Paul is Dead? Differences and Similarities before and after Paul McCartney's Supposed Death. Stylometric Analysis on Transcribed Interviews

Antonio Pascucci, Raffaele Manna, Johanna Monti
L'Orientale University of Naples
UNIOR NLP Research Group
Naples, Italy
[apascucci,rmanna,jmonti]@unior.it

Vincenzo Masucci
Expert System Corp. Naples, Italy
vmasucci@expertsystem.com

Abstract

English. In this paper, we show the results of a stylometric analysis conducted on Paul McCartney's interview transcriptions using three different approaches in order to detect differences and similarities in his speeches before and after 9th November 1966, the date of his supposed death. Our research is based on the *Let IT Corpus*, a corpus of Paul McCartney's Interview Transcriptions. The corpus is a collection of texts from the *Beatles Interviews Database*, a repertoire of one-hundred sixty-three Beatles' interviews freely available on the Web (http://www.beatlesinterviews.org/), and from other interviews available on YouTube.

Italiano. In questo contributo mostriamo i risultati di un'analisi stilometrica con tre ap-procci differenti operata sulle trascrizioni delle interviste di Paul McCartney con lo scopo di individuare analogie e differenze stilistiche nelle trascrizioni delle interviste fatte prima e dopo il 9 novembre 1966, data della sua presunta scomparsa. La ricerca si basa sul *Let IT Corpus*, un corpus composto da trascrizioni di interviste fatte a Paul McCartney che abbiamo costruito con le trascrizioni delle interviste presenti sul *Beatles Interviews Database*, una raccolta di centosessantatré interviste ai Beatles disponibile su internet (http://www.beatlesinterviews.org/) e di altre interviste disponibili su YouTube.

1 Introduction

Paul McCartney's supposed death (dated 9th November 1966 because of a car accident) represents a legend which does not belong to the music business only but embraces other worlds, both for Paul McCartney's fame given by Beatles' everlasting success, and because of many stories born around this episode. Paul is dead (PID) theory represents one of the most controversial legends in the history of music, enough to be still debated after more than half a century, during which numerous stories are born, some feeding and some other damping its truth. In this paper, we show the results of a stylometric analysis conducted on Paul McCartney's interview transcriptions using three different approaches in order to detect differences and similarities in his speeches before and after 9th November 1966. Our research is based on:

- the Beatles Interviews Database (http://www.beatlesinterviews.org/), a collection of one-hundred sixty-three interviews from 1962 to 1984;
- YouTube subtitles, which we manually corrected, if necessary, listening to the audio.

To the best of our knowledge, this research represents one of the very first stylometric analyses on interview transcriptions. In this paper, we also present the *Let IT Corpus* (Paul McCartney's Interview Transcriptions), composed of fifty-two documents concerning interviews before 9th November 1966 and fifty-two documents concerning interviews after 9th November 1966. *Let IT Corpus* is still in its embryonic stage: we foresee to expand it with further texts so that it can be used for more accurate analyses in the near future.

The strongest supporters of PID theory claim that immediately after his death, Paul McCartney has been replaced by a lookalike. There are several theories that even today sustain the veracity of PID, many of which spread by the Beatles themselves, who sometimes enjoyed including subliminal messages in their songs. For example the celebrated John Lennon's whisper in the song *I'm so tired*, if listened backwards, seems to say *Paul is dead, man: miss him! miss him! miss him!*. The Abbey Road's album cover also shows at least ten references to Paul McCartney's death. On the other hand, some theories remove all doubts, claiming that Paul McCartney never died and all PID hypotheses are nothing but a business choice, which has contributed to add extra charm to Beatles' success. In October 1969, the Beatles' press office categorically denied PID rumours, labelling them as *a load of old rubbish*. PID theory has been investigated in literature (Cartocci, 2005) and in automatic-recognition (Holland et al., 2014). The present contribution is organized as follows: in Section 2 we show Related Work. The *Let IT Corpus* is described in Section 3, in Section 4 we describe three different approaches adopted in the analysis and their results. In Section 5 the stylistic differences and similarities detected by the linguistic analysis are thoroughly discussed. Conclusions are in Section 6. In Section 7 we introduce Future Work.

