimf',[1.32 1.71 2.09]
MF2='3nota':'trapmf',[0 0 1.32 1.52]
MF3='3notb':'trapmf',[1.9 2.09 100 100]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=5
MF1='4a':'trapmf',[1.94 2 2.1 2.16]
MF2='4b':'trimf',[2.27 2.59 2.91]
MF3='4nota':'trapmf',[0 0 1.95 2]
MF4='4notb':'trapmf',[2.15 2.2 2.27 2.425]
MF5='4notc':'trapmf',[2.75 2.91 100 100]
```

Figure 21: Script for classes of Four Sets.

As can be deduced by the crosscheck of the preceding table (Table 10) and the above script of Figure 21, the data points of Set One (2.1%) and Two (7.9%) are both actual and ‘not/none’ actual. This can be traced in the shapes-membership functions of Set One ([Input1] in Matlab script) and Two ([Input2] in Matlab script) in Figure 22.

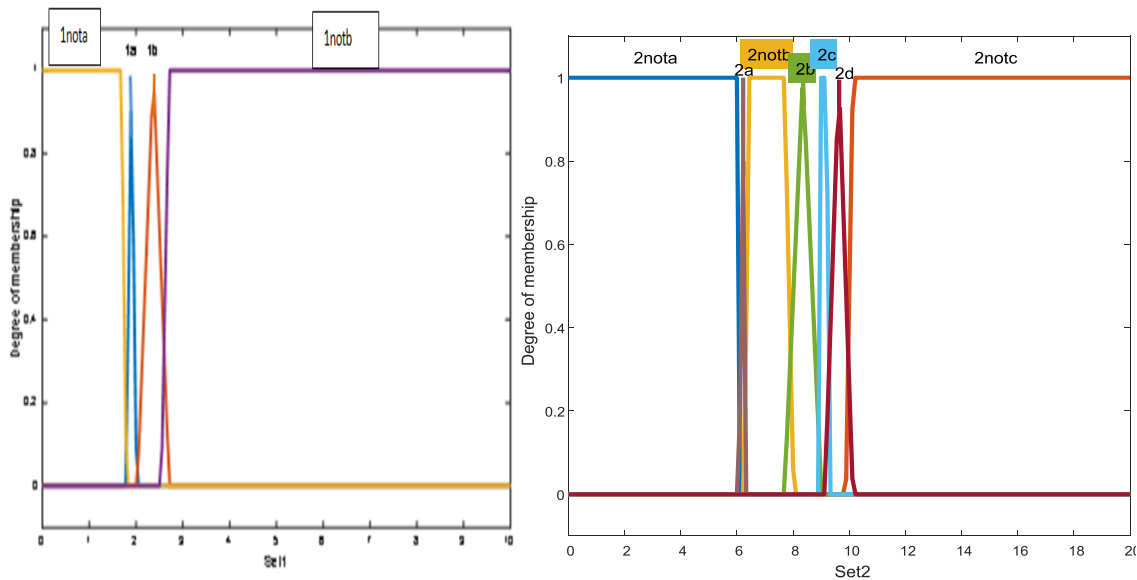


Figure 22: Membership functions of Set One and Two.

Particularly for Set Two [Input2] it is visible how the class of ‘2b’ (a trapezoid in orange-‘trapmf’ in Matlab script) and ‘2notb’ (a triangle in green-‘trimf’ in Matlab script) overlap and include the x -value of 7.9. (The occasional slight curve of straight lines at some points is due to Matlab graphics’ deficiency to represent many membership functions in a short interval of x -values.) In the Technical Appendix (Section 7.3.14) you can view the actual and not actual classes of all five sets. By comparing the numbers in Table 15 with a set of membership functions we can see if these numbers fall inside the shapes of actual or ‘not/none’ actual Shakespearianess.

It has been already stated in the design of membership functions that the actual classes are formed--depending on the principle of full coverage--by one or two SD below and above the mean, and in the second case (of two SD) the ‘not/none’ class overlaps with the actual class from the end of the first to the end of the second SD below and above the mean. Consequently, if any set’s data point of *The Staple of News* has an x -value after the first SD, and the actual class is extended to the second SD, then this data-point is assessed also as a ‘not/none’ class. In practice, the data points are mapped as actual and/or ‘not/none’ actual and

then are formed all the possible combinations between actual or ‘not/none’ actual classes of each of the four semantic sets. The same applies later with Set Five in the second layer of the Fuzzy program. In our case, the occurred combinations of the actual and ‘not/none’ classes of *Staple of News* are:

1b 2b 3na 4a
1nb 2nb 3na 4a
1nb 2b 3na 4a
1b 2nb 3na 4a

As there are no such combinations in the current set of 21 rules, it is necessary to turn these four combinations into antecedents of four rules and to add the proper consequences. Notice that the second combination (R23) contains three ‘not/none’ classes and one of them is Set Two, which, as stated above, is treated in the algorithm of the first layer as a special case (default SiS1-Low). The new rules formed are:

R22 IF 1b AND 2b AND 3na AND 4a THEN SiS1-Medium2
R23 IF 1nb AND 2nb AND 3na AND 4a THEN SiS1-Low
R24 IF 1nb AND 2b AND 3na AND 4a THEN SiS1-Medium1
R25 IF 1b AND 2nb AND 3na AND 4a THEN SiS1-Low

Therefore, the full set of rules is now formed as:

(Note: 'na' stands for ‘not/none’ a subclass, 'nb' for ‘not/none’ b subclass and so forth)

R1 IF 1b AND 2c AND 3 AND 4b THEN SiS1-High
R2 IF 1b AND 2d AND 3 AND 4b THEN SiS1-High
R3 IF 1a AND 2a AND 3 AND 4b THEN SiS1-High
R4 IF 1b AND 2b AND 3 AND 4b THEN SiS1-High
R5 IF 1b AND 2c AND 3 AND 4a THEN SiS1-High
R6 IF 1b AND 2d AND 3 AND 4a THEN SiS1-High

And the new combinations of ‘not/none’ classes from the tree search:

R7 IF 1na AND 2a AND 3 AND 4a THEN SiS1-Medium2

R8 IF 1a AND 2na AND 3 AND 4a THEN SiS1-Low
 R9 IF 1na AND 2na AND 3 AND 4a THEN SiS1-Low
 R10 IF 1na AND 2a AND 3na AND 4a THEN SiS1-Medium1
 R11 IF 1na AND 2na AND 3na AND 4a THEN SiS1-Low
 R12 IF 1na AND 2na AND 3 AND 4na THEN SiS1-Low

And the new combinations of ‘not/none’ classes of the 12 comedies:

R13 IF 1nc AND 2c AND 3 AND 4b THEN SiS1-Medium2
 R14 IF 1nc AND 2b AND 3 AND 4b THEN SiS1-Medium2
 R15 IF 1a AND 2a AND 3na AND 4b THEN SiS1-Medium2
 R16 IF 1b AND 2nb AND 3 AND 4b THEN SiS1-Low
 R17 IF 1b AND 2b AND 3nb AND 4b THEN SiS1-Medium2
 R18 IF 1nb AND 2nc AND 3 AND 4a THEN SiS1-Low
 R19 IF 1nc AND 2b AND 3 AND 4 nc THEN SiS1-Medium1
 R20 IF 1b AND 2c AND 3 AND 4nb THEN SiS1-Medium2
 R21 IF 1b AND 2b AND 3 AND 4nc THEN SiS1-Medium2

And the new combinations-rules produced by the data points of the play titled *The Staple of News*:

R22 IF 1b AND 2b AND 3na AND 4a THEN SiS1-Medium2
 R23 IF 1nb AND 2nb AND 3na AND 4a THEN SiS1-Low
 R24 IF 1nb AND 2b AND 3na AND 4a THEN SiS1-Medium1
 R25 IF 1b AND 2nb AND 3na AND 4a THEN SiS1-Low

Let us have a look at the plays’ index of cosine similarity of the average-ideal document-vector with *The Staple of News* that is included in the stage of validation of performance (Validation stage 2). Concerning the index of cosine similarity of the average-ideal document with the well-attributed to Ben Jonson comedy of *The Staple of News*, the following data are produced:

<i>The Staple of News</i> (9,953 words) Validation Stage	cosine similarity = 0.931 angle = 21.35 angle type = acute
---	--

Extracting all necessary data for the comedy of *The Staple of News* was necessary for the updating of the set of rules. After designing the testing stage in the next subsection, it is possible then proceed to the building of the second layer of the fuzzy classifier with the evaluation of the three additional input variables both for the validation and testing stage.

4.7 Design of the Testing Stage and Further Development of the Inference Mechanism of the Full-Fuzzy Stylistic Classifier.

In this stage I map the data points of the disputed play *All's Well That Ends Well*, for which, as we saw in Section 2.6, the *New Oxford Shakespeare* concludes that probably only a few dozen lines are Thomas Middleton's (Taylor 2017, 322–336; Loughnane 2017, 287-302). Furthermore, relevant findings for this play were also discussed in the primary experimentation and in the validation with the application of the technique of cosine similarity in the Technical Appendix (Section 7.3.9).

With the addition of four rules for *The Staple of News*, there is already a set of 25 rules. Let us map now the four sets' data points of this new play, *All's Well That Ends Well*, in order to assess whether they trigger any of the existing rules or it is necessary to form at least one new rule based on the principle of one-level-fall in the first algorithm (see Section 7.3.13 of Technical Appendix).

Table 11: Five Sets' data points of *All's Well That Ends Well*.

Play for Testing Stage	Set One %	Set Two %	Set Three %	Set Four %	Set Five-Total of Four Sets %
<i>All's Well That Ends Well</i>	1.8	8.3	1.6	2.4	14.1

```

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a':'trimf',[1.79 1.9 2.01]
MF2='1b':'trimf',[2.04 2.37 2.71]
MF3='1nota':'trapmf',[0 0 1.7 1.81]
MF4='1notb':'trapmf',[1.99 2.1 2.11 2.2]
MF5='1notc':'trapmf',[2.54 2.71 100 100]

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=7
MF1='2nota':'trapmf',[0 0 6 6.1]
MF2='2notb':'trapmf',[6.29 6.4 7.69 8.02]
MF3='2notc':'trapmf',[9.88 10.14 100 100]
MF4='2a':'trimf',[6.09 6.2 6.31]
MF5='2b':'trimf',[7.69 8.35 9.01]
MF6='2c':'trapmf',[8.94 9 9.2 9.26]
MF7='2d':'trimf',[9.13 9.63 10.14]

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=3
MF1='3':'trimf',[1.32 1.71 2.09]
MF2='3nota':'trapmf',[0 0 1.32 1.52]
MF3='3notb':'trapmf',[1.9 2.09 100 100]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=5
MF1='4a':'trapmf',[1.94 2 2.1 2.16]
MF2='4b':'trimf',[2.27 2.59 2.91]
MF3='4nota':'trapmf',[0 0 1.95 2]
MF4='4notb':'trapmf',[2.15 2.2 2.27 2.425]
MF5='4notc':'trapmf',[2.75 2.91 100 100]

```

Figure 23: Script for classes of Four Sets.

As can be deduced by the crosscheck of the preceding table (Table 11) and the above script, the data points of Set One (1.8%) and Set Four (2.4%) are both actual and ‘not/none’ actual. Therefore, the occurred combinations of the actual and ‘not/none’ classes for *All’s Well That Ends Well* are:

1a 2b 3 4b
1a 2b 3 4nb
1na 2b 3 4b
1na 2b 3 4nb

In other words, the data points of Sets Two and Three are Shakespearian but the data points of Sets One and Four are somewhat Shakespearian and somewhat non-Shakespearian as they

fall in the range of x -values between the end of a first standard and the end of a second deviation of the actual (triangular) classes respectively of Set One and Four. As there are no corresponding rules in the current set of 25 rules, it is necessary to turn these four combinations into antecedents of four rules and to add the proper consequents.

The new added rules are:

- R26 IF 1a AND 2b AND 3 AND 4b THEN SiS1-High
- R27 IF 1a AND 2b AND 3 AND 4nb THEN SiS1-Medium2
- R28 IF 1na AND 2b AND 3 AND 4b THEN SiS1-Medium2
- R29 IF 1na AND 2b AND 3 AND 4nb then SiS1-Medium1

Regarding R26, an interested finding is that although the added SiS1-High rules are those detected in the actual ordered combinations of the existing corpus of plays-comedies-based, a new SiS1-High rule is added only in case such a combination arises with a new play. This means a new SiS1-High rule should be formed because all the new play's four sets' data points have a non-zero membership in the Shakespearian actual classes of the four inputs-sets. Therefore, from the pool of the ten hypothetical SiS1-High combinations (from a total of theoretical maximum 16 SiS1-High rules) that were set aside before, it is necessary now to extract one, the existing combination-antecedent of 1a AND 2b AND 3 AND 4b for the play *All's Well That Ends Well*. By turning it into a rule (R26) I increase by one the number of the current six SiS1-High rules of the current set of the total of 29 rules.

Current SiS1-High rules are:

- R1 IF 1b AND 2c AND 3 AND 4b THEN SiS1-High
- R2 IF 1b AND 2d AND 3 AND 4b THEN SiS1-High
- R3 IF 1a AND 2a AND 3 AND 4b THEN SiS1-High
- R4 IF 1b AND 2b AND 3 AND 4b THEN SiS1-High
- R5 IF 1b AND 2b AND 3 AND 4a THEN SiS1-High
- R6 IF 1b AND 2d AND 3 AND 4a THEN SiS1-High

Let us now have a look at the index of cosine similarity of the average-ideal document-vector with the play *All's Well That Ends Well*:

<i>All's Well That Ends Well</i> (9,966 words)-Testing Stage	cosine similarity = 0.983 angle = 10.58 angle type = acute
--	--

Similar to the data extraction of the play *The Staple of News*, the process of extracting all necessary data for the comedy of *All's Well That Ends Well* was necessary for the updating of the full set of rules and the building of the second layer of the fuzzy classifier in the next subsection.

4.8 Building of the Second Layer of the Fuzzy Classifier.

Let us now add some new rules (R30-35) regarding the expressed algorithm about the three new inputs that constitute what I call the second layer of the fuzzy system-classifier of Shakespearianness of comedies. These three new inputs were described in Section 4.4 and are the fifth set of words, SiS1 as input and the index of cosine similarity. The data points of the fifth set for the two plays are 13.1 (for *The Staple of News*) and 14.1 (for *All's Well That Ends Well*) and they are both non-Shakespearian, but the indices of cosine similarity of the two plays are Shakespearian. I have not yet calculated the exact SiS1 scores for the two plays, but I have mapped their data points which formed multiple combinations-antecedents of four rules for *The Staple of News* and four rules for *All's Well That Ends Well*. Considering the algorithmic steps of the second layer of the fuzzy program, I can now add some new rules in order to complete the construction of the full Fuzzy Simulator Classifier.

Accounting for two general combinations and the fact that the data point of the fifth set of words is non-Shakespearian ('5notb') both for *The Staple of News* and *All's Well That Ends Well*, the new rules I add to our current set of rules are:

- R30 IF 5a AND SiS1-High AND High-CosS THEN SiS2-High (A general rule. See 4.5)
- R31 IF 5nb AND SiS1-Low AND High-CosS THEN SiS2-Low (A general rule. See 4.5.)
- R32 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Medium1
- R33 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Low

(R32-R33: These two rules form a common output area. Rules 32 to 33 are generated according to the algorithm by the data points of the three additional inputs of *The Staple of News*.)

R34 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium2 (SiS2)

R35 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium1 (SiS2)

Rules 34 and 35 are generated by the data points of the three additional inputs of *All's Well That Ends Well*. These two rules essentially are the components of a single rule described in the algorithm with the three additional input variables. What happens is that the SiS of the first layer, SiS1, from the four sets' data points is updated with the fall of one level from the current, SiS1-Medium2 to SiS1-Medium1 level, and then the evaluation of three new additional inputs at the second layer of the fuzzy classifier is combined with the result of SiS1 in order to produce the SiS2.

Practically, the SiS2 will be now produced from the centroid of the summed truncated areas of SiS2-Medium2 and SiS2-Medium1 classes. The same logic of updating applies also for rules 32 and 33. This kind of update of inference mechanism is not needed for rules 30 and 31 as no change of SiS-level is detected with them in the two distinct stages of evaluation, SiS1 and SiS2. As seen, Rule 30 is a general rule that is triggered when SiS1 (sixth input) is SiS1-High and the data point of inputs of the two additional inputs (the fifth and seventh input) are all Shakespearian (and so SiS2 will be also High). In simple terms, it can be argued that if this rule, R30, fires for a new play, it is 'almost certain' that this play is Shakespearian. Similarly, Rule 31 is also a general rule that is triggered when SiS1 (sixth input) is SiS1-Low and the data point of the fifth set is 'not/none' Shakespearian (and so SiS2 will be also Low). Simplifying again, we can say that if this rule, R31, fires for a new play, it is 'almost certain' that this play is not Shakespearian. For the description of these evidential intervals and their expression in linguistic terms (indicated by the use of quotation marks), see Section 7.3.15 of the Technical Appendix. Rules 32 and 33 are generated by the data points of the three additional inputs (fifth set, SiS1 as input and cosine similarity) of *The Staple of News*. Rules 34 and 35 are generated--according to the algorithm--by the data points of *All's Well That Ends Well*. SiS1 as input is the result of the evaluation of the data points of the four semantic sets and it is included in the antecedent of rules 30 to 35. When SiS1 and SiS2 are produced, the second SiS (SiS2) is produced by the minimum membership of the three new input variables. The first SiS (SiS1) has been derived from the minimum membership of the data points' memberships of the first four input variables, the four sets.

Having completed the building of the second layer of the fuzzy classifier I will now proceed, with the inference mechanism of 35 rules, to the experimentation in the next subsection with a well-attributed non-Shakespearian and an anonymous play in order to explore how the two SiS scores (SiS1, SiS2) are produced for each play.

4.9 Complete Fuzzy-System Output: Validation Stage 2 and Testing Stage.

This subsection includes the experimentation of the complete Fuzzy Stylistic Classifier with a well-attributed non-Shakespearian play in the second validation stage (Validation Stage 2) and with a disputed, anonymous play in the testing stage (Stage 3). For reasons of clarity and impartiality it was judged necessary to first form the new rules as generated by the two new plays' data points, and after forming the set of 35 rules, then to proceed with the same rules to the experimentation with the two new plays in the first and the second layer of the fuzzy system. In the second layer, SiS1 is further evaluated and a second SiS2 is produced. Therefore, the existence of two SiS scores provides us with an interval expressing the belief in a range of values-degrees of Shakespearianness attributed to a new input play. The transition from the first to the second layer of experimentation takes place gradually, as the data points of the four semantic sets provide the score of SiS1 which is used as input in the second layer of fuzzy evaluation.

4.9.1 Experimentation with Well-Attributed, Sole-authored Non-Shakespearian Comedies.

First, it is necessary to have a look at the newly formed set of our current 35 rules:

(Note: 'na' stands for 'not/none' a subclass, 'nb' for 'not/none' b subclass and so forth)

Six initial SiS1-High rules:

R1 IF 1b AND 2c AND 3 AND 4b THEN SiS1-High

R2 IF 1b AND 2d AND 3 AND 4b THEN SiS1-High

R3 IF 1a AND 2a AND 3 AND 4b THEN SiS1-High

R4 IF 1b AND 2b AND 3 AND 4b THEN SiS1-High

R5 IF 1b AND 2c AND 3 AND 4a THEN SiS1-High

R6 IF 1b AND 2d AND 3 AND 4a THEN SiS1-High

To these are added the new combinations of actual classes with at least one ‘not/none’ from the tree search whose formation is explained Sections 4.2 and 7.2.2:

R7 IF 1na AND 2a AND 3 AND 4a THEN SiS1-Medium2
R8 IF 1a AND 2na AND 3 AND 4a THEN SiS1-Low
R9 IF 1na AND 2na AND 3 AND 4a THEN SiS1-Low
R10 IF 1na AND 2a AND 3na AND 4a THEN SiS1-Medium1
R11 IF 1na AND 2na AND 3na AND 4a THEN SiS1-Low
R12 IF 1na AND 2na AND 3 AND 4na THEN SiS1-Low

To these are added the new combinations of actual and classes of the 12 comedies:

R13 IF 1nc AND 2c AND 3 AND 4b then SiS1-Medium2
R14 IF 1nc AND 2b AND 3 AND 4b then SiS1-Medium2
R15 IF 1a AND 2a AND 3na AND 4b then SiS1-Medium2
R16 IF 1b AND 2nb AND 3 AND 4b then SiS1-Low
R17 IF 1b AND 2b AND 3nb AND 4b then SiS1-Medium2
R18 IF 1nb AND 2nc AND 3 AND 4a then SiS1-Low
R19 IF 1nc AND 2b AND 3 AND 4 nc then SiS1-Medium1
R20 IF 1b AND 2c AND 3 AND 4nb then SiS1-Medium2
R21 IF 1b AND 2b AND 3 AND 4nc then SiS1-Medium2

To these are added the rules generated by the *Staple of News*:

R22 IF 1b AND 2b AND 3na AND 4a THEN SiS1-Medium2
R23 IF 1nb AND 2nb AND 3na AND 4a THEN SiS1-Low
R24 IF 1nb AND 2b AND 3na AND 4a THEN SiS1-Medium1
R25 IF 1b AND 2nb AND 3na AND 4a THEN SiS1-Low

To these are added the rules generated by *All's Well That Ends Well*:

R26 IF 1a AND 2b AND 3 AND 4b THEN SiS1-High
R27 IF 1a AND 2b AND 3 AND 4nb THEN SiS1-Medium2
R28 IF 1na AND 2b AND 3 AND 4b THEN SiS1-Medium2

R29 IF 1na AND 2b AND 3 AND 4nb then SiS1-Medium1

To these are added two general SiS2-rules (R30, R31) and four other new rules (R32-R36), all of which are related to the processing of the three additional inputs (input variables 5, 6, 7) of the two new plays:

R30 IF 5a AND SiS1-High AND High-CosS THEN SiS2-High

R31 IF 5nb AND SiS1-Low AND High-CosS THEN SiS2-Low

R32 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Medium1

R33 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Low

R34 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium2

R35 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium1

Let us now start the validation process and experimentation with the first new well-attributed, sole-authored play, *The Staple of News* by Ben Jonson. After the paradigmatic description and detailed graphical display of the experimentation process in deriving the SiS scores of *The Staple of News* for reasons of understanding of the process of fuzzification-defuzzification and truncation, the scores of the other plays will be only tabulated. Below there is the graphical display of the fired rule(s), the output of the truncated areas, the centroid and the SiS scores, which are produced after I input to the fuzzy system-classifier the four sets' data points of *The Staple of News*. There are two graphs that show the interaction with the data points of the four sets, as the increased number of rules did not allow to cover all the interaction in one figure (there are technical restrictions of minimisation of the window frames in Matlab software.) Initially, I start with the data points of the four semantic sets. The first figure below (Figure 24) prints out the interactive output until the 30th rule. The last two columns produce the scores of SiS1 and SiS2. The first of these two last columns produced as SiS1 number of 0.337. The SiS2 in the last column is in default mode (0.5) and there is no truncation yet of any blue shaded area of the output variable.



Figure 24: *The Staple of News*, SiS1 (R22, R23, R24, R25).

The second figure below (Figure 25) prints out the part/interaction that was not visible in the previous graph and at the bottom right corner you can see the centroid of the blue truncated shaded area of the first SiS1. As stated, when I input only the data points of the four sets, the SiS2 in the last column is in default mode (0.5) and there is no truncation yet, no blue shaded area of the second output variable.



Figure 25: *The Staple of News*, SiS1 (R22, R23, R24, R25).

Notice the white field at the bottom-left corner of the figures above where I entered the data points of *The Staple of News*. The first four entries correspond to the data points of the four sets and the remaining to the three additional inputs that have not yet been entered.

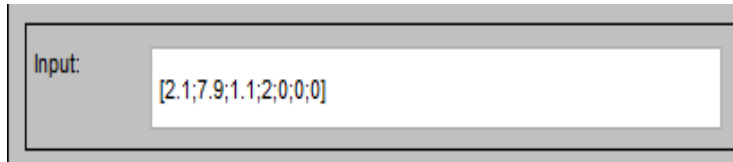


Figure 26: Input variables of *The Staple of News* for the production of SiS1.

As now the first SiS has been produced (SiS1=0.337), I can continue with the experimentation in the second layer and enter as input the data points of the three additional variables of the *Staple of News*. Now is produced the score of SiS2 for *Staple of News* and it is 0.27.

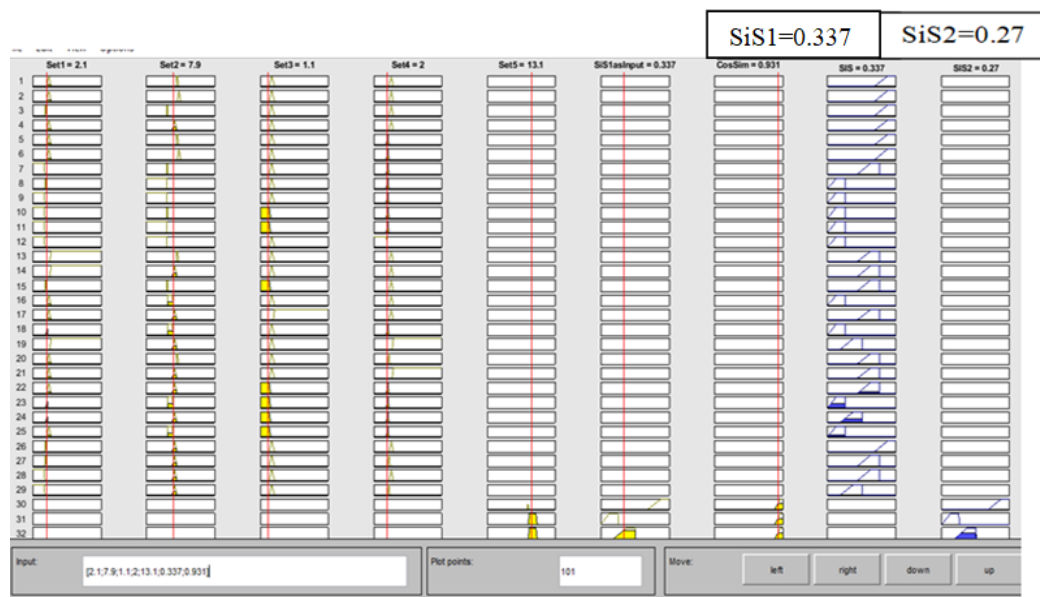


Figure 27: *The Staple of News*, SiS2 (R32, R33).

The Figure 28 below completes the output and prints out the part/interaction that was not visible in the previous graph and at the bottom-right corner you can now see the centroid of the blue truncated shaded areas of the output variable's class/classes that relate(s) to SiS1 but also to that of the centroid of SiS2. Both SiS scores are produced from more than one blue shaded area as more than one rule apply in each case (rules R22,R23, R24 and R25 for SiS1, rules R32 and R33 for SiS2).

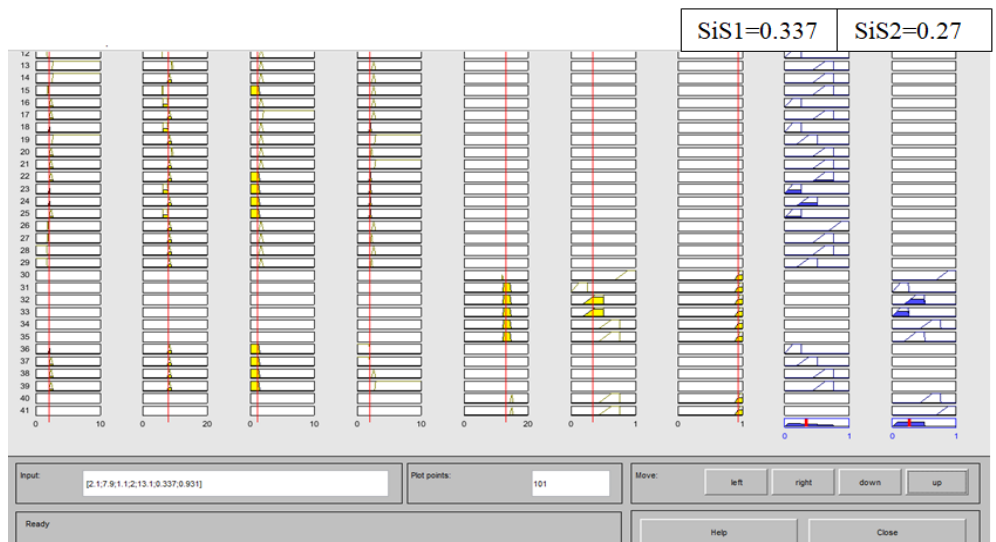


Figure 28: *The Staple of News*, SiS2 (R32, R33).

Notice the white field at the bottom-left corner of the two above figures where I entered as inputs the two data points of the fifth set, the SiS1 and the index of cosine similarity of *The Staple of News*. The first four entries correspond to the data points of the four sets, and the remaining three entries are filled with the data of the three additional inputs. (For instance, see the sixth entry of 0.337 which is the SiS1 and here plays the role of input. Each entry is separated by the next one with the use of a semi-colon).

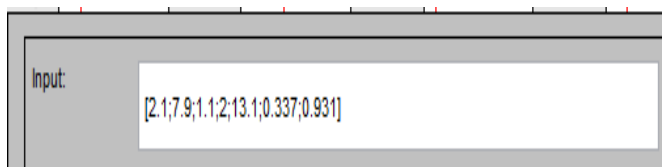


Figure 29: Input data points for the production of SiS2 of *The Staple of News*.

Therefore, the two SiS scores are 0.337 and 0.270 and so the evidential interval of 0.270 to 0.337 is produced and it expresses the limits of the degree of Shakespearianness of *The Staple of News*.

<i>The Staple of News</i>	Which rule fires	Output
0.337	R22, R33, R34, R35	SiS1
0.27	R32, R33	SiS2

Table 12: SiS1 and SiS2 results of *The Staple of News*.

This kind of interval assists us in avoiding the uncertainty of single values-based of SiS scores. The printed scores of the SiS interval, in fact the production of an SiS-Low interval, enhance the validity of our fuzzy classifier of Shakespearianness of comedies and square with the claim that the *Staple of News* is not a Shakespearian play.

4.9.2 Experimentation with Anonymous or Disputed Plays: Testing Stage.

The goal in this (Testing) stage is to follow exactly the same technical methodology with the experimentation with *The Staple of News* in the previous stage (Validation Stage 2, 4.9) and to evaluate if the interval based on the two SiS scores is congruent to the claims about *All's Well That Ends Well*, thus that this comedy is almost a sole-authored Shakespearian play. SiS1 will be produced based on the data points of the new plays' four sets-inputs and then the new plays' data points of three additional inputs will filter SiS1 and derive an SiS2 score. This will give us a second index of stylistic similarity of *All's Well That Ends Well* with the Fuzzy-Logic-based modelled Shakespearian style of the 12 comedies.

I start with entering to the fuzzy system the data points of the four semantic sets. The last two columns will again produce the scores of SiS1 and SiS2. The first of these two last columns produces as SiS1 the number 0.601 with firing rules R26, R27, R28 and R29. The SiS2 for the time being is in default mode (0.5) and there is no truncation yet of any output class. This is because it is necessary to estimate first the result for SiS1 and then to enter it as an input together with the two other inputs (fifth set and CosSim).

<i>All's Well That Ends Well</i>	Which rule fires	Output
0.601	R26, R27, R28, R29	SiS1
0.5	(Default, no firing of rules and no truncation yet)	SiS2

Table 13: SiS1 result of *All's Well That Ends Well*.

The first four entries of *All's Well That Ends Well* correspond to the data points of the four sets and the remaining last three entries (0; 0; 0) will be filled with the data points of the three additional inputs.

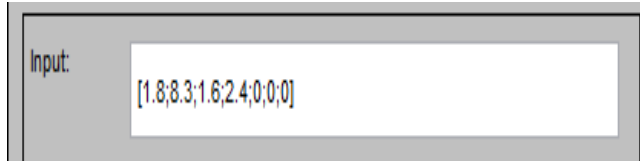


Figure 30: Input variables of *All's Well That Ends Well* for the production of SiS1.

Let us now add the data points for the three inputs of the second layer of the fuzzy program. The first of these produced, SiS1, is of 0.601 and the rules that fired for the production of this score were the rules R26, R27, R28 and R29. By entering the data points in the last three entries of the white box similar to the steps followed in the experimentation with *The Staple of News*, the system produces for SiS2 the score of 0.496 (with Rules 34 and 35 firing). So, the results of the two SiS scores for *All's Well That Ends Well* are:

<i>All's Well That Ends Well</i>	Which rule fires	Output
0.601	R26, R27, R28, R29	SiS1
0.496	R34, R35	SiS2

Table 14: SiS1 and SiS2 of *All's Well That Ends Well*.

The first score (SiS1) was derived from the overlapping area of SiS1-High and the SiS1-Medium2 class and the SiS2 is then produced from the classes of SiS2-Medium2 and SiS2-Medium1 (which have the same design with SiS1-Medium2 and SiS1-Medium1). Therefore, the two SiS scores are 0.601 and 0.496. This gives *All's Well That Ends Well* an evidential interval of 0.496 to 0.601. It contains a range of x -values of the output class of SiS2-Medium2. In comparison with the primary experimentation where the produced single SiS was around 0.85, the SiS scores of this testing stage for *All's Well That Ends Well* are lower. But this difference is rational as in this core experimentation many additional strict constraints were applied, such as the enlargement of 'not/none' areas with the use of the second standard deviation and the use of additional inputs and a second layer of evaluation. The additional input variables and strict constraints together with the existence of overlapping actual areas-classes with the neighbouring 'not/none' areas-classes improve the distinguishing power of the fuzzy classifier as these factors assist in exploiting the properties of values-data points' neighbourhood connectivity. In effect, this kind of connectivity of neighbourhood constraints should be construed in the context of a constraint-based local search technique and combinatorics (Hentenryck and Michel 2005, 3–10). In that sense, combinatorics here helps us to reduce crispness and uncertainty because they assess not only

the membership of the data points in a class (Shakespearian or not) but also the data points' proximity from any, left or right, if any, other neighbouring finite sets-classes.

The claim that *All's Well That Ends Well* is mostly a Shakespearian play with some small part written by another author is still valid as the upper bound of the evidential interval falls into the class of SiS1-Medium2 ([0.4375-0.75]). Its degree of Shakespearianness is close to the plays *A Midsummer Night's Dream* and *The Tempest*, which in the Validation Stage 1 of the first fuzzy layer have respectively SiS1scores of 0.623 and 0.612. Though these plays are clearly deviating from the range of x -values of the class of SiS1-High, they are above the other three lower classes. At the same time, the deviation of the SiS scores might indicate that the participation of a second author might be larger than a few dozen lines as concluded in the *New Oxford Shakespeare: Authorship Companion* (Taylor 2017, 322–336; Loughnane 2017, 287-302). To substantiate this claim, multiple runs of experiments would need to be performed.

4.9.3 Some more plays for Validation Stage 2 and Experimentation with a New Anonymous, Disputed Play.

In this subsection I deal with three more plays, consisting of two well-attributed, sole-authored non-Shakespearian comedies and one disputed comedy that is thought to be mostly Shakespearian (Taylor and Loughnane 2017). I form the new rules based on the data points of the three plays and then with the updated set of rules all three plays are assessed for their Shakespearianness.

These comedies are: *A Mad World, My Masters* which is well-attributed to Thomas Middleton, *The Wild Goose Chase* which is well-attributed to John Fletcher, and *Measure For Measure*, which is held by some to be an adaptation by Thomas Middleton of a previous work by Shakespeare (Taylor and Loughnane 2017, 555–68). The two well-attributed, sole-authored non-Shakespearian plays are employed for the further validation of the performance of the Fuzzy Stylistic Classifier and the third disputed play is used for the testing stage. Let us map the four sets' data points of the three plays in order to proceed to the formation of the new rules that will be added to the current set of 35 rules.

Play	Set One %	Set Two %	Set Three %	Set Four %	Set Five- Total of Four Sets %
Validation Stage:					
<i>A Mad World, My Masters</i>	2.1	8.3	1.2	1.9	13.5
<i>The Wild Goose Chase</i>	2.4	8.7	1.2	1.5	13.8
Testing Stage:					
<i>Measure For Measure</i>	2.2	8.3	1.4	2.9	14.8

Table 15: Sets' data points of three new comedies under scrutiny.

```

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a':'trimf',[1.79 1.9 2.01]
MF2='1b':'trimf',[2.04 2.37 2.71]
MF3='1nota':'trapmf',[0 0 1.7 1.81]
MF4='1notb':'trapmf',[1.99 2.1 2.11 2.2]
MF5='1notc':'trapmf',[2.54 2.71 100 100]

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=7
MF1='2nota':'trapmf',[0 0 6 6.1]
MF2='2notb':'trapmf',[6.29 6.4 7.69 8.02]
MF3='2notc':'trapmf',[9.88 10.14 100 100]
MF4='2a':'trimf',[6.09 6.2 6.31]
MF5='2b':'trimf',[7.69 8.35 9.01]
MF6='2c':'trapmf',[8.94 9 9.2 9.26]
MF7='2d':'trimf',[9.13 9.63 10.14]

