

A computer-assisted analysis of literary texts: a sample study

I Keshabyan-Ivanova*

Departamento de Filología Inglesa, Universidad de Murcia, Facultad de Economía y Empresa Campus de Espinardo, 30100 Murcia. España (Spain)

Received 1 March 2014, www.tsi.lv

Abstract

The overall aim pursued in this work is to demonstrate how quantitative data and a range of different corpus-based analytical techniques can be used in assessing an author's literary originality in relation to his texts' structures and meanings. With this in mind, the present study provides a sample of quantitative analysis of the two literary texts –Shakespeare's *Hamlet* [1685] and Sumarokov's *Gamlet* [1787]. Prior research has explored *Hamlet* and *Gamlet* in terms of historical, philosophical, language-based, etc. approaches that have existed to date. Taking into consideration the aforementioned visions of both plays, a special perspective on *Hamlet* and *Gamlet* is adopted herein. Given the importance assigned to computer-assisted analysis of literary texts, the current study is based on the idea that the texts under examination contain a certain number of particular characters that are distributed in a special way within and among the acts and intervene with a particular frequency specified by the authors. To achieve this aim, the texts are closely read and, then, computational and quantitative resources are applied. In general, the relevant findings unveil substantial structural deviations of the presence and interventions of all main characters, leading to noticeable diversions in the role and weight assigned by the authors to them per different acts inter-plays.

Keywords: Corpus-based, computational, quantitative, presence, interventions

1 Introduction

Corpus-based analytical techniques and specific ways in which corpus analysis has been applied to the study of literature have become more widespread over the recent decades. Interestingly, the works of such scholars as Stubbs [2005], Wynne [2006], Biber [2011], Johnson [2011], etc. demonstrate that it is increasingly becoming possible to test empirically claims about the language of literature, to search for and provide evidence from texts, to establish the norms of literary and non-literary style, and to have in-depth insights into the texts' structures and meanings.

The present research addresses a fundamental question concerning how quantitative data and a range of different computational and quantitative tools can be used in assessing an author's literary originality in relation to his texts' structures and meanings. It should not only benefit research on computer-assisted analysis of literary texts but also be of interest more generally to scholars of translation and comparative literature, leading to a more far-reaching understanding of many aspects of literature. At the same time, it should be noted that translation *stricto sensu* is not the main topic of the present investigation.

The overall aim pursued in this paper is to provide an example of quantitative analysis of the two literary texts – the Fourth Folio Edition of *The Tragedy of Hamlet Prince of Denmark* [1685] by Shakespeare and the English translated version of *Gamlet* [1787] by

Sumarokov, rendered into English by Richard Fortune in 1970.

For ease of reference, the abbreviations are used instead of the complete titles of the plays. For example, *Hamlet* or SH for Shakespeare's text, SG-R for Sumarokov's text, whereas *Gamlet* or SG for the English translation of the Russian text. However, one should keep in mind that in this research SG-R and SG are interchangeable, even though the general parameters of structural distinctions are explored between SH and SG and not between SH and SG-R. The selected texts are shown in Table 1.

TABLE 1 Texts selected for the analysis

Genre	Author	Title	Abbreviation
Drama	Shakespeare	<i>The Tragedy of Hamlet Prince of Denmark</i> (1685), the Fourth Folio Edition	SH
	Sumarokov	<i>Gamlet</i> (1787), in Russian (for reference) <i>Hamlet</i> (1970), in English	SG-R SG

The rationale behind my selection of Sumarokov's *Gamlet* for the current study is based on the fact that it was the first appearance of any of Shakespeare's plays in Russian culture, literature and theatre, although Shakespeare's name was nowhere mentioned [Lang 1948: 67] and the author himself denied any resemblance to Shakespeare's tragedy, 'apart from the monologue at the end of the third act and Claudius' falling down on his knees' [in Levitt 1994: 320]. The Fourth Folio Edition of *The Tragedy of Hamlet Prince of Denmark* [1685] was

*Corresponding author e-mail: kirina@um.es

selected for the comparison due to the fact that Sumarokov may have been acquainted with this edition before writing his *Gamlet* and, thus, may have been influenced by it [Levitt 1994: 322].

Prior research has analysed *Hamlet* and *Gamlet* in terms of historical, philosophical, language-based, etc. approaches that have existed to date. With respect to *Hamlet*, Bradley [1904], Wilson [1959], Johnson [1960], Eagleton [1986], Kermode [2000], to name but few, have provided studies by using previously mentioned perspectives. With regard to *Gamlet*, Tynianov [1929], Lang [1948], Billington [1970], Levitt [1994 and Gukovskii [2003], among others, have contributed relevant works based on the above-mentioned approaches.

