

# Examining an author's individual grammar

---

Dr Andrea Nini

[andrea.nini@manchester.ac.uk](mailto:andrea.nini@manchester.ac.uk)

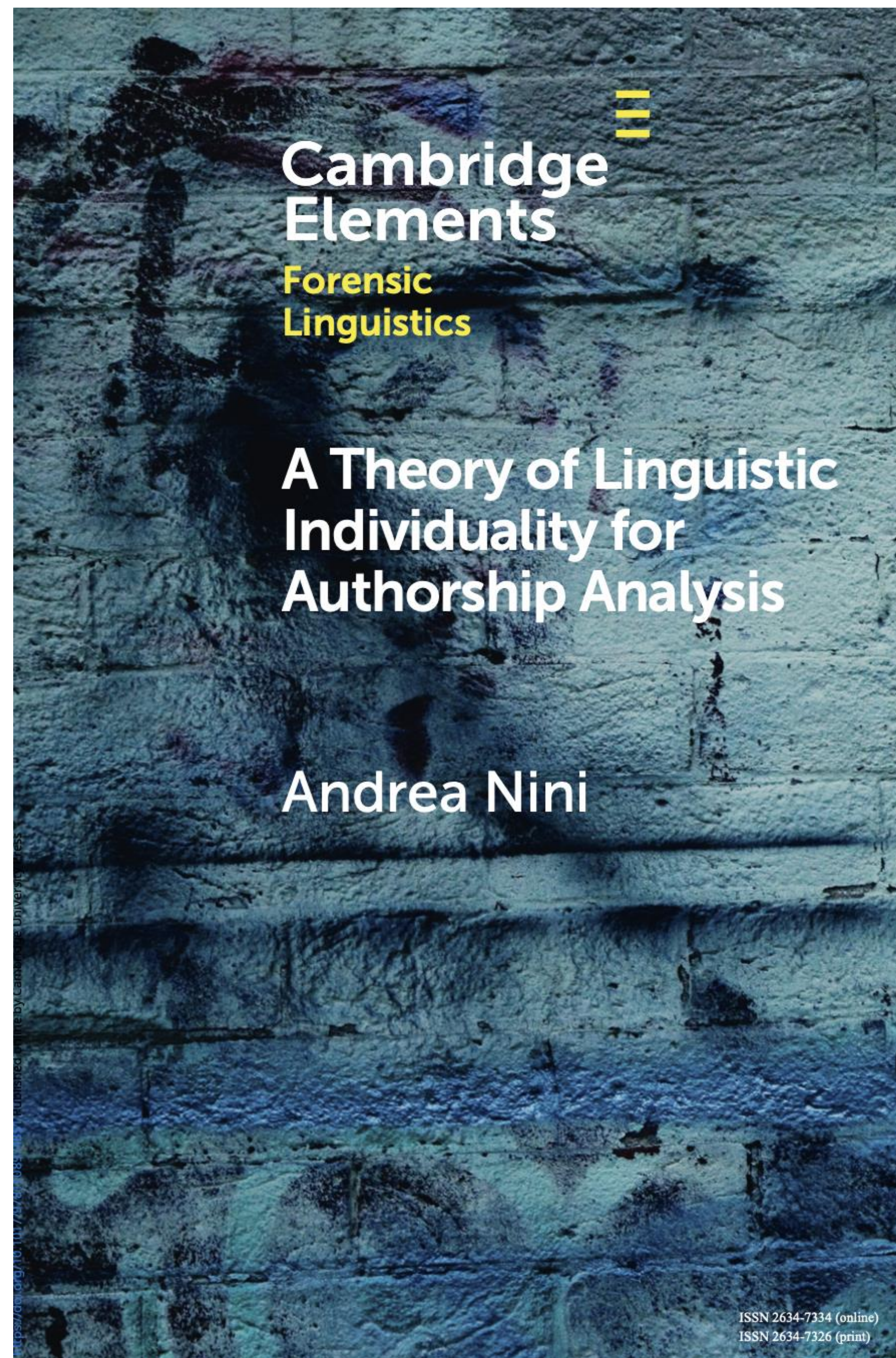
[www.andreanini.com](http://www.andreanini.com)

Comparative Literature Goes Digital Workshop

DH2025

14<sup>th</sup> July 2025





Free!



# Grammar as a Behavioral Biometric: Using Cognitively Motivated Grammar Models for Authorship Verification

**Andrea Nini** 

Linguistics and English Language  
University of Manchester  
Manchester, M13 9PL  
[Andrea.Nini@Manchester.ac.uk](mailto:Andrea.Nini@Manchester.ac.uk)

**Oren Halvani** 

Fraunhofer Institute for Secure Information Technology SIT  
Darmstadt, 64295  
[Oren.Halvani@SIT.Fraunhofer.de](mailto:Oren.Halvani@SIT.Fraunhofer.de)

**Lukas Graner** 

Fraunhofer Institute for Secure Information Technology SIT  
Darmstadt, 64295  
[Lukas.Graner@SIT.Fraunhofer.de](mailto:Lukas.Graner@SIT.Fraunhofer.de)