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=3
MF1='3':'trimf',[1.32 1.71 2.09]
MF2='3nota':'trapmf',[0 0 1.32 1.52]
MF3='3notb':'trapmf',[1.9 2.09 100 100]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=5
MF1='4a':'trapmf',[1.94 2 2.1 2.16]
MF2='4b':'trimf',[2.27 2.59 2.91]
MF3='4nota':'trapmf',[0 0 1.95 2]
MF4='4notb':'trapmf',[2.15 2.2 2.27 2.425]
MF5='4notc':'trapmf',[2.75 2.91 100 100]

```

Figure 31: Script of four sets' membership functions-classes.

As can be deduced by the crosscheck of the preceding table (Table 15) and the above script (Figure 31), some data points of some of the three plays' four sets are actual and 'not/none' actual. Therefore, it is necessary to form the combinations of the three plays' actual and 'not/none' classes. For *A Mad World, My Masters* there are two combinations:

1nb 2b 3na 4na
1b 2b 3na 4na

For *The Wild Goose Chase* there is only one combination:

1b 2b 3na 4na (This is the same combination as the second combination of *A Mad World, My Masters* so there is no need to duplicate it.)

For *Measure For Measure* there are four combinations:

1b 2b 3na 4b

1b 2b 3 4b (This combination is the antecedent of rule R4-SiS1-High so there is no need to duplicate it.)

1b 2b 3 4nc (This combination is the antecedent of rule R21 so there is no need to duplicate it.)

1b 2b 3na 4nc

It is necessary now to turn these new combinations into antecedents of new rules and to add the proper consequents. Therefore, four new rules must be added.

R36 IF 1nb AND 2b AND 3na AND 4na THEN SiS1-Low

R37 IF 1b AND 2b AND 3na AND 4na THEN SiS1-Medium1

R38 IF 1b AND 2b AND 3na AND 4b THEN SiS1-Medium2

R39 IF 1b AND 2b AND 3na AND 4nc THEN SiS1-Medium1

A total of 39 rules has now been formed (the previous 35 plus 4 new rules generated by the data points of the three new plays). By following the same technical methodology as before with *The Staple of News* and *All's Well That Ends Well* I can proceed to the experimentation with the current set of rules in order to produce SiS1 scores for the three new plays.

Let us now start the validation process and experiment with *A Mad World, My Masters* by Thomas Middleton. As before, the two rightmost columns of the inference engine produce the scores of SiS1 and SiS2. The first of these two columns produces a SiS1 score of 0.199. The SiS2 in the last column is in default mode (0.5) and there is no truncation yet of any area of the output variable. The rules that fire here for SiS1 are R36 and R37.

<i>A Mad World, My Masters</i>	Which rule fires	Output
0.199	R36, R37	SiS1
0.5	(Default, no firing of rules and no truncation yet)	SiS2

Table 16: SiS1 result of *A Mad World, My Masters*.

The next play I experiment with is *The Wild Goose Chase* by John Fletcher. As with the previous play, I experiment only with the data points of the four semantic sets and the focus is on the production of SiS1 for the three plays. The rightmost two columns produce the scores of SiS1 and SiS2. The first of these two columns produces an SiS1 score of 0.367. The SiS2 in the last column is in default mode (0.5) and there is no truncation yet of any area of the output variable. The rule that fires for the production of SiS1 is R37.

<i>The Wild Goose Chase</i>	Which rule fires	Output
0.367	R37	SiS1
0.5	(Default, no firing of rules and no truncation yet)	SiS2

Table 17: SiS1 result of *The Wild Goose Chase*.

The disputed play that I use for the testing stage--I have already experimented with *All's Well That Ends Well*--is the comedy *Measure For Measure*. As usual, the rightmost two columns of the inference engine produce the scores of SiS1 and SiS2. The first of these two columns produces as SiS1 the number 0.459. SiS1 is based on the evaluation only of the four sets' data points. The SiS2 is in default mode (0.5) and there is no truncation yet of any area of the output variable. The rules that fire for the production of SiS1 are the rules R21 and R39.

<i>Measure For Measure</i>	Which rule fires	Output
0.459	R21, R39 (R4 and R38 have minimal effect)	SiS1
0.5	(Default, no firing of rules and no truncation yet)	SiS2

Table 18: SiS1 result for *Measure For Measure*.

Let us now have a look at the index of cosine similarity of the average-ideal document-vector with each of the three comedies under scrutiny.

<i>A Mad World, My Masters Middleton</i> , (9,968 words)-Validation Stage	These sets are not orthogonal cosine similarity = 0.94 angle = 19.17 angle type = acute
<i>The Wild Goose Chase Fletcher</i> , (10,019 words)-Validation Stage	These sets are not orthogonal cosine similarity = 0.91 angle = 24.34 angle type = acute
<i>Measure for Measure</i> (10,099 words)- Testing Stage	These sets are not orthogonal cosine similarity = 0.97 angle = 14.71 angle type = acute

Table 19: Index of cosine similarity of the three new plays with the average-ideal document.

As can be viewed by the above table, *Measure for Measure* is far more Shakespearian than the two other plays even from the scope of the counts of words' frequencies. *The Wild Goose Chase* especially is close to the minimum (0.906) of the indices of cosine similarity of the 12 Shakespearian plays. At this point, it should be emphasised that though the angles of the three plays-vectors formed with the vector of the entries of the average-ideal document seem to be close, even an angle dissimilarity of four degrees corresponds to almost 0.03 of cosine, and this disparity here is noticeable as it equals almost the one third of the x -values of the High-CosS class (see again Figure 20). So, *The Wild Goose Chase* cannot be judged as highly Shakespearian from the scope of the counts of words' frequencies as it is situated close to the bottom-left part of the triangular class of high cosine similarity. In addition, there are many other complexities behind the term of the index of cosine similarity that will be discussed in the next two chapters. A non-Shakespearian play can even have an index of cosine similarity (in comparison with the average-ideal document) that is higher than a well-attributed Shakespearian play (*A Mad World, My Masters*) but if this index (0.94) does not fall close--in second decimal--to any of the detected patterns of the (12) Shakespearian indices of cosine similarities (see again Table 8), and if the sets' data points allow it, this can be even a sign of

non-Shakesperianness. A large difference of only a few words and a complete similarity in the frequencies counts of the majority of the 100 words of a play under scrutiny in comparison with the entries of an average-ideal document can affect dramatically the index of cosine similarity. On the other hand, a small difference of all 100 words from the scope of counts of frequencies can affect less severely the index of cosine similarity. In every case, the index of cosine similarity should be interpreted with caution.

The aim of the use of cosine similarity in this thesis is to create an additional discriminating threshold and we should always keep in mind the necessity of the measurement of the eventuality (of the paradox) of almost ‘completely’ similar sets’ counts (SiS1-High) but ‘excessively’ dissimilar counts of words’ frequencies (Low-CosSim) when examining a new play in comparison with the well-attributed Shakespearian plays and the three genre-based models. Furthermore, as the effort in this thesis is to support the discriminating power of the counts of words’ sets by employing the index of cosine similarity it can be shown that even if, in certain cases, counts of words’ frequencies do not discriminate style (for instance there is a small difference of cosine index of the average-ideal document with the *Wild Goose Chase* and with *The Tempest*, both around 0.91), counts of sets can.

Let us now add the three plays’ data points for the three inputs of the second layer of the fuzzy program.

Comedy title	Set Five-Total of Four Sets %	SiS1 (as Input)	Index of CosSim
<i>A Mad World, My Masters</i> (Middleton, 9,968 words)- Validation Stage	13.5	0.199	0.94

<i>The Wild Goose Chase</i> (Fletcher, 10,019 words)- Validation Stage	13.8	0.367	0.91
<i>Measure for Measure</i> (10,099 words)- Testing Stage	14.8	0.459	0.97

Table 20: Data points of three new plays under scrutiny.

The process followed is exactly the same as the one described in the previous subsection (4.9.1) for *The Staple of News* and *All's Well That Ends Well*. Let us see if the current set of 39 rules is adequate to produce SiS2 for the three plays or whether it is necessary to generate new rules. The current set of rules that can produce the SiS2 score are:

R30 IF 5a AND SiS1-High AND High-CosS THEN SiS2-High

R31 IF 5nb AND SiS1-Low AND High-CosS THEN SiS2-Low

R32 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Medium1

R33 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Low

R34 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium2

R35 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium1

As can be observed in the above rules, when Set Five of a new play is not Shakespearian then the SiS2 falls a level from SiS1 and SiS2 is produced from the truncation of the summed areas of SiS1 output class plus the class below that.

Let us start the experimentation in order to form the complete SiS interval (SiS1, SiS2) with the full set of rules for the first comedy under scrutiny, *A Mad World, My Masters*. The data points for the second stage are: 13.5 (count of total set), 0.199 (SiS1), 0.94 (CosS)

Index). The SiS2 that the system produces is 0.177-rule R31 fires-and, therefore, the results for *A Mad World, My Masters* are as follows:

<i>A Mad World, My Masters</i>	Which rule fires	Output
0.199	R36,R37	SiS1
0.177	R31	SiS2

Table 21: SiS1 and SiS2 of *A Mad World, My Masters*.

As the two SiS scores are 0.199 and 0.177, *A Mad World, My Masters* has an evidential interval of 0.177 to 0.199. The limits of the degree of Shakespearianness of this play falls into an SiS1-Low continuous interval of 0.177 to 0.199.

Let us now proceed to the next play and start the experimentation with *The Wild Goose Chase* in order to form the complete SiS interval (SiS1, SiS2) with the full set of rules for that comedy under scrutiny. The data points for the second stage are: 13.8 (count of total set), 0.367 (SiS1), 0.91 (CosS Index). The SiS2 that the system produces is 0.262 (R32, R33) and, therefore, the results for *The Wild Goose Chase* are as follows:

<i>The Wild Goose Chase</i>	Which rule fires	Output
0.367	R37	SiS1
0.262	R32, R33	SiS2

Table 22: SiS1 and SiS2 of *The Wild Goose Chase*.

Therefore, *The Wild Goose Chase* has the evidential interval of 0.262 to 0.367. The limits of the degree of Shakespearianness of this play are close to the SiS1-Low continuous interval of 0.262 to 0.367 (see design of SiS1-Low output class in 7.2.3).

Regarding the experimentation with the next play *Measure For Measure*, its data points do not trigger any of the current rules (R30-R35) of the algorithm that produces the SiS2 score (and so the default value of 0.5 is printed for SiS2.) The reason is that this play has its fifth set Shakespearian and because has not yet been formed any rule for that case it is necessary to apply the relevant to SiS2 algorithm and map new rules.

Let us first recall the membership functions-classes of the fifth set by looking at Figure 32:

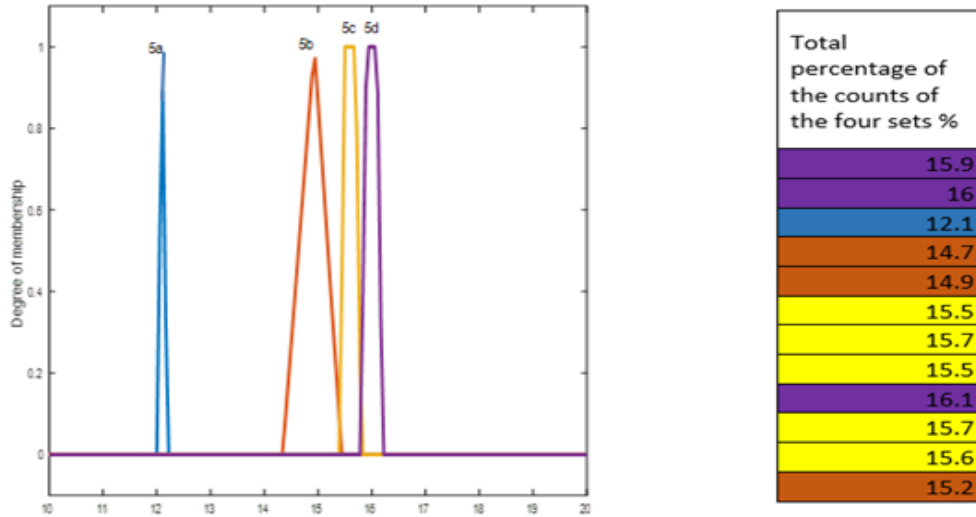


Figure 32: Classes of Set Five.

The data set of the fifth set for *Measure For Measure* is 14.8 % and falls into the Shakespearian class of ‘5b’ (see the triangle and cells in orange). The SiS1 (that will be used as input) is 0.459 and falls into the class of SiS1-Medium2 [0.4375-0.75]. In addition, the index of cosine similarity of *Measure For Measure* in comparison with the average-ideal document is also Shakespearian (0.97). It falls into the class of high cosine similarity [0.884-1]. Consequently, the two new rules to be added are:

R40 IF 5b AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium2

R41 IF 5b AND SiS1-Medium2 AND High-CosS THEN SiS2-High

These two rules (R40, R41) essentially are the components of a single rule described in the algorithm with the three additional input variables. What happens is that the SiS of the first layer (SiS1=SiS1-Medium2) with the four sets’ data points is updated with the evaluation of the three data points of the three new inputs (and so of the fifth set, too) and this process leads to the rise of one level (here to SiS1-High) from the current SiS1 level (here SiS1-Medium2) so that SiS2 is produced at the second layer of the fuzzy classifier based on the aggregated area of the previous (SiS1-Medium2) and the new updated (SiS1-High) output class. Practically, SiS2 will be now produced from the centroid of the truncated summed areas of the output classes of SiS2-Medium2 and SiS2-High. As a result, regarding the comedy *Measure For Measure*, SiS1 is 0.459 and SiS2 is 0.726.

<i>Measure For Measure</i>	Which rule fires	Output
0.459	R21, R39 (R4 and R38 have minimal effect)	SiS1
0.726	R40, R41	SiS2

Table 23: SiS1 and SiS2 of *Measure For Measure*.

With the experimentation of the five plays under scrutiny a complete Fuzzy Stylistic Classifier has been built with a total of 41 rules, some of which play off one another assisting the fuzzification process and the reduction of uncertainty by assessing if the plays' data points are somewhat actual and somewhat 'not/none' actual.

4.10 Validity of Performance and Testing Mechanism.

Assessing the SiS intervals for all five plays under scrutiny (*All's Well That Ends Well*, *Measure For Measure*, *The Staple of News*, *A Mad World*, *My Masters* and *The Wild Goose Chase*), the results tally with the general evidence of historical scholarships and the claim that *All's Well That Ends Well* and *Measure For Measure* are almost sole-authored Shakespearian plays, whereas *The Staple of News*, *The Mad World My Masters* and *The Wild Goose Chase* are non-Shakespearian plays. This stylistic classifier of approximate reasoning is based on the counts of sets of words as detected in the 12 well-attributed Shakespearian comedies. In addition, the sets-based part of the Fuzzy Stylistic Classifier does not contradict the Shakespearianness of the counts of words' frequencies and therefore does not produce the so-called paradoxical results (that is, the eventuality of almost 'completely' similar sets' counts (SiS1-High) but excessively dissimilar counts of words' frequencies (Low-CosSim)). The upper and lower bound values of the SiS interval of *The Staple of News* by Jonson, *A Mad World*, *My Masters* by Middleton and *The Wild Goose Chase* by Fletcher fall into or are located near the class of SiS1-Low [0-0.25]. If the SiS1 score is lower than the SiS2 score it is the lower bound and if it is higher than SiS2 it is the upper bound of the interval. The results of the experimentation with the five plays can be represented by a scatter plot where the lower SiS bound of each play under scrutiny is represented by the X-axis and the upper SiS bound by the Y-axis.

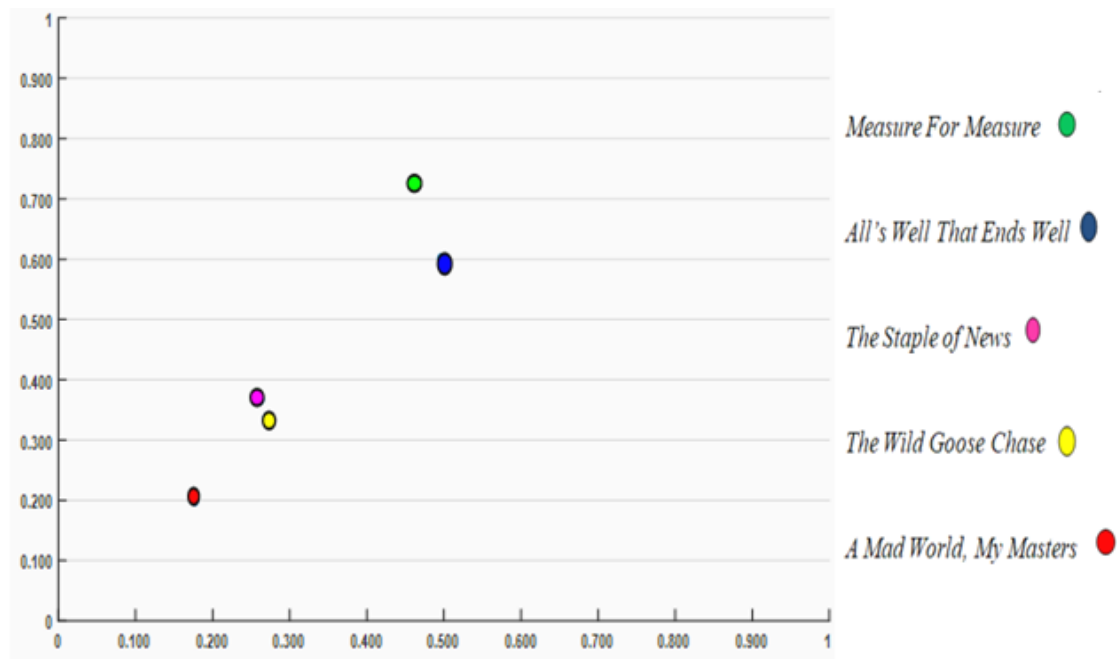


Figure 33: The scatter plot of the SiS interval of the data points of five comedies under scrutiny.

The comedies *All's Well That Ends Well* and *Measure For Measure* have at least one limit value (upper or lower) of their intervals situated in the SiS1-Medium2 area ([0.4375-0.75], see design in Sections 3.3 and 7.2.3). The goal of this core experimentation is to show that by employing the principles of Fuzzy Logic and counts of sets of words as stylistic markers it is possible to assess approximately (but with high interval-based certainty) Shakespearianess. It is also possible to assess the general stylistic tendency of comedies and if this tendency conforms to the features of the counts of the sets' data points as detected in the well-attributed, sole-authored plays of a specific author.

So, the general goal has been achieved, as by looking at Figure 33 the two plays (*All's Well That Ends Well*, *Measure For Measure*) are clearly distant (and so separated) from the three other plays that are classified as plays of low Shakespearianess. Moreover, another observation is that as the set of rules increases and approaches the number 40 (there are in total 41 rules), the inference mechanism becomes economical and there is a smaller need to generate new rules for the new plays as the existing combinations are adequate for the production of SiS scores of new plays whose combinations may coincide with the antecedents of the existing rules. Improvements and further discussion of how such a

genre-based model can be enriched at the second layer of the fuzzy simulator with new stylistic markers are discussed in the Chapter of Conclusions (Chapter Six).

4.11 Histories-Based Fuzzy-Stylistic Classifier.

This subsection contains the modelling and the experimentation with the well-attributed, sole-authored histories of Shakespeare. The histories-based classifier is based on the same method that was developed in the previous chapter and in most of the subsections, such as 4.1, of the comedies-based classifier. As with the comedies-based model, the building of the classifier is composed of three stages. First, the stage of the design of (possibly overlapping) class-membership functions of each input variable-data point of sets of words. Secondly, the stage of the formation of actual combinations of these classes in the well-attributed histories. And thirdly, the rules-based phase where the generation of rules takes place by associating the class-membership functions of inputs-total of stylistic markers (four semantic sets plus three additional variables) with the four output SiS1 classes. In every phase of classes-membership functions' modelling, the existence of different sequences or else the proximity of the data points of each set of words between them affects also the parametrisation of the design of the histories-based model.

4.11.1 The Histories-Based Fuzzy Classifier of Shakespearianness: Design of Membership Functions-Classes.

Let us now start building the second, histories-based, Fuzzy Stylistic Classifier of Shakespearianness. This model is based on the well-attributed, sole-authored six histories of Shakespeare.

HISTORIES (six plays)

Henry IV Part 1

Henry IV Part 2

Henry V

King John

Richard II

Richard III

For the building of this model, were employed again as stylistic discriminators the four semantic sets of words that were described in Section 2.4 and formed in the construction of the comedies-based classifier. An additional fifth set expressing the total of four sets' counts was also employed (similarly to the comedies-based experimentation). The counts of words' sets are represented in triangular or trapezoidal functions in accordance with the same design principles of the primary and comedies-based experimentation. Similarly to the previous models, the shapes of the membership classes are derived from the numbers of the data points of each set and the way these data are spread.

Looking at the table below (Table 24), data points in cells that have the same colour belong to the same class. In fact, there is one duplet (and so one trapezoidal class) and one triangular class for Set One, one duplet and three triangular classes for Set Two, two duplets and one triangular class for Set Three and two triplets for Set Four. Set Five is composed of three triangular classes.

History title	Set One %	Set Two %	Set Three %	Set Four %	Total percentage of the counts of the four sets %
<i>Henry IV Part 1</i>	2.3	8.4	1.5	2.9	15.1
<i>Henry IV Part 2</i>	2.1	8.3	1.7	2.3	14.4
<i>Henry V</i>	1.7	6.5	1.4	2.4	12
<i>King John</i>	2.1	7.6	2.2	2.5	14.4
<i>Richard II</i>	1.7	5.9	1.8	2.8	12.3
<i>Richard III</i>	1.8	8.6	2.2	2.7	15.3

Table 24: Data points of the five sets of the six Shakespearian histories.

I adopted the methodological steps set early in Chapter Three (Sections 3.2 and 3.3) and I also applied, when necessary, the relevant constraints (see in Section 3.2 the constraint a). Concerning the design of membership functions of the six histories, in certain cases, such as for Set Two (of the play *Richard III* with data point 8.6%) and for Set Five (of the plays *Henry IV Part 1* and *King John* both with data point 14.4%) I had to follow the approach of the hypothetical triangle. Recalling that constraint, it has been stated that if the distance of a x -value (a set's data point) to any other sets' data point above or below is more than 0.5 and there are more than one values before or after that value (so that these 'more than one' values can form a duplet, triplet or a non-hypothetical-triangular function), a separate single-point triangular class is created for the single, isolated x -value. This plays the role of the mean of two hypothetical values--one below $[-0.1]$ and one above $[+0.1]$ of this actual value-hypothetical mean which constitutes the apex of a separate membership function, an isosceles triangle of which the y -value is one and whose bottom-left and bottom-right corners are one SD above and one SD below this hypothetical mean.

In Set Two, as can be seen in the above Table 24, was met one particular case that was mentioned roughly in the design constraints set from the beginning in Section 3.2. There were more than two neighbouring isolated data points. Thus, there were three data points that each had a distance of more than 0.5 from a neighbouring value, forming an unmapped space of stylistic patterns. The particularity of the constraint for this specific sequence of three data points (non-continuous triplet) arises from the fact that the sets' data points (decimal numbers) theoretically can be spread in numerous different Euclidean distance-based ways between them and it would be uneconomical to define precisely all possible cases from the beginning. In other terms, the process of membership functions' design entails an implied variant mixture of combinatorics (that is, constraints with adaptability). These are influenced by the necessity to map out the sets' patterns for every detected sequence of data points taking into account the optimal continuity. Roughly speaking, this continuity reflects the distance of 0.1 for duplets and 0.1, 0.2 for triplets, or up to 0.5 for other triangular cases.

In order to adopt the same design principles, as in primary and comedies-based experimentation, I must find a way of representing the central stylistic tendency of the three neighbouring data points (5.9, 6.5, 7.6) of Set Two in the above Table 24 by taking into account the elaborate use of 'not/none' class in the core experimentation and without ignoring the distances (both in excess of 0.5) between them. The solution for the creation of membership functions for Set Two is based on a two-step process. The first step was the

creation of two triangular classes when there are no duplets or triplets present, but with a slight adaptation (see also the first constraint about the widely spaced data on page 100 in Section 7.2.1). I used twice the middle value (6.5) of the three data points (5.9, 6.5, 7.6) for the formation of two separate triangular classes that eventually overlapped. Instead of making only triangular actual classes, as will be seen below in step two, internal ‘not/none’ classes were also formed, and this adaptation facilitated the creation of rules that assess in a fuzzy mode data points that fall in specific intervals of x -values. In effect, as has been already seen in the comedies-based model, one rule (of an actual class) can play off one another rule (of a ‘not/none’ class) assisting the fuzzification process and the reduction of uncertainty by assessing the membership of (new) plays’ data points—that are 0.1 more distant from the three existing actual patterns of 5.9, 6.5, 7.6—in both the actual and ‘not/none’ actual class. Let us see the steps of the adaptation process regarding the creation of the membership classes for the area that is covered by these three data points (5.9 to 7.6).

Step One:

First, a matrix (M) was built with two vectors (V1, V2), of which the first is composed of the data points 5.9 and 6.5 and the second of 6.5 and 7.6 (In the case, V1 and V2 should be viewed as rows).

So, the data point 6.5 as the mid-point of the three values (5.9, 6.5, 7.6) was employed in two different vectors (and eventually for the creation of two different triangular classes.)

Then the mean and the first standard deviation of the data points of each vector were calculated.

The first mean (6.2) was employed for the formation of the first triangular class (‘2a’) and constitutes the apex of a separate membership function, an isosceles triangle of which the y -value is one and whose bottom-left and bottom-right corners are one SD below ($6.2 - 0.42 = 5.78 = x$ value) and one SD above ($6.2 + 0.42 = 6.62 = x$ -value) this mean of 6.2. The second mean (7.05) was employed for the formation of the second triangular class (‘2b’) and constitutes the apex of a separate membership function, an isosceles triangle of which the y -value is one and whose bottom-left and the bottom-right corners are one SD below ($7.05 - 0.78 = 6.27 = x$ -value) and one above ($7.05 + 0.78 = 7.83 = x$ -value) this mean of 7.05.

Figure 34 shows the membership functions of Set Two, and it is clear that these two first two triangular classes (‘2a’-the first triangle in light blue, ‘2b’-the second triangle in red) overlap and the problem of the existence of more than two (consecutive) isolated values has been resolved from the scope of the actual Shakespearianness.

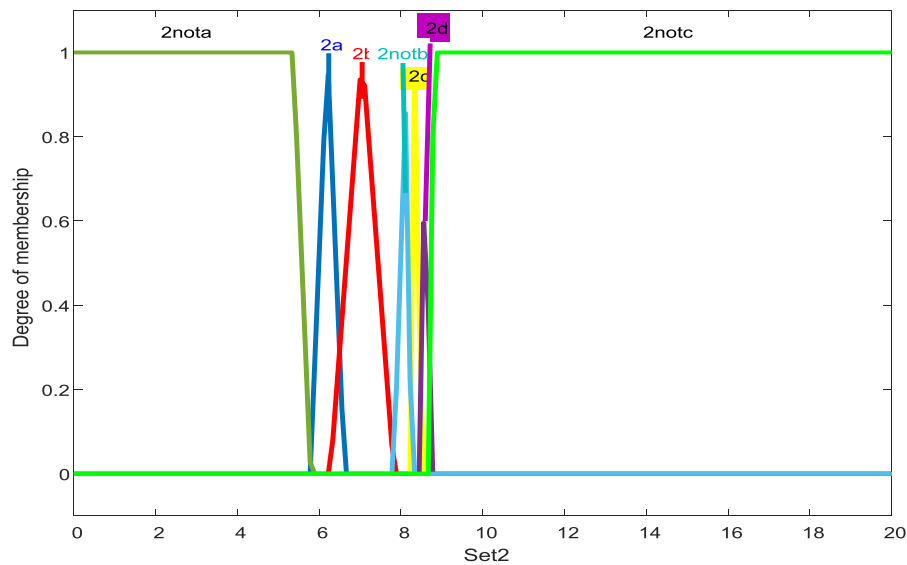


Figure 34: Basic membership functions for Set Two of the six Shakespearian histories.

Step Two:

At the same time, for the widely spaced data points 5.9, 6.5, and 7.6 I also applied the principle of the single-point-based hypothetical triangle (that were defined in the constraints on page 100 in 7.2.1 and also used in the comedies-based model) and I created three new (internal) ‘not/none’ subclasses. These three ‘not/none’ subclasses cover subareas of the range of x -values of the two actual classes ‘2a’, ‘2b’. Their interval of 5.78-7.83, and so (as parts of the large ‘actual’ interval) three ‘not/none’ subintervals are formed: 6.0-6.4, 6.6-7.5 and 7.7-7.83. These subintervals do not require the slight adaptation of 0.01 discussed in Section 7.3.14, because the ‘not/none’ classes’ edges are situated inside an area of an actual class. By forming in the interval 5.78-7.83 overlapping actual and ‘not/none’ actual subclasses to the formation of which contributed the non-continuous triplet 5.9, 6.5, 7.6 I manage to assess the lack of proximity of the three data points between them (> 0.5). Indeed, it seems that at the open continuous interval of the non-continuous triplet of 5.9, 6.5, 7.6--see 7.3.1 for the definition of intervals--it was necessary to apply the principle of partial truths-membership also in terms of the ‘not/none’ class. In fact, from the ‘not/none’ areas of the interval 5.78-7.83 formed by the two actual overlapping classes (‘2a’, ‘2b’) were excluded the single, isolated x -values of 5.9, 6.5, 7.6 plus one below (-0.1) and one above ($+0.1$) of the values of the triplet 5.9, 6.5, 7.6. Therefore, the complete set of values excluded from the classes of actual Shakespearianness was: 5.8, 6, 6.4, 6.6, 7.5, and 7.7. Consequently, in the interval 5.78-

7.83 of the two overlapping actual classes ('2a', '2b') of Set Two the decimal values a) 5.8, 5.9, 6, b) 6.4, 6.5, 6.6 and 7.5, 7.6, 7.7 are the only values with a membership in the actual Shakespearian class. The remaining values of the interval 5.78-7.83 are assessed both as actual and 'not/none'. This differentiation of actual with (internal) 'not/none' classes can be viewed by looking the triangles in black at the next two figures.

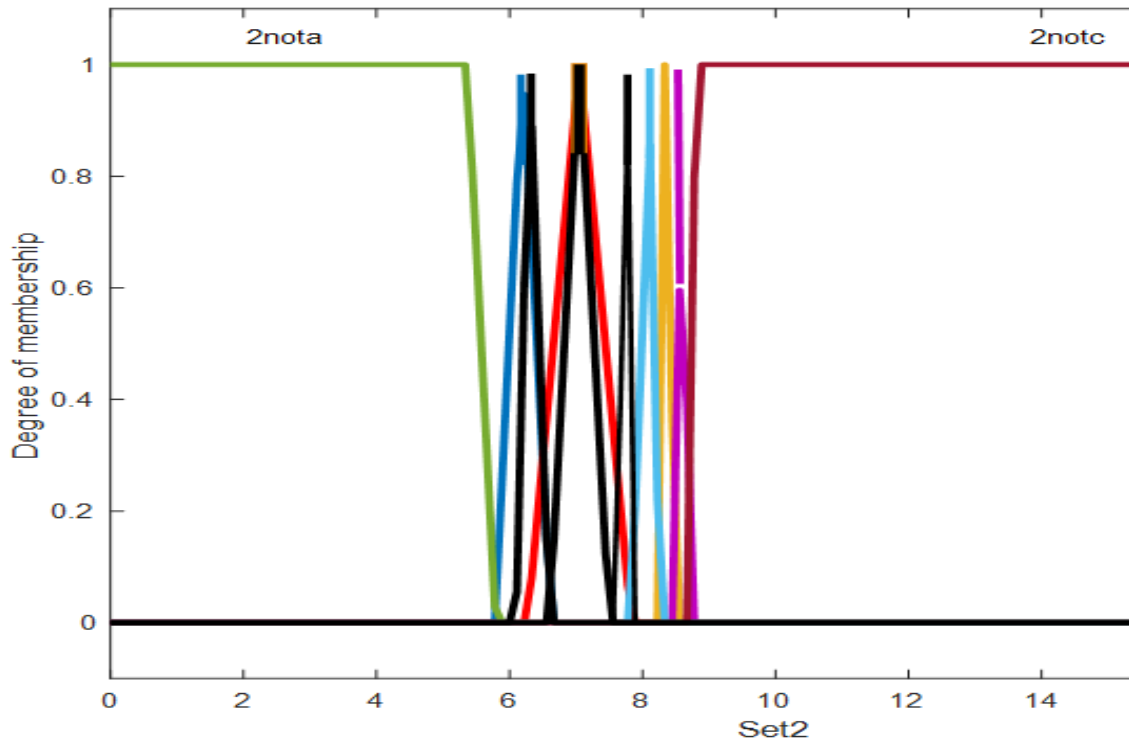


Figure 35: Membership functions including the 'not/none' classes (see black triangles) inside the classes '2a' (first blue triangle) and '2b' (red triangle) of Set Two of the six Shakespearian histories.



Figure 36: Zooming in the black triangles of the Figure 35.

Figure 35 and Figure 36 show that the three black triangles fall inside the area of the x -values of the classes '2a' (first triangle in blue) and '2b' (triangle in red). These three black triangular classes essentially are internal 'not/none' subclasses situated on ranges of x -values in the interval of actual classes 5.78-7.83. The 'not/none' parts-continuous intervals on the X -axis are 6.0-6.4, 6.6-7.5, and 7.7 to 7.83, whereas the full interval of '2a' and '2b' classes,

which overlap, is 5.78-7.83. This adaptation of the second step is necessary in order to evaluate in a fuzzy way any x -values that happen to fall into any part of the actual Shakespearian classes ('2a', '2b') but at the same time are somewhat distant (more than 0.1) from the detected three actual data points-patterns of 5.9, 6.5, and 7.6. For instance, if the data point of Set Two of a new play is 7.1 it will be evaluated for its membership in the actual '2b' class--somewhat actual-- but will be also assessed for its proximity-distance from the values of 6.5 and 7.6, somewhat 'not/none' actual.

In this experimentation the descriptive statistics of the comedies' sets' data points can be compared with those of the histories. Apart from the mean and standard deviation, the sets' data points of the histories are explored also from the scope of the relative standard deviation (RSD), which measures in percentage the spread of the data around the mean. The study of the RSDs can provide us with important information about the variance of the data points of each Set for a specific genre. The smaller a set's RSD is, the smaller the deviation of each set's counts from play to play, and, consequently, the smaller the variance of style regarding each set of words. The comparison of the RSDs of different sets of words can show which set's SD have the smallest distance from the mean.

Statistics of Histories-based model	Set One	Set Two	Set Three	Set Four	Set Five
Mean	1.95	7.55	1.80	2.60	13.92
STDEV	0.25	1.11	0.34	0.24	1.42
RSD %	12.87	14.77	18.92	9.10	10.20

Table 25: Mean, Standard Deviation and RSD of the five sets in the six histories.

Statistics of Comedies-based model	Set One	Set Two	Set Three	Set Four	Set Five
Mean	2.33	8.74	1.70	2.46	15.24
STDEV	0.21	0.97	0.20	0.27	1.07
RSD %	9.01	11.15	11.76	11.17	7.05

Table 26: Mean, Standard Deviation and RSD of the five sets in the 12 comedies.

In the comedies-based model (Table 26) the fifth set has the lowest RSD (7.05), as it is the set whose data points cluster more tightly around the mean. In histories (Table 25) though the Set Five has again a relatively low RSD (around 10%), it is Set Four that has the lowest RSD (9.1). Another interesting finding is that Set Three has again in histories as in

comedies the largest RSD (18.92) but it is much larger (around 60% more) than in the comedies-based model (11.76). The RSDs of Set One and Set Two are much larger in histories than in the comedies. An explanation for these discrepancies might be that the use of words that denote space-time (Set One) or Set Two's personal pronouns (and especially pronouns that indicate male and female actors) depend in each history on a number of factors that differ from play to play and affect an author's stylistic preferences when writing histories. Overall, Shakespeare uses at a more stable rate (and hence with smaller RSDs) the words of all five sets in the comedies than the histories. The explored well-attributed Shakespearian histories are half (6) the number of the comedies (12) and so someone might expect the preponderance of varied patterns for the comedies (12 comedies against six histories), given that they have twice the data in which to find them. Overall, this fact might be an implication that Shakespeare's style of writing in comedies is less variable and more consistent than that of histories.

By joining the tops of the bars of all RSD values of the five sets it can be viewed in the left graph of Figure 37 that there is kind of a symmetric function (skewed to the right) for Set One to Four with feet on Set One and Four and shoulders on Set Two and Three and this symmetricity is interrupted as the curve continues with the data point of Set Five. (In fact, the whole curve of the left graph in Figure 37 seems to be kind of an inverted z--function. The inverted z-function is a complex term of analytical science but what deserves to be noted here is that the peak in the left graph of Figure 37 could stand for a zero-point on the X-axis in case of a normal z-function.) The right graph of Figure 37 shows the curve of the comedies' RSDs, which was discussed in 4.1. It is also possible to investigate the importance of the symmetricity of the quantitative features of the selected stylistic markers.

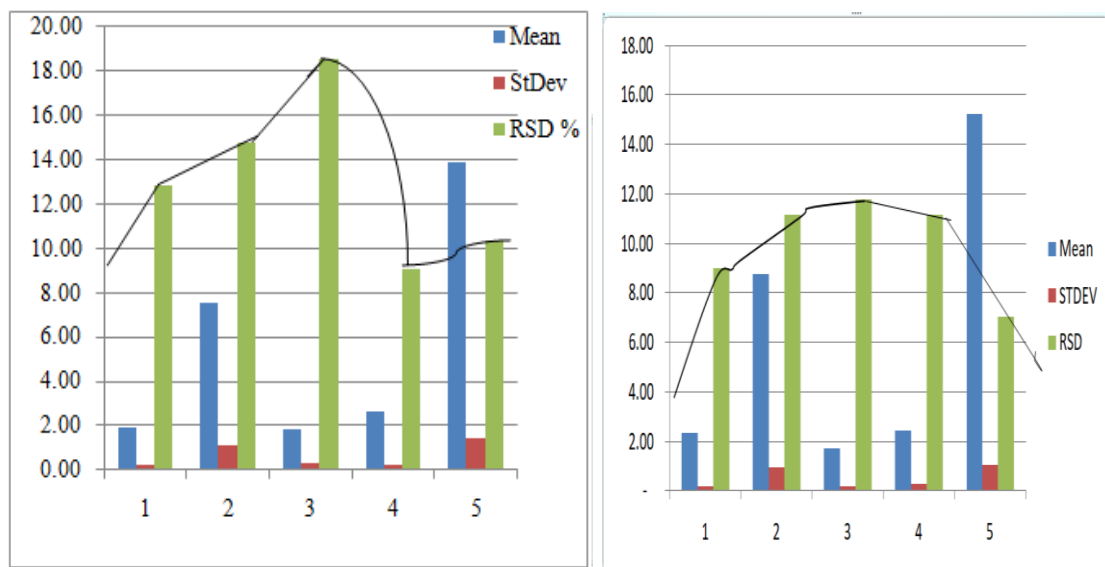


Figure 37: Y-axis expresses Mean and StDev in numbers. RSDs on Y-axis are measured in percentages. The numbers on X-axis correspond to the five sets. The left figure shows the tendencies for histories and the right for the previously analysed comedies-based model.

Let us now concentrate on Figure 38 below and view the RSDs of the two genres from the perspective of histograms of two variables.

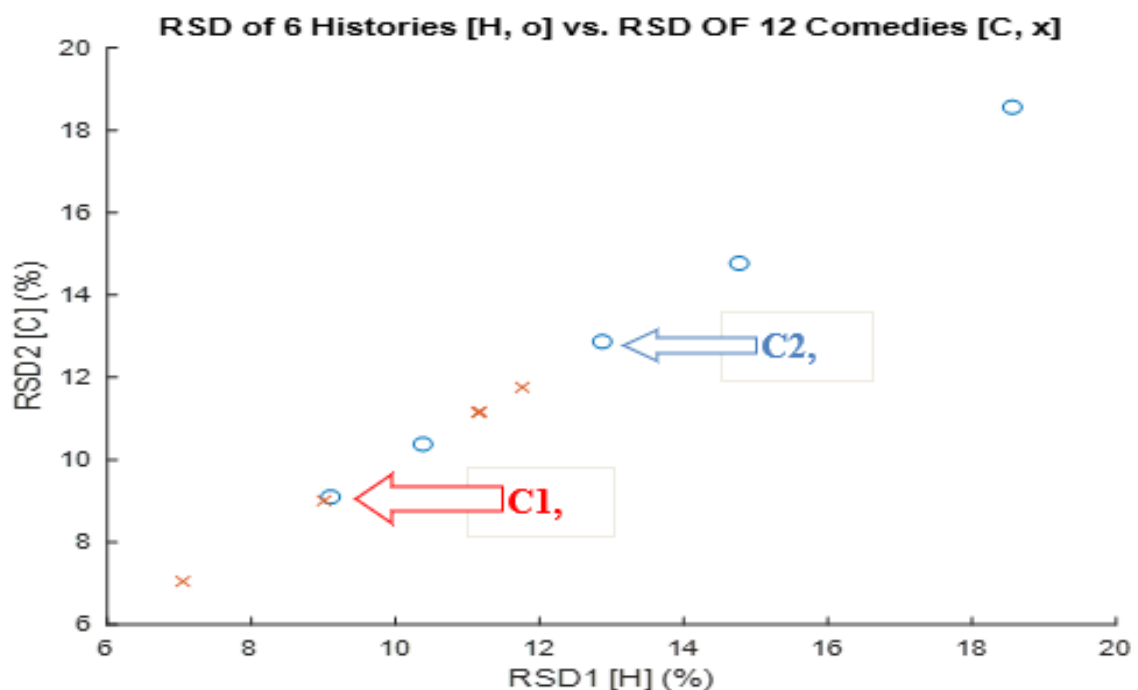


Figure 38: Comparison of the relative standard deviations (RSDs) of Comedies and Histories.

In Figure 38 there are a total of 10 data points, five representing sets' counts of comedies (marked as 'x') and five for the histories (marked as 'o'). Note that there are visible four instead of five red mark signs ('x') of comedies because the RSDs of Set Two and Four almost coincide (around 11.1 on the Y-axis) and so two red mark signs ('x') visually overlap. In Figure 38, the data points of the five sets' RSDs of comedies (indicated by the red mark sign 'x') are clustered much more tightly than those of histories (indicated by the blue mark sign 'o'). In other words, the sum of the distances of each set's RSD data point from the centroid of each set's data points (Centroid 1 [C1] for comedies and Centroid 2 [C2] for histories) is much smaller in comedies than in histories. This graph (Figure 38) confirms visually the observation made previously that Shakespeare's style of writing in comedies is less variant than that of histories as far as it concerns the measurement of Sets' data points. Let us now proceed to another comparative theorisation of the sets' data points as detected in the core experimentation.

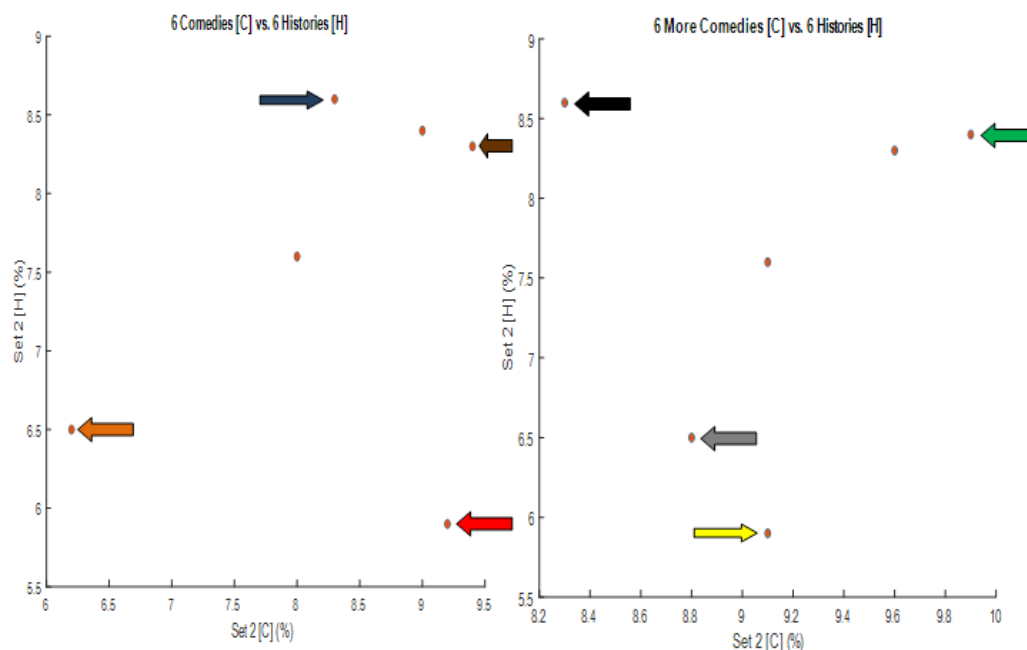


Figure 39: Histograms of data points of Set Two in Comedies (X-axis) and Histories (Y-axis).

In the above two graphs (Figure 39) Set Two's data points for the comedies [C] are represented by the X-axis and Set Two's data for the histories [H] by the Y-axis. The 12 Shakespearian comedies are divided into two sets of six plays (six comedies on the left graph

and six more comedies on the right graph). Each of the two distinct sets of comedies is compared with the same six Shakespearian histories. The red dots in the two graphs represent the total counts of the words' tokens of Set Two across a total of 18 plays (12 comedies and six histories.) The red dots are graphically pointed by coloured arrows for purposes of explanation. As can be viewed (Figure 39), the scale in each coordinate dimension differs. But we can see the correlation of Set Two's data points between each dimension or else between two the genres. For instance, look at the red dot point that in two graphs is situated on 6.4 of the *Y*-axis (look at the orange arrow on left graph and grey arrow on the right graph). In the left graph this value (6.4) on the *Y*-axis (Histories) corresponds to the value of 6.2 on the *X*-axis (Comedies) and on the right graph to the value of 8.8.

If there were only the six comedies on the left graph, and so the six histories were compared with only six comedies, the stylistic patterns of comedies and histories would look alike but with six more comedies this likeness is disrupted, as you can see in the right graph by comparing the different range of values on *X*-axis (comedies) and *Y*-axis (histories). On the left graph you can see that the data points of Set Two in histories range from 5.9 on the *Y*-axis (see red arrow) to 8.6 on the *Y*-axis (see blue arrow), whereas the data points in the first six comedies range from 6.2 on the *X*-axis (see orange arrow) to 9.4 on the *X*-axis (see brown arrow). On the right graph we can see that the data points of the same six histories range again from 5.9 on the *Y*-axis (see yellow arrow) to 8.6 on the *Y*-axis (see black arrow), whereas the data points of the second set of six comedies range from 8.3 on the *X*-axis (see black arrow) to 9.9 on the *X*-axis (see green arrow).

From the comparative graphical display of the two graphs it is clear that the higher counts of Set Two in comedies are 9.9 (see on the right graph the green arrow and *X*-axis) whereas in histories they are 8.6 (see blue arrow on the left graph and the black arrow on right graph, and the *Y*-axis). The lower counts in comedies are 6.2 (see orange arrow on the left graph and *X*-axis) whereas in histories they are 5.9 (see red arrow on the left graph and the *Y*-axis). By looking comparatively at the two graphs, it is possible to view the correlations of the two genres' data points. For example, look at the left graph: the first data point (see orange arrow) is at $x=6.2$ (a comedy) and $y=6.5$ (a history). The comparative view of the two 'dimensional' genres (x , y) in the two graphs assists in the identification of the distribution of the sets' data points as detected in comedies and histories. In terms of maximum, the higher counts of Set Two are detected in comedies.

By combining conceptually the technical principles of Fuzzy Logic with this graphical comparison it can be argued that for the building of a model's membership

functions-classes-Shakespearian patterns with which new, disputed plays are compared, it is not critical only the number of plays that form a genre-based canon (12 for comedies, six for histories) but there is complexity of several factors influencing the design properties of a fuzzy expert stylistic classifier. An interesting element is that the data points of 12 comedies cover a non-continuous interval of 3.6 % (9.9-6.2), and only six histories cover a non-continuous interval of 2.7 % (8.6-5.9). Though the histories' non-continuous interval is numerically smaller than that of comedies, considering their total number (only half of comedies), it is a quite large interval. A limited number of plays (six histories) in a relatively large non-continuous interval (5.9-8.6) have the disadvantage of containing evidence of Shakespearian styles from a small sample of plays whose data points are very dispersed (and form a large interval).

Set Two's data points of the 12 comedies map a larger interval (6.2 to 9.9) than the six histories (5.9 to 8.6). Since the ratio is two to one, (2:1, 12 comedies against six histories), it would be reasonable to expect a larger number of classes for the comedies-based model. But, the formed comedies' classes-membership functions are, in fact, fewer and more robust (than those of histories) since the data points of each class cluster more tightly around the mean of the total data points that contributed to the formation of each class. Overall, the two graphs display the fuzziness and complexities that arise from the extracted data of plays of different genres.

Let us now focus on the histories-based model and see which classes each of the play's data fall within for each of the five sets, expressed as a matrix of colours.

History title	Set One %	Set Two %	Set Three %	Set Four %	Total percentage of the counts of the four sets %
<i>Henry IV Part 1</i>	1b	2c	3a	4b	5c
<i>Henry IV Part 2</i>	1b	2c	3b	4a	5b
<i>Henry V</i>	1a	2a/2b	3a	4a	5a

<i>King John</i>	1b	2b	3c	4a	5b
<i>Richard II</i>	1a	2a	3b	4b	5a
<i>Richard III</i>	1a	2d	3c	4b	5c

Table 27: Data points of six well-attributed Shakespearian histories and their contribution to actual classes. Similarly-coloured cells represent the distinct classes.

Note in the above table the data point of Set Two in *Henry V* that contributed to the formation of two classes ‘2a/2b’, as I showed in the two steps process about the formation of two overlapped triangular classes. As discussed previously, the construction of these overlapped classes is based on the non-continuous triplet of three plays’ distant data points, that is 5.9, 6.5, 7.6 and the use of a middle value (6.5) for the formation of the two classes ‘2a’, ‘2b’ and the three internal ‘not/none’ classes of Figure 35 and Figure 36. Due to this adaptation there will be two SiS1-High combinations for *Henry V*, as we will see on the next page.

As in the design of comedies-based fuzzy classifier (4.1) and the primary experimentation (7.2), the process of generating membership functions from sets’ data points is represented with a matrix of colours. Looking at Table 27, for Set One there are two shapes=colours=classes and hence in the matrix the column for Set One has two background colours: a yellow background showing that *Henry V*, *Richard II* and *Richard III* contributed the data that formed the first (triangular) class (‘1a’) for Set One and a red background indicating that *Henry IV Part 1*, *Henry IV Part 2*, and *King John* contributed the data that formed the second (triangular) class (‘1b’). Recalling the principle explained above, each background colour represents a class (which is a membership function) and the data points of each column of five sets that share the same background colour have contributed to the same class-shape.

Looking at the number of the six histories’ actual shapes of each of four sets--we revisit the design of the classes of Set Five in Section 4.11.4--there are two possibilities for Set One times the four possibilities for Set Two times the three of Set Three times the two possibilities for Set Four (48 in total for the six histories, which is three times the 16 possibilities for the 12 comedies-based model.) But again, as in primary experimentation, I ignore (for now) the theoretically possible combinations because they were not combinations

‘selected’ by the six histories that were employed for the design of the membership functions-classes (see complete explanation in Section 7.2.2). So, for the time being, seven SiS1-High rules are derived from the sets-based stylistic patterns of six histories:

<i>Henry IV Part 1</i>	1b	2c	3a	4b
<i>Henry IV Part 2</i>	1b	2c	3b	4a
<i>Henry V</i>	1a	2a	3a	4a
	1a	2b	3a	4a
<i>King John</i>	1b	2b	3c	4a
<i>Richard II</i>	1a	2a	3b	4b
<i>Richard III</i>	1a	2d	3c	4b

Therefore, there are seven unique combinations and so the following seven SiS1-High rules (R1-R7) are formed:

R1 IF 1b AND 2c AND 3a AND 4b THEN SIS1-High
R2 IF 1b AND 2c AND 3b AND 4a THEN SIS1-High
R3 IF 1a AND 2a AND 3a AND 4a THEN SIS1-High
R4 IF 1a AND 2b AND 3a AND 4a THEN SIS1-High
R5 IF 1b AND 2b AND 3c AND 4a THEN SIS1-High
R6 IF 1a AND 2a AND 3b AND 4b THEN SIS1-High
R7 IF 1a AND 2d AND 3c AND 4b THEN SIS1-High

Looking at the script below (Figure 40) and the premise parameters of the input classes of Set One, it is clear that the data point of Set One (2.3) of the first play, thus *Henry IV Part 1*, is somewhat ‘not/none’ apart from actual (Sets’ data of all comedies have already been displayed in the beginning of this Section in Table 24). *Henry IV Part 1* constitutes the only case that the 2SD principle of coverage needs to be applied, as all other sets’ data points of the other five comedies fall inside only an actual class.

```

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a':'trapmf',[1.64 1.7 1.8 1.86]
MF2='1b':'trimf',[1.936 2.167 2.398]
MF3='1nota':'trapmf',[0 0 1.58 1.65]
MF4='1notb':'trapmf',[1.85 1.92 1.936 2.052]
MF5='1notc':'trapmf',[2.282 2.398 100 100]

```

```

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=10
MF1='2a':'trimf',[5.78 6.2 6.62]
MF2='2b':'trimf',[6.27 7.05 7.83]
MF3='2c':'trapmf',[8.23 8.3 8.4 8.47]
MF4='2d':'trimf',[8.49 8.6 8.71]
MF5='2nota':'trapmf',[0 0 5.36 5.79]
MF6='2notb':'trimf',[7.82 8.16 8.24]
MF7='2notc':'trapmf',[8.69 8.8 100 100]
MF8='2notd':'trimf',[6 6.2 6.4]
MF9='2note':'trimf',[6.6 7.05 7.5]
MF10='2notf':'trimf',[7.7 7.765 7.83]

```

```

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=7
MF1='3a':'trapmf',[1.33 1.4 1.5 1.57]
MF2='3b':'trapmf',[1.63 1.7 1.8 1.87]
MF3='3c':'trimf',[2.09 2.2 2.31]
MF4='3nota':'trapmf',[0 0 1.26 1.34]
MF5='3notb':'trimf',[1.57 1.6 1.64]
MF6='3notc':'trapmf',[1.86 1.94 2 2.11]
MF7='3notd':'trapmf',[2.29 2.4 100 100]

```

```

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=4
MF1='4a':'trapmf',[2.19 2.3 2.5 2.61]
MF2='4b':'trapmf',[2.59 2.7 2.9 3.01]
MF3='4nota':'trapmf',[0 0 2.1 2.21]
MF4='4notb':'trapmf',[2.99 3.1 100 100]

```

Figure 40: Script for actual and ‘not/none’ classes of the four sets’ data points. With particular focus on Set One [Input1].

So it is necessary to form a new additional rule:

R8 1nc AND 2c AND 3a AND 4b THEN SiS-Medium2

4.11.2 Inference Mechanism of the Histories-Based Fuzzy Classifier of Shakespearianess.

Apart from the SiS1-High rules, there is a necessity as with the comedies-based model to define an algorithm adding rules that invoke the ‘not/none’ classes. This is necessary in

case a new play's data fall within some but not all the shapes specified in the above seven SiS1-High rules. As before, except for the two steps process regarding Set Two's widely spaced data points, outside of the shapes that represent the data from the six histories are areas that have been defined by 'not/none' functions. (See the explanation for the overlapping of actual with 'not/none' classes in Sections 7.3.7 and 7.3.8 of the Technical Appendix.) When a new play is not quite in the class of most Shakespeare-like (SiS1-High), it falls instead in a lower SiS1-level. In general, the 'not/none' classes of the histories' fuzzy classifier contain the x -values outside of the range of x -values that fall within each (actual) class. (See also the adaptation in Section 7.3.14 of the Technical Appendix.) The design of the 'not/none' classes of the data points of the six histories meets the criteria set in the constraints of Section 3.2 and, if data points allow, there can be an overlapping of the actual with the left and right neighbouring 'not/none' subclasses. And, the closer the x -values are to the left/right actual class, if any, the less their membership in the 'not/none' class. The algorithm for the derivation of rules follows the same approach used in the comedies-based model: it is the principle of one-level-fall, as in Section 7.3.13.

At this stage, it is necessary to add the 'not/none' rules. First of all, there are 48 ways ($2 \times 4 \times 3 \times 2$) that a play could in theory match the actual Shakespearian shapes but if we add one negative ('not/none') to the existing actual shapes for each of the four sets there are 180 ways ($3 \times 5 \times 4 \times 3$) to combine the negated and actual shapes. In fact, only seven of the total 48 ways of combining actual shapes (not including negated sets) are seen in the Shakespearian real-world of the histories, the ones used above to generate the rules leading to a SiS1-High classification. Therefore, similar to the process adopted in primary experimentation and the comedies-based fuzzy classifier, it is necessary to discount the block of 48 SiS1-High rules. Of the possible 180 combinations, seven correspond to the SiS1-High classification of real-world histories and 41 correspond to hypothetical SiS1-High matches that we have not (yet) seen in real-world histories.

If data points of any new play in the validation or testing stage match with possible actual combinations of the Shakespearian four sets' classes, then the real world of Shakespearian plays will be updated with a new SiS1-High rule, in addition to the current seven SiS1-High rules. For now, that leaves 138 ($180 - (48 - 7) - 1$) possible ordered combinations, since we subtract from the total of 190 initial ordered combinations the 41 hypothetical SiS1-High matches that we have not yet seen in real-world histories and the one 'not/none' combination of *Henry IV Part 1*. I will follow exactly the same methodology with the comedies-based model and therefore in order to enlarge the database of rules (currently

seven SiS1-High) it is necessary to form the possible combinations of data points for the total of four sets as a tree structure in which the root node has four branches emerging from it, each one of which represents one of the following four possibilities:

- 1) That for one set the matched shape is 'not/none'
- 2) That for two sets the matched shapes are 'not/none'
- 3) That for three sets the matched shapes are 'not/none'
- 4) That for four sets the matched shapes are 'not/none'

Each of these selections represents a collection of ways in which the criterion it embodies can be met (See Section 7.2.2 in the detailed primary experimentation). The selection is made here, as before in the primary experimentation and comedies-based model, considering that Set One has two actual shapes and one 'not/none' shape, Set Two has four actual shapes and one 'not/none', Set Three one actual and one 'not/none' shape and Set Four two actual and one 'not/none'. (Note that the aggregation of 'not/none' class experimentation is technically complex and entails the notion of subclasses, see Section 7.3.14 of the Technical Appendix.) I follow the same search and harvesting process adopted in comedies-based model (Section 4.2) and the primary experimentation (Section 7.2.2). Taking into account the first (or single) actual classes of the sets-inputs, the six resulting combinations of the rule's antecedents with 'not/none' ('n') classes are:

1na 2a 3a 4a
 1a 2na 3a 4a
 1na 2na 3a 4a
 1n 2a 3na 4a
 1n 2na 3na 4a
 1n 2na 3a 4na

These antecedents are turned into rules by adding the appropriate consequences (SiS1-Low, SiS1-Medium1, SiS1-Medium2 and SiS1-High). By applying the algorithm of the special case of Set Two and the principle of one-level-fall (7.3.13) from the SiS1-High for the negated Sets One, Three and Four, six new rules are formed (R7 to R12):

IF 1na AND 2a AND 3a AND 4a THEN SiS1-Medium2

IF 1a AND 2na AND 3a AND 4a THEN SiS1-Low
 IF 1na AND 2na AND 3a AND 4a THEN SiS1-Low
 IF 1na AND 2a AND 3na AND 4a THEN SiS1-Medium1
 IF 1na AND 2na AND 3na AND 4a THEN SiS1-Low
 IF 1na AND 2na AND 3a AND 4na THEN SiS1-Low

After this, there are in total 14 rules (7 SiS1-High + 1 ‘not/none’ of *Henry IV Part 1* + 6 ‘not/none’ rules):

R1 IF 1b AND 2c AND 3a AND 4b THEN SiS1-High
 R2 IF 1b AND 2c AND 3b AND 4a THEN SiS1-High
 R3 IF 1a AND 2a AND 3a AND 4a THEN SiS1-High
 R4 IF 1a AND 2b AND 3a AND 4a THEN SiS1-High
 R5 IF 1b AND 2b AND 3c AND 4a THEN SiS1-High
 R6 IF 1a AND 2a AND 3b AND 4b THEN SiS1-High
 R7 IF 1a AND 2d AND 3c AND 4b THEN SiS1-High
 +
 R8 1nc AND 2c AND 3a AND 4b THEN SiS-Medium2
 +
 R9 IF 1na AND 2a AND 3a AND 4a THEN SiS1-Medium2
 R10 IF 1a AND 2na AND 3a AND 4a THEN SiS1-Low
 R11 IF 1na AND 2na AND 3a AND 4a THEN SiS1-Low
 R12 IF 1na AND 2a AND 3na AND 4a THEN SiS1-Medium1
 R13 IF 1na AND 2na AND 3na AND 4a THEN SiS1-Low
 R14 IF 1na AND 2na AND 3a AND 4na THEN SiS1-Low

The output variable of the histories-based Fuzzy Stylistic Classifier is designed in the same way as in the comedies-based model (Section 4.1) and the primary experimentation, and it is formed by the four trapezoidal membership functions of the classes of the Low, Medium1, Medium2 and High index of Shakespearian Similarity (SiS). As described in more detail in Section 3.3, the design of the output variable hinges upon the concept of the mean of maxima (MOM) and four overlapping classes-intervals ([0-0.25], [0.1875-0.50], [0.4375-0.75], [0.6875-1]).

4.11.3 Validation Stage 1: Validation of Inference Mechanism.

Let us now see how the histories-based fuzzy program with the current inference mechanism of 14 rules classifies the six well-attributed histories, whose four sets’ data points

contributed to the modelling parameters of the first layer of the fuzzy classifier (SiS1) of histories.

History title	Which Rule fires	SiS1
<i>Henry IV Part 1</i>	1, 8	0.794
<i>Henry IV Part 2</i>	2	0.877
<i>Henry V</i>	3, 4	0.86
<i>King John</i>	5	0.86
<i>Richard II</i>	6	0.859
<i>Richard III</i>	7	0.887

Table 28: Validation SiS1 scores of the six Shakespearian histories.

Below for illustration purposes are displayed the two rightmost columns, of which the first produces the SiS1 result of *Henry IV Part 1* (SiS2 in the last column defaults to 0.5 as there have not been generated any relevant rules yet for the production of SiS2 score.)

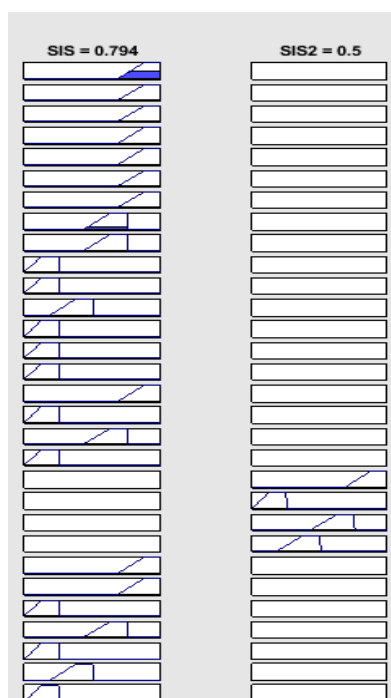


Figure 41: *Henry IV Part 1*'s SiS1-score (0.794) produced by rules R1 and R8. SiS2 defaults to 0.5.

In general terms, as can be deduced by the above SiS1 scores, a robust and coherent model is built, as all SiS scores for the six histories are indicative of high Shakespearianness.

As stated already in the comedies-based model, SiS1 corresponds to the SiS score of the primary experimentation, and this score is produced before the further evaluation (SiS2) in the second layer of the fuzzy system with the three additional input variables-stylistic markers (fifth set of words, SiS1 as input and CosSim).

The above results in Table 28 are produced with the set of the seven SiS1-High rules, the one ‘not/none’ rule (of *Henry IV Part 1*) and the six ‘not/none’ rules generated through the harvest and search process. As it was explained before, when the classes are formed by the two SD below and above the mean, the ‘not/none’ class overlaps with the actual class from the end of the first to the end of the second SD below and above the mean. Consequently, if the data point of any of the six well-attributed histories has an x -value after the first SD, then the actual class is extended to the second SD and this data point is assessed also as a ‘not/none’ class. Regarding the production of the SiS1 scores, the centroids of all six histories (unlike the centroids of the comedies) give on the X -axis values that lie within the class of SiS1-High [0.6875-1] and in fact lie within the part (0.76 to 1) of the class of SiS1-High that does not overlap with the SiS1-Medium2. The interpretation for these high results (SiS1-High) in the first stage of validation derives from the fact that the sets’ data points, apart from Set Two, of the six histories are closely bunched and because they are represented mainly by trapezoidal classes there is no need, apart from the case of the data point of Set One of *Henry IV Part 1*, to employ the second standard deviation as the principle of full coverage is met with the exact patterns-data points (X -axis) of duplets or triplets that form the upper side of the trapezoids ($y=1$).

It is necessary to gather some more data and look at the comparison of the counts of the individual 100 words’ frequencies of each vector-history with the counts of the average-ideal Shakespearian document-vector of the six histories. The angle formed between the two documents-vectors each time is assessed in terms of the cosine function (see Section 1.3.1.2.2 and the explanation of Table 8 in Section 4.3). The indices of cosine similarity of each vector (of words’ frequencies’ counts) of the six histories in comparison with the average-ideal Shakespearian document-vector are displayed below:

History title	Cosine index of similarity with the average-ideal history
<i>Henry IV Part 1</i>	These sets are not orthogonal cosine similarity = 0.98 angle = 11.51 angle type = acute
<i>Henry IV Part 2</i>	These sets are not orthogonal cosine similarity = 0.96 angle = 15.41 angle type = acute
<i>Henry V</i>	These sets are not orthogonal cosine similarity = 0.93 angle = 20.99 angle type = acute
<i>King John</i>	These sets are not orthogonal cosine similarity = 0.92 angle = 23.70 angle type = acute
<i>Richard II</i>	These sets are not orthogonal cosine similarity = 0.96 angle = 17.00 angle type = acute
<i>Richard III</i>	These sets are not orthogonal cosine similarity = 0.97 angle = 13.22 angle type = acute

Table 29: Cosine index of similarity of the six well-attributed histories with the average-ideal history.

As indicated by the results above (Table 29), all the plays as vectors of the counts of 100 words' frequencies in comparison with the vector of the respective entries of the average-ideal document (see Section 7.3.9 of the Technical Appendix) have a cosine similarity which is more than 0.92 (the minimum of all the cosine indices.) In fact, in comparison with the entries-words of the vector of the average-ideal document, the minimum cosine index of 0.92 for the six histories is the index of *King John*, and the maximum is 0.98 for the play *Henry IV Part 1*. It is interesting that *King John*, from the scope of words'

counts, seems to be the history with the least Shakespearian style. This evidence, after reconsidering the plays' isolated value of Set Two's counts, can be further explored in connection with historical scholarship and the claim that some parts of this play probably have been written by other authors, for instance George Peele (Merriam 2016; 2017; 2018). We should recall that *King John*'s counts for the second set have been included in the problematic non-continuous triplet of data points that contributed to the formation of '2a' and '2b' classes of Set Two. As with the comedies-based model, the indices of the cosine similarity together with the data points of the fifth set and SiS1 as input in the second level of the fuzzy-system produce the SiS2 score. SiS2, together with SiS1--produced from the four sets' counts--form the evidential interval of Shakespearianness. Simultaneously, a second general stylistic marker, the total counts of four sets, treated as fifth set, is included in the second layer of the fuzzy program.

Overall, in this subsection after experimenting with each of the six well-attributed, sole-authored Shakespearian histories I validated the inference mechanism of the first layer of the Fuzzy Stylistic Classifier based on the four sets' data points. I extracted all other necessary data, such as each history's cosine similarity index in comparison with the vector of the average-ideal document representing the average style of the six well-attributed histories. In the next subsection will be discussed the design of the second layer of the fuzzy classifier (SiS2 stage), which is based on the three additional stylistic markers-input variables (total of sets' counts, SiS1 as input, cosine index).

4.11.4 The Three Additional Input-Variables of the Two-Layered Complete Fuzzy-Logic Stylistic Classifier of Histories.

The second layer of the fuzzy system uses as inputs: 1) the frequencies of occurrence of a fifth set of words (being the fifth variable in the whole fuzzy system), 2) the SiS1 output (as the sixth variable in the whole fuzzy system), and 3) the index of cosine similarity to an average-ideal document (as the seventh variable in the whole fuzzy system).

The fifth set of words is composed of all the words in Set One to Set Four, and the membership classes are designed for this input in the usual way by processing the actual sets' data points, that is the frequencies of occurrences of these words in the six Shakespearian histories. Below are the data points of the fifth set and its membership functions: three triangular actual classes, '5a', '5b' and '5c'. Notice that there is a correspondence of the colours of the cells of the right column with the respective coloured membership class.

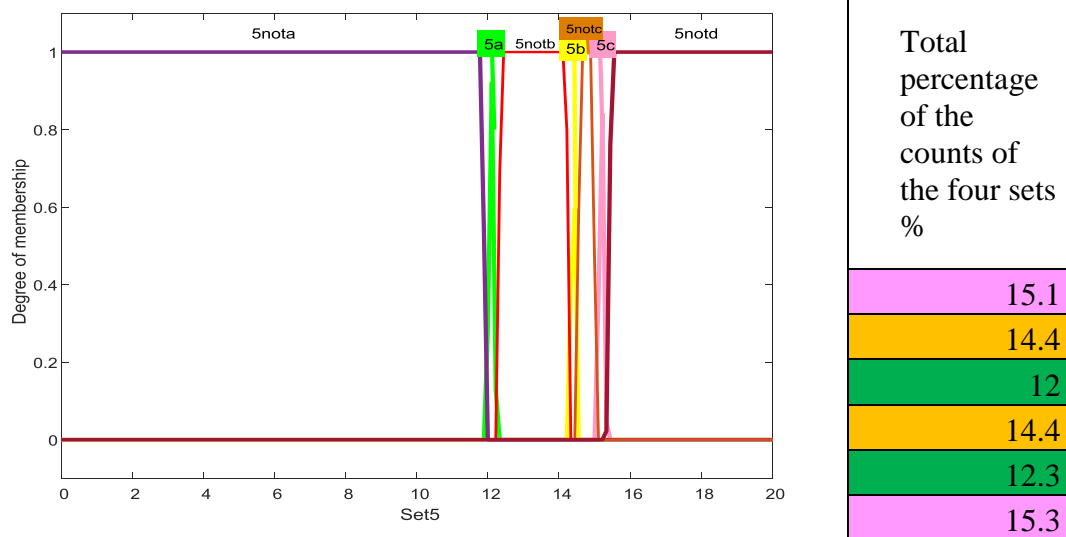


Figure 42: Memberships and data points of Set Five in the six histories.

This fifth set aggregates the counts for Sets One to Four. As you can see in Figure 42, based on the counts of the data points of the six well-attributed histories, the non-continuous interval of Set Five ranges from 12 to 15.3, whereas for comedies, as we saw in Section 4.4, the range was from 12.1 to 16.1. The interval of Set Five in Histories, as in comedies, covers a much larger area of ‘not/none’ Shakesperianness than any of the four semantic Sets. As for the output class of the SiS1, the classes are exactly the same with the classes of the output variable employed in the primary (3.3, 7.2.3) and core experimentation with the comedies (4.2).

Let us now, as in the comedies-based model, design the input variable of the index of cosine similarity for the second layer of the histories-based Fuzzy Stylistic Classifier of Shakesperianness. Below are the descriptive statistics that occurred after the comparison of the indices of cosine similarity of each of the six well-attributed Shakesperian histories with that of the average-ideal document (7.3.9):

Average index (mean) of the indices of Cosine Similarity (CosSim) of six histories	0.95
Standard Deviation	0.023
Minimum index (CosSim)	0.92
Left Corner (= Minimum – (minus) One Standard Deviation before the mean) of the actual triangular class of the Cosine Similarity	0.897

Table 30: Descriptive statistics of the index of Cosine Similarity of the six histories.

Based on the above evidence, and applying the methodology of the comedies-based experimentation, two classes will represent the actual and ‘not/none’ areas of the input variable of the index of cosine similarity of the six well-attributed Shakespearian histories. One class will be named the class of low cosine similarity (Low-CosSim), the non-Shakespearian class, and the second class will be the class of high cosine similarity (High-CosSim). As shown in Figure 82, the class of high cosine similarity will be a right trapezoidal class with its bottom-left corner at $x=0.897$ (which is one SD below the mean) and $y=0$, and its bottom-right corner will be $x=1, y=0$. The left shoulder point (with $y=1$) of the trapezoidal class of High-CosSim equals on the X-axis the average (0.95) of all values of the six cosine indices. The right shoulder point (again with $y=1$) on the X-axis is the value 1 and hence is directly above the bottom corner of the trapezoid. To the left of the trapezoidal High-CosSim class is the class of low cosine similarity (Low-CosSim) and it is represented by an almost orthogonal triangle with its bottom left corner at $x=0, y=0$. The peak ($y=1$) of this triangle gives on the X-axis the number 0.897. The bottom-right corner of this low-class triangle gives on the X-axis the value 0.898, thus it is located on X-axis 0.001 after the bottom-left corner of the triangular class of high cosine similarity (High-CosSim).

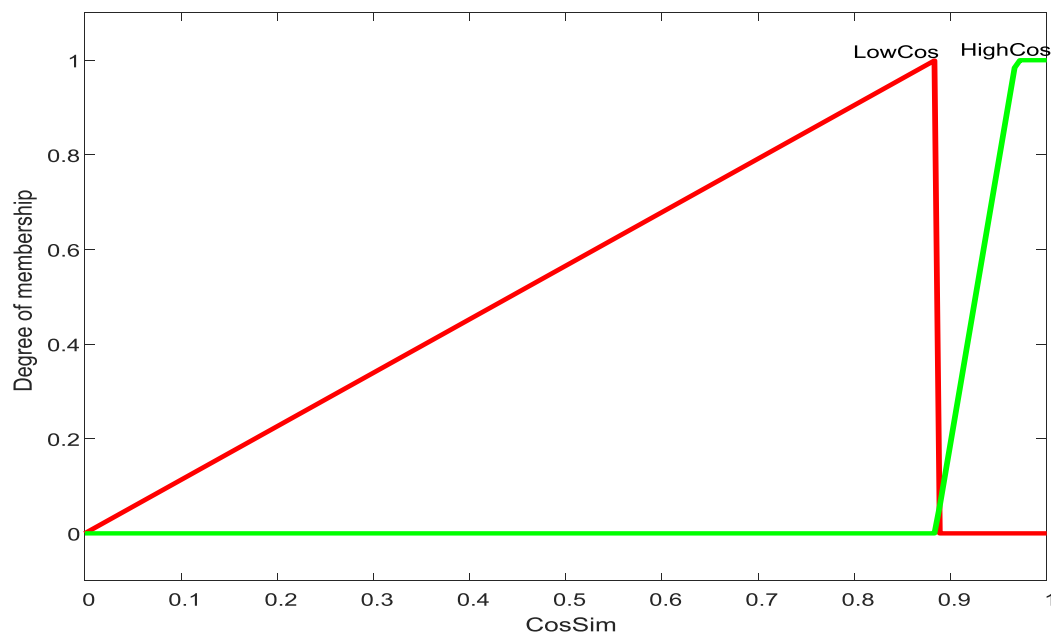


Figure 43: Membership functions of Cosine Similarity.

As the membership functions of the three additional input variables of the second layer of the histories-based stylistic classifier have been now designed, it is now feasible to proceed to Validation Stage 2 by experimenting with a non-Shakespearian play.

4.11.5 Design of the Validation Stage 2, Validation of Performance: Mapping A Non-Shakespearian Well-attributed Shakespearian Play and Further Development of the Inference Mechanism of the Histories-Based Fuzzy Stylistic Classifier.

As with the comedies' experimentation, in this stage are mapped the data points of a well-attributed, sole-authored non-Shakespearian play in order to validate the performance of the automated stylistic classifier by using as input to the system the new plays' data points. A play that will be used is the history play *Edward II* by Christopher Marlowe, which was first published in 1594 and because it is widely considered to be well-attributed and of his sole authorship (Boas 1960; Parks 1999; Craig and Kinney 2009, 215).

Let us recall that has been formed already a set of 14 rules (seven SiS1-High + one 'not/none' rule (of *Henry IV Part I*) + six other 'not/none' rules):

R1 IF 1b AND 2c AND 3a AND 4b THEN SiS1-High
R2 IF 1b AND 2c AND 3b AND 4a THEN SiS1-High
R3 IF 1a AND 2a AND 3a AND 4a THEN SiS1-High
R4 IF 1a AND 2b AND 3a AND 4a THEN SiS1-High
R5 IF 1b AND 2b AND 3c AND 4a THEN SiS1-High
R6 IF 1a AND 2a AND 3b AND 4b THEN SiS1-High
R7 IF 1a AND 2d AND 3c AND 4b THEN SiS1-High
R8 IF nc AND 2c AND 3a AND 4b THEN SiS1-Medium2

R9 IF 1na AND 2a AND 3a AND 4a THEN SiS1-Medium2
R10 IF 1a AND 2na AND 3a AND 4a THEN SiS1-Low
R11 IF 1na AND 2na AND 3a AND 4a THEN SiS1-Low
R12 IF 1na AND 2a AND 3na AND 4a THEN SiS1-Medium1
R13 IF 1na AND 2na AND 3na AND 4a THEN SiS1-Low
R14 IF 1na AND 2na AND 3a AND 4na THEN SiS1-Low

Let us now map the four sets' data points for *Edward II* to see if they trigger any of the existing 14 rules or it is necessary to form new rules based on the algorithmic principle of one-level-fall regarding the first layer of the fuzzy program (see Section 7.3.13 of Technical Appendix).

As you can see in Table 31, the produced sets' data points of *Edward II* are:

Play for Validation Stage	Set One %	Set Two %	Set Three %	Set Four %	Set Five- Total of Four Sets %
<i>Edward II</i>	2.2	9.1	1.2	2.3	14.8

Table 31: Five Sets' data points of *Edward II*.

Before producing the new rules, let us compare visually the sets' data points of *Edward II* of Table 31 with the sets' data of the six well-attributed Shakespearian histories that were previously (See Section 4.1) analysed in the coloured matrix below.

History title	Set One %	Set Two %	Set Three %	Set Four %	Total percentage of the counts of the four sets %
<i>Henry IV Part 1</i>	2.3	8.4	1.5	2.9	15.1
<i>Henry IV Part 2</i>	2.1	8.3	1.7	2.3	14.4
<i>Henry V</i>	1.7	6.5	1.4	2.4	12
<i>King John</i>	2.1	7.6	2.2	2.5	14.4
<i>Richard II</i>	1.7	5.9	1.8	2.8	12.3
<i>Richard III</i>	1.8	8.6	2.2	2.7	15.3

The data point of Set Two of *Edward II* is numerically larger than any of the counts of the respective set of any of the six well-attributed Shakespearian histories' data. For Set Three it is smaller than any of the respective sets' data points of the six Shakespearian histories. The counts, 14.8 %, of Set Five of *Edward II* are also clearly non-Shakespearian, as they do not fall into any of the three coloured classes formed respectively by the points 12-12.3, 14.4-14.4 and 15.1-15.3. Only data points of Set One and Set Four of *Edward II* have a

membership into an actual Shakespearian class of the respective set modelled on the six well-attributed histories. Recalling the analysis of RSDs in Section 4.11.1, where, as stated, Sets Two and Three have the largest RSDs, it might be deduced that counts of sets' data points with large RSDs (somewhat more than 10% and less than 20%) provide us with much discriminating information about an author's style of writing.

Crosschecking the data points of *Edward II* to see if they fall into actual or 'not/none' actual classes (or both, if that is allowed by the spread of the data and the membership functions), the combination of the actual and 'not/none' classes of *Edward II* is:

1b 2nc 3na 4a (where 'na' stands for 'not/none' 'a' subclass)

I turn this antecedent into a rule by adding the appropriate consequence (SiS1-Low) according to the algorithm of one-level-fall as described in 4.2 and 7.3.13, and consequently only one new rule must be added.

R15 1b AND 2nc AND 3na AND 4a THEN SiS1-Low

So, a total of 15 rules have now been formed (the previous 14 plus one new rule generated by the data points of *Edward II*).

Let us look at the index of cosine similarity of the entries of the histories-based average-ideal document-vector with those of *Edward II*:

<i>Edward II</i> - 10,028 words Validation Stage	cosine similarity = 0.953 angle = 17.548 angle type = acute
---	---

After extracting all data for *Edward II*, updating the set of rules and designing the next processing stage, the components of the second layer of the fuzzy classifier of histories can be built.

4.11.6 Design of the Testing Stage and Further Development of the Inference Mechanism of the Histories-Based Fuzzy Stylistic Classifier.

In this stage I map the data points of the disputed play *Henry VI Part 1*. From the primary model-classifier has already been produced an SiS1 score of 0.62 (see 7.2.4). As discussed in the literature review (1.2.1), Edmond Malone (1704-1774) argued that a large portion of *Henry VI Part 1* does not conform to the Shakespearian style of writing and also claimed that this history is a posterior version of a lost anonymous play that Shakespeare adapted (Malone 1787). Since then, similar claims have been expressed by many other researchers. The *New Oxford Shakespeare* has recently argued that *Henry VI Part 1* is probably a co-authored play and the Shakespearian contribution is rather minor (Holdsworth 2017, 338–365; Taylor and Loughnane 2017, 416–518).

Let us map now the four sets' data points of this play.

Play for Testing Stage	Set One %	Set Two %	Set Three %	Set Four %	Set Five- Total of Four Sets %
<i>Henry VI Part 1</i>	2.3	7.2	1.4	2.7	13.6

Table 32: Five Sets' data points of *Henry VI Part 1*.

