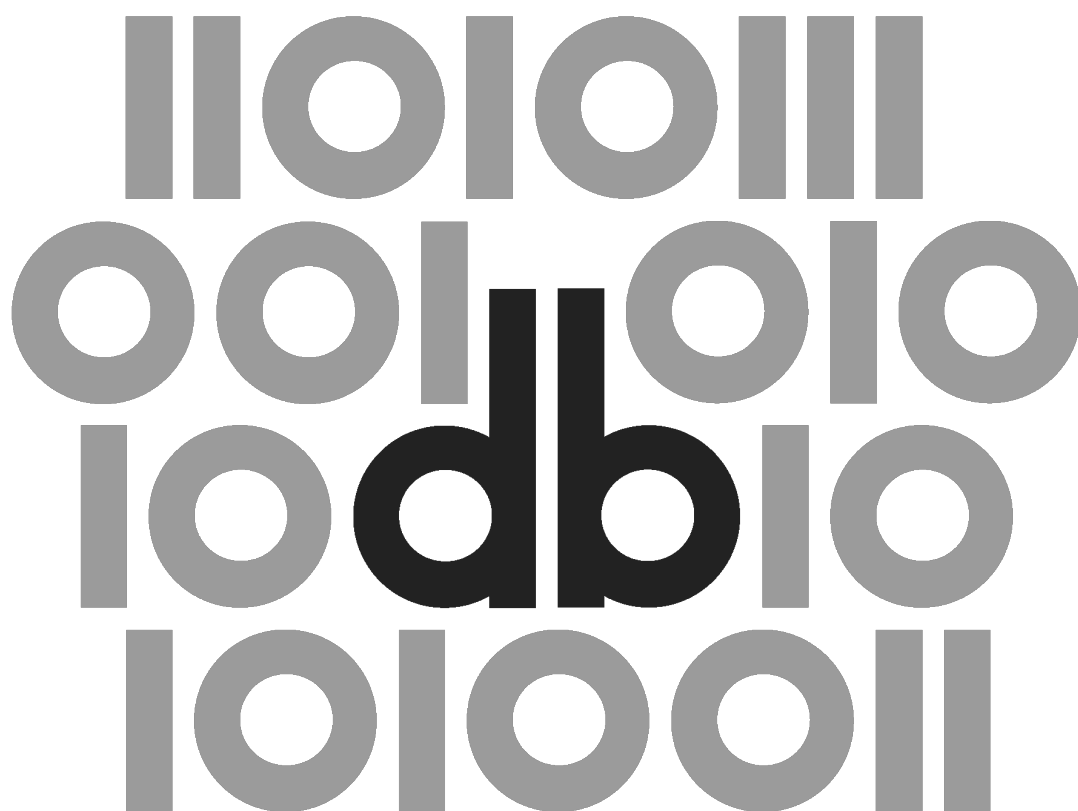


5 (2021) <DIGITÁLIS BÖLCSÉSZET>  
A krakkói Computational Stylistics Group  
(Különszám)

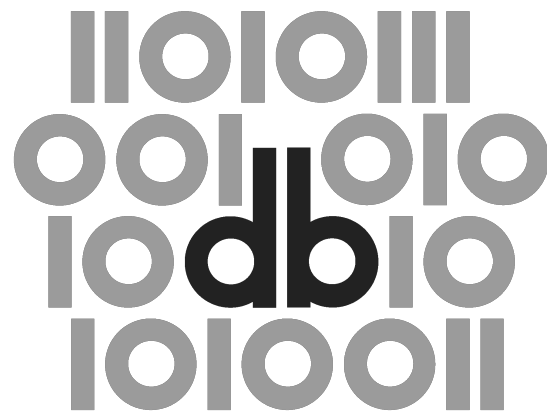


5 (2021) </DIGITÁLIS BÖLCSÉSZET>

**Digitális Bölcsészet**  
**2021., ötödik szám**

**A krakkói Computational Stylistics Group**  
**(Különszám)**

<DIGITÁLIS BÖLCSÉSZET>



5 (2021)

A krakkói  
Computational Stylistics Group

(Különszám)

A különszámot Szemes Botond szerkesztette.

**Felelős szerkesztő:**

Maróthy Szilvia

**Szerkesztőség:**

Kokas Károly, Parádi Andrea

**Rovatvezetők:**

*Tanulmányok:* Kiss Margit

*Műhely:* Péter Róbert

*Kritika:* Almási Zsolt

*Labor:* Mártonfi Attila

**Tanácsadó testület:**

Bartók István, Fazekas István, Golden Dániel, Horváth Iván, Palkó Gábor, Pap Balázs, Sass Bálint, Seláf Levente

**Korábbi munkatársaink:**

Bartók Zsófia Ágnes (szerkesztő, rovatvezető), Fodor János (szerkesztő),

†Labádi Gergely (szerkesztő, rovatvezető), †Orlovsky Géza (tanácsadó testület)

**ISSN 2630-9696**

**DOI 10.31400/dh-hun.2021.5**

Kiadja a Bakonyi Géza Alapítvány és az ELTE BTK Régi Magyar Irodalom Tanszéke (1088 Budapest, Múzeum krt. 4/A).

Felelős kiadó az ELTE BTK Régi Magyar Irodalom Tanszék vezetője.

Megjelenik az Open Journal Systems (OJS) v. 3. platformon, melynek működtetését az ELTE Egyetemi Könyvtár- és Levéltár biztosítja.

A különszám megjelenését a Wacław Felczak Alapítvány támogatta.



WACŁAW  
FELCZAK  
ALAPÍTVÁNY

FUNDACJA  
IM. WACŁAWA  
FELCZAKA



Ez a mű a Creative Commons *Nevezd meg! – Ne add el! – Így add tovább! 2.5 Magyarország Licenc* (<http://creativecommons.org/licenses/by-nc-sa/2.5/hu/>) feltételeinek megfelelően felhasználható.

Honlap: <http://ojs.elte.hu/digitalisbolcseszett>

Email cím: [dbfolyoirat@gmail.com](mailto:dbfolyoirat@gmail.com)

Olvasószerkesztő: Bucsics Katalin

Tördelés: Hegedüs Béla

Grafika: Hegyi Gábor

# Tartalom

ELŐSZÓ	1
Joanna Byszuk – Szemes Botond <i>A krakkói Computational Stylistics Group bemutatkozása</i> <i>Előszó a Digitális Bölcsészet folyóirat tematikus lapszámához . . . . .</i>	3
TANULMÁNYOK	1
Maciej Eder <i>Elena Ferrante: Egy „virtuális” szerző . . . . .</i>	3
Jan Rybicki <i>Vive la différence!</i> <i>Írók nemének azonosítása többváltozós szógyakorisági elemzések során . . . . .</i>	19
Greta Franzini – Mike Kestemont – Gabriela Rotari – Melina Jander – Jeremi K. Ochab – Emily Franzini – Joanna Byszuk – Jan Rybicki <i>Szerzőazonosítás Jacob és Wilhelm Grimm zajos, digitalizált</i> <i>levelezésében . . . . .</i>	39
Artjoms Šeļa – Boris Orekhov – Roman Leibov <i>Gyenge műfajok</i> <i>A költői versmérték és a jelentés közötti kapcsolat modellálása</i> <i>az orosz költészetben . . . . .</i>	69
Albert Leśniak– Zbigniew Pasek <i>Neoprotestáns és katolikus tanúságtételek a korpuszalapú</i> <i>diskurzuselemzés perspektívájából . . . . .</i>	91
Helena Grochola-Szczepanek – Ruprecht Von Waldenfels – Rafał L. Górski – Michał Woźniak <i>A szepességi lengyel nyelvjárás korpusznyelvészeti elemzése . . . . .</i>	113

