Applying General Impostors Method to the Ferrante Case

Michele A. Cortelazzo, George K. Mikros, and Arjuna Tuzzi

Abstract Elena Ferrante is the *nome de plume* of an anonymous writer who is highly successful on the international stage and whose success far exceeds that of other authors of contemporary Italian literature. In this study, we approach Ferrante's authorship investigation as a verification problem since we cannot be sure whether the real author behind Ferrante's pseudonym is among the candidates we have considered in previous studies. For this reason, we applied the General Impostors (GI) method using the Cosine Delta distance in both a corpus of 150 novels written by 40 authors (39 candidates and Elena Ferrante) and a non-literary corpus of 113 texts signed by 14 different entities (12 authors, a collective author, and Elena Ferrante). In the literary corpus, Starnone emerged as the most likely author of Ferrante's novels. Results were quite different in the second case: Starnone was not the only possible author since, in many non-literary texts, Raja, Martone as well as the E/O publishing house staff and publishers, seem to have authorial contributions. The GI method not only confirmed previous results but also improved our knowledge of this case since it provides a measure of the attribution strength.

Keywords Ferrante, authorship verification, stylometry, General Impostors method

1. Introduction

Elena Ferrante is the pen name of an Italian writer whose novels became a global phenomenon. Today she is perhaps the best-known Italian author on the international stage, and this result seems both deserved and peculiar given that Elena Ferrante is a secretive author. Elena Ferrante's identity has been kept secret for the last 30 years with

Cortelazzo, Michele A., Mikros, George K. and Tuzzi, Arjuna: Applying General Impostors Method to the Ferrante Case, in: Hesselbach, Robert, et al. (Eds.): Digital Stylistics in Romance Studies and Beyond, Heidelberg: Heidelberg University Publishing, 2024, pp. 299–313. DOI: https://doi.org/10.17885/heiup.1157.c19377

the strong support of her publishers, Sandro Ferri and Sandra Ozzola, the owners of the E/O publishing house.

In 2018, Elena Ferrante was the author of seven novels: *L'amore molesto*, *I giorni dell'abbandono*, *La figlia oscura*, *L'amica geniale. Infanzia, adolescenza*, *Storia del nuovo cognome. L'amica geniale volume secondo*, *Storia di chi fugge e di chi resta. L'amica geniale volume terzo* and *Storia della bambina perduta. L'amica geniale volume quarto*. The last four books represent episodes of the best-selling series of novels *L'amica geniale* (*My Brilliant Friend*).

Elena Ferrante has also written a children's story (*La spiaggia di notte*), and she is also the main contributor to a collection of non-literary texts (interviews, essays, and letters) published in a book entitled *La Frantumaglia* (Ferrante 2016). In 2019 a new collection of non-literary texts, *L'invenzione occasionale* (*Incidental Inventions*), appeared: it includes all columns that she published during her year-long collaboration with *The Guardian* in 2018 (Ferrante 2019). Moreover, a new novel was published by the E/O publishing house in November 2019 with the title *La vita bugiarda degli adulti*. This new novel is not the fifth episode of the famous *L'amica geniale* saga, though it is a further story set in Naples. In 2021 Elena Ferrante described in the essay "I margini e il dettato" the pleasure of reading and writing (Ferrante 2021).

Beyond the obvious, intriguing issue of her real identity, Elena Ferrante represents a relevant research object from both the stylistic and the stylometric standpoints. Ferrante's authorship problem is a complex research task as she is an active author not only in literature but also in non-fiction prose. Moreover, since it is a pseudonym, we cannot exclude the existence of processes of collective writing and/or ghostwriters behind the pen name. To better understand her linguistic production and model her writing style, we need to examine not only her literary works but also her articles, essays, and books that are primarily journalistic or autobiographical and represent a completely different genre.

Since this study aims at comparing results with the ones achieved in previous studies, we consider two corpora that have already been exploited:

1) a large literary corpus (Tuzzi and Cortelazzo 2018a, b, c);
2) a corpus of non-literary texts (Cortelazzo, Mikros, and Tuzzi 2018).

