_____

Research Methods for Attributing Authorship in English Studies
_____

Amit Kumar Rath, Research Scholar, PhD English Literature, Department of English Literature, The English and Foreign Languages University, Lucknow Campus, Luckow, Uttar Pradesh, India

_____

Abstract

The past decade has witnessed a metamorphic rise in the number of courses, papers, and workshops being organized, written, and sponsored in the various departments of English and literary studies, but this was not the case in the decade prior and this need for the formulation and introduction of standardized research methods for the conduction of research in this field has been long felt but only recently acknowledged. Other sister subjects in the field of Humanities and Social Sciences have had written and theorized methods of research for several years now. The attitude towards research methods in English Studies has been of "learning via doing," one of traditional apprenticeship, but this is problematic in the conduction of fair research in any capacity. There are several problems with this way of doing research but the most important one is of conformational bias and fixed agenda. This problem is compounded when undertaking authorship attribution research in English Studies. The researcher is prone to bending textual evidence of scholarship that might not be completely apparent in favour of one of the several collaborators of a literary text. This creates the imperative problem of spurious attributions and can wreak chaotic havoc on our bibliographical understanding of the authorship of treasured texts. This paper aims to pose a study of available research methods for authorship attribution in English Studies that can eliminate this problem and lay the foundation for objective and ethical research in this area.

Keywords: "Authorship Attribution," "Research Methods," "ICT," "Discourse Analysis," and "(Auto)biography"

---

1. Research Methods for English Studies

In the area of research in English Studies, currently, there are several research methods that have been formulated for literary studies and are available at the disposal of the student of literature. Some of them deal with traditional textual analysis including methods such as close reading, (auto)biographical reading, studying oral origins, reading the visual as textual, and interviewing, etc., while others deal with newer research methods that have been either inspired and borrowed from allied subjects, have evolved from generic methods or have been newly devised for the sole purpose of assisting literary and linguistic research, some of them are discourse analysis, ethnographic methods, qualitative and quantitative methods, creative writing, and ICT as research methods etc. It is important for a scholar to have the functional knowledge of all these several methods of doing research in this field so that s/he may take into employment any one or several of these methods while conducting his or her research.

The traditional textual analysis methods primarily focus on the primary text in its purest form and employ literary theories to extract information from the source that may be latent and not apparent to the unkeen eye. Structuralist and deconstructionist methods use close reading to assess the weight of a particular word in context and the force it exerts on the text and they replace it with other feasible words to understand the difference in the imparted meaning and therefore its true meaning and manner of asserting and attributing a particular understanding. (Auto)biographical research method allows the scholar an insight into the possible translation of experiences of the author as expressed in their (auto)biography into their creation by un/conscious dexterity. Oral history as a research method is used in the analysis of the texts that primarily originated and were propagated as songs and folk tales. It studies the history of their oral dissemination in contrast to their effect as a written text. Paintings, dance forms, and sculptures can together be called visual art forms and they are also inspired by the same movements and schools of thought that shape literature, therefore, they are sometimes analyzed in a manner similar to textual analysis. Interviewing has originated as a research method in fields that deal with raw empirical data and have socio-anthropological significance. Subjects, as well as authors, are interviewed to establish graduation between reality and textuality.

The newer methods of researching originated as a supplement for achieving what traditional methods couldn't and filling the gaps left in traditional research. Discourse analysis concerns itself with "language in use" and analyses non-fictional linguistic texts to make apparent the sites and method of hegemonical hierarchy to attempt and bring a socio-political change. Ethnographic methods are an anthropological approach to texts and their sociological determinism. They study the

influence of the culture in which the text is written. Qualitative and quantitative methods use computational help to mark and retrieve excerpts from the text of where the element is of peculiar nature typical to the author or genre. Creative writing explores the process of creation of a text and the elements that influence it by mimicking a similar process of writing thus helping in the understanding of several determinants of creativity that may be unrecognized. Lastly, the advent of the space age and the increasing use and availability of computers in our lives have enabled their use in research methods in English Studies thus making research easier and more accurate. Information and Communication Technology (ICT) as a research method is the latest addition to this list of assets in literary research.

