The Translator's Other Invisibility: Stylometry in Translation

SLE 2013

Jan Rybicki
Jagiellonian University
Kraków, Poland

Translator's Invisibility

A translated text, whether prose or poetry, fiction or nonfiction, is judged acceptable by most publishers, reviewers, and readers when **it reads fluently**, when **the absence of any linguistic or stylistic peculiarities** makes it seem **transparent**, giving the appearance that it reflects the foreign writer's personality or intention or the essential meaning of the foreign text—the appearance, in other words, that the translation is not in fact a translation, but the "original."

Lawrence Venuti, *The Translator's Invisibility: A History of Translation*, 1995

2

Translator's Style?

Little or no attention has been paid so far to the possibility of describing the 'style' of a translator or group of translators in terms of what might be distinctive about the language they produce. (We should) investigate the question of style in literary translation — not in the traditional sense of whether the style of a given author is adequately conveyed in the relevant translation but in terms of whether individual literary translators can be shown to use distinctive styles of their own.

Mona Baker, *Towards a methodology for investigating the style of a literary translator*, 2000

3

Stylometric authorship attribution

- in a set of texts, all you need know to order the texts by their authors are the frequencies of the most frequent words (sometimes as few as 30).

- stylometric authorship attribution only needs to compare the series of (relative) word frequencies for a collection of texts to correctly order the texts by their authors.

- this series for one text by one author will be most similar to this series for another text by the same author.

Burrows's Delta Distance

"the mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text"

J. Burrows, *'Delta' a measure of stylistic difference and a guide to likely authorship*, LLC 2002

Burrows's Delta Distance

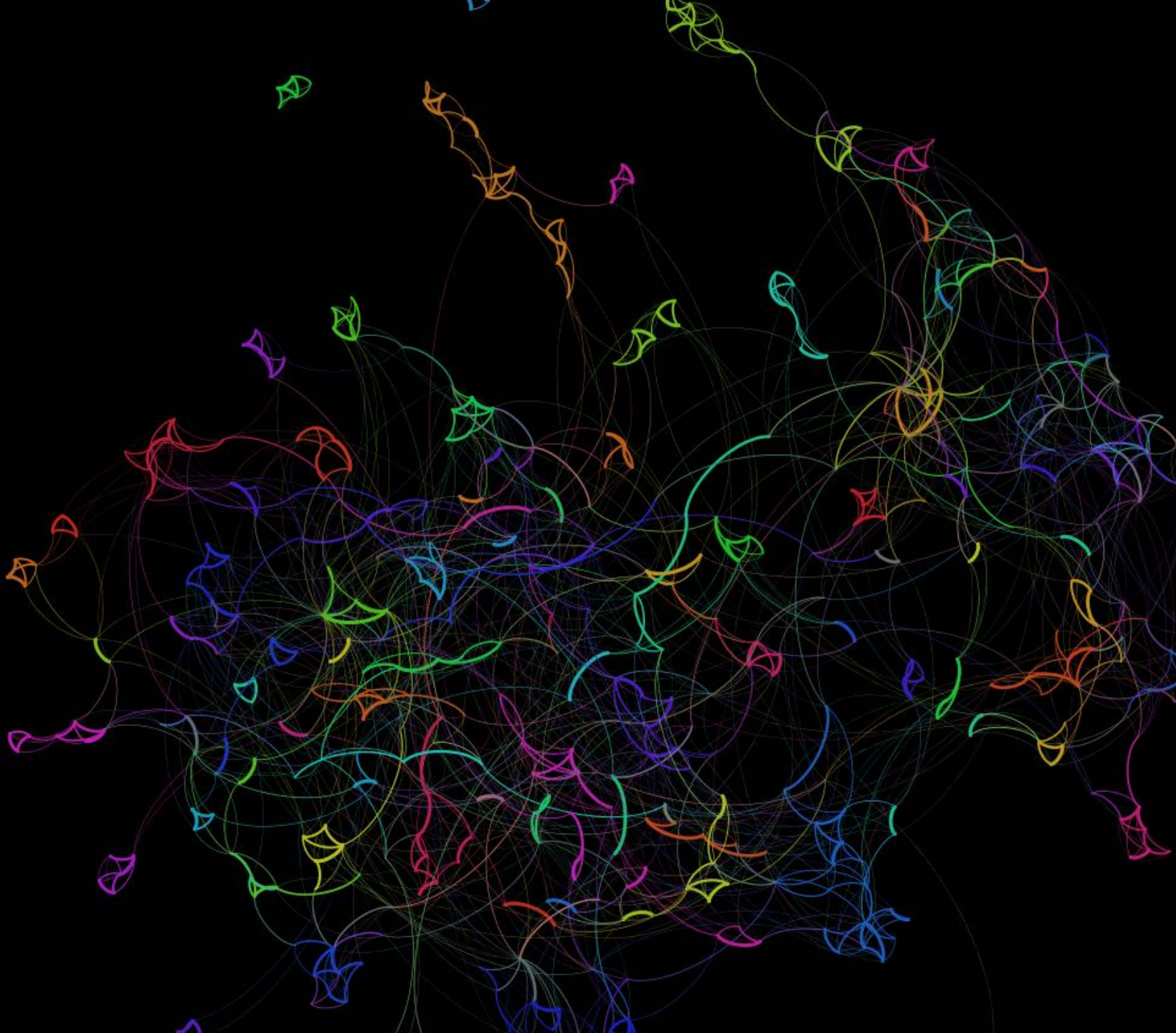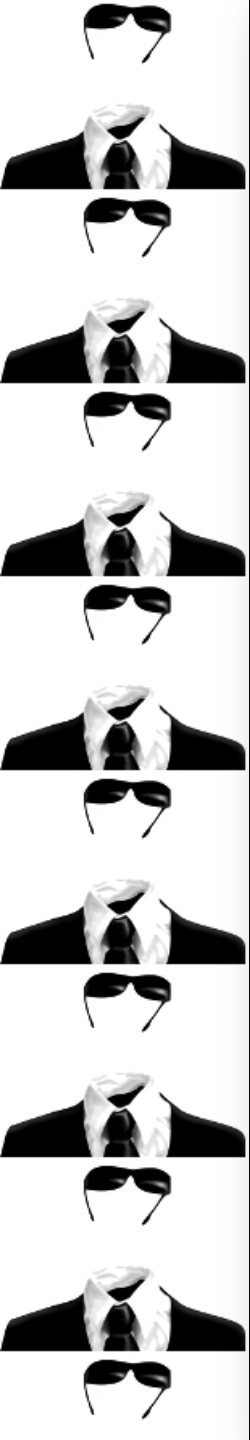For two texts *T* and *T1,* and for a set of *n* words,

$$\Delta(T, T_1) = \frac{1}{n} \sum_{x=1}^{n} \left| z(f_i(T)) - z(f_i(T_1)) \right|$$