Lately, original scholarly contributions within Shakespeare studies have appeared. Murphy [2007: 67] suggests that a popular interest in Shakespeare has been paralleled in the last decade by a rise of computer-assisted textual analyses of his plays, specifically in the field of literary stylistics, although met with resistance by some literary critics such as Fish [1996], Louw [1997], etc. From within the field of corpus linguistics, Stubbs [2005: 22] calls to combine the findings of corpus stylistics with close reading of texts.

Recent corpus-based studies on Shakespeare have examined distinct aspects of his plays such as characterisation in *Romeo and Juliet* [Culpeper 2002], the morpho-syntactic variability of the second person pronouns in the Shakespeare Corpus [Busse 2002], key semantic domains and metaphor in love tragedies and love comedies [Archer, Culpeper and Rayson 2009], the rhetoric of suicide in *Hamlet* [Anderson and Crossley 2011], etc.

Having taken the aforementioned visions of both plays into consideration, a special perspective on *Hamlet* and *Gamlet* has been adopted herein. Given the importance assigned to computer-assisted analysis of literary texts, the current study is based on the idea that the analysed texts contain a certain number of particular characters that are distributed in a special way within and among the acts and intervene with a particular frequency specified by the authors. Hence, to achieve this aim, the texts are closely read and, then, computational and quantitative resources are applied.

To identify the dimensions of structural divergences that are especially characteristic of *Hamlet* and *Gamlet*, this investigation focuses on those aspects of the plays that could be easily located, extracted, quantified and computerized –in other words, on the distribution patterns of the presence and interventions of all main characters per distinct acts inter-plays. In so doing, the researcher seeks to reveal probable commonalities and/or deviations in the weight the authors put on all main characters within the plays that have led Sumarokov to introduce substantial changes into the structure of his play compared to Shakespeare's original play *Hamlet*.

The two phases of the present research include the analysis and discussion of the data associated with the distribution patterns of the presence and total interventions of all main characters, namely Hamlet, Claudius, Polonius, Gertrude and Ophelia, per different acts inter-plays.

2 Methodology

Two variables (including presence and intervention categories) are chosen in the current work under the criteria that they are quantitative and require certain computational tools. Each text is analysed with respect to the occurrences of these features that are quantified. The quantification of presence and intervention variables is carried out manually by exploring the two text files. After, the extracted data is computerised, tabulated (intra-play), cross-tabulated (inter-plays) and presented in tables and figures. The tools used for the computational quantification and presentation of the data in tables and figures are SPSS V.15 and Excel (Office 2007).

For the purposes of analysis, the data shown as a percentage throughout various phases of the present investigation are considered more valid than the data given in figures.

However, the aim is to normalise the data quantitatively in order to provide more precise identification and comparison of the general trends employed by the two playwrights in relation to the distribution patterns of the presence and interventions of all main characters per distinct acts inter-plays.

The instrument used for the standardisation of the data and their presentation in figures is Pearson's Correlation Test [Altman 1991: 285-288]. Correlation is a kind of technique that summarises the strength of the connection between two variables. For example, in this study, the presence and intervention variables in one text (SG) are compared with the presence and intervention variables in another text (SH) –in other words, there is one variable in each text separately that is compared between two texts.

The general requirement for Pearson's correlation coefficient is the observation of the two variables, which are measured on an interval or ratio scale and the calculation is based on the actual values [Altman 1991: 285-288]. In the present research, Pearson's Correlation Test computes the quantitative correlation between the presence and intervention variables per act inter-plays.

3 Findings and discussion

3.1 PRESENCE VARIABLES

Table 2 and Figure 1 incorporate the data linked to the distribution patterns of the presence of all main characters per act intra-play and inter-plays. However, as has already been commented in section 2, only the data

presented as a percentage inter-plays are analysed and discussed.

TABLE 2 Distribution Patterns of the Presence of All Main Characters per Act

Act	SH	SH %	SG	SG %
	Presence	Presence	Presence	Presence
I	5	22.73	2	13.33
II	5	22.73	3	20.00
III	5	22.73	3	20.00
IV	4	18.18	3	20.00
V	3	13.64	4	26.67

Comparing these data, the following diversions can be observed: the distribution pattern varies a little in acts II, III and IV inter-plays, whereas, in acts I and V, the difference is the greatest as in act I it equals 22.73 % in SH against 13.33 % in SG and, in act V, 13.64 % in SH against 26.67 % in SG.