While dealing with authorship attribution in English literary studies three specific methods seem to be the most helpful. The three main research methods that would be discussed here are Discourse Analysis, (Auto)biography, and ICT as research methods.

2. Discourse Analytical Research Method
2.1. Objective of Discourse Analysis

The main objective of discourse analysis is to investigate the structure of the oral and written language in use and not a system or structure of language that is ideal or utopian. It varies from textual analysis in trying to analyze the way language is understood by the reader and presented by the writer. Van Dijk rightly calls discourse analysis the study of the invested language. It can be divided into two subparts. The first is the investigation into the patterns employed by the author in the creation of the text i.e. whether or not there is the presence of a particular pattern, for example in the use of pronouns for the denotation of self. The second deals with the use of this pattern in the language, for example, how much verbal space is given to the narrator in a narration in comparison to the protagonist or other characters. On the whole, this method of analysis aims at listing all the textual elements that have been used to create a stance of power and to find if it is possible to rectify the imbalance of power wherever possible through the use of alternative textual patterns.

Discourse has multiple meanings in its practice as well as in the dictionary. Discourse can be seen as single, separate words pronounced during the duration of a conversation, but it may also refer, at the same time, to the string of utterances made and the idea insinuated. It also may refer to a larger, communal ideology whether or not being realized by the utterer. The meaning of discourse is highly contextual and often accompanied by immediate effects. Such as if we consider the incantatory beginning of fairy tales beginning with "Once upon in time…" (Griffin 92) we would immediately be led to a set of other "facts" in mind such as the beginning of a mythical tale that is bound to be didactic in nature but also primarily meant for young

children as audience and not adults. But at the same time, Children's fiction is greatly consumed by adults as well with the imparted morale having a similar effect albeit on a different note and scale. Discourses have frameworks that are normative to their operation and predefine the effect of a particular discourse to some extent. This can be seen as one of the agendas of discourse, to normalize and naturalize a particular thought that might not otherwise be accepted so easily and readily. This normalization is achieved by iterative performance and this also serves as the distinction between one discourse and another. These discourses co-exist and often struggle with each other for the share of the same audience and "only a limited number of such discourses can coexist on the same terrain" (Griffin 93). These discourses can also, at times, become antithetical to each other's existence. The case of traditional and alternative medicine is often provided as an ideal example. There is always a constant struggle amongst these coexisting discourses and the monopoly of the thetical discourse is always challenged. One example of thetical discourse being challenged would be the feminists' claim of gendered characters being given treatments in fairy tales and this questions the imbalance of power created in fairy tales through their discourse. Such challenges can also when administered properly and rigorously, lead to changes in society.

2.2. Types of Discourse Analytical Research Methods

Griffin states that there are, mainly, six discourse analysis research traditions and they can be divided into conversation analysis and ethnomethodology, interactional sociolinguistics and the ethnography of communication, discursive psychology, critical discourse analysis and critical linguists, Bakhtinian research, and Foucauldian research. There are certain methods of operation that are shared by these research methods and they are namely language in use is the primary research object, the idea that language is not innocent and doesn't only function as a means of conveying a message but also creates the reality of the world we live in, the creation of this reality is under the structure of a hegemony that is created and maintained by language, these hegemonical structures are often biased and can be changed by changing the discourse and such discursive change can ultimately lead to change in the society. These are the similar assumptions that these research methods operate with but at the same time, there are several points of contention between them. They mainly differ in the degree to which they think discourse constructs our reality, the source and type of discourse, the degree to which subjects have or do not have the ability to control the use of their own language, the effects of discursive practices, context and the knowledge that comes out of such analyses. This elucidates the fact that discourse analysis is not a unified method of researching but has several sub-methods via which the research can be undertaken.