Where $z(f_x(T)) = \dfrac{f_x(T) - \mu_x}{\sigma_x};$

$f_x(T)$ = raw frequency of word *x* in text T;

μ_x = mean frequency of word *x* in a collection of texts;
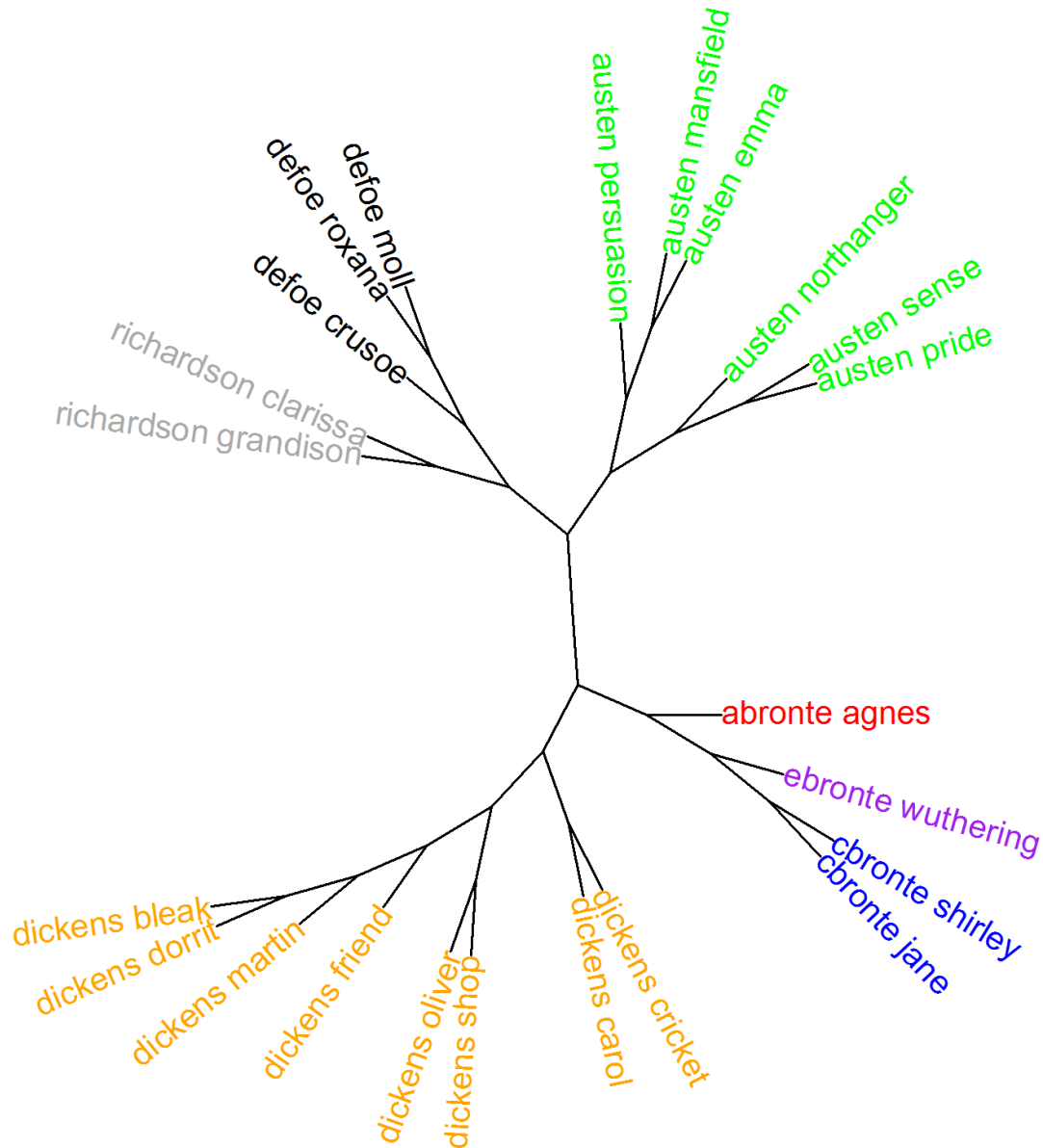
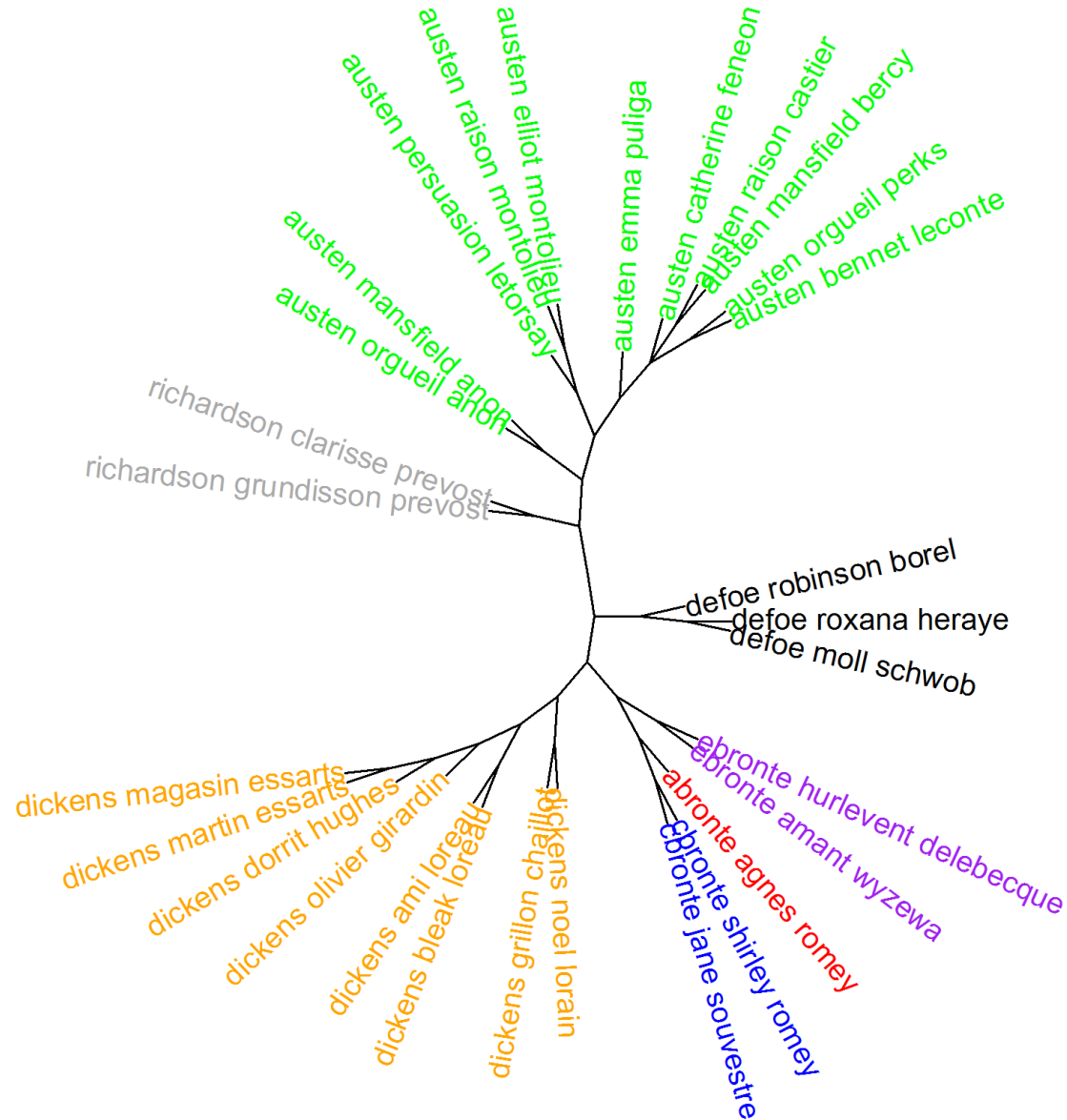σ_x = standard deviation of frequency of word *x*.