As a result, the difference in the distribution is not statistically significant inter-plays ($\chi^2 = 1.293$; $df = 4$; $p = 0.862$), which is a clear sign of divergence.

The data in Figure 1 display a probable quantitative correlation between the distribution patterns of the presence of all main characters per act inter-plays. These data show that the patterns of the presence of all main characters are negatively correlated per act inter-plays. The statistically non-significant correlation ($\rho = -0.790$; $df (8)$; $p = 0.111$) might indicate distinct distribution patterns.

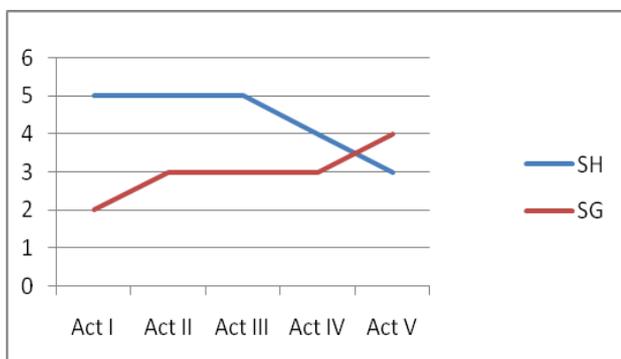


FIGURE 1 Quantitative Correlation between the Distribution Patterns of the Presence of All Main Characters per Act

It should be observed that Figure 1 clearly illustrates that the line, which corresponds to SH, remains in the same position in acts I-III as Shakespeare distributes all main characters equally in these acts whilst in acts IV and V it drops dramatically due to the reduction of the number of all main characters. Sumarokov behaves differently as the line goes up from act I to act II, remains in the same position in acts II-IV and again goes up in act

V. Even though the lines cross in act V inter-plays, the movement is downward in SH as opposed to the upward movement in SG.

As a result, the data in Table 2 and Figure 1 possibly demonstrate that both playwrights follow diverse distribution patterns of all main characters per act inter-plays, specifically in acts I and V.

3.2 INTERVENTION VARIABLES

Table 3 and Figures 2 and 3 comprise the data linked to the distribution patterns of the total interventions of all main characters per act and per full text intra-play and inter-plays.

TABLE 3 Distribution Patterns of the Total Interventions of All Main Characters per Act and per Full Text

Act	Frequency of Interventions	Percentage of Interventions	Frequency of Interventions	Percentage of Interventions
	SH	SH %	SG	SG %
I	100	9.17	32	16.67
II	145	13.30	22	11.46
III	197	18.07	47	24.48
IV	98	8.99	22	11.46
V	111	10.18	37	19.27
I-V	651	59.72	160	83.33

Comparing the data expressed as a percentage inter-plays, it can be seen that the frequency of total interventions of all main characters per act is higher in SH than in SG. As can be appreciated in Table 3 and Figure 2, the frequency of occurrence of total interventions is significantly higher in SH, although the percentage of total interventions is greater in SG, especially in acts I, III and V. For example, the difference is the highest in acts I and V as it equals 9.17 % against 16.67 % and 10.18 % against 19.27 % in SH versus SG, respectively. However, the distribution pattern is more or less alike in acts II and IV inter-plays.

Although the divergence in the distribution is statistically significant inter-plays ($\chi^2 = 9.407$; $df = 4$; $p = 0.049$), which might indicate similar distribution patterns, there appears to be a partial dissimilarity, particularly in acts I and V.

To identify a possible quantitative correlation between the distribution patterns of the total interventions of all main characters per act inter-plays, Figure 3 is generated. According to the data shown in this figure, the quantitative correlation between the distribution patterns of the total interventions of all main characters is statistically not significant per act inter-plays ($p = 0.608$; $df (8)$; $p = 0.276$); a clear sign of a distinct inter-play behaviour.