_____

Norman Fairclough is a universally accepted authority in the field of critical discourse analysis and his vision for it separates communication of everyday life into discursive and non-discursive. He pre-dominantly engages with the text created by institutional bodies to identify how these discourses might be creating hegemonies of unbalanced power. Griffin gives the example of Jane Austen's *Pride and Prejudice* to illustrate this fact

> When Jane Austen in the opening sentence of *Pride and Prejudice* claims that 'It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife', she is not so much proclaiming a neutral and objective fact as mustering all her rhetorical force to offer a particular construction of a specific kind of single man'. That rhetorical force sets out to both avow a certain view and simultaneously and implicitly to disavow other ways of seeing. (96)

An analysis of similar scope may only aim at a certain chosen section of a text that the analyst finds fertile but this can be applied to linguistic texts of larger scope as well. For such an analysis to be made the analyst must be aware of the implicit and unspoken rules at play in the text and the context that provides the statement with their truer meaning and thus making apparent the possibility of an alternate reading and outlook. Robert Scholes suggests that a literary text needs to be viewed as "the product of a person or persons, at a given point in human history, in a given form of discourse, taking its meanings from the interpretive gestures of individual readers using the grammatical, semantic, and cultural codes available to them" (qtd. in Griffin 16).

A major effect of such an analysis is that the linguistic text chosen is meant to be fertile in ideological contentions. But other texts can also be employed to be analyzed in research in English literary studies. The research tools should be selected beforehand while doing discourse analytical research. Different linguistic texts require different treatments. For example, dramatic texts are best analyzed with conversation analysis tools to understand "turn-taking" in conversation in the drama or the novel. The purpose of such an exercise is not limited to check whether or not a character is speaking out of turn but also to see if there is any dissent in the viewpoints expressed by the characters and what is the degree of such dissent. There are many issues that can be taken into consideration by the researcher while analyzing the text. These may be gender issues, race issues, class issues, caste issues, etc. A semantic classification of a people also needs to be addressed as this is primarily done to assign a place in the hegemony construct. Although, discourse analysts might vary in their opinion of hegemony is created by words but they all agree on their inter-relation. Discourses have power in that particular structure but may be relatively powerless in

others. Such analysis is done by identifying patterns in that particular structure. Patterns can be differentiated on the basis of the lexicon, grammar, and semantics. Such peculiar patterns can be marked and then compared with other literature of the same author, genre, theme, or time. Word repetitions need to be noted but the use of synonyms of such words needs to be marked as well and this can become tiresome when dealing with the large canon of work. This work is made easier by the employment of computer software.

2.3. Computer Assisted Qualitative Data Analysis

In the past two decades, computers have heavily been employed in the field of qualitative data analysis. The main caution that needs to be taken by the researcher while employing computer-aided data analysis methods is that utmost care needs to be taken while inputting the code for data retrieval from a linguistic text or canon. If the coding is biased or leaning towards a particular agenda then the output is only going to authenticate this conformational bias. Discourse analysis can be divided into qualitative and quantitative methods of analysis. The former deals with the text as a non-uniform text and the latter deals with the frequency of the usage of certain words, nouns, pronouns, grammatical structures, etc. to make apparent the motifs and themes of the text and to discuss the words more likely to be used by one author than the other.

The first step in computer-assisted discourse analysis research should be the coding of the material that is to be checked, summarised, and retrieved from the text in question. This process is critical in such research as these codes define the agenda with which the research material is approached as well as separate the material on which the research is based. There are two ways of coding, one is to apply all the general codes that are deemed relevant, for example, "marriage," "husband," and "wife" in Jane Austen's *Pride and Prejudice* will be relevant and return several results and then these can later be refined with recoding. Another method would be, to begin with, the perspective that is to be undertaken in the research. Socialist perspectives can begin by coding the word "class," for example. It is also helpful to create a codebook in which all the codes along with their definition and reasons for interest are listed. Some of the popular software packages used for this purpose are ATLAS. ti, Kwalitan, and WinMAX.