**Maciej Eder**  0000-0002-1429-5036

Institutu Języka Polskiego PAN

Uniwersytetu Pedagogicznego w Krakowie, Instytucie Filologii Polskiej

maciej.eder@ijp.pan.pl

## Elena Ferrante: Egy „virtuális” szerző\*

A jelen tanulmány az Elena Ferrante írói álnév alatt publikált regényeket vizsgálja azért, hogy kísérletet tegyen az álnév mögött rejtőző szerzőség azonosítására. A tanulmány ugyanakkor többre vállalkozik annál, minthogy egyszerűen újranyissa a szerzőazonosítás kérdését; a kutatás valódi jelentősége az, hogy magának a szerzői ujjlenyomatnak az állandóságát teszteli Ferrante műveinek vonatkozásában, rejtőzzen akárki is az írói álnév mögött. A kutatási kérdés megválaszolásához egy 150 regényből álló korpusz állt rendelkezésre, a stilisztikai hasonlóság méréséhez az úgynevezett Bootstrap Consensus Network eljárást használta a kutatás. Azon szerzők listája, akiknek a szerzői ujjlenyomata a leginkább egyezőnek bizonyult „Ferrante” lenyomataival – köztük Domenico Starnonéval a lista élén – az úgynevezett Rolling Classify (‘mozgó osztályozás’) metódus alkalmazásával vált hozzáférhetővé. Ez a módszer – amit Ferrante és Starnone regényein egymástól függetlenül alkalmaztunk – az irodalmi szövegek lokális stilisztikai sajátosságainak feltárására alkalmas, ami ezáltal részletesebb megfigyeléseket is megfogalmazhatóvá tesz. Az összkép, néhány kivételtől eltekintve, megerősíti, hogy Starnone és Ferrante stílusa megkülönböztethető egymástól, ami viszont erős érvnek mutatkozik a „virtuális szerzőség” hipotézise mellett. Úgy tűnik, hogy Domenico Starnone – különösen kései műveiben – képes megkülönböztetni egymástól saját és alteregója hangját.

Kulcsszavak:

stilometria, szerzőazonosítás, virtuális szerzőség, Bootstrap Consensus Network, Rolling Classify



### Bevezetés

Elena Ferrante szerzői nevét csaknem három évtizede tartják számon a nemzetközi irodalmi színtéren. A hét regény és a *The Guardian* heti rovatában rendszeresen publikált művek szerzőjének valódi kiléte azonban sajtóságos módon anonimitásba burkolózik. Az olvasóközönség ugyanis tisztában van azzal, hogy az „Elena Ferrante” egy, a szerző(k) által választott álnév, amely elfedi a valós szerző(k) kilétét. Azaz

\* Eredeti megjelenés: Maciej Eder, „Elena Ferrante: A Virtual Author,” in Arjuna Tuzzi and Michele A. Cortelazzo, eds., *Drawing Elena Ferrante’s Profile*, 31–47 (Padova: Padova University Press, 2018).

talán csak a közelmúltig volt ez így: egyre több cikkben olvasható ugyanis a feltevés, miszerint az írói álnév mögött Domenico Starnone szerzősége rejtőzik.<sup>1</sup> Az azóta megjelent tudományos munkákban, amelyek a legkorszerűbb statisztikai elemzéseket alkalmazzák vizsgálatuk során, szintén beigazolódott ez a feltevés.<sup>2</sup> Starnone 1943-ban született író és újságíró, aki számos, kritikusok által elismert regény és kisebb prózai művek alkotója. Érdekes tény, hogy a másik szerző, akinek a közvélekedés szintén az írói álnév alatt publikált regényeket tulajdonítja, nem más, mint Anita Raja – Domenico Starnone felesége, akit a német irodalom egyik fordítójaként ismerhetünk.<sup>3</sup>

A jelen tanulmány nem a fentiekben részletezett szerzői kérdést kívánja újra megnyitni – bár a Starnone-hipotézis most is a vizsgálat egyik előterében álló jelenség –, ehelyett a vizsgálat a szerzői „ujjlenyomat” stabilitásának tesztelésére tesz kísérletet, legyen akárki is Ferrante műveinek tényleges szerzője. Azontúl, hogy a vizsgálat törekszik azonosítani az anonimitásba burkolózó szerző nevét, mérhetővé teszi azt is, hogy az azonosított szerző neve alatt megjelent művek, valamint a Ferrante név alatt publikált prózai munkák milyen mértékű egyezést és eltérést mutatnak egymáshoz képest a stilisztika szempontjából. Tehát a kutatási kérdés a következő: vajon két elhatárolható, de egymással mégis összefüggő „szerzői” profillal állunk-e szemben, amelyek egyfelől az íróhoz, másfelől az író által konstruált szerzői identitáshoz tartoznak? A szerzőazonosításra alkalmazott módszerek az előbb vázolt kérdések kontextusában az alábbi hipotézist teszik lehetővé: ha a gépi tanulással működő klasszifikációs eljárás képes elválasztani egymástól a két különböző, de egymáshoz erősen kapcsolódó szerzői ujjlenyomatot – tehát a „Ferrante”-hez és a tényleges olasz íróhoz tartozót – az visszaigazolja, hogy egyetlen személy képes lehet egyszerre több stilisztikai perszónát kialakítani.

## A potenciális írók listájának szűkítése

A szerzőazonosításhoz elvégzendő vizsgálatok első lépése az, hogy olyan szerzők műveit gyűjtsük össze, akik az ismeretlen szöveg potenciális írói lehetnek. „Ferrante” esetében több, mint kockázatos azt feltételeznünk, hogy a lehetséges szerzők pontosan listázhatók, még akkor is, ha az előzetes vizsgálatok alapján úgy tűnik, Domenico

<sup>1</sup> Luigi Galella, „Ferrante-Starnone: Un amore molesto in via Gemito,” *La Stampa*, 2005. jan. 16., 27; Simone Gatto, „Una biografia, due autofiction. Ferrante-Starnone: cancellare le tracce,” *Lo Specchio di carta: Osservatorio sul romanzo italiano contemporaneo*, 2016. okt. 22, <https://losp ecchiodicarta.it/2016/10/22/una-biografia-due-autofiction-ferrante-starnone-cancellare-le-tracce/>.

<sup>2</sup> Michele A. Cortelazzo e Arjuna Tuzzi, „Sulle tracce di Elena Ferrante: Questioni di metodo e primi risultati,” in Giuseppe Palumbo, ed., *Testi, corpora, confronti interlinguistici: Approcci qualitativi e quantitative*, 11–25 (Trieste: Edizioni Università di Trieste, 2017), <http://doi.org/10.13137/978-88-8303-913-3/18478>; Michele A. Cortelazzo, George K. Mikros and Arjuna Tuzzi, „Profiling Elena Ferrante: A Look Beyond Novels,” in Domenica Fioredistella Iezzi, Livia Celardo and Michelangelo Misuraca, eds., *Jadt '18. Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*, 165–173 (Roma: UniversItalia, 2018); Anjura Tuzzi and Michele A. Cortelazzo, „What is Elena Ferrante? A Comparative Analysis of a Secretive Bestselling Italian Writer,” *Digital Scholarship in the Humanities* 33, 3. sz. (2018): 685–702, <https://doi.org/10.1093/lhc/fqx066>.

<sup>3</sup> Claudio Gatti, „Elena Ferrante: An answer?” *The New York Review of Books*, 2016. okt. 2., <http://www.nybooks.com/daily/2016/10/02/elena-ferrante-an-answer/>.