2 Related Work

CS is the statistical analysis of writing style (Zheng et al., 2006) and it is used to identify or profile the author of a text. The main assumption of Authorship Attribution (AA) is that each author operates choices which are influenced by sociological (age, gender and education level) and psychological (personality, mental health and being a native speaker or not) factors (Daelemans, 2013) which determine a unique writing style. In AA some studies are being conducted on speech transcriptions. In 2014 (Herz and Bellaachia, 2014) investigated the authorship of Barack Obama's speechwriters on a corpus composed by thirty-seven speech transcriptions. They based their research on the supposition that Barack Obama has four principal speechwriters and deal with the AA of Barack Obama's speeches with four different approaches, that reached different results, but still showing that CS can be used to differentiate authors who write in a similar style. (Airoldi et al., 2006) conducted a similar research on Ronald Regan's radio speeches. The corpus they used for their investigation is composed of a thousand thirty-two radio addresses delivered by Ronald Reagan between 1975 and 1979. The scholars focused the experiment on three-hundred twelve radio addresses for which no direct AA evidence is available, and they concluded that in 1975, Ronald Reagan drafted seventy-seven speeches and his collaborators drafted seventy-one, whereas over the years 1976-1979, Ronald Reagan drafted ninety speeches and his collaborator Peter Hannaford drafted seventy-four speeches. The study of (Herz and Bellaachia, 2014) and that of (Airoldi et al., 2006) share a problem: it is not possible to know the accuracy of the AA results of their study.

CS is also useful in studying changes in the style of an author over time. As argued by (Rybicki, 2015) time is one of the most significant factors for the evolution of the literary lexicon. With this in mind, some researches are conducted on stylochronometry (for a survey, see (Stamou, 2007)), namely the study of the change of style correlated to the passing of time. (Forsyth, 1999) differentiates the style of the poet William Butler Yeats between *younger Yeats* and *older Yeats*, devising along the way a measurement he calls a *youthful Yeatsian Index*. (Van Hulle and Kestemont, 2016) use sylometry to periodize Samuel Beckett's works, finding stylistically innovative change in his late style. Lastly, the findings of (Evans, 2018) show that the dramatic style of Aphra Behn over the course of her 20-year career, can be divided in three different phases. Obviously we must keep in mind that our analysis is based on transcription of speeches, and therefore not on written texts. Until now, to the best of our knowledge, no stylistic research analysis has been carried out to detect differences and similarities in interview transcriptions before and after Paul McCartney's supposed death.

3 *Let IT Corpus*

For our research we investigated the *Beatles Interviews Database*[1], a collection of one-hundred sixty-three transcription of Beatles' interviews from 1962 to 1984 created in 1997 by Jay Spangler and now managed by Jude Southerland Kessler and Suzie Duchateau. The website also contains a songwriting and

[1] http://www.beatlesinterviews.org/

recording database, a collection of Beatles' movies, quotes and pictures. We also investigated thirty-five Beatles' interviews available on YouTube: in this case we analyzed the automatic captions generated by speech recognition, and we corrected texts if necessary. In each interview, we isolated Paul McCartney's speeches and we created a document for each interview. The *Let IT Corpus* is a very small balanced corpus composed of one-hundred four documents belonging to two different classes: I) *before* (composed by fifty-two documents concerning interviews before 9th November 1966) and II) *after* (composed by fifty-two documents concerning interviews after 9th November 1966). A few texts belonging to the *after* class found on YouTube date after 2000. The majority of texts of the *Let IT Corpus* are from the *Beatles Interviews Database* (32 *before* texts and 25 *after* texts, including a few chunks). The corpus contains also texts from the *Beatles Interviews Database* concerning interviews involving the whole Beatles group, from which we isolated Paul McCartney's speeches. The remaining part of *Let IT Corpus* consists of Beatles' interviews freely available on YouTube. *Let IT Corpus* is still in its embryonic stage, since it is composed of approximatively one-hundred texts and it represents the first step in this field. Further work will be carried out as soon as *Let IT Corpus* will be expanded.

4 Our three approaches to stylometric analysis

We investigated this AA issue with three different approaches, in order to compare the results. For all the experiments we removed punctuation and symbols, and we lowercased all characters.