3. (Auto)Biographical Research Method

3.1. Objective of the (Auto)Biographical Research Method

The Western system of education has made academia research-centric i.e. teachers and professors need to be researchers as well in order to be able to survive in modern academia. Researchers who research into the lives of others via the use of (auto)biographies also lead lives of similar fluidity to their subjects, the veil between

their personal and professional lives is equally thin. This has become a discourse wherein people who are not open about their personal lives in public are thought to be harboring a vice as openness about personal life has come to be seen as a virtue of civility. This openness is always controlled and can be manipulated by, if intended, the subject of the (auto)biography and this raises the issue of the reliability of (auto)biographies as a source of information and knowledge. The self-revelation is itself subjective and completely in control of the subject. This is very similar to the increasing demand for the researcher to make apparent their standing in life in terms of their sexuality, politics, and ideology.

While undertaking (auto)biography as a research method it is important to take into consideration the fact that the research should just not be archived information about the subject's life. The focus is no more on the information about the subject's life because of that information already available rather the search is for the demarcation and creation of an ideological standpoint of the subject of the (auto)biography that indicates the belief system of the society or period to which this person belongs.

> In this change, we can perhaps see the search in (auto)biography transformed from a search for the person to a search for a convincing reading. Where once the search was for greater information about the person, now it has shifted to a search for an interpretation of how the individual could be located within a particular *zeitgeist*. The paradox of contemporary (auto)biography and research about it is that in large part the individual life becomes almost less important. (qtd. in Griffin 32)

The newer and older methods of (auto)biography research share some of their strengths as well as their weaknesses. The shared weaknesses are, being partial in the inclusion or exclusion of an account of the life of the subject, the difference in the thought process behind the writing of the (auto)biography and that of the interpretation, and lastly the definitive focus on one individual as a source of knowledge of their life as well as that of the period they lived in.

3.2. The Requirement of Caution

The definition of what is called a "fact" in an (auto)biography needs to be re-engaged. The authenticity of all research and their validity is challenged when additions are made to the (auto)biography of a person. It is indeed very difficult to demarcate the "truth" of one's life. Therefore, (auto)biography as a research method must be undertaken with the formal acceptance of the fact that there are certain aspects of the subject's life that the public hence, by extension, the researcher will never know. Another thing that needs to be kept in mind that bias in research can be removed by taking care to keep fantasies about the subject or their life at bay from

the study. Self-justification, at no cost, should be the driving force behind the conduction of (auto)biographical research. (Auto)biographies concerning Western individuals and their societies have the common denominator of individualism. Individualism is the basis on which the capitalist Western society is built but at the same time, it is a major factor in its undermining. The capitalist society stresses the achievement as well as the responsibilities of the individual and in the process leaves social ties, achievements and responsibilities out of the question. Charles Taylor writes "What is more, the primacy of self-fulfilment reproduces and reinforces some of the same negative consequences as instrumentalism. Community filiations, the solidarities of birth, of marriage, of the family, of the polis, all take second place" (qtd. in Griffin 72). Here the main concern is the impact of individualism on society. This same concern extends to (auto)biography and the question about the extent of social reflection in the (auto)biography arises.

3.3. Individualism, Society, and (Auto)biography

The most apparent way of individualism affecting (auto)biography, as per Evans, is in the psychological restlessness in discovering one's self whether that be of own or other.