Starnone lapul az írói álnév mögött. Ilyen esetekben érdemes a jelöltek nyílt halmazát, tágabb körét vizsgálnunk, amely során megbízhatóan magas számú művet viszünk be a korpuszba – jelen esetben kortárs olasz szerzőktől –, azért, hogy a stilometriai vizsgálatok minél szélesebb spektrumon legyenek elvégezhetőek, és ezáltal kiszűrhetővé váljanak azok az alkotók, akik szerzői ujjlenyomata potenciálisan megfeleltethető a kérdéses szerzőségű szöveg írójának lenyomatával. A jelen tanulmányban bemutatott vizsgálat tesztkorpusza összesen 150 olasz 20. századi regényt tartalmaz, 40 különböző szerzőtől, köztük hét olyat, amely a Ferrante álnév alatt került publikálásra. A tesztkorpusz megegyezik azzal a gyűjteménnyel, amelyet Michele Cortelazzo és Arjuna Tuzzi használt egy korábbi, Ferrante írói kilétének kérdéséhez kapcsolódó vizsgálatban.<sup>4</sup>

Mivel a cél a potenciális jelöltek listájának szűkítése, nem pedig egyetlen szerző hozzárendelése a kérdéses szerzőségű szövegekhez – ez olyan eset, ahol a nagy merítés, azaz a felidézés (*recall*) fontosabb, mint a precízió (*precision*) –, ezért egy meglehetősen egyszerű eljárást érdemes alkalmaznunk. Annak ellenére, hogy a kifinomultabb gépi tanuláson alapuló módszerek (pl. Support Vectors Machines) rendkívül hatékonyak bizonyulnak a hasonló többdimenziós problémák feltárásában,<sup>5</sup> a szerzőazonosításban a viszonylag egyszerű távolságalapú eljárások is kellően jól teljesítenek,<sup>6</sup> különösen akkor, amikor magas a vizsgált szerzők száma.<sup>7</sup> A Ferrante-regények lehetséges szerzőinek kiválasztásához a Bootstrap Consensus Network (BCN) módszert<sup>8</sup> használta a kutatás. Ez a metódus a klaszteranalízisnek olyan továbbfejlesztett változata, melynek célja egy adott korpusz többszörös kiértékelése különböző jellemzők mentén (jelen esetben a leggyakoribb szavak értékei 100–1000 szavas tartományban). Az iterációkban a stilometriában használatos távolságalapú metrikák mérik a szövegek közelségét; a cél ekkor az egymáshoz stilárisan kapcsolódó szövegek azonosítása.

A stiláris kapcsolat e koncepciója lényegében abból indul ki, hogy a korpusz minden szövege jellemezhető aszerint, hogy többé-kevésbé hasonlít (stilometriailag) az összes többi szöveghez, ezáltal mindegyiknek létezik egy szomszédja, amely a legmagasabb egyezést mutatja vele. E feltételezés geometriai interpretációja azt jelenti, hogy a szövegek egy többdimenziós térben (a dimenziók számát a vizsgált jellemzők száma adja meg) pontokként reprezentálódnak; a pontok közötti térbeli távolság pedig azt

<sup>4</sup> Cortelazzo e Tuzzi, „Sulle tracce di Elena Ferrante”; Cortelazzo, Mikros and Tuzzi, „Profiling Elena Ferrante”; Tuzzi and Cortelazzo, „What is Elena Ferrante?”; Michele A. Cortelazzo, Paolo Nadalutti, Stefano Ondelli and Arjuna Tuzzi, „Authorship Attribution and Text Clustering in Contemporary Italian Novels: Does Elena Ferrante’s and Domenico Starnone’s Regional Origin Play a Role?” in Lu Wang, Reinhard Köhler and Arjuna Tuzzi, eds., *Structures, Properties, and Interrelations: Selected Papers from Qualico 2016*, 1–14 (Lüdenscheidt: RAM Verlag, 2018).

<sup>5</sup> Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R* (New York: Springer, 2013), <https://doi.org/10.1080/24754269.2021.1980261>.

<sup>6</sup> Matthew Jockers and Daniela Witten, „A Comparative Study of Machine Learning Methods for Authorship Attribution,” *Literary and Linguistic Computing* 25, 2. sz. (2010): 215–223, <https://doi.org/10.1093/llc/fqq001>.

<sup>7</sup> Koen Luyckx and Walter Daelemans, „The Effect of Author Set Size and Data Size in Authorship Attribution,” *Literary and Linguistic Computing* 26, 1. sz. (2011): 35–55, <https://doi.org/10.1093/llc/fqq013>.

<sup>8</sup> Maciej Eder, „Visualization in Stylometry: Cluster Analysis Using Networks,” *Digital Scholarship in the Humanities* 31, 1. sz. (2017): 50–64, <https://doi.org/10.1093/llc/fqv061>.



szimbolizálja, hogy a szövegek egymáshoz képest milyen stilometriai hasonlóságot mutatnak. Ezért a szövegek közelsége meghatározható geometrikusan, távolságalapú metrikák alapján. A jelen tanulmányban a Manhattan-távolságot alkalmaztuk a standardizált szógyakoróságokra (*z-score*), amelyeket a dimenziók számával (vagyis az összes vizsgált szó számával) osztottunk el. Ez a mérési procedúra Burrows Deltaként vált ismertté a stilometriában.<sup>9</sup>

A BCN-eljárás minden egyes iterációja egy listát eredményez, amelyek a megadott számú jellemzők alapján a szövegekre vonatkozó szomszédsági viszonyokat tartalmazzák. Az első ilyen „pillanatkép” a 100 leggyakoribb, míg a második iteráció a 200 leggyakoribb szóra irányult, majd a 100-as lépésköz további alkalmazásával eljutottunk a 300-as, 400-as, és egészen az 1000 leggyakoribb szót figyelembe vevő beállításig. A következő lépésként ezeknek a pillanatképeknek az eredményeit szinkronizáljuk egymással, és egy hálózatalapú leképezést hozunk létre belőlük. A hálózat élei a szövegek (csomópontok) közötti legközelebbi szomszédos kapcsolatokat jelölik. Pontosabban, minden egyes csomóponthoz három él rendelődik – a hasonlóság alapján a szöveg három legközelebbi szomszédja. Ebből kifolyólag a BCN-eljárás több szempontból megegyezik a *k*-NN klasszifikációval,<sup>10</sup> amely egy szöveg *k* legközelebbi szomszédja alapján hozza meg klasszifikációs döntését.

A fenti módon generált hálózat meglehetősen informatív számunkra, ugyanakkor az így kirajzolódó ábra szabad szemmel nehezen olvasható, így érdemes a hálózatot erőirányított gráfként (*force-directed graph*) is vizualizálni. Ebben az esetben az eljárás részét képezi a csoportokba rendezett csomópontok (azaz a szövegek), illetve a még nagyobb csoportokba rendezett csoportok kézi ellenőrzése is. A kialakuló klaszterek emberi értelmezése, több más magyarázó módszerhez hasonlóan, meglehetősen egyszerűvé teszi ennek használatát.

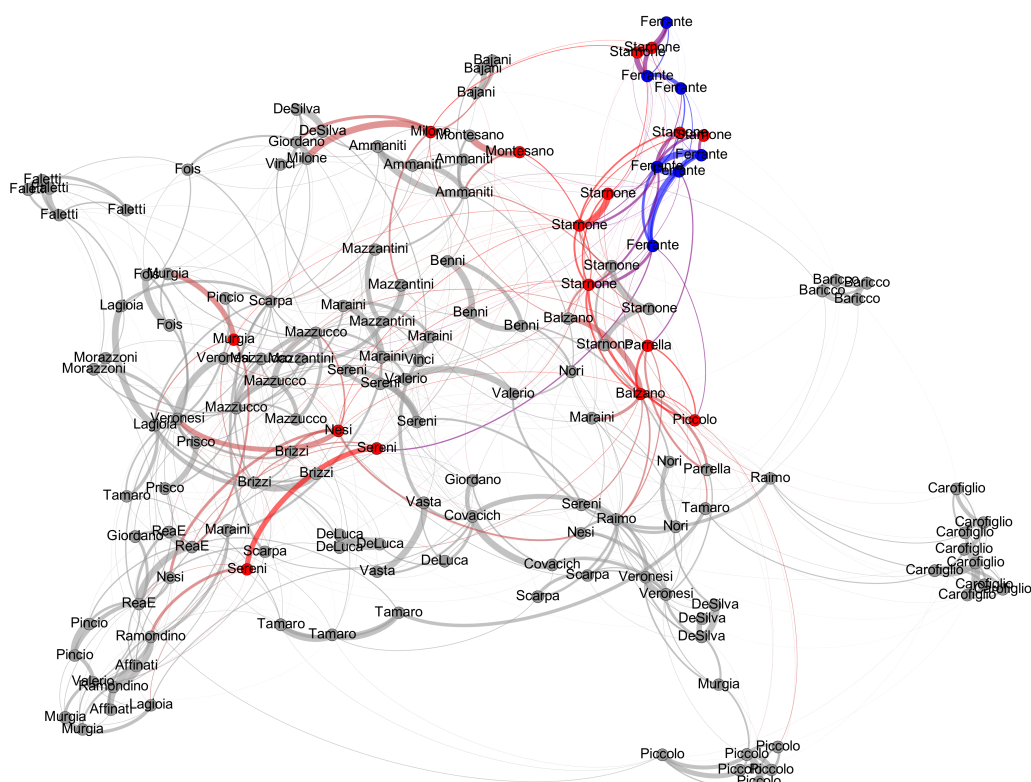
A tanulmányban közzétett összes számítási elemzés elvégzéséhez – beleértve a stilometriai hálózatok megrajzolását is – az R program *Stylo* bővítményét alkalmaztuk.<sup>11</sup> A végleges hálózatok rendezéséhez pedig a *Gephi* szoftvert és annak a beépített ForceAtlas2 algoritmusát használtuk.<sup>12</sup>

<sup>9</sup> John Burrows, „Delta’: A Measure of Stylistic Difference and a Guide to Likely Authorship,” *Literary and Linguistic Computing* 17, 3. sz. (2002): 267–287, <https://doi.org/10.1093/l1c/17.3.267>.