**Valerio Gherardi** 

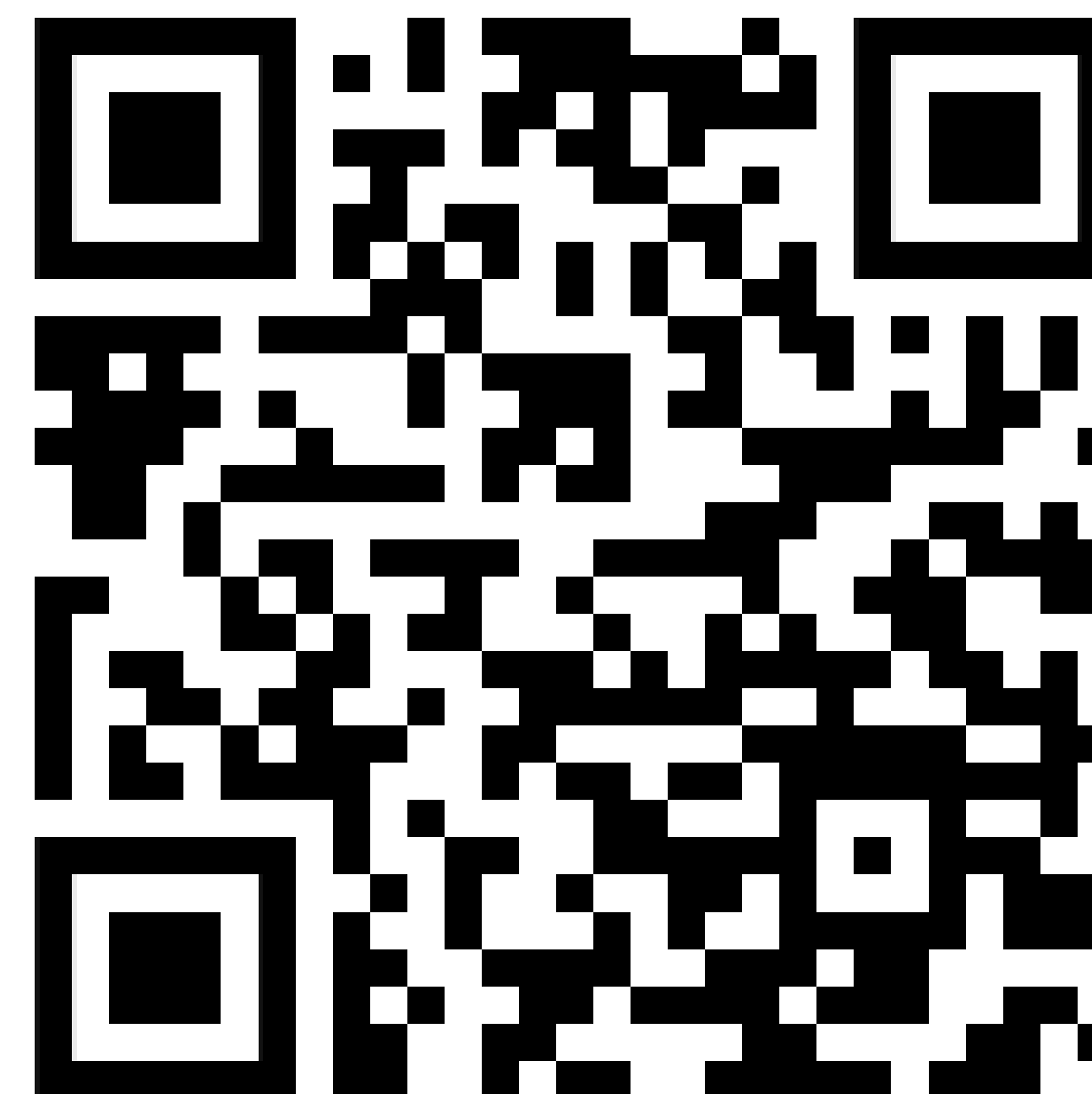
Cavanilles Institute of Biodiversity and Evolutionary Biology  
University of Valencia  
C/Catedrático José Beltrán 2, 46980 Paterna, Valencia, Spain  
[VGherard840@GMail.com](mailto:VGherard840@GMail.com)

**Shunichi Ishihara** 

Speech and Language Laboratory  
ANU College of Asia and the Pacific  
The Australian National University  
[Shunichi.Ishihara@ANU.edu.au](mailto:Shunichi.Ishihara@ANU.edu.au)

The Likelihood  
Ratio of  
Grammar  
Models

$\lambda_G$   
*LambdaG*



# Content masking

The cat sat on the mat and the dog sat on the chair.

The **N** **V** on the **N** and the **N** **V** on the **N** .

