

Stylometric investigations into translationese: The Baby-Sitters Club across languages

Joanna Byszuk¹, Quinn Dombrowski²

¹ Institute of Polish Language of the Polish Academy of Sciences – joanna.byszuk@ijp.pan.pl

² Stanford University – qad@stanford.edu

10 Computational 01
01 Stylistics 0101000
11 Group 011010110



PAN **JP** Institute of Polish Language Polish Academy of Sciences



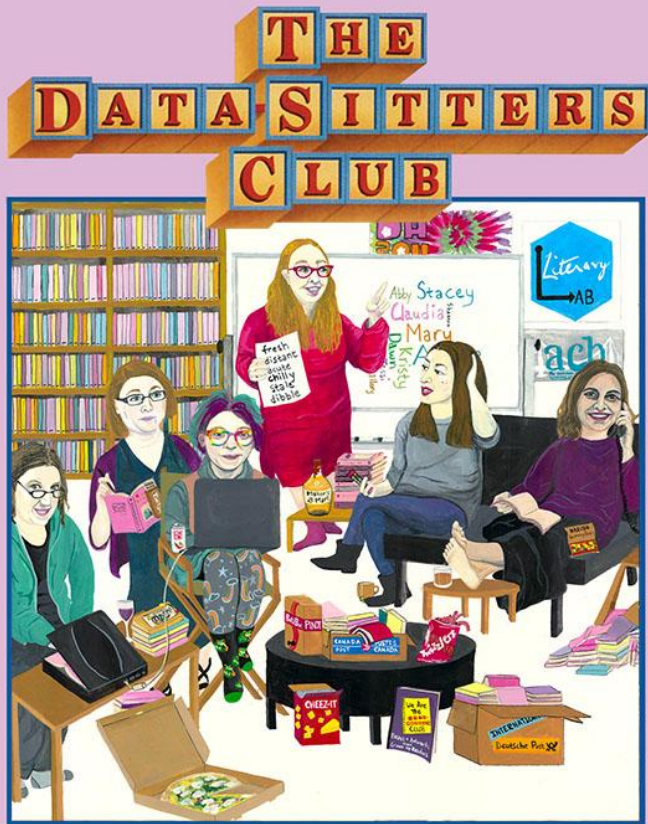
A translated text (...) is judged acceptable by most publishers, reviewers, and readers when it reads fluently, (...) the appearance, in other words, that the translation is not in fact a translation, but the "original." (...) The more fluent the translation, the more invisible the translator, and, presumably, the more visible the writer or meaning of the foreign text.

(Venuti 1995: 1-2)

Our research question

- is the impact of ghostwriters stylometrically visible?
- do translators have visible style? (cf. Jan Rybicki on Virginia Woolf, discussion on Anita Raya being Elena Ferrante in *Drawing Elena Ferrante's Profile*, ed. A. Tuzzi, M.A. Cortelazzo)



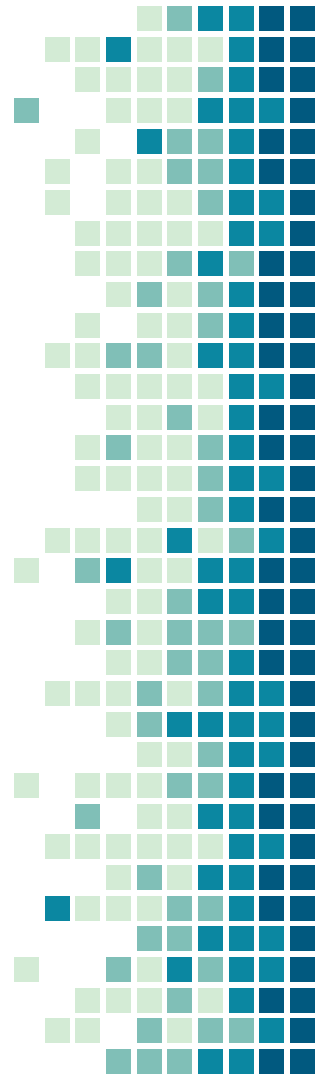


The Fun and Colloquial Guide to DH Computational Text Analysis

Lee Skallerup Bessette, Katherine Bowers,
Maria Sachiko Cecire, Quinn Dombrowski, Anouk Lang,
and Roopika Risam

Goals:

- apply DH methods to this corpus
- explain how they work in easy terms
- collect all the translations and compare them