4.1 Hybrid approach

In this section we describe the first approach to stylometric analysis, namely a hybrid approach based on CS, Linguistic Rules and Machine Learning (ML). Thanks to the analysis of approximatively five thousand English documents from a variety of sources (newspapers, social media and books) we identified several stylistic features that we used to write linguistic rules for English. Here we report a short list of stylistic features: sentence length (Argamon et al., 2003), vocabulary richness (De Vel et al., 2001), word length distributions (Zheng et al., 2006), punctuation (Baayen et al., 1996), use of a specific class of verbs or adjectives, use of first/third person. The hybrid approach of CS, Linguistic Rules and ML consists in the following steps: I) *Linguistic Definition of Stylometric Features*: starting from the assumption that each author operates different grammatical choices when writing a text (Daelemans, 2013), we organized the grammatical characteristics of the case-study language (in this case, English) in a taxonomy. The work was carried out thanks to COGITO® by Expert System Corp., a semantic analysis software based on Artificial Intelligence algorithms. In each limb of the taxonomy it is possible to write linguistic rules concerning the language of the case study in order to recognize the grammatical characteristics of the analyzed texts (i.e. to detect modal verbs, we create the limb "modal verbs" and we associate to it linguistic rules that allow to find modal verbs in the texts); II) *Semantic Engine Development*: Expert System's semantic engine is trained in order to extract the aforementioned features from texts and is implemented thanks to COGITO®'s semantic network (called *Sensigrafo*); III) *Features Extraction*: texts are analyzed and all features (based on the grammatical characteristics of the texts) are extracted; IV) *Supervised ML Process*: the features extracted are used to train the model in order to detect the features in the untagged texts. For ML process we exploit WEKA (Hall et al., 2009), a software with ML tools and algorithms for data analysis.

The hybrid approach is evaluated through the 10-folds Cross Validation method. We tested two different algorithms, Random Forest (RF) and Tree J48 (J48). During previous AA investigations RF resulted to be the most performing algorithm for a binary classification. The results we obtained for 10-folds Cross Validation test confirm this result and Table 1 presents the performances in terms of Precision, Recall and F-Measure for both algorithms (namely, RF and J48). In order to evaluate the performances of the classifier, after this process, we tested both RF and J48, as well as 10-folds Cross-Validation (Table 1). Compared to the results obtained for the 10-folds Cross Validation (see Tables 2 and 3), J48 performances (Table 3) are better than RF performances (Table 2).

10-folds Cross Validation (RF)	Precision	Recall	F-Measure
	0.815	0.824	0.808
10-folds Cross Validation (J48)	Precision	Recall	F-Measure
	0.779	0.784	0.781

Table 1: 10-folds Cross Validation on the whole corpus with RF and J48

Test Set 80-20 (RF)	Precision	Recall	F-Measure
	0.781	0.750	0.764

Table 2: Performances in terms of Precision, Recall and F-Measure with 80% of the corpus as Training set and the remaining 20% as Test set randomly selected with the support of RF algorithm

Test Set 80-20 (J48)	Precision	Recall	F-Measure
	0.853	0.800	0.819

Table 3: Performances in terms of Precision, Recall and F-Measure with 80% of the corpus as Training set and the remaining 20% as Test set randomly selected with the support of J48 algorithm

4.2 Support Vector Machine (SVM)

For our second approach we exploited SVM with a Bag-of-Words (BoW) features set created using TF-IDF vectorization. As stated by (Diederich et al., 2003) SVM is capable to process thousands of inputs, which allows to use all the words of a text directly as features. SVM involves building a decision boundary to separate the data into classes (in our case, *before* and *after*), which may be non-linear if the kernel trick is used to transform our existing data into a higher dimensional space. As such, the right choice to take when fitting an SVM classifier is kernel in addition to others hyperparameters specific to that kernel. In applying SVM to AA, (Schwartz et al., 2013) used a linear kernel, while (Diederich et al., 2003) examined a range of different kernels. Since our AA is a binary classification problem we used the linear kernel for our model and considered C values in the set {1, 10, 100}. The optimal value of C was determined using GridSearchCV function with a default 3-fold Cross-Validation and accuracy used as the scoring metric. The optimal C value was determined to be C = 1. Results are in Tables 4 and 5.

SVM-BoW	Precision	Recall	F-Measure
	0.785	0.761	0.773

Table 4: 10-folds cross validation SVM - BoW features set.

SVM-BoW	Precision	Recall	F-Measure
	0.885	0.809	0.818

Table 5: Performances in terms of Precision, Recall and F-Measure with 80% of the corpus as Training set and the remaining 20% as Test set randomly selected.