```

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a': 'trapmf', [1.64 1.7 1.8 1.86]
MF2='1b': 'trimf', [1.936 2.167 2.398]
MF3='1nota': 'trapmf', [0 0 1.58 1.65]
MF4='1notb': 'trapmf', [1.95 1.92 1.936 2.052]
MF5='1notc': 'trapmf', [2.282 2.398 100 100]

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=10
MF1='2a': 'trimf', [5.78 6.2 6.62]
MF2='2b': 'trimf', [6.27 7.05 7.83]
MF3='2c': 'trapmf', [8.23 8.3 8.4 8.47]
MF4='2d': 'trimf', [8.49 8.6 8.71]
MF5='2nota': 'trapmf', [0 0 5.36 5.79]
MF6='2notb': 'trimf', [7.82 8.16 8.24]
MF7='2notc': 'trapmf', [8.69 8.8 100 100]
MF8='2notd': 'trimf', [6 6.2 6.4]
MF9='2note': 'trimf', [6.6 7.05 7.5]
MF10='2notf': 'trimf', [7.7 7.765 7.83]

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=7
MF1='3a': 'trapmf', [1.33 1.4 1.5 1.57]
MF2='3b': 'trapmf', [1.63 1.7 1.8 1.87]
MF3='3c': 'trimf', [2.09 2.2 2.31]
MF4='3nota': 'trapmf', [0 0 1.26 1.34]
MF5='3notb': 'trimf', [1.57 1.6 1.64]
MF6='3notc': 'trapmf', [1.86 1.94 2 2.11]
MF7='3notd': 'trapmf', [2.29 2.4 100 100]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=4
MF1='4a': 'trapmf', [2.19 2.3 2.5 2.61]
MF2='4b': 'trapmf', [2.59 2.7 2.9 3.01]
MF3='4nota': 'trapmf', [0 0 2.1 2.21]
MF4='4notb': 'trapmf', [2.99 3.1 100 100]

```

Figure 44: Script for classes of Four Sets.

As can be deduced by the crosscheck of the preceding Table 32 and the above script (Figure 44), the data point of Set One (2.3) and Set Two (7.2%) of *Henry VI Part 1* is somewhat actual and somewhat ‘not/none’ actual. Therefore, the occurred combinations of the actual and ‘not/none’ classes for *Henry VI Part 1* are:

1b 2b 3a 4b
1b 2ne 3a 4b
1nb 2b 3a 4b
1nb 2ne 3a 4b

Note that since there are six ‘not/none’ subclasses (‘na’, ‘nb’, ‘nc’, ‘nd’, ‘ne’, ‘nf’), ‘ne’ stands for ‘not’ subclass ‘e’, which in fact is a ‘not/none’ area inside the actual class ‘2b’. This recalls the ‘notness’ of the area on the X-axis between the end of the first and end of the second SD.

Consequently, according to the principle of one-level-fall (4.2 and 7.3.13), the new added rules are:

- R16 IF 1b AND 2b AND 3a AND 4b THEN SiS1-High
- R17 IF 1b AND 2ne AND 3a AND 4b THEN SiS1-Low
- R18 IF 1nc AND 2b AND 3a AND 4b THEN SiS1-Medium2
- R19 IF 1nc AND 2ne AND 3a AND 4b THEN SiS1-Low

Let us extract also the index of cosine similarity of the average-ideal document-vector with the play *Henry VI Part 1*:

<i>Henry VI Part 1</i> (10,087 words) – Testing Stage	cosine similarity = 0.951 angle = 17.924 angle type = acute
---	---

By extracting all necessary data from *Edward II* and *Henry VI Part 1* a total of 19 rules have been formed preceding the building of the second layer of the fuzzy classifier of histories.

4.11.7 Building of the Second Layer of the Fuzzy Classifier.

The three new inputs of the second layer of the fuzzy system-classifier have been described in Section 4.11.4 and, similar to the comedies-based model, are the fifth set of words, SiS1 as input and the index of cosine similarity. The data points of the fifth set for the two histories are 14.8 for *Edward II* and 13.6 for *Henry VI Part 1*. They do not fall into any of the Shakespearian actual classes. On the other hand, the indices of cosine similarity of these two histories fall into the second, actual Shakespearian triangular class of the cosine variable (High-CosSim). In the previous subsection we saw that after mapping the two histories' data points three more rules were formed, one rule for *Edward II* and two for *Henry VI Part 1*. For the second layer of the fuzzy program, as before in the experimentation with comedies, new rules are needed in order to evaluate the three additional input variables and the produce the SiS2 score.

Due to fact that the data point of the fifth set of words for the two histories is non-Shakespearian for *Edward II* ('5notc') and *Henry VI Part 1* ('5notb') the new rules to be added, forming the current set of 23 rules, are:

R20 IF 5a AND SiS1-High AND High-CosS THEN SiS2-High (A general rule. See description in the experimentation with comedies in Sections 4.5 and 4.9.)

R21 IF 5nc AND SiS1-Low AND High-CosS THEN SiS2-Low (A general rule. See Section 4.5)

R22 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium2

R23 IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium1

Rules R22 and R23 form a common output area and are generated according to the algorithm by the data points of the three additional inputs of *Henry VI Part 1*. The general rule R21 coincidentally is identified with the data points of *Edward II* and consequently there is no need to proceed to the duplication of this combination. In general, the same principles apply as those defined with the comedies-based modelling in 4.5 and 4.8.

After the building of the second layer of the histories-based fuzzy classifier and the completion of the inference mechanism of 23 rules, in the next subsection follows the actual experimentation with *Edward II* and *Henry VI Part 1*. These two plays are employed only for testing since the histories that created the model are: *Henry IV Part 1*, *Henry IV Part 2*, *King John*, *Richard II*, *Richard III*. The SiS1 and SiS2-based interval will define the limits of the two plays' Shakespearianness.

4.11.8 Complete Fuzzy-System's Output: Validation Stage 2 and Testing Stage.

Let us now start the validation process (Validation Stage 2) of the model and the experimentation with the well-attributed, sole-authored history *Edward II* by Marlowe. As described before in the comedies-based experimentation, after entering to the input system the data of the four sets, the first of the two rightmost columns of the inference engine produces for *Edward II* as SiS1 the number 0.152 and the last column gives the SiS2, which for now is a default value of 0.5. Before the production of SiS1 and the addition of the counts of the three additional variables, the SiS2 score defaults 0.5 since for the time being there is no truncation of the classes of the output variable.

<i>Edward II</i>	Which rule fires	Output
0.152	R15	SiS1
0.5	(default value, no truncation yet)	SiS2

Table 33: SiS1 score of *Edward II*.

As before, I can now add three new entries with the data points of the fifth set (14.8), the SiS1 (0.152) and the index of cosine similarity (0.953) of *Edward II*. After entering the new data to the input's system, the SiS2 of 0.155 is produced and the rule that fires for SiS2 is R21.

<i>Edward II</i>	Which rule fires	Output
0.152	R15	SiS1
0.155	R21	SiS2

Table 34: SiS1 and SiS2 of *Edward II*.

Therefore, the two resulting SiS scores are 0.152 (SiS1) and 0.155 (SiS2) and so an evidential interval of 0.152 to 0.155 is produced. The limits of the degree of Shakespearianness of *Edward II* fall into the specific SiS1-Low continuous interval. These limits of the SiS1-Low interval (0.152-0.155) enhance the validity of the fuzzy classifier of Shakespearianness of histories, as *Edward II* is widely considered a well-attributed Marlovian play (Craig and Kinney 2009, 215; Boas 1960; Parks 1999) and there is no serious claim that Shakespeare contributed to it.

The next history with which I experiment is the disputed play *Henry VI Part 1*. By employing the histories-based fuzzy classifier, the SiS1 of *Henry VI Part 1* is produced and it is 0.544. For the time being, the SiS2 defaults to 0.5 and there is no truncation yet of any area of any of the output variable's class.

<i>Henry VI Part 1</i>	Which rule fires	Output
0.544	R16, R17, R18, R19	SiS1
0.5	(default value, no truncation yet)	SiS2

Table 35: SiS1 score of *Henry VI Part 1*.

By filling in for *Henry VI Part 1* the entries of the last three fields (the total of four sets= 13.6, SiS1= 0.544, cosine index = 0.951) of the input's system, the rule mechanism is re-activated (R22 and R23 fire) for the production of SiS2 score.

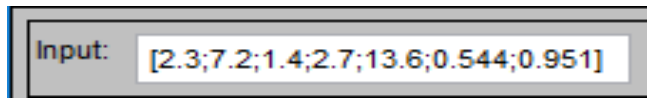


Figure 45: Data points of input variables of *Henry VI Part 1* for the production of SiS2.

The system produces for *Henry VI Part 1* the SiS2 score (0.496). Therefore, as the two SiS scores are 0.544 (SiS1) and 0.496 (SiS2), *Henry VI Part 1*'s Shakespearianness has an evidential interval of 0.496 to 0.544.

<i>Henry VI Part 1</i>	Which rule fires	Output
0.544	R16, R17, R18, R19	SiS1
0.496	R22, R23	SiS2

Table 36: SiS1 and SiS2 of *Henry VI Part 1*.

These limits (0.496, 0.544) fall into the overlapped area by the SiS1-Medium1 and SiS1-Medium2 classes. This result is close to the findings in primary experimentation, where the fuzzy holistic model produced as an authorship verdict (in favour of Shakespeare) the score 0.62. Of course, the differentiation of the core with the primary experimentation is the use of the second standard deviation in the former case for the Validation Stage 1 and the sophisticated 'not/none' areas that cover the distance between the end of the first and end of the second standard deviation that are defined as somewhat actual and somewhat 'not/none'. In fact, the SiS scores of *Henry VI Part 1* have validating power, too, since they enhance the validity of the fuzzy classifier of Shakespearianness of histories, because researchers think that Shakespeare wrote only a portion *Henry VI Part 1*. As noted in the literature review (1.2.1), this finding is supported by recent work on the topic (Holdsworth 2017, 338–365; Taylor and Loughnane 2017, 416–518),

4.11.9 Experimentation with Three More Disputed Histories (Testing Stage).

In order to proceed swiftly to the experimentation with three new disputed histories, the data of three plays--*Edward III*, *Henry VI Part 2*, and *Henry VI Part 3*--have been collated in order to add in one process any new rules to the 23 created so far.

History title	Set One %	Set Two %	Set Three %	Set Four %	Set Five-Total of Four Sets %
<i>Edward III</i>	2.1	6.4	1.8	2.7	13

Table 37: Data points of Sets of *Edward III*.

History title	Set One %	Set Two %	Set Three %	Set Four %	Set Five-Total of Four Sets %
<i>Henry VI Part 2</i>	1.8	6.8	1.8	2.3	12.7

Table 38: Data points of Sets of *Henry VI Part 2*.

History title	Set One %	Set Two %	Set Three %	Set Four %	Set Five-Total of Four Sets %
<i>Henry VI Part 3</i>	1.9	6.9	2.1	2.6	13.5

Table 39: Data points of Sets of *Henry VI Part 3*.

```

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a':'trapmf',[1.64 1.7 1.8 1.86]
MF2='1b':'trimf',[1.936 2.167 2.398]
MF3='1nota':'trapmf',[0 0 1.58 1.65]
MF4='1notb':'trapmf',[1.85 1.92 1.936 2.052]
MF5='1notc':'trapmf',[2.282 2.398 100 100]

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=10
MF1='2a':'trimf',[5.78 6.2 6.62]
MF2='2b':'trimf',[6.27 7.05 7.83]
MF3='2c':'trapmf',[8.23 8.3 8.4 8.47]
MF4='2d':'trimf',[8.49 8.6 8.71]
MF5='2nota':'trapmf',[0 0 5.36 5.79]
MF6='2notb':'trimf',[7.82 8.16 8.24]
MF7='2notc':'trapmf',[8.69 8.8 100 100]
MF8='2notd':'trimf',[6 6.2 6.4]
MF9='2note':'trimf',[6.6 7.05 7.5]
MF10='2notf':'trimf',[7.7 7.765 7.83]

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=7
MF1='3a':'trapmf',[1.33 1.4 1.5 1.57]
MF2='3b':'trapmf',[1.63 1.7 1.8 1.87]
MF3='3c':'trimf',[2.09 2.2 2.31]
MF4='3nota':'trapmf',[0 0 1.26 1.34]
MF5='3notb':'trimf',[1.57 1.6 1.64]
MF6='3notc':'trapmf',[1.86 1.94 2 2.11]
MF7='3notd':'trapmf',[2.29 2.4 100 100]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=4
MF1='4a':'trapmf',[2.19 2.3 2.5 2.61]
MF2='4b':'trapmf',[2.59 2.7 2.9 3.01]
MF3='4nota':'trapmf',[0 0 2.1 2.21]
MF4='4notb':'trapmf',[2.99 3.1 100 100]

```

Figure 46: Script for the classes of Four Sets.

As can be deduced by the crosscheck of the three plays' preceding tables and the above script (Figure 46), the data points of Set Two for the two plays (*Henry VI Part 2*, 6.8%, *Henry VI Part 3*, 6.9%) are somewhat actual and somewhat 'not/none' actual, whereas *Edward III* marginally avoids the membership in the 'not/none' actual class and it is only of actual class. By checking the data points with each of the classes of the script, the occurred combinations of the actual and 'not/none' classes for each play are:

Edward III

1b 2b 3b 4b

Henry VI Part 2

1a 2b 3b 4a

1a 2ne 3b 4a

Henry VI Part 3

1nb 2b 3c 4a

1nb 2ne 3c 4a

1nb 2b 3nc 4a

1nb 2ne 3nc 4a

1nb 2b 3c 4b

1nb 2ne 3c 4b

1nb 2b 3nc 4b

1nb 2ne 3nc 4b

I turn these antecedents into rules by adding the appropriate consequences (SiS1-Low, SiS1-Medium1, SiS1-Medium2, SiS1-High) and applying the algorithm of the special case of Set Two and one-level-fall from the SiS1-High for the negated sets, so that the following new rules are formed:

Edward III

R24 IF 1b 2b 3b 4b THEN SiS1-High

Henry VI Part 2

R25 IF 1a 2b 3b 4a THEN SiS1-High

R26 IF 1a 2ne 3b 4a THEN SiS1-Low

Henry VI Part 3

R27 IF 1nb 2b 3c 4a THEN SiS1-Medium2

R28 IF 1nb 2ne 3c 4a THEN SiS1-Low

R29 IF 1nb 2b 3nc 4a THEN SiS1-Medium1

R30 IF 1nb 2ne 3nc 4a THEN SiS1-Low

R31 IF 1nb 2b 3c 4b THEN SiS1-Medium2

R32 IF 1nb 2ne 3c 4b THEN SiS1-Low

R33 IF 1nb 2b 3nc 4b THEN SiS1-Medium1

R34 IF 1nb 2ne 3nc 4b THEN SiS1-Low

The fuzziest data points are those of *Henry VI Part 3* and for a single play it is necessary to generate seven more rules (R27-R34). The data points of *Edward III* constitute a new SiS1-High combination, and likewise one combination of *Henry VI Part 2* increases further the number of the SiS1-High rules to nine. At the same time the data point of Set Two of *Henry VI Part 3*, apart from being Shakespearian, falls into one of the internal ‘not/none’ triangles and therefore it generates an SiS1-Low rule due to the gravity of Set Two that is treated as a special case of major importance. Below are printed out the tables for the SiS1 scores (and the default 0.5 values of SiS2) for the three plays *Edward III*, *Henry VI Part 2*, and *Henry VI Part 3*:

<i>Edward III</i>	Which rule fires	Output
0.854	R24	SiS1
0.5	(default value, no truncation yet)	SiS2

Table 40: SiS1 and SiS2 results of *Edward III*.

<i>Henry VI Part 2</i>	Which rule fires	Output
0.604	R25, R26	SiS1
0.5	(default value, no truncation yet)	SiS2

Table 41: SiS1 and SiS2 results of *Henry VI Part 2*.

<i>Henry VI Part 3</i>	Which rule fires	Output
0.381	R27, R28, R29, R30, R31, R32, R33, R34	SiS1
0.5	(default value, no truncation yet)	SiS2

Table 42: SiS1 and SiS2 results of *Henry VI Part 3*.

After the production of the SiS1 scores of the three histories, it is necessary to proceed to the addition of the counts of the three additional variables.

Let us first recall the membership functions-classes of the fifth set by looking at Figure 47:

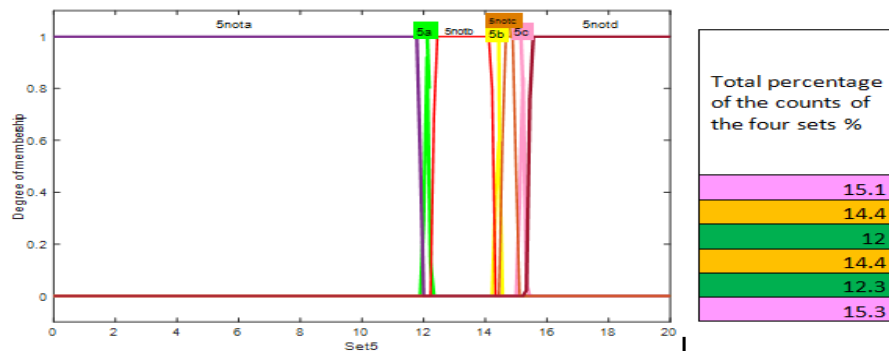


Figure 47: Membership functions of Set Five.

The data points of the fifth set for the three histories (*Edward III*, 13, *Henry VI Part 2*, 12.7, *Henry VI Part 3*, 13.5) fall into the Shakespearian class of ‘5nb’ (see the third class, the red trapezoid after the green triangle of ‘5a’). As we saw above, the three plays’ SiS1 scores (that will be used as inputs) are 0.854, 0.604 and 0.381. It is now necessary to produce the indices of cosine similarity of the three histories under scrutiny in comparison with the average-ideal vector-document. In the table below (Table 43), particular attention should be paid to the first play (*Edward III*) bearing in mind the limits of the class of high cosine similarity [0.884-1].

History title	Cosine index of similarity with the average-ideal history
<i>Edward III</i>	cosine similarity = 0.867 angle = 29.859 angle type = acute
<i>Henry VI Part 2</i>	cosine similarity = 0.950 angle = 18.234 angle type = acute
<i>Henry VI Part 3</i>	cosine similarity = 0.934 angle = 20.979 angle type = acute

Table 43: Cosine indices of *Edward III*, *Henry VI Part 2* and *Henry VI Part 3* in comparison with the average-ideal document.

In the first layer of the histories-based fuzzy program the data points of the four semantic sets led to a score that is situated in the SiS1-High class of the output variable. By adding in the second layer the assessment of three additional variables we see the importance of the index of the cosine similarity. It assists in the detection of large deviation/dissimilarity of the counts of words' frequencies of a new play in comparison with the average counts of words' frequencies as they have been found in the six histories. Recall the case where the sets' counts of a new play all have a membership in the actual class but their sets' counts of words' frequencies are deviating considerably from the average words' counts, as they are detected in the plays that were used to create the histories-based model (see also Section 7.3.9 of the Technical Appendix). Indeed, *Edward III* has fully Shakespearian counts of the four semantic sets but the individual words' counts (Low-CosSim) of these four sets weaken this impression by producing in the second layer of the fuzzy program a much smaller SiS2 than SiS1, as will be seen with the further experimentation.

In the graph below (Figure 48) are displayed the counts of 100 individual words' frequencies detected in *Edward III* in comparison with the average counts of the six well-attributed Shakespearian histories. The Y-axis expresses the counts of the words' frequencies and the X-axis the four sets' individual words, which are numbered from 1 to 100. You can view that between the two documents represented as two lines there is significant

dissimilarity for, in general, frequently used words (see x -values around words number 20-25 and 50-55, which, in fact, are mainly personal and relative pronouns).

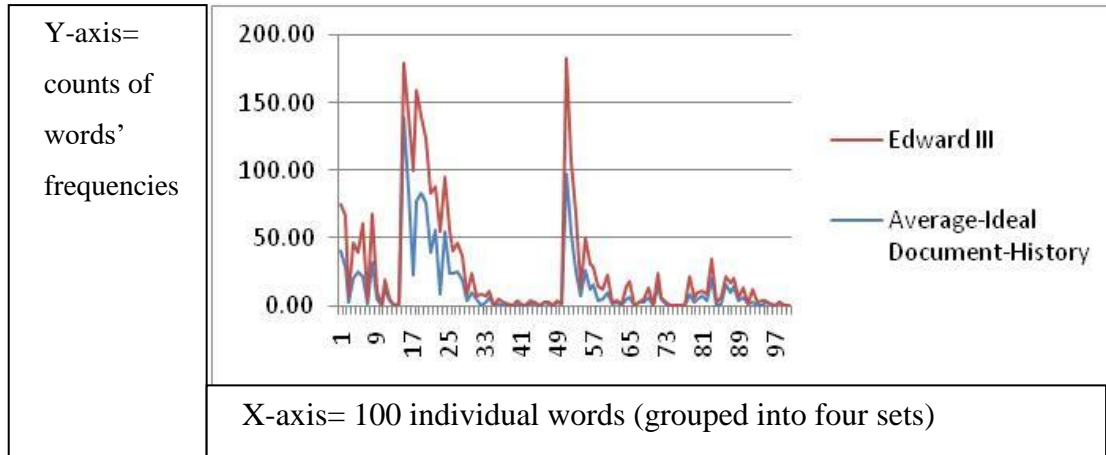


Figure 48: Counts of words' frequencies of *Edward III* and the average-ideal document.

Recalling the algorithm of the formation of new rules (Section 4.5), we can remind ourselves of the eventuality of the detection of a low cosine index (Low-CosSim) for a new play, which brings as a consequent the generation, and then activation of a single SiS1-Low rule (Section 4.5). But let us describe for the three plays the actual process of new rules' formation and experimentation of the second layer. Based on the indices of cosine similarity for the three plays (Table 43), the fifth Set's data points and the initially produced SiS1 scores, the following new combinations are produced:

IF 5nb AND SiS1-High AND Low-CosS THEN SiS2-Low

IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium2 (This combination rule has been already generated by R22 and so there is no need to duplicate it.)

IF 5nb AND SiS1-Medium2 AND High-CosS THEN SiS2-Medium1 (This combination rule has been already generated by R23 and so there is no need to duplicate it.)

IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Medium1

IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Low

So, only three rules are added to the current set of (34) rules.

R35 IF 5nb AND SiS1-High AND Low-CosS THEN SiS2-Low

(R35: Recall the description of the algorithm for SiS2 in Section 4.5. Since the cosine index of *Edward III* is low, only one rule is formed and the output is SiS2-Low irrespective of the data point of the fifth set, which in any case is also ‘not/none’ actual.)

R36 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Medium1

R37 IF 5nb AND SiS1-Medium1 AND High-CosS THEN SiS2-Low

The complete inference engine of the fuzzy classifier is composed now of a total of 37 rules. The system of the histories-based classifier is even more economical than the first comedies-based classifier, which needed 41 rules.

As we saw in the beginning of this subsection, the SiS1 scores of the three plays (*Edward III*, 0.854, *Henry VI Part 2*, 0.604, *Henry VI Part 3*, 0.381) have already been produced. Let us now start the experimentation with these three disputed plays in order to produce the SiS2 scores, and consequently to form the interval of belief in the degree of Shakesperianness. I input to the system the data points of the three additional variables in the second stage and the simulation, classification is initiated. The three tables below show the results and the rules that fired for the production of these two SiS scores:

<i>Edward III</i>	Which rule fires	Output
0.854	R24	SiS1
0.153	R35	SiS2

Table 44: SiS1 and SiS2 results of *Edward III*.

<i>Henry VI Part 2</i>	Which rule fires	Output
0.604	R25, R26	SiS1
0.508	R22, R23	SiS2

Table 45: SiS1 and SiS2 results of *Henry VI Part 2*.

<i>Henry VI Part 3</i>	Which rule fires	Output
0.381	R27, R28, R29, R30, R31, R32, R33, R34	SiS1
0.275	R36, R37	SiS2

Table 46: SiS1 and SiS2 results of *Henry VI Part 3*.

Particular attention should be paid to the first of the three plays, *Edward III*, where the limits (0.153-0.854) of the SiS interval equal much more than 0.25 due to the fact that the data points of the four Sets are Shakespearian but the index of cosine similarity is of low Shakespearianness (the class of Low-CosSim).

With the experimentation of these three additional disputed plays I managed to build the second genre-based Fuzzy Stylistic Classifier, for the genre of history plays, with a total of 37 rules. Some of these rules play off one another to reduce the uncertainty by assessing whether and how much the plays' data points are somewhat actual and somewhat 'not/none' actual.

4.11.10 Validity of Performance and Testing Mechanism.

Assessing the SiS intervals for all five plays under scrutiny--*Edward II*, *Henry VI Part 1*, *Edward III*, *Henry VI Part 2*, and *Henry VI Part 3*--we find that the results are consistent with the evidence of historical scholarship and the orientation of recent authorship attribution assertions. The history *Edward II* has been rightly judged by the fuzzy classifier as a non-Shakespearian play since its SiS interval is 0.152-0.155 (SiS1-Low). The SiS interval for *Henry VI Part 1* in the core experimentation almost iterates the authorship verdict (0.62) of the primary experimentation (which produces only the SiS1 score) since the formed SiS interval is 0.496 to 0.544. This is a result that verifies the fact that Shakespeare has written only a portion of the play. This view of the play's authorship is discussed in the beginning of the literature review in 1.2.1 where the claims of Malone are analysed, and the New Oxford Shakespeare's findings lead to the same conclusion (Holdsworth 2017, 338–365; Taylor and Loughnane 2017, 416–518).

It is worth mentioning that the SiS intervals-based conclusions drawn from the experimentation with the histories-based classifier regarding *Henry VI Part 3*, *Edward III* and *The Jew of Malta* are similar to the results produced in the article by Matthews and Merriam (1994) that was discussed in the Literature Review in section 1.2.10.1. Matthews and Merriam constructed a neural network in order to discriminate between the plays of Shakespeare and Marlowe (Matthews and Merriam 1994). According to the classification made by the neural network, the play *Henry VI Part 3* seems to be a revised version of the Marlovian original titled *The True Tragedy*. This evidence does not deviate from the SiS interval (= 0.275-0.381, an interval that falls into SiS1-Medium1 class) that the histories-based Fuzzy Stylistic Classifier has produced for *Henry VI Part 3*. A second conclusion of

the same article (Matthews and Merriam 1994) is that in the disputed play *Edward III* there are stylistic features which are similar to the Shakespearian style, a verdict that is corroborated by the four Sets' based SiS1 verdict ($=0.854$) of the experimentation in this section, though this classification is set under further examination due to the lack of a compact output of SiS interval (*Edward III*'s SiS interval= $0.153-0.854$). As far as it concerns *The Jew of Malta*, the results of Matthews and Merriam's experimentation are congruent with Burrows and Craig's claim that 'the test can be deceived by a play like *The Jew of Malta*, which evidently departs from the Marlowe style' (Burrows and Craig 2017, 212). The same judgments for the *Jew of Malta* seem to apply in the primary and histories-based experimentation, since the results are not identified neither with the SiS-High nor the SiS-Low class (SiS1= 0.35 in primary experimentation and SiS interval of core experimentation= $0.152-0.601$).

In addition, *Henry VI Part 2* is judged much more Shakespearian than *Henry VI Part 3* since the SiS interval for the former play is $0.508-0.604$ (SiS1-Medium2), whereas for the latter it is $0.275-0.381$ (SiS1-Medium1). The SiS interval for *Edward III* is $0.153-0.854$. *Edward III* is generally considered to be a co-authored play and some researchers argue that Shakespeare has contributed to its writing (Taylor and Loughnane 2017, 598). As I already showed, the data points of the four semantic sets of *Edward III* are all Shakespearian. On the other hand though, this is the play with the lowest tokens' indicator-index of cosine similarity. *Edward III*'s non-set based stylistics features are less Shakespearian from any other play examined in this thesis. This paradox and the detection of an especially wide SiS interval of $0.153-0.854$, indicating uncertainty, merit further discussion. In intervals that are more than 0.25 but less than approximately 0.5 there is a way to construct more robust interval classes, as shown in Section 7.3.16. The prevalence of the low cosine index (0.867) and the SiS2 (0.153) against the SiS1 score of 0.854 for *Edward III* can be explained by the counts of words' frequencies for the approximately 10,000 words of the histories, as displayed in the table below:

History title	you	me	her	thou	this	he	thee	him	she	we	they	them	us	these	none	those	himself	herself	themselves	nothing
Henry IV Part 1	143	84	14	109	80	89	47	52	4	56	41	50	30	16	5	12	4	0	3	5
Henry IV Part 2	211	116	20	82	57	110	38	71	7	33	22	16	12	17	3	6	8	0	4	7
Henry V	147	43	14	19	62	74	21	39	11	85	31	26	48	14	2	6	5	2	2	5
King John	81	57	45	112	150	48	47	42	18	64	12	18	21	27	4	12	6	0	1	6
Richard II	71	72	8	64	70	54	33	62	0	53	21	11	24	19	5	10	6	1	1	12
Richard III	182	131	38	81	78	84	47	73	24	37	19	19	17	21	6	9	3	1	2	3
Edward III	40	56	77	81	57	47	44	31	46	41	32	17	21	18	4	15	2	8	5	5

Table 47: Counts of words' frequencies of six well-attributed Shakespearian histories (Rows 1-6) and *Edward III* (last row).

The six initial rows of the above table contain the counts of the 20 words' frequencies (20 columns) that compose Set Two. Most of these words are pronouns and, in general, they are considered as words that carry rich stylistic information. By focusing on the white cells of the seventh row that contain the counts of words of Set Two of *Edward III*, it is possible to justify the low cosine index and to assess if the final verdict of SiS2 (0.153) is more reasonable than SiS1 (0.854). The above table suggests various possibilities. One is that apart from the cosine index and SiS2, there is also strong partial evidence about the existence in *Edward III* of stylistic features that clearly are not Shakespearian (see *Edward III*'s only 40 counts of 'you', 77 counts of 'her', 46 counts of 'she', and eight counts of 'herself'). Another is that there exist a few stylistic similarities with the Shakespearian style and these are supported by the total counts of Set Two (see also individual counts, such as of 'me', 'thou' and 'this'. Combining these possibilities, with the assessment of sets' counts and cosine index, there seems to be an 'invisible hand' that distorts the Shakespearianness of approximately the first 10,000 words of the play *Edward III*. In other words, it is claimed by the present investigator that these 10,000 words of *Edward III* contain sets-based stylistic similarities and some words' counts strong dissimilarities. Therefore, as a whole *Edward III* might not be considered a sole-authored Shakespearian history. Equally, it can be viewed as a special case because of the nature of Shakespeare's involvement in certain scenes of a particular kind: those concerning the attempted wooing of the Countess of Salisbury. These

scenes make heavier use of feminine pronouns than is usual in a history play (Severdia 2019). The output SiS interval (0.153-0.854) of *Edward III* implies that further investigation is needed--for instance, by exploring styles of other candidate authors (such as George Peele) and smaller textual parts--in order to detect if individual styles can be located in certain parts of *Edward III* or if there is a justification of the distorted but actual Shakespearianess that is indicated by the counts of specific words.

Similar to the comedies-based model, the results of the experimentation with the four histories other than *Edward III* can be represented by a scatter plot where the lower SiS bound of each play is represented by the *X*-axis and the upper SiS bound by the *Y*-axis. This scatter plot shows that *Henry VI Part 1* and *Henry VI Part 2* ‘tend’ to be Shakespearian. *Henry VI Part 1* has an SiS marginally less than 0.5 and a second SiS somewhat more than 0.5; *Henry VI Part 2* has both SiS scores somewhat more than 0.5 in both axes; *Henry VI Part 3* ‘tends’ to be non-Shakespearian; and *Edward II* is almost certainly non-Shakespearian. For the linguistic categorisation of tendencies-evidential degrees of Shakespearianess see Section 7.3.15 of the Technical Appendix.

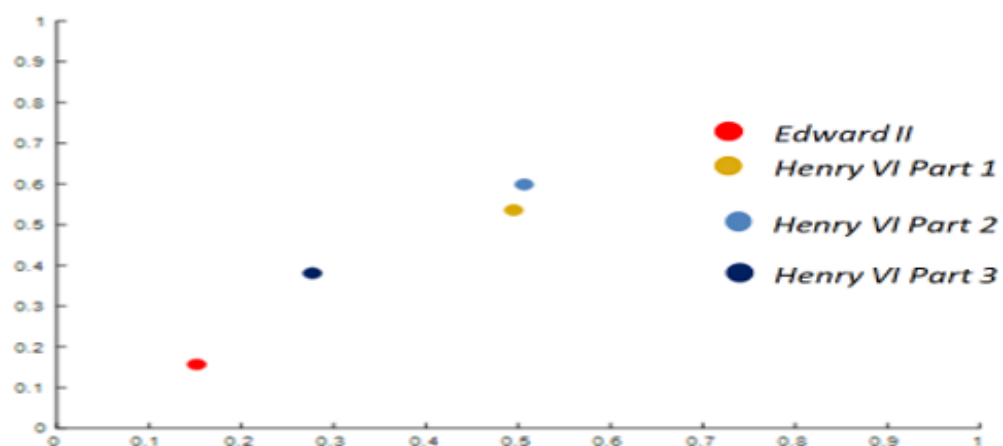


Figure 49: The scatter plot of the SiS interval of the data points of four histories under scrutiny.

In conclusion, the SiS intervals of *Henry VI Part 1* and *Henry VI Part 2* are situated in the SiS1-Medium2 output class, just one below the SiS1-High class, whereas that of *Henry VI Part 3* falls into the SiS1-Medium1 output class and is judged less Shakespearian than the first two parts of *Henry VI*. The current results of the SiS intervals corroborate Malone’s

claims (Malone 1787) and recent findings of the New Oxford Shakespeare (Holdsworth 2017, 338–365; Taylor and Loughnane 2017, 416–518). In addition, our investigation highlights the need of investigating further *Edward III* based on additional stylistic markers.

For the sake of economy a condensed version of the tragedies-based model will be provided in the next section. The logic and methodologies are identical with those I employed in the previous comedies- and histories-based stylistic classifiers.

4.12 Tragedies-Based Fuzzy-Stylistic Classifier.

This subsection contains the modelling and the experimentation with the nine well-attributed, sole-authored tragedies of Shakespeare. The tragedies-based classifier is based on the principles of the methodology that has been applied for the building of the comedies- and histories-based classifiers. As we are familiar with the methods of fuzzification and defuzzification, the description of the steps will be condensed.

4.12.1 The Tragedies-Based Fuzzy Classifier of Shakespearianness: Design of Membership Functions-Classes.

This model is based on these nine well-attributed, sole-authored tragedies of Shakespeare:

Antony and Cleopatra

Coriolanus

Cymbeline

Hamlet

Julius Caesar

King Lear

Othello

Romeo and Juliet

Troilus and Cressida

At this point it should be mentioned that in computational stylometry it is necessary to categorise plays in terms of genre based on the stronger available evidence or the predominant verdict. Of course, in certain Shakespearian plays, various mixed stylistic tendencies can be detected because the author's prevailing intention may not be. Consequently, it is not always easy to define the nature-genre of a play and its distinct categorisation. Frederick S. Boas (1862-1967) expressing a personal opinion (but based on valid premises) pointed out that *Troilus and Cressida* is a 'problem-play' (a category in which he also put *All's Well that Ends Well*, *Measure for Measure*, and *Hamlet*) and it is not easy to discern or quantify the satirical, comical, tragic or even epic poetic features since a part of the play is based on the Homeric epic of the *Iliad* (which was written probably around the eighth century BC). This notion of 'problem plays', particularly for *Troilus and Cressida*,

Measure for Measure, and *Hamlet*, has been further discussed and it has been theorised until nowadays as a diachronic problem (Schanzer 1963; Fowler 1985; Danson 2000; Marsh 2002) though critique has also been expressed about the fact that the categorisation of plays into a genre was mainly a problem that arose in the past due the intention of forming rigid, immutable rules of categorisation (Danson 2000, 4). There are also some deviations from Boas's definitional context, as for instance Ernest Schanzer relates the problematic nature of the 'problem plays', and particularly of *Measure for Measure*, to 'the moral bearings' of the audience towards the role of the protagonists (Schanzer 1963, 5-10). Similar to these conceptions about 'problem plays' might be the claim that the inclusion of the latest (1623 Folio) of the editions of *Hamlet* in the corpus of the well-attributed Shakespearian plays entails a risk of losing parts of the original style. Indeed, there are so many versions of *Hamlet*--Q1 1603, Q2 1604, Q3 1605, Q4 1605, Q5 1611 (Dowden 1899, ix)--and with so many differences between them that someone would doubt that the version of F1 is the ideal one for analysing the Shakespearian style.

The Folio version of *Hamlet* was selected because all the other plays of the corpus in this experimentation are Folio versions and consistency is needed when forming a corpus. Because the 1623 Folio was a single printing project occurring at a particular historical moment (1622-23) and executed by a relatively small number of workers in a single printshop, it offers the best hope of minimizing the happenstance variability of production to which early printed texts are necessarily subject. Analysing small extracts of *Hamlet* could be even more problematic. In the analysis of the first approximately 10,000 words of that play in the Validation Stage 1 the classifier produced for *Hamlet* the second lowest SiS1 score, thus 0.479 (see section 4.3). The implication here is that 'problem plays' not only in the sense of Boas's arguments and but also from the scope of the most appropriate version (when many alternative and revised versions exist) can be detected with the Fuzzy-Logic-based stylistic classifiers. Experimenting with the other versions could give us-in combination with historical scholarship-some evidence about which version seems to approach the style of Shakespeare as viewed in the other comedies. But this is a complex problem and exceeds the purposes of this thesis.

Such problematic cases can be resolved with the building of a holistic fuzzy classifier (and perhaps the auxiliary use of a strict genre-based classifier) as has been carried out in the primary experimentation or, perhaps, by examining such a play in all available versions or genres it relates due to such play's mixed nature. On the other hand, a holistic model lacks the genre-based stylistic specialisation. But this subtle issue of the

‘problem plays’ (Boas 1910, 344–408) and their debatable genre exceeds the goal of this thesis and so I proceed to more general categorisation, such as that *Troilus and Cressida* is a tragedy. The same concern may be raised about *Cymbeline*, which appears in Tragedies section of the 1623 Folio but ends essentially happily for its main protagonists.

Let us now look at the next two tables showing the spread of the sets’ data and the formed membership functions-classes of the nine tragedies:

Tragedies	Set One %	Set Two %	Set Three %	Set Four %	Total of the four sets/ Set Five %
<i>Antony and Cleopatra</i>	2.4	8.8	2.2	3.1	16.5
<i>Corolianus</i>	1.9	9.2	1.3	3	15.4
<i>Cymbeline</i>	2.2	8.6	1.5	2.6	14.9
<i>Hamlet</i>	2.2	7.5	1.8	2.5	14
<i>Julius Caesar</i>	2.4	8.4	1.8	2.9	15.5
<i>King Lear</i>	2.1	8.8	1.5	2.9	15.3
<i>Othello</i>	2.1	8.2	1.6	2.5	14.4
<i>Romeo and Juliet</i>	2.6	7.9	2.4	2.8	15.7
<i>Troilus and Cressida</i>	2.3	7.7	1.9	2.7	14.6

Table 48: Data points of the five sets of the nine Shakespearian tragedies.

Tragedies	Set One	Set Two	Set Three	Set Four	Total of four sets/Set Five
<i>Antony and Cleopatra</i>	1b	2	3d	4	5c
<i>Corolianus</i>	1a	2	3a	4	5b
<i>Cymbeline</i>	1b	2	3b	4	5a

<i>Hamlet</i>	1b	2	4c	4	5a
<i>Julius Caesar</i>	1b	2	4c	4	5b
<i>King Lear</i>	1b	2	3b	4	5b
<i>Othello</i>	1b	2	3b	4	5a
<i>Romeo and Juliet</i>	1c	2	3d	4	5c
<i>Troilus and Cressida</i>	1b	2	4c	4	5a

Table 49: Data points of nine well-attributed Shakespearian tragedies and their contribution to actual classes. Similarly coloured cells represent the distinct classes.

As you can see in Table 49, data points in cells that have the same colour are members of the same class. The Shakespearian tragedies' actual data points have activated a variety of design constraints described early in the preliminary discussion of the design method and clustering principles (see page 100). Because in many cases the sets' data points form non-continuous intervals and that the data points are varied--being above our 0.1 threshold for proximity but below 0.5--and hence because the number of distinct duplets or triplets, and so trapezoidal classes, is limited, the Shakespearian style of tragedies is less distinguishable than the two other genres. That is, there are not many distinct patterns of sets' counts but mainly a general representation of the central stylistic tendency by triangular classes.

As was mentioned previously in the design of the membership functions of the comedies (4.1) and histories (4.11.1), the relative standard deviation (RSD) measures how far the SD is from the mean in terms of percentage of the mean value.

Below are displayed the descriptive statistics of the five sets of the three genres:

Statistics of Comedies-based model	Set One	Set Two	Set Three	Set Four	Set Five
Mean	2.33	8.74	1.70	2.46	15.24
STDEV	0.21	0.97	0.20	0.27	1.07
RSD %	9.01	11.15	11.76	11.17	7.05
Statistics of Tragedies-based model	Set One	Set Two	Set Three	Set Four	Set Five
MEAN	2.24	8.34	1.78	2.78	15.14
STDEV	0.21	0.57	0.35	0.22	0.76
RSD	9.22	6.78	19.84	7.80	5.00
Statistics of Histories-based model	Set One	Set Two	Set Three	Set Four	Set Five
Mean	1.95	7.55	1.80	2.60	13.92
STDEV	0.25	1.11	0.34	0.24	1.42
RSD %	12.87	14.77	18.92	9.10	10.20

Table 50: Mean, Standard Deviation and RSDs of the five sets in the three genres.

Interestingly, in the tragedies-based classifier, the RSDs of three (S2, S3, S4) of the five sets' data points, as can be viewed in Figure 50, form almost a (Gaussian-like, normal distribution) curve membership function.

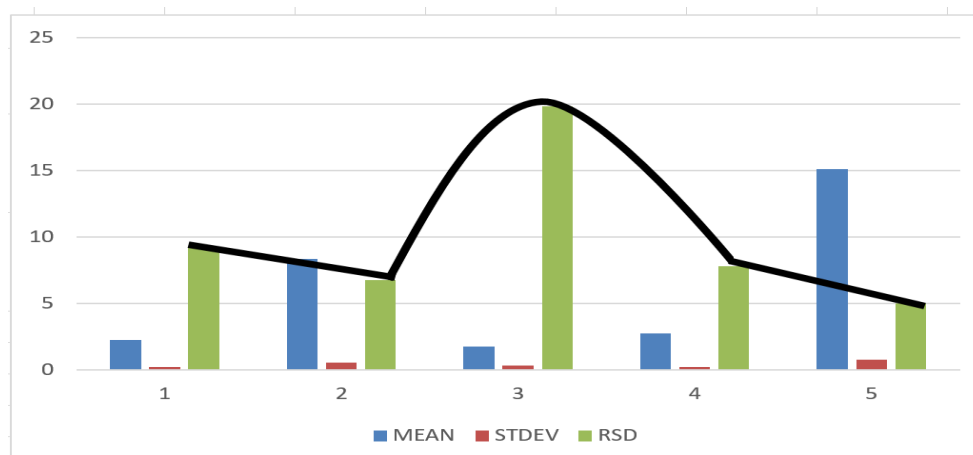


Figure 50: Y-axis expresses Mean and StDev in numbers. RSDs on Y-axis are measured in percentages. The numbers on X-axis correspond to the five sets.

4.12.2 Inference Mechanism of the Tragedies-Based Fuzzy Classifier of Shakespearianess.

Looking at the number of the nine tragedies' actual shapes of each of four sets there are three possibilities for Set One times one for Set Two times the four of Set Three times one for Set Four, giving 12 possibilities. By contrast, I found 16 possibilities for the 12 comedies-based model and 48 for the histories-based model). But again, as constantly through the whole thesis, I ignore the theoretically possible combinations because they were not combinations 'selected' by the nine tragedies that were employed for the design of the membership functions-classes (see complete explanation in Section 7.2.2). The following list of the tragedies' combinations of four sets' actual classes shows in colours the combinations that are found in more than one play:

<i>Antony and Cleopatra</i>	1b	2	3d	4
<i>Coriolanus</i>	1a	2	3a	4
<i>Cymbeline</i>	1b	2	3b	4
<i>Hamlet</i>	1b	2	3c	4
<i>Julius Caesar</i>	1b	2	3c	4
<i>King Lear</i>	1b	2	3b	4
<i>Othello</i>	1b	2	3b	4
<i>Romeo and Juliet</i>	1c	2	3d	4
<i>Troilus and Cressida</i>	1b	2	3c	4

Table 51: Combinations that are found in more than one tragedy.

So, for the time being, only the following SiS1-High rules are derived from the sets-based stylistic patterns of nine tragedies:

1b	2	3d	4
1a	2	3a	4
1b	2	3b	4
1b	2	3c	4
1c	2	3d	4

I turn these five combinations into antecedents of five rules and add the proper consequents.

R1 IF 1b AND 2 AND 3d AND 4 THEN SiS1-High
 R2 IF 1a AND 2 AND 3a AND 4 THEN SiS1-High
 R3 IF 1b AND 2 AND 3b AND 4 THEN SiS1-High
 R4 IF 1b AND 2 AND 3c AND 4 THEN SiS1-High
 R5 IF 1c AND 2 AND 3d AND 4 THEN SiS1-High

In the next stage of generating new negated rules I follow the same search and harvesting process adopted in the primary experimentation (Section 7.2.2) and taking into account the first (or single) actual classes of the sets-inputs, the six resulting combinations of the rule's antecedents with 'not/none' ('n') classes are:

1na 2 3a 4
 1a 2na 3a 4
 1na 2na 3a 4
 1na 2 3na 4
 1na 2na 3na 4
 1na 2na 3a 4na

Then, as usual with the principle of one-level-fall (7.3.13) from the SiS1-High, I turn these antecedents into rules by adding the appropriate consequents (SiS1-Low, SiS1-Medium1, SiS1-Medium2, SiS1-High), and so six new rules are formed (R6 to R11):

IF 1na AND 2 AND 3a AND 4 THEN SiS1-Medium2
 IF 1a AND 2na AND 3a AND 4 THEN SiS1-Low
 IF 1na AND 2na AND 3a AND 4 THEN SiS1-Low
 IF 1na AND 2 AND 3na AND 4 THEN SiS1-Medium1
 IF 1na AND 2na AND 3na AND 4 THEN SiS1-Low
 IF 1na AND 2na AND 3a AND 4n THEN SiS1-Low

After this, there are in total 11 rules (5 SiS1-High and 6 ‘not/none’ rules):

R1 IF 1b AND 2 AND 3d AND 4 THEN SiS1-High
 R2 IF 1a AND 2 AND 3a AND 4 THEN SiS1-High
 R3 IF 1b AND 2 AND 3b AND 4 THEN SiS1-High
 R4 IF 1b AND 2 AND 3c AND 4 THEN SiS1-High
 R5 IF 1c AND 2 AND 3d AND 4 THEN SiS1-High
 +
 R6 IF 1na AND 2 AND 3a AND 4 THEN SiS1-Medium2
 R7 IF 1a AND 2na AND 3a AND 4 THEN SiS1-Low
 R8 IF 1na AND 2na AND 3a AND 4 THEN SiS1-Low
 R9 IF 1na AND 2 AND 3na AND 4 THEN SiS1-Medium1
 R10 IF 1na AND 2na AND 3na AND 4 THEN SiS1-Low
 R11 IF 1na AND 2na AND 3a AND 4na THEN SiS1-Low

The output variable of the histories-based Fuzzy Stylistic Classifier is designed in the same way as in the primary experimentation (Section 7.2.3) and comedies- and histories-based models, and it is formed by the four trapezoidal membership functions of the classes of the Low, Medium1, Medium2 and High index of Shakespearian Similarity (SIS).

4.12.3 Validation Stage 1: Validation of Inference Mechanism.

Before proceeding to the production of the SiS1 for the nine Shakespearian tragedies, I should consider the tragedies one by one, as I have done for the comedies (Section 4.3) and histories (Section 4.11.3), and see if any of their data points are also somewhat ‘not/none’. If they are, then I have to proceed to the mapping of these points and update our rules’ mechanism--currently with 11 rules--and add new rules for each tragedy, so that the final

result takes into account the centroid of more than one output variable's class apart that of SiS1-High.

Antony and Cleopatra

R12 IF 1b AND 2 AND 3d AND 4nb THEN SiS1-Medium2

R13 IF 1nc AND 2 AND 3d AND 4nb THEN SiS1-Medium1

R14 IF 1nc AND 2 AND 3d AND 4 THEN SiS1-Medium2

Coriolanus

R15 IF 1a AND 2nb AND 3a AND 4nb THEN SiS1-Low

R16 IF 1a AND 2nb AND 3a AND 4 THEN SiS1-Low

R17 IF 1a AND 2 AND 3a AND 4nb THEN SiS1-Medium2

Julius Caesar R18 IF 1nc AND 2 AND 3c AND 4 THEN SiS1-Medium2

King Lear R19 IF 1nb AND 2 AND 3b AND 4 THEN SiS1-Medium2

Othello

R20 IF 1nb AND 2 AND 3b AND 4na THEN SiS1-Medium1

R21 IF 1b AND 2 AND 3b AND 4na THEN SiS1-Medium2

IF 1nb AND 2 AND 3b AND 4 THEN SiS1-Medium2 (No need to duplicate-same as combination of R19)

Troilus and Cressida R22 IF 1b AND 2na AND 3c AND 4 THEN SiS1-Low (This combination maps also a case for *Hamlet*.)

After having updated the rules' mechanism with 11 new rules (to make 22 rules in all), let us now see how the tragedies-based fuzzy program with the current inference mechanism of 22 rules classifies the nine well-attributed Shakespearian histories, whose four sets' data points contributed to the modelling parameters of the first layer of the fuzzy classifier (SiS1) of tragedies.

Tragedy title	Which Rule Fires	SiS1
<i>Antony and Cleopatra</i>	1, 12, 13,14	0.598
<i>Corolianus</i>	2, 15, 16, 17	<u>0.437</u>
<i>Cymbeline</i>	3	0.872
<i>Hamlet</i>	4, 22	<u>0.479</u>
<i>Julius Caesar</i>	4, 18	0.714
<i>King Lear</i>	3, 19	0.820
<i>Othello</i>	3, 19, 20, 21,	0.635
<i>Romeo and Juliet</i>	5	0.859
<i>Troilus and Cressida</i>	4, 22	0.704

Table 52: Validation SiS1 scores of the nine Shakespearian tragedies.

As can be observed by the above Table 52, six out of nine SiS scores for the nine Shakespearian tragedies are indicative of high Shakespearianness. The only plays that its SiS1 score falls into the SiS1-Medium2 classes are the tragedies of *Hamlet* and *Corolianus* whose SiS1 scores are respectively 0.479 and 0.437. (In fact, *Corolianus* falls marginally into the SiS1-Medium1 area, since SiS1-Medium2 class starts at the point that gives on X-axis the value of 0.437). The result (SiS1=0.479) of *Hamlet* might have a connection to the arguments of Boas about the problematic single genre categorisation of this play, whereas *Corolianus*'s medium score (0.437) might have been derived due to the fact that this play is one of the very late tragedies, if not the last written tragedy of Shakespeare, and so it can be argued that the set-based deviation of style is due to the stylochronometric changes (see relevant discussion about this term in Section 1.3.1.1) that occur during the life of a writer, particularly towards a career's end. Consequently, *Hamlet* and *Corolianus* are considered the least Shakespearian tragedies based on the SiS1 score. In addition, the SiS score of *Troilus and Cressida* falls into the overlapping area of the SiS1-Medium2 class--close to its right corner--and the SiS1-High class.

The above results in Table 52 are produced with the combined set of 22 rules: the five SiS1-High rules, the six added ‘not/none’ rules and the 11 additional rules generated by some of the Shakespearian tragedies’ data points that needed to be assessed for at least one of their data’s membership both in an actual and ‘not/none’ class. Similarly to the histories-based model (Section 4.11.3)--and unlike the centroids of the comedies (Section 4.3)--the centroids of six of the nine tragedies give on the X-axis values that lie within the class of SiS1-High [0.6875-1] and in three of them the centroids are located in the part (= 0.76 to 1) of the class of SiS1-High that does not overlap with the SiS1-Medium2. The interpretation for the high results (SiS1-High) for the histories in the first stage of validation derived from the fact that most of the sets’ data points are closely bunched with the extensive use of duplets or triplets. But for the tragedies-model, and in relation to the discussion about the RSDs of data, the majority of the high-results are because two sets’ data points are represented by a single triangular class. The values of this class on the X-axis cover most of the Shakespearian tragedies in the area of the first (rather than second) standard deviation. Of course, the fact that a large area of the Set Two (from the end of the first to the end of the second deviation before and after the mean) is mapped also as a ‘not/none area’ caused the low results for *Coriolanus* and *Hamlet* and had a negative effect for other plays, too.

The next step is to look at the comparison of the counts of the individual 100 words’ frequencies of each vector-tragedy with the counts of the average-ideal Shakespearian document-vector of the nine tragedies. The indices of cosine similarity of each vector (of words’ frequencies’ counts) of the nine tragedies in comparison with the average-ideal Shakespearian document-vector can be viewed in the table below:

Tragedy title	Cosine index of similarity with the average-ideal tragedy
<i>Antony and Cleopatra</i>	These sets are not orthogonal cosine similarity = 0.987 angle = 9.348 angle type = acute
<i>Coriolanus</i>	These sets are not orthogonal cosine similarity = 0.943 angle = 19.416 angle type = acute

<i>Cymbeline</i>	These sets are not orthogonal cosine similarity = 0.969 angle = 14.196 angle type = acute
<i>Hamlet</i>	These sets are not orthogonal cosine similarity = 0.971 angle = 13.729 angle type = acute
<i>Julius Caesar</i>	These sets are not orthogonal cosine similarity = 0.972 angle = 13.559 angle type = acute
<i>King Lear</i>	These sets are not orthogonal cosine similarity = 0.976 angle = 12.617 angle type = acute
<i>Othello</i>	These sets are not orthogonal cosine similarity = 0.974 angle = 13.145 angle type = acute
<i>Romeo and Juliet</i>	These sets are not orthogonal cosine similarity = 0.935 angle = 20.740 angle type = acute
<i>Troilus and Cressida</i>	These sets are not orthogonal cosine similarity = 0.952 angle = 17.768 angle type = acute

Table 53: Cosine index of similarity of the well-attributed Shakespearian tragedies with the average-ideal tragedy.

As can be deduced by Table 53, the minimum cosine similarity index of the nine tragedies is that of 0.935 of the play *Romeo and Juliet* and the maximum is 0.987 of *Antony and Cleopatra*.

4.12.4 The Three Additional Input-Variables of the Two-Layered Complete Fuzzy-Logic Stylistic Classifier of Tragedies.

As in the experimentation with comedies, the inputs in the second layer of the fuzzy system are a fifth set of words (fifth variable), the SiS1 output (sixth variable) and the index of cosine similarity to an average-ideal document (seventh variable of the complete fuzzy-system classifier).

Below are the data points of the fifth set and its membership functions: two triangular actual classes, '5a', '5c', and one trapezoidal class, '5b'. Notice that there is a correspondence of the colours of the cells of the right column with the respective coloured membership class. The second graphical representation of Set Five's membership functions in the left side does not include the intermediate 'not/none' subclasses and so the three actual classes, '5a', '5b' and '5c', are more clearly visible.

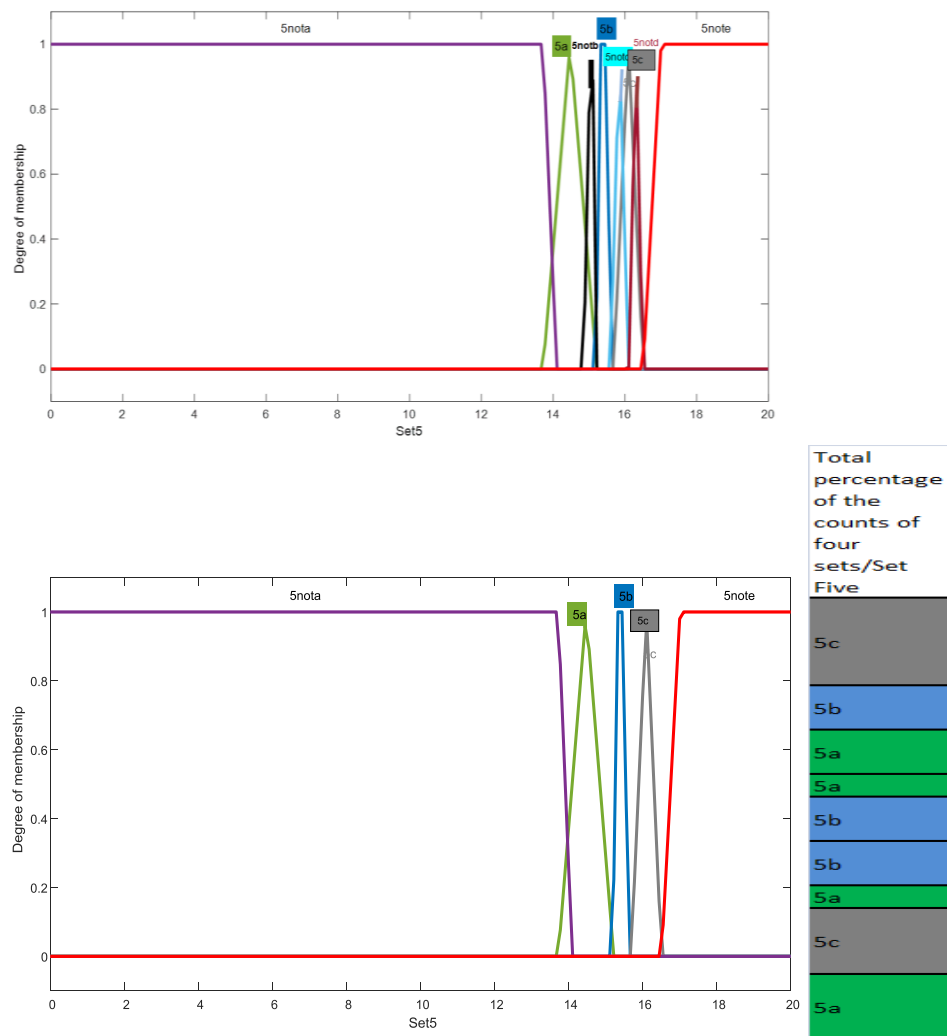


Figure 51: Memberships and data points of Set Five in the nine Tragedies.

This fifth set aggregates the counts for Sets One to Four. As you can see in Figure 51, based on the counts of the data points of the nine well-attributed tragedies, the non-continuous interval of Set Five ranges approximately from 13.8 to 16.5, whereas for comedies, as seen in Section 4.4, the range was from 12.1 to 16.1. As for the output class of the SiS1, which plays the role of an input to the SiS2 layer, the classes are exactly the same with the classes of the output variable employed in the primary (Section 3.3) and core experimentation with the comedies (Section 4.2).

Let us now, as with the comedies-based model, design the input variable of the index of cosine similarity for the second layer of the tragedies-based Fuzzy Stylistic Classifier of Shakespearianness. Below are the descriptive statistics that occurred after the comparison of the indices of cosine similarity of each of the nine well-attributed Shakespearian tragedies with that of the average-ideal document (see Section 7.3.9):

Average index (mean) of the indices of Cosine Similarity (CosSim) of nine tragedies	0.964
Standard Deviation	0.0161
Minimum index (CosSim)	0.935
Left Corner = Minimum – (minus) One Standard Deviation before the mean of the actual triangular class of the Cosine Similarity	0.9189 (rounded to 0.919)

Table 54: Descriptive statistics of the indices of Cosine Similarity of nine tragedies.

Based on the above data, and applying the methodology of the comedies- and histories-based experimentation, I designed the ‘not/none’ actual (triangle in blue) and actual areas (triangle in orange) of the input variable of the index of cosine.

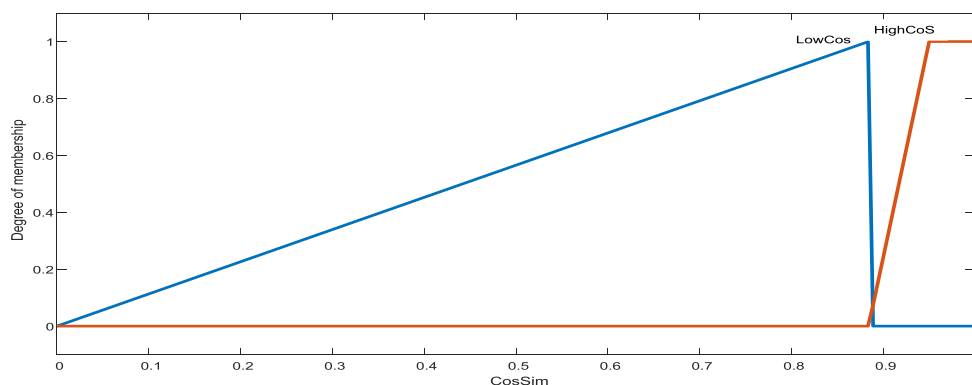


Figure 52: Membership functions of Cosine Similarity.

After the completion of design of the membership functions of the three additional input variables of the second layer of the tragedies-based stylistic classifier and by employing the same algorithm as for the comedies and histories of producing SiS2 scores (see Section 4.5), the next step is to proceed in the next stage with the Validation Stage 2 by experimenting with two well-attributed, sole-authored non-Shakespearian plays.

4.12.5 Experimentation with two Well-attributed, Sole-authored Non-Shakespearian Tragedies (Validation Stage 2).

Let us now start the validation process (Validation Stage 2) and experimentation with the two new well-attributed, sole-authored non-Shakespearian tragedies, *The Jew of Malta* by Marlowe, a play written around 1590 and *The Spanish Tragedy*, the version of 1592 written between 1582 and 1592, by Thomas Kyd (Craig and Kinney 2009, 217–18). *The Jew of Malta* was also assessed for its Shakespearianness in the primary experimentation (Section 7.2.4) with the holistic model of the nine Shakespearian plays. For this new play, the SiS interval of the newly-produced tragedies-based classifier is anticipated to be close (producing an output SiS interval from SiS1-Low to SiS1-Medium2 at most) to the single SiS1 score (0.35) that was produced with the holistic model, as *The Jew of Malta* is a non-Shakespearian play. Below are the two tragedies' gathered data points:

Tragedy title	Set One %	Set Two %	Set Three %	Set Four %	Set Five (Total of four sets) %
<i>The Jew of Malta</i>	2.5	8.2	1.6	2.2	14.5
	Cosine index : These sets are not orthogonal cosine similarity = 0.917 angle = 23.525 angle type = acute				
Tragedy title	Set One %	Set Two %	Set Three %	Set Four %	Set Five (Total of four sets) %
<i>The Spanish Tragedy</i>	2.5	7.7	2	2.6	14.8

	Cosine index : These sets are not orthogonal cosine similarity = 0.844 angle = 32.388 angle type = acute				
--	--	--	--	--	--

Table 55: Sets' data points and cosine indices of two well-attributed, non-Shakespearian tragedies. *The Jew of Malta* and *The Spanish Tragedy*.