Dataset: The Baby-Sitters Club

- a series of middle-grade novels written by Ann M. Martin, published from 1986 to 2000,
- translated into numerous languages, becoming international bestsellers
- our corpus: 142 translations into 6 language versions (distinguishing three French versions, next to Italian, Spanish, and Polish translations)



Method

Distant reading

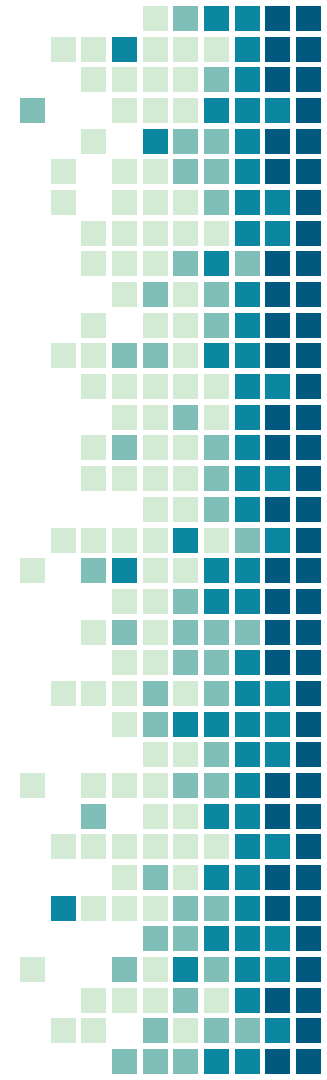
- **Analysing big literary collections** based on not text but metadata / research literature etc.

(as conceptualized by Franco Moretti 2000)

- **Analysing language / literature “objectively”**
 - **at a distance**, looking at particular features within the texts

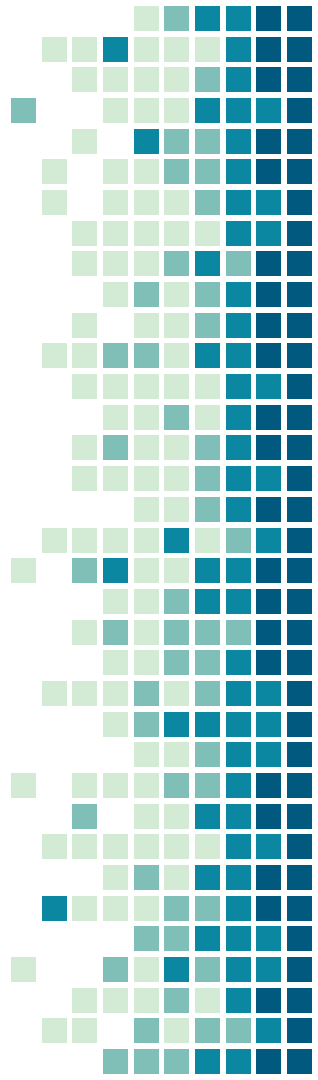
(Mendenhall 1887, Lutosławski 1890)

(but also Lorenzo Valla 1440, Augustus de Morgan 1851)



What is stylometry?

Stylometry =
use of quantitative methods
to examine similarities and differences
within a group of [texts]



Our approach

- cluster analysis
- bootstrap consensus tree + network visualization and additional modularity test
- 100-1000 MFW
- Cosine Delta and Burrows's Delta
- for EN and FR also tests with culling