2. Corpora Used in This Study

To examine Ferrante's fiction style, we utilized a corpus of contemporary Italian literature that contains 150 novels from 40 different authors,[1] most of them written between 1987 and 2016 and totaling 9,837,851 tokens[2] and 159,149 types. The corpus consists of texts of variable length (Mdn = 50,840.5 words, Mean = 65,586, St.Dev = 39,120, Min = 8,129, Max = 199,839 words) and Ferrante is represented by all her seven novels (635,819 tokens, 33,158 types) which are also variable in size (Mdn = 97,893 words, Mean = 90,831, Min = 36,784, Max = 142,215). The corpus is composed of 50 books from 13 female authors (including Ferrante) and 11 authors from the Campania region (including Ferrante) with 46 books. It contains not only the authors suspected to be behind Ferrante's name but also a wider range of authors that offer a more varied picture of literary production. In that sense, the specific corpus can be used to explore Ferrante's position in the larger framework of contemporary Italian literature and to model author profiles with a more generic coverage.

The selected novels and novelists are all ascribed to one (or more) of these categories:

— Elena Ferrante's novels.
— Novels written by authors from the same area (Naples and its surroundings).
— Novels written by novelists suspected to be Elena Ferrante.
— Blockbusters (best sellers, award-winning novels).
— Novels written by authors who enjoyed the praise of literary criticism.

Furthermore, we compiled a second corpus of Ferrante's non-fiction texts along with a comparable non-fiction corpus with some of the candidate authors behind Ferrante's pseudonym (Cortelazzo, Mikros, and Tuzzi 2018). This non-fiction corpus is composed of 113 texts (143,695 word tokens[3] and 19,020 word types, Mdn = 779 words, Mean = 1,272, St.Dev = 1,406, Min = 228, Max = 8,987). It includes letters, interviews, and additional material written by different authors that can be compared with a selection of texts of *La Frantumaglia* by Elena Ferrante (last Italian version 2016) (Ferrante 2016).

1 The authors contained in this corpus are: Affinati, Ammaniti, Bajani, Balzano, Baricco, Benni, Brizzi, Carofiglio, Covacich, De Luca, De Silva, Faletti, Ferrante, Fois, Giordano, Lagioia, Maraini, Mazzantini, Mazzucco, Milone, Montesano, Morazzoni, Murgia, Nesi, Nori, Parrella, Piccolo, Pincio, Prisco, Raimo, Ramondino, Rea, Scarpa, Sereni, Starnone, Tamaro, Valerio, Vasta, Veronesi, Vinci.
2 Calculations performed with Taltac software (Bolasco 2010) ver. 2.10.
3 Calculations performed with Taltac software (Bolasco 2010) ver. 2.10.

The subcorpus of non-Ferrante texts contains 86 texts (87,458 tokens, 14,308 types, Mdn = 723.5 words, Mean = 1,017, St.Dev = 965, Min = 228, Max = 4,777), and it is composed mainly of articles in newspapers and magazines, essays published in various media, interviews, letters, and texts posted on the Web. 78 of these texts were written by 12 authors (Laura Buffoni, Gianrico Carofiglio, Sandro Ferri, Goffredo Fofi, Marcella Marmo, Mario Martone, Sandra Ozzola, Valeria Parrella, Francesco Piccolo, Anita Raja, Clara Sereni, Domenico Starnone) and eight by a collective subject (E/O) that represents the editorial staff of E/O publishing house (Sandro Ferri and Sandra Ozzola are the owners of E/O).

The subcorpus of Ferrante's non-fiction works includes 27 texts signed by Elena Ferrante and it is distributed across six essays, seven interviews and 14 letters (56,237 tokens, 21,293 types, Mdn = 1,001 words, Mean = 2,083, St. Dev = 2,138, Min = 298 words, Max = 8,987 words).

3. Methodology

3.1 Ferrante and the Need for Applying Authorship Verification Methods

The authorship identification in Elena Ferrante's case is a complex task since it involves several unknown parameters regarding the exact nature of the problem. It can be viewed as an authorship attribution case, i.e., a closed-class classification problem where a standard text classification task can be used. Machine learning algorithms can be trained on a corpus where texts belong to known authors, and the model developed can be further verified using a hold-out set. Then it can predict the authorship in the collection of texts of unknown authorship. When we apply this pipeline to Ferrante's fiction corpus (Mikros 2018), we get as the most probable author behind Elena Ferrante, Domenico Starnone, with an accuracy of over 96 percent. However, this approach is based on a very unstable assumption, namely, that the rest of the 39 authors (including Starnone) who are represented in the fiction corpus are indeed a set of authors which includes, beyond any doubt, the real author of the Ferrante novels. However, we do not have any external evidence that this is the case. A high accuracy attribution further confirmed by different stylometric methods applied to the same data (Tuzzi and Cortelazzo 2018a; Savoy 2018b) is a sign of increased consensus and high reliability. However, we still cannot exclude the possibility that the real author behind Ferrante is someone outside our initial large corpus of contemporary Italian literature.