# Authorship verification with LambdaG

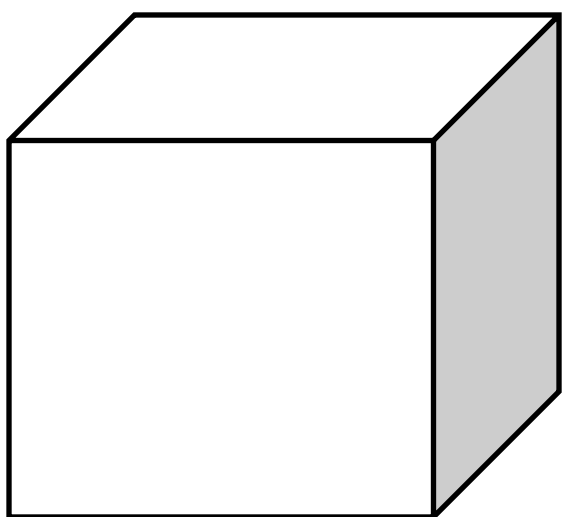
POS-tag and  
mask content

N-gram model  
training

Grammar  
model

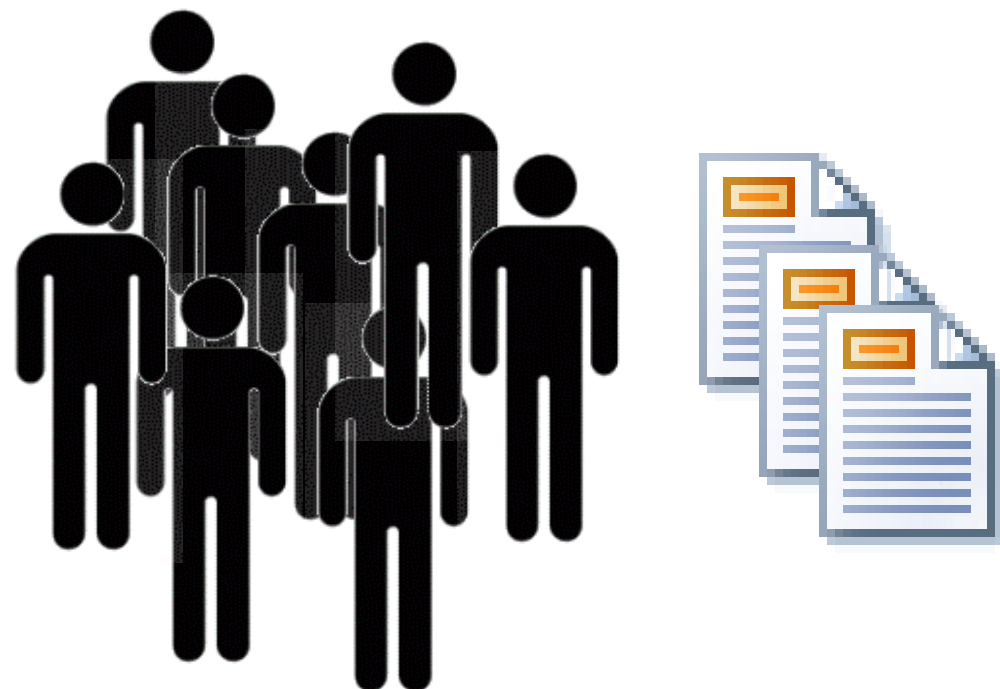
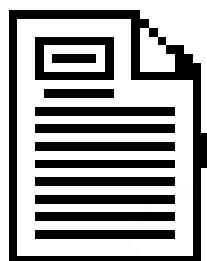


The N V on the N  
and the N V on  
the N .

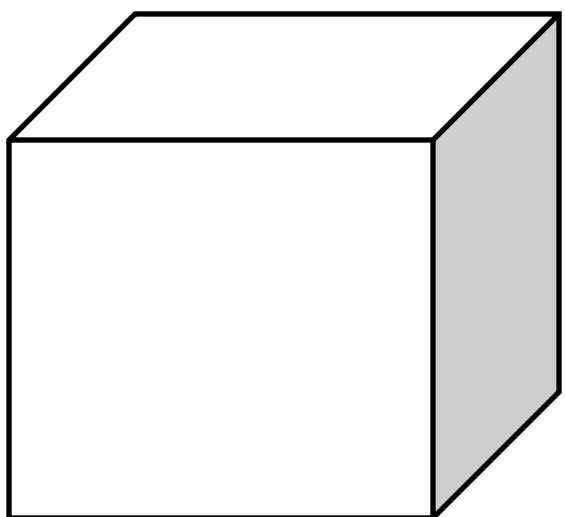


$G_A$

Disputed  
text



The N V on the N  
and the N V on  
the N .



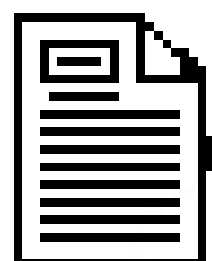
$G_r$

POS-tag and  
mask content

Grammar  
model

$G_A$

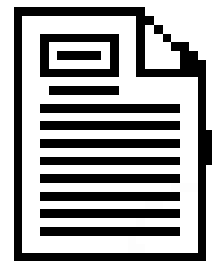
Disputed  
text



The **N** **V** on the **N**  
and the **N** **V** on  
the **N** .

$G_r$

Disputed  
text



The **N** **V** on the **N**  
and the **N** **V** on  
the **N** .

POS-tag and  
mask content



$P_A(\textit{disputed text})$

$P_r(\textit{disputed text})$

Grammar  
model

$G_A$

$G_r$

$$\begin{array}{l} \text{Likelihood} \\ \text{ratio of} \\ \text{Grammar} \\ \text{Models} \end{array} = \frac{P_A(\textit{disputed text})}{P_r(\textit{disputed text})}$$

