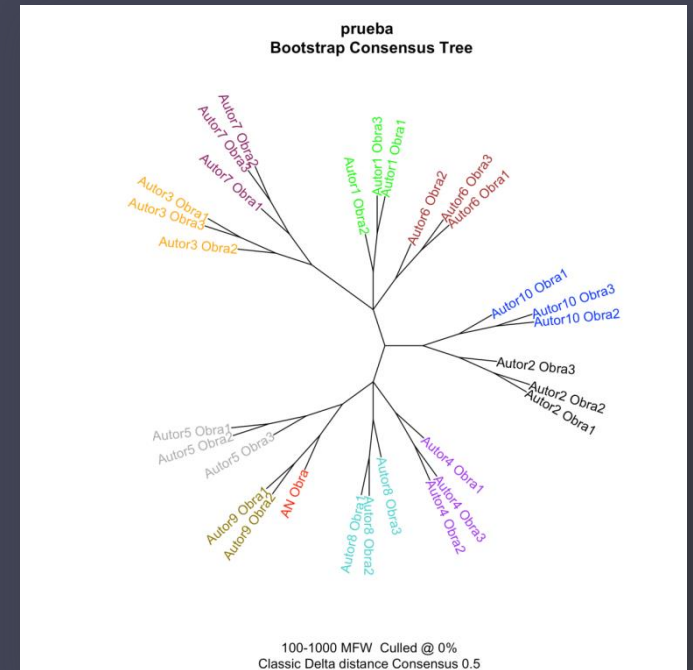


# Profundización en stylo. Verificación de autoría



*Laura Hernández Lorenzo*

Humanidades Digitales. Del corpus a la interpretación:  
Estilometría con R (Universidad de Burgos)

# Contenidos

- Atribución / verificación de autoría
- Verificación de autoría
- General Imposters Framework
- Práctica 6

# Métodos de atribución de autoría

- Se intenta encontrar el autor de un texto anónimo dentro de un conjunto de candidatos
- Es decir:
  - *De todos estos autores, ¿quién es el autor más probable de la obra anónima?*
  - *De todos estos autores, ¿cuál es el más cercano estilísticamente al texto anónimo?*

# Métodos de atribución de autoría

- Problema: ¿y si el autor del texto anónimo no está entre los candidatos que se han recogido para el análisis?
  - Problema del vecino más cercano → siempre habrá un candidato estilísticamente más cercano al texto anónimo

# Métodos de verificación de autoría

- Se trata de confirmar o rechazar si el principal candidato o sospechoso de haber escrito el texto anónimo es efectivamente el autor
- Se supera el problema del ‘vecino más cercano’

# Métodos de verificación de autoría

- Destaca el marco de los Impostores -‘General Imposters (GI) framework’-
  - Introducido por primera vez en Estilometría y problemas de autoría por Koppel y Winter (2014).

## Determining If Two Documents Are Written by the Same Author

Moshe Koppel and Yaron Winter

Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel 52900.  
E-mail: {moishk, yaron.winter}@gmail.com

Almost any conceivable authorship attribution problem can be reduced to one fundamental problem: whether a pair of (possibly short) documents were written by the same author. In this article, we offer an (almost) unsupervised method for solving this problem with surprisingly high accuracy. The main idea is to use repeated feature subsampling methods to determine if one document of the pair allows us to select the other from among a background set of “impostors” in a sufficiently robust manner.

problem, regardless of the number of candidates. All we need to do is ask if the anonymous text was written by each of the respective candidates; we will get a positive answer for the true author and a negative answer for all the others. On the other hand, the verification problem is strictly harder than the attribution problem: the fact that we solve a closed-set attribution problem offers no guarantees that we can solve an open-set verification problem. It is thus not surprising that, with a single limited exception (see below), no satisfactory solution has


# Impostores

- En estudios literarios, destaca su aplicación a los problemas autoriales con las obras de Julio César (Kestemont et al., 2016).

Expert Systems With Applications 63 (2016) 86–96


---

Contents lists available at ScienceDirect

 Expert Systems With Applications

journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)

---




---

## Authenticating the writings of Julius Caesar

Mike Kestemont<sup>a,\*</sup>, Justin Stover<sup>b</sup>, Moshe Koppel<sup>c</sup>, Folgert Karsdorp<sup>d</sup>, Walter Daelemans<sup>e</sup>

<sup>a</sup> University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium  
<sup>b</sup> University of Oxford, All Souls College, Oxford OX1 4AL, United Kingdom  
<sup>c</sup> Bar-Ilan University, 52900 Ramat-Gan, Israel  
<sup>d</sup> Center for Language Studies, Radboud University, P.O. Box 9103, NL-6500 HD, Nijmegen, The Netherlands  
<sup>e</sup> University of Antwerp, Belgium, Prinsstraat 13, B-2000 Antwerp, Belgium

 CrossMark

---

### ARTICLE INFO

Article history:  
Received 10 March 2016  
Revised 27 May 2016  
Accepted 14 June 2016

### ABSTRACT

In this paper, we shed new light on the authenticity of the *Corpus Caesarianum*, a group of five commentaries describing the campaigns of Julius Caesar (100–44 BC), the founder of the Roman empire. While Caesar himself has authored at least part of these commentaries, the authorship of the rest of the texts

# Impostores

- Implementado en mayo de 2018 en el paquete ‘stylo’ en R de Eder, Rybicki y Kestemont
  - <https://computationalstylistics.github.io/blog/imposters/>
- Valores resultantes entre 0 y 1
  - **0** significa que con toda seguridad **no** es el autor
  - **1** significa que con toda seguridad **sí** es el autor



# Práctica 6

- Abre en RStudio el script “taller7.R”
- ¡Manos a la obra!