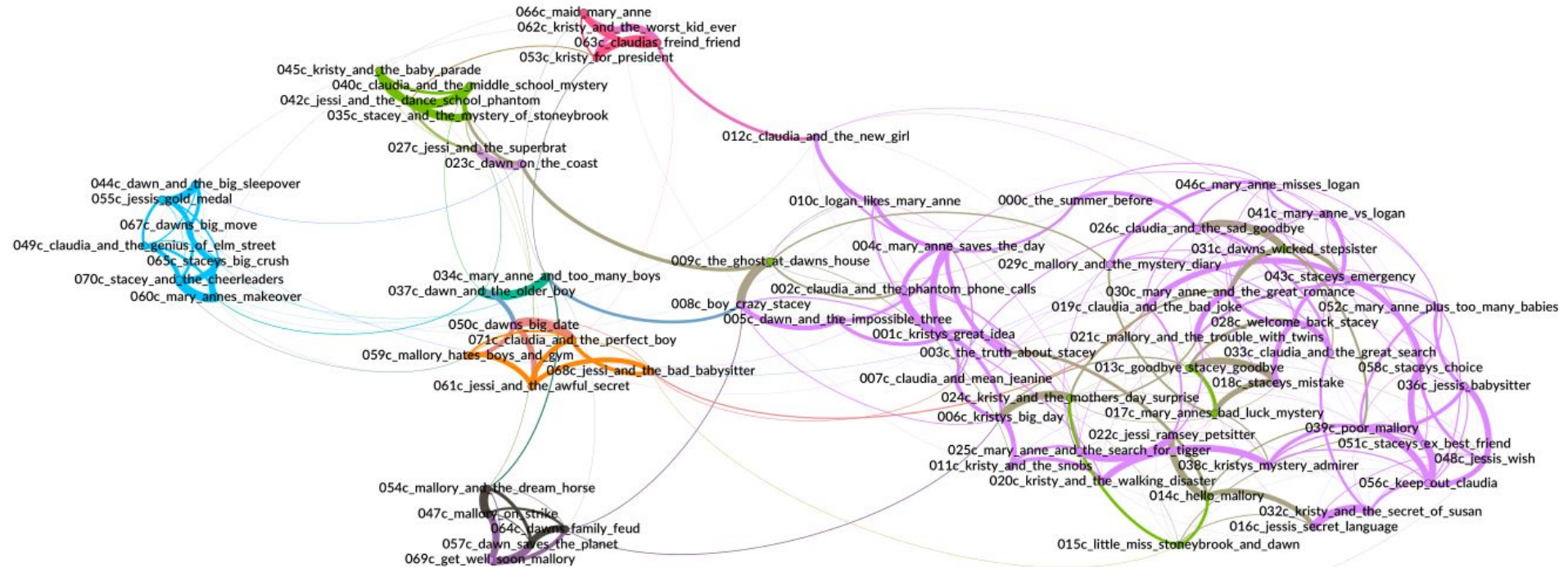
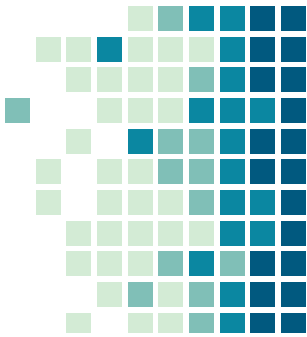
1. English as a baseline for the relations between texts,
2. French, considering all three language variants together,
3. Italian,
4. Polish,
5. Spanish.