$\lambda_G$



# Likelihood ratio of Grammar Models

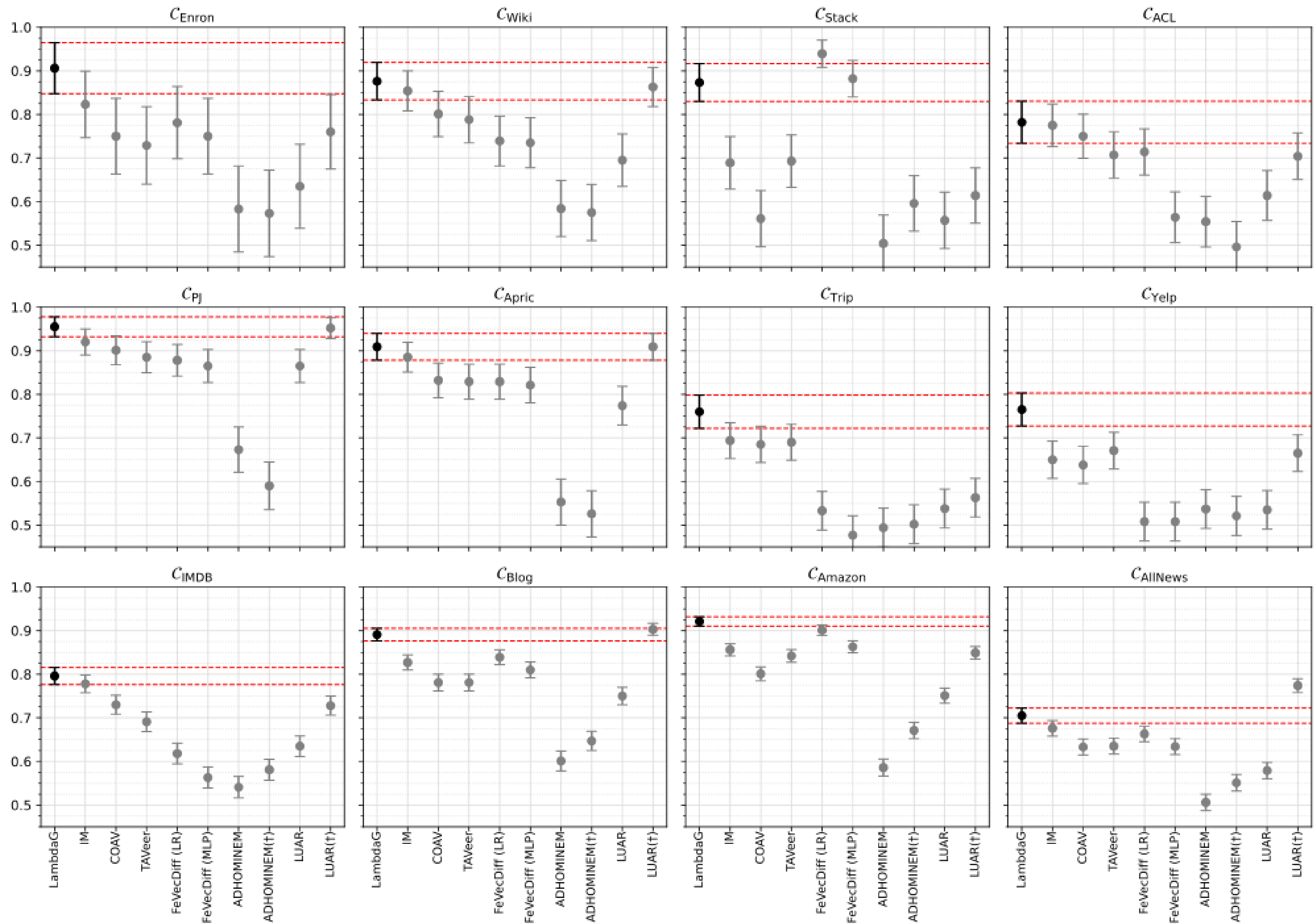
$\lambda_G$

12 corpora comprising: academic articles,  
emails, chat logs, forum posts, blog posts,  
reviews, news