> It is in this sense that contemporary (auto)biography becomes something akin to a detective story. We want to find out why we and others act because through this knowledge we might be better achievers in the remorseless achievement culture of the Western twenty-first century. (qtd. in Griffin 36)

The measurement of the individual's achievements is that scale for the authenticity and goodness of their person in today's world. And this has become the normative order of the day. The society can be studied within the actions of the individual and in the rewards or punishments doled for it. (Auto)biography is similar to detective fiction in the manner that in the course of the narration several layers of a person's emotions are revealed and are apparent by their introspection or actions. (Auto)biography makes it possible to understand the way the individual is judged by the members of the culture and by those outsiders to it. Different accounts of the same life somewhere reveal the failure of the culture in judging the individual. But the writing of these (auto)biographies also fails in encompassing the then contemporary collective force of the culture in which life was lived. These writings help us in understanding the person, the culture, and the person in the culture.

4. Information and Communication Technology (ICT) Research Method
4.1. Objective of ICT

With the advent of computers, the manner of conducting research in the field of humanities has changed greatly. The speed and accuracy of computers have greatly helped us in saving time and making our research more accurate; details that could

be missed by an individual researcher receive attention with the thoroughness of technology. All this has happened in the past couple of decades. The use of these new resources has enabled researchers to include several new steps and methods in their research such as better analysis and interrogation of primary manuscripts, access to multiple online databases, archives, dictionaries, thesauruses and libraries. Often ICT functions as an aid and catalyst in research but sometimes it functions as the tool for the creation of transformative discovery and research and in such cases, it is called "digital scholarship."

4.2. Scope of ICT

Computer-aided research methods have emerged at an opportune moment in time for English Studies as in the past few decades English Studies have transferred their focus from traditional subjects to several other sister fields and everyday media including magazines, newspapers, political statements, and advertisements etc. The access to primary source materials has also increased in recent years, benefitting English scholars greatly as earlier manuscripts which were rare were extremely inaccessible for students and other faculties but now their digital copies ensure that they would be available for unlimited viewing globally without damaging the original copy. They also allow for the possibility of tertiary digital reconstruction of lost material in original manuscripts damaged over time and by faulty repairs. Libraries, universities, and other educational institutions have increasingly participated in digitization programs to help in the preservation of these manuscripts. Libraries are also working together, worldwide, to create digital libraries that are interoperable and function of similar standards that are operable with common tools to ensure maximum reach and accessibility.

4.3. Marking Data

Now there are varied needs in research that make data mark-up necessary and they are not just limited to data exchange. There are three ways of marking up data: descriptive, structural, and content. The first type deals with defining the basic scholarship of the work. It mentions details that are expected to be found on the cover page of any work today such as the name of the author, text, publisher, and year of publication etc. The second deals with marking of the units of the work, for example, line in a verse, scene of an act and chapter of a novel etc. The last marking helps us in separating a particular pattern of text from the rest by marking it up for systematic retrieval later. This can be useful while creating an index of any sort. Encoding is usually further divided into shallow and deep encoding. While the former deals with superficial encoding for instance in the case of applying publication guidelines to a text the latter deals with word patterns and units in depth to allow systematic analysis for instance in the case of a stylistic study.

_____

The text forms the base of the research and the markup provides us with a manner to embed a greater amount of identified and sorted detail into the text as an overlay. These markups can then be retrieved or ignored according to the specific needs of the research. It can be helpful to list all the marked-up elements in a spreadsheet or database management software in order to enable statistical analysis of the elements of interest in the particular text. To make the data retrieval process easy and accurate the markers in the text should be made explicit in a literary text where there is much possibility of implicitness as unlike us computers do not have their own intelligence to deal with implicitness and can alter the result to a great extent. Therefore, terms like "husband," "wife," and "spouse" should explicitly be linked to terms like "marriage" and shouldn't be left at the discretion of the software program.