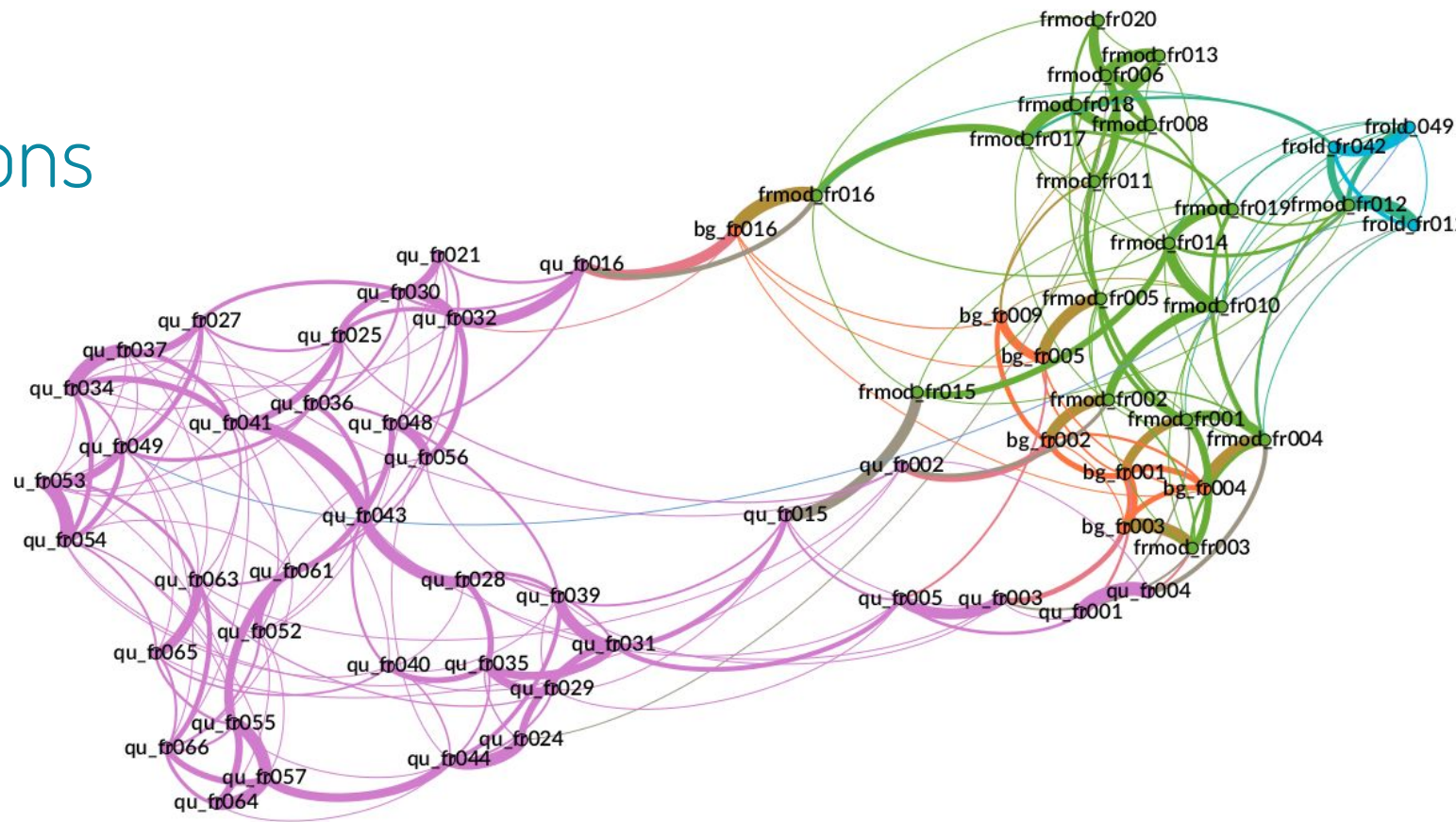
Results

More on: <https://github.com/JoannaBy/BSC-translationese/>

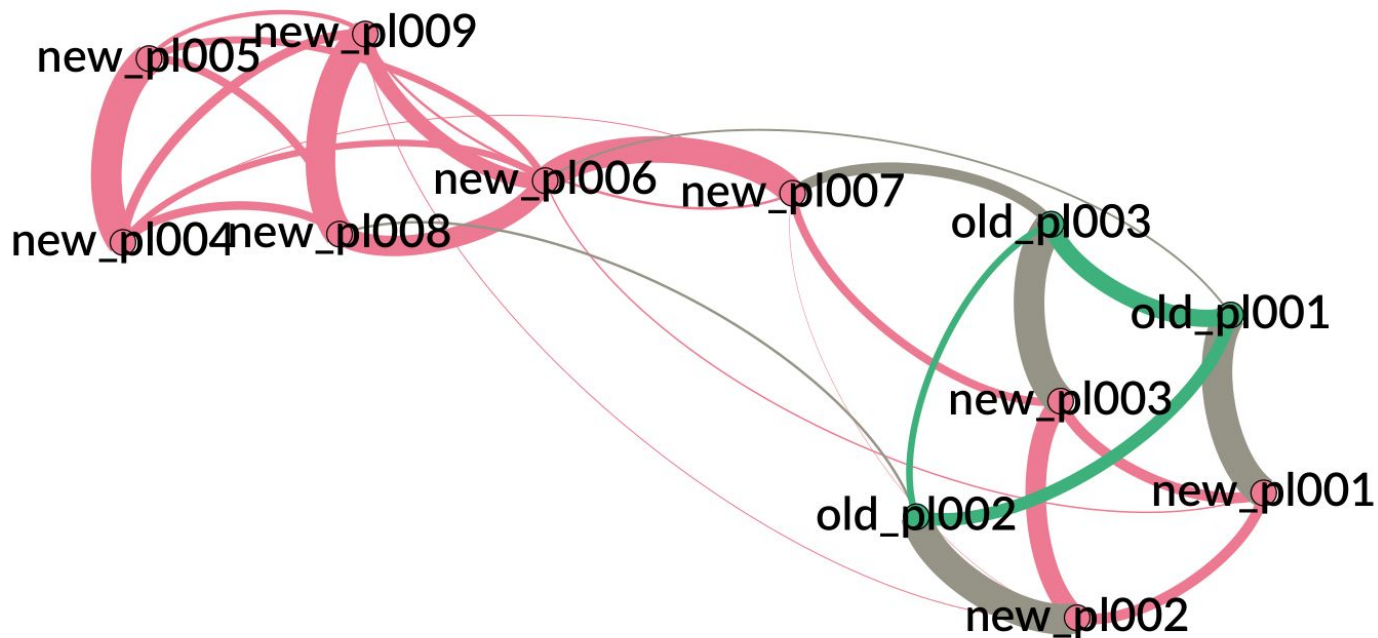
Originals (colors indicating (ghost –) writers