$\lambda_G$  overall outperforms other methods

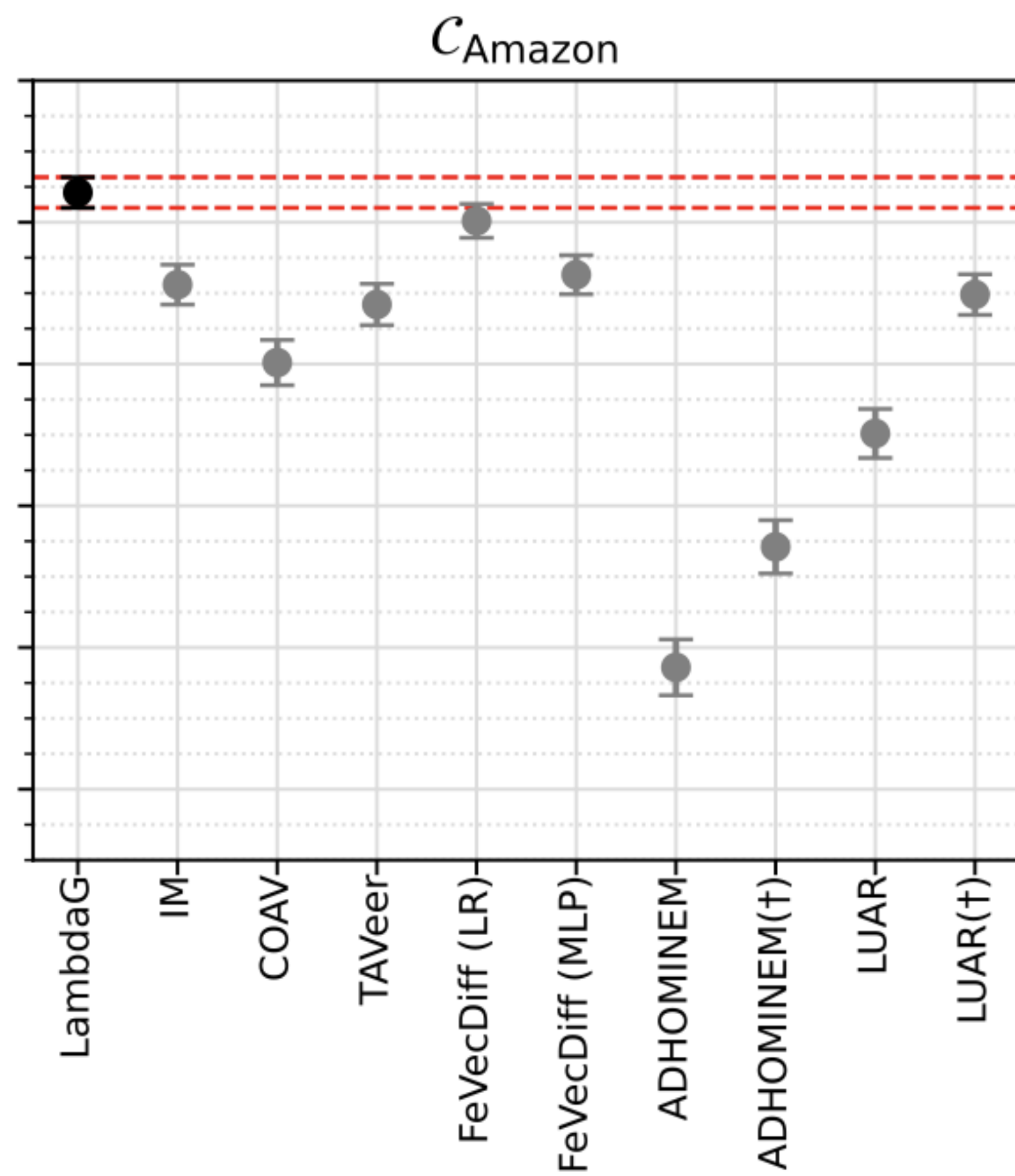
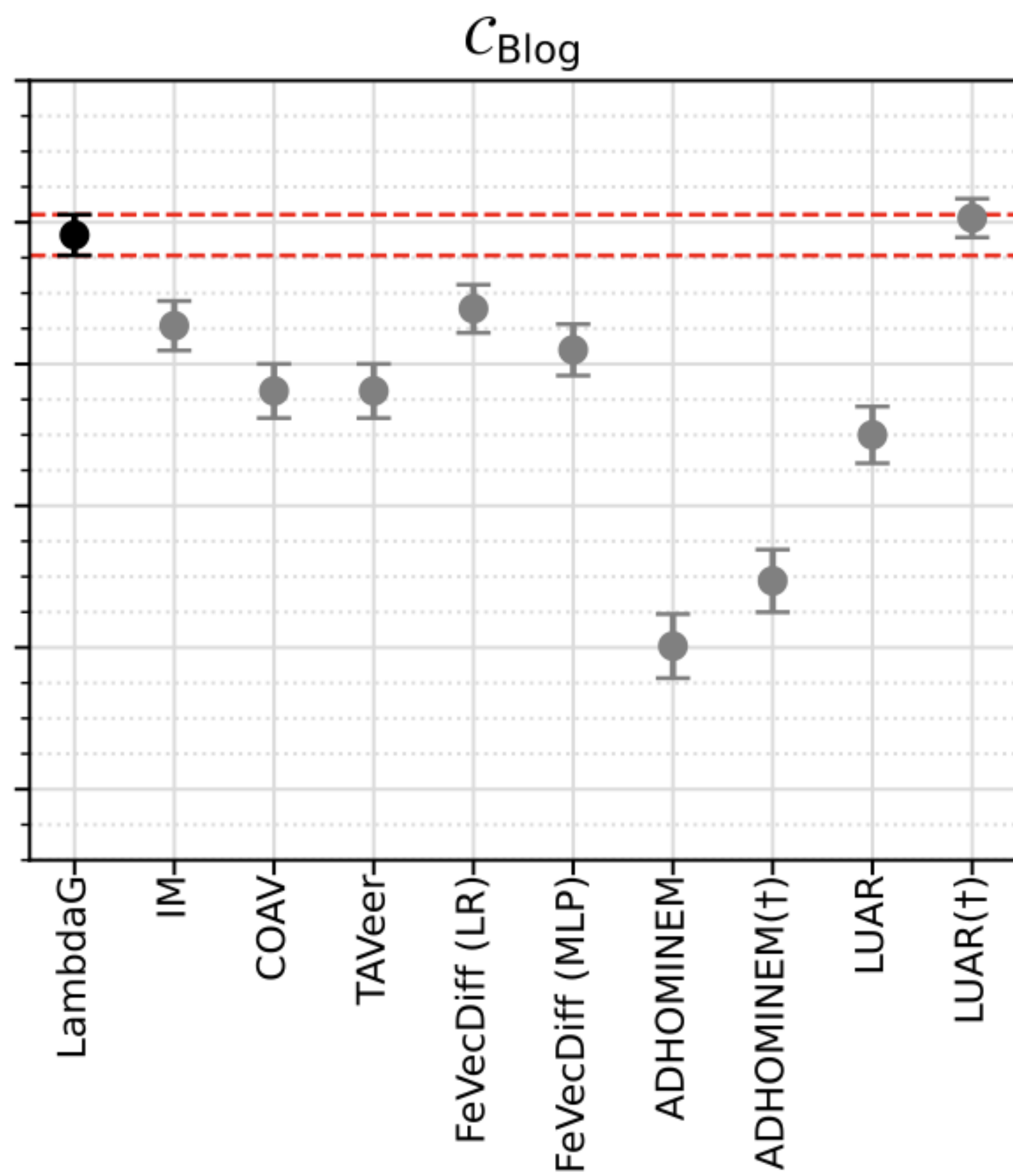
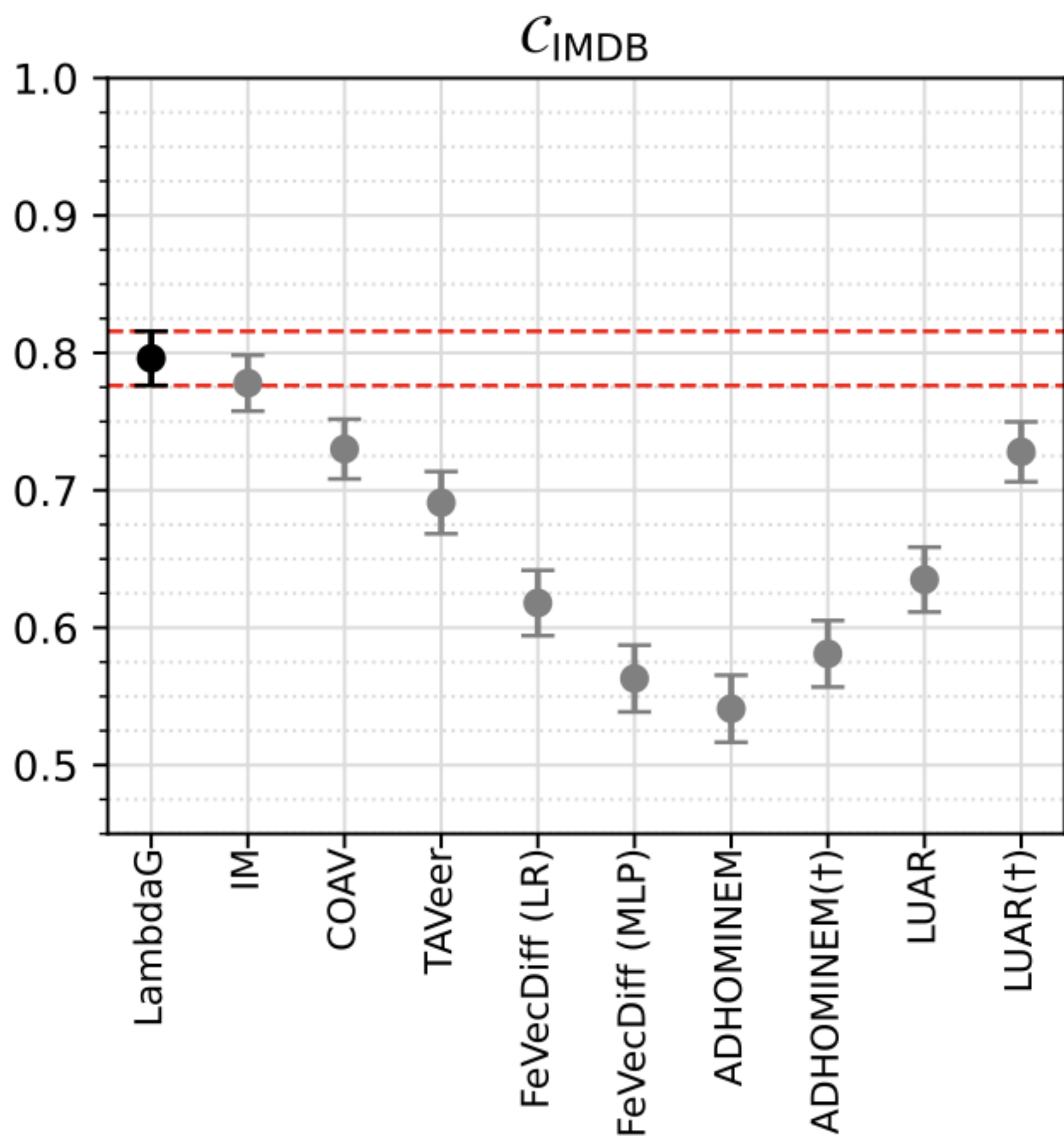
Corpus	Genre	Topic	Partition	$ \mathcal{C} $	$\text{avg}(\mathbb{D}_{\mathcal{A}})$	$\text{avg}(\mathcal{D}_{\mathcal{U}})$
$\mathcal{C}_{\text{Enron}}$	E-mails	Related	$\mathcal{C}_{\text{train}}$	64	783 (3866)	906 (4432)
			$\mathcal{C}_{\text{test}}$	96	763 (3808)	882 (4391)
$\mathcal{C}_{\text{Wiki}}$	Wikipedia talk pages	Related	$\mathcal{C}_{\text{train}}$	150	467 (1266)	614 (1654)
			$\mathcal{C}_{\text{test}}$	226	492 (1328)	626 (1699)
$\mathcal{C}_{\text{Stack}}$	Q & A posts	Cross	$\mathcal{C}_{\text{train}}$	150	2204 (11247)	1972 (9956)
			$\mathcal{C}_{\text{test}}$	228	2306 (11803)	2114 (10700)
$\mathcal{C}_{\text{ACL}}$	Scientific papers	Related	$\mathcal{C}_{\text{train}}$	186	1707 (9497)	2553 (14196)
			$\mathcal{C}_{\text{test}}$	280	1479 (8250)	2492 (13913)
$\mathcal{C}_{\text{PJ}}$	Chat logs	Related	$\mathcal{C}_{\text{train}}$	208	594 (2329)	752 (2946)
			$\mathcal{C}_{\text{test}}$	312	592 (2324)	754 (2967)