4.3 Convolutional Neural Network (CNN)

To deal with the problem of AA of speech transcriptions, our third approach consists in a two-class text classification based on a deep CNN. We built a neural network that exploits the morpho-syntactic

information to improve the classification and correctly identify the given samples. The input data are preprocessed and tagged with linguistic information using the Part-of-Speech (PoS) tagger provided by the free NLP open-source library *Spacy*. Given the importance of function words (Kestemont, 2014), conjunctions, prepositions, interjections, adverbs and auxiliary verbs were taqken into account for this analysis. In fact, as proved by (Mosteller and Wallace, 1963) and confirmed by (Koppel et al., 2006), function words are discriminators of authorship, since the usage variations of such words are a strong reflection of stylistic choices. Our proposed architecture receives a sequence of tagged texts as input and then is transformed into padded sequences of fixed length. The sequences are then processed by four modules: an embedding module, a convolutional module and two max pooling layers to consolidate the output of the convolutional layer. The output of the three modules are processed by one Dense layer and an output layer. Results are shown in Tables 6 and 7.

CNN-PoS	Precision	Recall	F-Measure
	0.681	0.734	0.706

Table 6: 10-folds cross validation CNN + PoS.

CNN-PoS	Precision	Recall	F-Measure
	0.692	0.818	0.750

Table 7: Performances in terms of Precision, Recall and F-Measure with 80% of the corpus as Training set and the remaining 20% as Test set randomly selected.

5 Differences and Similarities before and after 9th November 1966

Thanks to the linguistic analysis of the texts belonging to the two different classes, we detected stylistic differences and similarities in the speech transcriptions before and after 9th November 1966. We started by dividing Paul McCartney's interviews into separate sentences. A number of stylistic features are extracted from these sentences and then all features are used for K-Means clustering. Here we report a list of some features extracted: number of function words, number of verbs and a number of interjections. For clustering, the average of each feature is calculated. Further, a SVM classifier is trained on 70% of the interviews and tested on the remaining 30%. Performing this process means to see whether a link is present and consistent over time through Paul McCartney's style. Accuracy is shown in Table 8.

Accuracy on test set
0.561

Table 8: SVM to test stylometric similarity

Here we report some examples of interview transcriptions before and after 9th November 1966 and we highlight the most noticeable differences and similarities. It is very important to consider that all the interviews collected are from different sources (TV, radio, newspaper), that means that speeches can differ from source to source as well as according to the historical moment in which they were done.

- You know like, number one records, Sunday Night At The Palladium, Ed Sullivan Show, go to America, you know. All kinds of ambitions like that. (**Carnegie Hall - New York, 1964 February 12th**);
- The only thing is that we've gotta do a lot from London, 'cuz a lot of the TV shows are down in London, you know. And so, we're forced to do a lot down in London. I mean, it's like someone said the other day Why doesn't Harry Secombe go to Cardiff? You know, he never does. But no

one ever moans about Harry never going...You know what I mean? (**BBC-TV by Gerald Harrison - Liverpool, 1964 July 10th**);

- Personal differences, business differences, musical differences, but most of all because I have a better time with my family. Temporary or permanent? I don't really know. (**Break-up - 1970 April 10th**);
- It was like a gesture to Russia because normally records are released first in America and England in Europe and then Russia gets them last and because Gorbachev and Reagan were talking about glasnost and we're talking about arms reduction. I think a lot of us in Europe were very happy to hear this so I had the opportunity to release this record so I wrote a little note on the record saying this is the peace gesture the hand of friendship from the west to the east and I just felt it might just help a bit of glasnost it's my little bit of glasnost. (**Flemish Public Television Interview - 1989**)

As we can see, slang expressions and fillers such as *'cuz*, *I mean*, *You know?* and *You know what I mean?* completely disappear in interviews after 9th November 1966. The use of slang disappears also in other interviews after this date, in which we can find a different Paul McCartney, who seems to be more serious and not only because of an older age. Changes can be brought about by the different topics addressed in the interviews, but we also believe that speech preserves some characteristics (such as slang) in different contexts. In texts belonging to the *after* class, sentences are longer compared to those of the *before* class. We noticed also that in texts belonging to the *after* class style changes occur continously not allowing for the identification of a specific style. For these reasons we also report the date and the source. Our research highlights some similarities in *before* and *after* texts: the overuse of expressions such as *We are gonna do* and *a lot/a lot of* is confirmed in both periods. These represent the most used expressions by Paul McCartney in his speeches. In the interviews in the *Let IT Corpus* we also noticed that Paul McCartney is inclined to rely on lists both in *before* and in *after* periods.

