Language Features of Author's Styles: A Comparison of the Work of Two Authors

Nuridinova Sabiya Alahidinovna

Teacher of Kimyo International University of Tashkent (Samarkand Branch)

Abstract. *In this article, the language features of the author's styles, their uniqueness, the importance of comparing the works of two authors are highlighted, and it is considered very relevant in the fields of culture, language, literature, art, science.*

Key words: *language, literature, art, enlightenment, linguistics, creative excellence, mistakes, author's styles.*

When an author writes, it's not just the words that have meaning. Instead, it's also how those words are stitched together to create a coherent work. After all, every author has a unique voice, or way of writing, that in turn affects the way with which their style is rendered. **Style** refers to how an author writes. When you read a text, do you imagine someone highly educated reading it to you? Or perhaps you imagine someone a little less refined. In either case, this is because the author intended for the text to be read with a particular voice. **Voice** refers to the distinct qualities that make a writer's work unique. So how do we figure out what the author wants us to know about style and voice? Luckily, we've got some clues to help us. First, let's make sure that the definition of style is completely understood. The best way of doing that is through looking at writing that has contrasting styles. To focus more on meaning, we'll look at this through warnings. Take, for example, this warning: "You should stay away from the dog." It is considered to be informal because it directly addresses the reader, but also uses a rather simple language. The style is easily mimicked, but what about voice? This is a bit harder. Voice, after all, refers to the overall way that a specific writer produces work. When you read a book by Mark Twain, you can almost hear an older gentleman with a deep drawl telling you the story. That's voice. Let's put it like this. Whereas style may describe the particular style of clothing someone wears, voice describes that person's personality. So how can we tell a writer's voice? Simple, by looking at many of the same clues given to us for style. Again, the way that they use language is important here, but it goes deeper. Does the author favor a very complex way of speaking, using compound-complex sentences with abandon and intricately weaving context clues for what is to come? Or is a simpler approach preferred? Maybe clues about the author's voice can be found in details as small as their preferred conjunctions? In any event, while style can often be described in a much broader way, voice involves closely examining language. In this lesson, we looked at how to analyze the language that an author uses in order to better understand their style and voice. Remember that **style** is the broad way in which an author writes, while **voice** refers to the specificities that make an author's writing stand out. Often, you'll have to look into word origins, sentence structure, and context clues to fully understand the author's voice, but a simple read-through will give you most of the information necessary to pinpoint style. What is Author's Style?

➢ Audience-listeners, readers, viewers of work

➢ Author's voice-sets writing apart from others; uniqueness of writing

> Genre-category/organization pattern of writing

> Purpose-intent of writing

> Mood-reader's emotion to the writing

> Tone-author's attitude about subject matter

> Sentence fluency-cadence, flow, transitions used

> Sentence variety-lengths/mixture of sentences

> Word choices-descriptions, figurative language, repetitions, synonyms...

Style and Language Analysis Guide

An explanation of writing style. An author's writing style is not incidental, superficial, or supplementary style identifies how ideas are embodied in language. In other words, the effect of how an author uses words and literary elements is important for understanding the meaning of a text. An author's writing style includes all of the items on the list below, including specific word choice (diction), kind of tone, use of formal or informal language, etc. The author adopts a variety of style elements depending on his or her purpose, audience, and genre. Analyzing an author's style involves understanding the particular way a text is written. Style in writing is not what is said but how it is said. Analyzing an author's style involves analyzing the writer's unique way of communicating ideas. Styles in writing are created deliberately by the author to convey a specific mood or effect. Style is often aligned with pathos, since its figures of speech are often employed to persuade through emotional appeals. However, style has just as much to do with ethos, for an author's style often establishes or mitigates one's authority and credibility. But it should not be assumed, either, that style simply adds on a pathetic or ethical appeal to the core, logical content. Style is very much part of the appeal through logos (appeal to logic and reason), especially considering the fact that schemes of repetition (e.g. outlines) serve to produce coherence and clarity, which are attributes of the appeal to reason. In other words, most pieces of writing have all three appeals (pathos, ethos, logos), but one or the other may be more dominate depending on the purpose of the piece of writing. In order to analyze a piece of writing, go through this list to evaluate how an author is using these styles elements. Then, choose which ones are most dominate in a particular piece that will help you to interpret the meaning of the piece, which is ultimately the entire point of doing a close evaluation of a piece of writing. Vocabulary/word choice: Are the words simple or fancy? Long/short. simple/complex, many modifiers/few modifiers? Are they technical, flowery, colloquial, formal, cerebral, lively, exciting, vivid, etc? Use of dialect, standard, non- standard English? Does the text or this passage make use of shocking, taboo language? Does the author pile on the details? Does author use slang or jargon specific to the topic? For example, does the writer utilize sports jargon to describe. Style- So, when analyzing an author's style, we need to consider:

> point-of-view,formal or informal writing,

> organization/structure of text,

> level of complexity in the writing, text.