```

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=7
MF1='1a':'trimf',[1.79 1.9 2.01]
MF2='1b':'trimf',[1.988 2.24 2.497]
MF3='1c':'trimf',[2.49 2.6 2.71]
MF4='1nota':'trapmf',[0 0 1.7 1.81]
MF5='1notb':'trapmf',[1.988 1.989 2 2.11]
MF6='1notc':'trapmf',[2.37 2.38 2.4 2.497]
MF7='1notd':'trapmf',[2.7 2.8 100 100]

[Input2]
Name='Set2'
Range=[0 20]
NumMFs=3
MF1='2':'trimf',[7.212 8.34 9.476]
MF2='2nota':'trapmf',[0 0 7.212 7.778]
MF3='2notb':'trapmf',[8.91 9.476 100 100]

[Input3]
Name='Set3'
Range=[0 10]
NumMFs=7
MF1='3a':'trimf',[1.19 1.3 1.41]
MF2='3b':'trapmf',[1.442 1.5 1.6 1.657]
MF3='3c':'trapmf',[1.742 1.8 1.9 1.957]
MF4='3d':'trimf',[2.158 2.3 2.441]
MF5='3nota':'trapmf',[0 0 1.1 1.21]
MF6='3notb':'trapmf',[1.956 2.014 2.017 2.159]
MF7='3notc':'trapmf',[2.44 2.58 100 100]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=3
MF1='4':'trimf',[2.344 2.778 3.211]
MF2='4nota':'trapmf',[0 0 2.344 2.561]
MF3='4notb':'trapmf',[2.994 3.211 100 100]