6 Conclusions

In this paper we have presented the *Let IT Corpus*, namely a corpus of one-hundred four transcriptions from speech to text of Paul McCartney's interviews collected from the *Beatles Interviews Database* and YouTube. The aim of this research is to detect possible differences and similarities in Paul McCartney's speeches before and after 9th November 1966 (date of his supposed death). For this reason texts have been organised in two classes: I) *before* and II) *after*. We investigated three different text classification approaches and we detected that all methods achieved high percentage of accuracy classifying texts in two different classes referring to two different periods. To reinforce these results and on the basis of the analysis of the stylistic features set out above, it is clear that the way of modulating the words of Paul McCartney is quite distinguishable between the two periods examined.

7 Future Work

The corpus is in its embryonic stage, since it is composed of approximatively a hundred texts. Future work therefore concerns the expansion of the *Let IT Corpus*, so to allow a more thorough investigation. To corroborate our hypothesis it might be interesting to see if the differences we detected between the two classes represent a pure coincidence. A possible experiment in this respect can be carried out considering a different temporal division of the texts.

Acknowledgements

This research has been partly supported by the PON Ricerca e Innovazione 2014-20 and the POR Campania FSE 2014-2020 funds. Authorship contribution is as follows: Antonio Pascucci is author of Sections 1, 2, 3, 4.1 and 7 and Raffaele Manna is author of Sections 4.2 and 4.3. Section 5 is in common. This research has been developed in the framework of two Innovative Industrial PhD projects in Computational Stylometry (CS) by "L'Orientale" University of Naples in cooperation with Expert System Corp. We are grateful to Vincenzo Masucci and Expert System Corp. for providing COGITO® for research and to Prof. Johanna Monti for supervising the research.

References

Edoardo M Airoldi, Annelise G Anderson, Stephen E Fienberg, Kiron K Skinner, et al. 2006. Who wrote ronald reagan's radio addresses? *Bayesian Analysis* 1(2):289–319.

Shlomo Argamon, Marin Šarić, and Sterling S Stein. 2003. Style mining of electronic messages for multiple authorship discrimination: first results. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pages 475–480.

Harald Baayen, Hans Van Halteren, and Fiona Tweedie. 1996. Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution. *Literary and Linguistic Computing* 11(3):121–132.

Glauco Cartocci. 2005. *Il caso del doppio Beatle: il dossier completo sulla" morte" di Paul McCartney*. Robin Edizioni IT.

Walter Daelemans. 2013. Explanation in computational stylometry. In *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, pages 451–462.

Olivier De Vel, Alison Anderson, Malcolm Corney, and George Mohay. 2001. Mining e-mail content for author identification forensics. *ACM Sigmod Record* 30(4):55–64.

Joachim Diederich, Jörg Kindermann, Edda Leopold, and Gerhard Paass. 2003. Authorship attribution with support vector machines. *Applied intelligence* 19(1-2):109–123.

Mel Evans. 2018. Style and chronology: A stylometric investigation of aphra behn's dramatic style and the dating of the young king. *Language and Literature* 27(2):103–132.

Richard S Forsyth. 1999. Stylochronometry with substrings .

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explorations* 11(1):10–18.

Jonathan Herz and Abdelghani Bellaachia. 2014. The authorship of audacity: Data mining and stylometric analysis of barack obama speeches. In *Proceedings of the International Conference on Data Mining (DMIN)*. The Steering Committee of The World Congress in Computer Science, Computer . . . , page 1.

Jerremy Holland, Jan Erik Solem, and William E Hensler. 2014. Auto-recognition for noteworthy objects. US Patent 8,755,610.

Mike Kestemont. 2014. Function words in authorship attribution. from black magic to theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*. pages 59–66.

Moshe Koppel, Navot Akiva, and Ido Dagan. 2006. Feature instability as a criterion for selecting potential style markers. *Journal of the American Society for Information Science and Technology* 57(11):1519–1525.

Frederick Mosteller and David L Wallace. 1963. Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed federalist papers. *Journal of the American Statistical Association* 58(302):275–309.

Jan Rybicki. 2015. Vive la différence: Tracing the (authorial) gender signal by multivariate analysis of word frequencies. *Digital Scholarship in the Humanities* 31(4):746–761.

Roy Schwartz, Oren Tsur, Ari Rappoport, and Moshe Koppel. 2013. Authorship attribution of micro-messages. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. pages 1880–1891.

Constantina Stamou. 2007. Stylochronometry: Stylistic development, sequence of composition, and relative dating. *Literary and Linguistic Computing* 23(2):181–199.

Dirk Van Hulle and Mike Kestemont. 2016. Periodizing samuel beckett's works: A stylochronometric approach. *Style* 50(2):172–202.

Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. 2006. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American society for information science and technology* 57(3):378–393.