6

The 100 most frequent words

THE TO AND OF I A IN THAT IT YOU HER HE WAS MY AS HIS FOR NOT SHE WITH BE HAD HAVE ME BUT IS AT HIM THIS SO WHICH ON BY ALL SAID WOULD IF WILL YOUR FROM WHAT MR NO OR WHEN THEY ONE AN WERE BEEN WHO VERY THERE ARE UPON DO MORE COULD THEM WE SHOULD NOW THAN ANY OUT SUCH AM THEIR MAN THEN UP LITTLE CAN MUCH HAS INTO MUST MISS KNOW OWN SOME MAY THINK GOOD LADY HOW WELL SAY MRS NEVER SEE TIME OTHER BEFORE DID SHALL MADE ABOUT
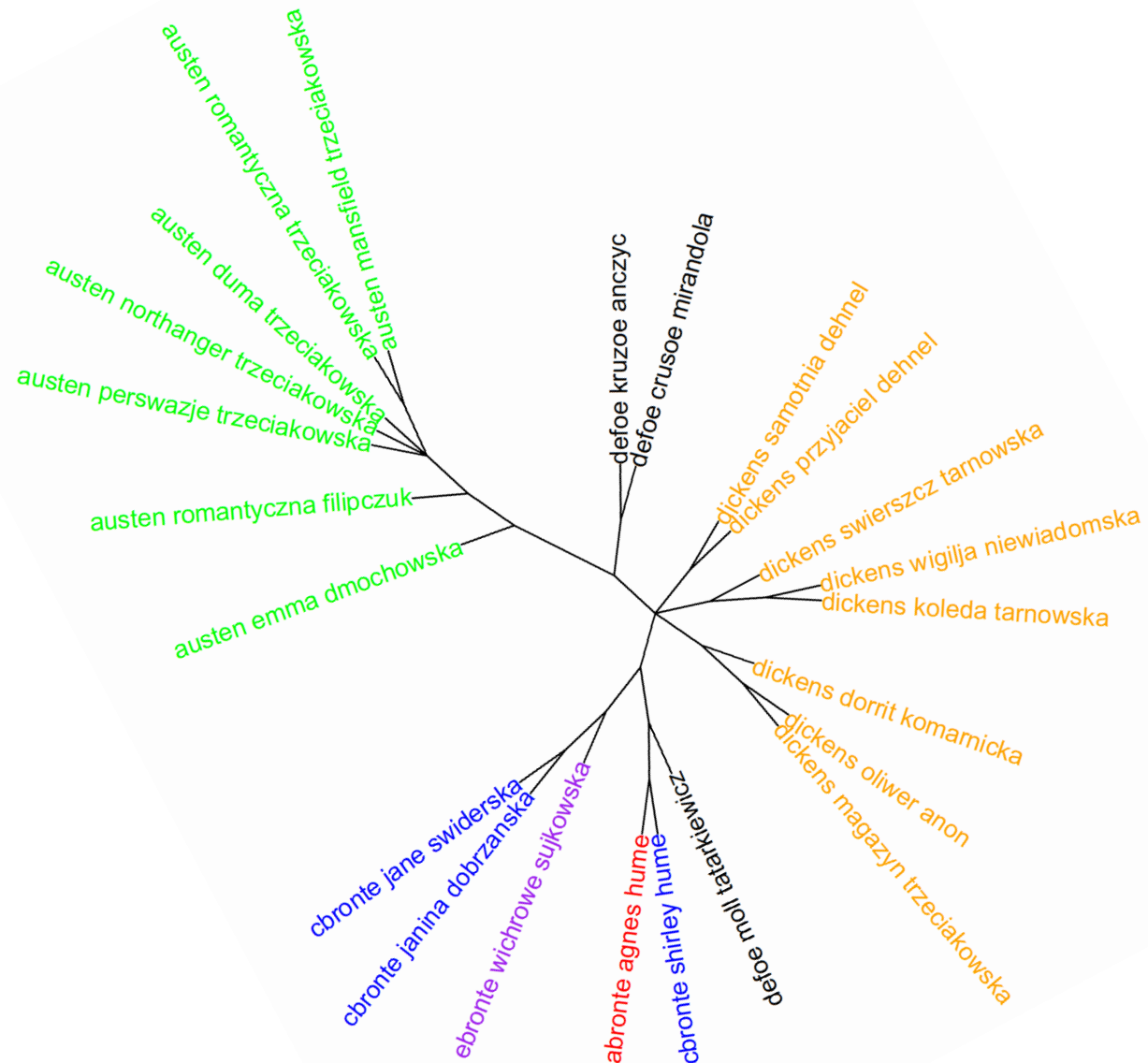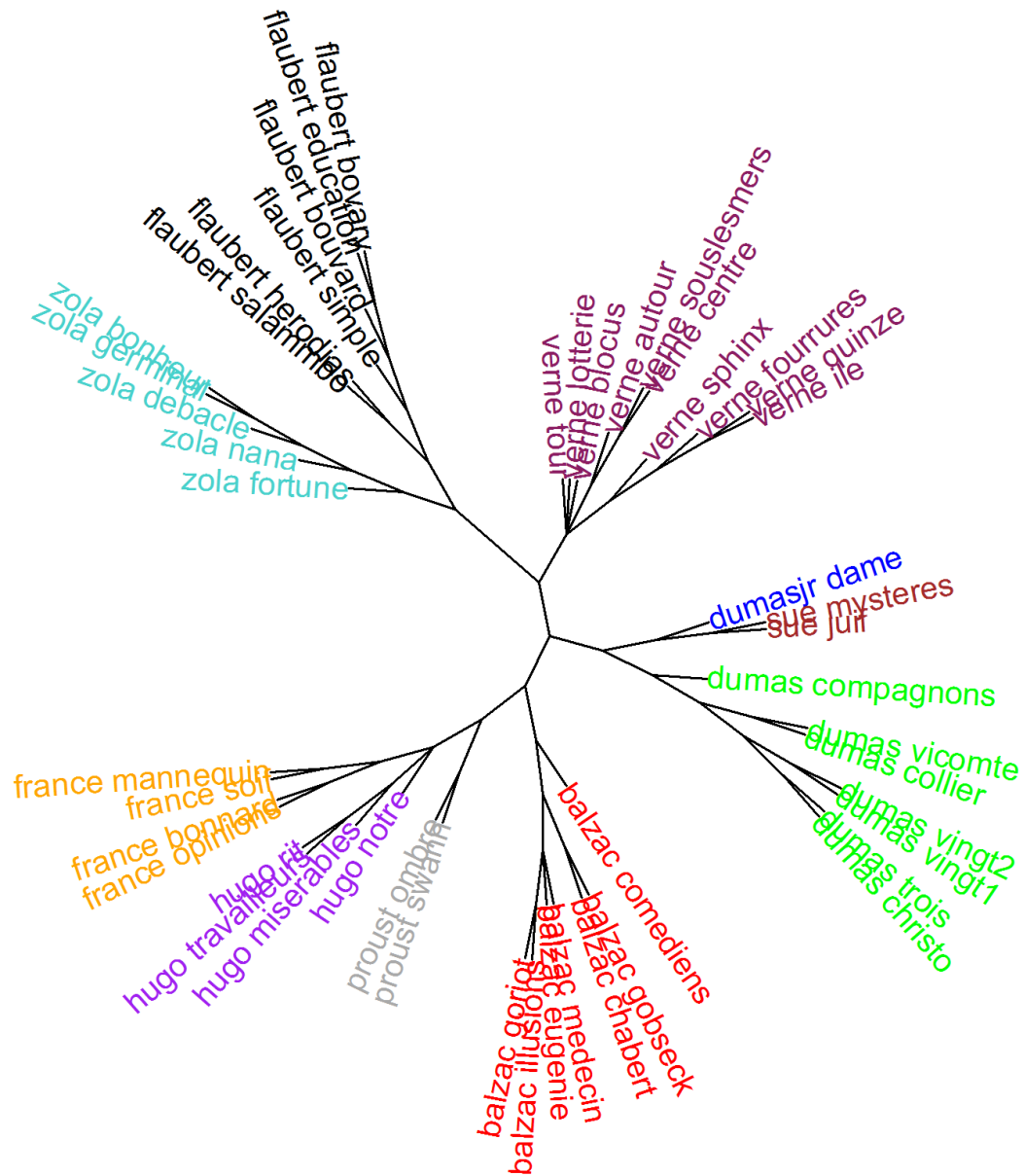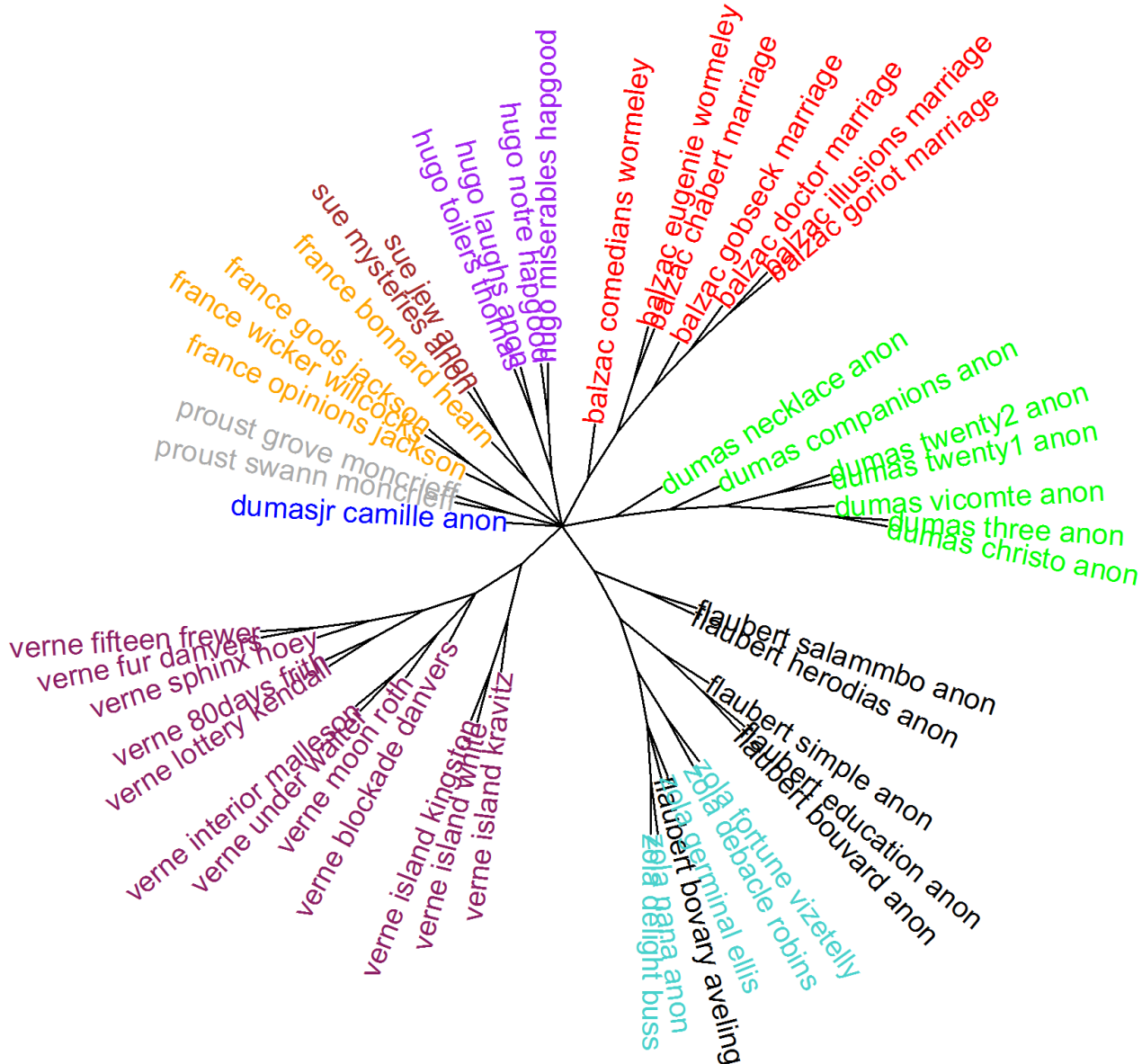
29 English novels…