[Input5]
Name='Set5'
Range=[0 20]
NumMFs=8
MF1='5a':'trimf',[13.72 14.475 15.22]
MF2='5b':'trapmf',[15.2 15.3 15.5 15.6]
MF3='5c':'trimf',[15.69 16.1 16.51]
MF4='5nota':'trapmf',[0 0 13.72 14.1]
MF5='5notb':'trapmf',[14.85 15.04 15.1 15.2]
MF6='5notc':'trimf',[15.6 15.85 16.1]
MF7='5notd':'trimf',[16.11 16.31 16.5]
MF8='5note':'trapmf',[16.51 17.01 100 100]

```

Figure 53: Script displaying the membership functions of five of the seven inputs of the tragedies-based model.

As can be deduced by the crosscheck of the two plays' preceding Table 55 and the above script (Figure 53), the data point of Set Four (2.2) of *The Jew of Malta* is only of 'not/none' actual (see blue arrow). The same applies with Set Three (2) of *The Spanish Tragedy* (see yellow arrow) whereas its data point of Set Two (7.7) is somewhat actual and somewhat 'not/none' class (see red arrows). Consequently, it is necessary to form for these two new tragedies new rules mapping these 'not/none' combinations as antecedents. But the main differentiating factor with the known Shakespearian tragedies is that both new

non-Shakespearian plays have a cosine index that is situated in the low class of the seventh variable. The index of *The Spanish Tragedy* is especially low, 0.844 (and it is the least Shakespearian of all the known and disputed plays that were examined in this thesis). This shows that the counts of the words' frequencies of Kyd's play are very dissimilar to those that are detected in the well-attributed nine Shakespearian tragedies. This is a case in which is being activated the seventh input variable as a checking filter. By checking the data points with each of the classes of the script, the occurred combinations of the actual and 'not/none' classes for each play are described below.

Regarding the formation of the rules for the production of SiS2 scores, I will avoid adding the two general rules of the second layer that were employed in the other two genres, since the production of the SiS is straightforward with only one rule in the second layer of the fuzzy program, and for reasons of brevity. From the data points of the *Spanish Tragedy* are produced the followed rules:

R22 IF 1c AND 2na AND 3nb AND 4 THEN SiS1-Low

R23 IF 1c AND 2 AND 3nb AND 4 THEN SiS1-Medium2

R24 IF 5a AND SiS1-Low AND Low-CosS THEN SiS2-Low

R25 IF 5a AND SiS1-Medium1 AND Low-CosS THEN SiS2-Low

R26 IF 5a AND SiS1-Medium2 AND Low-CosS THEN SiS2-Low

In relation to Rule 22 (R22) the data point of Set Two falls just inside the 'not/none' class--instead of only actual--and the data point for Set Four falls just outside the 'not/noneness' and is assigned a membership only in actual class. In Rule 24 (R24), the data point of Set Five ('5a') falls just outside the 'not/noneness' and is assigned a membership only in actual class. Rules 22 and 23 assess Set Two's data points both as somewhat 'not/none' class (R22) and as actual class (R23).

Of course, the crucial factor is the existence of a low index of cosine similarity that leads to the formation of the Rule 24 (R24) and Rule 26 (R26) according to the algorithm of the rules' formation (as described in 4.5). As we can see, R24 contains in its antecedent part as input the SiS1 which is the consequent-output (SiS1-Low) of Rule 22 (R22) and Rule 26 (R26) the consequent-output of SiS1 (SiS1-Medium2) of Rule 23 (R23). As these two rules (R24, R26) play off each other and the result is going to be produced from the area of SiS2-Medium1 it is necessary to define, according to the algorithm, also a rule (R25) for SiS2-Medium1 as an antecedent which brings also as in Rules 24 and 26 as a consequent the

SiS2-Low (that is the case of any combination of antecedents in the second layer when the seventh input falls into the class of low cosine similarity).

From the data points of the play *The Jew of Malta* is produced the followed rule:

R27 IF 1c AND 2 AND 3b AND 4na THEN SiS1-Medium2

A second combination occurs, that is of ‘IF 5a AND SiS1-Medium2 AND Low-CosS THEN SiS2-Low’ but this is the same combination with R25 that was added to the set of our rules with *The Spanish Tragedy*, and so there is no need to duplicate it. Looking at the structure of this rule, we should once more recall that when there is Low-CosSim the SiS2-Low is produced irrelevant of the other data points, for instance here Set Five is the first actual class ‘a’, as described in the algorithm in Section 6.5.

Let us now see the results of the experimentation with these two well-attributed non-Shakespearian tragedies:

<i>The Spanish Tragedy</i>	Which rule fires	Output
0.391	R22, R23	SiS1
0.144	R25	SiS2

Table 56: SiS1 and SiS2 of *The Spanish Tragedy*.

<i>The Jew of Malta</i>	Which rule fires	Output
0.601	R27	SiS1
0.152	R26	SiS2

Table 57: SiS1 and SiS2 of *The Jew of Malta*.

Consequently, the output SiS interval of *The Spanish Tragedy* is compact (the limits have almost a distance of 0.25) and ‘it is almost certain’ that this play is non-Shakespearian. (An explanation of how such verdicts arise from the numbers is given in Section 7.3.15 of the Technical Appendix.) For *The Jew of Malta*, a less compact interval is produced, that is the interval 0.152 (SiS2)-0.601 (SiS1). Though SiS1 was higher than the SiS2 (0.35) of the holistic model of the primary experimentation, the final authorship verdict of SiS2 updates the initial score and practically with the evaluation of the low cosine index, resettles the initial attribution of relatively high Shakespeariness (0.601). Further analysis of this produced interval (0.151-0.601) in relation to the result of the primary experimentation

reveals that the middle value of this interval is 0.376 and this is very close to the verdict of the unique SiS1 (0.35) in the primary experimentation.

4.12.6 Experimentation with Two Disputed Tragedies (Testing Stage).

In this section there will be an experimentation with two tragedies that are widely believed to be co-authored and for which there is evidence that Shakespeare has contributed to their writing (Vickers 2002). The first of these two plays is *Timon of Athens* (included in the First Folio edition of 1623), which is widely believed that has been co-authored by Shakespeare and Middleton (Pruitt 2017), and the second is *Titus Andronicus* (the version of 1594) and it is generally considered that this second tragedy was the product of collaboration between Shakespeare and George Peele. *Timon of Athens* is also considered as one of these ‘problem plays’ as it was discussed on the basis of Boas’ arguments (Boas 1910, 344–408). This mixed nature of these problem plays perhaps necessitates the combination of stylistic classifications based on holistic and strictly genre-based models. Below are the two tragedies’ gathered data points:

Tragedy title	Set One %	Set Two %	Set Three %	Set Four %	Set Five (Total of four sets) %
<i>Timon of Athens</i>	2.2	8.8	1.4	2.4	14.8
	Cosine index : These sets are not orthogonal cosine similarity = 0.975 angle = 12.933 angle type = acute				
Tragedy title	Set One %	Set Two %	Set Three %	Set Four %	Set Five (Total of four sets) %
<i>Titus Andronicus</i>	2.2	8.1	2.2	2.4	14.9

	Cosine index : These sets are not orthogonal cosine similarity = 0.890 angle = 27.186 angle type = acute				
--	---	--	--	--	--

Table 58: Sets' data points and cosine indices of two disputed tragedies. *Timon of Athens* and *Titus Andronicus*.

As can be deduced by the crosscheck of the two plays' preceding Table 58 and the script (Figure 118) and checking the data points of all inputs with each of the classes of the script, the occurred combinations of the actual and 'not/none' classes for each play are derived:

Timon of Athens

R28 IF 1b AND 2 AND 3a AND 4 THEN SiS1-High

R29 IF 1b AND 2 AND 3a AND 4na THEN SiS1-Medium2

R30 IF 5a AND SiS1-High AND High-CosS THEN SiS2-High

R31 IF 5a AND SiS1-Medium2 AND High-CosS THEN SiS2-High (The actual fifth set causes the rise of one level from the consequent of R29)

Titus Andronicus

IF 1b AND 2 AND 3d AND 4 THEN SiS1-High (This combination coincides with the first (R1) of the five SiS1-High rules I formed in Section 4.12.2, and consequently there is no need to duplicate R1).

R32 IF 1b AND 2 AND 3d AND 4na THEN SiS1-Medium2

R33 IF 5a AND SiS1-High AND Low-CosS THEN SiS2-Low

R34 IF 5a AND SiS1-Medium2 AND Low-CosS THEN SiS2-Low

R35 IF 5notb AND SiS1-High AND Low-CosS THEN SiS2-Low

R36 IF 5notb AND SiS1-Medium2 AND Low-CosS THEN SiS2-Low

The experimentation with these two disputed (and believed to be co-authored) tragedies (*Timon of Athens*, *Titus Andronicus*) produces the following results:

<i>Timon of Athens</i>	Which rule fires	Output
0.726	R28, R29	SiS1
0.871	R30,R31	SiS2

Table 59: SiS1 (R28, R29) and SiS2 (R30, R31) of *Timon of Athens*.

<i>Titus Andronicus</i>	Which rule fires	Output
0.684	R1,R32	SiS1
0.14	R26, R34, R36	SiS2

Table 60: SiS1 (R1, R32) and SiS2 (R26, R34, R36) of *Titus Andronicus*.

In the two above tables can be viewed the two SiS scores for *Timon of Athens*, which are 0.726 and 0.871, expressing the belief that the Shakespearianness of this tragedy falls into an output area that starts at the overlapping area of SiS1-Medium2 and SiS1-High and ends close to the middle of the SiS1-High class where $x=0.871$. (Note that the maximum possible centroid on the *X-axis* is around 0.87-88). Consequently, the output SiS interval of *Timon of Athens* is very compact (the limits have a distance much less than a quarter of 1) and the tragedy's extract of the approximately first 10,000 words can be characterised as highly Shakespearian (the end of the extract is close to the medium of Act III, Scene VI, just before the long monologues of *Timon*, and it ends as follows: '...but time will, and so. I do conceive'.)

Though in our experimentation the analysis is focussed only on Shakespeare's stylistic habits, the high results derived from the tragedies classifier can revive Karl Klein's arguments about the sole Shakespearian authorship of *Timon of Athens* (Klein 2001). At the same time the consensus about the co-authorship of *Timon of Athens* in the book of *The New Oxford Shakespeare: Authorship Companion* is based on the exploration of structural linguistic differences than 'upon individual words' (Taylor and Loughnane 2017, 435) or sets of words and their conclusions have not led always to the strict separations of scenes (but mostly set of lines) to Shakespeare and Middleton. Apart from Scene 2 of Act 2, there does not exist any other complete part that is attributed to Middleton (Taylor and Loughnane 2017, 287) and Middleton's share is considered to be 'at most' around 40% (Holdsworth 2017, 374). This final statement of the maximum percentage attribution does not seem to deviate much from the SiS1 (0.726) results of our experimentation.

For *Titus Andronicus*, a less compact range of SiS interval values is produced, that is the interval 0.14 (SiS1-Low)-0.684 (SiS1-Medium2). The first extract of the approximately 10,000 of *Titus Andronicus* ends almost in the middle of Act III, Scene 1 ('...As frozen water to a snake'). So, according to the current results, *Titus Andronicus* cannot certainly be

considered as highly Shakespearian and this does not contradict with the general claim that this tragedy was been co-authored by Shakespeare and Peele and that at least the first scene of Act I and Act II, which are parts of the selected first extract of the play, have probably been written by George Peele, as shown by Brian Vickers (Vickers 2002).

We can recall even the recent historical theorisation of the *The New Oxford Shakespeare: Authorship Companion* that very early researchers, such as Malone, reached even to the point of ‘evicting’ *Titus Andronicus* from the Shakespearian canon (Taylor and Duhaime 2017, 71). It should be also mentioned that this play constitutes a landmark in the playwright’s stylistic development because *Titus Andronicus*, written around or some years before 1590, was probably Shakespeare’s first tragedy. (We should recall at this point the discussion in literature review in Sections 1.1.2 and 1.3.1.1 about ‘stylochronometry’.) On the other hand, *Timon of Athens* is classified as highly Shakespearian. Because the selected sets have a core semantic role and that most of the individual 100 words are widely accepted to be rich in stylistic information (as we saw in Section 7.2.1) it can be argued here that the SiS1-High classification essentially does not deviate from an early statement of a member of the *New Shakspeare Society* (1873), Frederick Gard Fleay, that ‘the nucleus, original and only valuable part’ of *Timon of Athens* is Shakespearian and that this ‘valuable part’ was further processed for the theatrical scene ‘by a second and inferior hand’ (Haug 1940, 227). It must be highlighted that the researchers were led to the conclusions about Middleton’s contribution based on the analysis of some non-lexical features as the punctuation. Such an input variable does not exist in the experimentation with the fuzzy tragedies-based classifier though it would be feasible to formalise and add it as a new stylistic marker.

As with comedies and histories, the results of the experimentation with the four tragedies (two in the Validation Stage 2 and two in the Testing Stage) can be represented by a scatter plot where the lower SiS-bound of each play is represented by the X-axis and the upper SiS bound by the Y-axis.

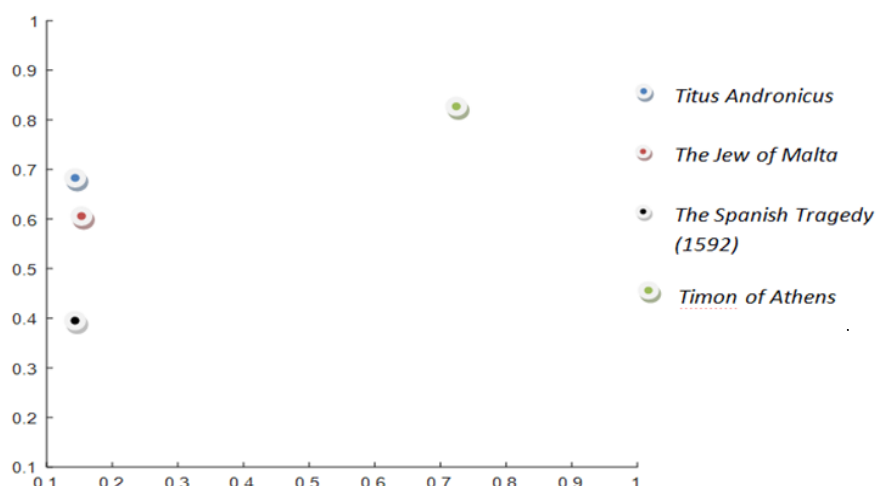


Figure 54: The scatter plot of the SiS interval of the data points of four tragedies under scrutiny.

In the above scatter plot the tragedies *The Jew of Malta*, *The Spanish Tragedy* and *Titus Andronicus* have at least one lower limit value of their intervals situated in the SiS1-Low area ([0-0.25], see Section 3.3). The tragedy *Timon of Athens*, according to the produced centroids, proved to be, as far as it concerns the first extract of approximately 10,000 words, more Shakespearian than any of the other plays of the Validation Stage 2 and Testing Stage (*The Jew of Malta*, *The Spanish Tragedy* and *Titus Andronicus*). In this tragedies-based fuzzy classifier it can be argued that by employing the principles of Fuzzy Logic it has been shown that *Timon of Athens* is a tragedy that resembles with high certainty the Shakespearian style of the nine well-attributed Shakespearian tragedies from the scope both of the counts of sets of words and counts of words' frequencies. Furthermore, *The Spanish Tragedy* is classified in the range of SiS1-Low to SiS1-Medium1 output values and so it would be unreasonable to claim that there is evidence suggesting the existence of Shakespearian stylistic habits in the version of 1592 of *The Spanish Tragedy*.

For the two remaining tragedies (*The Jew of Malta*, *Titus Andronicus*), there is a problem that resembles the lack of compactness of the range of the output SiS1 and SiS2 values that were produced in *Edward III* in the histories-based model but with much less uncertainty, since for *Edward III* the output SiS interval was 0.153-0.854, whereas in the tragedies-based experimentation the output intervals are 0.152-0.601 for *The Jew of Malta* and 0.14-0.684 for *Titus Andronicus*. In the case of *Edward III* it was mentioned that such uncertain results (widely spaced SiS intervals) can lead to further examination and it would be possible to either add new input variables or to experiment with a holistic model. For the *Jew of Malta* it can be claimed that there is the corroborating evidence of the primary

experimentation of SiS1 (0.35) which falls almost at the mid-point of the SiS interval produced in the core tragedies-based experimentation (and so it can be postulated that this tragedy ‘tends’ to be non-Shakespearian). For *Titus Andronicus*, as its cosine similarity falls into the low class, the results of the fuzzy classifier indicate that it is not a play with highly or even medium Shakespearian stylistic features.

With the finalisation of the building of the third tragedies-based stylistic classifier (with a total of only 36 rules), it has been shown how it is possible to model the style of Shakespeare based on the features of each genre and how feasible it is to automatically derive an interval of degrees of Shakespearianness for a disputed play. Taking also into account the potential of a similar to Mamdani type of fuzzy simulator and exploring the possibility of extracting authorship conclusions in the context of comparative analysis of SiS results from fuzzy programs with different structures and algorithmic processing, it is feasible to demonstrate the plethora of techniques Fuzzy Logic and fuzzy expert systems have at their disposal.

5 Chapter Five: Comparison with an Automated Sugeno-ANFIS (Adaptive Neuro-Fuzzy Inference System).

In this Chapter the aim is to describe the potential of combining the results of core Mamdani-based experimentation with an automated Sugeno-ANFIS, thus building an adaptive neuro-fuzzy inference system (Jang 1993; “Neuro-Adaptive Learning and Sugeno-ANFIS” 2015; Mathur, Glesk, and Buis 2016). There will be a brief description of the Sugeno-ANFIS system. The goal in this chapter is to assess the functional properties of an automated fuzzy neural-network simulator exploiting some elements of the previous three Mamdani classifiers.

5.1 Application of Sugeno-ANFIS Adaptive Neuro-Fuzzy Inference System in Stylometry.

The adaptive fuzzy neural inference system that is known widely as Sugeno-ANFIS (for differentiation to the other types of ANFIS, among which the Mamdani-ANFIS) was invented by M. Sugeno, T. Takagi and G. T Kang (Takagi and Sugeno 1985) and employs the same structural elements as the Mamdani type for building the rule mechanism and the same fuzzy operators but in a more elaborate form (see Section 7.3.17 of the Technical Appendix). But the Sugeno-ANFIS, unlike the Mamdani system, employs a set of training data (in our case the five sets’ counts) assigning them a specific value by joining the membership functions of all the rules’ antecedents. The Sugeno-ANFIS does not produce an output by finding a centroid or some other value derived from the truncation of specific rules’ output classes. Instead, the resulting value is produced from all the existing rules and it is not a result of defuzzification but the result of a weighted average of all rules’ output classes in the form of a linear equation of the form $ax+by+c$ (in which c is a constant). For instance, say that I build a fuzzy model with only two inputs as stylistic markers, Set One and Two, with x and y denoting the input variables Set1 and Set2 and a and b representing the parameters in the output function. (In fact, there is also the term of the zero-order Sugeno model (Sugeno and Kang 1988) where $a = b = 0$, and so the output is a constant c .)

Let us now explain step by step the features of Sugeno-ANFIS model by creating a holistic (that is, genre-blind) fuzzy Sugeno-ANFIS stylistic classifier built upon the data of five sets’ counts and a single SiS1 value of 24 of the 27 of well-attributed Shakespearian plays from the three genres plus ten from the Validation Stage 2 and the Testing Stage of the core experimentation. The reason I have selected plays also from the Validation Stage 2 and

the Testing Stage is because it is necessary for the Sugeno-ANFIS to have available for training a variety of input data points (variety in the senses of the range of data values), since the 24 well-attributed plays have an SiS1 score in the SiS1-High class and the fuzzy neural network mechanism will not work properly with the limited range of SiS1-High target values.

I decided to select only 24 plays from the total of 27 and I left out three of the Shakespearian ‘problem plays’ (*Hamlet*, *Troilus and Cressida*, and *King John*) implying that these plays could be employed for a testing stage. In addition, four plays of the second corpus (well-attributed non-Shakespearian or disputed plays) were left out because in the core experimentation they have either a very dispersed non-compact SiS interval (0.153-0.854 for *Edward III*, 0.152-0.601 for *The Jew of Malta* and 0.14-0.684 for *Titus Andronicus*) or are classified as highly Shakespearian (0.726-0.871 for *Timon of Athens*). Regarding the three plays with the non-compact interval it is implied that they could be used for classification in the testing stage in order to derive more concrete results through a holistic theorisation (model). Therefore, for the Sugeno-ANFIS experimentation I selected the majority, which is 36, of the total of the 41 plays that were employed in the core experimentation (Chapter Four).

King John perhaps should not be included in the canon of the Shakespearian histories, given the evidence adduced above. Furthermore, as discussed extensively in the beginning of Section 4.12.1, *Troilus and Cressida* and *Hamlet* are considered ‘problem plays’ (Boas 1910, 344–408) and so it could be possible-though this exceeds the purpose of this thesis-to evaluate their Shakespearianness stylistically again but this time with a holistic, irrespective of genre, Sugeno-ANFIS model. It would be also feasible to experiment in the testing stage with *Timon of Athens* and *The Jew of Malta* (but this again exceeds the purposes of this thesis).

Let us now start building the Sugeno-ANFIS system. The training corpus contains the data points of the counts of the five sets (so there are five inputs) and the output for the data points of the sets of 24 of the 34 plays is the value of 0.87 (SiS1-High). This is the highest SiS1 value a Shakespearian play can be attributed taking into account that the truncation of the SiS1-High output class (0.75-1) can give, at maximum, a value on the X-axis that is close to this interval’s central value (centroid) on the X-axis, approximately 0.87-0.88 (rounded to second decimal). The ten other plays are assigned a SiS1 or SiS2 score derived in the core experimentation of Chapter Four. In some cases, such as *The Spanish Tragedy*, when the SiS2 (0.144) derived in the second layer was smaller than SiS1 (0.391), I selected as a target the SiS2 score instead of SiS1. This was necessary for creating varied target output patterns,

varied in the sense of having output values in the range from 0.144 to 0.87. Therefore, the counts of the ten plays' five sets of words are the inputs and these are paired with the targets, the scores of SiS1 or SiS2 produced respectively in the first or second layer of the relevant genres-based fuzzy program. Below is a table with the training data of the selected plays.

Title of Play	Set One	Set Two	Set Three	Set Four	Set Five	Output-Target-SiS1-
<i>Antony and Cleopatra</i>	2.4	8.8	2.2	3.1	16.5	0.87
<i>Coriolanus</i>	1.9	9.2	1.3	3	15.4	0.87
<i>Cymbeline</i>	2.2	8.6	1.5	2.6	14.9	0.87
<i>Julius Caesar</i>	2.4	8.4	1.8	2.9	15.5	0.87
<i>King Lear</i>	2.1	8.8	1.5	2.9	15.3	0.87
<i>Othello</i>	2.1	8.2	1.6	2.5	14.4	0.87
<i>Romeo and Juliet</i>	2.6	7.9	2.4	2.8	15.7	0.87
<i>Henry IV, Part 1</i>	2.3	8.4	1.5	2.9	15.1	0.87
<i>Henry IV, Part 2</i>	2.1	8.3	1.7	2.3	14.4	0.87
<i>Henry V</i>	1.7	6.5	1.4	2.4	12	0.87
<i>Richard II</i>	1.7	5.9	1.8	2.8	12.3	0.87
<i>Richard III</i>	1.8	8.6	2.2	2.7	15.3	0.87
<i>As You Like it</i>	2.6	9	1.8	2.5	15.9	0.87
<i>The Comedy of Errors</i>	2.3	9.4	1.7	2.6	16	0.87
<i>Love's Labour's Lost</i>	1.9	6.2	1.5	2.5	12.1	0.87
<i>The Merchant of Venice</i>	2.4	8	1.6	2.7	14.7	0.87
<i>The Merry Wives of Windsor</i>	2.2	9.2	1.5	2	14.9	0.87
<i>A Midsummer Night's Dream</i>	2.5	8.3	2.1	2.6	15.5	0.87
<i>Much ado about Nothing</i>	2.1	9.9	1.6	2.1	15.7	0.87
<i>The Taming of The Shrew</i>	2.2	9.6	1.6	2.1	15.5	0.87
<i>The Tempest</i>	2.6	8.8	1.9	2.8	16.1	0.87
<i>Twelfth Night</i>	2.4	9.1	1.9	2.3	15.7	0.87
<i>The Two Gentlemen of Verona</i>	2.5	9.1	1.5	2.5	15.6	0.87
<i>The Winter's Tale</i>	2.3	8.3	1.8	2.8	15.2	0.87
<i>The Staple of News</i>	2.1	7.9	1.1	2	13.1	0.27
<i>All's Well That Ends Well</i>	1.8	8.3	1.6	2.4	14.1	0.496

<i>A Mad World, My Masters</i>	2.1	8.3	1.2	1.9	13.5	0.177
<i>The Wild Goose Chase</i>	2.4	8.7	1.2	1.5	13.8	0.262
<i>Measure for Measure</i>	2.2	8.3	1.4	2.9	14.8	0.459
<i>Edward II</i>	2.2	9.1	1.2	2.3	14.8	0.152
<i>Henry VI Part 1</i>	2.3	7.2	1.4	2.7	13.6	0.496
<i>Henry VI Part 2</i>	1.8	6.8	1.8	2.3	12.7	0.508
<i>Henry VI Part 3</i>	1.9	6.9	2.1	2.6	13.5	0.275
<i>The Spanish Tragedy</i>	2.5	7.7	2	2.6	14.8	0.144

Table 61: Training corpus of Sugeno-ANFIS composed of 24 well-attributed Shakespearian plays and 10 non-Shakespearian or disputed plays (see the 10 plays in grey background). In the first five columns with data there are the inputs of the five sets' counts and the last column is the target-assigned SiS1.

The data of these 34 plays in the Table 61 represent the historical data. The building of a Sugeno-ANFIS system necessitates a lot of pre-processing, trial and error-based primary experimentation with the modelling parameters since it is not easily discernible what is the appropriate number and type of membership functions for the particular training input data set. The use of Gaussian membership functions is usually a good solution proposed by many researchers (Mathur, Glesk, and Buis 2016) for the design of input variables but I will stick with the membership functions employed until now and so the selection of triangular classes will be made.

Let us now see the distribution of the data points of the 34 plays. The X-axis in the graph below (Figure 55) represents the number of the cell in the first column of the Table 61 and it contains one play. The Y-axis denotes the SiS1 score each of the 34 plays is assigned, the target. For the first 24 well-attributed Shakespearian plays you can see that the SiS1 score is identical (0.87) and there is a sequence of 24 dots. The next 10 dots are scattered around in the bottom-right corner as their target SiS scores differ.

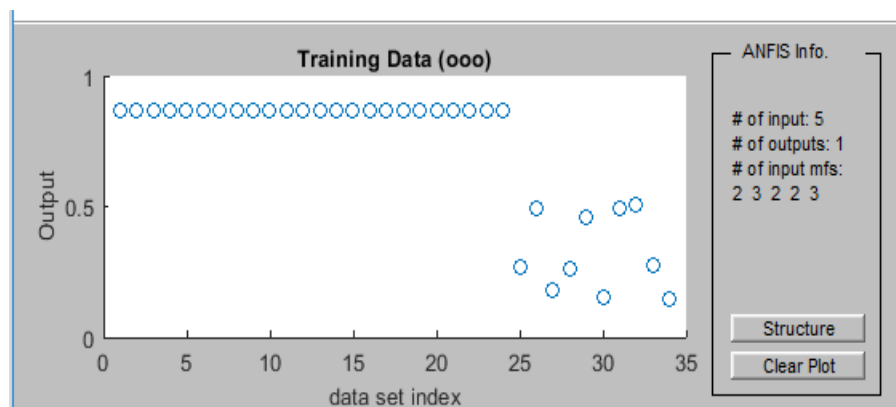


Figure 55: Distribution of the 34 plays and their target SiS score.

Before starting the generation of the Sugeno-ANFIS structure, I decided that triangular functions ('trimf') will represent each of the five input variables-sets.

The screenshot shows the configuration window for a Sugeno-ANFIS model. In the 'INPUT' section, the 'Number of MFs' is set to '2 3 2 2 3' and the 'MF Type' is 'trimf'. A note states: 'To assign a different number of MFs to each input, use spaces to separate these numbers.' In the 'OUTPUT' section, the 'MF Type' is set to 'constant' and 'linear'.

Figure 56: Selection of the number and type of membership functions for each of the five sets of Sugeno-ANFIS model.

The generation of the fuzzy system starts with the creation of a total of 12 triangular classes (2, 3, 2, 2, 3) and there is no differentiation of actual and 'not/none' class. The default design of the membership functions for each input follows a default estimation of the areas of x -values that will be assigned in each class. In fact, the second class of the second and fifth input variable covers much of the area of the first and third triangular class and so it creates a relatively large overlapping area.

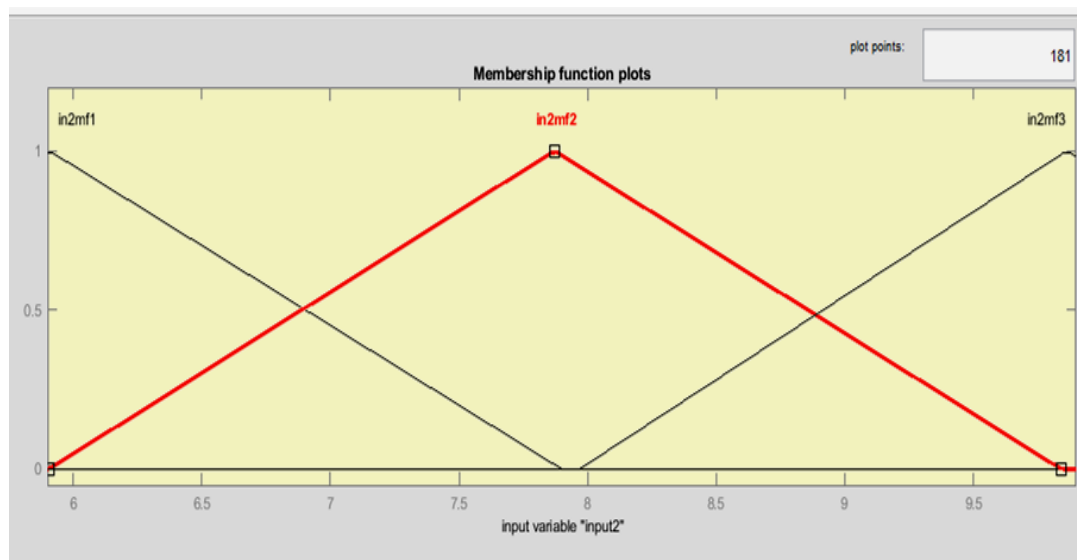


Figure 57: Membership functions of Set Two's data points in the Sugeno-ANFIS model.

It seems that the relatively small amount of data, as there are only five columns-variables with 34 rows-observations, the appropriate selection of the modelling parameters, particularly the selection of the triangular classes, and the adopted hybrid training algorithm, which minimises the total error from the beginning, produce optimal results for our Sugeno-ANFIS classifier very early, thus essentially from the second epoch.

The error for finding the solution during training, 0.000001, can be judged as minimal. Each epoch equals one forward and then backward training session-pass from all input data to the output data. For the definition of epoch see Section 7.3.18 of the Technical Appendix.

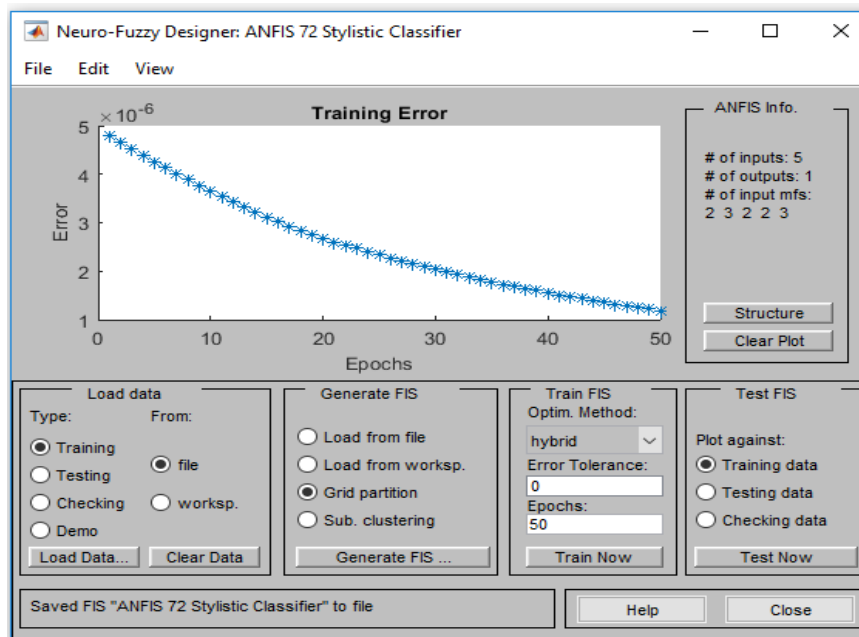


Figure 58: Sugeno-ANFIS training of five sets' data associated with their SiS1 scores of the core experimentation.

5.2 Evaluation of Sugeno ANFIS.

The three stylistic classifiers of the previous chapter function efficiently with a set of only 41 rules mapping the relations of seven inputs to two output variables. With the generation of the membership functions of only five input variables the produced Sugeno-ANFIS contains 72 rules but if I add one more class to the current classes of any of the variables which is composed of two classes, the number of rules will be 243 rules making the prediction-production of output (SiS1) dependable on numerous complexities. The stylistic Sugeno-ANFIS functions in general as a dynamic non-linear system. Although it

contains linear equations, the final output does not change linearly with the change of the value of input data points and that is why it should be construed as a non-linear system.

Let us now see the properties and the structure of the produced Sugeno-ANFIS:

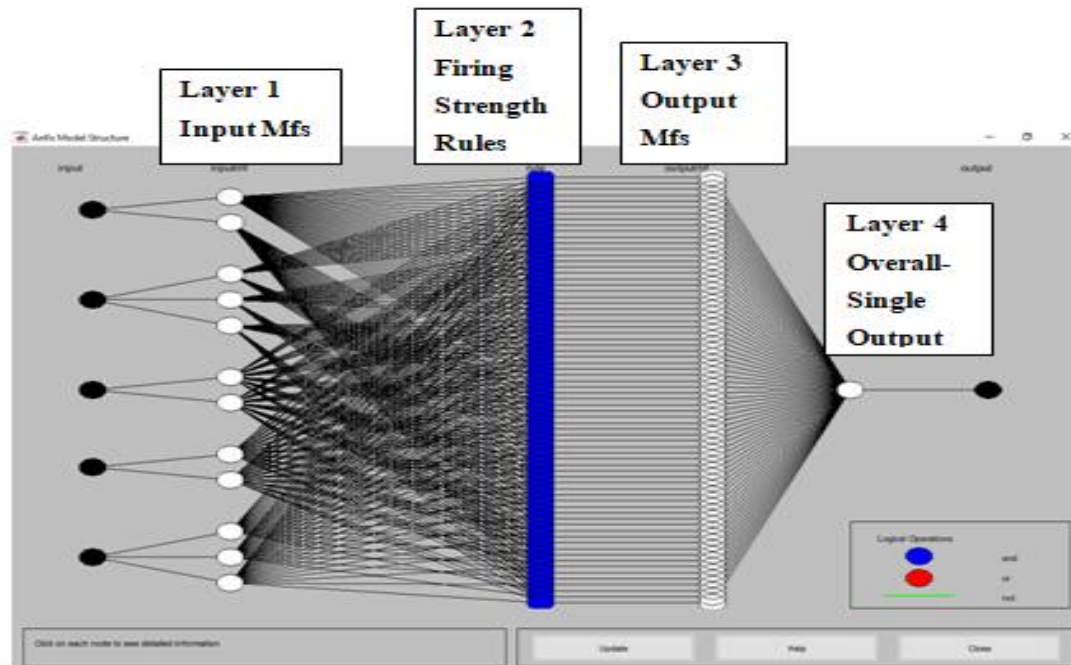


Figure 59: The structure of the adaptive fuzzy neuro-inference mechanism (Sugeno-ANFIS).

Along with the production of the structure of the Sugeno-ANFIS, the command line prints out some basic information of the structural properties of the layers of the Sugeno-ANFIS:

ANFIS info:

Number of nodes: 176
 Number of linear parameters: 432
 Number of nonlinear parameters: 36
 Total number of parameters: 468
 Number of training data pairs: 34
 Number of checking data pairs: 0
 Number of fuzzy rules: 72

Warning: number of data is smaller than number of modifiable parameters.

The so-called premise parameters are the 36 non-linear parameters that form the (12) triangular classes of the five input variables in the first layer of the neural network. The 432 linear parameters are the consequent parameters of the third layer that match as consequents the antecedents of the input variables, sets' counts here. But why are there so many--432--linear parameters? The answer is that since in our output-first order polynomial-linear model there are five inputs-sets and a constant, the total of the rules (72) have six (6) parameters and the product of them gives $72 \times 6 = 432$ linear parameters. (Recall the a and b parameters that was mentioned above for the function $ax+by+c$ in the description of the Sugeno-ANFIS.)

There is also available a validation-checking stage, but I avoided describing that for reasons of economy. I chose to run until epoch 50, so the training continues to try to improve but, in practice, the goal is reached very early. The neuro-fuzzy network (Figure 59) contains 176 nodes or neuron-units. Each node, apart from the first input layer, receives input information from the other neighbour nodes during the forward and backward pass and is associated with a function that calculates the product of the weight of its own input nodes. As shown in Figure 59, the neural network is composed of four layers (from a different perspective it can be claimed that there are five layers but, here, I ignore the layer with the input data points.) The second and third layers, as in the classic neural networks, are the hidden layers. In the first layer, after the input, the 12-input membership triangular functions are generated. In the second layer the membership functions of the input variables are formed as antecedents of 72 rules, and in the third layer 72 output membership functions (in the form of polynomial equations of first order) are derived. In the fourth layer there is only a crisp value, an output function (f). Each unit in each layer produces an input for the next layer and has a specific weight.

This Sugeno-ANFIS system is called adaptive because the nodes of the second and fourth layer are employing an adaptive updating function producing the weights on the basis of the degree of membership of each data point (here sets' counts) in each input's (triangular) class and based on the product of all these degrees of memberships. Essentially, the degrees of membership functions affect the role of the rules' weights-firing strength and the goal of the learning-training process is to normalise these weights from layer to layer. The third layer of the fuzzy neural network contains a mechanism that calculates the firing strength of each of the 72 rules and its contribution-ratio in relation to the total of all the rules' firing strengths

and this is forwarded to the next layer. The last layer contains a single node and practically calculates the sum of all incoming signals by employing a polynomial equation of the first order. (Practically every signal in the fourth layer equals the sum of the products of every rule's weight and the result of a relevant polynomial equation divided by the sum of all weights.) The fundamental idea behind the Sugeno-ANFIS is to build a neuro-fuzzy neural network that will produce new output values (SiS1 score) for new input values (sets' data points of new unseen plays) using a training stage where the target is the production of output values that are the right ones for the magnitude of the values of the data points of the input variables.

During the training, the Matlab system notifies us about the correlation of the Sugeno-ANFIS' parameters, printing out a warning that the 'number of data is smaller than the number of modifiable parameters'. There are 34 training data but 36 non-linear parameters, the latter being 12 (the total of membership functions of the five classes) multiplied by 3 (the number of parameters of a triangle: the peak, left and right bottom corners, see Section 7.3.5 of the Technical Appendix.) With fewer membership functions, for instance 6 ($1+2+1+1+1$) in our Sugeno-ANFIS, only two rules would be produced. The same problem would occur if the Sugeno-ANFIS system had fewer inputs. With an increased number of inputs, say the counts of the ten individual words-input variables for which there are only two membership functions, the system would probably produce thousands of rules, which could well be incalculable in practice. In the Sugeno-ANFIS system (as produced after the fuzzy neural-network training) there is no defuzzification process. The final result is derived from the mathematical function that sums up the results from all separate rules.

In the Chapter of Conclusion there will be a comparison of these ANFIS features with the three genre-based models of Chapter Four.

6 Chapter Six: Conclusions

This thesis aimed to develop a completely novel computational stylometric method for assessing the Shakespeariness of disputed, possibly co-authored, plays. Three fuzzy expert stylistic Mamdani-based classifiers of Shakespeariness were built and functioned in the various experimentation phases as universal stylistic estimators of Shakespeariness. (This term, universal estimator, is usually attributed to Sugeno-ANFIS systems but aptly applies for the three Mamdani-classifiers.) These systems provided original approaches to Shakespearian authorship attribution.

The examination of the similarity of anonymous and co-authored texts to the works of Shakespeare was carried out through three fuzzy simulator programs that have modelled stylistic features of, in total, 27 well-attributed, sole-authored Shakespearian plays. The introductory chapter (Chapter One) provided an analysis of the foundational principles of Fuzzy Logic, such as natural maths, approximate reasoning, Set Theory and partial truths-memberships, and gave an extensive literature review and a critical discussion of stylometric techniques starting from the rather unsystematic approaches of the middle of the eighteenth century up to the machine learning techniques of the first two decades of the twenty-first century. In Chapter Two the focus was on the methodological approach and the plan to cover existing gaps in digital stylometry by examining if the extraction of numerical occurrences of words and sets of words can give new attribution insights when fed through novel functional tools. The methodological approach of this thesis is of a mixed type, as I noted. It is inductive, deductive and exploratory. There is an attempt to explore comparatively the partial conclusions from the different phases of the experimentation and, in connection with related historical evidence for disputed plays, to assess the discriminating power of the generic properties of sets-based fuzzy expert systems. These offer a new theoretical arsenal and tool of computational stylometry.

In Chapter Three there was a comparison by analogy of the design and function of a fuzzy fan controller with the principles of designing a fuzzy stylistic classifier of Shakespearian style. This analogical description showcases the properties of the inference mechanism of fuzzy expert systems and the fuzzification-defuzzification processing through which stylistic classification takes place. An element of the cohesiveness of this thesis is the fact that the principles of Fuzzy Logic, Set Theory and Boolean Algebra are applied in the

stylometric domain in the same manner as in the real world domain of micro-electronics as can be viewed with the comparison of the fuzzy fan controller's inputs of humidity and temperature and output of fan's speed with the inputs of stylistic markers and the SiS result. Overall, in the primary experimentation (Chapter Three and 7.2) a holistic (that is, genre-blind) model was built based on the counts of six sets of a total of 102 individual words of nine well-attributed Shakespearian plays. This model did not contain any validation stage and was used to test two disputed and a well-attributed, non-Shakespearian play: *All's Well That Ends Well* (SiS 0.85), *Henry VI Part 1* (SiS 0.601) and *The Jew of Malta* (SiS 0.35).

In the core experimentation of Chapter Four three fuzzy stylistic controllers were built, one for each of the genres of tragedy, comedy and history. Each of the three controllers in this thesis has been programmed using the counts of four sets of a total of 100 words and an index of the counts of the words' frequencies in the known Shakespearian plays of each genre (12 comedies, six histories, nine tragedies). These counts are represented in the fuzzy controller through appropriate membership functions, recorded as geometrical shapes with certain mathematical properties. The core experimentation is composed of two validation stages (Validation of Inference Mechanism, Stage 1, and Validation of Performance, Stage 2) and a Testing Stage. The purpose in the core experimentation of Chapter Four and Chapter Five, in which a Sugeno-ANFIS was built, was the generation of new Fuzzy-Logic-based theoretical models of stylistic analysis and the construction of expert systems of detection of Shakespearian style.

The significant duality of the stylistic markers of counts of words' frequencies and sets' counts is safeguarded through the use and adaptation (see Section 7.3.9 of the Technical Appendix) of what is known as term frequency and the cosine similarity index between documents as vectors. During the experimentation with the 14 new plays (in Validation Stages 1 and 2, and Testing Stage), the three stages yielded intuitive and plausible inferences. The outcomes, in the form of SiS intervals, are almost in full agreement with the major claims of other researchers and the existing scholarship evidence.

With the first classifier were examined *The Staple of News* (SiS interval: 0.270-0.337), *All's Well That Ends Well* (0.496-0.601), *The Mad World My Masters* (0.177-0.199), *The Wild Goose Chase* (0.262-0.367), and *Measure For Measure* (0.459-0.726). The derived centroids-based-intervals more or less agree with the general evidence of historical scholarships and the claim that *All's Well That Ends Well* and *Measure For Measure* are almost entirely sole-authored Shakespearian plays, whereas *The Staple of News*, *A Mad World, My Masters* and *The Wild Goose Chase* are well-attributed non-Shakespearian plays.

With the second classifier, for histories, were investigated *Henry VI Part 1* (SiS interval: 0.496-0.544), *Edward II* (0.152-155), *Edward III* (0.153-0.854), *Henry VI Part 2* (0.508-0.604), and *Henry VI Part 3* (0.275-0.381). The derived SiS intervals of *Edward II* agree with the view that this play is a well-attributed, sole-authored Marlovian play whereas *Henry VI Part 1* and *Henry VI Part 2* are classified with SiS scores from the SiS1-Medium2 output class, just one below the SiS1-High class. As for *Henry VI Part 3*, it falls into the SiS1-Medium1 output class, just one output class-level below that of the two other parts (*Part 1*, *Part 2*) of *Henry VI*. *Edward III*'s interval score was assessed for its particularities and it is a case where it might be applicable to re-adjust the parameters and the kind of input variables of the fuzzy classifier for reasons of further examination. I have already explained in Section 4.12.6 why this play has certain particularities arising from the repetitiveness of erotic scenes that cause the contradiction of the Shakespearian sets' counts with the non-Shakespearianness of certain words' counts. So, even when the resulting classification of a play is significantly uncertain, when we find a large SiS interval, we can invoke for that extraordinary case the principle of adaptability, which is an advantage of the technicalities of the genre-based fuzzy expert systems.

As for the tragedies, I experimented with two well-attributed non-Shakespearian and two disputed plays: *The Spanish Tragedy* (SiS interval: 0.144-0.391), *The Jew of Malta* (0.152-0.601), *Timon of Athens* (0.726-0.871) and *Titus Andronicus* (0.14-0.684). The object of the listing of these 14 plays of the experimentation of Chapter Four and the derived SiS intervals is to show that the fuzzy stylistic classifiers have a discriminating role that allows the extraction of authorship verdicts on the basis of distinct decision boundaries, that is, SiS intervals, which are reasonable and align with serious scholarly evidence. A number of plays were selected for the Validation Stage 2 on the ground of their well-attributed non-Shakespearianness and a number of the other plays were selected due to their disputed, and possibly co-authored, status. This thesis does not imply that all existing non-Shakespearian plays of the early modern period would be assessed definitely as non-Shakespearian plays by the current fuzzy stylistic classifiers, as the similarities of styles is a complex problem and there can be a few cases where stylistic markers of an author's play of this period might resemble the style of Shakespeare. The objective of this layering of experimentation stages in Chapter Four is not orientated towards forming new authorship canons but towards qualifying the computationally functional and, from the scope of historical evidence, reasonable character of the three stylistic fuzzy classifiers-assessors.

Each of the two limits of the SiS interval evaluates quantitatively the stylistic features of the selected sets' counts and, indirectly, the words' counts of frequencies. These SiS intervals can be construed as intervals of general belief in the degree of Shakespearianness of each play under examination. Using a SiS interval instead of one score, as most researchers do, allows us to estimate the limits of our uncertainty. The benefit of such an approach is that we are avoiding the binary focus on Shakespearianness. Rather than simply asking if a play is Shakespearian or not, we can instead describe it in terms of an upper and lower limit of a degree of Shakespearianness. This 'allows uncertainty to be expressed, not just in information, but also in the reasoning process' (Kendal and Green 2007, 241–42). Even if almost the same SiS interval result is produced for two different plays, the reason for the result in each case may be different, since it depends on the different way that classes of variables in the antecedent and consequent part can be (Kendal and Green 2007, 241–42).

The three classifiers make autonomous decisions, as I extensively showed through the core parts of this thesis in the experimentation with each known (Validation Stage 1), well-attributed non-Shakespearian or disputed play. The step-by-step analysis of the phases of the experimentation was performed in order to show how stylometric data was formed into a knowledge base that supports the links of the inference engine. Taking into account the general axiom that expert systems 'model the higher order cognitive functions of the brain; and mimic the decision making of human experts' (Kendal and Green 2007, 21), this higher order of cognitive ability and classification is combined with Fuzzy Logic to produce solutions applicable to the real world. This approach is especially useful for authorship attribution research, where accuracy is distorted by numerous uncertain parameters (as in Mamdani classifiers) or associations of parameters (as in Sugeno-ANFIS).

The three Fuzzy Stylistic classifiers described here meet the three principles of the Knowledge Engineering, that is, compatibility, derivative and remedy knowledge (Kendal and Green 2007, 202). Thus, with the fuzzy design of the input and output membership functions and the prediction only for existing relations, the application of relevant algorithms, the creation of the actual and 'not/none' actual Shakespearian classes and the definition of constraints, the fuzzy stylistic simulators associate satisfactorily only for existing variables' classes the antecedents with consequents. This satisfactory association enacts the compatibility principle in a direct causal relation with a defuzzified crisp result, the derivative principle in our case. This kind of compatible design and the derivation of the result after defuzzification manage to detect violations of the system's principles applying remedy actions (the remedy principle), when this is needed, as for example in the case of Set Two or

of the paradoxical results with words' counts contradicting the previously SiS1 assessment of the Shakespearian four sets' counts. In the core experimentation, the algorithms of the two layers of each fuzzy program and the rules are those that set the preference (SiS1-High, SiS1-Medium2, SiS1-Medium1) or elimination criteria (SiS1-Low).

From a comparative approach, the significant features of the three Mamdani classifiers in contrast to the primary and Sugeno-ANFIS model are the use of the elaborate 'not/none' classes, the efficient mapping mechanism of the rules' formation, the adaptive constraints of the manual clustering and the intervals-based classification of the degrees of Shakespearianness. In the Sugeno-ANFIS experimentation, in Chapter Five, instead of prior knowledge-based algorithms, there is a training process through the structure of a fuzzy neural network. The Sugeno-ANFIS probably would necessitate probably much more than 100 rules in order to produce the authorship verdicts, whereas in the three genre-based classifiers almost around 40 were enough. As for the primary model, there is a lack of fuzzy transition from the actual to the 'not/none' actual classes and this may affect dramatically the output classification not assessing in a fuzzy mode the closeness of a data point between an actual and a 'not/none' actual class. In addition, the fact that in the Sugeno-ANFIS system the non-linear parameters can be more than the training data may cause overfitting for data that are quite dissimilar to those contained in the training data. (Note: Overfitting means that the fuzzy neural network functions well with the training data and the error during the training is driven to a small value. But when new data are presented to the network the error is much larger, and this indicates that the network has been adapted well for specific input patterns but the goal of 'generalised learning' has not been achieved. A major cause of overfitting is the existence of small or not representative/varied sets of input data. See analytical graphical illustration in Section 7.3.19 of Technical Appendix.)

In simple terms, the state of having less training data than non-linear parameters indicates that unnecessary complexity has been added to the system of Sugeno-ANFIS and this may cause the production of results that, however reasonable they may appear, are not right. But when data of new plays are not completely different (for all five sets) with those in the training corpus, such as in our case in the first subsection of Chapter Five, this danger seem to be minimised. On the other hand, since the default Sugeno-ANFIS mechanism does not allow the combination of different membership functions for the same input variable and the creation of more than one output, a complex and not necessarily efficient script would be needed for the Sugeno-ANFIS simulator to achieve a similar design to that of the three genre-based classifiers. Another disadvantage of the Sugeno-ANFIS model in comparison to those

of the Mamdani type is that it needs a large number of training data but a limited number of inputs (a maximum of three) in order to satisfy the principle-constraint of the correlation of the non-linear parameters and the training data. The advantage of Sugeno-ANFIS is that with its use there is no space for arguing that there can be an unintended bias from predefined constraints or design principles (as for instance treating Set Two as a special case with a SiS1-Low effect).

It could be argued that a disadvantage of the three genre-based classifiers is the creation of a so-called hesitancy effect, that is the fact that there is not an exact complementary relation--as in primary experimentation with the default 'not' classes--between neighbouring actual and 'not/none' actual classes. This can be construed in terms of strict mathematical formalisation as a weakness of the system in detecting precisely how actual and how 'not' actual a data point is in terms of the unity. The use of the second standard deviation in the core, but not the primary, experimentation for the design of the 'not/none' actual classes satisfies the proportionality effect: as we approach from the 'not/none' actual class to the neighbouring actual class, the membership in the 'not/none' class is reduced according to the standard deviation from the mean on the basis of which the actual class has been designed. And the lack of the precise mathematical formalisation is acceptable in systems that function under high uncertainty. In plain English, we cannot say that if something is 60% positive it is also 40% negative; rather we may hold that something is 60% positive, 30% negative and 10% something else that we cannot define; this third class may not be explicitly defined or designed.

In addition, the division of the output variable into trapezoidal classes and units of one-quarter assist in avoiding the sharp differentiation of the results when input data points of the same class differ slightly. In the default 'not' classes of the primary experimentation the hesitancy effect does not exist in contradistinction to the core experimentation and the manually designed 'not/none' classes on the ground of the second standard deviation for validation purposes. Though the lack of hesitancy effect is considered technically as an advantage of mathematical formalisation, in practice such a design in our case (core experimentation) would have as a result the completely different classification (by a whole SiS output class) in cases where, for instance, two sets' data points fall just merely 0.1 outside the first deviation-based actual classes of two respective input variables but they are inside the 'not/none' class, and vice versa. (Indeed, this difference of counts can arise simply from the different processing power of two different computers with which I counted the frequencies.)

The core genre-based models manage to detect these subtleties and assess the input data in a much fuzzier mode (than in the primary and the Sugeno-ANFIS experimentation), reducing the uncertainty in the derived authorship conclusions. The model of the primary experimentation explores the general-average Shakespearian stylistic tendencies and can provide us with a general view about the Shakesperianness of plays under examination. It is, though, less valid than the three genre-based classifiers. In effect, without the validation principle of the use of the second standard deviation and the selection of fewer and of larger counts sets this approach cannot always efficiently model plays that have peculiarities but are also Shakespearian. For instance, recall the discussion on page 100 (Section 3.2) and in Section 4.11.1 about the isolated data values and the single-point-based hypothetical triangles that was used in the core experimentation for modelling Shakespearian plays' data that do not look like the other plays.

The three genre-based fuzzy classifiers developed here could also function with more than seven input variables and in order to bring into consideration additional features such as punctuation marks, lexical density ratios or the counts of some words' frequencies. The increased variety of the stylistic markers-input variables would potentially increase the credibility to the final SiS results. A disadvantage of the design method in the core experimentation in Chapter Four might be the lack of the reshuffling. That is, the change of the way that the 100 individual words compose each of the four sets' (overcoming slightly the semantic criterion) before their final selection as it seems that the most distinguishing sets are those with a medium Relative Standard Deviation of about 10% to 20%. Such reshuffling of words was avoided as it would add much more pre-processing burden and extra material to this thesis.

By using Fuzzy-Logic-based expert systems it was possible to represent all the relations of the data between them, that is the relations-associations of the input with the output variables and this would not be feasible with PCA. Furthermore, if the PCA of the Set Two (in practice, of the individual counts of frequencies of this set's words) was employed in the core experimentation, there would be no major gain since it has already been formulated that Set Two is treated as a special case and has a significant role for the attribution of Shakesperianness. At the same time, by employing the cosine metric I managed to use vectors in order to represent the large dataset of the counts of 100 words without proceeding to any reductionism, which should be avoided since all words' counts of frequencies need to be assessed. In addition, the universal approximation capabilities of the fuzzy expert systems and the flexible structure (that allows the adjustment of the premise parameters of input and

output variables) enable us, in contradistinction to PCA, to capture complex non-linear behaviours, such as Set Two treated as special case.

Furthermore, the Fuzzy-Logic based models are considered universal in the sense that they can be efficiently applied in every kind of stylometric categorisation by approximating stylistic features whereas as noted in the Literature Review (page 47 in Section 1.2.10.2), this is not feasible with PCA. The reason is that with PCA even slight changes in the data of input variables might give a completely different result, and this is something that does not happen with the trapezoidal and triangular membership functions and the partial truths representation of the counts. In other words, with PCA it is not possible to build a universal model that can be applied to any disputed or anonymous play detecting the distances from the unity in the actual-Shakespearian classes of each membership function. In fact, PCA is generally a good observational tool that can be employed for specific stylometric problems. But PCA indices can be fed into the input system of the fuzzy simulators taking the form of input stylistic markers. For instance, an investigator could replace a set's count with a component variable that expresses the most significant information as produced by the counts of this set's words, but, again, caution is advised since the representation of the original problem (raw data of counts) in that case is transformed.

PCA's main utility is in reducing large numbers of variables to smaller numbers for ease of processing and visualisation, and in the present thesis the sets are relatively very few and manageable. Finally, as PCA is mainly an observational tool, it provides the investigator with neither a computerised and simulated environment nor a plethora of design options as far as it concerns the representation of the input (antecedents) and output (consequent) variables through the alternative types of membership functions. Particularly in connection with the properties of the computerised environment, another advantage of the Fuzzy Logic-based Matlab environment is the full visualisation of the variables' properties, the aggregation and the truncation process and that kind of feedback is not available with PCA, which has mainly a low dimensional observational role. Overall, the general disadvantages of PCA methodology is the loss of information, the lack of interpretability and the fact that it does not enable the investigators to receive feedback of the subtle differential patterns formed by the variables. Of course, it has to be noted that Fuzzy-Logic programs are expert systems and they are designed based on prior knowledge and density-based patterns analysis, whereas the method of PCA is not related to any kind of 'expertise' and does not aim at classifying every value from the subspace of input variables but instead constitutes a data-reduction tool.

Overall, regarding the three genre-based stylistic classifiers of Chapter Four, the building of a knowledge base is associated with the triggering mechanism of the inference engine through a semantically clear and economical set of propositions that are expressed by the antecedents. As for the inference engine, two algorithms assist the building of the ‘casual influences’ and ‘class-property relationships’ (Pearl 1988, 77) between the input (sets’ counts, SiS1, cosine index) and output variables’ classes (SiS1, SiS2). Lastly, one more advantage of the three genre-based models is that the inference mechanism and the rule base are characterised by uniformity and modularity, since computationally were applied the principles of conjunction (‘and’, ‘min’) and negation (‘not/none’).

Fuzzy Logic can be exploited widely in the computational stylometry, as it offers many options and technicalities of designing the ‘real-world’ of the closed authorship problems. When implemented in computational tools (toolboxes of Matlab Fuzzy and Neuro-Fuzzy), it offers a clear visual interface (supplemented by command lines option), plus apt and quick mathematical and geometrical data points’ formalisation in the representation of membership functions. This kind of formalisation and consequent user computer interaction cannot be derived by the mechanical analysis and the statistical methods of the twentieth century or even the machine learning techniques of twenty-first century described in the literature review. Besides, it is also possible to extract and combine multiple authorship conclusions by employing fuzzy systems with different structures, such as the intuitive Mamdani and the computationally efficient Sugeno-ANFIS system.

Other researchers can build upon the current three genre-based classifiers and model the style of any other author of the early modern period by simply changing the premise parameters of the input membership functions depending on the actual data points’ counts as found in the plays of the new author. The current three stylistic classifiers can also be extended to Bayesian probabilistic models if all the candidates of a disputed play, viewed in the context of a closed authorship problem, are modelled. The inference mechanism and the association of classes-membership functions of the input and output variables can assist researchers in the boundary detection of the style of various authors. This is because the openness of the structure of the input mechanism of the fuzzy systems can allow even the insertion of additional sophisticated stylistic markers, such as the Delta and z-scores of John Burrows’s approach or the ratios of lexical density, which can contribute to the generation of important conclusions about an author’s style.

The environment of the (Matlab-based) fuzzy logic simulator and the general structure of the models-classifiers applied in this thesis could be also used with markers used

by other researchers studying the work of early modern dramatists. The method developed in this thesis could be employed in order to explore which scenes of *The Two Noble Kinsmen* and *Henry VIII* have been written by Shakespeare and which by Fletcher. As noted in the literature review (in Section 1.2.1), it seems that Shakespeare in contradistinction with Fletcher had no strong preference for double or ‘feminine’ line-endings (Spalding 1833, 11-12) and his style is characterised by an increased number of enjambments. So, the first input variable could contain classes with measurements of ‘feminine endings’ in terms of proportion of the textual segments-extracts (total of words) and the second variable could contain classes of enjambments represented by continuous intervals of values. The modelling of these two input variables presupposes the design of the membership functions, classes based on the evidence-data of segments (or whole scenes) of some of the well-attributed, sole-authored plays of each candidate author and according to specific constraints, such as using plays from the same genres and approximate periods as those in the present study.

(As mentioned in the literature review in Section 1.2.1, Spedding did some relevant work and explored the metrical similarities between *Cymbeline* and *The Winter’s Tale*, intending to find the average rate of use of redundant syllables in verse lines (Rolfe 1884, Appendix, 14)). The third and fourth set-input variable could be respectively the number of rhymes and the number of pauses as a proportion of a segment or a whole scene (measured in words). By adopting a simple Fuzzy-Logic-based version, two such separate stylistic models-classifiers could be built, one for Shakespeare and one for Fletcher. Regarding the investigation of a disputed segment or a whole scene (of *The Two Noble Kinsmen*), the candidate with the higher SiS would be also the most probable author of the disputed scene under scrutiny. In a more complex version of fuzzy stylistic classifier, these two models (Shakespeare, Fletcher) could be merged into one, so that a unified system produces two SiS outputs, one for Shakespeare and one for Fletcher. The first SiS could produce a degree of Shakespeareanness (SiS1) and the second a degree of Fletchererness (SiS2). The higher of these two SiS outputs would be identified with the attribution of the examined scene to one of these two authors.

Another option for other researchers adopting the present method would be to use texts of different plays or segments of the same text of a specific author and to find the descriptive statistics of the length-based classes of words. By constructing four input variables based on the stylistic markers that Mendenhall employed (Mendenhall 1897;1901), thus, for example, the length of the two-letters, three-letters, four-letters and five-letters words, it is possible to model in a fuzzy mode the words’ length-based style of an author of

the early modern English period (as Mendenhall did with Marlowe) and then to detect its stylistic similarity with extracts of a disputed play. It would be also possible to apply the same logic by employing as markers n-grams.

It is also feasible to apply the method developed in this thesis for detecting the style of a disputed text among a limited number of authors or poets, for example five of the 25 English poets of the Restoration period (1660-1685), adapting the design of Burrows's relevant stylometric research (Burrows 2002). After the analysis of a number of texts or segments of each poet's artefact, the investigator can build four input variables-sets of words. The criteria for the formation of these sets of words can be different from those adopted in this thesis. The criteria can be based on the Delta (Burrows 2002), Zeta and Iota (Burrows 2007) procedures adopted by Burrows. Thus, Set One could contain classes of say 6-10 common words of generally high frequency. Each class of Set One would represent the generalised style of one of these five poets in our case (and in that way would be formed five classes, one for each poet). The counts of the words that would contribute to the formation of each class should not be too distanced between them, for example regarding Set One the minimum word's averaged counts could not be less than 40 and not more than 50) and the peak of each formed class could be the average/centroid of the averages of the counts of these (6-10) words. So, the (five) classes of that input variable would model the average tendencies of these words in poems written by the five of the 25 English poets of the Restoration. The other three sets of words (Set Two, Set Three, Set Four) under the same premises could model the averaged counts of words that are sporadic and rare in known poems of one poet and respectively rare and not existing in the styles of the other four English poets.

By applying such a design, two input variables of different degrees of sporadicity and one input variable of rarity could be formed, and by opting out for that design four in total input variables would be built (for the definitions of sporadicity and rarity, see Section 1.3.1.4). The SiS output variable can contain five classes, one for each candidate poet, and each class could produce for each author an index of stylistic similarity. Entering to the fuzzy classifier the input data of a disputed poem could produce one SiS for each of the five candidate poets. The poet with the highest SiS would be also the most probable author of the disputed play. (The same method could be employed for the other 20 poets of the Restoration period and in that manner from five fuzzy classifiers could be derived five respective highest SiS and deductively the highest of all candidate poets can be detected. From this scope, the methodology applied in this thesis can be also applied for open authorship problems.)

Evidently, these methodological steps can be followed for authorship problems of theatrical plays, too.

Furthermore, though degrees of memberships have a different meaning than probabilities (see the explanation in the Technical Appendix, last paragraph of Section 7.3.15), it is possible to ‘turn’ (the proper term is associate) the initial results of the Fuzzy Stylistic Classifiers into probabilistic estimators on the basis of Type-2 logic. (In general, Type-2 design can generate a second membership degree of uncertainty for the first membership degree and give us a probability for the veracity of Type-1 result, the SiS1 score). There are various ways for accomplishing that goal and it necessitates mathematical formalisation (Dubois et al. 2004; Liu and Li 2005). Let us refer to a specific and simplified example for reasons of illustration. Suppose that the intention is to investigate the Marlovianess and Shakespearianess of *Henry VI Part 3* and that historical scholarship lead us to investigate the degrees of stylistic similarity of this play to each candidate author’s style and, in this context (Shakespeare ‘AND/OR’ Marlowe) it is necessary to explore also what are the probabilities that it is a co-authored play. Therefore it is necessary initially to build one Fuzzy Logic Mamdani-based model for Shakespeare and one for Marlowe. (A merged classifier is another option but the aim here is to illustrate the potential and not the complexities.)

Suppose that both separate classifiers are two layered. The first layer in each classifier can have similar to the thesis’ input and output variable’s membership functions but the output classes of SiS2 in the second layer should have classes represented by at least a curve function such as the so-called ‘gauss2mf’ or ‘ Π -Shaped mf’.

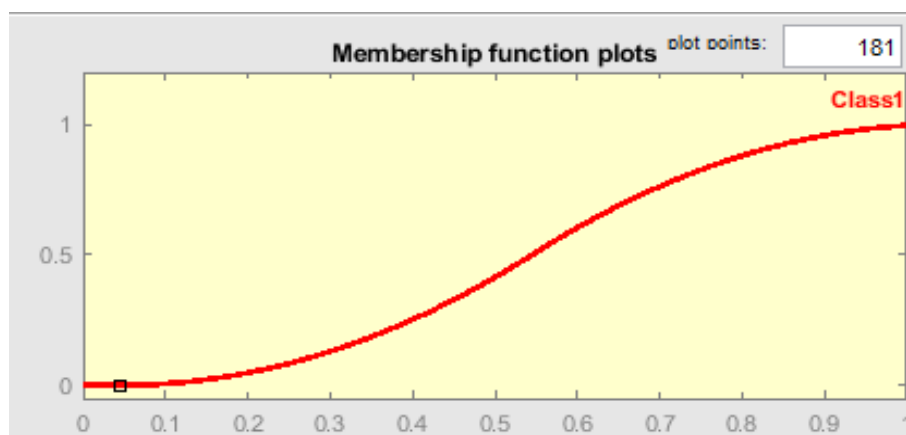


Figure 60: A Π -Shaped membership function.

When are entered to the two separate fuzzy classifiers the data points (of any kind of markers, as discussed in the previous paragraphs) of the disputed play, the first layer in each fuzzy classifier will produce, as in the core experimentation of Chapter Four, one output SiS (SiS1) for Shakespeare and one SiS (SiS1) for Marlowe. The second layer in each separate fuzzy classifier can take as input each previously produced SiS1 and based on mathematical formalisation can ‘turn’ SiS1 into a probabilistic index SiS2 (kind of Type-2 fuzzy logic value), and, therefore, one SiS2 (in % or 0.1 to 1) for Shakespeare and one SiS2 (in % or 0.0 to 1) for Marlowe can be derived. (Of course, the relation of SiS2/Type-2 value with SiS1/Type-1 value necessitates further analysis in the sense that this value as a certainty index might have various meanings: generally certainty of the degree of Type-1/SiS1 or certainty of each candidate author being the single author of the disputed play?)

Now suppose in actual experimentation the two classifiers are run and the two SiS1 and two SiS2 scores for each candidate author are produced. I keep only the SiS2 scores—probabilistic indices (Type-2 values) that are the Type-2 values/transformations of the two respective SiS1 scores. In the next step, a new fuzzy expert system can be built with a very simple structure, as it can have two input variables, each composed of classes with x -values in the range of 0 to 1 (where 1 represents 100% certainty). For reasons of illustration suppose there is only one output class. Each unique class of the two input variables (one for Shakespeare’s and one for Marlowe) in the new fuzzy classifier should have the same design with the output classes of the second layer of the two separate fuzzy classifiers. In this second fuzzy program a fuzzy ‘probabilistic OR’ (‘probor Or’) is applied in order to produce a new SiS score. The ‘probabilistic OR’ for two inputs (a, b) takes their memberships and instead of the ‘min’ operation applies the function $a+b-a*b$ (See Section 7.3.17 of Technical Appendix). This method optionally is applied as a type of a probabilistic ‘OR’ method instead of the minimum method in the selection of the membership of all inputs and the centroid in the defuzzification process is produced from the output area formed after the calculation of the membership produced with the ‘probor Or’ function applied on the inputs’ membership function.

With the proper formation of a few rules (and essentially two or three classes of the output variable) the new fuzzy program produces a probabilistic index of co-authorship of *Henry VI Part 3* and in that way both input variables and their difference is taken into account (whereas with the use of minimum operation there would not be probabilistic approach). If, for instance, the two separate fuzzy classifiers produced as SiS1 for Shakespeare 0.8 with SiS2 0.8 (Type-2 value, say 80% certainty) and as SiS1 for Marlowe

0.3 with SiS2 0.1 (Marlovian Type-2 value, say 10% certainty), then after entering as inputs these two SiS2 values to the new fuzzy classifier and by applying the ‘probor OR’ defuzzification method, the membership (y-axis) of the new SiS in the output class would be $0.1+0.8-(0.8-0.1)=0.2$ (= y) and the new SiS probabilistic score on x-axis, depending on the parameters of the membership function, could be around 0.4 (= 40%).

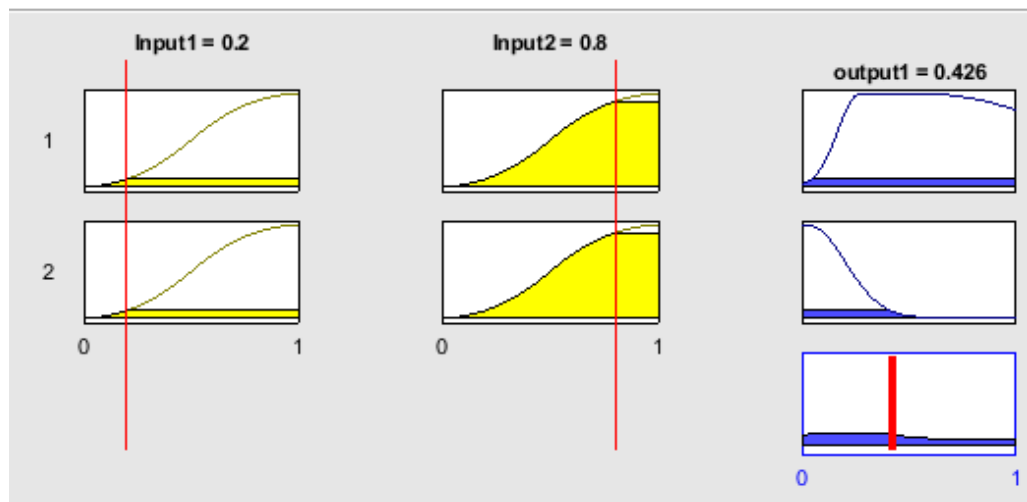


Figure 61: An example of a probabilistic co-authorship estimator.

The question that arises is what this probabilistic index might denote? The answer is that it can show what are the probabilities for *Henry VI Part 3* of being a co-authored by Shakespeare and Marlowe play. Therefore, the first conclusion is that the degree of Marlovianess (0.3) detected in the first layer of the first two separate fuzzy classifiers for this play falls near the SiS1-Low class, whereas Shakespearianess is high (0.8) falling into the class of SiS1-High. The second conclusion that is drawn from this new additional probabilistic index of the new fuzzy program can be construed as ‘it is less than 50% probable’ that *Henry VI Part 3* is co-authored by Shakespeare and Marlowe and that it ‘tends’ (See Section 7.3.15 for the linguistic categorisation) that Marlowe has not contributed to its writing. The variety of the markers-input variables, as employed by other researchers, and the possible combination of degrees of membership with probabilistic estimation show the potential of further exploitation of the Fuzzy-Logic tools (with the proper mathematical formalisation) in the area of the computational stylometry.

In general, there are numerous other alternative ways of applying the Fuzzy-Logic-based methodology of this thesis and with the proper adaptations these alternatives can be used for a plethora of problems of authorship verification and attribution of the early modern

English drama, whether these problems concern the assignation of whole plays, such as any of the Parts of the trilogy of *Henry VI* or extracts such as the Scene 2 of Act 2 of the play *Timon of Athens*.

The goal of this thesis was primarily to develop a novel, automated, robust and fully operational new type of stylistic classifier of Shakespearianness with all the benefits of automation and natural mathematics approximation. Secondly, it sought to provide some evidence about the profile of the plays it examined or the styles of their authors. More generally, this thesis provides experimentally an extensive survey of Fuzzy-Logic-based methodologies that can be applied for the resolution of authorship attribution problems but also potentially of any kind of stylistic analysis of various types of texts. For example, the present approach could be used in the context of profiling in clinical psychology and preventative forensic research for the detection of expressed neuroticism or aggressiveness. It could be applied to the prevention of aggressive acts by students in academic environments in the United States and in the context of gun licensing. It is also significant that with any of the fuzzy classifiers (holistic or genre-based) it is feasible to assess the Shakespearianness of texts of authors beyond the Elizabethan/Jacobean period.

From simple essays of undergraduate students to modern theatrical plays in English, the fuzzy expert systems can assess them for the resemblance to the style of Shakespearian plays. These tools, for instance, could have been used by Gary Taylor when preparing his adaptation of Shakespeare's lost play '*Cardenio*'. Taylor could have employed the tools used here to test his adaptation's Shakespearianness (Bourus, Terri, and Taylor 2013). Another interesting use of the deliverable product of this thesis is as an automated classifier for the comparison of Shakespeare's and his contemporaries' styles to the style of authors that have written in other languages or other eras. A comparison, after proper processing and translation of the corresponding stylistic markers, of Ancient Greek tragedies with the tragic Shakespearian style of the sixteenth century could reveal if the quality of a written artefact has been diachronically the result of a largely quantifiable recipe of individual words formed into semantic sets with a general, set-based and a specific, words' frequencies-based analogy using thresholds of cosine index.

Regarding the emphasis laid on the role of words' sets' and not only of words' frequencies counts as stylistic markers for the Shakespearian authorship attribution, the alternative hypothesis (H_1) I set in Section 2.5 has been verified. There are similarities regarding the writing style of a specific author (in our case Shakespeare) and these can be uncovered by exploring through Fuzzy Logic not only the counts of distinct words but also

the counts of sets of semantically related words. The counts of sets of words carry a lot of stylistic information since the selection and grouping of words with various features and of various frequencies assist naturally the formation of different strata of words that an author employs in his plays. The random-but deterministic, due to the criteria of the selection of the words, formation of different frequencies' strata and the individual words' intra-set compensatory function convert these 'collective' sets into individual stylistic patterns that, as shown throughout the core experimentation (Validation Stages 1-2 and Testing stage), contribute to the formation of authorship verdicts that do not contradict historical evidence. The present investigator's claim is that the combination of sets-markers with the markers of the counts of words' frequencies assists stylistic discrimination much better than counts of words' frequencies alone.

Moreover, following the track of the emerging machine learning techniques in the beginning of the twenty-first century (see Section 1.3.1), I succeeded in this thesis in employing Fuzzy Logic to overcome the limitations of the methods of the descriptive and multivariate statistics. Such limitations are, for example, the lack of simulation and functionalities arising from the geometrical representation of the exact positions of the data points in a two-dimensional coordinate system. By overcoming these limitations, Fuzzy-Logic-based stylistic classifiers can have a key universal role for almost any kind of authorship problem since a plethora of stylistic features can be not only qualitatively assessed but also quantitatively represented in terms of partial truths. The major advantage of the principles of Fuzzy Logic is that they can assist the finding of solutions for complex problems in environments with many uncertain parameters. Fuzzy Logic principles can provide approximate, and so less precise, but safer calculations that are based on the worst-case scenario which, in practice, is enacted by the minimum operation applied on the data points of the input variables.

In conclusion, this thesis overcame the limitations of the conventional mathematical tools by initiating various tools of Fuzzy Logic. It built three sophisticated genres-based automated stylistic classifiers of Shakesperianness. These classifiers reduced the uncertainty of authorship verdicts with the use of an evidential interval of indices of Shakespearian Similarity. The key achievement of this systematic approach is that disputed plays' counts of stylistic features are not simply measured or represented but they are exploited for the direct production of authorship verdicts through an interactive computer environment that can be adjusted and re-adjusted for a variety of authorship conclusions apart from the strictly defined closed or open problems of authorship attribution. The deliverable concrete result of this

thesis is the creation of automated stylistic expert systems that mimic the human experts in this field.

- 7 Appendices: List of Plays' Versions, Textual Addendum and Technical Appendix.
- 7.1 List Addendum: Plays (and their Date Versions) of the Core Experimentation (Chapter Four).

Below are the versions of well-attributed Shakespearian plays that were used in the core experimentation of the three genre-based models. The plays are listed in alphabetical order. The texts of these plays are XML-encoded in TEI-conformant versions that were modernised and regularised by researchers at the Centre for Literary and Linguistic Computing (CLLC) of the University of Newcastle in Australia under the guidance of Hugh Craig, an expert in sixteenth and seventeenth century literature. The texts have been all regularised either in a single R or SR mode. SR stands for silent regularisation, a slightly differentiated, milder technique of regularisation (R). In cases where R versions did not exist, the existing SR versions were used. This differentiation of regularisation and minor regularisation has generally a minor effect for the purposes of the investigation in this thesis. In effect, the words-markers of the sets have a stable form and are not affected as far as it concerns their detection by the AntWordProfiler.

A Midsummer Night's Dream (Com., 1623, 10,073 words) R

Anthony and Cleopatra (Trag., 1623, 10,094 words) R

As You Like It (Com., 1623, 10,004 words) R

Coriolanus (Trag., 1623, 10,043 words), SR

Cymbeline (Trag., 1623, 10,057 words) R

Hamlet (Trag., 1623, 10,013 words) R

Henry IV, Part 1 (Hist., 1623, 10,042), SR

Henry IV, Part 2 (Hist., 1623, 10,077 words), SR

Henry V (Hist., 1623, 9,995 words) R

Julius Caesar (Trag., 1623, 10,017 words), SR

King John (Hist., 1623, 10,094 words) R

King Lear (Trag., 1623, 10,092 words) R

Love's Labour's Lost (Com., 1623, 10,002 words) R

Much Ado About Nothing (Com., 1623, 10,008 words) R

Othello (Trag., 1623, 10,082 words) R

Richard II (Hist., 1623, 9,945 words) R

Richard III (Hist., 1623, 10,058 words) R
Romeo and Juliet (Trag., 1623, 10,116 words) R
The Comedy of Errors (Com., 1623, 10,145 words), SR
The Merchant of Venice (Com., 1623, 10,012 words) R
The Merry Wives of Windsor (Com., 1623, 10,037 words) R
The Taming of The Shrew (Com., 1623, 10,136 words) R
The Tempest (Com., 1623, 10,013 words), SR
The Two Gentlemen of Verona (Com., 1623, 9949 words) R
The Winter's Tale (Com., 1623, 10,069 words), SR
Troilus and Cressida (Trag., 1623, 10,057 words) SR
Twelfth Night (Com., 1623, 9,966 w), SR

Versions of well-attributed non-Shakespearian or anonymous-disputed plays (Validation Stage 2 and Testing Stage):

All's Well That Ends Well (anonymous-disputed, Com., 1623, 9,966 words) R
A Mad World, My Masters (Thomas Middleton, Com., 1608, 9,968 words), SR
Edward II (Christopher Marlowe, 1594, Hist., 10,028 words) R
Edward III (anonymous-disputed, 1596, Hist., 10,047 words) R
Henry VI, Part 1 (anonymous-disputed, 1623, Hist., 10,087 words) R
Henry VI, Part 2 (anonymous-disputed, 1623, Hist., 10,037 words) R
Henry VI, Part 3 (anonymous-disputed, 1623, Hist., 10,004 words) R
A Measure for Measure (anonymous-disputed, Com., 1623, 10,099 words) R
The Jew of Malta (Christopher Marlowe, Trag., 1633, 9,989 words) SR
The Spanish Tragedy (Thomas Kyd, Trag., 1592, 10,071 words), SR
The Staple of News (Ben Jonson, Com., 1631, 9,953 words) SR
The Wild Goose Chase (John Fletcher, Com., 1652, 10,019 words), SR
Timon of Athens (anonymous-disputed, Trag., 1623, 10,087 words) SR
Titus Andronicus (anonymous-disputed, Trag., 1594, 10,084 words), R

7.2 Textual Addendum: Reaching Stylometric Conclusions Quickly and Efficiently by Using Fuzzy Logic.

Based on the paradigm of the fuzzy fan controller (see Section 3), by analogy I experimented with a real authorship attribution problem. The received feedback was exploited for the formation of the building of corpus and the design of the core experimentation of the thesis. This subsection gives a detailed account on how we can model the counts of sets of words by using trapezoidal and triangular membership functions.

7.2.1 Corpus Building and Sets of Words.

Three comedies, three tragedies, and three history plays were chosen at random from the well-attributed, sole-authored Shakespeare canon and approximately the first 10,000 words of each were extracted. It was initially decided that two styles of two plays for which Shakespeare's sole authorship is in doubt, the comedy of *All's Well That Ends Well* and the history play of *Henry VI, Part 1* would be compared with the style of the large, well-attributed, sole-authored canon. Then a new play was added for the Testing phase, a well-attributed play of Christopher Marlowe, *The Jew of Malta*. The texts of these plays are modernised versions from the *Internet Shakespeare Editions*, or, if not available there, from the Folger Shakespeare Library websites. The extract of the play *The Jew of Malta* was taken from the standardised database of Hugh Craig. (This database and its features are extensively described in Chapter Two in Section 2.4).

The texts were made of almost equal length--the first approximately 10,000 words--in order to avoid bias by length. As with the approach in our later, core experimentation, the texts were not chopped blindly at the threshold of 10,000 words but instead I set some constraints. The rationale and arguments for these constraints and the avoidance of automatic cuts at the threshold of 10,000 words have been described in detail in Chapter Two and the analysis of the building of corpora and pre-processing of texts of the core experimentation (Section 2.4). As you can see below, the average of the texts of the nine well-attributed Shakespearian plays is 10,068 words. The play *All's Well That Ends Well* was cut exactly at the end of an act and that's why it exceeds the approximate upper limit of 10,150 words.

<i>A Midsummer's Night</i>	10,006
<i>Antony and Cleopatra</i>	10,112
<i>As You Like It</i>	9,986
<i>Coriolanus</i>	10,071
<i>Cymbeline</i>	10,077
<i>Henry IV Part 1</i>	10,098

<i>King John</i>	10,101
<i>Richard II</i>	10,003
<i>The Winter's Tale</i>	10,159
Mean	10,069
Max	10,164
Min	9,986
Testing Stage:	
<i>All's Well That Ends Well</i>	10,186
<i>Henry VI Part 1</i>	10,053
<i>The Jew of Malta</i>	9,989

Table 62: Size of extracts in primary experimentation.

The texts of the 11 plays (apart the already regularised play of *The Jew of Malta*) were regularised in their spelling by the present investigator using the Vard 2 software developed at Lancaster University (Archer et al. 2003; Baron and Rayson 2008; *VARD 2*, n.d.).

The method of style analysis adopted here is to count the frequencies of certain words that the investigator has first, subjectively but based on semantic criteria, put into certain categories so that as well as counts for each word I can measure counts for each set of words. A total of 102 words were used and they formed six sets. In the core experimentation, as described in the section of Methodology (Section 2.4), only 100 words were selected, excluding the words 'brains' and 'moons' which anyway were judged to be of low stylistic information. Furthermore, in the core experimentation, in contrast to the approach adopted here, the separate sets Three, Five and Six were unified and constituted the Set Three of the Four Sets in order to avoid having fuzzy membership functions modelled on sets with counts less than 1%. So, I managed to have the same words--apart from two--as stylistic markers. The words used and the six sets formed are the following:

Set One if, then, else, here, more, less, now, night, nights, there, day, days, time, tonight

-Total: 14 words-types

Set Two you, me, her, thou, this, he, thee, him, she, we, they, them, us, these, none, those, himself, themselves, herself, nothing

-Total: 20 words-types

Set Three white, black, red, yellow, colour/color, colours/colors, odours/odors, speed, smell, touch, liquid, matter, strong, strength, weak, weakness

-Total: 16 words-types

Set Four with, what, when, which, where, who, within, whom, whose, wherein, since, things, people

-Total: 13 words-types

Set Five eye, eyes, eyed, ear, ears, face, faces, hand, hands, heart, hearts, brain, brains, throat, throats, bones, bone, man, wife, son, daughter, married, part, world

-Total: 24 words-types

Set Six life, death, shine, shines, fire, earth, sea, water, sun, moon, moons, bright, dark, past, present

-Total: 15 words-types

In authorship attribution, as already said, we may count various textual features at the same time, each forming a separate ‘input’ to our method of analysis. Words of Set One mainly signify space or time, such as ‘here’, ‘there’, ‘day’, ‘night’, ‘time’, Set Two is pronouns (personal, demonstrative and reflexive), Set Three is ‘physical properties’ including colour, like ‘black’ or ‘white’, but also words that relate to some of the human senses, such as ‘touch’, Set Four is relative pronouns, and Set Five is composed of hyponyms whether these denote body parts, such as ‘hand’, ‘hands’, ‘brain’, ‘face’, or members of a ‘family’, such as ‘son’, ‘daughter’, ‘wife’. Finally, Set Six includes words which indicate mainly environmental setting or Heraclitus’s elements, such as ‘water’, ‘sea’ and ‘earth’ or metaphorically ‘life’ and ‘death’. This last set, as stated (Section 2.4) is indicative of building, for reasons of stylometric experimentation, a set of non-function words with a philosophical connotation.

Overall, the main goal was to construct sets of words with different attributes, whether these concern grammatical categories, the singular or plural form, or the classification of the words’ as very frequent, low frequent, and rare (Burrows 2007). A basic distinction in stylometry is that of ‘tokens’ and ‘types’. Tokens relate to words counted once for each time they occur and the ‘types’ refer to the singular occurrence of distinct words.

Thus, in the phrase ‘to be or not to be’ there are six tokens in all but only four types--‘to’, ‘be’, ‘or’, and ‘not’--since ‘to’ and ‘be’ are repeated.

Each word type in the six sets occurs a number of times in each of our extracts of Shakespearian writing. For example, the word-type ‘then’ occurs 46 times in the extract from *A Midsummer Night’s Dream*. Each of these 46 occurrences constitutes a ‘token’ so that the type ‘then’ has a token count of 46 for that 10,000-token extract. As a general principle, the total token count for all the occurrences of all the types in the six sets did not exceed 17% of the token count of each extract (that is, it was under 1,700 tokens).

I count for each set the number of tokens in the extract that are types in that set, so for Set One I sum the counts of the occurrences of the tokens ‘then’, ‘if’, ‘here’, and so on, in an extract, and I express this as a percentage. So, for example, if I find in a 10,000-token extract that there are 200 tokens corresponding to the types in Set One, I record Set One contributing 2% of the tokens for that extract. If for another 10,000-token extract I find Set One contributing 300 tokens that is recorded as 3% for Set One for that extract.

A first glance at the results in next figure suggests that the Marlovian play is like the Shakespearian plays for the words in Set Two but not the words in sets Three and Four where the Marlowe scores are lower than all the Shakespeare scores. The Fuzzy-Logic approach to these data will allow us to make much subtler discriminations than can be made by the human eye simply comparing the heights of the bars, one with another.

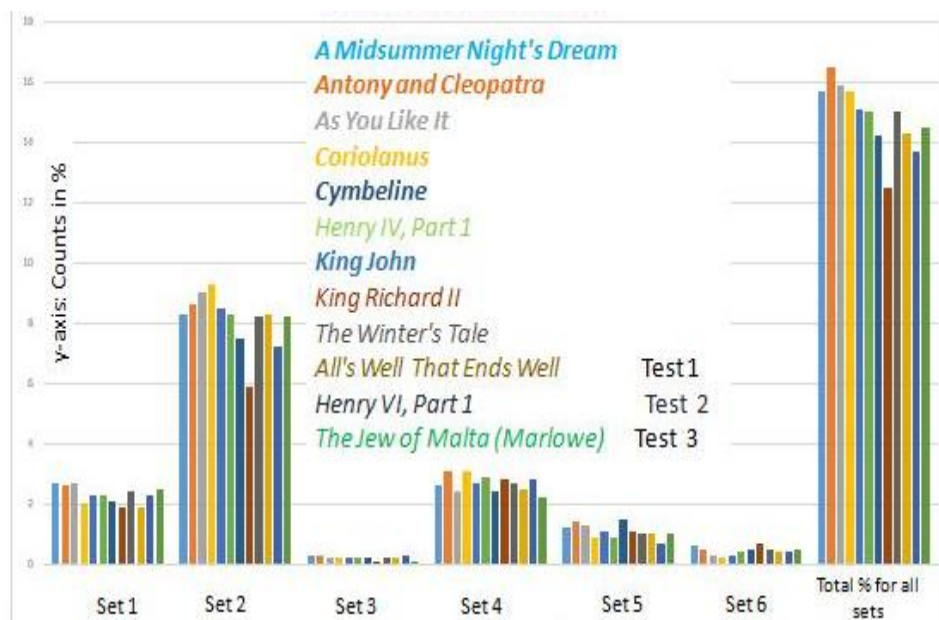


Figure 62: Twelve plays and six sets of words.

Play title (C for Comedy, T for Tragedy, H for History)	Set One %	Set Two %	Set Three %	Set Four %	Set Five %	Set Six %	Total of six Sets %
<i>A Midsummer Night's Dream</i> C	2.7	8.3	0.3	2.6	1.2	0.6	15.7
<i>Antony and Cleopatra</i> T	2.6	8.6	0.3	3.1	1.4	0.5	16.5
<i>As You Like It</i> C	2.7	9	0.2	2.4	1.3	0.3	15.9
<i>Coriolanus</i> T	2	9.3	0.2	3.1	0.9	0.2	15.7
<i>Cymbeline</i> T	2.3	8.5	0.2	2.7	1.1	0.3	15.1
<i>Henry IV, Part 1</i> H	2.3	8.3	0.2	2.9	0.9	0.4	15
<i>King John</i> H	2.1	7.5	0.2	2.4	1.5	0.5	14.2
<i>Richard II</i> H	1.9	5.9	0.1	2.8	1.1	0.7	12.5
<i>The Winter's Tale</i> C	2.4	8.2	0.2	2.7	1	0.5	15

Table 63: Data points of nine well-attributed Shakespearian plays (C= Comedies, T= Tragedies, H= Histories).

Play title	Set One %	Set Two %	Set Three %	Set Four %	Set Five %	Set Six %	Total of six Sets %
<i>All's Well That Ends Well</i>	1.9	8.3	0.2	2.5	1	0.4	14.3
<i>Henry VI Part 1</i>	2.3	7.2	0.3	2.8	0.7	0.4	13.7

<i>The Jew of Malta</i>	2.5	8.2	0.1	2.2	1	0.5	14.6
-------------------------	-----	-----	-----	-----	---	-----	------

Table 64: Data points of the three plays' six sets in the Testing Stage.

As said previously in Chapter Three, in the fuzzy fan controller example there were two variables and two membership classes, but for the application to Shakespeare's text there are six input variables, one for each of the six sets of words. In this experimentation, I employed only nine known plays, not the 15 used in the illustration of the method, and so each set had only nine counts of percentages. Applying the above algorithm to find triplets and duplets in the data and draw trapezoidal and triangular membership functions from them, the nine plays create membership functions for the six sets.

Set One's membership functions comprise three trapezoids, Set Two's two trapezoids and two triangles, Set Three's one trapezoid, and Sets Four, Five, and Six one triangle each. The scales of some of these six graphs are quite different because their data points are quite dissimilar, with each set containing words from low to very high frequencies, and therefore it is necessary to represent them focusing on the actual scale and producing a clear visual effect of the actual membership functions. When the algorithm is applied to Sets Three to Six, the resulting graph in each case contains only one membership-function class, which I will call the 'actual class' since it embodies all the data for that set.

Now that I have derived the class-membership functions for each set--comprising triangles and trapezoids--from the data points provided by the nine plays' occurrences of the words in each set from Set One to Set Six, I must derive the rules by which a new (previously untested) play's frequency of occurrence of the words in each set is judged to show that this new play is like or unlike Shakespeare's. These rules are derived systematically from the same data points that provided the class-membership functions.

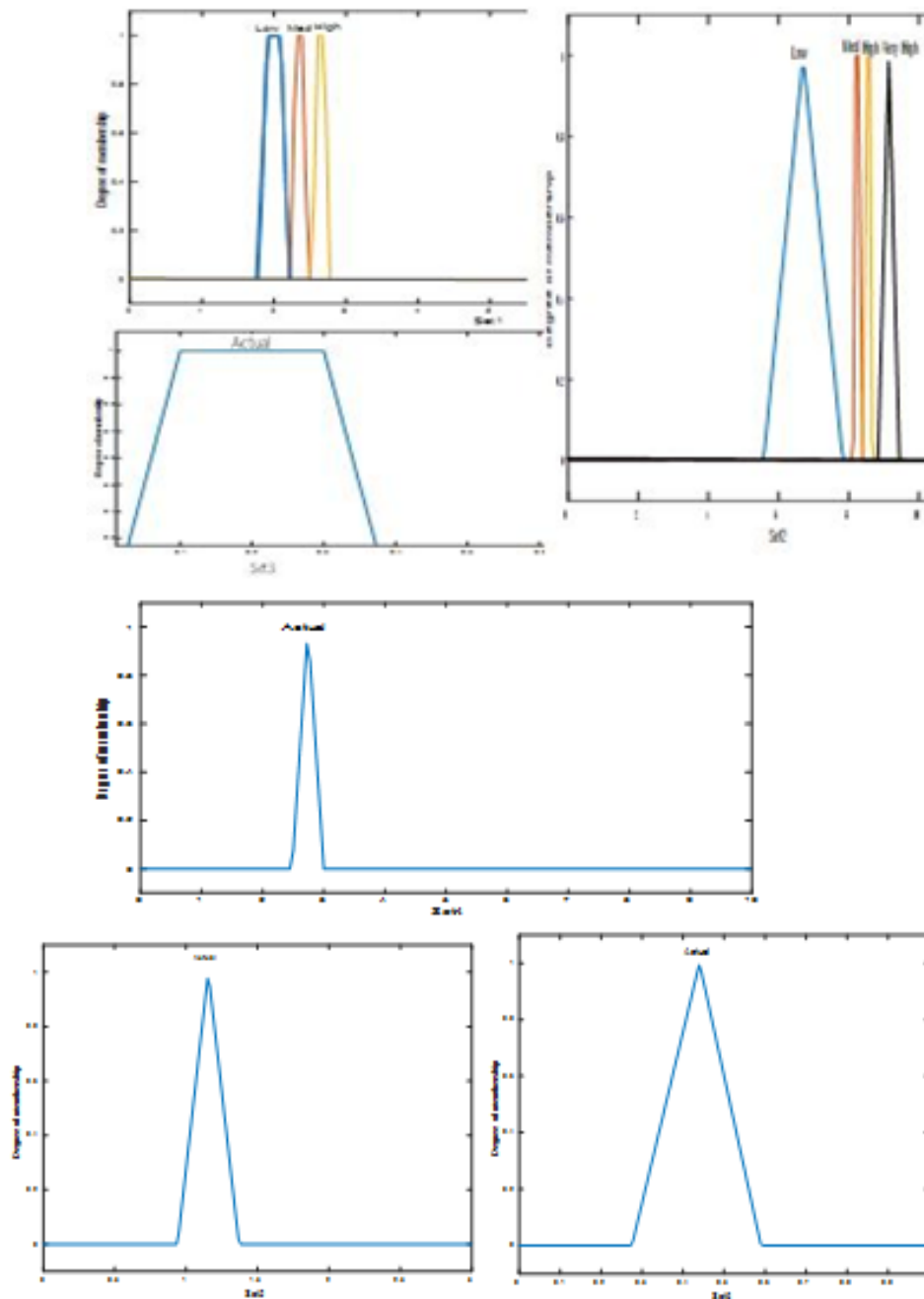


Figure 63: Membership functions of six sets.

Looking at the six pictures that show the class-membership functions for each of the six sets, we see that for some sets there is only one shape, the actual class, and for others there is more than one shape. These shapes reflect the actual data I got from the nine Shakespearian plays that form our ‘model’ for Shakespearian style, so any new play for which I get data points that, for every set, fall within one of the shapes (or the only shape) for that set is, by definition, a new play that matches perfectly (that is, falls entirely within) the

data derived from the nine Shakespearian plays. Such a play would, in relation to our model, be most highly Shakespearian. Indeed, if this new play's x -value for every set corresponds to a y -value of 1.0 in that set, then it would be entirely Shakespearian. There must be a distinction for new play's y -values being less than 1.0 (but more than 0.0) and all the possible ways that a new play might produce x -values that fall within the triangles and trapezoids across the six sets.

Each of the nine plays (across the X -axis of the matrix) produced an actual data point for each of the six sets (across the Y -axis of the matrix). Because the shape for each of Sets Three, Four, Five, and Six is a single trapezoid or isosceles triangle, we can be sure that the data for all nine plays is embodied in that single trapezoid or triangle, so the entire rows for these sets are coloured in a single colour showing that the play could not, as it were, 'choose' between shapes for those sets.

MND, Ant., AYL, Cor., CH, 1H4, Jn., R2, WT.

	<i>A Midsummer's Night=MND</i>	<i>Antony and Cleopatra=Ant.</i>	<i>As You Like It=AYL</i>	<i>Coriolanus=Cor.</i>	<i>Cymbeline=CH</i>	<i>Henry IV, Part I=1H4</i>	<i>King John=Jn.</i>	<i>King Richard the Second=R2</i>	<i>The Winter's Tale=WT</i>
S1	1c	1c	1c	1a	1b	1b	1a	1a	1a
S2	2b	2c	2d	2d	2c	2b	2a	2a	2b
S3	3	3	3	3	3	3	3	3	3
S4	4	4	4	4	4	4	4	4	4
S5	5	5	5	5	5	5	5	5	5
S6	6	6	6	6	6	6	6	6	6

Figure 64: Eight combinations of Shakespearian style.

For Set One and Two there are respectively three and four shapes and hence in the matrix the row for Set One has three background colours: a light-green background showing that *A Midsummer Night's Dream*, *Antony and Cleopatra*, and *As You Like It* contributed the data that formed formation the third trapezoid ('1c' or 'high' class), a blue background showing that the data of plays *Coriolanus*, *King John*, *Richard II*, *The Winter's Tale* contributed to the formation of the first trapezoid ('1a', the 'low' class), and a brown background showing that the data of plays *Cymbeline* and *Henry IV Part I* contributed to the formation of the middle trapezoid ('1b' or 'medium' class). For Set Two, the grey

background shows that the plays *A Midsummer Night's Dream*, *Henry IV Part 1*, and *The Winter's Tale* contributed the data that formed the left one of the two central trapezoids ('2b', 'medium' class), the white background shows that the plays *Antony and Cleopatra* and *Cymbeline* contributed to the right one of the two central trapezoids ('2c', 'high' class), the green background shows that the plays *As You Like It* and *Coriolanus* contributed to the right triangle ('2d', 'very high' class), and the light blue background shows that the plays *King John* and *Richard II* contributed the data that formed the left triangle ('2a', 'low' class). Notice that in the data that formed the model, not every possible combination of shapes is necessarily represented. Let us look at which shape each play contributed a data point to:

MND: 1c, 2b, 3, 4, 5, 6

Ant.: 1c, 2c, 3, 4, 5, 6

AYL: 1c, 2d, 3, 4, 5, 6

Cor.: 1a, 2d, 3, 4, 5, 6

Cym.: 1a, 2c, 3, 4, 5, 6

IH4: 1b, 2b, 3, 4, 5, 6

Jn.: 1a, 2a, 3, 4, 5, 6

R2: 1a, 2a, 3, 4, 5, 6

WT: 1a, 2b, 3, 4, 5, 6

As *King John* (*Jn*) and *Richard II* (*R2*) have the same combination, there are only eight possible combinations of the shapes of the six sets of nine plays: *A Midsummer Night's Dream*, *Antony and Cleopatra*, *As You Like It*, *Coriolanus*, *Cymbeline*, *Henry IV Part 1*, *King John*, *Richard II*, *The Winter's Tale*.

This table is derived by looking at each row of the matrix in turn and asking 'which shape did each play contribute a data point to?' For the first play, *A Midsummer Night's Dream*, the matrix shows that it contributed to shapes 3, 4, 5 and 6 because every play did. But where there is a choice of shapes within a set, as in Sets One and Two, we see from the matrix that it contributed to shape '1c' and '2b'. I repeat this operation for each play. At this point there are 12 theoretical ways to combine the shapes: the three possibilities for Set One times the four possibilities for Set Two times the single possibilities for each of Sets Three, Four, Five, and Six. But (for now) I ignore the theoretically possible combinations that were not in fact combinations 'selected' by the nine Shakespeare plays that I used to create the shapes. I derive our rules from the eight combinations (listed above) that our nine plays represented, and since I build one rule for each combination, there are eight rules. If a new

play that I test provides data that, for each Set, falls within the shapes for that set and in doing so ‘selects’ shapes to form one of these eight combinations (derived from nine known-to-be-Shakespeare plays) then we should say that it belongs to the membership class of highest Shakesperianness, having what I call the highest possible Shakespearian-Index-of-Similarity (SiS). That is what the eight rules must achieve: they must select the SiS1-High shape in the output membership function for any play that matches the known Shakespeare plays in this way.

Why do I use only the eight combinations of shapes that actual Shakespeare plays selected rather than all theoretically possible combinations? The main reason is that with real-world examples using only slightly larger datasets, using all theoretically possible combinations quickly produces a very large number of rules and this may not be needed. If I experimented with a larger number of plays, say a holistic model with the total of 27 well-attributed Shakespearian plays in the corpus of our main experimentation with each set having on average two shapes for each of the six Sets, I would have theoretically 64 combinations and so 64 rules for SiS1-High (because $2^6 = 64$). Therefore, a range of actual combinations of shapes selected by the 27 may well be only a quarter of this theoretical maximum, say 16 combinations giving a much more manageable and necessary 16 rules. Unless data points of new plays map new combinations of actual classes, the set of the rules is defined by the current eight actual SiS1-High combinations as detected in Shakespearian plays. So, from the above list of ways of combining the shapes I form eight rules:

R1 IF 1c AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R2 IF 1c AND 2c AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R3 IF 1c AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R4 IF 1a AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R5 IF 1a AND 2c AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R6 IF 1b AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R7 IF 1a AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R8 IF 1a AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

(In the actual model I could identify the shapes by slightly different names, such as low, medium, high, very high instead of a, b, c, d, but that need not concern us here: the principle is the same.) So much for the rules that select the output shape SiS1-High.

7.2.2 Rules and Inference Mechanism for Decision Making in the Stylometric Area.

First, it is necessary to finish making our rules. What about a new play that mostly matches the scores from our nine Shakespeare plays, but perhaps for just one of the six sets this new play's data point falls outside the shapes for that set? Such a new play is not quite in the class of most Shakespeare-like ('SiS1-High') but falls instead into a slightly lower class (say, 'SiS-Medium'). Because the rules must identify actual membership classes that new pieces of data are tested against (and cannot simply refer to the failure to match a given class), it is necessary to generate new 'not' or 'none' classes that represent the x -values outside of the range of x -values that fall within each class.

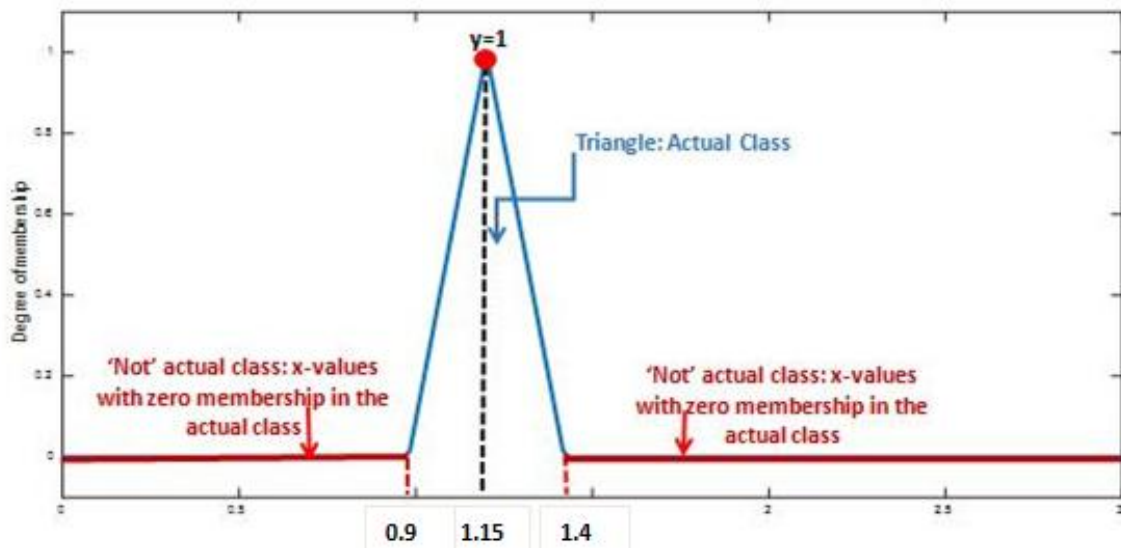


Figure 65: An example of 'Not' Class.

If, now, we look at the triangular shape of Set five in the figure above, the left leg of this triangle meets the X -axis at the point $x=0.9$ and the right leg meets the X -axis at the point $x=1.4$. (In Matlab script I round to the second decimal place and the exact values are 0.94 and 1.37 but I use here first decimal for purposes of illustration. As it stands, this adaptation does not affect the experimentation as the AntWord Profiler software counts the sets' counts to the first decimal place). This means that x values greater from 0.9 and less than 1.4 all have positive y values, with the maximum y value of 1 being reached by the x value of 1.15 (because $x=1.15, y=1.0$ is the apex of the triangle). The 'not' class that corresponds to this shape is the class of x -values from 0 to 0.9 and from 1.4 to infinity (see the red segments of

X-axis): these are all the x -values for which the corresponding y value is 0.0, indicating non-membership of the class that the triangle defines.

If we imagine a shape that is the negation of the shape for Set Five, it would have $y=1$ for those x -values (namely 0 to 0.9 and 1.4 to infinity) for which Set Five's shape has $y=0$. But what y -values should this 'not' shape have for those x -values between 0.9 and 1.4? It is reasonable to say that within this range the 'not' shape for Set Five's shape should take a y -value that is 1 minus the y -value for Set Five's shape. To see why this is so, imagine that 0-to-1 is the scale for an object's membership of the class 'heavy', with 0 corresponding to 'not at all heavy' and 1 corresponding to 'entirely heavy'. We can imagine a 'not heavy' class as the negation of this 'heavy' class, so that an object that is 'entirely heavy' has a membership of 0 this 'not heavy' class and an object that is 'not at all heavy' has a membership of 1 of this 'not heavy' class. An object that has a membership of 0.25 of the 'heavy' class is one-quarter on the way to being 'entirely heavy' and hence is three-quarters (0.75) of the way towards being 'not at all heavy'. Likewise, an object that has a membership of 0.66 of the 'heavy' class is, almost, only one-third (0.33) of the way towards being 'not at all heavy'. The degree to which something is 'not heavy' is 1 minus the degree to which it is 'heavy', and vice versa.

If we apply this principle to the shape for Set Five we can see its negation as an inverted triangle. When we apply this negation principle to sets that have multiple shapes (Set One and Two) our negation takes the name 'none' (of any of actual classes) instead of 'not' because we are aggregating categories of 'notness'. (Technically, as it is described parenthetically in the formation of rules, for Sets One and Two I had to design separate 'none' subclasses, see Section 7.3.7 of the Technical Appendix.) For Set One, and similarly for Set Two, our 'none' shape would have $y=1$ for x values that are not within (that is, correspond to $y=0$ for) the shape 'Lo' (or '1a'), and not within the shape 'Med' (or '1b'), and not within the shape 'Hi' (or '1c').

We saw above that eight rules depict the ideal combinations of Shakespearian patterns based on the nine known plays. Let us now say that I begin the experimentation with the plays of the testing corpus and I have only these eight SiS1-High rules. We must note that the fuzzy system produces a default value that is outputted whenever a new play's data points (derived from counting the words it has from Sets One through Six) trigger none of our rules. This default value (0.5) is the average value (0.5), since our values on X-axis range from 0 to 1. In fact, if none of the current rules is triggered, the index of 0.5 is produced in the system, meaning in a way 'neither one thing nor the other'. This is a useful technical feature of the

automated software and it tells us about a new play's classification and the system's inaction, meaning 'none of current rules is triggered' for the data points of a tested play. Of course, when there is not printed a default mean value, but one or more rules apply, then the centroid of the area is produced, and in the that case there is the complexity of differential calculus (see Section 7.3.4 of the Technical Appendix.)

The algorithm for the derivation of rules from the scope of constraints relies heavily on data points of Set Two. Of all the sets, Set Two has for each of the nine well-attributed Shakespearian plays the highest percentage of words' frequencies' counts, that is 8.1%. In fact, Set Two accounts for more than half of all six sets' data points. But for the production of an SiS1-High score, all the other data points of the five sets of the testing play have to be of non-zero membership in the actual class of the respective variables of our known model, as modelled by the nine well-attributed plays of Shakespeare. If the tested play is not detected as SiS1-High, the system must detect the play's 'none' or/and 'not' classes and measure how far the data of the sets of the tested play, as a single combination, are from the class of SiS1-High and the actual patterns of the nine well-attributed Shakespearian plays.

The idea, in rough terms, is that the detection of each 'none' or 'not' class of the six sets in the new testing plays causes the fall of one level from the SiS1-High class. Thus, if a new play triggers no 'none' or 'not' classes I select SiS1-High. If it triggers one 'none' or 'not' class I select SiS1-Medium2. If it triggers two 'none' or 'not' classes I select SiS1-Medium1. And lastly if the new testing play triggers three 'none' or 'not' classes I select SiS1-Low. Here is the algorithm that expresses these ideas, with Set Two treated as a special case because it accounts for some much of the data I am working with:

[START OF PROGRAM—INFERENCE MECHANISM IS READY TO ASSESS
INPUT VARIABLES]

Step 1

Does the six-sets data for the new play have, for each set's shapes, a non-zero membership?

If 'yes' then select SiS1-High and stop. If 'no', continue to Step 2.

Step 2

Does the Set Two data for the new play have a non-zero membership?

If 'no' then select SiS1-Low and stop. If 'yes' then continue to Step 3.

Step 3

Considering the five sets (1, 3, 4, 5, 6), does the new play's data fall within the 'not/none' class for only one of them?

If 'yes' select SiS1-Medium2 and stop. If 'no' continue to Step 4.

Step 4

Considering the five sets (1, 3, 4, 5, 6), does the new play's data fall within the 'not/none' class for two of them?

If 'yes' select SiS1-Medium1 and stop. If 'no' continue to Step 5.

Step 5

Considering the five sets (1, 3, 4, 5, 6), does the new play's data fall within the 'not/none' class for three of them, including Set Three (3) and/or Set Six (6). If 'yes' select SiS1-Medium1. If 'no' continue to Step 6.

Step 6

Select SiS1-Low and stop.

[END OF PROGRAM]

Notice that in Step 5 I have introduced a new criterion involving Sets Three and Six. Sets Three and Six contain words of low counts of frequencies. When the new testing play has only one or two 'not/none' classes, it falls one level from the SiS1-High level and each of the six Sets accounts equally for the SiS classification. Now, as the 'not/none' classes increase to three, I check if Sets Three and Six (the lowest-frequency sets) are included in those three and if they are I assign the play to the class of SiS-Medium1. In any other case with three (say Set One, Set Four and Set Five) or more 'not/none' sets, the new play falls into the lowest possible class, thus the SiS1-Low. That is, I 'reward' a play whose three 'not/none' matches include one or both of the two lowest-frequency sets, Sets Three and Six, by classifying it as slightly more Shakespearian (one class higher) than a play whose three 'not/none' matches include neither Set Three nor Set Six. In general, the six sets of this experimentation can be categorised into sets of words with very high (average 8.1% for Set Two), high (av: 2.4% and 2.7% respectively for Set one and Set Four), medium (av: 1.2% for Set Five) and low counts of frequencies (av: 0.2% and 0.4 respectively for Set Three and Set Six).

There are 320 ways to match the actual and the negative shapes for the six sets, because there are three actual and one negated shape for Set One (making 4), there are 4 actual and one negated for Set Two (making 5), and 1 actual and 1 negated for each of Sets Three, Four, Five, and Six. So that's $4 \times 5 \times 2 \times 2 \times 2 \times 2 = 320$ ways. (Technically, the automated

software's negated, 'none' class for Set One and Two is based on the void theorem of Set Theory, and it is interpreted simultaneously as none of actual and none of not actual classes. So, I had to design manually separate none trapezoidal subclasses for the areas not covered by the multiple actual classes in Set One and Two, see Section 7.3.7 of the Technical Appendix.)

I have developed eight SiS1-High rules so far that correspond to the matches of nine real Shakespearian plays:

R1 IF 1c AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R2 IF 1c AND 2c AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R3 IF 1c AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R4 IF 1a AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R5 IF 1a AND 2c AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R6 IF 1b AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R7 IF 1a AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
R8 IF 1a AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

Since Set One has three actual shapes and Set Two has four actual shapes, there are 12 ways for a play to match both an actual in Set One and an actual shape in Set Two, and I have so far (in the above eight rules) only accounted for eight of them. So that leaves four more possible ways to match the actual shapes for all six sets. These four are:

IF 1c AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
IF 1c AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
IF 1b AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-High
IF 1b AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

Let us put these four rules aside as a block of rules that I call 'hypothetical matches'. Because until now none of the real-world plays matched the antecedents of these rules, I can discount this block of four and say that of the possible 320 combinations, eight correspond to SiS1-High classification of real-world plays and four correspond to hypothetical SiS1-High matches that I have (yet) not seen in real-world plays. That leaves 308 possible combinations for which I should produce more rules. Next, I run three new plays--*All's Well That Ends Well*, a comedy, *Henry VI, Part 1*, a history, and the well-attributed, sole-authored Marlovian

play *The Jew of Malta*--through the rule-generating algorithm. One of them, *All's Well that Ends Well*, triggers an existing rule and has as its consequence the SiS1-High classification, so that play generates no new rules. The other two plays trigger none of the above 12 rules, so it is necessary now to generate new rules that correspond to their matches to the shapes for the six sets. Using the algorithm for rule generation, these two plays generate two new rules:

R9 IF 1b AND 2a AND 3 AND 4 AND 5-NOT AND 6 THEN SiS1-Medium2
 R10 IF 1none AND 2b AND 3 AND 4-NOT AND 5 AND 6
 THEN SiS1-Medium1

These two plays in terms of classification/Shakespearianness are assigned the output classes of SiS1-Medium2 and SiS1-Medium1 respectively, on account of the first play having one 'not' match and the second play having two 'not/none' matches. That leaves 306 possible combination of shape-matching that I have not yet considered. These combinations form the complete pool of the total of rules' antecedents. By applying the algorithm, I then form the consequent part of the rules.

For each rule it is necessary to define the states of all six sets, requiring a match on all six to trigger the execution of that rule. Fuzzy Logic systems work by iterative pattern matching, comparing the incoming data against each rule in turn. It is desirable that each rule is specified in terms of all six sets--that is, I give explicit conditions for each set--so that new rules can be added to the collection of rules without the investigator needing to worry about new rules inadvertently causing spurious matches against previous, incompletely defined, rules.

Rule 9 is triggered when Set Two is non-zero and corresponds to the algorithm, as described in the pseudocode, not stopping in Step 2 but rather going on to Step 3. Rule 10 corresponds to the algorithm stopping in Step 4 because a 'not/none' class is matched for two of the sets from this list: Set One, Set Three, Set Four, Set Five, and Set Six. I will add four new rules (R11-R14). The new rules (R11-14) will arise from the application of the pseudocode algorithm and the selection of some of the shapes of new plays' data, apart from those three of our testing corpus, by applying a heuristics-based search of the remaining 306 theoretical combinations.

Indeed, there is technically the option of gradually adding new rules to our system and building a larger and more representative model. But instead of defining all the possible theoretical combinations from the beginning, something that requires a high processing cost,

in practice I follow a case by case strategy and exploit a transition stage, in which the fuzzy program of the model of the known corpus prints out a default medium value for a new play, if this new play is not SiS1-High and does not trigger any of the existing, six in our case, non-SiS high rules. Whenever a new play triggers none of our rules, the Fuzzy Logic system outputs a default value as its classification (a value of 0.5 in our system), and whenever I get this default value I can consider whether to add a new rule to cover the particular combination of classes that this new play matches. I then proceed to the adoption of our algorithm considering the ‘not/none’ classes of the new play, as I have done with Rules 9 and 10.

In order to see how the model of the corpus of nine well-attributed Shakespearian plays behaves in the real world, as with other plays of the early modern period, and to test the model’s plausibility as a classifier of degree of Shakespearianness, I formed a new corpus with three plays: *All’s Well That Ends Well*, a comedy, *Henry VI Part 1*, a history play, and the well-attributed, sole-authored Marlovian play *The Jew of Malta*. The first of these three plays triggered one of the initial eight SiS1-High rules and for the next two generated respectively Rule 9 and 10. I employed the corpus of the new three plays under scrutiny in order to assess the model’s results in connection with the general evidence of historical scholarships and the claim that *All’s Well That Ends Well* is almost entirely a sole-authored Shakespearian play, whereas *Henry VI Part 1* is probably a play of at least two authors, one of them being Shakespeare. *The Jew of Malta* is of widely agreed to be entirely by Marlowe (Craig and Kinney 2009, 215).

Two of the three plays of our testing corpus, *Henry VI Part 1* and *The Jew of Malta*, are non-highly Shakespearian, they do not trigger any of the eight SiS1-High rules. Because one of the new testing plays is highly Shakespearian and the two others are non-highly Shakespearian (non-SiS1-High), it might seem that that it is necessary to form only ten rules, the eight SiS1-High rules and two non-SiS1-High rules for the two non-highly Shakespearian plays. Instead of adding only two rules to the initial pool of eight SiS1-High rules, I decided to select a more general strategy and added four more non-SiS1-High rules from the total of all possible theoretical combinations (306) in order to show the logic of a full model. Let us discuss how I arrived at these four new rules.

How did I select four rules from the large pool of 306 non-SiS1-High possible theoretical combinations that play the role of rules’ antecedents? Which rules should I select each time when aiming at updating and enlarging the database of rules? There are 306 combinations, which in mathematics are called (non-repetitive) ordered permutations.

Now I will deal with the process of enlarging the database of rules and it is possible to express the possible combinations of data for the total of six sets as a tree structure in which the root node has six branches (technically known as ‘edges’) emerging from it, each one of which represents one of the following six possibilities:

- 1) That for one set the matched shape is ‘not/none’
- 2) That for two sets the matched shapes are ‘not/none’
- 3) That for three sets the matched shapes are ‘not/none’
- 4) That for four sets the matched shapes are ‘not/none’
- 5) That for five sets the matched shapes are ‘not/none’
- 6) That for all six sets the matched shapes are ‘not/none’

That list can be expressed pictorially, with ‘O’ as the symbol for the root node, thus:

```
O
+--For one set the matched shape is ‘not/none’
+--For two sets the matched shapes are ‘not/none’
+--For three sets the matched shapes are ‘not/none’
+--For four sets the matched shapes are ‘not/none’
+--For five sets the matched shapes are ‘not/none’
+--For all six sets the matched shapes are ‘not/none’
```

Consider the first case, in which for just one set the matched shape is ‘not/none’. How many ways can that happen? Or, in other words, in our tree how many edges will emerge from this node? The answer is seven, because there are six ways in which the data for a play match the ‘not/none’ shape for just one set:

```
O
+--For one set the matched shape is ‘not/none’
| +--For Set One the play matches the ‘not/none’ shape
| +--For Set Two the play matches the ‘not/none’ shape
| +--For Set Three the play matches the ‘not/none’ shape
| +--For Set Four the play matches the ‘not/none’ shape
| +--For Set Five the play matches the ‘not/none’ shape
| +--For Set Six the play matches the ‘not/none’ shape
|
+--For two sets the matched shapes are ‘not/none’
+--For three sets the matched shapes are ‘not/none’
+--For four sets the matched shapes are ‘not/none’
+--For five sets the matched shapes are ‘not/none’
```

+--For all six sets the matched shapes are 'not/none'

What about the second main possibility, that for two sets the matched shapes are 'not/none'? How many ways are there for this to happen, or in other words how many edges emerge from this node? It is possible to break this down by asking first what are the possibilities if Set One is one of the two sets for which the matched shapes are 'not/none'?

The other set of the two can be only Set Two, Set Three, Set Four, Set Five, or Set Six. Pictorially, that is:

```
O
+--For one set the matched shape is 'not/none'
| +--For Set One the play matches the 'not/none' shape
| +--For Set Three the play matches the 'not/none' shape
| +--For Set Four the play matches the 'not/none' shape
| +--For Set Five the play matches the 'not/none' shape
| +--For Set Six the play matches the 'not/none' shape
|
+--For two sets the matched shapes are 'not/none'
| +--For Set One and Set Two the play matches the 'not/none' shape
| +--For Set One and Set Three the play matches the 'not/none' shape
| +--For Set One and Set Four the play matches the 'not/none' shape
| +--For Set One and Set Five the play matches the 'not/none' shape
| +--For Set One and Set Six the play matches the 'not/none' shape
| +...
|
+--For three sets the matched shapes are 'not/none'
+--For four sets the matched shapes are 'not/none'
+--For five sets the matched shapes are 'not/none'
+--For all six sets the matched shapes are 'not/none'
```

By abbreviating the descriptions of the nodes and the sets 'Set One' becomes 'S1' in order to produce this tree:

```
O
+--One set matched as 'not/none'
| +--S1 matched as 'not/none'
| +--S2 matched as 'not/none'
| +--S3 matched as 'not/none'
| +--S4 matched as 'not/none'
| +--S5 matched as 'not/none'
| +--S6 matched as 'not/none'
|
+--Two sets matched as 'not/none'
| +--S1+S2 matched as 'not/none'
```

```

| +--S1+S3 matched as 'not/none'
| +--S1+S4 matched as 'not/none'
| +--S1+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'
| +...
|
+--Three sets matched as 'not/none'
+--Four sets matched as 'not/none'
+--Five sets matched as 'not/none'

```

What if Set One is not the first of the two sets to match as 'not/none' but instead Set Two is?

That gives us:

```

O
+--One set matched as 'not/none'
| +--S1 matched as 'not/none'
| +--S2 matched as 'not/none'
| +--S3 matched as 'not/none'
| +--S4 matched as 'not/none'
| +--S5 matched as 'not/none'
| +--S6 matched as 'not/none'
|
+--Two sets matched as 'not/none'
| +--S1+S2 matched as 'not/none'
| +--S1+S3 matched as 'not/none'
| +--S1+S4 matched as 'not/none'
| +--S1+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'
| +--S2+S3 matched as 'not/none'
| +--S2+S4 matched as 'not/none'
| +--S2+S5 matched as 'not/none'
| +--S2+S6 matched as 'not/none'
| +--S3+S4 matched as 'not/none'
| +--S3+S5 matched as 'not/none'
| +--S3+S6 matched as 'not/none'
| +...
|
+--Three sets matched as 'not/none'
+--Four sets matched as 'not/none'
+--Five sets matched as 'not/none'

```

Notice that it is not necessary to record as a separate node the combination of Set Two and Set One because that was already recorded earlier in the tree as the combination of Set One and Set Two. When I turn to the possibility of Set Three being the first of the two sets that matches I do not have to go back and consider the combination of Set Three and Set One or

Set Three and Set Two because these were already recorded as combinations earlier in the tree, so the tree gains just two new nodes:

```

O
+--One set matched as 'not/none'
| +--S1 matched as 'not/none'
| +--S2 matched as 'not/none'
| +--S3 matched as 'not/none'
| +--S4 matched as 'not/none'
| +--S5 matched as 'not/none'
| +--S6 matched as 'not/none'
|
+--Two sets matched as 'not/none'
| +--S1+S2 matched as 'not/none'
| +--S1+S3 matched as 'not/none'
| +--S1+S4 matched as 'not/none'
| +--S1+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'
| +--S2+S3 matched as 'not/none'
| +--S2+S4 matched as 'not/none'
| +--S2+S5 matched as 'not/none'
| +--S2+S6 matched as 'not/none'
| +--S3+S4 matched as 'not/none'
| +--S3+S5 matched as 'not/none'
| +--S3+S6 matched as 'not/none'
| +--S4+S5 matched as 'not/none'
| +--S4+S6 matched as 'not/none'
| +...
|
+--Three sets matched as 'not/none'
+--Four sets matched as 'not/none'
+--Five sets matched as 'not/none'

```

There is only one more possible combination involving two sets matching as 'not/none': Set Five and Set Six are the matching pair. So, that makes the full tree for the two-sets-match scenario this:

```

O
+--One set matched as 'not/none'
| +--S1 matched as 'not/none'
| +--S3 matched as 'not/none'
| +--S4 matched as 'not/none'
| +--S5 matched as 'not/none'
| +--S6 matched as 'not/none'
|
+--Two sets matched as 'not/none'

```

- | +--S1+S2 matched as 'not/none'
- | +--S1+S3 matched as 'not/none'
- | +--S1+S4 matched as 'not/none'
- | +--S1+S5 matched as 'not/none'
- | +--S1+S6 matched as 'not/none'
- | +--S2+S3 matched as 'not/none'
- | +--S2+S4 matched as 'not/none'
- | +--S2+S5 matched as 'not/none'
- | +--S2+S6 matched as 'not/none'
- | +--S3+S4 matched as 'not/none'
- | +--S3+S5 matched as 'not/none'
- | +--S3+S6 matched as 'not/none'
- | +--S4+S5 matched as 'not/none'
- | +--S4+S6 matched as 'not/none'
- | +--S5+S6 matched as 'not/none'
- |
- +--Three sets matched as 'not/none'
- +--Four sets matched as 'not/none'
- +--Five sets matched as 'not/none'

Let us consider the scenarios in which three sets match as 'not/none'. Here is the full list of possibilities:

O

- +--One set matched as 'not/none'
- | +--S1 matched as 'not/none'
- | +--S2 matched as 'not/none'
- | +--S3 matched as 'not/none'
- | +--S4 matched as 'not/none'
- | +--S5 matched as 'not/none'
- | +--S6 matched as 'not/none'
- |
- +--Two sets matched as 'not/none'
- | +--S1+S2 matched as 'not/none'
- | +--S1+S3 matched as 'not/none'
- | +--S1+S4 matched as 'not/none'
- | +--S1+S5 matched as 'not/none'
- | +--S1+S6 matched as 'not/none'
- | +--S2+S3 matched as 'not/none'
- | +--S2+S4 matched as 'not/none'
- | +--S2+S5 matched as 'not/none'
- | +--S1+S6 matched as 'not/none'
- | +--S3+S4 matched as 'not/none'
- | +--S3+S5 matched as 'not/none'
- | +--S3+S6 matched as 'not/none'
- | +--S4+S5 matched as 'not/none'
- | +--S4+S6 matched as 'not/none'
- | +--S5+S6 matched as 'not/none'


```

|
+--Three sets matched as 'not/none'
| +--S1+S2+S3 matched as 'not/none'
| +--S1+S2+S4 matched as 'not/none'
| +--S1+S2+S5 matched as 'not/none'
| +--S1+S2+S6 matched as 'not/none'
| +--S1+S3+S4 matched as 'not/none'
| +--S1+S3+S5 matched as 'not/none'
| +--S1+S3+S6 matched as 'not/none'
| +--S1+S4+S5 matched as 'not/none'
| +--S1+S4+S6 matched as 'not/none'
| +--S1+S5+S6 matched as 'not/none'
| +--S2+S3+S4 matched as 'not/none'
| +--S2+S4+S5 matched as 'not/none'
| +--S2+S5+S6 matched as 'not/none'
| +--S3+S4+S5 matched as 'not/none'
| +--S3+S4+S6 matches as 'not/none'
| +--S3+S5+S6 matched as 'not/none'
| +--S4+S5+S6 matched as 'not/none'
|
+--Four sets matched as 'not/none'
+--Five sets matched as 'not/none'

```

Let us fill in the four-sets-matched possibilities:

```

O
+--One set matched as 'not/none'
| +--S1 matched as 'not/none'
| +--S2 matched as 'not/none'
| +--S3 matched as 'not/none'
| +--S4 matched as 'not/none'
| +--S5 matched as 'not/none'
| +--S6 matched as 'not/none'
|
+--Two sets matched as 'not/none'
| +--S1+S3 matched as 'not/none'
| +--S1+S4 matched as 'not/none'
| +--S1+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'
| +--S2+S3 matched as 'not/none'
| +--S2+S4 matched as 'not/none'
| +--S2+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'
| +--S3+S4 matched as 'not/none'
| +--S3+S5 matched as 'not/none'
| +--S3+S6 matched as 'not/none'
| +--S4+S5 matched as 'not/none'
| +--S4+S6 matched as 'not/none'
| +--S5+S6 matched as 'not/none'
|

```

```

+--Three sets matched as 'not/none'
| +--S1+S2+S3 matched as 'not/none'
| +--S1+S2+S4 matched as 'not/none'
| +--S1+S2+S5 matched as 'not/none'
| +--S1+S2+S6 matched as 'not/none'
| +--S1+S3+S4 matched as 'not/none'
| +--S1+S3+S5 matched as 'not/none'
| +--S1+S3+S6 matched as 'not/none'
| +--S1+S4+S5 matched as 'not/none'
| +--S1+S4+S6 matched as 'not/none'
| +--S1+S5+S6 matched as 'not/none'
| +--S2+S3+S4 matched as 'not/none'
| +--S2+S4+S5 matched as 'not/none'
| +--S2+S5+S6 matched as 'not/none'
| +--S3+S4+S5 matched as 'not/none'
| +--S3+S4+S6 matches as 'not/none'
| +--S3+S5+S6 matched as 'not/none'
| +--S4+S5+S6 matched as 'not/none'
|
+--Four sets matched as 'not/none'

| +--S1+S2+S3+S4 matched as 'not/none'
| +--S1+S2+S4+S5 matched as 'not/none'
| +--S1+S2+S4+S6 matched as 'not/none'
| +--S1+S3+S4+S5 matched as 'not/none'
| +--S1+S4+S5+S6 matched as 'not/none'
| +--S2+S3+S4+S5 matched as 'not/none'
| +--S2+S3+S4+S6 matched as 'not/none'
| +--S3+S4+S5+S6 matched as 'not/none'
|
+--Five sets matched as 'not/none'

```

Let us fill in the five-sets-matched possibilities:

```

O
+--One set matched as 'not/none'
| +--S1 matched as 'not/none'
| +--S2 matched as 'not/none'
| +--S3 matched as 'not/none'
| +--S4 matched as 'not/none'
| +--S5 matched as 'not/none'
| +--S6 matched as 'not/none'
|
+--Two sets matched as 'not/none'
| +--S1+S2 matched as 'not/none'
| +--S1+S3 matched as 'not/none'
| +--S1+S4 matched as 'not/none'
| +--S1+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'

```

```

| +--S2+S3 matched as 'not/none'
| +--S2+S4 matched as 'not/none'
| +--S2+S5 matched as 'not/none'
| +--S1+S6 matched as 'not/none'
| +--S3+S4 matched as 'not/none'
| +--S3+S5 matched as 'not/none'
| +--S3+S6 matched as 'not/none'
| +--S4+S5 matched as 'not/none'
| +--S4+S6 matched as 'not/none'
| +--S5+S6 matched as 'not/none'
|
+--Three sets matched as 'not/none'
| +--S1+S2+S3 matched as 'not/none'
| +--S1+S2+S4 matched as 'not/none'
| +--S1+S2+S5 matched as 'not/none'
| +--S1+S2+S6 matched as 'not/none'
| +--S1+S3+S4 matched as 'not/none'
| +--S1+S3+S5 matched as 'not/none'
| +--S1+S3+S6 matched as 'not/none'
| +--S1+S4+S5 matched as 'not/none'
| +--S1+S4+S6 matched as 'not/none'
| +--S1+S5+S6 matched as 'not/none'
| +--S2+S3+S4 matched as 'not/none'
| +--S2+S4+S5 matched as 'not/none'
| +--S2+S5+S6 matched as 'not/none'
| +--S3+S4+S5 matched as 'not/none'
| +--S3+S4+S6 matches as 'not/none'
| +--S3+S5+S6 matched as 'not/none'
| +--S4+S5+S6 matched as 'not/none'
|
+--Four sets matched as 'not/none'

| +--S1+S2+S3+S4 matched as 'not/none'
| +--S1+S2+S4+S5 matched as 'not/none'
| +--S1+S2+S5+S6 matched as 'not/none'
| +--S1+S3+S4+S5 matched as 'not/none'
| +--S1+S3+S4+S6 matched as 'not/none'
| +--S1+S3+S5+S6 matched as 'not/none'
| +--S2+S3+S4+S5 matched as 'not/none'
| +--S2+S3+S5+S6 matched as 'not/none'
| +--S3+S4+S5+S6 matched as 'not/none'
|
+--Five sets matched as 'not/none'
| +--S1+S2+S3+S4+S5 matched as 'not/none'
| +--S2+S3+S4+S5+S6 matched as 'not/none'

```

And finally, there is just one way for all six sets to match as 'not/none':

O

+--One set matched as 'not/none' = 6 ways
 | +--S1 matched as 'not/none'
 | +--S2 matched as 'not/none'
 | +--S3 matched as 'not/none'
 | +--S4 matched as 'not/none'
 | +--S5 matched as 'not/none'
 | +--S6 matched as 'not/none'
 |
 +--Two sets matched as 'not/none' = 15 ways
 | +--S1+S2 matched as 'not/none'
 | +--S1+S3 matched as 'not/none'
 | +--S1+S4 matched as 'not/none'
 | +--S1+S5 matched as 'not/none'
 | +--S1+S6 matched as 'not/none'
 | +--S2+S3 matched as 'not/none'
 | +--S2+S4 matched as 'not/none'
 | +--S2+S5 matched as 'not/none'
 | +--S2+S6 matched as 'not/none'
 | +--S3+S4 matched as 'not/none'
 | +--S3+S5 matched as 'not/none'
 | +--S3+S6 matched as 'not/none'
 | +--S4+S5 matched as 'not/none'
 | +--S4+S6 matched as 'not/none'
 | +--S5+S6 matched as 'not/none'
 |
 +--Three sets matched as 'not/none' = 17 ways
 | +--S1+S2+S3 matched as 'not/none'
 | +--S1+S2+S4 matched as 'not/none'
 | +--S1+S2+S5 matched as 'not/none'
 | +--S1+S2+S6 matched as 'not/none'
 | +--S1+S3+S4 matched as 'not/none'
 | +--S1+S3+S5 matched as 'not/none'
 | +--S1+S3+S6 matched as 'not/none'
 | +--S1+S4+S5 matched as 'not/none'
 | +--S1+S4+S6 matched as 'not/none'
 | +--S1+S5+S6 matched as 'not/none'
 | +--S2+S3+S4 matched as 'not/none'
 | +--S2+S4+S5 matched as 'not/none'
 | +--S2+S5+S6 matched as 'not/none'
 | +--S3+S4+S5 matched as 'not/none'
 | +--S3+S4+S6 matches as 'not/none'
 | +--S3+S5+S6 matched as 'not/none'
 | +--S4+S5+S6 matched as 'not/none'
 |
 +--Four sets matched as 'not/none' = 8 ways
 | +--S1+S2+S3+S4 matched as 'not/none'
 | +--S1+S2+S4+S5 matched as 'not/none'
 | +--S1+S2+S4+S6 matched as 'not/none'
 | +--S1+S3+S4+S5 matched as 'not/none'