This suspicion can be further supported by the puzzling author profiling results we obtained when we examined the non-fiction corpus and tried to evaluate Ferrante's gender and age. Author profiling methods were used with considerable success in the fiction corpus defining Ferrante as a male author over 60 and coming from the Naples area (the sole candidate with these characteristics was Domenico Starnone) (Mikros 2018). However, when the same methods were applied to the non-fiction corpus, the results were inconclusive (Cortelazzo, Mikros, and Tuzzi 2018). The gender, age, and region profiling results can be found in Figure 1.

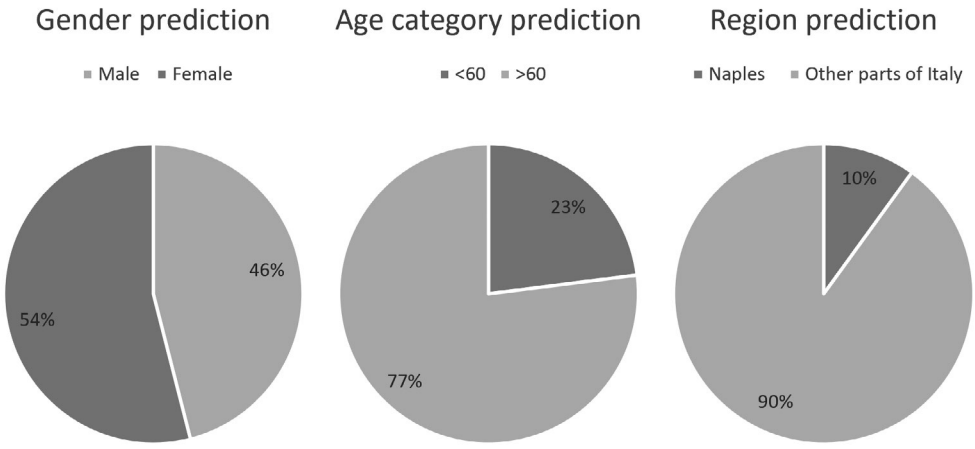

Fig. 1 Pie charts visualizing profiling results (gender, age, region) in Ferrante's non-fiction texts. The reported percentages correspond to the portion of the profiling characteristic predicted in Ferrante's texts. E.g., from the 27 texts signed by Ferrante, the algorithm predicted that 12 (46 percent) were written by a male author. (Cortelazzo, Mikros, Tuzzi, CC BY).

As shown in Figure 1, nearly half of the non-fiction texts are attributed to a female (46 percent) and the other half to a male (54 percent). A somewhat less but still intense variation can be seen in the age profiling, where 77 percent of the texts are attributed to a person over 60 years old and 23 percent to someone less than 60 years old. A more stable attribution appears with the region profiling since 90 percent of the texts are classified as belonging to someone from Naples. A reasonable hypothesis that emerges from these results is that the non-fiction Ferrante texts represent a collective work of more than one author employing authors of both genders and some variation in age.

The above research outcomes reinforce the need for employing explicitly designed methodologies for handling open authorship problems and do not require the existence of real authors inside the training corpus. This specific set of methods is designated by the general term of authorship verification. They can be classified into two broad categories (Potha and Stamatatos 2017):

— Intrinsic methods: perform analysis only on the documents under investigation and handle the verification problem as a one-class classification task. These methods are robust since they do not require external resources and fast since they analyze only a few documents. Examples of these approaches can be found in Jankowska, Milios, and Kešelj (2014), Halvani, Winter, and Pflug (2016), and Mikros and Perifanos (2011).
— Extrinsic methods: these methods analyze an additional set of external documents and transform the verification problem into a binary classification task. They are usually more effective, especially when the set of external documents has been carefully compiled. Characteristic examples of this approach are Koppel and Winter (2014), Seidman (2013), and Kestemont et al. (2016a).

The verification problem is considered the most challenging among the authorship identification tasks and, over the recent years, has attracted considerable research attention, including the organization of two PAN competitions in 2014 (Stamatatos et al. 2014, Stamatatos et al. 2015).