…their 29 French translations…

austen catherine feneon
austen elliot montolieu
austen raison montolieu
austen persuasion letorsay
austen emma puliga
austen raison castier
austen mansfield bercy
austen orgueil perks
austen bennet leconte
austen mansfield anon
austen orgueil anon

richardson clarisse prevost
richardson grundisson prevost

defoe robinson borel
defoe roxana heraye
defoe moll schwob

dickens magasin essarts
dickens martin essarts
dickens dorrit hughes
dickens olivier girardin
dickens ami loreau
dickens bleak loreau
dickens grillon chailly
dickens noel lorain

cbronte hurlevent delebecque
cbronte amant wyzewa
abronte agnes romey
cbronte shirley romey
cbronte jane souvestre

10

…and their 24 Polish translations

48 French novels…

12

...their 51 English translations

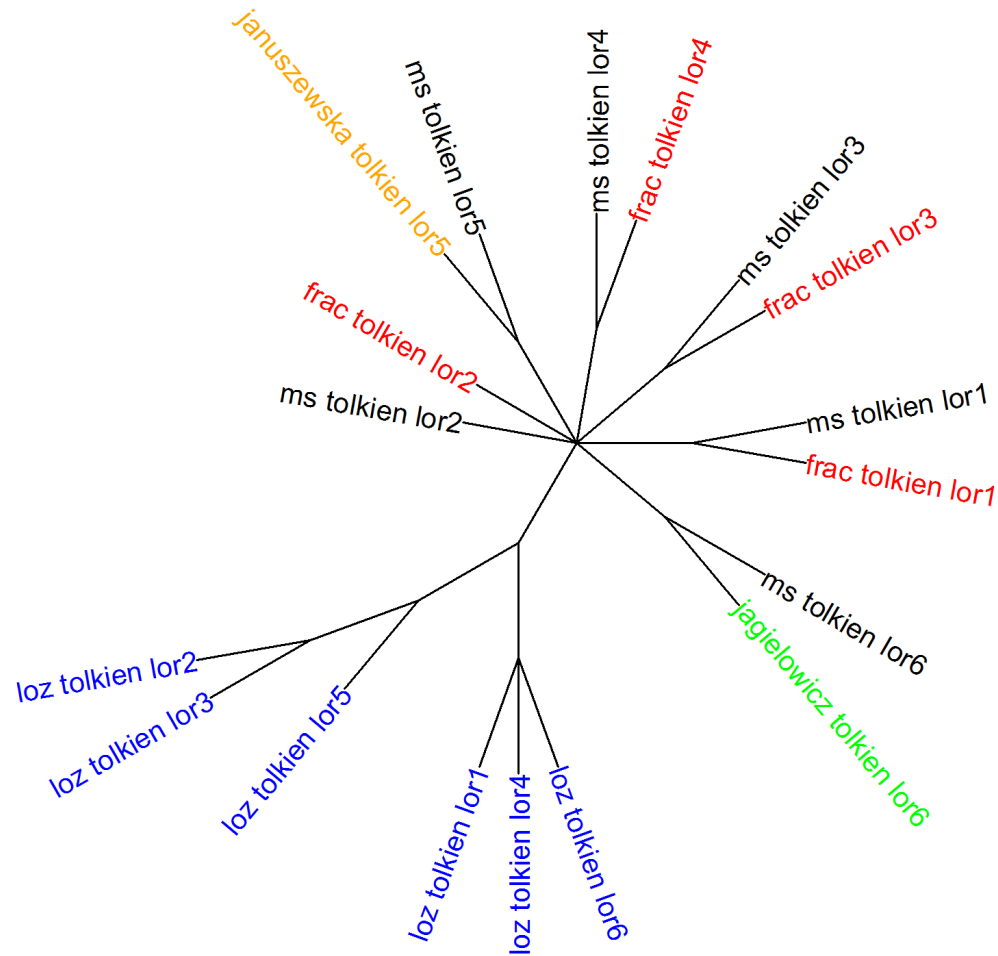# …and their 50 Polish translations
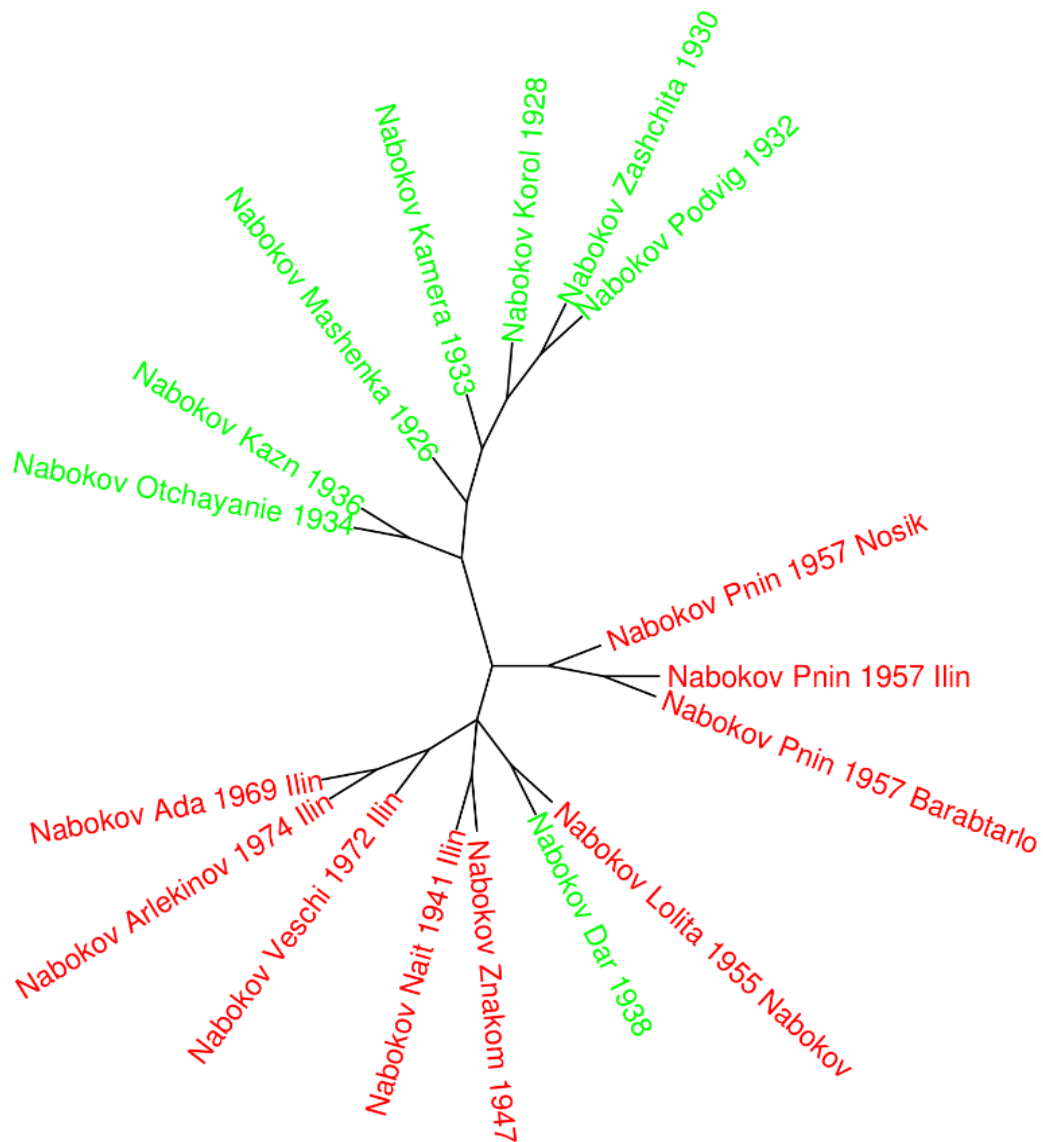
…and their 50 Polish translations
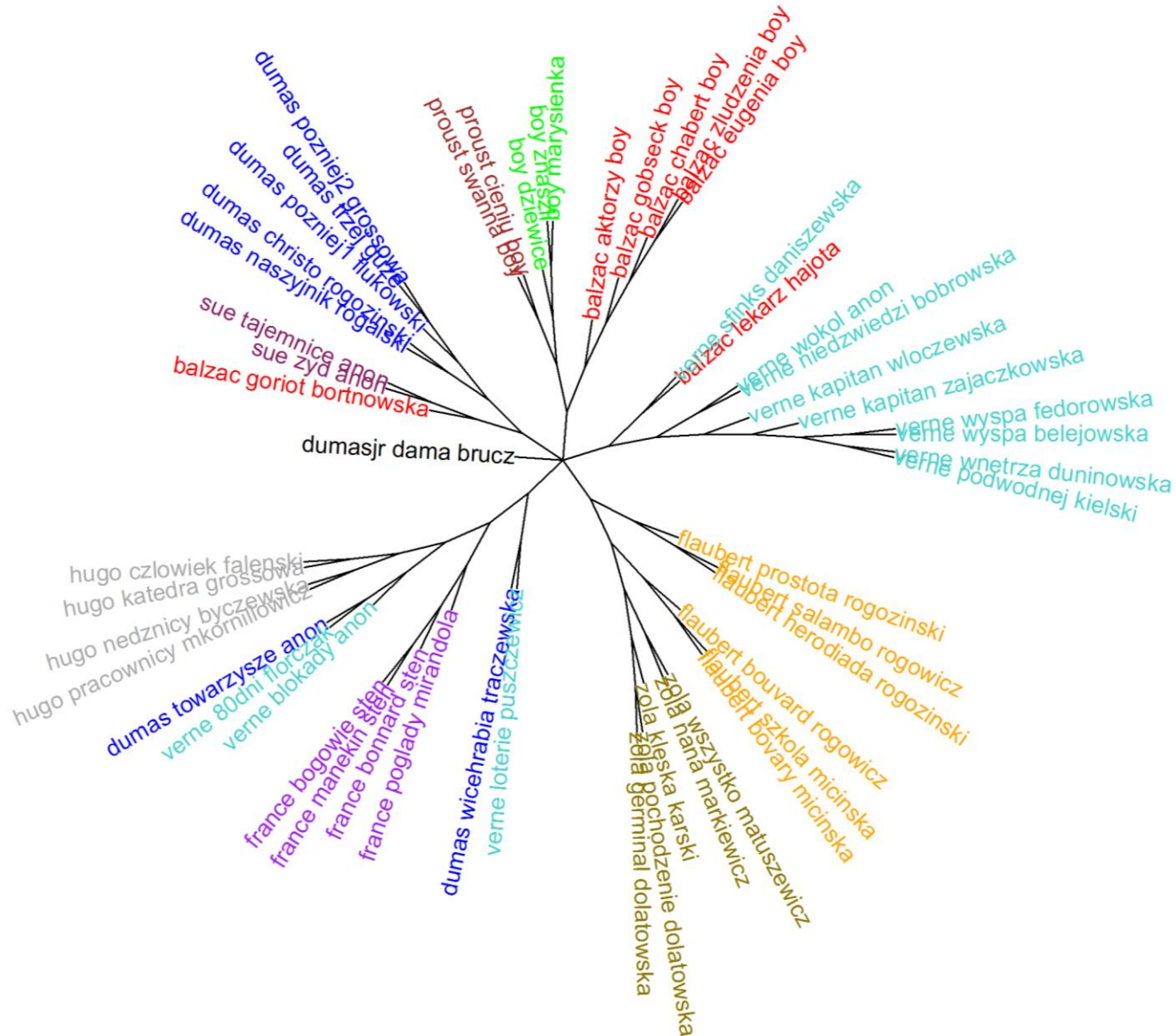
3 Polish translations
of Tolkien's *Lord of the Rings*
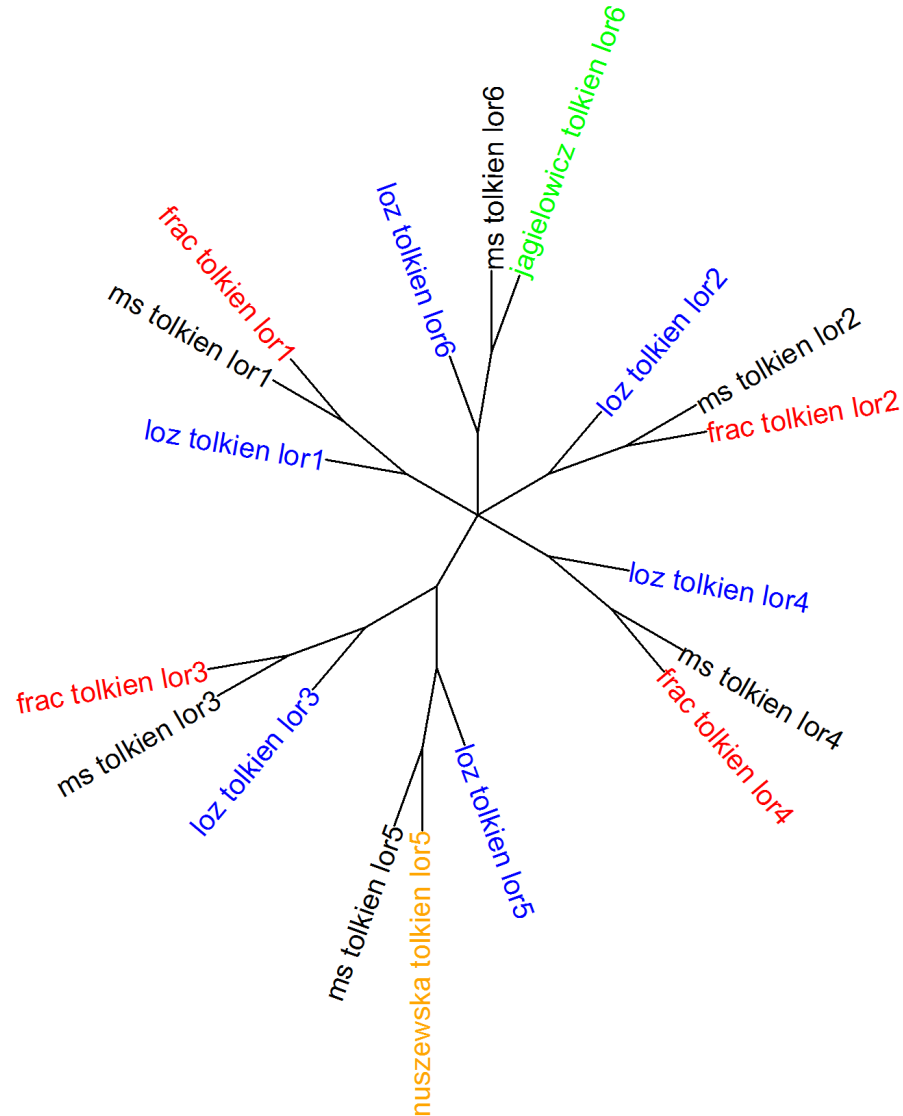
3 Polish translations
of Tolkien's *Lord of the Rings*