```

| +--S1+S4+S5+S6 matched as 'not/none'
| +--S2+S3+S4+S5 matched as 'not/none'
| +--S2+S3+S4+S6 matched as 'not/none'
| +--S3+S4+S5+S6 matched as 'not/none'
|
| +--Five sets matched as 'not/none' = 2 ways
| +--S1+S2+S3+S4+S5 matched as 'not/none'
| +--S2+S3+S4+S5+S6 matched as 'not/none'
|
| +--Six sets matched as 'not/none' = 1 way
+--S1+S2+S3+S4+S5+S6 matched as 'not/none'

```

The above tree shows the 49 ways in which a play's data may match the 'not/none' shape for one, two, three, four, five, or all six sets. How do I use this tree to add new rules to our model? Observe that the tree does not cover the possibility of a play matching the 'not/none' shape for none of the sets. But that possibility is already covered by the 12 rules I have already devised in which an 'actual' rather than a negated shape was matched for each of the six sets; eight (R1-R8) of those 12 rules arose from the real-world data (from nine plays) and four additional rules I decided were hypothetical possibilities and set them aside. So, none of these 12 rules could emerge from this 49-way tree in any case, since the tree covers only the various ways that one or more 'not/none' shapes may be matched.

I developed further two rules (R9, R10) using three further real-world plays that did indeed match 'not/none' shapes for certain sets, and these two rules represent combinations that may indeed come up again as I use the 49-way tree to consider all the remaining 308 possible ways of matching shapes. Of the 308 ways, I want to consider the 306--that is, 320 minus eight for R1 to R8, minus four hypothetical, minus R9 and R10--for which no rule has yet been mapped. The process for using the 49-way tree involves traversing this tree systematically to select leaf nodes representing possible ways to match one or more 'not/none' shapes and then expanding those leaf nodes to explicitly list all the combinations that they represent.

Let us start with the means by which I traverse the tree. I start at the tree's origin (the node labelled 'O' above) that we can say is at Level 0, because it is a distance of zero steps from the origin. I descend to the first node Level 1, which is the node labelled 'One set matched as 'not/none' = 6 ways'. I select the first two Level 2 children of this node (the first two of the six 'ways') and they are 'S1 matched as 'not/none'' and 'S2 matched as 'not/none''. Then I ascend the tree to consider the second node at Level 1, which is 'Two sets matched as 'not/none' = 15 ways'. I select the first two Level 2 children of this node, and

they are ‘S1+S2 matched as ‘not/none’ and ‘S1+S3 matched as ‘not/none’’. I then ascend the tree again and do the same for the remaining four Level 1 nodes, selecting the first two of their children in each case, giving 11 selections in all. (I select two of the Level 2 children for each of the first five Level 1 nodes, but the last Level 1 node, ‘Six sets matched as ‘not/none’ = 1 way’, has just one ‘way’ I can select.)

Each of these selections represents a collection of ways in which the criterion it embodies can be met. Bearing in mind that Set One has three actual shapes and one ‘not/none’ shape, and that Set Two has four actual shapes and one ‘not/none’ (with Sets Three to Six have one actual and one ‘not/none’ each), the first selection ‘S1 matched as ‘not/none’ can be met in four possible ways:

1n 2a 3 4 5 6

1n 2b 3 4 5 6

1n 2c 3 4 5 6

1n 2d 3 4 5 6

(Here ‘1n’ represents that the ‘not/none’ shape is matched for Set One, and ‘2a’ to ‘2d’ represents matching each of the four actual shapes for Set Two.) From these four possible ways that a play might meet the criterion ‘S1 matched as ‘not/none’ I select just the first and make that into a new rule:

R11 IF 1NONE AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-Medium2

Then I move to the second one of the selection of 11 leaves from the 49-way tree, which was ‘S2 matched as ‘not/none’. There are just three ways for this to happen:

1a 2n 3 4 5 6

1b 2n 3 4 5 6

1c 2n 3 4 5 6

Again I take just the first of these, and turn it into the rule:

R12 IF 1a AND 2NONE AND 3 AND 4 AND 5 AND 6 THEN SiS1-Low

And so on for the remaining nine of the 11 leaves selected from the 49-way tree, until I have created as many new rules as I want to add. Using the two-from-each-Level-2 principle, I selected 11 leaves, but of course I could have chosen more than two each time; this ‘two’ is just a changeable variable in the tree-traversing process. Another variable is how far down (counted in levels) I descend before I start harvesting nodes, and yet another variable is how many rules I generate before I stop the process.

For the purposes of this explanation, let us say that I run the tree-traversing procedure until it has generated four new rules. That gives this set of 14 rules in all:

R1 IF 1c AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R2 IF 1c AND 2c AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R3 IF 1c AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R4 IF 1a AND 2d AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R5 IF 1a AND 2c AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R6 IF 1b AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R7 IF 1a AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R8 IF 1a AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

R9 IF 1b AND 2a AND 3 AND 4 AND 5-NOT AND 6 THEN SiS1-Medium2

R10 IF 1NONE AND 2b AND 3 AND 4-NOT AND 5 AND 6

THEN SiS1-Medium1

R11 IF 1NONE AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-Medium2

R12 IF 1a AND 2NONE AND 3 AND 4 AND 5 AND 6 THEN SiS1-Low

R13 IF 1NONE AND 2NONE AND 3 AND 4 AND 5 AND 6 THEN SiS1-LOW

R14 IF 1NONE AND 2a AND 3NOT AND 4 AND 5 AND 6

THEN SiS1-Medium1

In relation with a previous parenthetical remark about the fact that I had to form technically separate ‘none’ trapezoidal classes for the areas not covered by the multiple actual classes in Set One and Two, see Section 7.3.7 of the Technical Appendix.

7.2.3 Classes of Output Variable and Experimentation with the Three Plays of the Testing Corpus.

In order to understand how the inference mechanism functions, it is necessary to describe also the classes of the output variable, what I have called Shakespearian-Index-of-Similarity (SiS). In the experimentation with the fuzzy fan controller, the output variable was the speed of the fan. In the stylometric experimentation, the output variable expresses the degree of stylistic Shakespearian Similarity (SiS) of the play to be tested and it is measured in the range of 0.00 to 1. The output variable of the fuzzy model, as shown in the figure below, is represented by four trapezoids, the classes of the Low, Medium1, Medium2 and High index of Shakespearian Similarity (SIS), as also described in Section 3.3.

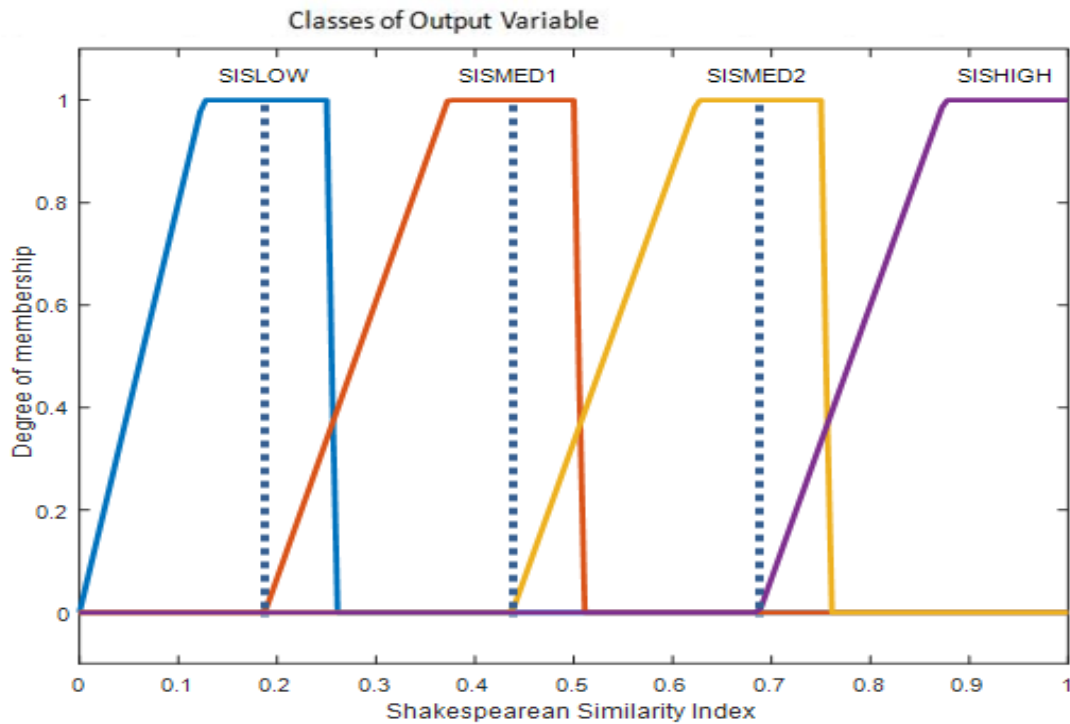


Figure 66: SiS and four classes of output variable.

7.2.4 Results

The measurement-index of the Shakespearianness of each of the three tested plays is derived with a specific process similarly to the way the optimal fan's speed was produced based on the truncation of one or more of the classes of the output variable. First, I input to the fuzzy system the values-counts of the data points (percentages) of the six sets of the play

that is tested for its degree of Shakespearianness (SiS). In the actual experimentation, only one of the fourteen rules fires for each of the three plays under scrutiny, and each rule specifies which one of the four output classes (SiS1-Low, SiS1-Medium1, SiS1-Medium2, or SiS1-High) is used to classify the play. At this point, I have to clarify that for a single play more than one rule could be triggered, if I designed the rules in a different way or in case there were large overlapped areas between the classes of the first and second set-variable. But with the current design of this experimentation and the database of the 14 rules it is not technically possible to have more than one rule activated because the rules are mutually exclusive. As seen above in the first part of the explanation of the scenarios of the mechanism of fuzzy inference, we say that a rule fires if its antecedents are met. Of the six sets-inputs, as we saw also with the fuzzy fan controller, only one, the input with minimum membership, is selected. By 'selected' I mean that it is possible to draw a horizontal line on the *Y*-axis of the output variable's SiS graph starting from this point of the specific input's minimum membership and use this line to chop the top off the trapezoid representing the respective SiS class. The resulting trapezoid has a centroid, equivalent to its centre of mass if the trapezoid were made of a thin sheet of material of uniform density. This centroid has an *x*-value, so a line dropped from the centroid to the *X*-axis gives the SiS for this play.

For the comedy *All's Well That Ends Well* there are six data and they correspond to the proportion of tokens of types of each of six set out of the total of (initial) approximately 10,000 words of each play: Set One amounts to 1.9%, Set Two to 8.3%, Set Three to 0.2%, Set Four to 2.5%, Set Five to 1% and Set Six to 0.4%.

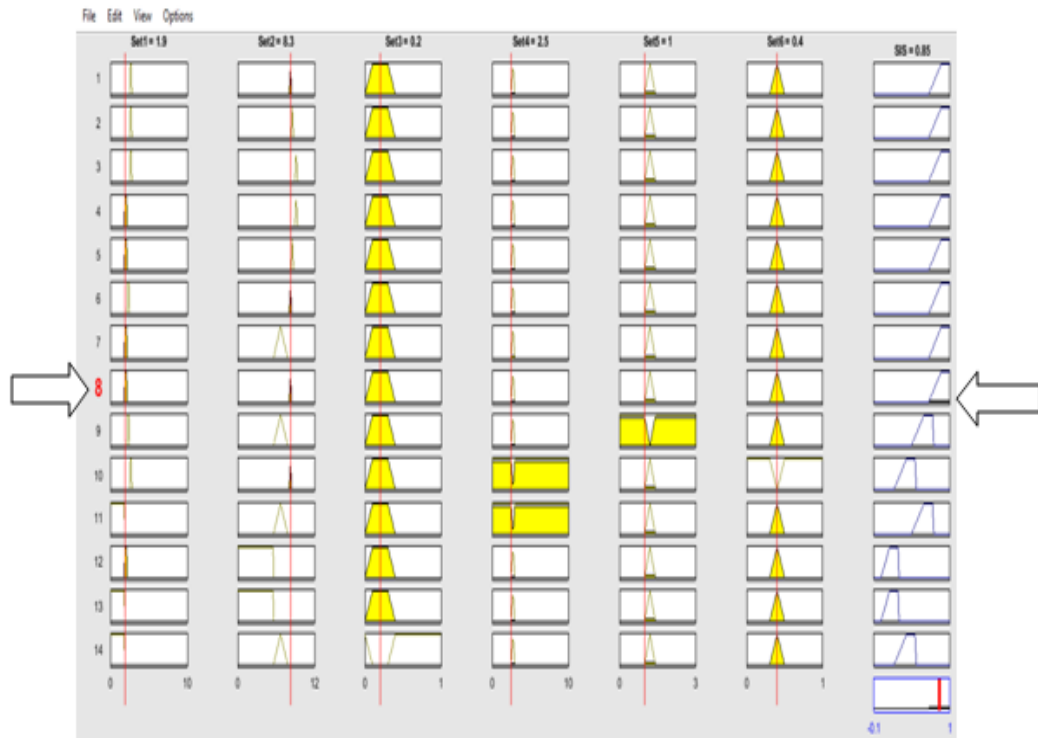


Figure 67: Trigger of Rule 8 for the SiS of *All's Well That Ends Well*.



Figure 68: Minimum of memberships of data points of *All's Well That Ends Well*.

The data points of *All's Well That Ends Well* match the shapes-combination of the six sets in the following order:

1a AND 2b AND 3 AND 4 AND 5 AND 6

As you can see in this combination, there is not any ‘not/none’ set. This pattern represents the antecedent of rule eight (R8), which produces as a consequent a centroid from SiS1-High class. To be more specific, rule eight is defined as:

R8 IF 1a AND 2b AND 3 AND 4 AND 5 AND 6 THEN SiS1-High

The six data values for *All's Well that Ends Well* all match actual shapes in the six sets. That is, they all correspond to membership values above zero. But not all correspond to membership values that are fully 1.0. The membership values yielded by these six data are 1, 1, 1, 0.2, 0.2, and 1 for Sets One through Six respectively. Of these memberships, the lowest are the 0.2 values for Set Four and Set Five, so $y=0.2$ is the truncation line that is drawn across the output shape that is triggered by Rule 8 which is the SiS1-High shape. The resulting trapezoid has a centroid from which a line is dropped vertically to the X-axis and gives the SiS result for *All's Well that Ends Well* of 0.85.

The same process for *Henry VI Part 1* matches the ‘not’ class of Set Five, which triggers this rule:

R9: IF 1b AND 2a AND 3 AND 4 AND 5-NOT AND 6 THEN SiS1-Medium2

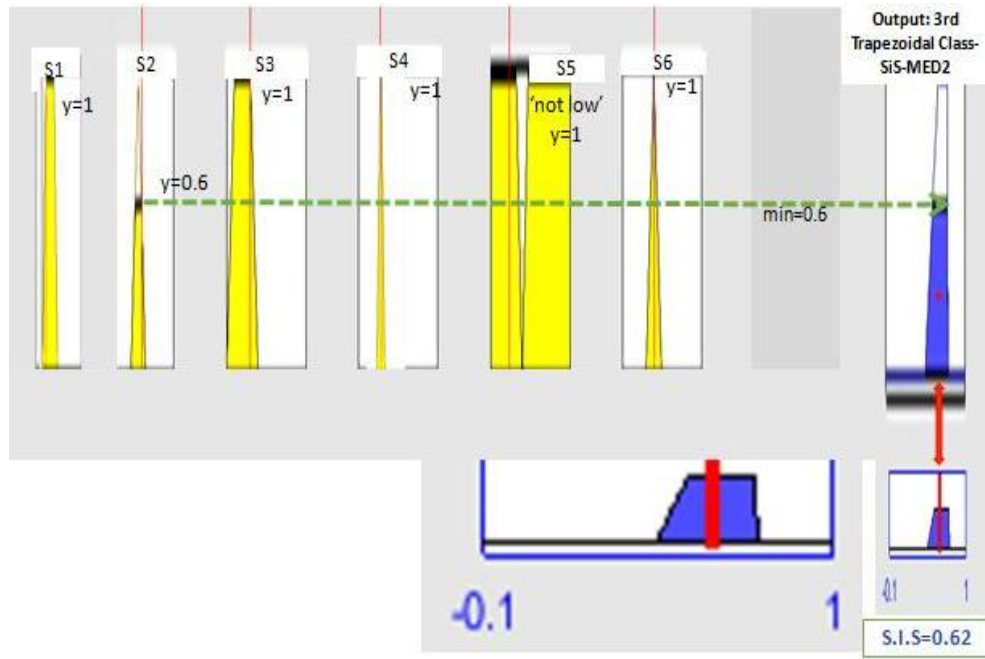


Figure 69: Minimum of memberships of data points of *Henry VI Part 1*.

Therefore, only five of the six data values for *Henry VI Part 1* match actual shapes in five of the six sets. That is, five of them correspond to membership values above zero. But not all correspond to membership values that are fully 1.0. The membership values yielded by these five data are 1, 0.6, 1, 1 and 1 for Sets One, Two, Three, Four, and Six respectively. There is also a complementary membership of Set Five as 'not', which in this case is fully 1.0. Of these memberships, including that of Set Five, the lowest is the membership of Set Two at $y=0.6$, so that is the truncation line that is drawn across the output shape that is triggered by Rule 9 which is the SiS-Medium2 shape. The resulting trapezoid has a centroid from which a line is dropped vertically to the X -axis and gives the SiS result for *Henry VI Part 1* of 0.62.

The data points of *The Jew of Malta* match the shapes-combination of the six sets in the following order:

R10 IF 1None AND 2b AND 3 AND 4-NOT AND 5 AND 6 THEN SiS1-Medium1

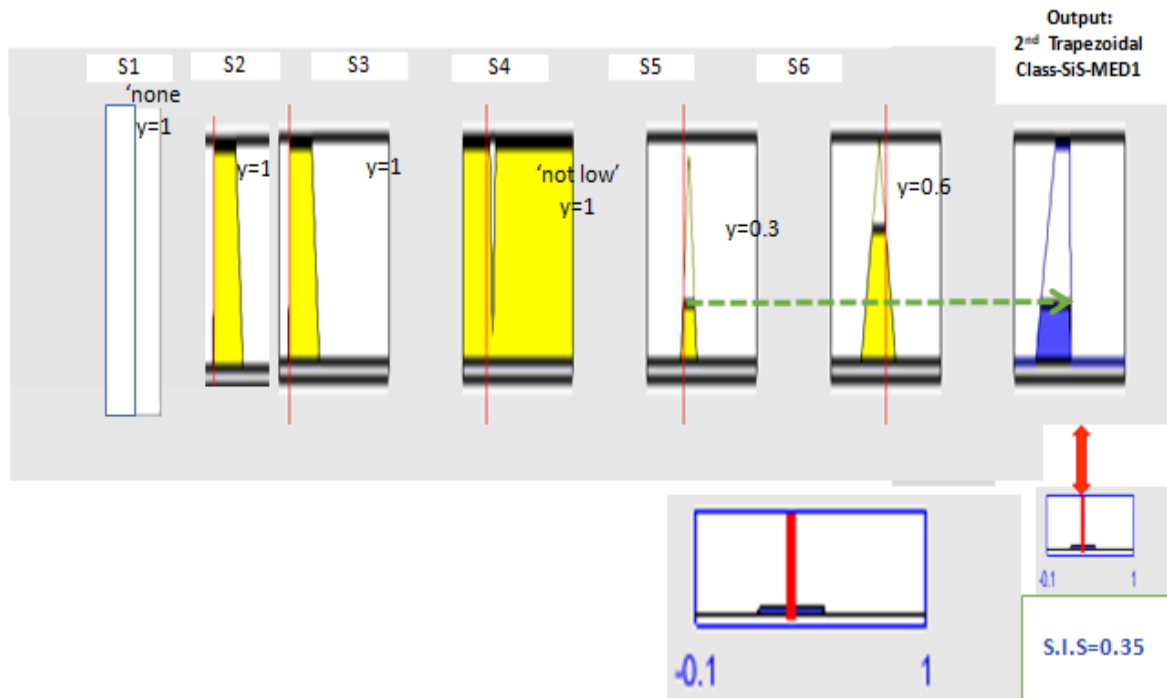


Figure 70: Minimum of memberships of data points of *The Jew of Malta*.

As you can see in this combination, there are two ‘not/none’ sets, that is Set One and Set Four. (Set One is marginally ‘none’ as the data point of 2.5 of *The Jew of Malta* falls exactly on the bottom-left edge of an actual Shakespearian class ($y=0$) and in the beginning of a ‘none’ subclass, and so it has the membership of 1, $y=1$, in the ‘none’ class). This abruptness of ‘not/none’ classes is tackled in the core experimentation and the risk it entails is discussed further in the Chapter of Conclusions. This pattern represents the antecedent of rule 10, which produces as a consequent a centroid from SiS1-Medium1 class.

Only four of the six data values for *The Jew of Malta* match actual shapes in four of the six sets. That is, four of them correspond to membership values above zero. But not all correspond to membership values that are fully 1.0. The membership values yielded by these four data are 1, 1, 0.3 and 0.6 for Sets Two, Three, Five and Six, respectively. There is also a complementary membership of Set One and Four, which in this case are both fully 1.0. Of these six memberships, including this of Set Four, the lowest is the membership of Set Five, at $y=0.3$, so that is the truncation line that is drawn across the output shape that is triggered by Rule 10 which is the SiS1-Medium1 shape. The resulting trapezoid has a centroid from which a line is dropped vertically to the X-axis and gives the SiS result for *The Jew of Malta* of 0.35.

Summarising the results for the three plays, these are:

All's Well That Ends Well

Similarity Index to Shakespearian model/style: 0.85

Henry VI Part I

Similarity Index to Shakespearian model/style: 0.62

The Jew of Malta

Similarity Index to Shakespearian model/style: 0.35

7.2.5 Conclusions of the Primary Experimentation.

A major argument for the selection of semantic sets as stylistic discriminators in this experimentation is that it is feasible to investigate and represent numerically the strong existence of Shakespearian stylistic patterns. This serves in reducing the distorted effect of the differentiation of genre particularities, as in the case of over-predominance of male features 'he', 'him', 'himself' in history plays, replacing the 'expected counts' of 'she', 'her', 'herself'. Measuring the counts of sets of words, it is feasible to assess the counts of a set of words, whose own counts of frequencies may in certain cases differ dramatically from play to play. Of course, the evaluation of this differentiation perhaps could require additional critical analysis based on the existing particularities, for instance relevant historical scholarship, plays' chronologies and genre characteristics. John Burrows's principles have been also employed regarding the necessity to avoid employing only words of high frequencies as stylistic markers.

Further validation of the Fuzzy Model and the Sets-based approach has been carried out with the measure of cosine similarity, which is an index that, in our case, expresses the similarities of counts of the frequencies of each of the words of the six sets of the average-ideal document-play, representing the average style of the nine known Shakespearian plays, with the respective counts of frequencies of words of each of the three plays under scrutiny (see Section 7.3.9 of the Technical Appendix).

7.3 Technical Appendix

This appendix facilitates the understanding of technical terms and complex technical specifications. There is a clarification of scientific concepts and the terms of Fuzzy Logic. Matlab Fuzzy Tool's technicalities are described in detail.

7.3.1 Open and Closed Intervals.

An interval of two values-numbers (a, b) is the set of numbers that are included between these two values-numbers. In this thesis, I refer to intervals of real numbers that constitute a sequence of numbers in (the continuity of) first decimal. For instance, data points such as 2.0, 2.2, 2.1, 2.3, 2.4 constitute a closed (continuous) interval of real numbers in first decimal between 2.0 and 2.4 and this is indicated in terms of notation as [2.0, 2.4]. A closed interval is expressed by putting brackets at the two values-endpoints of the interval (and it means the two values-endpoints inclusive), whereas an open interval is denoted by enclosing the interval within parentheses (and it means the two values-endpoints exclusive). The closed interval of the example [2.0, 2.4] is interpreted as the interval of the numbers that are greater or equal to 2.0 and less or equal to 2.4, whereas the open interval (2.0, 2.4) is interpreted as the interval of the numbers that are greater than 2.0 and less than 2.4. There are also semi-closed and semi-open intervals ([a, b) or (a, b]). There are also continuous functions on intervals. Though this is a complex mathematical concept and it exceeds the purposes of our study, we can talk about micro-continuity. For instance, say there is an interval such as [2.1, 2.4] and a function such as $f(x)=1/x$. We can assess if small changes in the input/value of x (x_1 as 2.1, then x_2 as 2.2, then x_3 as 2.3, then x_4 as 2.4) leads to small changes in the output/function ($y_1=0.47$, then $y_2=0.45$, then $y_3=0.43$, then $y_4=0.41$, in second decimal precision) and a continuous function-curve can be formed for all the points (x, y) of the interval [2.1, 2.4] in the X, Y coordinate system.

7.3.2 'Ant' Statistical Analyser.

The AntWord Profiler essentially is a statistical analyser of counts of selected by the user individual or grouped into sets lexical features. For instance, by running the AntWord Profiler program for the first 10,028 words of a history under scrutiny, *Edward II*, the following graphical output is produced:

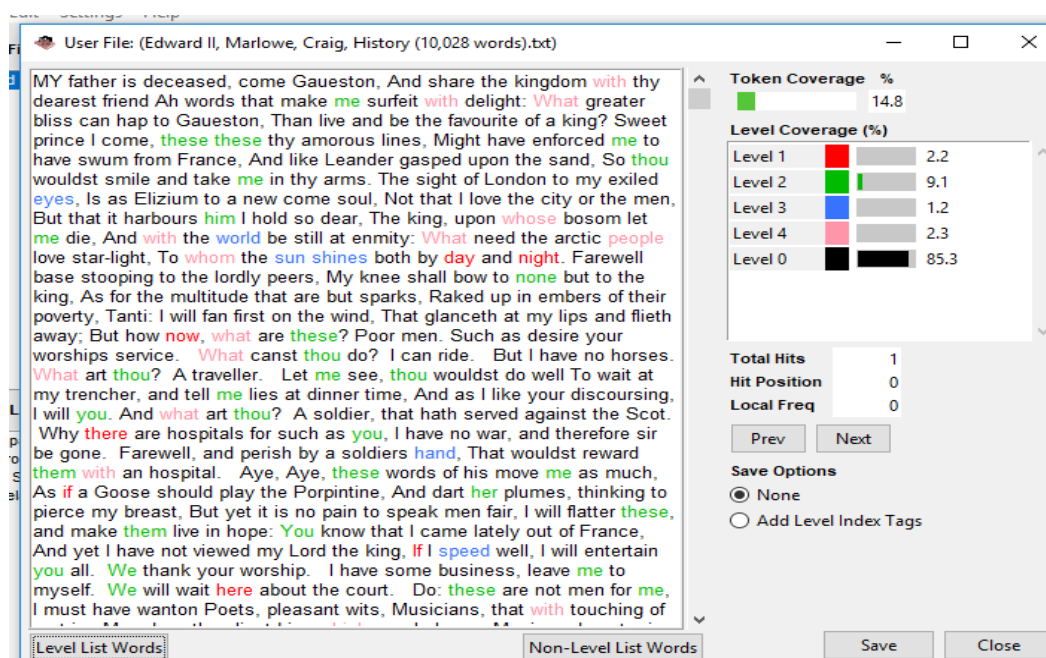


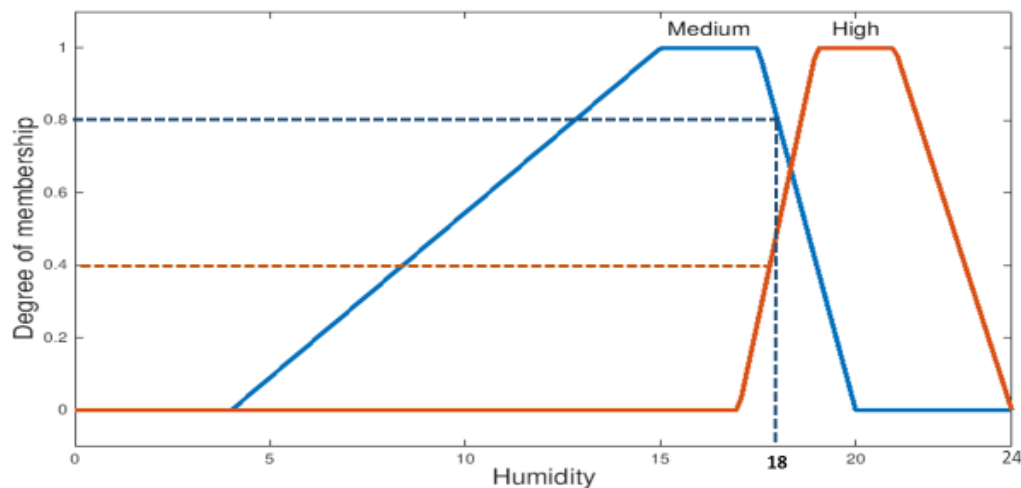
Figure 71: ‘Ant’ Statistical Analyser.

The fields of Level 1 to 4 correspond to the counts in percentage of each of the four semantic sets in sequence from Set One (bar in red colour) to Set Four (bar in rose) and the token coverage (see green in white field) equals the total of counts of the four sets containing the counts of the individual 100 words-types. The field named Level 0 (Non-Level List Words, see bar in black) denotes the counts of the words’ tokens that are not included in any of the four sets.

7.3.3 Absolute and Relative Humidity

In science, the concept of humidity has various terminologies, and the two prevailing are those of the absolute and relative humidity. In the second case, the humidity is not defined as an autonomous quantity, like every physical quantity in its absolute definition. Roughly speaking, as the temperature increases, the air starts to become less capable of holding water vapour. Other complexities are also applied, such as the saturation of air and the relation of what is called dew point in association with the current temperature. In addition, there are other physical quantities or artificial factors, such as the outdoors temperature and the structure of the building and the room that affect particularly indoors temperature and humidity control. In the current experimentation, our goal is, mainly, educative and it is to show the aspects of the functionality of the application of Fuzzy Logic. So, in the examples

with the function of the fan controller and the design of the membership function of humidity, we adopted a general, absolute, approach of the term humidity though, by the way I designed the membership function of the class of high humidity, I tried also to incorporate a sense of this kind of relativity with high temperature, as can be viewed by the linear fuzziness of the right leg of the second trapezoidal class of the variable of humidity.



:

Figure 72: Membership functions-classes of humidity.

Thus, the x -values of humidity that are greater than 22 grams have a linearly decreased membership in the class of high humidity. On the contrary, this is not the case with the design of x -values of the class of high temperature. This differentiation in the design of the two variables is an indirect way to express the term of relative humidity, that is, the possible particularities of the interrelationship between high humidity and high temperature.

7.3.4 Centroid of Area (COA).

The typical formula for the calculation of the centroid (COA) is: $COA = \frac{\int_{u_{min}}^{u_{max}} f(x) \cdot x \, dx}{\int_{u_{min}}^{u_{max}} f(x) \, dx}$. I have made here a slight adaptation of the formula mentioned in the toolkit of Fuzzy Logic for COA (“LabVIEW 2011 PID and Fuzzy Logic Toolkit Help” 2011), and I replaced ‘ x ’ with the term AVM, average-mean value. Explaining the above formula, the denominator ($\int f(x) \, dx$) produces the area of the truncated shape, and the numerator is a number that when divided by the measurement of that area produces the centroid value on the X -axis. The values of ‘ u_{max} ’ and ‘ u_{min} ’ are the limits on the X -axis of the truncated (here trapezoid) shape. The minimum is the u -value on the left of origins of the

X-axis and the maximum is the right point on X-axis of the shape. The sign of the integral (\int), definite as there are limits, denotes that in fact in order to find the area of the truncated shape the output shape can be divided into very narrow rectangular strips (assume here definite, though it is not). In practice each rectangular strip is added to the next one, and they are summed together, and so practically the centroid each time is shifted. The operation of $f(x) dx$ refers to the multiplication of the height, the membership on the Y-axis, by y the width, the residual of maximum minus the minimum point on the X-axis, of each added rectangular strip. Therefore, with the calculation of the integral of the denominator we manage to measure the area of the truncated shape. In the numerator the same operation is used as in the denominator but there is also another operand-number which is called average-mean value (AVM) of each thin rectangle. The calculation of AVM is performed by applying specific continuous functions in the interval of maximum and minimum values on the X-axis and this average value is a point between u_{maximum} and u_{minimum} . This AVM can be found by dividing one by the residual of u_{max} minus u_{min} on the X-axis and then multiplying that by the area of the truncated shape. In terms of a formula, the last operation is represented as the function:

$$f(\text{AvM}) = 1 / (u_{\text{max}} - u_{\text{min}}) * \int (f(x)) dx \text{ (where as said } \int (f(x)) dx \text{ equals the sum of areas of the thin rectangular strips).}$$

7.3.5 Trapezoidal and Triangular Functions.

Trapezoidal and triangular shapes are called piecewise linear functions. In fact, trapezoids are a kind of triangles where the left minimum point of the upper side of the trapezoid equals with the maximum point (Viattchenin, Reyhane, and Damaratski 2013).

7.3.6 Calculation of memberships (Y-axis) of values (u) on X-axis.

Trapezoidal function

Note: The X-axis expresses the counts (in percent %) of a set of words and the Y-axis the degrees of their memberships in the actual class-trapezoid.

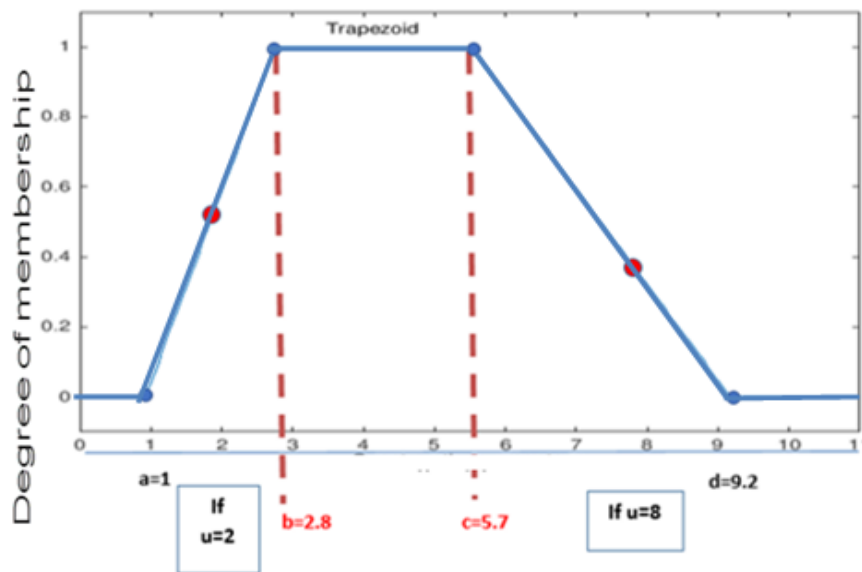


Figure 73: Premise parameters of a trapezoidal class and calculation of the degree of membership (y-value) and x-value.

As can be viewed in the above graph, the Y-axis expresses the degree of membership of each value on X-axis in the actual class-trapezoidal function. If u on the X-axis (graph above), as a value of counts of a set of words, then, its membership is (that is, y equals) 1 for u in $b \leq u \leq c$. (Note: The segment on the X-axis of the open interval (b, c) is parallel to the upper side of the trapezoid.)

The membership of any other value (u) on the X-axis in the actual trapezoidal class that is not included in the segment-closed interval $[a-d]$ of the X-axis is 0 ($y=0$).

For u and any value on the X-axis that falls in the area below of the left slope ($a \leq u \leq b$), the results of memberships derive from the equation $y = \frac{u-a}{b-a}$ (where y is the membership of u in the actual trapezoid).

For u and any value on the X-axis that falls in in the area below of the right slope ($d \leq u \leq c$), the results of memberships derive from the equation $y = \frac{d-u}{d-c}$ (where y the membership of u in the actual trapezoid).

Thus, in the above graph, for $u=2$ (X-axis) the membership is

$$y = \frac{2-1}{2.8-1} = \frac{1}{1.8} = 0.55/1.$$

and for $u=8$, the membership is

$$y=9.2-8/9.2-5.7=1.2/3.5=0.34/1.$$

Triangular function

Note: The X -axis expresses the counts of a set of words (in percent %) and the Y -axis the degrees of their memberships in the actual class-triangle.

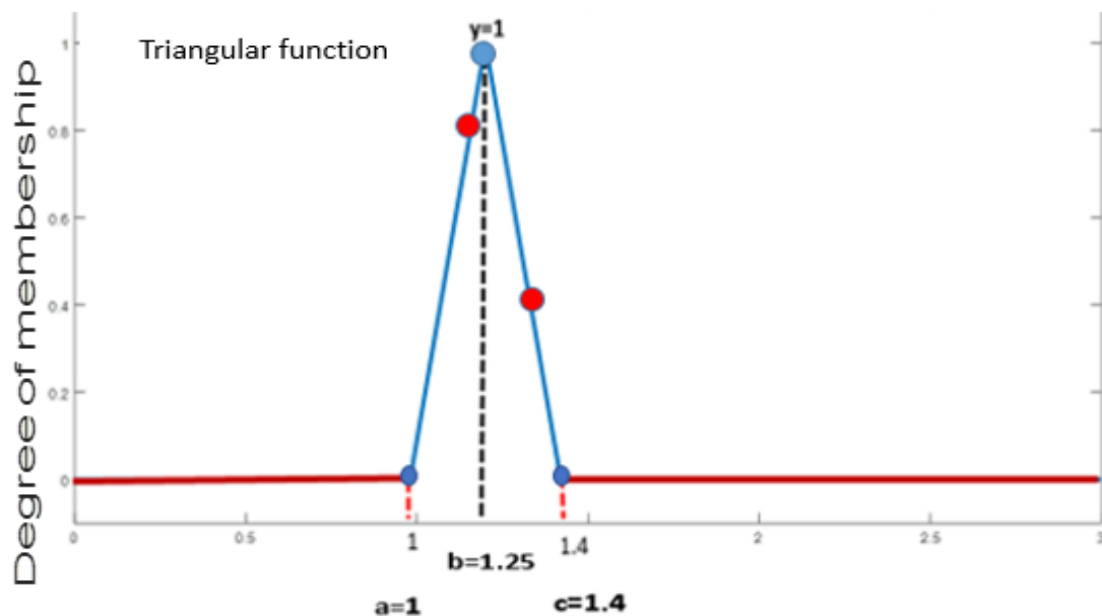


Figure 74: Premise parameters of a triangular class and calculation of the degree of membership (y-value) and x-value.

In the graph above, there are three points on the X -axis: $a=1$ (the triangle's bottom-left edge point), $b=1.25$ (the parallel point on the X -axis to the triangle's peak point [Y -axis]) and $c=1.4$ (the triangle's bottom-right edge point).

The membership of the b point ($b=1.25$) on the X -axis in the actual triangular function is 1 ($y=1$). The membership of any other value on the X -axis not included in the segment-closed interval $[a, c]$ in the actual triangular function is 0 ($y=0$).

Similar to the calculation of trapezoids' memberships, we compute the memberships for values on the X -axis that fall in the area below the left and right slope in a triangular function. In fact, these memberships (y -values) are calculated based on the distance of a

x -value (u) to at least two of the three points (a, b, c) on the X -axis. Therefore, based on the above graph:

For u in the area of left slope: $y = u - a / b - a$, ($a \leq u \leq b$) and

For u in the area of right slope: $y = c - u / c - b$, ($b \leq u \leq c$).

Thus, for $u = 1.2$ (X -axis) the membership is

$$y = 1.2 - 1 / 1.25 - 1 = 0.8 / 1$$

and for $u = 1.3$, the membership is

$$y = 1.4 - 1.3 / 1.4 - 1.25 = 0.66 / 1.$$

7.3.7 ‘None’ Subclasses of an Aggregated ‘None’ Class.

Consequently, the number of total combinations with all the separate negative subclasses is typically much larger than presented in the search and harvest process but that complexity can be avoided by employing in our analysis the term of simply one ‘none’ class, meaning none of actual classes, for Set One and Two, without this affecting at all the results for the three new plays under scrutiny. Regarding Set One, whether instead of 1none, we write ‘1nonea’ or ‘1noneb’ or ‘1nonec’ class in the antecedents of rules, there is no substantial difference and there is no need to include all separate ‘none’ classes. The differentiation would be necessary only if any of the three new plays, which is not the case, had Set One or/and Set Two with counts falling into the separate ‘none’ classes. In that case, I would have to generate new rules for the play under scrutiny, as I did for the two of our three plays. The x -values that fall into any of the ‘none’ classes have a membership (y) of 1 in the negated class and a 0 in any of the actual classes (see the first left in red trapezoid of the graph below). The application of a different smoother and fuzzier design of ‘none’ subclasses, like that of ‘not’ classes and its supplementary values, is discussed in Chapter Four and the core experimentation.

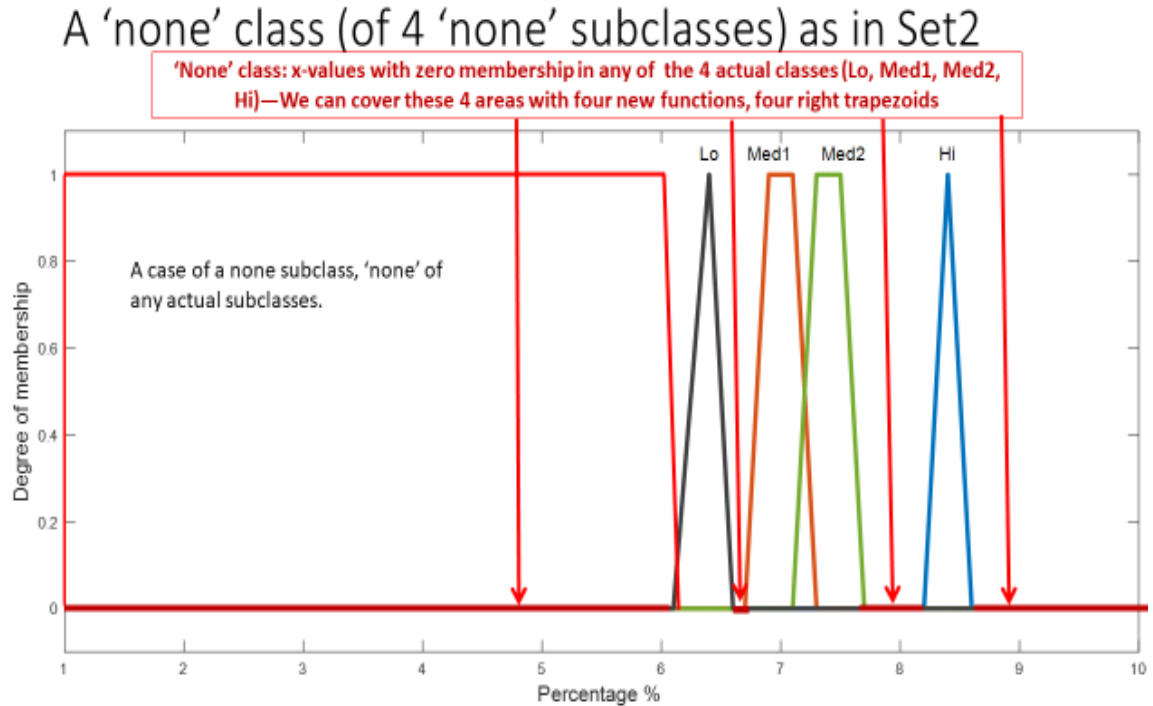


Figure 75: An aggregated 'none' class.

7.3.8 Technicalities of Adaptation of 'None' Classes (Subclasses).

Therefore, in the Matlab program I added the four rules as:

- R11 IF 1NONEa AND 2a AND 3 AND 4 AND 5 AND 6 THEN SiS1-Medium2
 R12 IF 1a AND 2NONEa AND 3 AND 4 AND 5 AND 6 THEN SiS1-Low
 R13 IF 1NONEa AND 2NONEa AND 3 AND 4 AND 5 AND 6 THEN SiS1-LOW
 R14 IF 1NONEa AND 2a AND 3NOT AND 4 AND 5 AND 6 THEN SiS1-Medium1

The results, in our case, would be the same whatever the last index of 1NONE and 2NONE, for example 1NONEa or 1NONEb. This index corresponds to an area, a range of values on the X-axis uncovered by any of the actual classes.

7.3.9 An Informal Validation of the Fuzzy Model and the Sets-based Approach--Cosine Similarity.

For further verification of the results of the fuzzy program with the six sets-input variables can be applied the measure of cosine similarity, which is an index that expresses the

similarities of counts of the frequencies of each of the six sets' words of the average-ideal document-play with the respective counts of frequencies of words of each of the three plays under scrutiny. This index of cosine measurement ranges in the scale of 0 to 1 and it is the cosine of the degree of the angle that two documents--each of three tested texts and the average-ideal document of the nine plays--as vectors form.

Vectors can be viewed in a simple way as a single row or column of the counts of the frequencies of 102 words of the six sets. The vector of the average-ideal document contains the standardised average counts of the frequencies (tokens) of the 102 words-types as detected in the nine known Shakespearian plays. In fact, I applied a version of an Inverse Document Frequency Standardisation. This means that the tokens of each of the 102 words-types were summed, divided by nine and finally were weighted negatively if they were not detected in all nine documents. For instance, let us assume the average of the tokens of the word-type 'day' is ten, meaning there are a total of 90 counts-tokens of 'day' in the whole corpus of nine plays. If these counts were detected in only seven out of the nine known Shakespearian plays, the entry of that word-element in the average-ideal document-vector of 102 elements-dimensions would be $10 \times 7/9$, thus 'a bit less than eight' counts. If the cosine of the angle ($\cos \theta$) formed by the average-ideal document and a tested document is 0, thus the angle is 90° . In this case, the documents are completely dissimilar. If the cosine of the angle is 1, thus an angle of 0° formed and the documents-plays are completely similar. In the first case (90°), when there is orthogonality, two documents represented as vectors are characterised by perpendicular direction. Here, the angle of difference with the average-ideal document for *All's Well That Ends Well* is 15° and for each of the two plays, that is *Henry VI Part 1* and *The Jew of Malta*, corresponds to approximately 22° (respectively 22.035° and 21.4°). So, the results of the cosine similarity in comparison with the final index (SIS) of the fuzzy model for the three tested plays are:

All's Well That Ends Well

Cosine Similarity: 0.964

Similarity Index to Shakespearian model/style: 0.85

Henry VI Part 1

Cosine Similarity: 0.926

Similarity Index to Shakespearian model/style: 0.62

The Jew of Malta

Cosine Similarity: 0.931

Similarity Index to Shakespearian model/style: 0.35

Clearly, the cosine result for *All's Well That Ends Well* does not contradict the SiS1-High index of our sets-based experimentation. These cosine indices validate our SiS1-High result for *All's Well That Ends Well* and show that sets of words can be reliable stylistic markers for solving authorship attribution problems.

I produced the results of cosine similarity by employing one of the many available automated utilities of cosine similarity calculator. (The formula of cosine similarity between two documents-vectors was explained in detail in the literature review in subsection 1.3.1.2.) But in the core experimentation and the genre-(tragedies)-based model, *The Jew of Malta* has a much lower index of cosine similarity (0.917), which is an indication that the genre-based models agree more fully with the evidence of historical scholarship and the fact that *The Jew of Malta* is widely considered a well-attributed Marlovian play.

Note: See below the specifications and the printout of the comparison of the two vectors (Row1 and Row2) that contain as entries the compared counts of the frequencies of 102 words respectively of *All's Well That Ends Well* and those of the average-ideal document of the nine well-attributed, sole-authored Shakespearian plays.

Cosine Similarity Calculator

Please click to add a row.

Row 1:

Row 2:

This blog post calculates the pairwise Cosine similarity for a user-specifiable number of vectors. All vectors must comprise the same number of elements.

Simply click on the link near the top to add text boxes. Each text box stores a single vector and needs to be filled in with comma separated numbers. All rows need to have the same number of samples.

Alternatively, you can choose two file entry methods:-

1. Select multiple single column CSV files to populate the text boxes by repeatedly pressing the Choose File button - there must be one distinct (and differently named) file for each text box i.e. one file per group. Each file can have a different number of samples.
2. Select a single multi-column CSV file by pressing the Choose File button once, where the number of columns equals the number of vectors.

Cosine Similarity between vectors 1 and 2 is 0.9642, cosine of 0.2684 radians

Figure 76: An on-line tool of calculation of the cosine similarity between two documents-vectors.

For the core experimentation (Chapter Four) a similar, more usable, online tool, a calculator of cosine similarity was employed, thus <http://www.minerazzi.com/tools/cosine-similarity/cosine-similarity-calculator.php>. (Set A and Set B are the two documents whose entries of words' frequencies are counted.)

Two simple examples, one with the counts of three words and another with four words:

Cosine Similarity Calculator

This tool computes the cosine similarity between two data sets and determines whether these are uncorrelated, orthogonal, or both.

[Try This Example](#)

Set A	Set B
5 5 5	5 5 5

Delimiter: Decimal places:

Mode: ☐ centered ☒ raw



These sets are not orthogonal

cosine similarity = 1.00 angle = 0.00 angle type = acute

Cosine Similarity Calculator

This tool computes the cosine similarity between two data sets and determines whether these are uncorrelated, orthogonal, or both.

[Try This Example](#)

Set A	Set B
1 17 28 58	58 28 17 1

Delimiter: Decimal places:

Mode: ☐ centered ☒ raw



These sets are not orthogonal

cosine similarity = 0.29 angle = 72.88 angle type = acute

Figure 77: An example of calculation of cosine index.

And below a graphical output from the actual experimentation with the comparison of the counts of 100 words' frequencies between two vectors-documents A (= average-ideal document/comedy) and B (= *The Tempest*). (Here the result is rounded to the second decimal place and the counts are not integers since the process of standardisation has taken place.)

Cosine Similarity Calculator

This tool computes the cosine similarity between two data sets and determines whether these are uncorrelated, orthogonal, or both.

[Try This Example](#)

Set A				Set B			
57.25	29.42	6.08	29.58	50.15	28.09	11.03	36.11
28.50	30.08	1.69	32.17	30.09	44.14	3.01	47.15
7.00	0.17	7.25	1.67	4.01	0.00	9.03	2.01
0.81	0.62	202.58		1.00	1.00	118.27	

Delimiter: Decimal places:

Mode: ☐ centered ☒ raw



These sets are not orthogonal

cosine similarity = 0.91 angle = 25.09 angle type = acute

Figure 78: An example of calculation of the cosine index from actual experimentation.

7.3.10 Script of Fan Controller (3.1).

[System]

Name='fan4118grams' % Test program with Input1-temperature 41 Celsius and
% Input2-humidity 18 grams.

```

Type='mamdani'
Version=2.0
NumInputs=2
NumOutputs=1
NumRules=2
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'

```

```

[Input1]
Name='Temperature'
Range=[-5 45]
NumMFs=2
MF1='Medium': 'trapmf',[12 18 25 37]
MF2='High': 'trapmf',[28 38 46 46]

```

```

[Input2]
Name='Humidity'
Range=[0 24]
NumMFs=2
MF1='Medium_Humidity': 'trapmf',[4 14 17 20]
MF2='High_Humidity': 'trapmf',[17 20 23 24]

```

```

[Output1]
Name='Rotations/Minute'
Range=[0 250]
NumMFs=2
MF1='Medium_Speed': 'trimf',[84.38 166.7 166.7]
MF2='High_Speed': 'trimf',[149 221 221]

```

```

[Rules]
2 2, 2 (1) : 1
1 1, 1 (1) : 1

```

7.3.11 Script of Fuzzy Stylistic Classifier in the Primary Experimentation.

```

[System]
Name='ShakespSiSPPrimaryExp'
Type='mamdani'
Version=2.0
NumInputs=6
NumOutputs=1
NumRules=14
AndMethod='min'
OrMethod='max'
ImpMethod='min'

```

AggMethod='max'
DefuzzMethod='centroid'

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=7
MF1='1a': 'trapmf',[1.8 1.9 2.1 2.2]
MF2='1b': 'trapmf',[2.21 2.3 2.4 2.47]
MF3='1c': 'trapmf',[2.53 2.6 2.7 2.77]
MF4='1NONEa': 'trapmf',[0 0 1.8 1.9]
MF5='1NONEb': 'trapmf',[2.1 2.2 2.21 2.3]
MF6='1NONEc': 'trapmf',[2.46 2.47 2.53 2.54]
MF7='1NONEd': 'trapmf',[2.76 2.77 100 100]

[Input2]
Name='Set2'
Range=[0 12]
NumMFs=9
MF1='2a': 'trimf',[5.57 6.7 7.83]
MF2='2b': 'trapmf',[8.13 8.2 8.3 8.37]
MF3='2c': 'trapmf',[8.43 8.5 8.6 8.67]
MF4='2d': 'trimf',[8.94 9.15 9.36]
MF5='2Nonea': 'trapmf',[0 0 5.57 5.58]
MF6='2Noneb': 'trapmf',[7.82 7.83 8.13 8.14]
MF7='2Nonec': 'trapmf',[8.36 8.37 8.43 8.44]
MF8='2Noned': 'trapmf',[8.66 8.67 8.94 8.95]
MF9='2Nonee': 'trapmf',[9.35 9.36 100 100]

[Input3]
Name='Set3'
Range=[0 1]
NumMFs=1
MF1='Actual': 'trapmf',[0.093 0.1 0.3 0.37]

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=1
MF1='Actual': 'trimf',[2.48 2.74 3]

[Input5]
Name='Set5'
Range=[0 3]
NumMFs=1
MF1='Actual': 'trimf',[0.94 1.155 1.37]

[Input6]
Name='Set6'
Range=[0 1]

```

NumMFs=1
MF1='Actual': 'trimf',[0.275 0.44 0.59]

[Output1]
Name='SIS'
Range=[0 1]
NumMFs=4
MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1.1]

```

```

[Rules]
3 2 1 1 1 1, 4 (1) : 1
3 3 1 1 1 1, 4 (1) : 1
3 4 1 1 1 1, 4 (1) : 1
1 4 1 1 1 1, 4 (1) : 1
1 3 1 1 1 1, 4 (1) : 1
2 2 1 1 1 1, 4 (1) : 1
1 1 1 1 1 1, 4 (1) : 1
1 2 1 1 1 1, 4 (1) : 1
2 1 1 1 -1 1, 3 (1) : 1
6 2 1 -1 1 1, 2 (1) : 1
4 1 1 -1 1 1, 3 (1) : 1
1 5 1 1 1 1, 1 (1) : 1
4 5 1 1 1 1, 1 (1) : 1
4 1 -1 1 1 1, 2 (1) : 1

```

7.3.12 Script of Comedies-Based Fuzzy Stylistic Classifier.

```

[System]
Name='41-- R Comedies classifier SiS1-2'
Type='mamdani'
Version=2.0
NumInputs=7
NumOutputs=2
NumRules=41
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'

[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a': 'trimf',[1.79 1.9 2.01]

```

```
MF2='1b': 'trimf', [2.04 2.37 2.71]
MF3='1nota': 'trapmf', [0 0 1.7 1.81]
MF4='1notb': 'trapmf', [1.99 2.1 2.11 2.2]
MF5='1notc': 'trapmf', [2.54 2.71 100 100]
```

```
[Input2]
Name='Set2'
Range=[0 20]
NumMFs=7
MF1='2nota': 'trapmf', [0 0 6 6.1]
MF2='2notb': 'trapmf', [6.29 6.4 7.69 8.02]
MF3='2notc': 'trapmf', [9.88 10.14 100 100]
MF4='2a': 'trimf', [6.09 6.2 6.31]
MF5='2b': 'trimf', [7.69 8.35 9.01]
MF6='2c': 'trapmf', [8.94 9 9.2 9.26]
MF7='2d': 'trimf', [9.13 9.63 10.14]
```

```
[Input3]
Name='Set3'
Range=[0 10]
NumMFs=3
MF1='3': 'trimf', [1.32 1.71 2.094]
MF2='3nota': 'trapmf', [0 0 1.32 1.52]
MF3='3notb': 'trapmf', [1.901 2.094 100 100]
```

```
[Input4]
Name='Set4'
Range=[0 10]
NumMFs=5
MF1='4a': 'trapmf', [1.94 2 2.1 2.16]
MF2='4b': 'trimf', [2.27 2.59 2.91]
MF3='4nota': 'trapmf', [0 0 1.95 2]
MF4='4notb': 'trapmf', [2.15 2.2 2.27 2.425]
MF5='4notc': 'trapmf', [2.75 2.91 100 100]
```

```
[Input5]
Name='Set5'
Range=[0 20]
NumMFs=8
MF1='5a': 'trimf', [11.89 12.1 12.21]
MF2='5b': 'trimf', [14.43 14.93 15.43]
MF3='5c': 'trapmf', [15.39 15.5 15.7 15.81]
MF4='5d': 'trapmf', [15.79 15.9 16.1 16.21]
MF5='5nota': 'trapmf', [0 0 11.8 11.91]
MF6='5notb': 'trapmf', [12.19 12.3 14.43 14.68]
MF7='5notc': 'trimf', [15.18 15.43 15.43]
MF8='5notd': 'trapmf', [16.19 16.3 100 100]
```

```
[Input6]
Name='SiS1asInput'
```

```

Range=[0 1]
NumMFs=4
MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

```

```

[Input7]
Name='CosSim'
Range=[0 1]
NumMFs=2
MF1='LowCos': 'trimf',[0 0.884 0.885]
MF2='HighCos': 'trapmf',[0.884 0.968 1 1]

```

```

[Output1]
Name='SIS'
Range=[0 1]
NumMFs=4
MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

```

```

[Output2]
Name='SIS2'
Range=[0 1]
NumMFs=4
MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

```

```

[Rules]
2 6 1 2 0 0 0, 4 0 (1) : 1
2 7 1 2 0 0 0, 4 0 (1) : 1
1 4 1 2 0 0 0, 4 0 (1) : 1
2 5 1 2 0 0 0, 4 0 (1) : 1
2 6 1 1 0 0 0, 4 0 (1) : 1
2 7 1 1 0 0 0, 4 0 (1) : 1
3 4 1 1 0 0 0, 3 0 (1) : 1
1 1 1 1 0 0 0, 1 0 (1) : 1
3 1 1 1 0 0 0, 1 0 (1) : 1
3 4 2 1 0 0 0, 1 0 (1) : 1
3 1 2 1 0 0 0, 1 0 (1) : 1
3 1 1 3 0 0 0, 1 0 (1) : 1
5 6 1 2 0 0 0, 3 0 (1) : 1
5 5 1 2 0 0 0, 3 0 (1) : 1
1 4 2 2 0 0 0, 3 0 (1) : 1
2 2 1 2 0 0 0, 1 0 (1) : 1
2 5 3 2 0 0 0, 3 0 (1) : 1

```

4 2 1 1 0 0 0, 1 0 (1) : 1
 5 5 1 5 0 0 0, 2 0 (1) : 1
 2 6 1 4 0 0 0, 3 0 (1) : 1
 2 5 1 5 0 0 0, 3 0 (1) : 1
 2 5 2 1 0 0 0, 3 0 (1) : 1
 4 2 2 1 0 0 0, 1 0 (1) : 1
 4 5 2 1 0 0 0, 2 0 (1) : 1
 2 2 2 1 0 0 0, 1 0 (1) : 1
 1 5 1 2 0 0 0, 4 0 (1) : 1
 1 5 1 4 0 0 0, 3 0 (1) : 1
 3 5 1 2 0 0 0, 3 0 (1) : 1
 3 5 1 4 0 0 0, 2 0 (1) : 1
 0 0 0 0 1 4 2, 0 4 (1) : 1
 0 0 0 0 6 1 2, 0 1 (1) : 1
 0 0 0 0 6 2 2, 0 2 (1) : 1
 0 0 0 0 6 2 2, 0 1 (1) : 1
 0 0 0 0 6 3 2, 0 3 (1) : 1
 0 0 0 0 6 3 2, 0 2 (1) : 1
 4 5 2 3 0 0 0, 1 0 (1) : 1
 2 5 2 3 0 0 0, 2 0 (1) : 1
 2 5 2 2 0 0 0, 3 0 (1) : 1
 2 5 2 5 0 0 0, 2 0 (1) : 1
 0 0 0 0 2 3 2, 0 3 (1) : 1
 0 0 0 0 2 3 2, 0 4 (1) : 1

7.3.13 Pseudo-code of the Algorithm of One-Level-Fall from SiS1-High.

[START OF PROGRAM—INFERENCE MECHANISM IS READY TO ASSESS
INPUT VARIABLES]

Step 1

Does the four-sets' data for the new play have, for each set's shapes, a non-zero membership?

If 'yes' then select SiS1-High and stop. If 'no', continue to Step 2.

Step 2

Does the Set Two data for the new play have a non-zero membership?

If 'no' then select SiS1-Low and stop. If 'yes' then continue to Step 3.

Step 3

Considering the three sets (1, 3, 4), does the new play's data fall within the 'not/none' class for only one of them?

If 'yes' select SiS-Medium2 and stop. If 'no', continue to Step 4.

Step 4

Considering the three sets (1, 3, 4), does the new play's data fall within the 'not/none' class for two of them? If 'yes' select SiS-Medium1 and stop. If 'no', continue to Step 5.

Step 5

Select SiS1-Low and stop.

[END OF PROGRAM]

7.3.14 Design of Bottom Corners of Actual and 'Not/None' Subclasses.

As I used (AntProfiler) software for measuring the counts of Sets, it produced at the first level statistical results rounded to one decimal place. I employed these counts as they satisfy the criteria of approximate reasoning and decimal patterns (trapezoids). (In fact, decimal rounding has been also used in the experimentation described in *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan (Taylor and Egan 2017).) According to the constraints of the design of membership functions (see Section 3.3), the triangular actual classes started and end one decimal (0.01) before the actual values of one SD below and above the class. For instance, for the triangular class of Set One with peak the point that gives on the X -axis 1.9, I used as left-bottom and right-bottom corners of the triangle the x -values of 1.79 and 2.01, so that counts of 1.8 and 2 are assessed as values--but with very low membership--of the actual class. This adaptation assists us in avoiding the creation of the so-called fuzzy singletons (classes where one data point has the membership of one and the all other data zero membership). In addition, the 'not/none' subclasses, as described in Sections 4.11.1 and 4.11.2, are designed in a more complex but fuzzier mode than in primary experimentation. Below are displayed the actual the 'not/none' classes of the five sets' data points of the comedies-based classifier model of the 12 well-attributed Shakespearian comedies.

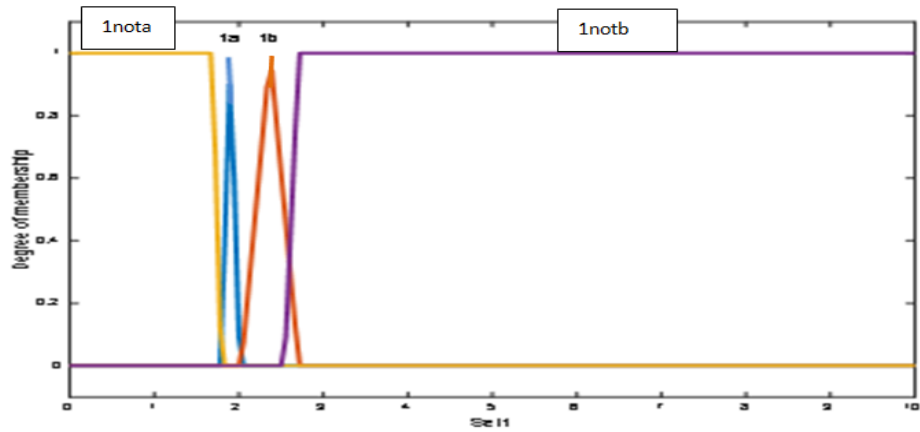


Figure 79: Membership functions of Set One for the comedies-based model.

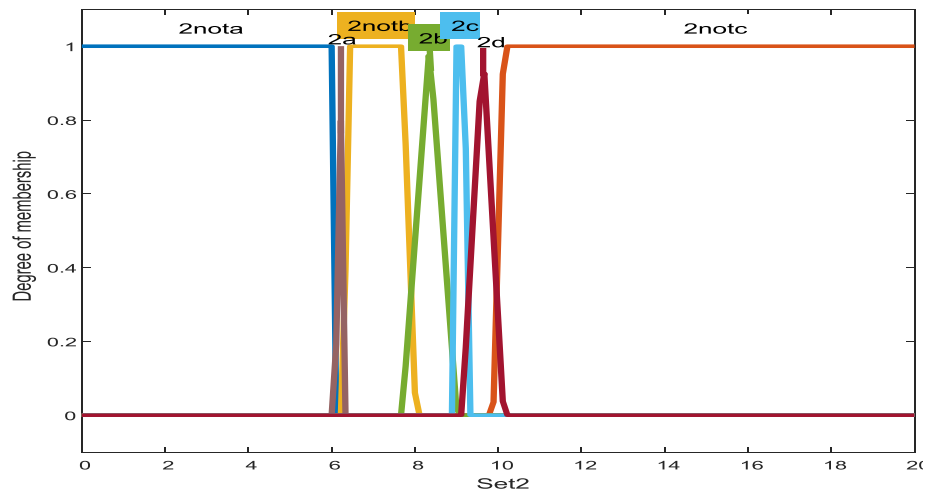


Figure 80: Membership functions of Set Two for the comedies-based model.

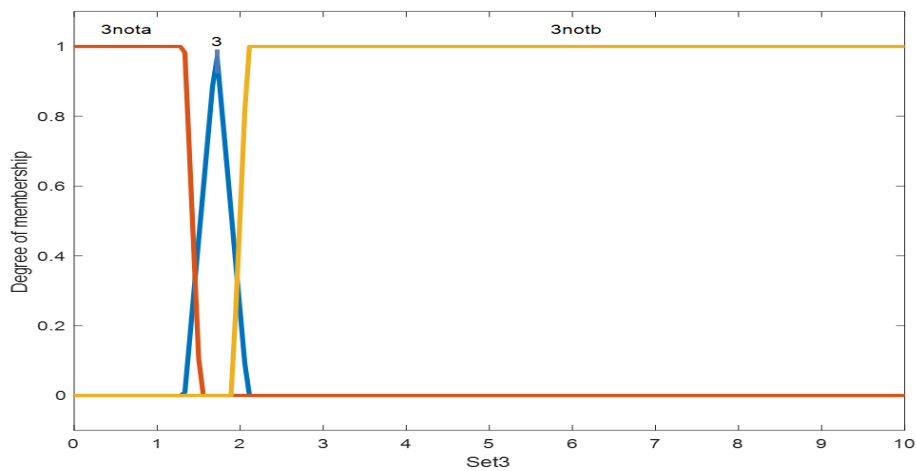


Figure 81: Membership functions of Set Three for the comedies-based model.

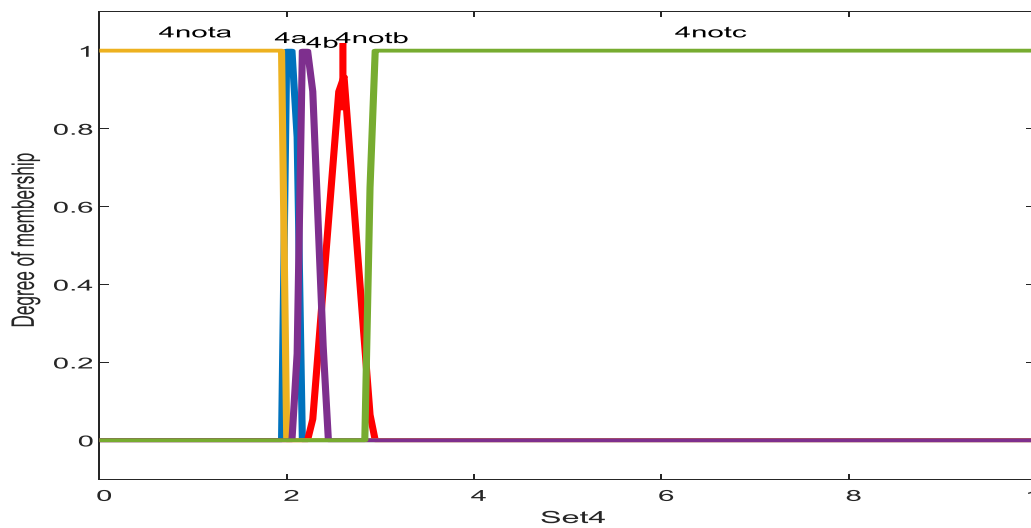


Figure 82: Membership functions of Set Four for the comedies-based model.

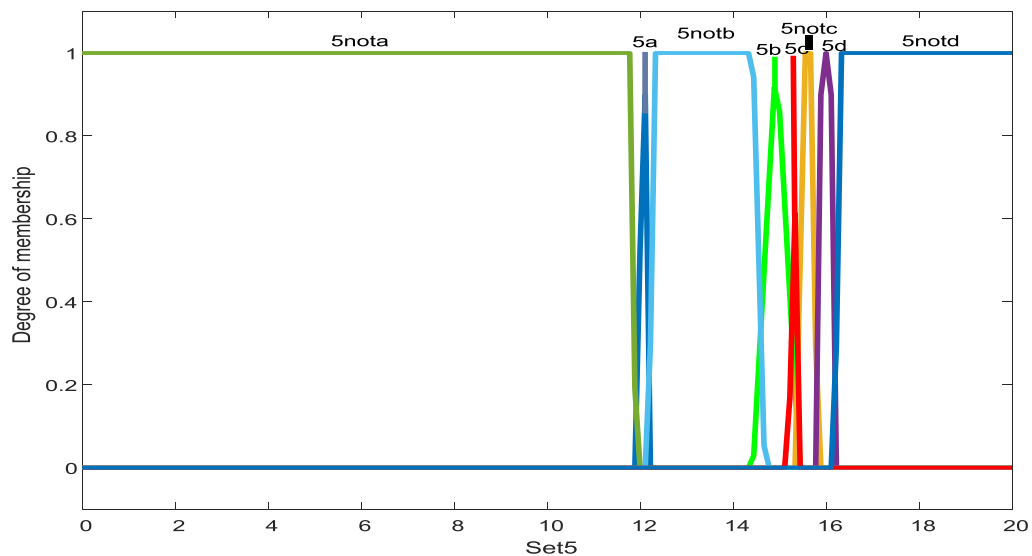


Figure 83: Membership functions of Set Five for the comedies-based model.

7.3.15 The Concepts of ‘Evidential Intervals’ of Belief and ‘Evidential Intervals’ of Degrees of Shakesperianness Expressed in Linguistic terms.

Many researchers (Giarratano and Riley 2005, 286–88; Kandell 1992, 216–18) refer to the so-called evidential intervals of belief, which can be described in linguistic terms and express the probability of an event happening. In our case, the resulting output SiS interval each time represents evidential degrees of Shakesperianness. Through this kind of interval it is also possible to describe qualitatively with probabilistic formulations a new plays’ SiS interval-based score though not employing the term belief in its full probabilistic mode. The

focus in our experimentation is on the general belief in an interval of degrees of Shakespeariness. But though the belief in the core experimentation (Chapter Four) is not identified with the explicit expression of likelihood and network representations (Pearl 1988, 12–16, 50–51), implicitly qualitative comparative relationships (kind of probabilistic conclusions) come up with the detection of a play's interval of Shakespeariness.

Let us proceed to the explanation. Each SiS score constitutes by itself a discreet output of the fuzzy expert system and the two output SiS scores jointly form together an output interval. If the output SiS interval (of degrees of Shakespeariness) for a play under scrutiny is of 0.1-0.25, then it can be argued that this SiS1-Low interval expresses the fact that it is 'almost certain' that the input play is not of Shakespearian style. If the SiS interval is 0.26 to 0.49 it is found that the play's stylistic tendency is not of Shakespearian style. In addition, if a play's SiS interval is exactly 0.50 it can be claimed that the play is of 'medium' Shakespeariness. If it is of 0.51-0.75 then the play's features show a tendency of Shakespearian style; if the SiS interval is 0.76-0.9, then it can be argued that it is 'almost certain' the play under scrutiny is of Shakespearian style. And, of course, if the resulting SiS interval is near 1 (0.91-1) the implication is that it is close to 'absolutely certain' that the play is of Shakespearian style. If the SiS interval is near 0 (0.01-0.09) it can be claimed it is close to 'absolutely uncertain' that the play is of Shakespearian style (or 'absolutely certain that it is not of Shakespearian style').

There are also other cases, such as that of *Edward III*, where the output SiS interval is interpreted as indicative of undetermined Shakespeariness because the two limits are very distant. Even though the notion of probabilities ('an event can or cannot occur') and evidential intervals of belief are not identical with membership degrees and partial truths, in case we can model the styles of all possible candidate authors for a specific disputed play, it is feasible to produce probabilistic belief networks for authorship attribution by combining the 'evidence of support', in our case membership in actual classes, and 'evidence of ignorance', membership in 'not/none' classes (Giarratano and Riley 2005, 287–88).

7.3.16 Method of Forming Compact SiS Intervals if SiS Intervals are Larger than 0.25.

By multiplying each limit of a new plays' SiS interval with the exact low cosine index of a play, we can get for a play under testing two new limits and a new SiS interval. (This is the present investigator's method based on fuzzy mathematics). Let us explain it with an example. Say that the produced SiS interval of a play under scrutiny is 0.547 (SiS1)-0.854

(SiS2), an interval whose two limits have a distance that is more than 0.25. We should then proceed to the necessary multiplications:

$$\text{SiS1} \times \text{CosSim} = 0.547 \times 0.867 = 0.474$$

$$\text{SiS2} \times \text{CosSim} = 0.854 \times 0.867 = 0.740$$

Now a second, an updated SiS interval is produced:

$$1) \text{ SiS interval (produced by the fuzzy program)} = 0.547 - 0.854$$

$$2) \text{ Updated SiS interval} = 0.474 - 0.740$$

Then, I apply one max (union) and one min (intersection) fuzzy operation in two new combinations and construct two reshuffled intervals, of which the more compact (interval with the closest limits) will be selected as the final SiS interval for the play under scrutiny. The new available intervals are:

$$1) \min(\text{updated SiS interval}) - \max(\text{SiS interval}) = 0.474 - 0.854$$

$$2) \min(\text{SiS interval}) - \max(\text{updated SiS interval}) = 0.547 - 0.740$$

The final SiS interval for the new play can now be 0.547-0.740 because it is less than 0.25. Why do I proceed to this adaptation? The answer is that if the fuzzy program produces for a play under scrutiny very large intervals, say 0.4 to 0.8, the authorship verdict becomes very uncertain and it tends to be useless. If the exact cosine index of a play is multiplied by each of the two SiS scores we can form a more compact SiS interval by detecting the exact relation between the counts of the Shakespearian sets and the cosine index. Consequently, by proceeding to the minmax process it is feasible to build a new updated, more reasonable and more robust SiS interval.

7.3.17 Algebraic Product and Algebraic Sum.

In the Sugeno-ANFIS system more elaborate fuzzy operations are used. These are built on the basic fuzzy operations of the intersection (\min , \cap), union (\max , \cup) and complement (A'). In fact, for the Sugeno-ANFIS-based learning process, instead of the intersection (\min) or the union (\max) the options of the algebraic product-probabilistic product (Prod) of the inputs'

membership functions are available. (Thus for two inputs A and B or, more succinctly, for an element (x of the Universe X) with a grade of membership in set-class A and another element (x of the same Universe X) with a membership in set-class B their product is $\text{prod}=\mu_A \times \mu_B$ or more formally $\text{prod}=\mu_A(x) * \mu_B(x)$) or the so-called algebraic probabilistic sum ($\text{sum}=\mu_A+\mu_B - \mu_A * \mu_B$ or again more formally $\text{sum}=\mu_A(x) + \mu_B(x) - \mu_A(x) * \mu_B(x)$).

7.3.18 Epoch and Hybrid Learning Algorithm (in ANFIS experimentation).

Epoch is one initial forward and then backward pass of all the training examples for learning purposes. As far as it concerns the training and the learning process, in our actual Sugeno-ANFIS experimentation a two-steps hybrid learning algorithm is employed. In practice what happens is that after the signals' forward pass to the fourth layer the least square method (LSM) is applied and the error in the parameters of the output membership functions is minimised. (It is common to say that LSM looks for the best-fitting curve to the actual data points.) In the second step of the learning process, there is a backwards algorithmic propagated update (a gradient descent) of the premise parameters, that is the parameters of the triangular input membership functions. In other words, the triangular classes of the input variables are redesigned with the application of derivatives.

7.3.19 Explanation of the Effect of 'Overfitting'

In 'overfitting', the attempt to account for a series of data points tries too hard to account for each one exactly. In this example, a complex polynomial equation could describe the red line that joins every data point, but this equation would have little predictive point regarding the relationship between the x and y values for data points not present in the graph. This is because the equation and its line focus on the detail of the data points that we have and do not capture the general trend that probably underlies them.

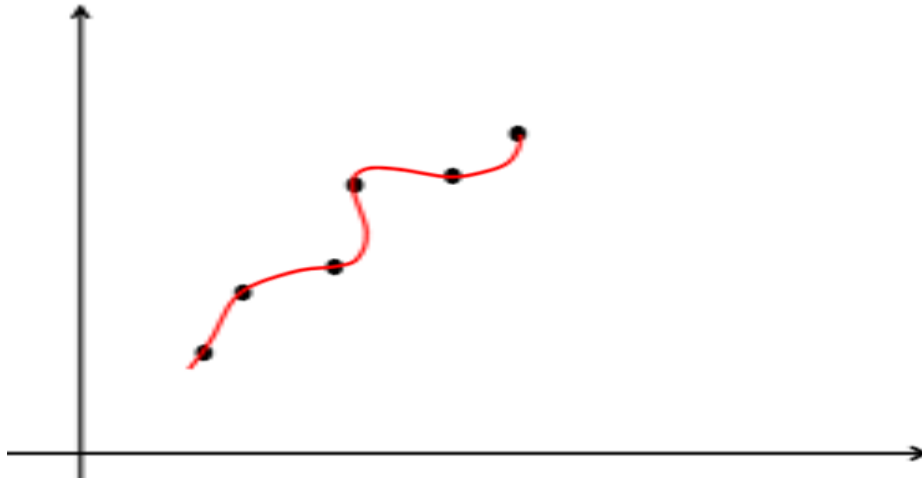


Figure 84: ‘Overfitting’ in the form of a complex polynomial equation.

Although it does not account so precisely for each of the data points we have, this blue line is probably better than the red line in the above picture at capturing the underlying relationship between x and y values in this graph.

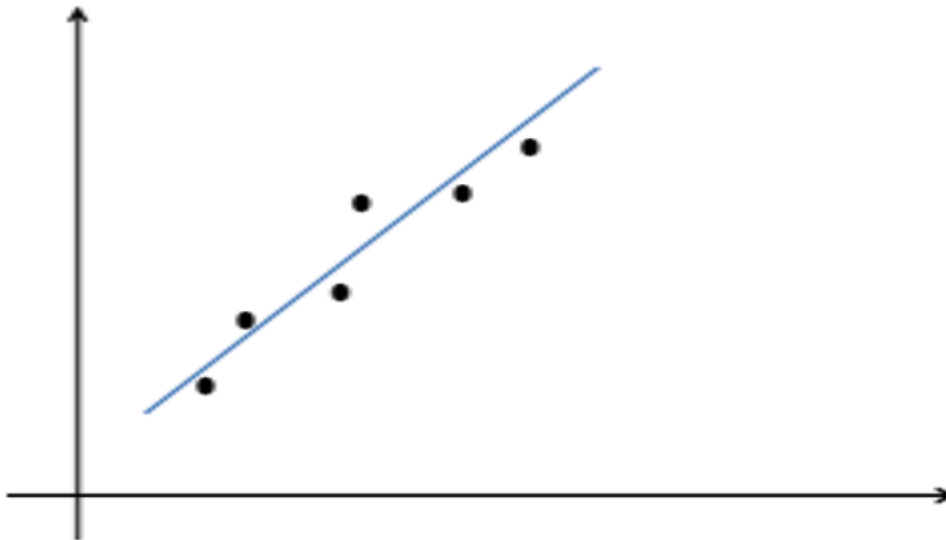


Figure 85: ‘Overfitting’ in terms of a linear relation.

On the assumption that the underlying relationship is probably simple and linear, and that the data points' deviations from the line are probably only random fluctuations, not real facts about the x/y relationship, the blue line (and the equation it represents) are likely to have better predictive power for points not present in the data. These assumptions may of course

be invalid and the red line might more correctly account for the complexity of the x/y relationship; but in many real-world applications the presumption of simplicity is valid.

7.3.20 Script of Histories-Based Fuzzy Stylistic Classifier.