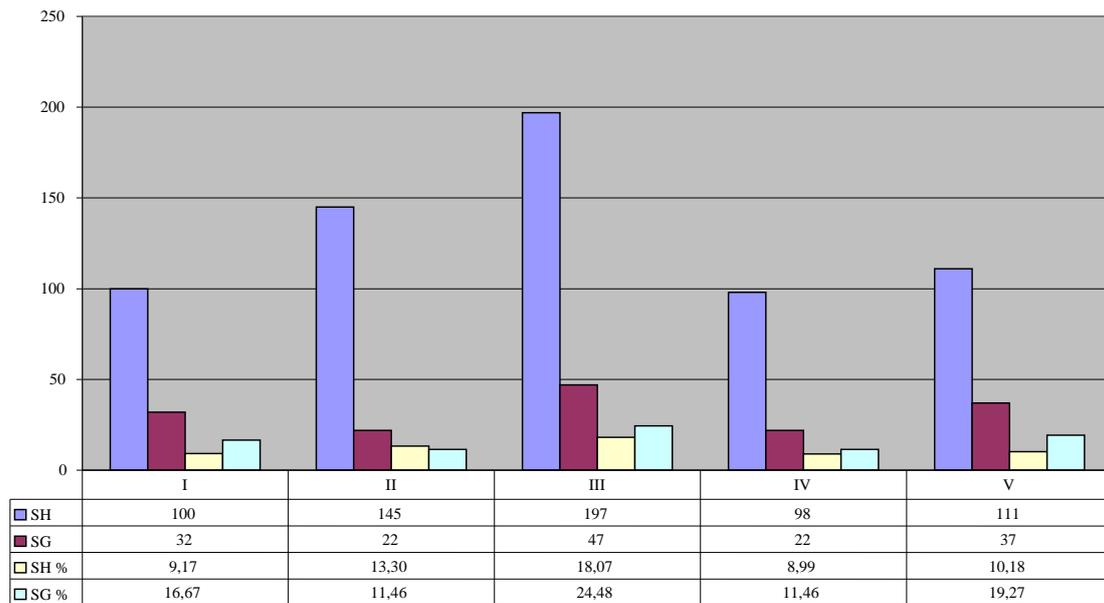


FIGURE 2 Distribution Patterns of the Total Interventions of All Main Characters per Act

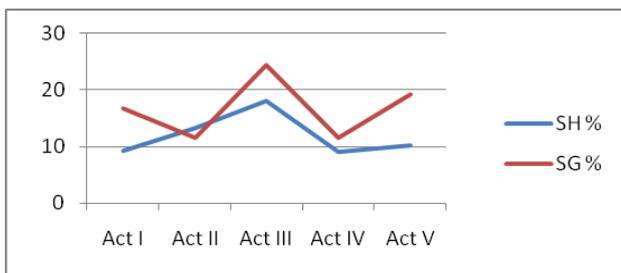


FIGURE 3 Quantitative Correlation between the Distribution Patterns of the Total Interventions of All Main Characters per Act

In fact, the line that displays the total interventions of all main characters in SH rises from act I to act II, goes up considerably in act III, falls in act IV and rises a little in act V. In SG, the line falls and rises more strikingly. Moreover, in act III, the upward movement is more significant in SG than in SH.

As a result, Shakespeare and Sumarokov apparently follow dissimilar distribution patterns of the total interventions of all main characters in all acts, especially in acts I and V.

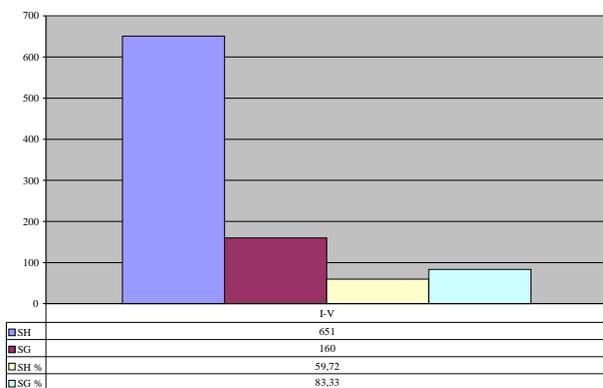


FIGURE 4 Distribution Patterns of the Total Interventions of All Main Characters per Full Text

To better illustrate the distribution patterns of the total interventions of all main characters per full text, Figure 4 is designed. The data in this figure show that the frequency of occurrence of total interventions of all main characters per full text equals 651 in SH in contrast to 160 in SG. However, the total percentage of interventions of all main characters is substantially higher in SG than in SH, that is, 83.33 % against 59.72 %, respectively.

Indeed, the data in Table 3 and Figures 2-4 seemingly point towards the fact that all main characters are assigned a more notable role in SG as opposed to SH.