$\mathcal{C}_{\text{Apric}}$	Forum posts	Related	$\mathcal{C}_{\text{train}}$	228	796 (3900)	825 (4023)
			$\mathcal{C}_{\text{test}}$	340	799 (3921)	817 (4020)
$\mathcal{C}_{\text{Trip}}$	Travel reviews	Related	$\mathcal{C}_{\text{train}}$	120	658 (3169)	249 (1186)
			$\mathcal{C}_{\text{test}}$	480	698 (3344)	252 (1202)
$\mathcal{C}_{\text{Yelp}}$	Hotel/restaurant reviews	Related	$\mathcal{C}_{\text{train}}$	320	139 (637)	167 (767)
			$\mathcal{C}_{\text{test}}$	480	139 (640)	167 (768)
$\mathcal{C}_{\text{IMDB}}$	Movie reviews	Mixed	$\mathcal{C}_{\text{train}}$	400	658 (3151)	374 (1796)
			$\mathcal{C}_{\text{test}}$	1600	690 (3291)	366 (1753)
$\mathcal{C}_{\text{Blog}}$	Blogs posts	Mixed	$\mathcal{C}_{\text{train}}$	1200	742 (3416)	881 (4086)
			$\mathcal{C}_{\text{test}}$	1800	740 (3385)	882 (4061)
$\mathcal{C}_{\text{Amazon}}$	Product reviews	Mixed	$\mathcal{C}_{\text{train}}$	1600	861 (4010)	869 (4041)
			$\mathcal{C}_{\text{test}}$	2400	860 (4009)	868 (4041)
$\mathcal{C}_{\text{AllNews}}$	News articles	Mixed	$\mathcal{C}_{\text{train}}$	1776	675 (3629)	941 (5064)
			$\mathcal{C}_{\text{test}}$	2662	680 (3665)	942 (5071)

