

Measuring Authorship - A Tribute to Forensic Discourse Analysis

Nejla Kalajđžisalihović
Faculty of Philosophy, Sarajevo
English Department
nejlak@gmail.com

Abstract: It is believed by many that our fingerprints are as unique as our DNA. Owing to the advances in modern technology and the aid of computers, it is possible to use software that is able to measure all the probabilities of occurrence of identical fingerprints, DNA, written or spoken discourse. In recent years, forensic discourse analysis experts and linguists have been trying to measure the degree to which every individual is unique. These findings are especially relevant for analysing the content of suicide letters, testimonies, testaments, ransom demands, confessions, SMS messages, diary entries etc. The quest of forensic discourse analysis is to apply the linguistic knowledge to the legal context with the aim of deciding on the authorship of the above-mentioned short notes. In applying the linguistic knowledge to the analysis of suicide letters, for instance, it is of great importance to determine whether there is a murder behind such a letter, viz. whether the letter is a genuine suicide letter. Another interesting phenomenon is related to testimonies, viz. the degree to which the interrogators added written content to the oral confession, or the degree to which the testimony, based on the linguistic evidence, is false. In this process, experts apply various methods of measuring the degree to which the testimonies are authentic. Some of these methods involve measuring sentence length average, word length average, collocations analysis, and forensic transcription.

The aim of this paper is to pay tribute to forensic discourse analysis of English texts and focus on some of its methods that are particularly related to the application of the linguistic knowledge. In doing so, we shall focus on a brief analysis of two well-known cases, Derek Bentley and Susan Smith.

Key words: forensic, transcription, word length average, sentence length average, collocation

Introduction

In recent years, there has been a rapid growth of interest in forensic linguistics, or forensic discourse analysis. The term 'forensic English', however, was first used in 1949 by Philbrick in *Language and the Law: Semantics of Forensic English* (Coulthard, 2007:5). In 1968, Jan Svartvik analysed the statements given by Timothy Evans, who was accused of murdering his wife and child. Svartvik, who used the term 'forensic linguistics' first, concluded that Evans did not give all the statements provided in the record written down by police. Namely, some of the statements were clearly distinctive due to their more formal style.

Another important founding father of forensic linguistics is Roger Shuy, whose contribution to the science is related to Miranda rights. Even today, a lot of research is being done on whether immigrants understand their rights. Shuy made it clear that an individual cannot testify non-voluntarily, especially if he/she does not understand his/her rights. Therefore, one of the major contributions of forensic linguistics to police interrogations is making sure that an individual's words are recorded correctly and not paraphrased.

Apart from police and courtroom-related issues, the main concerns of forensic linguistics are related to detecting plagiarism and attributing authorship to pieces of different types of written discourse. Especially popular is attributing authorship to SMS messages as it is sometimes found that a criminal is sending messages from a victim's phone. A similar analysis is applied when it comes to attributing authorship to e-mails.

Therefore, we can say that the focus of forensic linguistics is applying linguistic knowledge to the context of legal documents, courtroom interaction, speaker identification (SMS, e-mail, phone calls) and detecting plagiarism.

Methods applied in measuring authorship

A lot of emphasis has been placed on finding the best method for measuring and attributing authorship. So far, numerous statistical methods have been applied on finding the most accurate and

most reliable method for attributing authorship. One of the first scientists who attempted to compare two texts using forensic linguistics methods was Augustus de Morgan, 'who used the word length average as a marker' (Olsson, 2004:11). As for the sentence length average, it was U. Yule who considered it to be a 'viable marker' for attributing authorship (Olsson, 2001: 12). It is quite clear why word/ sentence length average cannot be considered the vital marker for attributing authorship unless we take into consideration the differences between spoken and written discourse, and some other factors.

A. Q. Morton claims that there is not much difference between the two if it is about the speaking/writing habits of an individual. This author is famous for his 'Qsum' or 'Cusum method'. Namely, in analysing shorter texts, Morton looks for vowel-initial words and two or three-letter words. After finding these values, one has to measure their distribution in the text, as well as in the sentence. These values should correspond to the values related to the average sentence length. If there is a discrepancy, that should imply that a piece of text has been inserted by another author. This method is purely statistical and, according to many experts, not considered to be reliable as it lacks the dimension of intuition, which is often important for the analysis.

Perhaps the most interesting approach is that of *forensic stylistics*. Namely, forensic stylistics compares texts in terms of (mis)spelling, the design of the pages, the space between the words, handwriting, collocations, word frequency, capitalisation, indentation, etc. (See: G.R. McMenamin, 2002).