3.3 SUMMARY OF THE MAIN FINDINGS

With regard to the distribution patterns of the presence, I have observed that Shakespeare keeps more or less the same number of all main characters from act I to act III, decreasing their presence dramatically in acts IV and V. By contrast, Sumarokov gradually increases their number, reaching its peak in act V.

With respect to the distribution patterns of the total interventions of all main characters, I have found out that all main characters intervene more frequently and, therefore, carry more weight, specifically in acts I and V in SG. In act III, Sumarokov, as opposed to Shakespeare, distributes them in a more striking way, which proves that there exists a considerable partial divergence in this act inter-plays.

The aforementioned findings fit with the results in the previous research linked to the analysis of the presence, interventions and interactions of the main and secondary characters as well as of the topics dealt with in SH and SG [Keshabyan-Ivanova 2011].

To give an example of clear correlations between the results of the quantitative approach and Sumarokov’s literary originality in relation to his text’s structure and

meaning, I shall compare the findings from this work as well as from the above-mentioned investigation linked to act I in SH and SG.

As a result of these studies, in act I Shakespeare's Hamlet intervenes much more than the other main characters, namely Claudius, Polonius, Gertrude and Ophelia; however, not more than the secondary characters [Keshabyan-Ivanova 2011: 92-95, 128, 129, 137 and 138]. Moreover, he interacts only two times with Claudius and three times with Gertrude [Keshabyan-Ivanova 2011: 137].

In comparison to Shakespeare, only two main characters –Hamlet and Gertrude appear and socialise with each other in act I in SG. At the same time, Hamlet interacts, although with a lower frequency, with his confidant Armands –a secondary character that is present only in SG [Keshabyan-Ivanova 2011: 157].

Consequently, the interrelation between Hamlet and secondary characters is more prominent in act I in SH than in SG where the linkage between two main characters –Hamlet and Gertrude, who represent all main characters in this act, is of major interest.

Furthermore, the results, obtained through the quantitative and qualitative comparison of the distribution patterns of the most frequent content words in act I, provide evidence to the fact that the topic of religion, with its traditional moral values, represents a greater appeal for Sumarokov as opposed to Shakespeare who is drawn to it to only some extent [Keshabyan-Ivanova 2011: 290-291].

Indeed, Sumarokov, through his main character Hamlet, tries to improve the morality of another main character Gertrude –in other words, in line with Sumarokov, to save the sinner's soul. Shakespeare also links the human soul to God's instructions, although he is less straightforward and explicit than Sumarokov is in his moral plan [Keshabyan-Ivanova 2011: 270].

References

- [1] Stubbs M 2005 *Language and Literature* 14(1) 5–24
- [2] Wynne M 2006 *Encyclopedia of Language & Linguistics* 12 ed Keith Brown Oxford: Elsevier 223-6
- [3] Biber D 2011 *Scientific Study of Literature* 1(1) 15-23
- [4] Johnson J 2011 *Target* 23(1) 62-76
- [5] Shakespeare W 1685 *The Tragedy of Hamlet Prince of Denmark* Mr William Shakespeare's Comedies, Histories, and Tragedies London: H Herringman, E Brewster, R Chiswell and R Bentley 59-86
- [6] Sumarokov A 1787 *Gamlet Tragediya Polnoe sobranie vseh sochineny v stikhakh i proze v 10-ti tomakh Aleksandra Petrovicha Sumarokova Vol 3* ed N Novikov 10 vols V Moskva: V Universitetskaya Tip 61-134 (in Russian)
- [7] Sumarokov 1970 *Hamlet A Tragedy Selected Tragedies of A P Sumarokov* ed Henry M Nebel Jr transl R Fortune Evanston: Northwestern UP 87-134
- [8] Lang D 1948 *Modern Language Review* 43(1) 67-72
- [9] Levitt M 1994 *Slavic and East European Journal* 38(2) 319-41
- [10] Bradley A C 1904 *Shakespearean Tragedy* London: Macmillan
- [11] Wilson J D 1959 *What Happens in Hamlet* Cambridge: Cambridge UP

4 Conclusions

The current study has explored the frequency of the distribution patterns of the presence and total interventions of all main characters in *Hamlet* and *Gamlet*. Obviously, the analysis indicates that Shakespeare and Sumarokov set rather distinct aims associated with all main characters.

Compared to Shakespeare, Sumarokov pays greater attention to the main characters, that is, people of a high social rank. At the same time, these aims had a great impact on the structure of the plays, leading to noticeable diversions in the role and weight assigned by the authors to them per different acts inter-plays.