3.2 General Impostors Method

Among the various approaches proposed for solving the verification problem, we selected the GI method, and we used the version implemented in the *stylo* R package (Eder, Rybicki, and Kestemont 2016). The GI method is based on earlier work of Koppel's Many Candidates method (Koppel, Schler, and Argamon 2011), which was further enhanced and tested using different variations by Koppel and Winter (2014) and Seidman (2013) and optimized by Kestemont et al. (2016b). As Kestemont et al. (2016b, 88) explain:

> The general intuition behind the GI is not to assess whether two documents are simply similar in writing style, given a static feature vocabulary, but rather, it aims to assess whether two documents are significantly more similar to one another than other documents, across a variety of stochastically impaired feature spaces (Stamatatos 2006; Eder 2012) and compared to random selections of so-called distractor authors (Juola 2015), also called "imposters."

The procedure is based on a bootstrapped approach in which repeated samples of stylometric features (usually words or n-grams) are used in distance-based comparisons between an anonymous text and a random selection of *impostor* documents the original author did not write. The score calculated represents not only how different the anonymous text is from the other texts of the candidates but also how consistent the stylistic differences are between them.

The process is based on a second-order metric termed O2 in Kestemont et al. (2016b) in the sense that the distance metric is further processed and transformed into a proportion metric which is the final metric used in the algorithm. More specifically, the algorithm finds the distance between the vector of an anonymous document to the centroid vector of a list of the documents of candidate authors. It also finds the distance between the vector of the anonymous text and a list of random non-relevant to the authorship problem texts. Then the GI algorithm starts a bootstrapping procedure in which it samples a random subset of the linguistic features used and a random subset of impostors. In each iteration, it determines whether the vector of the anonymous document is closer to the vector of the candidate author's texts or the vector of the distractors' texts. GI then calculates the proportion of times the vector of the anonymous document was found closer to the vector of the candidate author compared to the vector of the impostors' documents. The proportion is normalized in the 0–1 scale, and since it is based on the distance metric, which should be first calculated, is considered a second-order metric.

The GI method is a versatile technique as the researcher can use a variety of distances, including some well-established in authorship research like Delta (Burrows 2002), Cosine Delta (Evert et al. 2017), Min-Max (Kestemont et al. 2016a), etc. Moreover, it was the winning method in the PAN authorship identification contest in 2013 (Seidman 2013) and 2014 (Khonji and Iraqi 2014), performing considerably better than other authorship verification methods.

4. Results and Discussion

We applied the GI method to Ferrante's authorship problem and used both the fiction and the non-fiction corpus. Although there is a consensus among recent stylometric research (Tuzzi and Cortelazzo 2018a; Cortelazzo and Tuzzi 2020; Savoy 2018b) that the stylometric profile of Domenico Starnone is the closest to Ferrante's among other candidates in the fiction corpus, we used the GI method to evaluate this claim further. Moreover, authorship verification methods have not yet been applied to the Ferrante case, and this research would fill in the existing gap in the stylometric quest to reveal her authorship.

We used the GI method, utilizing the 5,000 most frequent words as stylometric features and applying the Cosine Delta (aka Würzburg) distance as this seems to be experimentally more robust across many languages and varied sizes of most frequent words sets (Evert et al. 2017). We decided not to use any impostors' texts and to focus on the texts of the 39 Italian authors of the fiction corpus, treating each one as a possible candidate for being the Ferrante author. In this way, we did not make any assumptions regarding Ferrante's authorship and let the algorithm examine each of the 39 authors as candidates with equal probabilities of being the real author. Although this procedure is time-consuming, the results obtained are far more valid since the proposed methodology can be considered an alternative cross-validation strategy that checks attribution scores across all possible author pairs with the questioned document.

Since GI scores fall into the normalized range of 0–1, there is also a need to determine the attribution threshold, i.e., the score which marks a positive or a negative attribution to the questioned document. E.g., in our dataset, the "Letter to Ozzola (no. 2)" written by Ferrante was assigned a score of 0.43 when tested against the collective authorship of the E/O publishing house. What sort of evidence is 0.43? Can it be translated to a specific attribution or not? Is it high or low? The only way for us to answer these questions is to thoroughly examine a given corpus to calculate the average proximity between any texts written by the same author and the average proximity between a text by a given author and any text written by someone else. This procedure will define a margin where a classifier is (on average) wrong. *stylo* has adopted a score-shifting algorithm (Kestemont et al., 2016b), based on the c@1 measure of the classifier's performance (Peñas and Rodrigo 2011).

Using the algorithm mentioned above, we calculated the optimal decision thresholds for rejecting or accepting the attribution using the Cosine Delta distance through the relative function offered by the Impostors method implemented in the *stylo* package. Given our corpus, the lower value was calculated as 0.49 and the upper value as 0.51. This means that GI scores under 0.49 can be considered *low* and cannot be used as evidence for attributing a tested text to a specific author. Moreover, GI scores over 0.51 are considered high enough to be translated as positive evidence for attributing the test document to the specific author.