In this study we investigate the style change of two Turkish authors, Cetin Altan and Yaşar Kemal, in their old and new works. In Altan's case old and new works are, respectively, defined as the columas published in the ten-year period 1960-1969, and the year 2000. In Kemal's case as an old work we use his 1971 novel "Bin Boğalar Efsanesi," (Kemal, 1971) and as a new work we use his 1998 novel "Firat Suyu Kan Akıyor Baksana" (Kemal, 1998). The objective measures of this study show that there is a significant difference between the old and new works of these authors within the context of the chosen works. The change can be attributed to a natural theme shift in the new and old works or a conscious style choice made by the writers. For example, Kemal indicated that in some different novels he used completely different languages because of their locales (Naci, 1999, p. 29). However, since what we use in style analysis is statistical patterns, which are unconsciously chosen by the authors, we hypothesize that style changes as suggested by the style markers are due to the time gap between the works, Altan and Kemal, whose works we study in this article, are among the

most welknown writers in Turkey. For example, Naci, in his collection of the reviews of one hundred Turkish novels of the century, included one work of Altan, and five works of Kemal, and all together sixty-two different writers (Naci, 1999). Writers would have more chance of (unconsciously) controlling their word lengths in agglutinative languages such as Turkish. Therefore, we hypothesize that word length occurrence frequency information is a good measure to use in stylo metric investigations in such languages. In this study first we compared the average word length between the old and new works of fixed size blocks for each author using a t-test. For a long time various statistical markers have been used to investigate the char- acteristics of artifacts in the humanities and line arts (Sedelow, 1970). A detailed. overview of the stylometry studies in literature within a historical perspective can be seen in Holmes (1994). It gives a critical review of numerous style markers. It also reviews works on the statistical analysis of change of style with time. A solid critique of many authorship studies is provided in Rudman (1998). The coverage of these studies is extensive and they come with broad reference lists. In stylometry studies about 1,000 style markers have been identified (Rudman, 1998). One of the oldest style markers is word length. For example, in 1901 Mendenhall published a well-known study using the word length frequencies and concluded that due to their style difference it was unlikely that Bacon could have written works attributed to Shakespeare. A later work showed that this conclusion could be false, since the style difference of these two authors could be due to different types of works used for comparison (prose of the former and verse of the latter author). This conclusion was based on the fact that in the writings of Sir Philip Sidney, a contemporary of Bacon and Shakespeare, the differences in word lengths between his prose and verse are very close to the differences found between Bacon's prose and Shakespeare's verse (Williams, 1975). However, Holmes (1985) does not give a positive recommendation for the use of word length frequencies in authorship attribution by pointing out the characteristics of Zipf's first law (Zipf, 1932). Another work (Tallentire, 1972) discusses the difficulty of using word length. frequencies in authorship studies. It is hard to find similar studies on word length as a style marker in agglutinative languages such as Turkish. In this study we use this style marker and show that it can be utilized successfully as suggested by its comparable (but not as good) results with those of most frequent word occurrence statistics. Our approach of using various style markers also matches Rudman's point that many different style markers need to be taken into account jointly (Rudman, 1998). An extensively used style marker is the frequency count of "context free" word(or similarly "most frequent words," and "function words"). The paper (Forsyth and Holmes, 1996) studies the use of five style markers (letters, most frequent words, most frequent digrams, and two methods of most frequent substring selection approaches) in ten stylometry problems (such as authorship, chronology, subject matter, etc.) with various levels of success.

The study states that frequencies of syntactic constructs lead to a higher classification accuracy. The work also states that syntax based methods are computationally expensive since they require syntactically annotated corpora. In this study we utilize the usage rate of most frequent words as a style marker. For analyzing the occurrence patterns of style markers various statistical methods are used. One popular technique in stylometric studies is principal component analysis. It can be easily appreciated using plots usually in two dimen- sions. For example, the Binongo and Smith (1999) study illustrates the use of the principal component analysis technique using occurrence frequency counts of two words, explains its intuition, and then uses it with several words in the author- ship study of one of Shakespeare's romances. We used this statistical technique to visually see the separation between old and new works using our style markers. Another statistical technique we used in this study is discriminant analysis that is also used in the literature for various purposes. For example, a recent work (Stamatatos et al., 2001) uses discriminant analysis and attacks the authorship detection problem using low-level measures (e.g., sentence length, punctuation mark count, etc.), syntax-based measures (e.g., noun phrase count, verb phrase count, etc.), and a set of style markers obtained by a natural language processing tool (e.g., percentage of rare or foreign words, a measure that indicates the morpho logical ambiguity, etc.). Additionally, they also use frequencies of most frequent words. The study is especially interesting due to its use of rich combination of style markers. The Baayen et al. (1996) paper mentioned before applies discrim- inant analysis to determine authorship attribution using syntax-based methods. In Holmes and Singh's (1996) paper, a

discriminant analysis is conducted to determine what measures of linguistic ability best discriminate aphasic patients from the normal person. The study reported in Holmes and Forsyth (1995) uses discriminant analysis to determine which vocabulary richness measures best discriminated between the papers written by Alexander Hamilton and those by James Madison. The study reported in Martindale and Tuffin (1996) uses it to find differences between Homer's Illiad and Odyssey.Another multivariate technique used was logistic regression. This technique isuseful for studying curvilinear relationships between a binary response variable and one or more predictor variables.