In this respect, it should be noted that to produce a critical analysis of the structure or of the subject of study, that is, to answer the question what quantitative differences mean with regard to literary analysis, more variables should be incorporated. For example, the method could measure presence alone or along other characters, at which points within acts the characters occur or intervene, the relevance of the intervention and/or their speech/actions in the development of each act, the different importance of each character for a distinct purpose, etc.

Finally, it should be highlighted that the present research represents only a sample of quantitative analysis of literary texts as, with this paper, I have aimed to suggest that this kind of empirical work is needed to underpin qualitative literary analysis. However, limitations of computer-assisted textual analysis should always be taken into consideration.

Acknowledgments

I would like to thank Dr Pascual Cantos-Gómez who was the source of the invaluable knowledge needed to carry out this study.

- [12] Johnson S 1960 *Samuel Johnson on Shakespeare* ed W K Wimsatt Jr New York: Hill
- [13] Eagleton T 1986 *William Shakespeare* Re-reading Literature Series Oxford: Oxford UP
- [14] Kermode F 2000 *Shakespeare's Language* Harmondsworth: Penguin
- [15] Tynianov Yu 1929 *Oda kak oratorsky zhanr* Arkhaisty i novatory Moskva Leningrad: Priboy (in Russian)
- [16] Billington J 1970 *The Icon and the Axe: An Interpretative History of Russian Culture* New York: Vintage
- [17] Gukovskii G 2003 *Russkaya literatura XVIII veka* Moskva: Aspekt P (in Russian)
- [18] Murphy S 2006 *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching* 1 66-85
- [19] Fish S 1996 1973 *The Stylistics Reader. From Roman Jakobson to the Present* ed J J Weber London: Arnold 94-116
- [20] Louw B 1997 *Teaching and Language Corpora* ed A Wichman, S Fligelstone, T McEnery and G Knowels London: Longman 240-51
- [21] Culpeper J 2002 *Conversation in Life and in Literature: Papers from the ASLA Symposium 15* ed U Melander-Marttala, C Ostman and M Kyto Uppsala: Association Suedoise de Linguistique Appliquee 11-30

- [22] Busse U 2002 *Linguistic Variation in the Shakespeare Corpus: morpho-syntactic variability of second person pronouns*. Amsterdam: John Benjamins
- [23] Archer D, Culpeper J, Rayson P 2009 *What's in a Word-list? Investigating word frequency and keyword extraction* ed A Dawn Farnham: Ashgate 137-158
- [24] Anderson Th, Crossley S 2011 *Stylistics and Shakespeare: Transdisciplinary Approaches* ed M Ravassat and J Culpeper London: Continuum 192-214
- [25] Altman D 1991 *Practical Statistics for Medical Research*. London: Chapman & Hall 285-8
- [26] Keshabyan-Ivanova I 2011 *Shakespeare's Hamlet versus Sumarokov's Gamlet: A Corpus-Based Perspective* Saarbrücken: LAP LAMBERT

Authors

I Keshabyan-Ivanova, born on November 25, 1952, Sochi (Russia)



Current position, grades: Lecturer in the English Department at the University of Murcia (Spain)

University studies: She studied English and German at the Pyatigorsk State University of Foreign Languages (Russia) and at the Yerevan Linguistics University (BA Hons). She holds a PhD in English/Corpus and Applied Linguistics from the University of Murcia (Spain).

Scientific interest: Her research interests focus on the application of computer-aided corpus-based methodologies to literary texts – specifically to structural and lexical analyses. Recently, she has also developed an interest in frame semantics. Since 2009 she has been a member of the LACELL Research Group at the University of Murcia.

Publications: Her publications include a book, entitled *Shakespeare's Hamlet versus Sumarokov's Gamlet: A Corpus-Based Perspective*, published by LAP LAMBERT Academic Publishing; articles in journals such as *IJES*, *RCEI*, etc. In collaboration with Dr Angela Almela, she has co-edited a special-themed issue on *A New Approach to Literature: Corpus Linguistics* in the *International Journal of English Studies* (2012).

Experience: She has been a Lecturer of English at the University of Murcia since 2001. She teaches in different areas of English Philology such as Philology and Culture of English-speaking Countries, Business English, Legal English, etc. She was previously a Lecturer of English in the Department of Intensive Methods of Teaching Foreign Languages at the Yerevan Pedagogical University and a Lecturer in Phonetics and Oral Practice at the Yerevan Linguistics University.