<sup>10</sup> Gareth, Witten, Hastie and Tibshirani, *An Introduction to Statistical Learning*, 104.

<sup>11</sup> Maciej Eder, Jan Rybicki and Mike Kestemont, „Stylometry with R: A Package for Computational Text Analysis,” *R Journal* 8, 1. sz. (2016): 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.

<sup>12</sup> Mathieu Bastian, Sébastien Heymann and Mathieu Jacomy, „Gephi: An Open Source Software for Exploring and Manipulating Networks,” in *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM. San Jose, California, May 17–20, 2009*, 361–362 (Menlo Park, CA: The AAAI Press, 2009), <http://doi.org/10.13140/2.1.1341.1520>.



1. ábra. Bootstrap Consensus Network: 150 kortárs olasz regény. A legközelebbi szomszédsági hasonlóságok hálózati kapcsolatokként ábrázolva. A Ferrantéhoz leginkább hasonló szövegek kékkel kiemelve.

A vizsgálatba bevont 150 kortárs olasz regény konszenzushálózata az 1. ábrán látható. Attól eltekintve, hogy az így kirajzolódó regényhálózat egészének elemzése további izgalmas eredményeket szolgáltatathat, a jelen vizsgálat középpontjában azon hálózatrészek állnak, amelyek Ferrante regényei környezetében fordulnak elő. A hálózatot szemlélve feltűnő lehet számunkra Ferrante és Starnone közelsége. Ebből következtethetünk arra, hogy az ábrán látható írók közül Starnone az, akinek a regényei stilometriailag leginkább egyezést mutatnak Ferrante műveivel. Megállapítható még, hogy lényegében mind a hét Ferrante-regény szoros kapcsolatban áll Starnone egyik művével. Azonban vannak más olyan szerzők is a hálózatban, akik időnként kapcsolódnak Ferrante regényeihez, tehát nem zárhatók ki a további stilometriai mérésekből. E regényírók a következők: Rosella Milone, Giuseppe Montesano, Valeria Parrella, Francesco Piccolo, Clara Sereni, Michela Murgia és Edoardo Nesi. A felsorolás vegyesen tartalmaz női és férfi szerzőket, ennek ellenére mindegyiket potenciális jelöltnek tekinthetjük az Elena Ferrante írói álnév mögött megbúvó szerzősége.

## Ferrante, darabokban

Az előző részben ismertetett vizsgálat eredményeit áttekintve megerősíthetjük azt a hipotézist, miszerint Elena Ferrante valóban Domenico Starnone írói alteregója lehet, továbbá azonosíthatóvá váltak olyan szerzők, akik művei ugyan kevésbé mutatnak

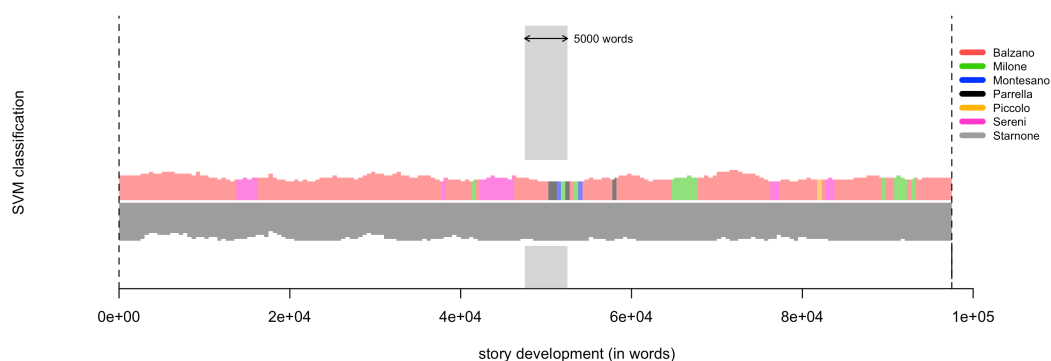
egyezést stilometriai szemszögből Ferrante regényeivel, azonban a térbeli közelség miatt mégsem zárhatók ki a potenciális szerzőjelöltek listájáról. Az így listázott művek közötti lehetséges kapcsolatok feltárásához ezután az előzőektől alapjaiban eltérő, újszerűnek számító kísérleti módszert alkalmazunk. A Rolling Stylometry eljárásban a fókuszpontban lévő (anonim) szövegeket kisebb részekre szegmentáljuk, majd ezeket szekvenciális elemzéssel dolgozzuk fel.<sup>13</sup> Ez a módszer azon feltételezésre épül, hogy a stilometriai lenyomatoknak nem szükségszerűen kell egyenletes mértékűnek lenniük egy adott szöveg különböző részeiben. Gondoljunk csak arra, hogy egy regényben vélhetően a narratív szövegrészek a dialógusoktól eltérő stiláris profillal bírnak. Hasonlóképpen gondolkodhatunk azokról a szövegegységekről, amelyek kiterjedtebb intertextuális kapcsolatokat hordoznak; az alkalmazott módszeren keresztül ezen szekvenciákat is detektálni tudjuk, hiszen mindegyik regény szegmenseit külön, ugyanakkor egymás után rendezve teszteljük.

Két oka van, hogy a Rolling Classify eljárást alkalmazzuk a regények vizsgálatára. Az első, hogy könnyen lehet, hogy a Ferrante álnév mögött nem egy, hanem több szerző áll. A második pedig, hogy az előzőekben bemutatott vizsgálat mintegy madártávlatból enged betekintést a szövegek hálózatrendszerébe, míg az imént vázolt eljárás egyfajta mikroszkópként is működhet, hiszen itt a kulcs az, hogy a szövegek minden szekvenciáját egyenként vetjük analízis alá.

Az alkalmazott eljárási mód ellenőrzött gépi tanuláson alapuló klasszifikációs módszer, amely használatához szükségünk van egy tanítókorpuszra (*training set*), azaz egy manuálisan válogatott, ismert szerzőségű szövegekből álló gyűjteményre, amely reprezentatívnak bizonyul az általa tartalmazott szerzői csoportokra vonatkozóan; és természetesen szükségünk van magára a tesztkorpuszra (*test set*), mely tartalmazza a vizsgálni kívánt szövegtesteket. A különbség abban rejlik a szokásos ellenőrzött beállítások és a Rolling Classify között, hogy az utóbbi a bemeneti (ismeretlen szerzőségű) szövegeket egyenlő méretű szekvenciákra darabolja fel, majd a tesztkorpuszt ezekkel a diszkrét szövegdarabokkal tölti fel, megtartva azok eredeti sorrendjét. Az eljárás kiegészül egy komplex vizualizációval is, amely az eredeti szöveget, vagyis a darabjainak sorrendjét színes csíkok formájában mutatja. A csík *alsó* része a klasszifikációs metódus döntését ábrázolja: minél szélesebb a csík, annál megbízhatóbb az osztályzás.

A jelen vizsgálatban az összes szekvenciális kísérletet ugyanazon hiperparaméter-beállításokkal hajtottuk végre. A tanítókorpusz 30 regényt tartalmazott az előző mérések által azonosított hét különböző szerzőtől: Marco Balzano, Rossella Milone, Giuseppe Montesano, Valeria Parrella, Francesco Piccolo, Clara Sereni és Domenico Starnone. A tesztek során Ferrante minden regényét 5000 szavas blokkokra daraboltuk fel, úgy, hogy a blokkok közötti lépésköz 500 szó legyen; tehát a blokkokat meghatározó vizsgálati „ablak” mindig 500 szóval haladt tovább az eredeti szövegen.