Much research work in field of humanities concerns itself with scholarship, authorship attribution, and bibliography. Originally computers were intended to be used in the helping of the bibliographer, during their initial introduction to humanities. Now the pervasiveness and use of computational technologies in humanities are so expansive that most critical works are published on the World Wide Web in journals that operate primarily online. It is virtually impossible now to find critical work that was achieved without the help of ICT methods from its inception to production. The marking up of relevant elements adds meaning to the text that, perhaps, might have been unseen earlier. They help in the situation of these works in a relevant political, social, and historical context. For example, in English Studies, one can consider the Canterbury Tales Project in which not only the entire extant primary source but analytical works of repute have been digitized and stored as well. Another example would be the Cambridge collection of the Jonsonian canon that illustrates a peculiar relationship between conventional and modern electronic formatting. The primary texts are conventionally stored in a stable format while the analytical works are provided with editing abilities for users to create a dynamic text analysis along with a stably stored source.

5. Conclusion

The discourse analytical method of research focuses on the sites of unequal power distribution in discourses created in texts. It identifies the discourses of hegemony and hierarchy and makes apparent the so-called "common-sense" or naturalized injustices in the collective ideology of the society. The (auto)biographical research method can be used to identify the socio-cultural practices and ideology of the time contemporary to the authors in order to find these influences in the collaborated canon and separate from the fertile fields of reflections of personal experiences and interpretation of the culture. Notes can be made from the

(auto)biographies to indicate events of trauma or experience in the lifetime of these authors that might have been epiphanic and might have left impressions in their work. This can be validated by matching these impressions with those of the individual canon to ensure the validity of the implication of scholarship. Different versions of (auto)biographies can be matched to discern the change in temperament of the person or the age and the evidence of any latent episode. ICT as a research method has appeared in English Studies and allied fields because of the advent of computers into our lives. The use of ICT in research in this field holds many merits. First of all, it makes access to versions of the original manuscript possible and affordable, often free of any cost, to several thousand researchers simultaneously across the globe. In the process, it allows for the proper preservation of these invaluable texts. It also helps researchers save time by making the mechanical aspects of the research machine dependent to ensure accuracy and clinical authenticity. ICT allows for marking up and retrieval of pertinent excerpts from the text as evidence of the presence of a pattern of writing and thought. Used with proper coding and care ICT can help researchers in this field in unraveling age-old mysteries of authorship and scholarship. These three research methods can be undertaken together without conflict because of their similar nature and areas of inspection. All these methods will be used in complementation to indicate the presence of trace evidence of individuality in the collaborated canon as noticed in the individual canon. The use of three different methods minimizes the risk of any possible bias in the research and helps in achieving the ultimate goal of impartial analysis and accurate authorship attribution with a plethora of supporting evidence.

References:

Beach, Richard. "Research and the Literature Methods Class." *English Education*, vol. 8, no. 4, 1977, pp. 234–44. *JSTOR*, http://www.jstor.org/stable/23017259.

Cooper, Charles R. "Research Roundup: General English Curriculum and Methods." *The English Journal*, vol. 64, no. 5, 1975, pp. 94–96. *JSTOR*, https://doi.org/10.2307/814921.

Da, Delia, and W. R. Owens. *The Handbook to Literary Research*. Routledge [In Association With] The Open University, 2010.

Dawson, Catherine. *Introduction to Research Methods: A Practical Guide for Anyone Undertaking a Research Project*. 2002. 5th ed., Robinson, 2019.

Griffin, Gabriele, editor. *Research Methods for English Studies*. 2nd ed., Edinburgh University Press, 2013.

L'Eplattenier, Barbara E. "An Argument for Archival Research Methods: Thinking Beyond Methodology." *College English*, vol. 72, no. 1, 2009, pp. 67–79. *JSTOR*, http://www.jstor.org/stable/25653008.

PROF. Nilesh Popatlal Sathvara. *Research Methodology in English Literature*. 2020.

Sinha, M P. *Research Methods in English*. New Delhi, Atlantic Publishers & Distributors (P) Ltd, 2018.

Svenja Adolphs, and Dawn Knight. *The Routledge Handbook of English Language and Digital Humanities*. Abingdon, Oxon ; New York, Ny, Routledge, 2020.