Nabokov in Russian

Virginia Woolf, *Noc i dzień*

- Original, 1919
- Polish translation, 2010
- Translators:
  - Anna Kołyszko (d. 2009)
    - Rushdie, H. Miller, Roth, Nabokov, Kerouac, Lessing…
  - Magda Heydel
    - Woolf, Conrad, G. Swift, Heaney, Walcott…
- Collaboration:
  - Some two thirds translated by Anna Kołyszko
  - Magda Heydel translated the rest and edited the entire text

Virginia Woolf, *Noc i dzień*

Heydel:

It is for the readers to see whether there is or there is not a rift in the middle of the book where one translator took over from another. I hope my editing made the narration coherent as far as the style was concerned. It was a truly unique experience to confront my own intuition of the authorial voice in the text with that of another translator.
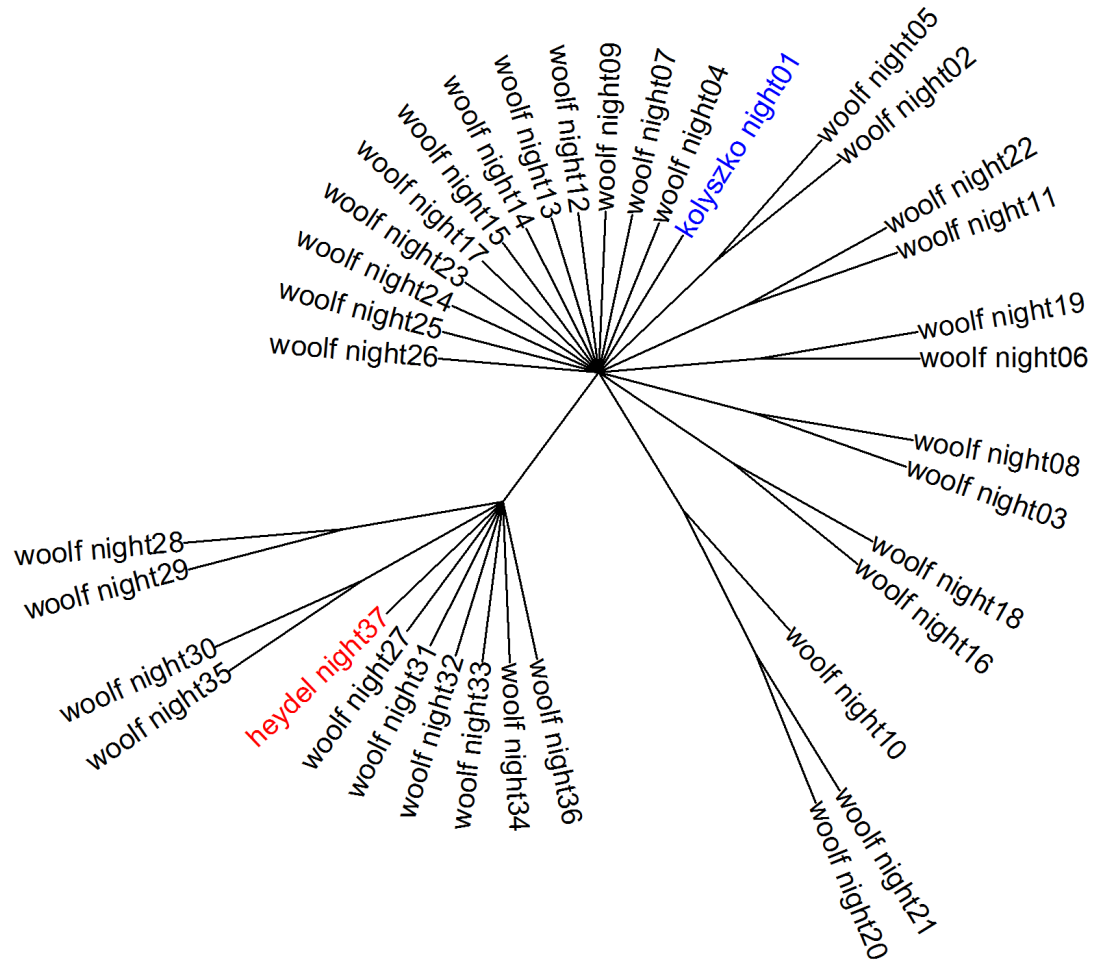
20