<sup>13</sup> Maciej Eder, „Rolling Stylometry,” *Digital Scholarship in the Humanities* 31, 3. sz. (2016): 457–469, <https://doi.org/10.1093/lhc/fqv010>.



2. ábra. Ferrante *L'amica geniale* című regénye, részletenként a szűkített korpusz 7 szerzőjével szembeállítva. Jellemzők: Rolling Classify módszer, SVM-osztályozás, a 100 leggyakoribb szóra.

A *L'amica geniale* (2011) című regényre vonatkozó eredményeket a 2. ábra szemlélteti. A vonal alatt látható egységes csík nem mutat stiláris különbséget, tehát az azonosító eljárás a regény minden szegmensét Domenico Starnone szerzőségéhez rendeli. A vizsgálatban minden kapott vizualizáció azonos képet vázol fel: a klasszifikációs metódus kivétel nélkül Starnonénak tulajdonítja a vizsgált szövegtesteket, amely erős bizonyítékul szolgál az előzetes, szerzősége irányuló hipotézis megerősítéséhez.

### Egy virtuális szerző?

Noha a szerzőazonosításra irányuló vizsgálatok sohasem zárhatók le biztos eredményekkel, a vizsgálat statisztikai alapokon működő eljárásai stabil empirikus bizonyítékokat nyújtottak annak megerősítésére, hogy az anonim szerzőség mögött Domenico Starnone szerzői ujjlenyomatát sejtthessük. Bármennyire is érdekes ez, azzal tisztában kell lennünk, hogy önmagában a szerzőazonosítási eljárás nem biztosít rálátást a vizsgált művek keletkezésének körülményeire. Ugyan izgalmas feltárni azt, ki áll az ismeretlen szerzőségű szövegek mögött, még izgalmasabb kérdés, hogy egy adott szerző mennyire képes szándékosan megváltoztatni a saját írói stílusát. Tételezzük fel, hogy Ferrante azonos Starnonéval. Vajon milyen kapcsolat áll fenn a „két” szerzői életmű között, és milyen stiláris, stilometriai különbséget mutathatunk ki közöttük?

Természetesen a kettős identitás problematikája nem új keletű. Az egyik legismertebb példa a világirodalomban Romain Gary (1914–1980) francia regényíró életműve, aki szerzői pályafutásának egy bizonyos időszakában úgy döntött, hogy személyét „elmaszkolva”, Émile Ajar álnéven publikálja műveit. Ő az egyetlen olyan szerző, akinek kétszer ítélték oda a Goncourt-díjat (a díj csak egyszer nyerhető el). Ez úgy történhetett meg, hogy egyszerűen nem derült ki, hogy valójában egyetlen szerző áll a két életmű mögött. A ritmikai mintázatokon alapuló stilometriai vizsgálatok ráadásul azt sugallták, hogy a két szerzői profil valóban mutat különbséget, még akkor is, ha az nem minden esetben egyértelmű.<sup>14</sup>

<sup>14</sup> Adam Pawłowski, *Séries temporelles en linguistique: Avec application à l'attribution de textes Romain Gary et Emile Ajar* (Lausanne: Slatkine, 1996).

A saját szerzői profil megváltozásának azonban nem feltétlen kell szándékosnak lennie. Érdeemes megemlíteni azokat az eseteket is, amikor ezek a stílári elmozdulások organikusan jönnek létre az idő múlásával.<sup>15</sup> Ezt a jelenséget figyelhetjük meg például Henry James életművében is.<sup>16</sup> Érdekesség továbbá, hogy a stilometriai módszer egyik első alkalmazása hasonló módon Platon szerzői ujjlenyomata időbeli alakulásának feltárására irányult.<sup>17</sup> A szerzői lenyomat megváltozása mögött olykor betegségek is állhatnak, mint például a demencia Agatha Christie esetében,<sup>18</sup> de a változás akkor is kimutatható lehet, amikor egy szerző segítséget kér a művei lejegyzéséhez, majd a lejegyző stílusa dominánsabbá válik az eredeti szerzőéhez képest – példaként említhető Francis Bacon pályájának egy szakasza.<sup>19</sup>

A jelen tanulmány esetében releváns hivatkozási pont a Mike Kestemont és szerzőtársai által végezett kutatás. Ennek során két középkori latin nyelven lejegyzett látomást vetettek analízis alá, amelyeket a szakirodalom hagyományosan Bingeni Hildegardnak tulajdonít.<sup>20</sup> A kutatók megfigyelései alapján azonban kijelenthető, hogy Hildegard és lejegyzője, Gibrólux-i Guibert együttműködése elkülönülő szerzői profilt alkot, amely nem feleltethető meg egyszerűen Hildegard és Guibert stílusának kombinációjaként. Ebben a példában a kooperációt nevezhetjük inkább egy virtuális „harmadik” szerző profiljának is. A virtuális szerzőség létezésének felvetése a szinergiahypotézis elméleti keretéből származtatható,<sup>21</sup> miszerint egy közösen írt mű szerzői lenyomata mutathat hasonlóságot a domináns szerzői kéz ujjlenyomatával, azonban akár el is térhet mindegyik alkotó önálló szerzői profiljától.

A következő vizsgálat arra irányult, hogy feltárja, vajon Starnone és Ferrante esetében beszélhetünk-e ilyen virtuális szerzőségről, valamint hogy Starnone saját neve alatt megjelent írásai mennyire különböznek el a Ferrante név alatt publikált szövegektől. Ez utóbbi kérdés valójában igen összetett, amely mentén további kutatói kérdések fogalmazhatók meg. Például az, hogy a Ferrante név alatt publikált szövegek több női szerzőre jellemző tulajdonságot mutatnak-e fel, mint azok, amelyeket az író a saját neve alatt jelentetett meg. Vajon a tényleges szerzőnek sikerült kilépnie a saját nyelvi

<sup>15</sup> Constantina Stamou, „Stylochronometry: Stylistic Development, Sequence of Composition, and Relative Dating,” *Literary and Linguistic Computing* 23, 2. sz. (2008): 181–199, <https://doi.org/10.1093/llc/fqm029>.

<sup>16</sup> David L. Hoover, „Corpus Stylistics, Stylometry, and the Styles of Henry James,” *Style* 41, 2. sz. (2008): 174–203.

<sup>17</sup> Wincenty Lutosławski, *The Origin and Growth of Plato’s Logic: With an Account of Plato’s Style and of the Chronology of his Writings* (London: Longmans, Green & Co, 1897).