We compared each of Ferrante's books (7 novels) with all the books by 39 authors (143 novels). In this task, Domenico Starnone was indicated as the author of these novels with a probability GI score of 1, which can be explained as a perfect match across all features' subsets. This result further confirms all the previously reported stylometric research stating that Domenico Starnone's stylometric profile is the closest to Elena Ferrante's writings.

After establishing the validity of the GI method in the fiction corpus, we tested Ferrante's authorship in the non-fiction corpus. Since the non-fiction corpus is smaller than the fiction one, we used the 2,000 most frequent words as stylometric features

and applied the Cosine Delta distance. Moreover, we calculated the optimal decision thresholds for the decline and the acceptance of the attribution using the cosine distance. Using the relative optimization algorithm employed in the *stylo* package, the lower value was calculated to be 0.4 and the upper value to be 0.52, i.e., any GI score over 0.52 produced in a comparison of a known authorship text and the anonymous text could indicate that the author of the known authorship text is also the author of the unknown text.

We compared each of Ferrante's texts (27 letters, interviews, essays) with the rest of the non-fiction texts written by 12 authors and one collective writer (staff of the E/O publishing house). The non-literary Ferrante texts (included in *La Frantumaglia*) seem like they may have been written by multiple authors. Among them are Starnone, Raja, Martone, Parella, Ozzola, and the rest of the staff from the E/O publishing house. Other candidates (Buffoni, Carofiglio, Ferri, Fofi, Marmo, Piccolo) seem entirely irrelevant to the writing of these documents, and seven texts out of 27 do not have an exact author match.

These results confirm our previous study in the same corpus using profiling methods, and closed-class supervised machine learning algorithms (Cortelazzo, Mikros, and Tuzzi 2018). To reduce the algorithm's search space and explore in detail the most probable candidates, we repeated the whole procedure maintaining E/O, Martone, Parrella, Raja, and Starnone in the candidates' pool and recalculating the GI decision boundaries and the GI scores for each of Ferrante's non-fiction texts. We used the same feature set (2,000 most frequent words), and the GI threshold values were calculated as 0.38 (rejection threshold) and 0.45 (attribution threshold). Table 1 reports the GI scores for each text and each candidate.

The GI scores calculated confirm our initial attribution. Using the GI verification method, all of Ferrante's non-fiction texts have been assigned to one or more than one author (six out of 27 have been assigned to two authors and one of them to three). Both Starnone and Raja seem to have written a number of these texts. The extended staff of the E/O publishing house now has ten attributions, confirming our suspicion that part of *La Frantumaglia* is a collective production of the staff of the E/O publishing house. The distributed authorship hypothesis can be visualized in Figure 2, which displays the GI scores.

In Figure 2, we used a stacked bar chart to standardize the magnitude of GI scores which were over 0.45 and are considered above the threshold of positive authorship attribution. E.g., if two authors had GI scores above the attribution threshold in one text, these two scores were normalized in relative percentages. For example., in *lettero2,* there was a GI score of 0.52, attributing this text to the E/O publishing house team, and a GI score of 0.65 that attributes the text to Martone. The stacked bar converted these scores to 44 percent and 56 percent correspondingly so that the bar adds up to 100 percent, and we can compare the relative magnitude of each GI score across the

Table 1 GI scores calculated for the Ferrante non-fiction texts and the attributed candidate authors. The attribution scores can be seen in bold under the candidate's column.

Testing documents	E/O	Martone	Parrella	Raja	Starnone
essay01	0.06	0.42	0.01	**0.77**	0.18
essay02	0.06	**0.93**	0.12	0.37	0.26
essay03	0.06	**0.95**	0.11	0.52	0.08
essay04	0.2	0.25	0.01	**0.86**	0.05
essay05	0.01	**0.79**	0.13	**0.49**	0.23
essay06	0.29	0.05	**0.63**	0	**0.61**
interview01	**0.92**	0.04	0.02	0.01	0.14
interview02	**0.86**	0	0.02	0.3	0.09
interview03	0.43	0.39	0	**0.68**	0
interview04	0.26	0.06	**0.93**	0	0.29
interview05	0.12	0.38	0	**0.87**	0.11
interview06	0.34	0.07	0.02	**0.85**	0.07
interview07	**0.46**	0.07	0.19	**0.47**	**0.53**
letter01	0.18	0.05	0.07	0	**0.89**
letter02	**0.52**	**0.65**	0.22	0	0.11
letter03	**0.86**	0.17	0.01	0	**0.52**
letter04	0.43	0.19	0.18	0	**0.56**
letter05	**0.65**	0.15	0.19	0	0.33
letter06	**0.96**	0.16	0.07	0	0.11
letter07	**1**	0.13	0	0	0.1
letter08	0.2	0.12	0.22	**0.45**	0.24
letter09	0.15	0.08	**0.86**	0	**0.54**
letter10	**0.66**	0.28	0.32	0	0.34
letter11	**0.57**	0.17	0.23	0	0.43
letter12	0.32	0.3	0.24	0	**0.59**
letter13	0.4	0.27	0.06	0.13	**0.47**
letter14	0.26	0.28	0.06	0	**0.67**