95% Confidence Interval for Accuracy



AV Method





# idiolect

---



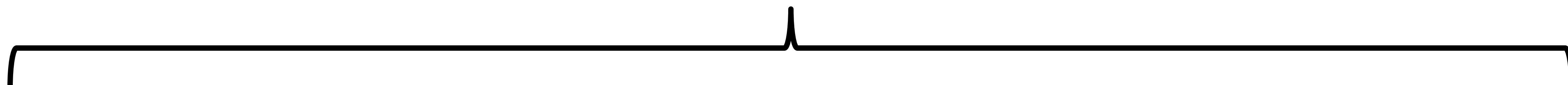
## Overview

---

The `idiolect` R package is designed to provide a comprehensive suite of tools for performing comparative authorship analysis within a forensic context using the likelihood ratio framework (e.g. Ishihara 2021). The package contains a set of authorship analysis functions that take a set of texts as input and output scores that can then be calibrated into likelihood ratios. The package is dependent on `quanteda` (Benoit et al. 2018) for all Natural Language Processing functions.



<https://andreanini.github.io/idiolect/>

$\Sigma$  $\lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G \lambda_G$ 

it is J enough to be B V in any J N . [EOS]

 $=$  $\lambda_G(\textit{text})$

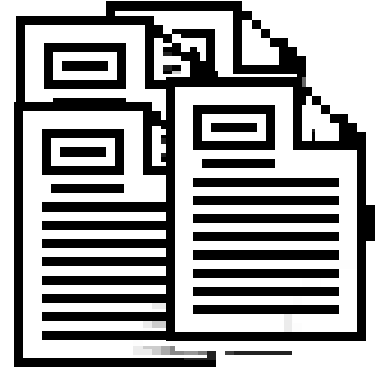


# Dickens case study

Full tutorial here: <https://github.com/andreanini/lambdaG-case-study-DH2025/>



# Dickens case study



75 novels (3 novels x 25 authors)  
*Project Gutenberg*

English *refcor* corpus

Dickens:

*Bleak House, Great  
Expectations, Oliver  
Twist*

13,057,183 words

Other authors:

Barclay, Blackmore, Braddon, Burnett,  
C. Bronte, Chesterton, Collins, Corelli,  
Doyle, Eliot, Forster, Gaskell, Gissing,  
Haggard, Hardy, James, Kipling, Lytton,  
Meredith, Morris, Stevenson, Thackeray,  
Trollope, Ward

i could not have V in such N without seeing it .  
my P , my P , i know it would , ' said i to him , ' and it would have been B J for me never to have heard the N of your J N , but  
B for me , i ca n't V the past , and the past V me here !  
i do n't know what may V if i am V beyond myself at last !  
P , ' the P P told me last week , ' not to V your N here , and to V , B V , down in P .  
to the last , i 'll show myself in that N to its N .  
P said that he V with him with all his N and that he set up no N himself in being B V by this J N .  
now , P .  
his N had , perhaps for N , become so set in its J N that it did not V , even now when he was J .  
my P knows nothing of it .  
it would be J for me , they tell me , if i V myself .  
P , in your N there are N and N V , where in mine there are N .  
P with no N of his N .  
it first came on after two N .  
you 're not J of me , P , are  
i am told on all N , it 's the N .  
i have done !  
now V me out N B !  
i ought to know it .

random 1,000 word sample from Dicken's *Bleak House*



i could not have V in such N without seeing it .

dickens_bleak.txt	fearful . I could not have believed in such rage	without seeing it	. " I have done ! " he said ,
dickens_expectations.txt	my brain , I could not think of a place	without seeing it	, or of persons without seeing them . It is
james_ambassadors.txt	friend replied , " that she can't effectually recognise it	without seeing it	for herself . Let Chad go over and show her

i could not have V in such N without seeing it .

dickens_bleak.txt	fearful . <u>I could not</u> have believed in such rage	without seeing it	. " I have done ! " he said ,
dickens_expectations.txt	my brain , <u>I could not</u> think of a place	without seeing it	, or of persons without seeing them . It is
james_ambassadors.txt	friend replied , " that she can't effectually recognise it	without seeing it	for herself . Let Chad go over and show her

my P , my P , i know it would , ' said i to him , ' and it would have been B  
J for me never to have heard the N of your J N , but B for me , i ca n't V the past , and the  
past V me here !

dickens_bleak.txt	' and it would have been far better for me	never to have	heard the name of your high office , but unhappily
dickens_expectations.txt	it would have been so much the better for me	never to have	entered , never to have seen . Chapter XLIV In
dickens_expectations.txt	much the better for me never to have entered ,	never to have	seen . Chapter XLIV In the room where the dressing-table
burnett_garden.txt	two worst things as can happen to a child is	never to have	his own way - - or always to have it
meredith_richmond.txt	, be the most unfortunate of human beings , than	never to have	aimed at a star . You laugh at me ?
lytton_what.txt	to be repaid to you . " " Am I	never to have	the Royal York Theatre ? Ambition of my life ,
gaskell_wives.txt	general frankness to have overheard what you did , and	never to have	mentioned it to me ? I could have told you



my P , my P , i know it would , ' said i to him , ' and it would have been B  
J for me never to have heard the N of your J N , but B for me , i ca n't V the past , and the  
past V me here !