The study reported in Kessler et al. (1997) employs logistic regression in genre detection and compares its performance with a neural network approach. In our work, this method appears appropriate when the responsevariable is the classification of a work being "old" or "new" and the predictors are frequencies of word lengths or of the usage of "most frequent" words. In this section we first compared the average token and type lengths between the old and new works for each author using a t-test (or a one way analysis of variance). The results indicated that the average token and type length of the new works was significantly larger than that of the old for both authors. This gave us the motivation to perform a regression analysis using Altan's data to show that, as the age of the work increases, the average token length decreases. Then for each author, a series of logistic regressions were conducted to test for differences of token and type length frequencies between the old and new works. Then we compared the rate of usage (i.e., usage frequency) of most frequent words between old and new for both authors. We next performed a principal component analysis using our all style markers and then created scatterplots of principal component scores for cach data corresponding to a text block. This is to graphically illustrate the differ- ences hetween the old and new works by looking at their principal component scores. Finally we conducted a stepwise discriminant analysis to determine the best discriminators and then used cross validation to determine the success rate using these discriminators. All of these analyses were conducted using the SAS for Windows softwareA one way analysis of variance for each author was conducted on the blocks of the two work types to determine if there are differences in the average token and type lengths between the two. The results are summarized in Tables VI and VII for Altan and Kemal.

For both token and type, the average word lengths for each author were signiti- cantly larger for the new writings compared to the old. However, for Kemal the difference was not as dramatic as that for Altan. The above results motivated us to perform a regression analysis to determine whether a relationship exists between average token length and age of the work. We could only conduct the analysis using Altan's columns as a data source since they were written during several time points.

REFERENCES:

1. Altan C. (1997-20011 Seytanın Gör Dediği. Sabah Newspaper (hop://www.sabah.com.tr). Baayen, R.11. (2001) Wood Fagron Pos Kluwer Academic, Dordrecht. Boston Baayın H., Van Haltesen H. Tweedie F. (1996) Outside the Cave of Shadows. Using Syntactic Annotation to Enhance Authorship Attribution. Lisemcy and Linguine Compning, 1153) pp. 121-131

2. Binango JN.G.. Smith M.W.A. (1999) The Application of Principal Compoment Analysis to

3. Stylometry. Literary and Lingvistic Computing, 1441, 10445-400

4. Çiftlikçi, R. (1997) Nezar Kenmad Yuzu-Eser-asp Turk urk Tarih Konumu Basımevı, Antara, Fayyad U, Pitetsky Shapiro G., Smyth P. (1996) The KDD Process for Extracting Useful Knowledge from Volumes of Data, Communications of the ACM, 39011), pp. 27-34 Foltz P.W. Domais ST (1992) Personalized Information Delivery an Analysis of lnformation

5. Filtering Methods. Cut of the ACM. 35(12), pp. 51 60. Fonyth R.S., Holmes DL (1990) Face-finding Computing, H4, pp. 103-174 i for Text Classification. Literary and Linguime Hakkani Tür Z. (2000) Sutistical Modeling of Agglutinative Languages Ph.D. Dissertation, Department of Computer Engineering, Bilkent University, Ankara, Turkey. Heaps H.S. (1978) Гormation Retrieval Computational and Theoretical Aspects. Academic Press Inc., New York, New York

6. Hoimes D.J., Forsyth R.S. (1995) The Federalivi Revisted New Directions in Authorship Attributice. Literary and Lingrats Computing, 1012), pp. 111-1127, Holmes D.J. (1954:

Authorship Attribution, Computers and the Hammitres, 25.21. pp. 57-100 Holmes D.1. (1985) The Analysis of Literary Style A Review of the Royal Statical

7. 3ociers, Sener A. 148(4), pp. 328-341. Jain A.K., Marty M.N., Flyan PJ. (1999) Data Clustering: A. Review. ACM Comparing Surveye, 3031, pp. 264

8. Kemal Y. (1998) Fraoul Saye Kan Alvor Bak Adam Yaimları, Istanbul Kemal Y. (1971) Bin Boğafar Efronest. Cem Yaymeen, Istanbul. Kessler B. Geoffrey N. Schatze H. (1997) Automatic Detection of Text Genre. Proceedings of the

9. Anna Mering of the Association for Computational Lingausties and Me Conderence of the European Chapter of the Association for Computational Lineutomer, Madrid, pp. 32-38. Kiksal A. 119731 Automatic Morphological Anséesis of Toks Ph.D. Dissertation, Haarttepo University, Ankara, Turkey Kucera H., Francis W.N. (1967) Computil Alysis of Present Dey Amevilen En Brown University Press, Rhode Island

10. Martindale C. Tuffin P. (1995) If Homer is the Pon of the find then. He may not be the Poet of the Odyssey. Lherary and Linguistic Computing, 11(3), pp. 109-120. Noci F. (1999) Yazarlin 100 Tark Roman. Adam Yayınları, Istanbul