The scientific evidence or what is considered to be valid at courts is sometimes not clearly defined. As Olsson points out, although there are thousands of references to the subject of forensic linguistics, the question of whether or not we *have* the linguistic fingerprint is still unsolved. Namely, in *Studies and Authorship Recognition: A Corpus-based Approach* (1998), Hänlein discusses the possibility of recognizing the stylistic profile of an individual (Olsson, op.cit. p. 27). The recognition of the stylistic profile depends on whose stylistic profile it is as more language aware individuals are more able to switch codes or adapt their linguistic choices to the register or the context. In the lines that follow, we shall focus on the approaches given by J. Olsson, as he is one of the experts who thoroughly analyses most of the above-mentioned theories. We are, namely, going to focus on the already solved cases of Derek Bentley and Susan Smith to show how it is possible to determine whether some parts of a testimony have or have not been inserted.

Forensic transcription - calculating word and sentence length average

The piece of text that can be processed for forensic analysis may be a page from a diary, an e-mail, a post-it note, a letter, etc. In case it is necessary to transcribe a text that was hand-written, one has to observe certain regulations when transcribing icons (e.g. smiley), exclamation or question marks, parts or words that were erased, etc. One of the most reliable methods of forensically transcribing a piece of text is transcribing it manually to a Word document. In addition to that, there is a number of software platforms that ease the transcription process as they are designed to find collocations (viz. colorcations) or calculate probability (e.g. Copy Catch).

The word length average is calculated by counting the number of characters in a text from which all the punctuation marks have been removed and dividing the number with the number of words. The number is usually reduced to two decimals. As for the sentence-length average, it is calculated by counting the number of words in a sentence and dividing it with the number of sentences in the text.

Data analysis- Susan Smith confession and Derek Bentley statement

In 1994, Susan Smith, while trying to end both her life and the life of her two sons, admitted to having killed her children by letting her car roll down into a lake. After being interrogated, Susan Smith wrote a confession and confessed the crime.

The average sentence length of the whole text is 14 words, whereas the average word length is 3.9 words. In total, the text contains 568 words, 2.173 characters and 39 sentences. Part one contains 25 sentences. The average word length for Part 1 is 13 words, whereas the average word length is 3.7. Part 2 contains 14 sentences. The average sentence length is 17 words, whereas the average word length is 3.9.

We know for sure that Susan Smith herself wrote the statement using some formal phrases/collocations she may have heard (e.g. *emotionally distraught*) from police officers. The whole text is an emotional rollercoaster ending in statements of justification and self-evaluation. The author is also inconsistent spelling-wise ('he' vs. 'He'),

which suggests that she is either distraught or unsure about correct spelling. The deleted parts occur at very specific spots and seem to be deliberate at hiding something that is not supposed to be revealed. Also, it is unusual that the words referring to religious imagery occur quite often. The second part has word length longer by 0.2 and sentence average by 4 words. Although there are variations in sentence and word length average, her idiolect is quite specific in terms of repetition and consistency in grammar and spelling errors (See: Appendix A).

As for Derek Bentley, a British teenager who was sentenced to death for being in the company of the juvenile Christopher Craig when Craig shot a police officer (1952), it has been confirmed that Bentley's statement given to the police was not properly recorded (See: Appendix, B). Analysing the text, and keeping in mind that Bentley suffered from epilepsy and had low intelligence quotient, it is apparent that the confession is not genuine, viz. as given by Derek Bentley. It is apparent that there are at least two authors of the statement as it is unlikely that Bentley could have remembered the date, the time and other details related to the event at the time of interrogation.

In the part that we refer to as DB1 (the text that starts with 'I have known Craig' and ends with 'I then ran after them'), there are 8 sentences and 483 characters. The average word length is 4.0, whereas the average sentence length is 15 words. In the part that we shall refer to as DB 2 ('There was a little iron gate' to 'he was going to use the gun'), there are 9 sentences and 346 characters. The average word length is 3.9 and the average sentence length is 10 words.

DB1 is significantly different from DB2 for several reasons. First of all, in DB1, it seems that the author had enough time to precisely remember the date and the time, as well as the order of events. Secondly, since the author had enough time to think about the setting, he can use tense agreement properly, viz. he is using *Past Perfect Tense* together with *Past Simple Tense*. Also, he is using afterthoughts, separated with dashes. The use of indirect speech shows that this is not an immediate reaction to interrogation. Also, in DB2, the personal pronoun 'I' occurs nine times, whereas in DB 2, 'I' occurs only three times (Coulthard points out that the 'I then' string is found in police-written statements). However, in DB2, the author is using shorter sentences, resembling spoken language. The events are put in an order, and they seem more immediate to the reader. The author is using direct speech and *Past Simple Tense*, a well-known pattern of economy in language when retelling recent events in the past. He is also not as precise as the author of DB1 ("for about ten minutes").