dickens_bleak.txt	' and <u>it would have been</u> far <u>better for me</u>	never to have	heard the name of your high office , but unhappily
dickens_expectations.txt	<u>it would have been</u> so much the <u>better for me</u>	never to have	entered , never to have seen . Chapter XLIV In
dickens_expectations.txt	much the better for me never to have entered ,	never to have	seen . Chapter XLIV In the room where the dressing-table
burnett_garden.txt	two worst things as can happen to a child is	never to have	his own way - - or always to have it
meredith_richmond.txt	, be the most unfortunate of human beings , than	never to have	aimed at a star . You laugh at me ?
lytton_what.txt	to be repaid to you . " " Am I	never to have	the Royal York Theatre ? Ambition of my life ,
gaskell_wives.txt	general frankness to have overheard what you did , and	never to have	mentioned it to me ? I could have told you

i do n't know what may V if i am V beyond myself at last !

dickens_bleak.txt	, here - - I may ! I don't know	what may	happen if I am carried beyond myself at last !
dickens_expectations.txt	ever call you mine , Estella . I am ignorant	what may	become of me very soon , how poor I may
dickens_expectations.txt	lay hold of his portable property . You don't know	what may	happen to him . Don't let anything happen to the
dickens_oliver.txt	it's found out , ' said Monks . ' And	what may	that be ? ' asked the matron . ' The
dickens_oliver.txt	woman with much deliberation . ' The first is ,	what may	the communication be worth ? ' ' Who the devil
james_ambassadors.txt	She gave a strange light laugh . " Putting aside	what may	really hurt HIM . " But her friend looked at
stevenson_arrow.txt	" said the priest , somewhat recovering , " and	what may	this be ? What enemy hath done this ? "
stevenson_arrow.txt	I , Richard Shelton , tide what betideth , come	what may	, will see you safe in Hollywood . The saints
stevenson_arrow.txt	tall chimney marked the ruins of a house . "	What may	this be ? " whispered Matcham . " Nay ,
stevenson_arrow.txt	he cried again , " I like that little .	What may	this betoken ? Let us go , by the mass
collins_basil.txt	into this hotel a little while ago . " "	What may	your business be ? " - - He hesitated ,
collins_basil.txt	added in an altered tone , " I mean ,	what may	you want with them , Sir ? " " I
kipling_captains.txt	It's all new to you , but we never know	what may	come to us . Now , take me - -
eliot_daniel.txt	scruple , having a conscience of what must be and	what may	be ; whereas Ignorance is a blind giant who ,
corelli_innocent.txt	! " she said . " One never can tell	what may	happen in this strange world ! But we have to
hardy_jude.txt	beside them . " Well , my lad , and	what may	you want up here ? " " I wanted to
lytton_kenelm.txt	who had saved from the general wreck of their fortunes	what may	be called a good squire's estate , - - about
doyle_lost.txt	" Well , " he cried , expectantly , "	what may	it run to ? I'm thinking , young man ,
gaskell_lovers.txt	I beg your pardon , miss ; I was thinking	what may	the price of yon crimson ribbon be ? ' Hester

i do n't know what may V if i am V beyond myself at last !

dickens_bleak.txt	, here - - I may ! I <u>don't know</u>	what may	<u>happen</u> if I am carried beyond myself at last !
dickens_expectations.txt	ever call you mine , Estella . <u>I am ignorant</u>	what may	become of me very soon , how poor I may
dickens_expectations.txt	lay hold of his portable property . You <u>don't know</u>	what may	<u>happen</u> to him . Don't let anything happen to the
dickens_oliver.txt	it's found out , ' said Monks . ' And	what may	that be ? ' asked the matron . ' The
dickens_oliver.txt	woman with much deliberation . ' The first is ,	what may	the communication be worth ? ' ' Who the devil
james_ambassadors.txt	She gave a strange light laugh . " Putting aside	what may	really hurt HIM . " But her friend looked at
stevenson_arrow.txt	" said the priest , somewhat recovering , " and	what may	this be ? What enemy hath done this ? "
stevenson_arrow.txt	I , Richard Shelton , tide what betideth , come	what may	, will see you safe in Hollywood . The saints
stevenson_arrow.txt	tall chimney marked the ruins of a house . "	What may	this be ? " whispered Matcham . " Nay ,
stevenson_arrow.txt	he cried again , " I like that little .	What may	this betoken ? Let us go , by the mass
collins_basil.txt	into this hotel a little while ago . " "	What may	your business be ? " - - He hesitated ,
collins_basil.txt	added in an altered tone , " I mean ,	what may	you want with them , Sir ? " " I
kipling_captains.txt	It's all new to you , but we never know	what may	come to us . Now , take me - -
eliot_daniel.txt	scruple , having a conscience of what must be and	what may	be ; whereas Ignorance is a blind giant who ,
corelli_innocent.txt	! " she said . " One never can tell	what may	happen in this strange world ! But we have to
hardy_jude.txt	beside them . " Well , my lad , and	what may	you want up here ? " " I wanted to
lytton_kenelm.txt	who had saved from the general wreck of their fortunes	what may	be called a good squire's estate , - - about
doyle_lost.txt	" Well , " he cried , expectantly , "	what may	it run to ? I'm thinking , young man ,
gaskell_lovers.txt	I beg your pardon , miss ; I was thinking	what may	the price of yon crimson ribbon be ? ' Hester

# Examining an author's individual grammar

---

Dr Andrea Nini

[andrea.nini@manchester.ac.uk](mailto:andrea.nini@manchester.ac.uk)

[www.andreanini.com](http://www.andreanini.com)

Comparative Literature Goes Digital Workshop

DH2025

14<sup>th</sup> July 2025