```
[System]
Name='37 R HISTORIES f classifier SiS1-2'
Type='mamdani'
Version=2.0
NumInputs=7
NumOutputs=2
NumRules=37
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'
```

```
[Input1]
Name='Set1'
Range=[0 10]
NumMFs=5
MF1='1a': 'trapmf',[1.64 1.7 1.8 1.86]
MF2='1b': 'trimf',[1.936 2.167 2.398]
MF3='1nota': 'trapmf',[0 0 1.58 1.65]
MF4='1notb': 'trapmf',[1.85 1.92 1.936 2.052]
MF5='1notc': 'trapmf',[2.282 2.398 100 100]
```

```
[Input2]
Name='Set2'
Range=[0 20]
NumMFs=10
MF1='2a': 'trimf',[5.78 6.2 6.62]
MF2='2b': 'trimf',[6.27 7.05 7.83]
MF3='2c': 'trapmf',[8.23 8.3 8.4 8.47]
MF4='2d': 'trimf',[8.49 8.6 8.71]
MF5='2nota': 'trapmf',[0 0 5.36 5.79]
MF6='2notb': 'trimf',[7.82 8.16 8.24]
MF7='2notc': 'trapmf',[8.69 8.8 100 100]
MF8='2notd': 'trimf',[6 6.2 6.4]
MF9='2note': 'trimf',[6.6 7.05 7.5]
MF10='2notf': 'trimf',[7.7 7.765 7.83]
```

```
[Input3]
Name='Set3'
Range=[0 10]
```



```

NumMFs=7
MF1='3a': 'trapmf',[1.33 1.4 1.5 1.57]
MF2='3b': 'trapmf',[1.63 1.7 1.8 1.87]
MF3='3c': 'trimf',[2.09 2.2 2.31]
MF4='3nota': 'trapmf',[0 0 1.26 1.34]
MF5='3notb': 'trimf',[1.57 1.6 1.64]
MF6='3notc': 'trapmf',[1.86 1.94 2 2.11]
MF7='3notd': 'trapmf',[2.29 2.4 100 100]

```

```

[Input4]
Name='Set4'
Range=[0 10]
NumMFs=4
MF1='4a': 'trapmf',[2.19 2.3 2.5 2.61]
MF2='4b': 'trapmf',[2.59 2.7 2.9 3.01]
MF3='4nota': 'trapmf',[0 0 2.1 2.21]
MF4='4notb': 'trapmf',[2.99 3.1 100 100]

```

```

[Input5]
Name='Set5'
Range=[0 20]
NumMFs=7
MF1='5a': 'trimf',[11.94 12.15 12.36]
MF2='5b': 'trimf',[14.29 14.4 14.51]
MF3='5c': 'trimf',[15.06 15.2 15.34]
MF4='5nota': 'trapmf',[0 0 11.82 11.97]
MF5='5notb': 'trapmf',[12.23 12.38 14.2 14.31]
MF6='5notc': 'trapmf',[14.49 14.6 14.92 15.07]
MF7='5notd': 'trapmf',[15.33 15.48 100 100]

```

```

[Input6]
Name='SiS1asInput'
Range=[0 1]
NumMFs=4
MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

```

```

[Input7]
Name='CosSim'
Range=[0 1]
NumMFs=2
MF1='LowCos': 'trimf',[0 0.884 0.885]
MF2='HighCos': 'trapmf',[0.884 0.95 1 1]

```

```

[Output1]
Name='SIS'
Range=[0 1]
NumMFs=4

```

```
MF1='SISLOW':trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1':trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2':trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH':trapmf',[0.6875 0.875 1 1]
```

```
[Output2]
Name='SIS2'
Range=[0 1]
NumMFs=4
MF1='SISLOW':trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1':trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2':trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH':trapmf',[0.6875 0.875 1 1]
```

```
[Rules]
2 3 1 2 0 0 0, 4 0 (1) : 1
2 3 2 1 0 0 0, 4 0 (1) : 1
1 1 1 1 0 0 0, 4 0 (1) : 1
1 2 1 1 0 0 0, 4 0 (1) : 1
2 2 3 1 0 0 0, 4 0 (1) : 1
1 1 2 2 0 0 0, 4 0 (1) : 1
1 4 3 2 0 0 0, 4 0 (1) : 1
5 3 1 2 0 0 0, 3 0 (1) : 1
3 1 1 1 0 0 0, 3 0 (1) : 1
1 5 1 1 0 0 0, 1 0 (1) : 1
3 5 1 1 0 0 0, 1 0 (1) : 1
3 1 4 1 0 0 0, 2 0 (1) : 1
3 5 4 1 0 0 0, 1 0 (1) : 1
3 5 1 3 0 0 0, 1 0 (1) : 1
2 7 4 1 0 0 0, 1 0 (1) : 1
2 2 1 2 0 0 0, 4 0 (1) : 1
2 9 1 2 0 0 0, 1 0 (1) : 1
5 2 1 2 0 0 0, 3 0 (1) : 1
5 9 1 2 0 0 0, 1 0 (1) : 1
0 0 0 0 1 4 2, 0 4 (1) : 1
0 0 0 0 6 1 2, 0 1 (1) : 1
0 0 0 0 5 3 2, 0 3 (1) : 1
0 0 0 0 5 3 2, 0 2 (1) : 1
2 2 2 2 0 0 0, 4 0 (1) : 1
1 2 2 1 0 0 0, 4 0 (1) : 1
1 9 2 1 0 0 0, 1 0 (1) : 1
4 2 3 1 0 0 0, 3 0 (1) : 1
4 9 3 1 0 0 0, 1 0 (1) : 1
4 2 6 1 0 0 0, 2 0 (1) : 1
4 9 6 1 0 0 0, 1 0 (1) : 1
4 2 3 2 0 0 0, 3 0 (1) : 1
4 9 3 2 0 0 0, 1 0 (1) : 1
4 2 6 2 0 0 0, 2 0 (1) : 1
4 9 6 2 0 0 0, 1 0 (1) : 1
0 0 0 0 5 4 1, 0 1 (1) : 1
```

0 0 0 0 5 2 2, 0 2 (1) : 1
0 0 0 0 5 2 2, 0 1 (1) : 1

7.3.21 Script of Tragedies-Based Stylistic Classifier.

```
[System]
Name='3--style tragedies fuzzy classifier SiS1-2'
Type='mamdani'
Version=2.0
NumInputs=7
NumOutputs=2
NumRules=36
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'
```

```
[Input1]
Name='Set1'
Range=[0 10]
NumMFs=7
MF1='1a': 'trimf', [1.79 1.9 2.01]
MF2='1b': 'trimf', [1.988 2.24 2.497]
MF3='1c': 'trimf', [2.49 2.6 2.71]
MF4='1nota': 'trapmf', [0 0 1.7 1.81]
MF5='1notb': 'trapmf', [1.988 1.989 2 2.11]
MF6='1notc': 'trapmf', [2.37 2.38 2.4 2.497]
MF7='1notd': 'trapmf', [2.496 2.642 100 100]
```

```
[Input2]
Name='Set2'
Range=[0 20]
NumMFs=3
MF1='2': 'trimf', [7.212 8.34 9.476]
MF2='2nota': 'trapmf', [0 0 7.212 7.778]
MF3='2notb': 'trapmf', [8.91 9.476 100 100]
```

```
[Input3]
Name='Set3'
Range=[0 10]
NumMFs=7
MF1='3a': 'trimf', [1.19 1.3 1.41]
MF2='3b': 'trapmf', [1.442 1.5 1.6 1.657]
MF3='3c': 'trapmf', [1.742 1.8 1.9 1.957]
MF4='3d': 'trimf', [2.158 2.3 2.441]
MF5='3nota': 'trapmf', [0 0 1.1 1.21]
MF6='3notb': 'trapmf', [1.956 2.014 2.017 2.159]
```

MF7='3notc': 'trapmf',[2.44 2.58 100 100]

[Input4]

Name='Set4'

Range=[0 10]

NumMFs=3

MF1='4': 'trimf',[2.344 2.778 3.211]

MF2='4nota': 'trapmf',[0 0 2.344 2.561]

MF3='4notb': 'trapmf',[2.994 3.211 100 100]

[Input5]

Name='Set5'

Range=[0 20]

NumMFs=8

MF1='5a': 'trimf',[13.72 14.475 15.22]

MF2='5b': 'trapmf',[15.2 15.3 15.5 15.6]

MF3='5c': 'trimf',[15.69 16.1 16.51]

MF4='5nota': 'trapmf',[0 0 13.72 14.1]

MF5='5notb': 'trapmf',[14.85 15.04 15.1 15.2]

MF6='5notc': 'trimf',[15.6 15.85 16.1]

MF7='5notd': 'trimf',[16.11 16.31 16.5]

MF8='5note': 'trapmf',[16.51 17.01 100 100]

[Input6]

Name='SiS1asInput'

Range=[0 1]

NumMFs=4

MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]

MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]

MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]

MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

[Input7]

Name='CosSim'

Range=[0 1]

NumMFs=2

MF1='LowCos': 'trimf',[0 0.919 0.92]

MF2='HighCos': 'trapmf',[0.919 0.96 1 1]

[Output1]

Name='SIS'

Range=[0 1]

NumMFs=4

MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]

MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]

MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]

MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

[Output2]

Name='SIS2'

```

Range=[0 1]
NumMFs=4
MF1='SISLOW': 'trapmf',[0 0.125 0.25 0.26]
MF2='SISMED1': 'trapmf',[0.1875 0.3725 0.5 0.51]
MF3='SISMED2': 'trapmf',[0.4375 0.625 0.75 0.76]
MF4='SISHIGH': 'trapmf',[0.6875 0.875 1 1]

```

```

[Rules]
2 1 4 1 0 0 0, 4 0 (1) : 1
1 1 1 1 0 0 0, 4 0 (1) : 1
2 1 2 1 0 0 0, 4 0 (1) : 1
2 1 3 1 0 0 0, 4 0 (1) : 1
3 1 4 1 0 0 0, 4 0 (1) : 1
4 1 1 1 0 0 0, 3 0 (1) : 1
1 2 1 1 0 0 0, 1 0 (1) : 1
4 2 1 1 0 0 0, 1 0 (1) : 1
4 1 5 1 0 0 0, 2 0 (1) : 1
4 2 5 1 0 0 0, 1 0 (1) : 1
4 2 1 1 0 0 0, 1 0 (1) : 1
2 1 4 3 0 0 0, 3 0 (1) : 1
6 1 4 3 0 0 0, 2 0 (1) : 1
6 1 4 1 0 0 0, 3 0 (1) : 1
1 3 1 3 0 0 0, 1 0 (1) : 1
1 3 1 1 0 0 0, 1 0 (1) : 1
1 1 1 3 0 0 0, 3 0 (1) : 1
6 1 3 1 0 0 0, 3 0 (1) : 1
5 1 2 1 0 0 0, 3 0 (1) : 1
5 1 2 2 0 0 0, 2 0 (1) : 1
2 1 2 2 0 0 0, 2 0 (1) : 1
3 2 6 1 0 0 0, 1 0 (1) : 1
3 1 6 1 0 0 0, 3 0 (1) : 1
0 0 0 0 1 1 1, 0 1 (1) : 1
0 0 0 0 1 2 1, 0 1 (1) : 1
0 0 0 0 1 3 1, 0 1 (1) : 1
3 1 2 2 0 0 0, 3 0 (1) : 1
2 1 1 1 0 0 0, 4 0 (1) : 1
2 1 1 2 0 0 0, 3 0 (1) : 1
0 0 0 0 1 4 2, 0 4 (1) : 1
0 0 0 0 1 3 2, 0 4 (1) : 1
2 1 4 2 0 0 0, 3 0 (1) : 1
0 0 0 0 1 4 1, 0 1 (1) : 1
0 0 0 0 1 3 1, 0 1 (1) : 1
0 0 0 0 5 4 1, 0 1 (1) : 1
0 0 0 0 5 3 1, 0 1 (1) : 1

```

7.3.22 Script of a Trained Sugeno-ANFIS Classifier.

```

[System]
Name='ANFIS 72 Stylistic Classifier'

```

Type='sugeno'
Version=2.0
NumInputs=5
NumOutputs=1
NumRules=72
AndMethod='prod'
OrMethod='probor'
ImpMethod='prod'
AggMethod='sum'
DefuzzMethod='wtaver'

[Input1]
Name='input1'
Range=[1.7 2.6]
NumMFs=2
MF1='in1mf1': 'trimf',[0.8 1.82546923659109 2.56123859006031]
MF2='in1mf2': 'trimf',[1.88949128333065 2.7189808661727 3.5]

[Input2]
Name='input2'
Range=[5.9 9.9]
NumMFs=3
MF1='in2mf1': 'trimf',[3.9 5.90918949335825 7.90220680143771]
MF2='in2mf2': 'trimf',[5.90785941482862 7.86848355970716 9.84184704869518]
MF3='in2mf3': 'trimf',[7.97655077337991 9.85817247985928 11.9]

[Input3]
Name='input3'
Range=[1.1 2.4]
NumMFs=2
MF1='in3mf1': 'trimf',[-0.2 0.945792517241339 2.2699160811336]
MF2='in3mf2': 'trimf',[1.2035920150414 2.22004782542378 3.7]

[Input4]
Name='input4'
Range=[1.5 3.1]
NumMFs=2
MF1='in4mf1': 'trimf',[-0.1 1.55809547309618 3.04862792459937]
MF2='in4mf2': 'trimf',[1.55237944125349 3.15599303386188 4.7]

[Input5]
Name='input5'
Range=[12 16.5]
NumMFs=3
MF1='in5mf1': 'trimf',[9.75 12.0134903296061 14.2821698474086]
MF2='in5mf2': 'trimf',[12.0021872537885 14.2255636581562 16.3629188418431]
MF3='in5mf3': 'trimf',[14.4108167597317 16.4610913103153 18.75]

[Output1]
Name='output'

Range=[0.144 0.87]
 NumMFs=72
 MF1='out1mf1': 'linear', [0.00219780101535946 0.0743111085365907
 0.0112109074090125 0.0141030472731687 0.102340890696232 0.00830902312701587]
 MF2='out1mf2': 'linear', [0.00394045958971887 0.016567651887886
 0.0038999159922368 0.00473332773699406 0.0292207723720127
 0.00224390016513438]
 MF3='out1mf3': 'linear', [-3.84387749101343e-05 -0.000118391426772647 -
 3.0751019921978e-05 -3.99763259443572e-05 -0.00022755754774674 -
 1.53755099574363e-05]
 MF4='out1mf4': 'linear', [-0.0223611286699938 0.0107152545382739 -
 0.000152914751756291 -0.0058222937017728 -0.0152049182856276 -
 0.0015426367976315]
 MF5='out1mf5': 'linear', [0.00487697377951409 0.0180329688168252
 0.00358575127228144 0.00679107088101502 0.0336571798122697
 0.00260867276170667]
 MF6='out1mf6': 'linear', [-8.34312891541747e-05 -0.000256968370511479 -
 6.67450313514908e-05 -8.67685409508753e-05 -0.00049391323205103 -
 3.33725156757454e-05]
 MF7='out1mf7': 'linear', [0.00349928571773824 0.04419811459333 0.0110207360392164
 0.0100905266484568 0.0696651387073142 0.00555292271433873]
 MF8='out1mf8': 'linear', [-0.0101201487003835 -0.033713480598409 -
 0.0122505225702145 -0.0144888666879676 -0.0704417149868003 -
 0.00470562842717861]
 MF9='out1mf9': 'linear', [-0.000147746225845902 -0.000455058376132578 -
 0.000118196980539033 -0.000153656074796935 -0.000874657656125155 -
 5.90984902695164e-05]
 MF10='out1mf10': 'linear', [0.00749379154435054 0.0646687408938618
 0.0232282023005395 0.0328150629001438 0.132200547593271 0.0112121478504183]
 MF11='out1mf11': 'linear', [-0.0342745885066466 -0.122616308456759 -
 0.0391519049287029 -0.0451219088525274 -0.240552288041992 -
 0.0169910104746208]
 MF12='out1mf12': 'linear', [-0.000320682906906081 -0.000987703354926095 -
 0.000256546325365731 -0.00033351022332434 -0.00189844281368068 -
 0.000128273162682865]
 MF13='out1mf13': 'linear', [-0.0154535526275362 -0.0793284415232496 -
 0.00327721062993697 0.00652507805353186 -0.0915341270949592 -
 0.00304482913939159]
 MF14='out1mf14': 'linear', [0.0223702275317422 0.000645774935344116
 0.0342911022079515 0.0978373819796616 0.155144486104701 0.00439869727484427]
 MF15='out1mf15': 'linear', [0.0214287190844131 0.0918456461605712
 0.0156158995890375 0.0224063510586858 0.151296615586321 0.0105315871394764]
 MF16='out1mf16': 'linear', [0.0244108414518413 0.0949968622952917
 0.0197812121213087 0.0377839834816422 0.176972900230218 0.0158270965828602]
 MF17='out1mf17': 'linear', [-0.010836185026897 -0.0866769217545988
 0.0615872055058767 -0.114353704352918 -0.150279605814219 -
 0.00975185416302159]
 MF18='out1mf18': 'linear', [0.0150478742294653 0.0342126772943787
 0.00263773041086559 -0.00153343140474657 0.0503648512166825
 0.00129542930569052]

MF19='out1mf19': 'linear', [0.0131604318106753 0.0532902090673611
 0.0122471937151128 0.0165143429317797 0.095212177636932 0.00836005026278422]
 MF20='out1mf20': 'linear', [-0.0280670013602668 -0.146897663006677 -
 0.0860120031183109 -0.00786226027939785 -0.26883892837807 -
 0.0189483620203051]
 MF21='out1mf21': 'linear', [-0.0533345851670315 -0.200440466519247 -
 0.056170083981207 -0.0617349961850719 -0.371680132091412 -0.0230968614525604]
 MF22='out1mf22': 'linear', [-0.008483972672815 -0.0242872775270593 -
 0.010412645160906 -0.0126232922031656 -0.0558071875119544 -
 0.00306245709383396]
 MF23='out1mf23': 'linear', [0.0755415227370719 0.262175903450876 -
 0.0332603251214254 0.0662262979897491 0.370683399234348 0.0282944529135813]
 MF24='out1mf24': 'linear', [-0.035141808687633 -0.170797087590542 -
 0.0532455438710715 -0.042941578133464 -0.302126018411991 -0.0228740483239469]
 MF25='out1mf25': 'linear', [-0.034008275269064 -0.135461644764182 -
 0.0199141024027024 -0.0314699659792572 -0.220853988631666 -
 0.0163263198600448]
 MF26='out1mf26': 'linear', [-0.0181494896664061 -0.0770911401435668
 0.0176681046710265 -0.0278489520600531 -0.105421476726494 -
 0.0107687981353293]
 MF27='out1mf27': 'linear', [0.0115383332272227 0.0640495408107655
 0.0134367901669729 0.0104840600255957 0.0995087240221565
 0.00730722884928297]
 MF28='out1mf28': 'linear', [-0.0110000799888404 -0.0446377475077812 -
 0.00684539415774623 -0.0107079594468458 -0.0731911811437518 -
 0.00537804186153571]
 MF29='out1mf29': 'linear', [-0.0230176035692057 -0.135062091436725
 0.0155144483142414 -0.0330224124539532 -0.175587659023554 -
 0.0142186783150369]
 MF30='out1mf30': 'linear', [0.0411243738115626 0.165141804318262
 0.031007863837333 0.059754079670981 0.297028121447896 0.0192981613035033]
 MF31='out1mf31': 'linear', [-0.00111874017547319 -0.00515863527183565 -
 0.000994435710213272 -0.00149165356544941 -0.00876346467283384 -
 0.000621522320406969]
 MF32='out1mf32': 'linear', [0.050107495968839 0.212953908332882
 0.00929785457782511 0.0429007445189738 0.31526000363697 0.021374884356815]
 MF33='out1mf33': 'linear', [-0.0471886877851266 -0.163035000901789 -
 0.0475213185112801 -0.0471526765802302 -0.304897683690161 -
 0.0183328944660487]
 MF34='out1mf34': 'linear', [-0.00135886646796885 -0.0062658842802174 -
 0.00120788130040886 -0.0018118219540715 -0.010644453958791 -
 0.000754925811737379]
 MF35='out1mf35': 'linear', [0.0422112400186766 0.136570029923971 -
 0.00757767027334151 0.0371495773449331 0.208353176914651 0.0151487999993088]
 MF36='out1mf36': 'linear', [0.0377237970545869 0.122335400722987
 0.0209219158149917 0.0611826388230419 0.242163751792677 0.0136252220258593]
 MF37='out1mf37': 'linear', [-0.00039144394302937 -0.00144440621651881 -
 0.000575471768678911 -0.000726077697644675 -0.00313739961384359 -
 0.000332678973732045]

MF38='out1mf38': 'linear', [0.00184702897748743 0.00583861650966977
 0.000829154768967783 0.0023120811370118 0.0108268813952312
 0.000838178984144642]
 MF39='out1mf39': 'linear', [-0.000339912379666796 -0.0010469301304698 -
 0.000271929904030785 -0.000353508874717606 -0.00201228128556874 -
 0.000135964952015392]
 MF40='out1mf40': 'linear', [0.00127532535695654 0.00359756527464658
 0.00023209859072186 0.00107864697423263 0.00618363614160255
 0.000289899569628902]
 MF41='out1mf41': 'linear', [0.00686309824763231 0.0215641004566368
 0.00355186853769003 0.00839265422094944 0.0403717214750641
 0.00306114029153499]
 MF42='out1mf42': 'linear', [-0.000737779186008813 -0.00227235989435726 -
 0.000590223348561264 -0.000767290354675043 -0.00436765278149198 -
 0.000295111674394319]
 MF43='out1mf43': 'linear', [-0.00067373416246087 -0.00232004039961382 -
 0.000662844023019318 -0.000952958799117247 -0.00460957737968672 -
 0.000391163515925267]
 MF44='out1mf44': 'linear', [-0.00444935541318473 -0.0138650997807365 -
 0.00384859044332346 -0.00462879589923289 -0.0267918415251408 -
 0.00179806697198452]
 MF45='out1mf45': 'linear', [-0.00123981843560621 -0.00382141119261642 -
 0.000983646148808524 -0.00128694859105786 -0.00733182436299628 -
 0.000496953449435266]
 MF46='out1mf46': 'linear', [-0.000605781929558355 -0.00228770971932214 -
 0.000826689417792533 -0.000991744832539975 -0.0047119259102481 -
 0.000430715126627467]
 MF47='out1mf47': 'linear', [-0.00870337280860497 -0.027112413285899 -
 0.0077427601519438 -0.00893164829184536 -0.0524901945151216 -
 0.00348995919653433]
 MF48='out1mf48': 'linear', [-0.00252470849926452 -0.00778902379951281 -
 0.00198148049247874 -0.0026142109449316 -0.0149094237309498 -
 0.0010146691874071]
 MF49='out1mf49': 'linear', [-0.0157283152729586 -0.0722871436396834 -
 0.0110264027670691 -0.012051296703146 -0.111093158089708 -
 0.00768237230510237]
 MF50='out1mf50': 'linear', [0.00944163599472132 0.0559317907619507
 0.0249833641704603 0.0447971877434244 0.13515397876586 0.00704305108698302]
 MF51='out1mf51': 'linear', [0.025873976445016 0.0873905305091958
 0.0160594203006807 0.0203966815982445 0.149720608953638 0.00983063560279142]
 MF52='out1mf52': 'linear', [0.00660960947067935 0.0169731873791813
 0.00314665099496247 0.00862169986112565 0.0353511476930337
 0.0027977436095236]
 MF53='out1mf53': 'linear', [0.00833746338265965 0.0849032516115941
 0.0310824621469055 -0.0192155627846485 0.105107614978768 0.00405159521059815]
 MF54='out1mf54': 'linear', [-0.00503010791210567 0.0244309142544871 -
 0.0343007226314386 0.00294662357041627 -0.0119532926402874
 0.00239651753904882]

MF55='out1mf55': 'linear', [0.000405312170326741 0.00115446956985191
 0.000125796156524949 0.000417437929402793 0.00210301582463205
 0.000144923802725703]
 MF56='out1mf56': 'linear', [-0.00381661623570793 0.015528853728802 -
 0.023331794021164 0.00978399594819241 -0.00183556028110435 -
 0.000568084031300179]
 MF57='out1mf57': 'linear', [0.0257798635316876 0.0360392677915078
 0.00911382449189952 0.00550240813339923 0.076435363571511
 0.00510510380115109]
 MF58='out1mf58': 'linear', [0.00163164486996145 0.0048712049681584
 0.000746539362391497 0.00180114060878316 0.00905052986734
 0.000649482382473828]
 MF59='out1mf59': 'linear', [-0.0221016143854336 0.00232703978873953 -
 0.0373115128149002 -0.0224202195730944 -0.0795063071243896 -
 0.00410232468899601]
 MF60='out1mf60': 'linear', [0.00532280602596493 0.00594953639572874
 0.0714413469491136 0.0221722009434055 0.104885890770623 0.00401198564690331]
 MF61='out1mf61': 'linear', [-0.0125050260775039 -0.0496828579341865 -
 0.00720207049445323 -0.0115347413183587 -0.0809246958412272 -
 0.00600172541174049]
 MF62='out1mf62': 'linear', [-0.0106633918026695 -0.0423946159769897
 0.0194192587534589 -0.0173309808468021 -0.0509697301229672 -
 0.00411353541707305]
 MF63='out1mf63': 'linear', [0.00882258161874377 0.00542970779433153
 0.0117727867720173 0.000145752884304295 0.0261708294489141
 0.00145913229538715]
 MF64='out1mf64': 'linear', [-0.00373953044245737 -0.0147800489199122 -
 0.00213687454228491 -0.0033833846920098 -0.0240398386070162 -
 0.00178072878488858]
 MF65='out1mf65': 'linear', [0.0402151353205375 0.0997618479673321
 0.0519425481179303 0.0593610129745811 0.251280543875786 0.0161543144632321]
 MF66='out1mf66': 'linear', [-0.032346018429061 -0.12156382234854 -
 0.0216643251520445 -0.0338759919530583 -0.209450157701169 -
 0.0118706624758722]
 MF67='out1mf67': 'linear', [0 0 0 0 0 0]
 MF68='out1mf68': 'linear', [0.0469137462170517 0.186116736235438
 0.0155714958197306 0.0459564187057895 0.29455839722185 0.0206421482961807]
 MF69='out1mf69': 'linear', [0.00145844094159865 -0.0507885694688366 -
 0.0257037580882046 -0.0138228872755156 -0.088856773787364 -
 0.00562214924693681]
 MF70='out1mf70': 'linear', [0 0 0 0 0 0]
 MF71='out1mf71': 'linear', [0.0507374197631515 0.196018440838011
 0.0104147868166457 0.0573835679443241 0.314554215707355 0.0230643157749736]
 MF72='out1mf72': 'linear', [-0.00514649710185154 0.0375607629596208
 0.0133430552453368 0.0159908614428212 0.0617481816961527
 0.00160111565679075]

[Rules]

1 1 1 1 1, 1 (1) : 1

1 1 1 1 2, 2 (1) : 1

1 1 1 1 3, 3 (1) : 1
 1 1 1 2 1, 4 (1) : 1
 1 1 1 2 2, 5 (1) : 1
 1 1 1 2 3, 6 (1) : 1
 1 1 2 1 1, 7 (1) : 1
 1 1 2 1 2, 8 (1) : 1
 1 1 2 1 3, 9 (1) : 1
 1 1 2 2 1, 10 (1) : 1
 1 1 2 2 2, 11 (1) : 1
 1 1 2 2 3, 12 (1) : 1
 1 2 1 1 1, 13 (1) : 1
 1 2 1 1 2, 14 (1) : 1
 1 2 1 1 3, 15 (1) : 1
 1 2 1 2 1, 16 (1) : 1
 1 2 1 2 2, 17 (1) : 1
 1 2 1 2 3, 18 (1) : 1
 1 2 2 1 1, 19 (1) : 1
 1 2 2 1 2, 20 (1) : 1
 1 2 2 1 3, 21 (1) : 1
 1 2 2 2 1, 22 (1) : 1
 1 2 2 2 2, 23 (1) : 1
 1 2 2 2 3, 24 (1) : 1
 1 3 1 1 1, 25 (1) : 1
 1 3 1 1 2, 26 (1) : 1
 1 3 1 1 3, 27 (1) : 1
 1 3 1 2 1, 28 (1) : 1
 1 3 1 2 2, 29 (1) : 1
 1 3 1 2 3, 30 (1) : 1
 1 3 2 1 1, 31 (1) : 1
 1 3 2 1 2, 32 (1) : 1
 1 3 2 1 3, 33 (1) : 1
 1 3 2 2 1, 34 (1) : 1
 1 3 2 2 2, 35 (1) : 1
 1 3 2 2 3, 36 (1) : 1
 2 1 1 1 1, 37 (1) : 1
 2 1 1 1 2, 38 (1) : 1
 2 1 1 1 3, 39 (1) : 1
 2 1 1 2 1, 40 (1) : 1
 2 1 1 2 2, 41 (1) : 1
 2 1 1 2 3, 42 (1) : 1
 2 1 2 1 1, 43 (1) : 1
 2 1 2 1 2, 44 (1) : 1
 2 1 2 1 3, 45 (1) : 1
 2 1 2 2 1, 46 (1) : 1
 2 1 2 2 2, 47 (1) : 1
 2 1 2 2 3, 48 (1) : 1
 2 2 1 1 1, 49 (1) : 1
 2 2 1 1 2, 50 (1) : 1
 2 2 1 1 3, 51 (1) : 1
 2 2 1 2 1, 52 (1) : 1

2 2 1 2 2, 53 (1) : 1
2 2 1 2 3, 54 (1) : 1
2 2 2 1 1, 55 (1) : 1
2 2 2 1 2, 56 (1) : 1
2 2 2 1 3, 57 (1) : 1
2 2 2 2 1, 58 (1) : 1
2 2 2 2 2, 59 (1) : 1
2 2 2 2 3, 60 (1) : 1
2 3 1 1 1, 61 (1) : 1
2 3 1 1 2, 62 (1) : 1
2 3 1 1 3, 63 (1) : 1
2 3 1 2 1, 64 (1) : 1
2 3 1 2 2, 65 (1) : 1
2 3 1 2 3, 66 (1) : 1
2 3 2 1 1, 67 (1) : 1
2 3 2 1 2, 68 (1) : 1
2 3 2 1 3, 69 (1) : 1
2 3 2 2 1, 70 (1) : 1
2 3 2 2 2, 71 (1) : 1
2 3 2 2 3, 72 (1) : 1

Works cited

- Abbasi, Ahmed, and Hsinchun Chen. 2005. "Applying Authorship Analysis to Extremist-Group Web Forum Messages" 20 (5).
<https://arizona.pure.elsevier.com/en/publications/applying-authorship-analysis-to-extremist-group-web-forum-message>.
- Amira, Aloui, and Grissa Touzi Amel. 2015. "An Extension Of Protégé For An Automatic Fuzzy-Ontology Building Using Clustering And Fca." *International Journal of Computer Science & Information Technology (IJCSIT)* 7 (2).
<http://airccse.org/journal/jcsit/7215ijcsit02.pdf>.
- Archer, D., T. McEnery, P. Rayson, and A. Hardie. 2003. "Developing an Automated Semantic Analysis System for Early Modern English." In *Proceedings of Corpus Linguistics*, edited by D. Archer, P. Rayson, A. Wilson, and T. Mcenery, 16:22–31. Lancaster, U.K: Lancaster University.
- Aristotle. Est.: 335-323 AD. *Περὶ Ποιητικῆς (Poetics)*. Edited by Alexandros Galhnos (n.d). 'Απαντα Των Αρχαίων Ελλήνων Συγγραφέων 33. Athens: Πάπυρος (Papyrus).
- Asquith, Paul. 2018. "The Problem With Coffee Cup Sizes."
<https://www.sevenmiles.com.au/editorial/problem-coffee-cup-sizes/>. Accessed 1 March 2020.
- Ayer, Alfred J. 1978. *Logical Positivism*. Westport, Connecticut: Greenwood Press.
- Baron, A., and P. Rayson. 2008. "VARD 2: A TOOL for Dealing with Spelling Variation in Historical Corpora." In *Proceedings of the Postgraduate Conference in Corpus Linguistics*. Birmingham, U.K: Aston University.
- Bee, E. Ronald. 1973. "The Use of Statistical Methods in Old Testament Studies." *Vetus Testamentum* 23 (Fasc. 3): 257–72.
- Boas, Frederick S. 1910. *Shakspeare and His Predecessors*. Third. London: John Murray, Albermarle.
- Boas, Frederick S. 1960. *Christopher Marlowe: A Biographical and Critical Study*. Oxford: The Clarendon Press.
- Bourus, Terri, and Gary Taylor. 2013. *The Creation and Re-Creation of "Cardenio"*. New York: Palgrave Macmillan.
- Britannica. n.d. In. <https://www.britannica.com/biography/Frederick-James-Furnivall>.

- Brooker, Tucker. 1912. *The Tudor Drama Shakespeare: A History of English National Drama to the Retirement of Shakespeare*. London: Constable & Co.
- Burrows, John. 2002. "Delta: Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17(3): 267–87.
<https://doi.org/10.1093/lc/17.3.267>.
- . 2005. "Who Wrote *Shamela*? Verifying the Authorship of a Parodic Text." *Literary and Linguistic Computing* 20: 437–50.
- . 2007. "All the Way Through: Testing for Authorship in Different Frequency Strata." *Literary and Linguistic Computing* 22 (1): 27–47. <https://doi.org/10.1093/lc/fqi067>.
- Burrows, John, and Hugh Craig, eds. 2017. "The Joker in the Pack? Marlowe, Kyd, and the Co-Authorship of *Henry VI, Part 3*." In *The New Oxford Shakespeare: Authorship Companion* edited by Gary Taylor and Gabriel Egan, 194–217. Oxford: Oxford: Oxford University Press.
- Byrne, St Clare Muriel. 1932. "Bibliographical Clues In Collaborate Plays." *Library (London)* s4-XIII (1): 21–4: 21–48.
- Callan, Robert. 1998. *The Essence of Neural Networks*. London: Essence of Computing, Europe Prentice Hall.
- Carvalho, N., V. Minim, M. Nascimento, M. Vidigal, M. Ferreira, A. Conchalves, and A. Minim. 2015. "A Discriminant Function for Validation of the Cluster Analysis and Behavioral Prediction of the Coffee Market." *Food Research International, Elsevier* 77 (3): 400–407. <https://doi.org/10.1016/j.foodres.2015.10.013>.
- Chen, Ye-Sho, and Ferdinand F. Leimkuhler. 1989. "A Type-Token Identity in the Simon-Yule Model of Text." *Journal of the American Society for Information Science and Technology* 40 (1): 45–53.
- Cooke, E. Jacobs. 1961. *The Federalist*. Middletow, Connecticut: Wesleyan University Press.
- Cox, E. 1999. *The Fuzzy Systems Handbook: A Practitioners Guide to Building, Using, and Maintaining Fuzzy Systems*. San Diego, California: AP Professional.
- Craig, Hugh, and Arthur F. Kinney, eds. 2009. *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.
- Danson, Lawrence. 2000. *Shakespeare's Dramatic Genres*. Oxford: Oxford University Press.
- Drake, Bryan. 1972. "Unanswered Questions in Computerised Literary Analysis." *Journal of Biblical Literature* 91 (2): 241–42.

- Dolezel, Lubomil, and Richard W. Bailey. 1969. *Statistical Analysis and Style*. Mathematical Linguistics and Automatic Language Processing 6. New York: American Elsevier.
- Dowden, Edward. 1899. *The Works of Shakespeare, The Tragedy Of "Hamlet"*. London: Methuen and Co.
- Dubois, Didier, Laurent Foulou, Gilles Mauris, and Henri Prade. 2004. "Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities." *Reliable Computing/Kluwer Academic Publishers* 10: 273–297.
- Egan, Gabriel. 2017a. "A History of Shakespearian Authorship Attribution." In *The New Oxford Shakespeare: Authorship Companion*, 27–47. Oxford: Oxford University Press.
- . 2017b. "The Limitations of Vickers's Trigram Tests." In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 60–66. Oxford: Oxford University Press.
- Eisen, Mark, Santiago Segarra, Gabriel Egan, and Alejandro Ribeiro. 2016. "Stylometric Analysis of Early Modern Period English Plays." *Digital Scholarship in the Humanities* 33 (3): 500–528. <https://doi.org/1610.05670v1>.
- Elliot, Jack, and Brett Greatly-Hirsch. 2017. "Arden of Faversham, Shakespearean Authorship, and 'The Print of Many.'" In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 407–16. Oxford: Oxford University Press.
- Encyclopaedia Britannica, ed. 2018. "Georg Cantor." In *Encyclopaedia Britannica*. <https://www.britannica.com/biography/Georg-Ferdinand-Ludwig-Philipp-Cantor#ref199443>.
- Ford, L. Paul, and Edward Bourne. 1897. "The Authorship of the Federalist." *Oxford University Press, The American Historical Review*, 2 (4): 675–87.
- Forsyth, S. 1999. "Stylochronometry with Substrings: Or a Poet Young and Old." *Literary and Linguistic Computing* 11: 163–74.
- Foucault, Michel. 1969. "Qu'est -Ce Qu'un Auteur?" In *What Is an Author?/Qu'est-Ce Qu'un Auteur?*, 3:73–104. Buffalo University, United States of America: Bulletin de la Société française de philosophie and later published in *Textual Strategies*, Ithaca, Cornell University Press, 1979, pp. 141–160. <http://libertaire.free.fr/MFoucault319.html>.
- Fowler, Alistair. 1985. *Kinds of Literature*. Oxford: Clarendon.

- Furnivall, James Frederick, Edwin Abbot, Richard Simpson, B. Nicholson, and John Hales. 1874. "Discussion on Paper II (Mr Fleay's): On Metrical Tests as Applied to Dramatic Poetry." *Ubner & Co* 1–2: 73–85.
- Fyfe, W.H. 1932. *Translation of Aristotle's Poetics*. Vol. 23. Cambridge: Harvard University Press.<http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0056%3Asection%3D1450b>.
- Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics* 114 (3): 739–67.
- Gabaix, Xavier, and M. Yannis Ioannides. 2003. "The Evolution of City Size Distributions. Written for the Handbook of Urban and Regional Economics." *North-Holland Publishing Company* IV. <http://pages.stern.nyu.edu/~xgabaix/papers/evolution.pdf>.
- Giarratano, Joseph C., and Gary D. Riley. 2005. *Expert Systems, Principles and Programming*. Fourth. Boston, Massachusetts: Thomson Course Technology.
- Graham, Graham W. n.d. "Heraclitus." In *Internet Encyclopedia of Philosophy (A Peer-Reviewed Academic Resource)*, edited by James Fieser and Bradley Dowden. <https://www.iep.utm.edu>.
- Greg, W. Walter. 1923. *Shakespeare's Hand in the Play of Sir Thomas More. Papers by Alfred Press W. Pollard, W. W. Greg, J. Dover Wilson, E. Maunde Thompson & R.W. Chamber*, 57-141. Cambridge: Cambridge University Press.
- Gurney, Kevin. 1997. *An Introduction to Neural Networks*. London & New York: UCL Press, Taylor & Francis Group.
- Haque, Ansarul, and Tamjid Rahman. 2014. "Sentiment Analysis by Using Fuzzy Logic." *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)* 4 (1). <https://doi.org/arXiv:1403.3185>.
- "Harvard-IV-4 Psychosocial Dictionary." n.d. In *General Enquirer*. <http://www.wjh.harvard.edu/%7Einquirer/kellystone.htm>.
- Haug, Ralf A. 1940. "The Authorship of *Timon of Athens*." *The Shakespeare Association Bulletin* 15 (4): 227–48.
- Herdan, Gustav. 1960. *Type-Token Mathematics: A Textbook of Mathematical Linguistics*. Hague: Mouton & Co.
- Hecht-Nielsen, Robert. 1989. "Theory of the Backpropagation Neural Network." In *International Joint Conference on Neural Networks*, I: 593–605. Washington, USA: Neural Networks. <https://doi.org/10.1109/IJCNN.1989.118638>.

- Hentenryck, Pascal, and Laurent Michel. 2005. *Constrain-Based Local Search*. Cambridge, Massachussetts: MIT Press.
- Hickson, Samuel. 1847. "Review of: [William Spalding] A Letter on Shakespeare's Authorship of *The Two Noble Kinsmen*, a Drama Commonly Ascribed to John Fletcher (Edinburgh: Adam and Charles Black, 1833); [William Shakespeare] The Pictorial Edition of Shakespeare (London: Charles Knight and Company, 1841); and Alexander Dyce (Ed.) The Works of Beaumont and Fletcher (London: Moxon, 1846)." *Westminster and Foreign Quarterly Review*, Reprinted in the Shakspeare Society's Transactions (1874). Series 1, No 1-2, 48: 59–88.
- Holdsworth, Roger. 2017. "Shakespeare and Middleton: A Chronology for 1605-6." In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 366-384. Oxford: Oxford University Press.
- Holmes, David I. 1985. "The Analysis of Literary Style--A Review." *Journal of the Royal Statistical Society, Series A (General)*, 148 (4): 328–41. <https://doi.org/10.2307/2981893>.
- . 1991. "A Stylometric Analysis of Mormon Scripture and Related Texts." *Journal of the Royal Statistical Society, A*, 155 (1): 91–120.
- Holmes, David I., and S. Forsyth. 1995. "The Federalist Revisited: New Directions in Authorship Attribution." *Literary and Linguistic Computing* 10 (2). <https://doi.org/10.1093/lc/10.2.111>.
- Honore, A. 1979. "Some Simple Measures of Richness of Vocabulary." *Ass. Lit. Ling. Comput. Bull.* 7: 172–77.
- Hoover, David L. 2007. "Quantitative Analysis and Literary Studies, A Companion to Digital Literary Studies." Oxford: Blackwell. <http://www.digitalhumanities.org/companion/>.
- Information Resources Management, ed. 2017. *Artificial Intelligence: Concepts, Methodologies, Tools and Applications*. Hershey, Pennsylvania: IGI Global.
- Jang, J-S.R. 1993. "ANFIS: Adaptive Network Based Fuzzy Inference Systems." *IEEE Transactions on Systems, Man, and Cybernetics* 23 (3): 665–85.
- John, O.P., and L.A Pervin. 1990. "The 'Big Five' Factor Taxonomy: Dimensions of Personality in the Natural Language and in Questionnaires." In *Handbook of Personality: Theory and Research*, 66–100. New York: Guilford Press.
- Juola, Patrick. 2006. "Authorship Attribution." *Foundations and Trends in Information* 1. <https://doi.org/0.1561.15000000005>.
- Kandell, Abraham. 1992. *Fuzzy Expert Systems*. Boca Raton, Florida: CRC Press.

- Kendal, Simon, and Malcolm Green. 2007. *An Introduction to Knowledge Engineering*. London: Springer-Verlag.
- Kinney, Arthur F. 2009a. "Authoring *Arden of Faversham*." In *Shakespeare, Computers, and the Mystery of Authorship*, edited by Hugh Craig and Arthur F. Kinney, 78–99. Cambridge: Cambridge University Press.
- . 2009b. "Transforming *King Lear*." In *Shakespeare, Computers, and the Mystery of Authorship*, edited by Hugh Craig and Arthur F. Kinney, 181–203. Cambridge: Cambridge University Press.
- Klein, Karl. 2001. *Timon of Athens*. The New Cambridge Shakespeare. Cambridge: Cambridge University Press.
- Konchady, Manu. 2006. *Text Mining Application Programming*. Boston, Massachusetts: Charles Rivers Press.
- Kowalzyk, Alexandre. 2014. "SVM in Practice, Tutorial." <https://www.svm-tutorial.com/svm-tutorial/>.
- Krisnamourti, Ramesh. n.d. "Planes in 3-D Descriptive Geometry." Carnegie Mellon University. <https://www.andrew.cmu.edu/user/ramesh/teaching/course/48-175/lectures/4.PlanesInDescriptiveGeometry.pdf>.
- "LabVIEW 2011 PID and Fuzzy Logic Toolkit Help." 2011. National Instruments Corporation. http://zone.ni.com/reference/en-XX/help/370401H-01/lvpidmain/center_of_area/.
- Lamont, J.M. 2005. "The Historical Rise of the English Phrasal Verb." <http://homes.chass.utoronto.ca/~cpercy/courses/6361lamont.html>.
- Langworthy, A Charles. 1931. "Langworthy, A Charles. A Verse-Sentence Analysis of Shakespeare's Plays." *PMLA* 46 (3): 738–51.
- Laurence, Anthony. 2014. *AntWordProfiler*. Faculty of Science and Engineering at Waseda University. <http://www.laurenceanthony.net/software/antwordprofiler/>.
- Liu, Z, and H.X Li. 2005. "A Probabilistic Fuzzy Logic System for Modeling and Control." *IEEE Transactions on Fuzzy Systems* 13: 848–59.
- Malone, Edmond. 1787. "A Dissertation on the Three Parts of *King Henry VI* Tending to Shew That Those Plays Were Not Written Originally." London: Henry Baldwin.
- . 1821a. *William Shakespeare's Plays and Poems With the Corrections and Illustrations of Various Commentator's and an Enlarged History of The Stage*. Vol.

- XVIII. London: C. Baldwin.
<https://babel.hathitrust.org/cgi/pt?id=nyp.33433074972146;view=1up;seq=9>.
- . 1821b. *William Shakespeare's Plays and Poems With the Corrections and Illustrations of Various Commentator's and an Enlarged History of The Stage*. Vol. I. London: C. Baldwin.
- Mamdani, E.H, and S. Assilian. 1975. "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller." *International Journal of Man-Machine Studies* 7 (1): 1–13.
- Margolin, Uri. 2005. "Russian Formalism." In *The Johns Hopkins Guide to Literary Theory and Criticism* edited by Michael Groden, Martin Kreiswirth, and Imre Szeman. Baltimore: The Johns Hopkins University Press.
- Martindale, Colin, and Dean McKenzie. 1995. "On the Utility of Content Analysis in Author Attribution: *The Federalist*." *Computers and the Humanities* 29: 259–70.
- Marsh, Nicholas. 2002. *Shakespeare: Three Problem Plays*. London: Palgrave Macmillan.
- Mathur, Neha, Ivan Glesk, and Arjan Buis. 2016. "Comparison of Adaptive Neuro-Fuzzy Inference System (ANFIS) and Gaussian Processes for Machine Learning (GPML) Algorithms for the Prediction of Skin Temperature in Lower Limb Prostheses." *Medical Engineering & Physics* 38 (10): 1083–89.
- Matthews, A.J Robert, and Thomas Merriam. 1993. "Neural Computation in Stylometry I: An Application to the Works of Shakespeare and Fletcher." *Literary and Linguistic Computing* 8 (4): 203-9. <https://doi.org/doi: 10.1093/lc/8.4.203>.
- Mendenhall, C. Thomas. 1897. "The Characteristic Curves of Composition." *Science* 11: 237–49.
- . 1901. "A Mechanical Solution of a Literary Problem." *The Popular Science Monthly* LX (7): 97–105.
- Merriam, Thomas. 2016. "Feminine Endings and Function Words in *King John*." *Notes and Queries* 63 (3): 402–5. <https://doi.org/10.1093/notesj/gjw128>.
- . 2017. "Is It Time to Re-Think *King John*?" *Digital Scholarship in the Humanities* 32 (3): 591–601.
- . 2018. "A Contrast between Acts I and IV and Acts II and V in *King John*." *Notes and Queries* 65 (4): 524–30. <https://doi.org/10.1093/notesj/gjy151>.
- Mitrany, David. 1975. *The Functional Theory Of Politics*. New York: St. Martin's Press.
- Moshe, Koppel, Jonathan Schler, and Schlomo Argamon. 2008. "Computational Methods in Authorship Attribution." *Journal of the American Society for Information Science and*

- Technology*, Wiley Subscription Services VL (1): 9–26.
<https://doi.org/10.1002/asi.20961>.
- Mosteller, Frederick, and L. David Wallace. 1963. “Inference in Authorship Problem – a Comparative Study of Discrimination Methods Applied to Authorship of Disputed Federalist Papers.” *J.Am.Stat.Assoc*, no. 58: 275–309.
- . 1984. *Inference and Classical Inference: The Case of the Federalist Papers*. 2nd ed. New York: Springer-Verlag.
- Nance, John V. 2017. “Middleton and the King’s Speech in *All’s Well That Ends Well*.” In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 321–36. Oxford: Oxford University Press.
- NCSS, LLC. n.d. “K-Means Clustering.” Chapter 443, NCSS Statistical Software.
https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/K-Means_Clustering.pdf.
- Negnevitsky, Michael. 2011. *Artificial Intelligence, A Guide to Intelligent Systems*. Third. Harlow, England: Addison Wesley, Pearson.
- “Neuro-Adaptive Learning and ANFIS.” 2015. MATHWORKS.COM.
<https://uk.mathworks.com/help/fuzzy/neuro-adaptive-learning-and-anfis.html>.
- Omega, ed. 2018. “How Does a Fuzzy Logic Controller Work?” Omega Company.
<https://www.omega.co.uk/technical-learning/pid-fuzzy-logic-adaptive-control.html>.
- Osgood, C.E, G.J Suci, and P.H Tannebaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Parks, Joan. 1999. “History, Tragedy, and Truth in Christopher Marlowe’s *Edward II*.” *SEL Studies in English Literature 1500-1900* 39 (2): 275–90.
- Pascucci, Giuliano. 2017. “Using Compressibility as a Proxy for Shannon Entropy in the Analysis of Double Falsehood.” In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 407–16. Oxford: Oxford University Press.
- Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann.
- Peterson, D. Norman. 1969. “Textual Processing, a Computerised Aid for the Writer, Editor, and Reader.” *IEEE TRANSACTIONS ON ENGINEERING WRITING AND SPEECH*, VOL EWS-12 (1): 9–12.

- Pruitt, Anna. 2017. "Refining the LION Collocation Test: A Comparative Study of Authorship Test Results for *Titus Andronicus* Scene 6 (= 4.1)." In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 92-106. Oxford: Oxford University Press.
- Reyna, Valerie F. 2012. "A New Intuitionism: Meaning, Memory, and Development in Fuzzy-Trace Theory." *Judgm Decis Mak.* 7 (May): 332–359.
- Rolfe, William Ed. 1884. *The Two Noble Kinsmen, Written by the Memorable Worthies of Their Time, Mr. John Fletcher and Mr. William Shakespeare*. New York: Harper & Brothers.
- Rousseau, R., and Qiaoqiao Qiaoqiao. 1992. "Zipf's Data on the Frequency of Chinese Words Revisited." *Scientometrics* 24 (2): 201–20.
- Rowlett, Russ. 2001. "Shannon (Sh) & Bit." In *How Many? A Dictionary of Units of Measurement*. Chapel Hill: University of North Carolina at Chapel Hill.
- Satzinger, John, Robert Jackson, and Stephen Burd. 2015. *Systems Analysis and Design in a Changing World*. Seventh edition. Boston, Massachusetts: Cengage Learning.
- Saussure, Ferdinand de. 1995. *Cours de Linguistique Générale (Publié Par Charles Bally et Albert Sechehaye, Avec La Collaboration d'Albert Riedlinger; Éd. Critique Préparée Par Tullio De Mauro; Postface de Louis-Jean Calvet)*. Payot. Paris.
https://monoskop.org/images/f/f1/Saussure_Ferdinand_de_Cours_de_linguistique_generale_Edition_critique_1997.pdf.
- Schanzer, Ernst. 1963. *The Problem Plays of Shakespeare*. New York: Schocken.
- Segarra, Santiago, Mark Eisen, and Alejandro Ribeiro Ribeiro. 2014. "Authorship Attribution Using Function Words Adjacency Networks." *Department of Electrical and Systems Engineering, University of Pennsylvania*. <https://arxiv.org/abs/1406.0001> › cs.
- Severdia, Ron, ed. 2019. "PlayShakespeare.Com." The PlayShakespeare Team.
<https://www.playshakespeare.com/edward-iii/characters>.
- Shannon, Claude. 1948. *A Mathematical Theory of Communication*. Vol. 27. Bell System Technical Journal. 379-423, 632-656.
- Sichel, H.S. 1975. "On a Distribution Law for Word Frequencies." *Journal of the American Statistical Association* 70: 542–47.
- Siddhartha, Bhattacharyya, ed. 2015. *Handbook of Research on Advanced Hybrid Intelligent Techniques and Applications*. Hersey, Pennsylvania: IGI Global.

- Spalding, Williams. 1833. *A Letter on Shakespeare's Authorship of The Two Noble Kinsmen: A Drama Commonly Ascribed to John Fletcher*. 1. Edinburgh: A. Balfour and Co.
- Spedding, James. 1874. "The Shares of Shakspeare and Fletcher in *Henry VIII*." *Transactions of the New Shakspeare Society*, Reprint of "Who Wrote Shakspeare's Henry VIII", Gentleman's Magazine, Appendix: 50-57.
- Stein, Sterling, and Shlomo Argamon. 2006. "A Mathematical Explanation of Burrows' Delta." *Digital Humanities ALLC*.
- Steiner, Peter. 1984. *Russian Formalism, A Metapoetics*. London: Cornell University Press.
- Stoll, Robert R., and Herbert Enderton. 2018. "Set Theory." In *Encyclopaedia Britannica*, edited by Encyclopaedia Britannica. <https://www.britannica.com/science/set-theory>.
- Stone, P.G et al. 1966. *The General Enquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.
- Suen, Y. Ching. 1979. "N-Gram Statistics for Natural Language Understanding and Text Processing," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, PAMI-1 (2).
- Sugeno, M., and G.T Kang. 1988. "Structure Identification of Fuzzy Model." *Fuzzy Sets and Systems* 28 (1): 15–33.
- Sykes, H. Dugdale. 1923. *Sidelights on Elizabethan Drama: A Series of Studies Dealing with the Authorship of Sixteenth and Seventeenth Century Plays*. Ocford: Oxford University Press.
- Takagi, T., and M. Sugeno. 1985. "Fuzzy Identification of Systems and Its Applications to Modeling and Control." *IEEE Transactions on Systems, Man, and Cybernetics*, 15 (1): 116–32.
- Taylor, Gary. 1987. "The Canon and Chronology of Shakespeare's Plays." In *William Shakespeare: A Textual Companion*, edited by Stanley Wells, Gary Taylor, John Jowett, and William Montgomery, 69-144. Oxford: Clarendon Press.
- . 2017. "All's Well That Ends Well: Text, Date, and Adaptation." In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 322–36. Oxford: Oxford University Press.
- Taylor, Gary, and Doug Duhaime. 2017. "Who Wrote The Fly Scene?" In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 67-91. Oxford: Oxford University Press.

- Taylor, Gary, and Gabriel Egan, eds. 2017. *The New Oxford Authorship Companion*. Oxford, United Kingdom: Oxford University Press.
- Taylor, Gary, and Rory Loughnane. 2017. "The Canon and Chronology of Shakespeare's Works." In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 417–602. Oxford: Oxford University Press.
- Tweedie, F.J., S. Singh, and David I. Holmes. 1996. "Neural Network Applications in Stylometry: *The Federalist Papers*." *Computers and the Humanities* 30: 1–10.
- Vapnik, N.Vladimir. 2000. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
- VARD 2. n.d. Lancaster, U.K: Lancaster University. <http://ucrel.lancs.ac.uk/var2>.
- Vickers, Brian. 2002. *Shakespeare, Co-Author: A Historical Study of Five Collaborative Plays*. Oxford: Oxford University Press.
- . 2007. *Shakespeare, A Lover's Complaint, and John Davies of Hereford*. Cambridge: Cambridge University Press.
- Yule, G. Udny. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge: University Press.
- Zadeh, L. A. 1968. "Communication. Fuzzy Algorithms." *Information and Control* 12 (2): 94–102.
- Zadeh, L. A. 1965. "Fuzzy Sets." *INFORMATION AND CONTROL* 8 (338-353).
- Zipf, Kingsley George. 1932. *Selected Studies of the Principle of Relative Frequency in Language*. Massachusetts: Harvard University Press.
- . 1935. *The Psychobiology of Language*. Boston: Houghton-Mifflin.
- . 1949. *Human Behavior and the Principle of Least Effort*. Massachusetts: Addison-Wesley.