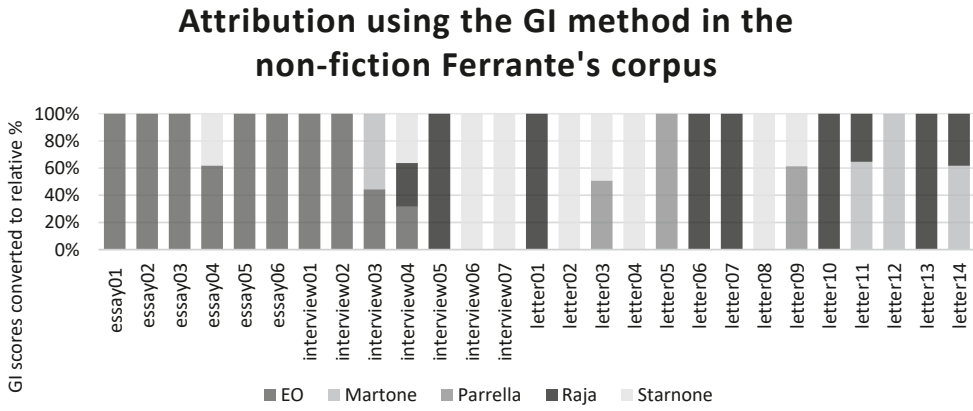

Fig. 2 Stacked bar chart of the GI scores obtained. The attributions of the Ferrante texts can be seen using the legend of the graph. (Cortelazzo, Mikros, Tuzzi 2023, CC BY).

various Ferrante's texts on a uniform scale way since all scores have now been transformed to the scale 0–100 percent.

Given the previous discussion and the results obtained by applying the GI method in Ferrante's non-fiction texts,[4] we can safely infer that Ferrante's non-fiction texts do not represent a homogeneous stylometric profile and could be attributed to various people working for the public relations of the Ferrante brand name.

5. Conclusion

Elena Ferrante's authorship remains a very interesting stylometric problem and one of the most complex cases of cross-genre attribution, as she is an active author in fiction and non-fiction texts. In this study, we tried to complement previous stylometric research and use an authorship verification technique called the GI method. The rationale behind this approach is that we need to approach Ferrante's authorship case as an open-class problem where the research question does not imply a set of predetermined candidates but leaves space for possibilities other than the ones we might have in mind.

4 The script used for applying the GI method to the Ferrante's texts is available on GitHub: https://github.com/gmikros/GI-Method-in-Ferrante-texts.

In both corpora (fiction and non-fiction), the GI method confirmed previous research results, but it also improved our knowledge since it provided a measure of the attribution strength. Domenico Starnone's stylometric profile was identified as the single undisputed match compared to the stylometric profile of Ferrante's novels. These results confirm and enrich results obtained by previous studies collected in Tuzzi and Cortelazzo (2018a) (cf. Eder 2018; Juola 2018; Lalli, Tria, and Loreto 2018; Mikros 2018; Ratinaud 2018; Rybicki 2018; Savoy 2018a). However, this clear-cut picture did not emerge when we examined the authorship of the non-fiction corpus. The GI method attributed some of the texts collected in *La Frantumaglia* to the staff or the owners of the E/O publishing house. For some other texts, Domenico Starnone, Anita Raja, Mario Martone, and Valeria Parrella were identified as possible authors. Moreover, in specific Ferrante non-fiction texts, we can detect patterns of co-writing as we observe attribution GI scores to more than one author.

Authorship verification methods are less accurate than the supervised classification pipelines, but, in our case, they can be used to complement the published research on this topic. They can shed light on research questions that a closed-class classification algorithm cannot answer. Both the mixed authorship signal detected in some of Ferrante's non-fiction texts and the distributed authorship hypothesis as part of an organized public communication project supporting Ferrante's name can be reliably investigated under the authorship verification framework. Enlarging available corpora with new works signed by Elena Ferrante could be the starting point for further investigations and new research questions.