A comparison between DB1 and DB2 indicates that DB2 is the original statement given by Derek Bentley, while DB1 was inserted afterwards, as DB1 has elements of precision found in police statements. As for events and actions, DB1 is more focused on the *events*, where the narrator is a patient (and not an agent). DB2, however, is more action-oriented, viz. both the narrator and his colleague are active participants in the event. We propose that it is possible to analyse another part of DB2, the part that we shall refer to as DB2.a (the string from 'The policeman dragged him' to the end), or the answers to police interrogation written down.

Conclusion

The above-given cases of Susan Smith and Derek Bentley are presented with the aim of stressing the importance of forensic discourse analysis when analysing statements given to the police or at court. In applying the linguistic knowledge to the analysis of corpora, it is of great importance to determine the degree to which the interrogators added written content to the oral confession, or the degree to which the testimony, based on the linguistic evidence, is false. In this process, experts apply various methods of measuring the degree to which the testimonies are authentic. Some of these methods involve measuring sentence length average, word length average, collocations analysis, and forensic transcription. However, there is not a single method that can be used for all the textual or phonetic evidence. Apart from linguistic and statistical evidence, profiling an individual's style and analysing the context or purpose for which a particular piece of text (or audio material) was created could be of great importance for discovering the vital cues. The two cases do not differ much in terms of the variations of sentence and word length. However, in terms of orthography, idiolect and style, it is evident why the authorship of the Bentley statement stirred so much debate. Further analysis of these and other texts (such as authorship reports in percentages) is beyond the scope of this paper, but it is important to point out that, for forensic discourse analysis, the roles of forensic stylistics and statistics are equally important.

In studies that follow, our aim is to apply the forensic knowledge to the context of students' papers and to the analysis of authorship and instances of plagiarism. Our aim is to stress the importance of proper language acquisition as it is a vital step towards increasing language awareness.

Note: Parts of the analysis presented above are taken from assignments the author of the article submitted to The Forensic Linguistics Institute (Powys, UK) in March 2010.

References:

- Coulthard, M. (1994). 'Forensic discourse analysis', in: *Advances in Written Text Analysis*, M. Coulthard (ed.). London: Routledge, pp. 242-258.
- Coulthard, M. & Johnson, A. (2007). *An Introduction to Forensic Linguistics, Language in Evidence*. New York, New York, NY: Routledge
- Coulthard, M. & Johnson, A. (2010). *The Routledge Handbook of Forensic Linguistics*. New York, NY: Routledge
- Gibbons, J. & Turell T. (2008). *Dimensions of Forensic Linguistics*. Amsterdam: John Benjamins.
- Hänlein, H. (1998). *Studies in Authorship Recognition- A Corpus-based Approach*. Frankfurt: Peter Lang.
- McMenamin, G.R. (2002). *Forensic Linguistics, Advances in Forensic Stylistics*. Florida: CRC Press
- Olsson, J. (2008). *Forensic Linguistics - The Language Detective*, Unit 1. Powys, UK: Forensic Linguistics Institute.
- Olsson, J. (2004). *Forensic Linguistics - An Introduction to Language, Crime and the Law*. London: Continuum.
- Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax*. Oxford: Blackwell.
- Svartvik, J. *The Evans Statements*. Gotheburg Studies in English No. 20.
- Yule, G. Udney (1944). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press

Web:

- Coulthard, M. "Identifying the Author." Web. 24 March 2011. <<http://clf.unige.ch/display.php?idFichier=168>>.
- "Analysis of Susan Smith's Confession." *LSI Laboratory for Scientific Interrogation, Inc.* Web. 14 Jan. 2011. <http://www.lsiscan.com/susan_smith_s_confession.htm>.
- "Forensic Linguistics Institute." *Forensic Linguistics Institute - The Home of Forensic Linguistics*. Ed. John Olsson. Web. 14 Jan. 2011. <[http://www.thetext.co.uk/cgi-bin/view_texts.pl?dir=&folder=Confessions&text=Derek Bentley's Police Statement.txt](http://www.thetext.co.uk/cgi-bin/view_texts.pl?dir=&folder=Confessions&text=Derek+Bentley's+Police+Statement.txt)>.
- "How Rare Is That Fingerprint? Computational Forensics Provides the First Clues - UB NewsCenter." *University at Buffalo*. 7 Dec. 2010. Web. 14 Jan. 2011. <<http://www.buffalo.edu/news/12073>>.

Articles:

- Dugandžija, M. (2011). Napisao je točku umjesto zareza. Po tome su otkrili ubojicu. *GLOBUS*, No. 1048, 68-7
- Durrant P. & A. Doherty (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2).
- Guillen-Nieto, V. (et.al.) (2008). Exploring State-of-the-Art Software for Forensic Authorship Identification. *IJES*, Vol. 8 (1), pp. 1-28.