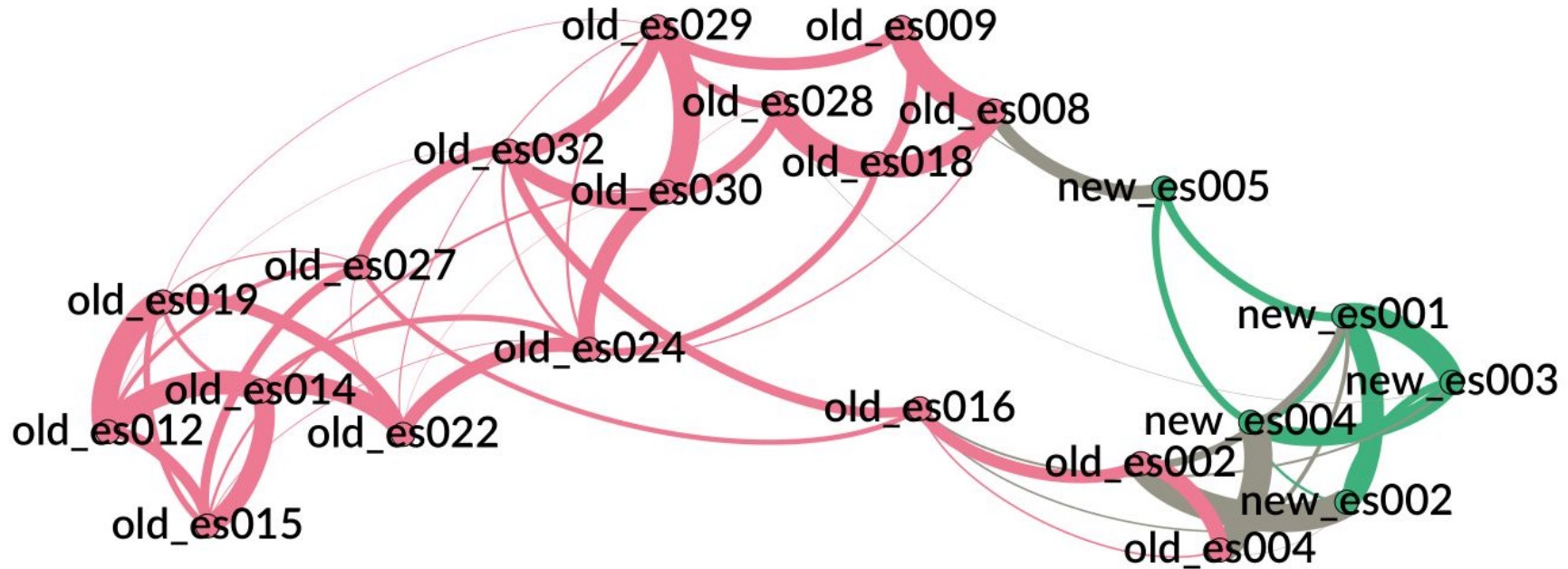
French translations



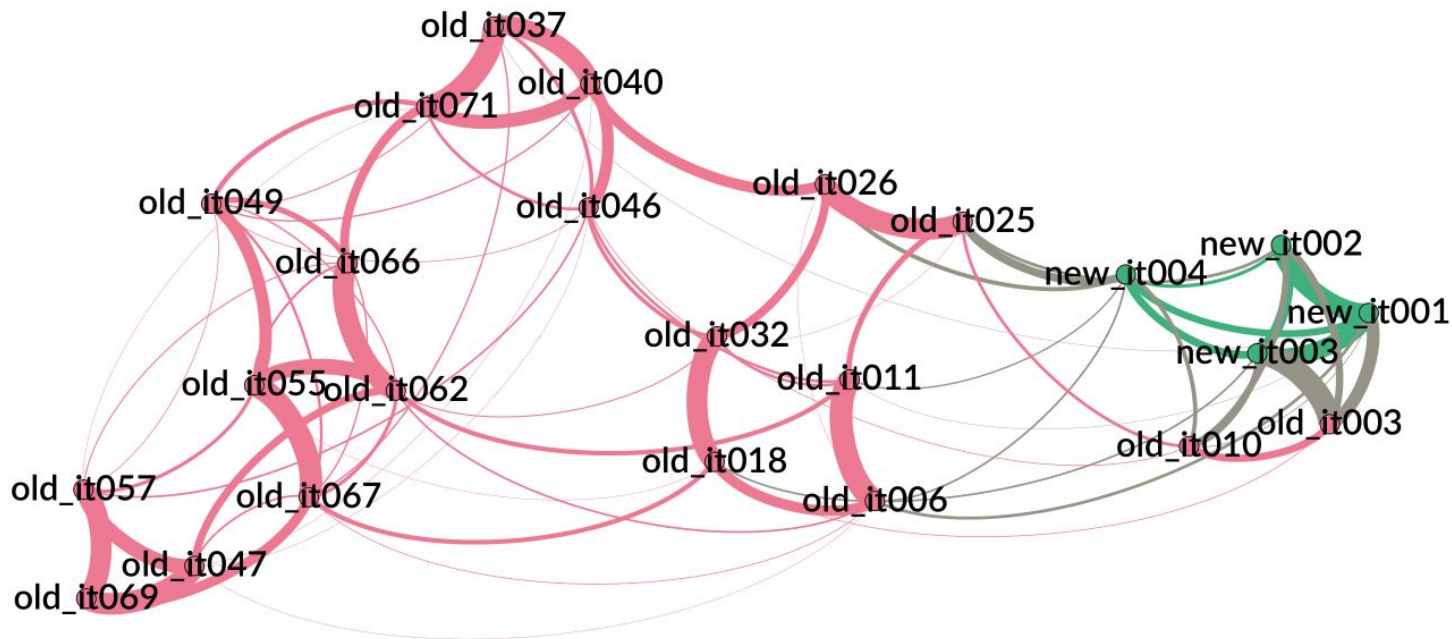
Polish translations



Spanish translations



Italian translations



Conclusions

- some translators are more visible than others
- more so in some language-circles in our corpus
= Spanish and Italian
- language variant carries strong stylistic signal
- ghostwriter is as visible as original author, also
in the translation



References

- Eder M. (2017). Visualization in stylometry: cluster analysis using networks. *Digital Scholarship in the Humanities*, 32(1): 50-64
- Eder M., Kestemont, M. and Rybicki, J. (2016). Stylometry with R: A package for computational text analysis. *R Journal*, 16(1): 107-121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
- Evert S., Proisl T., Vitt T., Schöch C., Jannidis F., and Pielström S. (2015). Towards a Better Understanding of Burrows's Delta in Literary Authorship Attribution. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 79–88.
- Lodge (2010): <https://www.publishersweekly.com/pw/by-topic/childrens/childrens-book-news/article/41524-the-baby-sitters-club-to-reconvene.html>
- Koppel M. and Ordan N. (2011). Translationese and its dialects. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11)*. Association for Computational Linguistics, USA, 1318–1326
- Newman M. EJ. (2006). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences* 103 (23): 8577–82.
- Rybicki J. and Heydel, M. (2013). The stylistics and stylometry of collaborative translation: Woolf's 'Night and Day' in Polish. *Literary and Linguistic Computing*, 28(4): 708-17
- Rybicki J. (2012). The great mystery of the (almost) invisible translator: stylometry in translation. In M. Oakley and M. Ji (eds.), *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam: John Benjamins, pp. 231-248
- Rybicki J. (2009). Translation and Delta revisited: when we read translations, is it the author or the translator that we really read? *Digital Humanities 2009: Conference Abstracts*, pp. 245-47.
- Venuti L. (2008). *The Translator's Invisibility: A History of Translation* (2nd ed.). Abingdon, Oxon, U.K.: Routledge
- Yuvaraj J. et al (2020). U.S. Copyright Termination Notices 1977–2020: Introducing New Datasets. *Journal of Empirical Legal Studies*, February 27, 2022. <https://doi.org/10.1111/jels.12310>.



Thank you for the attention!