ORCID®

Georgios Mikros https://orcid.org/0000-0002-4093-5973
Arjuna Tuzzi https://orcid.org/0000-0003-3795-5567

References

Bolasco, Sergio. 2010. *TaLTaC2.10 Sviluppi, esperienze ed elementi essenziali di analisi automatic dei testi*. Milano: LED.

Burrows, John F. 2002. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17 (3): 267–87.

Cortelazzo, Michele A., George Mikros, K., and Arjuna Tuzzi. 2018. "Profiling Elena Ferrante: A Look Beyond Novels." In *JADT 2018: Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*, edited by Domenica Fioredistella Iezzi, Livia Celardo, and Michelangelo Misuraca, 165–73. Rome: UniversItalia.

Cortelazzo, Michele A., and Arjuna Tuzzi. 2020. "A chi assomiglia Elena Ferrante? Un profilo stilometrico aggiornato." *Italica Wratislaviensia* 11 (1): 123–41.

Eder, Maciej. 2012. "Computational Stylistics and Biblical Translation: How Reliable Can a Dendrogram Be?" In *The Translator and the Computer*, edited by Tadeusz Piotrowski and Łukasz Grabowski, 155–70. Wrocław: WSF Press.

Eder, Maciej. 2018. "Elena Ferrante: A Virtual Author." In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi and Michele A. Cortelazzo, 31–45. Padova: Padova University Press.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8 (1): 107–21.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 2017. "Understanding and Explaining Delta Measures for Authorship Attribution." *Digital Scholarship in the Humanities* 32 (2): ii4–ii16. https://doi.org/10.1093/llc/fqx023.

Ferrante, Elena. 2016. *La Frantumaglia*. Rome: Edizioni E/O.

Ferrante, Elena. 2019. *L'invenzione occasionale*. Rome: Edizioni E/O.

Ferrante, Elena. 2021. *I margini e il dettato*. Rome: Edizioni E/O.

Halvani, Oren, Christian Winter, and Anika Pflug. 2016. "Authorship Verification for Different Languages, Genres and Topics." *Digital Investigation* 16: S33–S43. https://doi.org/10.1016/j.diin.2016.01.006.

Jankowska, Magdalena, Evangelos Milios, and Vlado Kešelj. 2014. "Author Verification Using Common N-Gram Profiles of Text Documents." In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 387–97. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Juola, Patrick. 2015. "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions." *Digital Scholarship in the Humanities* 30 (1): i100–i113. https://doi.org/10.1093/llc/fqv040.

Juola, Patrick. 2018. "Thesaurus-Based Semantic Similarity Judgements: A New Approach to Authorial Similarity?" In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi, and Michele A. Cortelazzo, 48–59. Padova: Padova University Press.

Kestemont, Mike, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans. 2016a. "Authorship Verification with the Ruzicka Metric." In *Digital Humanities 2016: Conference Abstracts*, 246–49. Krakow: Jagiellonian University & Pedagogical University.

Kestemont, Mike, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans. 2016b. "Authenticating the Writings of Julius Caesar." *Expert Systems with Applications* 63: 86–96. https://doi.org/10.1016/j.eswa.2016.06.029.

Khonji, Mahmoud, and Youssef Iraqi. 2014. "A Slightly-Modified GI-Based Author-Verifier with Lots of Features (ASGALF) – Notebook for PAN at CLEF 2014." In *CLEF 2014 Evaluation Labs and Workshop – Working Notes Papers, 15–18 September, Sheffield, UK, September 2014*, edited by Linda Cappellato, Nicola Ferro, Martin Halvey, and Wessel Kraaij, 977–83. CEUR Workshop Proceedings (CEUR-WS.org).

Koppel, Moshe, Jonathan Schler, and Shlomo Argamon. 2011. "Authorship Attribution in the Wild." *Language Resources and Evaluation* 45 (1): 83–94. https://doi.org/10.1007/s10579-009-9111-2.

Koppel, Moshe, and Yaron Winter. 2014. "Determining If Two Documents Are Written by the Same Author." *Journal of the Association for Information Science and Technology* 65 (1): 178–87. https://doi.org/10.1002/asi.22954.

Lalli, Margherita, Francesca Tria, and Vittorio Loreto. 2018. "Data-Compression Approach to Authorship Attribution." In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi, and Michele A. Cortelazzo, 61–83. Padova: Padova University Press.