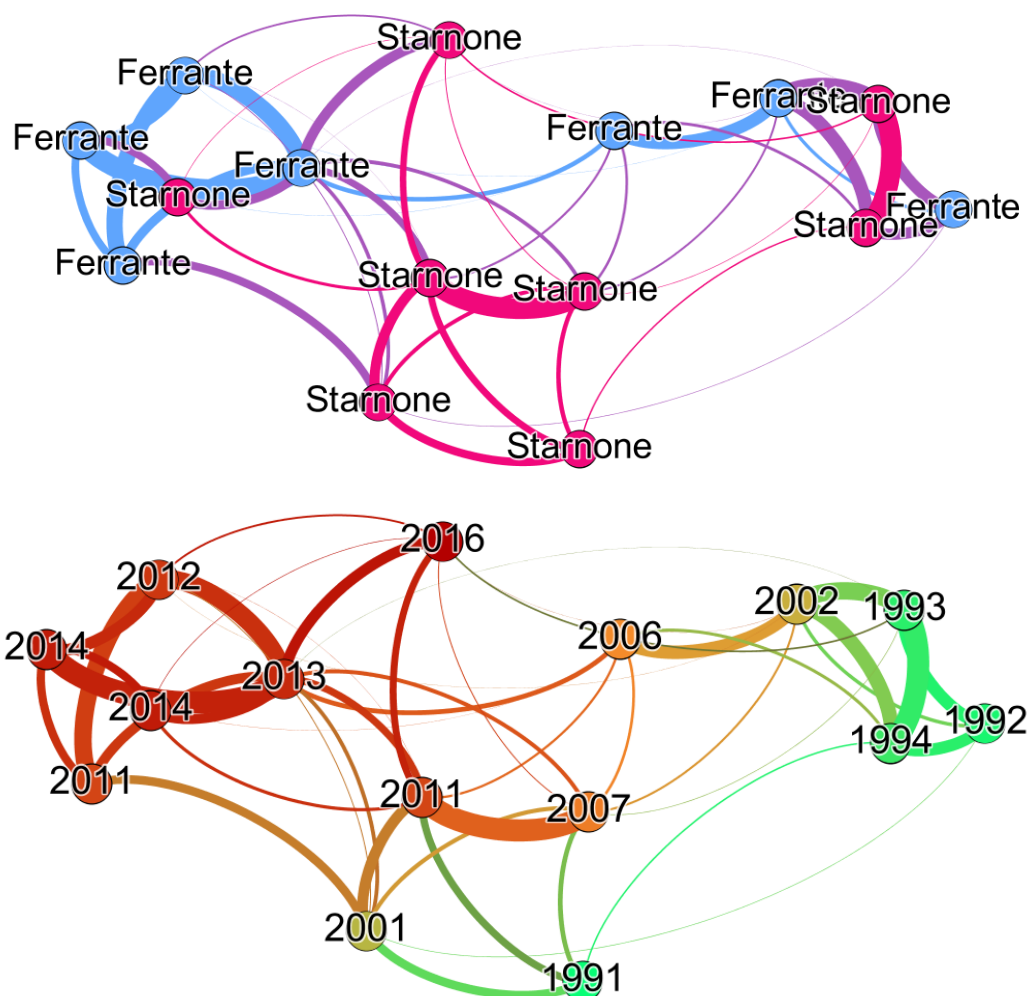
<sup>18</sup> Xuan Le, Ian Lancashire, Graeme Hirst and Regina Jokel, „Longitudinal Detection of Dementia through Lexical and Syntactic Changes in Writing: A Case Study of three British Novelists,” *Literary and Linguistic Computing* 26, 4. sz. (2011): 435–461, <https://doi.org/10.1093/llc/fqr013>.

<sup>19</sup> Noel B. Reynolds, Bruce G. Schaalje and John L. Hilton, „Who Wrote Bacon? Assessing the Respective Roles of Francis Bacon and his Secretaries in the Production of his English Works,” *Literary and Linguistic Computing* 27, 4. sz. (2011): 409–425, <https://doi.org/10.1093/llc/fqs020>.

<sup>20</sup> Mike Kestemont, Sara Moens and Jeroen Deploige, „Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux,” *Literary and Linguistic Computing* 30, 2. sz. (2015): 199–224, <https://doi.org/10.1093/llc/fqt063>.

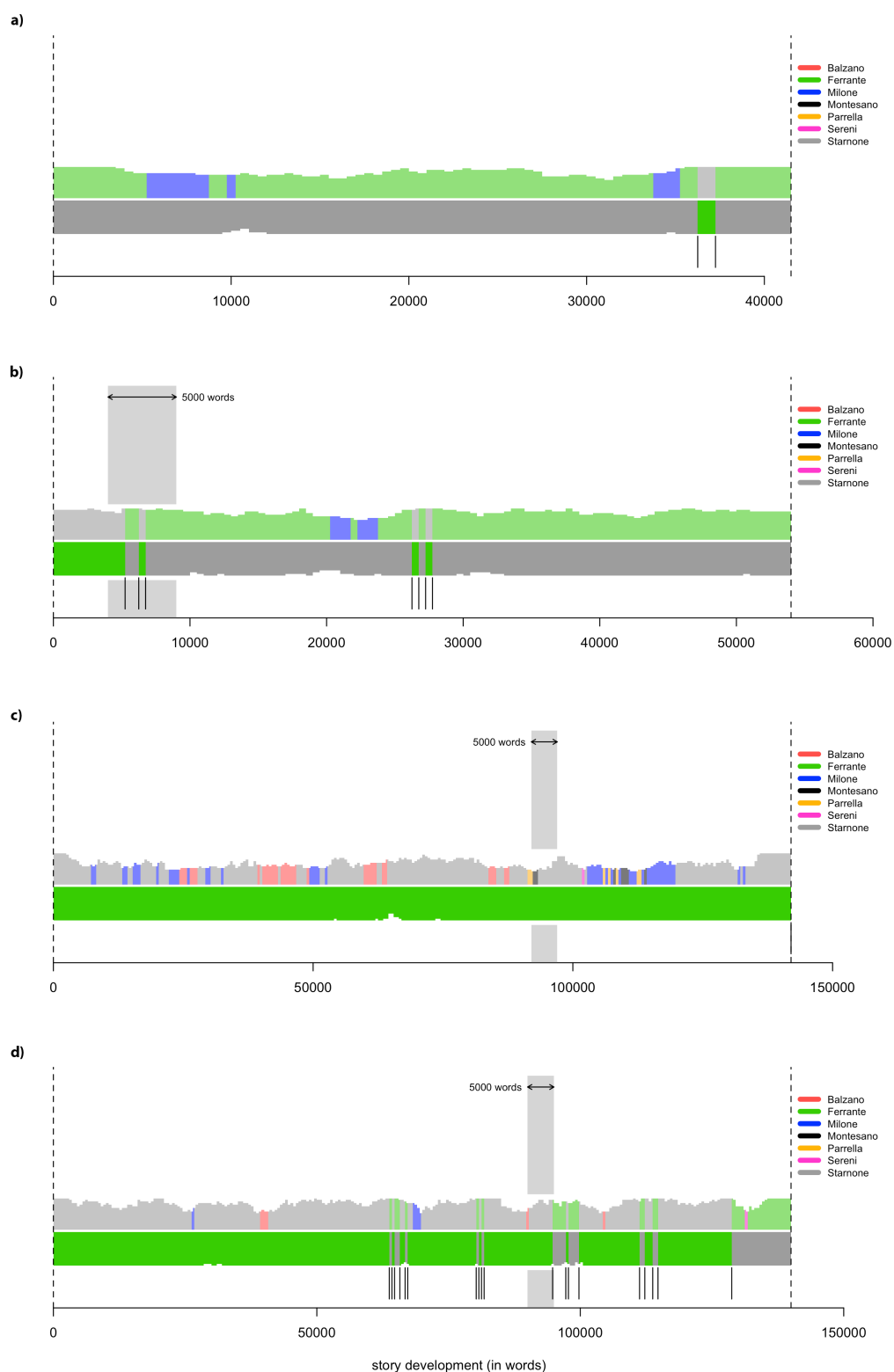
<sup>21</sup> James W. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us* (New York: Bloomsbury Press, 2011).

beágyazottságából? Azonban ezen kérdéseknek csupán töredéke az, amelyre a jelen tanulmányban választ tudunk adni.