Appendix:

A) Susan Smith confession

When I left my home on Tuesday, October 25, I was very emotionally distraught. I didn't want to live anymore! I felt like things could never get any worse. When I left home, I was going to ride around a little while and then go to my mom's. As I rode and rode and rode, I felt even more anxiety coming upon me about not wanting to live. I felt I couldn't be a good mom anymore but I didn't want my children to grow up without a mom. I felt I had to end our lives to protect us all from any grief or harm (deletion). I had never felt so lonely and so sad in my entire life. I was in love (underlined) with someone, very much, but he didn't love me and never would. I had a difficult time accepting that. But I had hurt him very much and I could see why he could never love me. When I was @ John D. Long Lake, (deletion) I had never felt so scared and unsure as I did then. I wanted to end my life so bad and was in my car ready to go down that ramp into the water and I did go part way, but I stopped. I went again and stopped.

I then got out of the car and (deletion) stood by the car (insertion>a) nervous wreck. Why was I feeling this way? Why was everything so bad in my life? I had no answers to these questions. I dropped to the lowest when I allowed my children to go down that ramp into the water without me. I took off running and screaming go back, but I knew it was too late. I was an absolute mental case! I

couldn't believe what I had done. I love my children w/ all my (icon: heart). That will never change. I have prayed to them for forgiveness and hope that they will find it in their (icon: heart) to forgive me. I never meant to hurt them!! I am SORRY for what has happened and I know that I need some help. I don't think I will ever be able to forgive myself for what I have done. My children, Michael and Alex, are with our Heavenly Father now and I know that they will never be hurt again. As a mom, that means more than words could ever say.

I knew from day one, the truth would prevail, but I was so scared I didn't know what to do. It was very tough emotionally to sit and watch my family hurt like they did. It was time to bring a piece of mind to everyone, including myself. My children deserve to have the best and now they will. I broke down on Thursday, November 3 and told Sheriff Howard Wells the truth. It wasn't easy, but after the truth was out, I felt like world was lifted off my shoulders. I know now that it is going to be a tough and long road ahead of me. At this very moment, I don't feel I will be able to handle what's coming, but I have prayed to God that he give me the strength to survive each day and to face (illegible) times and situations in my life that will be extremely painful. I have put my total faith in God and He will take care of me.

Susan V. Smith

11/3/94

B) Derek Bentley statement

I have known Craig since I went to school. We were stopped by our parents going out together, but we still continued going out with each other - I mean we have not gone out together until tonight. I was watching television tonight (2nd November 1952) and between 8pm and 9pm Craig called for me. My Mother answered the door and I heard her say I was out. I had been out earlier to the pictures and got home just after 7pm. A little later Norman Parsley and Frank Fazey called. I did not answer the door or speak to them.

My Mother told me that they had called and I then ran out after them. I walked up the road with them to the paper shop where I saw Craig standing. We all talked together and then Norman Parsley and Frank Fazey left. Chris Craig and I then caught a bus to Croydon. We got off at West Croydon and then walked down the road where the toilets are - I think it is Tamworth Road. When we came to the place where you found me, Chris looked in the window. There was a little iron gate at the side. Chris then jumped over and I followed. Up to then Chris had not said anything. We both got out on to the flat roof at the top. Then someone in a garden on the opposite side shone a torch up towards us. Chris said: "It's a copper, hide behind here." We hid behind a shelter arrangement on the roof. We were there waiting for about ten minutes. I did not know he was going to use the gun. A plain clothes man climbed up the drainpipe and on to the roof. The man said: "I am a police officer - the place is surrounded." He caught hold of me and as we walked away Chris fired. There was nobody else there at the time. The policeman and I went round a corner by a door. A little later the door opened and a policeman in uniform came out. Chris fired again then and this policeman fell down. I could see he was hurt as a lot of blood came from his forehead just above his nose.

The policeman dragged him round the corner behind the brickwork entrance to the door. I remember I shouted something but I forget what it was. I could not see Chris when I shouted to him - he was behind a wall. I heard some more policemen behind the door and the policeman with me said, "I don't think he has many more bullets left." Chris shouted "Oh yes I have" and he fired again. I think I heard him fire three times altogether. The Policeman then pushed me down the stairs and I did not see any more. I knew we were going to break into the place, I did not know what we were going to get - just anything that was going. I did not have a gun and I did not know Chris had one until he shot. I now know that the policeman in uniform is dead. I should have mentioned that after the plain clothes policeman got up the drainpipe and arrested me, another policeman in uniform followed and I heard someone call him 'Mac'. He was with us when the other policeman was killed.