Noc i dzień:
the experiment

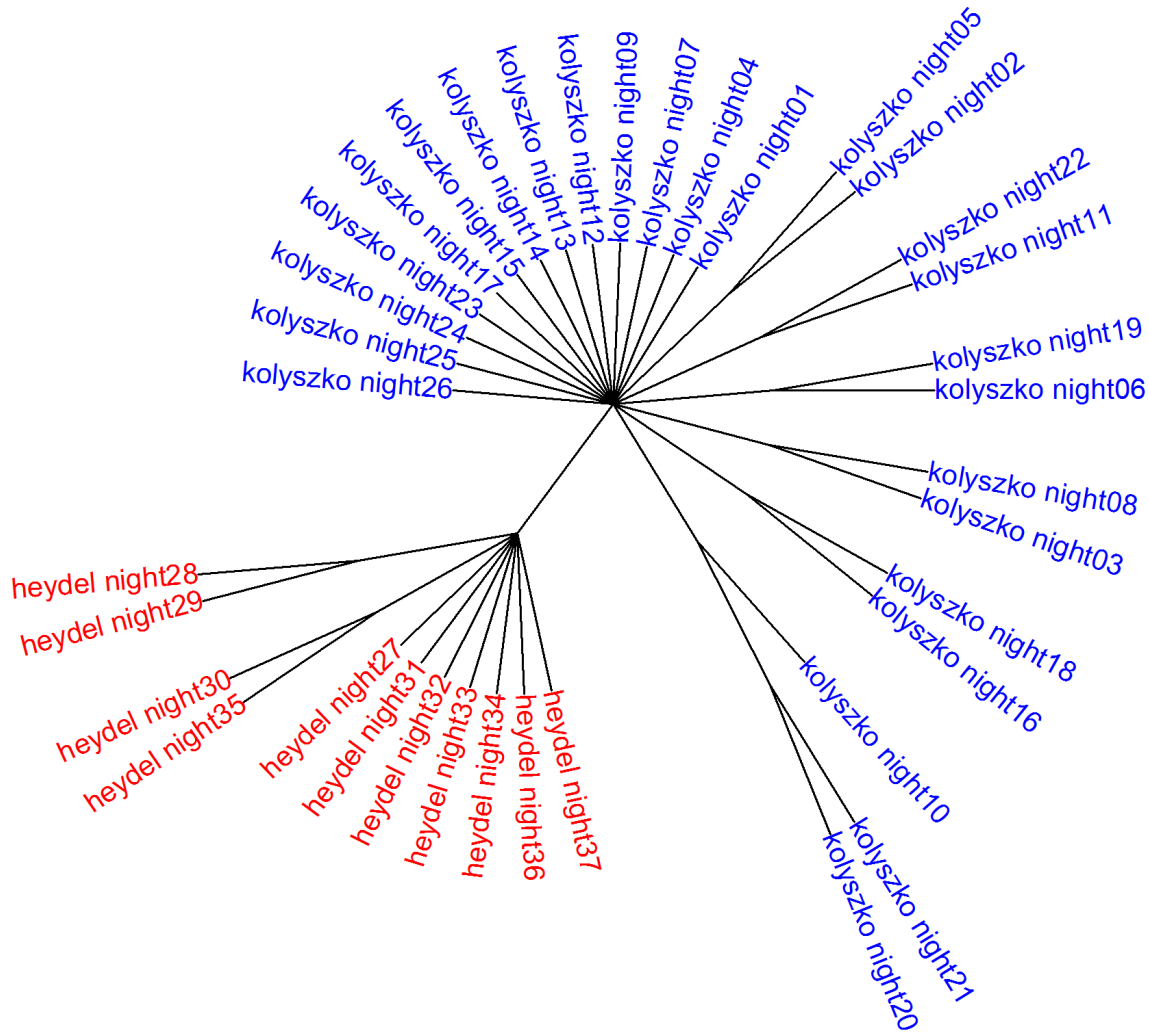Text divided into 37 equal sections (ca. 5000 word tokens)

Known:

- First section by Anna Kołyszko (kolyszko_night_01)
- Final section by Magda Heydel (kolyszko_night_37)

Other sections pretended unknown (woolf_night_02-36)

Noc i dzień:
the experiment

Noc i dzień: the experiment

100-1000 MFWs, Culled @ 0-100 %
Classic Delta distance

Conclusions

- In a large collection of translations, it is easier to detect original authors than translators with multivariate analyses of word frequency…

- …which might seem strange, since the former have little one-on-one counterparts in word frequency lists for original and translation.

- Translators can be sometimes detected when translations of texts by the same author are compared.

Conclusions

- The good news for the reading public is that, stylometrically speaking, translators usually remain more or less invisible.

- Some translators might be less enthusiastic.

Literature

1. Shlomo Argamon, 'Interpreting Burrows's Delta: Geometric and Probabilistic Foundations,' *Literary and Linguistic Computing* (2008) 23(2): 131-147
2. Mona Baker, 'Towards a methodology for investigating the style of a literary translator,' *Target* (2000) 12(2): 241-266
3. John F. Burrows, 'Delta' a measure of stylistic difference and a guide to likely authorship,' *Literary and Linguistic Computing* (2002) 17: 267-287
4. John F. Burrows, 'The Englishing of Juvenal: computational stylistics and translated texts,' Style (2002) 36: 677-699
5. Maciej Eder & J. Rybicki, *Stylometry with R* (Stanford: Digital Humanities conference, 2011), 308-311
6. Maciej Eder & J. Rybicki, Computational Stylistics, https://sites.google.com/site/computationalstylistics/
7. Magda Heydel, Wywiad w 'Czytelnia,' *TVP Kultura*, 12.02.2011
8. David Hoover, *Delta, Delta Prime, and modern American poetry: Authorship attribution theory and method* (Victoria: ALLC/ACH Conference, 2005)
9. David Hoover, 'Quantitative Analysis and Literary Studies,' in Ray Siemens, Susan Schreibman (eds) *A Companion to Digital Literary Studies* (Oxford: Blackwell, 2008), 517-33

Literature

10. Jan Rybicki, *A Computer-Assisted Analysis of Two English Translations of Henryk Sienkiewicz's Trilogy* (Kraków: Akademia Pedagogiczna, 2000)

11. Jan Rybicki, *Original, Translation, Inflation. Are All Translations Longer than Their Originals* (London: *Digital Humanities 2010*, 363-364)

12. Jan Rybicki, 'The Great Mystery of the (Almost) Invisible Translator,' in Michael P. Oakes, Meng Ji, *Quantitative Methods in Corpus-based Translation Studies*. (Amsterdam/Philadelphia: John Benjamins, 2012).

13. Peter Smith, & W. Aldridge, 'Improving Authorship Attribution: Optimizing Burrows's Delta Method,' *Journal of Quantitative Linguistics* (2011) 18(1): 63-88

14. Lawrence Venuti, *The Translator's Invisibility: A History of Translation*. (London: Routledge, 1995)

15. Brian Vickers, 'Shakespeare and Authorship Studies in the 21st Century,' *Shakespeare Quarterly* (2011) 62(1): 106-142