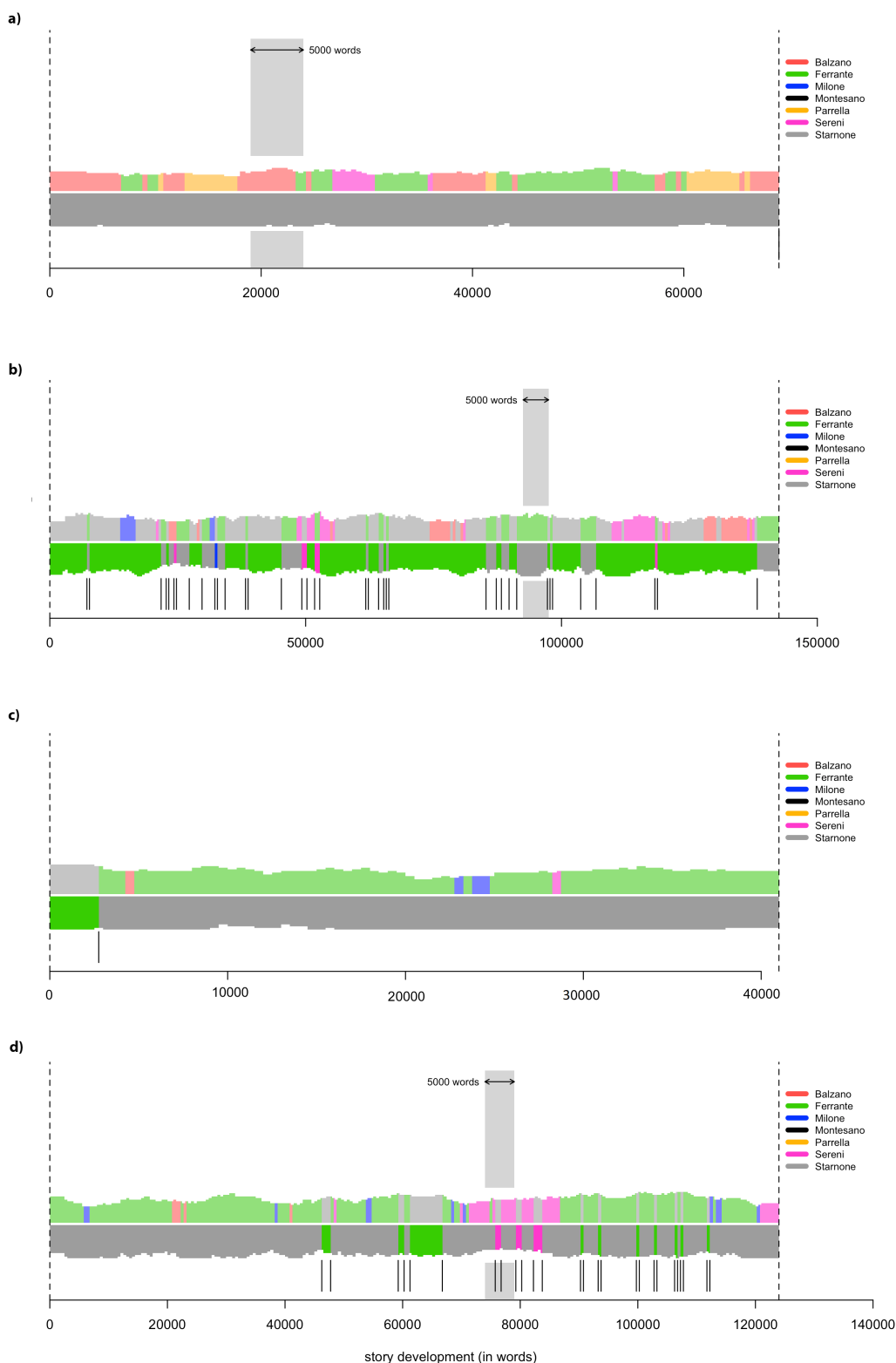


3. ábra. Bootstrap Consensus Network a Ferrante és Starnone által írt regények közötti hasonlóságokról: a) a két szerző közötti különbség: nem rajzolódik ki egyértelmű minta, b) egy időbeli jel megléte, amely erősebbnek tűnik, mint a két szerzői hang.

Az elemzés egy kis méretű konszenzushálózatból indul ki, amely kizárólag a Starnone és Ferrante regényei közötti kapcsolatok feltárására koncentrálódik. Ahogyan az a 3a. ábrán látható, a két szerzői profil különböző klasztereket alkot, de ezek a különbségek messze nem egyértelműek. Inkább az válik feltűnővé, hogy a regények időrendi mintába rendeződnek (lásd 3b. ábra). Még ha előzetesek is, ezek az eredmények egy további – és viszonylag erős – karakterjegyet fednek fel, amire érdemes odafigyelni: Starnone és alteregója feltételezett különálló hangját beárnyékolhatja a szerző általános stilisztikai változása az idők során.



4. ábra. Elena Ferrante regényeinek szekvenciális elemzése a Rolling Classify módszerrel. a) *L'amore molesto* (1992) b) *I Giorni dell'abbandono* (2002) c) *Storia del nuovo cognome* (2012) d) *Storia della bambina perduta* (2014). A tanítókorpusz 7 szerző 30 regényét tartalmazta, köztük Ferrante és Starnone műveit is.



5. ábra. Domenico Starnone regényeinek szekvenciális elemzése a Rolling Classify módszerrel. a) *Il salto con le aste* (1989) b) *Via Gemito* (2000) c) *Prima esecuzione* (2007) d) *Lacci* (2014). A tanítókorpusz 8 szerző 36 regényét tartalmazta, köztük Ferrante és Starnone műveit is.



A fent vázolt probléma feloldásához további vizsgálatssorozatokot végeztünk. Ezúttal a tanítókorpusz a feljebb említett hét szerző munkáit és Ferrante regényeit tartalmazta. Az így elvégzett vizsgálatok Ferrante és Starnone műveit külön csoportba sorolták. A kutatási kérdés a következő tehát: a klasszifikáció vajon azonosítja-e a „virtuális” Ferrantét? A hiperparaméter-beállítások ugyanazok maradtak, mint az előző vizsgálat esetében is: Support Vectors Machine módszer (SVM), 100 leggyakoribb szó, 5000 szóból álló vizsgálati „ablak”, 500 szavas csúsztatással. A kísérletek eredményeit a 4a–d. ábra szemlélteti; Ferrante elemzett regényei kronologikus sorrendben vannak feltüntetve. A mintázat nem rajzol ki teljesen egyértelmű képet: míg a korai Ferrante-regények viszonylag kevés hasonlóságot mutatnak a feltételezett „virtuális” Ferrante szerzői ujjlenyomatával, addig a kései művei esetében már nagyobb egyezést vélhetünk felfedezni a klasszifikációs eljárás alapján. A *L'amore molesto* (1992) című Ferrante-regény szinte minden szegmensét a klasszifikáció Starnonénak tulajdonítja, kivéve egy viszonylag rövid szövegrészletet a mű végén. A „virtuális” Ferrante-hang az *I Giorni dell'abbandono* (2002) című regény esetében élesebben körvonalazódik, ezúttal a szöveg elején. A *La figlia oscura*ban (2006) a ferrantei szegmensek aránya nagyjából megegyezik a Starnonénak tulajdonított szekvenciák arányával. A *L'amica geniale. Infanzia adolescenza* (2011) esetében a szerzői profil túlnyomórészt Ferrantére utal, ezen tendencia még erősebben kimutatható a *Storia del nuovo cognome* (2012) című regényben (lásd 4c. ábra). Ez utóbbi regényben már egy markáns virtuális szerzői hang azonosítható, hiszen a klasszifikációs eljárás az összes szegmenst Ferrantéhez kapcsolta. A két későbbi művében – *Storia di chi fugge e di chi resta* (2013) és *Storia della bambina perduta* (2014) – a szerzői ujjlenyomat már homályosabb foltokat is mutat (lásd 4d. ábra), azonban a virtuális szerzői hang jelenléte tagadhatatlan ezen szövegek esetében is.

Bármennyire is egyértelműnek tűnnek a kapott eredmények, akár némi kételyt is felvethetnek. A vizsgálat során kinyert eredmények elég erős bizonyítékként támasztják alá azt a hipotézist, hogy Ferrante „szerzői” profilja fokozatosan alakult ki, és vált uralkodóvá a kései regényeiben. Mielőtt azonban kétségek nélkül elfogadnánk ezen állítást, meg kell vizsgálnunk, hogy Starnone szerzői hangjának stabilitása fennáll-e a vizsgált periódusban. Hiszen felmerülhet egy olyan lehetőség is, amely szerint Starnone szerzői hangja egyszerűen „ferranteivé” változott (jelen esetben lényegtelen a publikáló név). Ez a felvetés az eddig bemutatott vizsgálatok alapján nem kizárható. Sőt az előzőekben bemutatott kísérlet (3. ábra) már rámutatott az időbeliség központi szerepére az adathalmaz vizsgálatakor.