Mikros, George, K. 2018. "Blended Authorship Attribution: Unmasking Elena Ferrante Combining Different Author Profiling Methods." In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi, and Michele A. Cortelazzo, 85–95. Padova: Padova University Press.

Mikros, George K., and Kostas Perifanos. 2011. "Authorship Identification in Large Email Collections: Experiments Using Features That Belong to Different Linguistic Levels." *Proceedings of PAN 2011 Lab, Uncovering Plagiarism, Authorship, and Social Software Misuse held in conjunction with the CLEF 2011 Conference on Multilingual and Multimodal Information Access Evaluation, 19–22 September 2011, Amsterdam*.

Peñas, Anselmo, and Alvaro Rodrigo. 2011. "A Simple Measure to Assess Non-response." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 1415–24. Portland, Oregon: ACL.

Potha, Nektaria, and Efstathios Stamatatos. 2017. "An Improved Impostors Method for Authorship Verification." In *Experimental IR Meets Multilinguality, Multimodality, and Interaction. 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017*, edited by Gareth J.F. Jones, Séamus Lawless, Julio Gonzalo, Liadh Kelly, Lorraine Goeuriot, Thomas Mandl, Linda Cappellato, and Nicola Ferro, 138–44. Cham: Springer.

Ratinaud, Pierre. 2018. "The Brilliant Friend(s) of Elena Ferrante: A Lexicometrical Comparison between Elena Ferrante's Books and 39 Contemporary Italian Writers." In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi and Michele A. Cortelazzo, 97–110. Padova: Padova University Press.

Rybicki, Jan. 2018. "Partners in Life, Partners in Crime?" In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi, and Michele A. Cortelazzo, 111–22. Padova: Padova University Press.

Savoy, Jacques. 2018a. "Elena Ferrante Unmasked." In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi, and Michele A. Cortelazzo, 123–41. Padova: Padova University Press.

Savoy, Jacques. 2018b. "Is Starnone Really the Author behind Ferrante?" *Digital Scholarship in the Humanities* 33 (4): 902–18. https://doi.org/10.1093/llc/fqy016.

Seidman, Shachar. 2013. "Authorship Verification Using the Impostors Method." In *Working Notes for CLEF 2013 Conference, Valencia, Spain, September 23–26, 2013*, edited by Pamela Forner, Roberto Navigli, Dan Tufis, and Nicola Ferro. Valencia: CEUR.

Stamatatos, Efstathios. 2006. "Authorship Attribution Based on Feature Set Subspacing Ensembles." *International Journal on Artificial Intelligence Tools* 15 (5): 823–38. https://doi.org/10.1142/S0218213006002965.

Stamatatos, Efstathios, Walter Daelemans, Ben Verhoeven, Benno Stein, Martin Potthast, Patrick Juola, Miguel A. Sánchez-Pérez, and Alberto Barrón-Cedeño. 2014. "Overview of the Author Identification Task at PAN 2014." In *Working Notes for CLEF 2014 Conference*, edited by Linda Cappellato, Nicola Ferro, Martin Halvey, Wessel Kraaij, Nicola Ferro, Martin Halvey, and Wessel Kraaij, 877–97. CEUR Workshop Proceedings (CEUR-WS.org).

Stamatatos, Efstathios, Walter Daelemans, Ben Verhoeven, Patrick Juola, Aurelio López-López, Martin Potthast, and Benno Stein. 2015. "Overview of the Author Identification Task at PAN 2015." In *Working Notes of CLEF 2015 – Conference and Labs of the Evaluation Forum*, edited by Linda Cappellato, Nicola Ferro, Gareth J. F. Jones, and Eric San Juan. CEUR Workshop Proceedings (CEUR-WS.org).

Tuzzi, Arjuna, and Michele A. Cortelazzo, eds. 2018a. *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*. Padova: Padova University Press.

Tuzzi, Arjuna, and Michele A. Cortelazzo. 2018b. "It Takes Many Hands to Draw Elena Ferrante's Profile." In *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, edited by Arjuna Tuzzi and Michele A. Cortelazzo, 9–29. Padova: Padova University Press.

Tuzzi, Arjuna, and Michele A. Cortelazzo. 2018c. "What Is Elena Ferrante? A Comparative Analysis of a Secretive Bestselling Italian Writer." *Digital Scholarship in the Humanities* 33 (3): 685–702. https://doi.org/10.1093/llc/fqx066.