A Ferrante regényeivel összefüggő eredmények kiértékeléséhez így további kísérletekre volt szükség, ezúttal Starnone regényeire helyezve a fókusz. Természetesen a paraméterbeállításokon nem módosítottunk az előző vizsgálatához képest. Az eredményeket az 5a–d. ábra mutatja (a tíz regényből négyet). Ezen ábrák egy egyedi, a virtuális ferrantei hangtól különböző starnonei hang létezéséről értesítenek, még ha a felvázolt kép nem is annyira egyértelmű, mint ahogyan azt előzetesen vártuk. Különösen akkor szembetűnő néhány stilisztikai sajátosság, ha időrendi sorrendbe rendezzük az elemzett regényeket. Nézzük először a legelső regényeket: az *Ex cattedra* (1987) szegmentált szövegteste egyértelműen Starnone szerzői profilját tükrözi, akárcsak az 1989-ben megjelent *Il salto con le aste* (5. ábra). A *Fuori registro* (1991) esetében nagyjából hasonló

a kép, eltekintve attól, hogy a regény elején, egy néhány ezer szavas szegmens ferrantei besorolást kapott a klasszifikációs eljárás során. Ezután nézzünk rá az *Eccesso di zelo* (1993) című műre, amelyben ismételtén Starnone szerzői ujjlenyomata érvényesül, noha két rövidebb rész Ferrante szerzőségét mutatja fel. Változó tendencia mutatkozik a *Denti* (1994) regény klasszifikációjánál, mivel az eljárás fele-fele arányban kapcsolta a szövegtestet Ferrantéhoz és Starnonéhoz, viszont a *Via Gemito* (2000) című műben a módszer már egyáltalán nem azonosította Starnonét. Meglepő, hogy a klasszifikáció az utóbb említett regényt még nem is csupán Ferrantéhoz rendelte, hanem néhány kisebb szövegszegmensét Clara Sereninek (!) tulajdonította. Ezen eredmény ellentétben áll azzal a hipotézissel, miszerint Starnone képes elhatárolni egymástól a szerzői hangokat. Másrésről a következő regény, nevezetesen a *Prima esecuzione* (2007), semlegesíti az 5c. ábrán felmerülő kételyt, hiszen Starnone szerzői hangja ebben újra tisztává válik, alig tartalmaz olyan részeket, amelyek esetében felmerül a „ferrantei” hang. Az *Autobiografia erotica di Aristide Gambia* című regényben a szerzőnek továbbra is adatolhatóvá válik a stabil írói profilja, még annak ellenére is, hogy néhány részlet véletlenszerűen Elena Ferrante vagy Clara Sereni szerzői hangjával kapcsolódik össze.

A két perszóna közötti váltás azonban nem egyértelmű a világhírű író esetében, erre bizonyíték a következő, *Lacci* (2014) című mű, amely egyértelműen tartalmaz olyan, viszonylagosan hosszú szegmenseket, amelyek már újra a ferrantei hangon szólalnak meg (lásd 5d. ábra). Azonban az öt éve megjelent (2016) *Scherzetto* már újra evidenciaként szolgál arra, hogy a szerző képes sikeresen elnémítani a ferrantei szólamokat: néhány marginális szövegdarabon kívül a mű túlnyomórésze egyértelműen Starnonénak tulajdonítható. Annak ellenére is sejthető a szerzői hangok mögötti tudatosság, hogy a fenti eredmények némelyike nem teszi lehetővé a határozott konklúzió megfogalmazását. A Ferrante- és a Starnone-regények jelentős többsége képes tiszta szerzői hangot megszólaltatni. Kiderült számunkra, hogy a nápolyi író irodalmi színjátéka több mint sikeresnek tekinthető, legalábbis a feltárt stilometriai eljárások felől közelítve.

## Konklúziók

A tanulmány arra vállalkozott, hogy olyan regények stilometriai alapú szerzőazonosítását mutassa be, amelyek Elena Ferrante írói álnév alatt jelentek meg. Ahogy azt vártuk, a már létező hipotézist, miszerint Domenico Starnone szerzősége lapul az álnéven publikált regények mögött, majdnem cáfolhatatlanul bizonyította a bemutatott vizsgálat. Azonban a tanulmány fő célja azt volt – a szerzőazonosítás viszonylag egyszerű kérdésén túl –, hogy választ találjon arra is, a szerzői álnév alatt megbújó író milyen mértékben, vagy egyáltalán képes-e stilisztikailag szétválasztani a két szerzői identitását.

A Rolling Classify tesztsorozat – függetlenül attól, hogy Starnone vagy Ferrante regényein alkalmaztuk – lehetővé tette az általános megfigyelések rögzítését. Az összkép, néhány kivételtől eltekintve, megerősíti, hogy Starnone és Ferrante hangja megkülönböztethető egymástól, és ez elég erős evidenciának tűnik ahhoz, hogy a virtuális szerzőség hipotézise fenntartható legyen. Domenico Starnone egyértelműen

bizonyítja – különösen a kései műveiben – hogy saját stílusprofilját képes elhatárolni szerzői alteregójától.

Azonban további vizsgálatokat igényel az a jelenség, hogy Starnone műveinek néhány részletét a klasszifikáció tévesen Clara Sereninek (*Via Gemito*), vagy Marco Balzanonak (*Lacci*) tulajdonította. A téves hozzárendeléseket magyarázhatja, hogy a klasszifikáció nem volt megfelelően betanítva a feladatra. Az ilyen példákat „hamis pozitívnak” nevezzük a gépi tanulásban. Másrészt itt foglalkozhatnánk a regények helyi stílusbeli sajátosságaival is. Hasonló módon például Harper Lee *Ne bántsátok a feketerigót!* című regényének stilometriai vizsgálatakor is lokálisan „beárnyékol” szerzői ujjlenyomatokat figyeltünk meg,<sup>22</sup> Truman Capote esetenként felismert stílusa azonban korántsem utal arra, hogy a vizsgált szöveg vegyes szerzőségű lenne. Valószínűbb, hogy inkább az intertextuális hangok (*intertextual voices*) jelenségével kell számolnunk, amelyek bizonyos szöveghelyeken átsugározhatnak az – általában átlátatlan – eredeti szerzői ujjlenyomaton. Vélhetően a *Via Gemito* és a *Lacci* esetében is hasonló jelenséggel állunk szemben.

A vizsgálat során feltárt eredmények azért is szembetűnőek, mert egy fokozatos – ugyanakkor sikeres – virtuális szerzői identitáskonstruálás művelete rajzolódik ki előttünk, amely a korai művekben alig fellelhető, míg a kései művekben már egyértelműen domináns jeggyé válik. Másrészt pedig az eredmények jelentős mértékben hozzájárulnak a fent említett szinergiahipotézis igazolásához.<sup>23</sup> Ez nemcsak két különböző szerzői hang stílus kombinációjára vonatkozhat, hanem arra is, hogy két különböző stílus személyiség rejtőzhet egyetlen tényleges szerző mögött.

Fordította: Bajzát Tímea

## Elena Ferrante: A Virtual Author

The present study scrutinizes the novels by Elena Ferrante, in order to discover the actual writer hidden behind the pseudonym. Rather than simply reopen the authorship question, however, the paper attempts at testing the stability of the authorial signal in the works by “Ferrante”, whoever the actual author might be. To address the research question, a network of 150 novels and their stylistic similarities has been computed using the Bootstrap Consensus Network method. A list of authors most similar to “Ferrante”, including Domenico Starnone at the first place, was then analyzed using the technique Rolling Classify, which was designed to detect local stylistic idiosyncrasies in literary texts. The series of Rolling Classify tests – performed independently for the novels by both Ferrante and Starnone – allows for formulating general observations. The overall picture confirms, with a few exceptions, that Starnone and Ferrante can

<sup>22</sup> Maciej Eder and Jan Rybicki, „Go Set a Watchman while We Kill the Mockingbird in Cold Blood, with Cats and Other People,” in *Digital Humanities: Conference Abstracts*, 184–186 (Kraków: Jagiellonian University & Pedagogical University, 2016).

<sup>23</sup> Pennebaker, *The Secret Life of Pronouns*.

be told apart, which, in turn, seems to be a strong argument in favor of the virtual author hypothesis. Apparently, Domenico Starnone demonstrates, particularly in his late works, the ability to differentiate his own stylistic profile and the voice of his *alter ego*.

Keywords:

stylometry, authorship attribution, virtual author, Bootstrap Consensus Network, Rolling Classify