# John Burrows: Delta

## A Measure of Stylistic Difference

Robert Paßmann

September 22, 2015

Arbeitsgruppe 2: **Who wrote the web?**
Sommerakademie der Studienstiftung in La Colle-sur-Loup

# TABLE OF CONTENTS

# THE DELTA PROCEDURE

# IN EASY WORDS...

**We have** a database of authors with some of their texts
a sample text of unknown authorship

**We want** to order the authors by likelihood of authorship

- Therefore, measure the difference of a sample text and an author by a single value – *Delta*.
- The most likely author will be the one with the least delta.

# HOW DOES IT WORK? AN EXAMPLE

| | A | B | C | D | F | G | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | **Main set** | | **Milton** | | ***Paradise Lost*** | | | |
| 2 | | | | | | | **count** | | | **30** |
| 3 | | | | | | | **sum** | | | **31.489** |
| 4 | | | | | | | **mean (= "delta")** | | | **1.050** |
| 5 | | | | | | | **stdev** | | | **0.770** |
| 6 | | | **Mean** | **Stdev** | **Scores** | **z-scores** | **Scores** | **z-scores** | **Diff.** | **Abs. diff.** |
| 7 | 1 | the | 4.242 | 0.630 | 4.719 | 0.757 | 4.091 | –0.239 | –0.996 | 0.996 |
| 8 | 2 | and | 3.770 | 0.501 | 4.407 | 1.272 | 4.165 | 0.789 | –0.483 | 0.483 |
| 9 | 3 | of | 1.821 | 0.315 | 2.420 | 1.905 | 2.769 | 3.015 | 1.110 | 1.110 |
| 10 | 4 | a | 1.601 | 0.430 | 0.893 | –1.645 | 0.696 | –2.103 | –0.458 | 0.458 |
| 11 | 5 | to(i) | 1.419 | 0.272 | 1.247 | –0.634 | 1.289 | –0.480 | 0.154 | 0.154 |
| 12 | 6 | in(p) | 1.358 | 0.189 | 1.554 | 1.035 | 1.720 | 1.916 | 0.881 | 0.881 |
| 13 | 7 | his | 1.154 | 0.323 | 1.062 | –0.284 | 1.532 | 1.171 | 1.454 | 1.454 |

J. F. Burrows, "Delta: a measure of stylistic difference and a guide to likely authorship", Literary and Linguistic Computing 17, pp. 267–287, 2002a.

# HOW DOES IT WORK?

1. For every text in the database, calculate the relative frequency or scores $f_{t_i}(w)$ of every (tagged) word $w$ in the text.
2. Calculate the means $\mu_{a_i}(w)$, $\mu(w)$ and standard deviations $\sigma_{a_i}(w)$, $\sigma(w)$ of the scores with respect to authors $(a_i)$ and the whole database.
3. Calculate the z-scores for every word of every author in the database:
$$z_{a_i}(w) = \frac{\mu_{a_i}(w) - \mu(w)}{\sigma(w)}$$
4. For the sample text $s$, calculate the mean frequencies $f_s(w)$ and their z-scores with respect to the mean frequencies in the whole database.
5. Calculate the delta for every author as:
$$\Delta_s(a_i) = \frac{1}{|M|} \sum_{w \in M} |z_s(w) - z_{a_i}(w)|$$
6. Finally, compare the deltas of the different authors.

## Burrows tested the method as follows:

- Using a main database of 25 english authors of the late seventeenth century
- He tested 200 english poems of 15 authors
  - 12 of 15 authors are in the database
  - no poem is contained in the database

## His observations were:

- The delta method works better than expected
- It works for closed- and open-class problems
- Great method for reducing the field of likely candidates
- It works best for longer texts ($>$ 1500 words)
- The method might fail for texts which are uncharacteristic for their authors or are far separated in time

# EXPERIMENTS AND RESULTS (II)

| | Poems sorted by length, showing authors' ranks (ex 25) | | | | | |
|---|---|---|---|---|---|---|
| | **1–500** | **501–** | **1001–** | **1501–** | **2001–** | **Totals** |
| | **100** | **40** | **20** | **20** | **20** | **200** |
| *150 words* | | | | | | |
| 1st | 27 | 18 | 13 | 17 | 19 | 94 |
| 1st–2nd | 40 | 26 | 14 | 18 | 20 | 118 |
| 1st–5th | 67 | 32 | 18 | 20 | 20 | 157 |
| 06 : 10 | 15 | 8 | 2 | 0 | 0 | 25 |
| 11 : 15 | 10 | 0 | 0 | 0 | 0 | 10 |
| 16 : 20 | 5 | 0 | 0 | 0 | 0 | 5 |
| 21 : 25 | 3 | 0 | 0 | 0 | 0 | 3 |

J. F. Burrows, "Delta: a measure of stylistic difference and a guide to likely authorship", Literary and Linguistic Computing 17, pp. 267–287, 2002a.

# EXPERIMENTS AND RESULTS (III)

| | Poems sorted by length, showing authors' ranks (ex 25) | | | | | |
|---|---|---|---|---|---|---|
| | 1–500 % | 501– % | 1001– % | 1501– % | 2001– % | Totals % |
| *150 words* | | | | | | |
| 1st | 27.0 | 45.0 | 65.0 | 85.0 | 95.0 | 47.0 |
| 1st–2nd | 40.0 | 65.0 | 70.0 | 90.0 | 100 | 59.0 |
| 1st–5th | 67.0 | 80.0 | 90.0 | 100 | 100 | 78.5 |
| 06 : 10 | 15.0 | 20.0 | 10.0 | 0 | 0 | 12.5 |
| 11 : 15 | 10.0 | 0 | 0 | 0 | 0 | 5.0 |
| 16 : 20 | 5.0 | 0 | 0 | 0 | 0 | 2.5 |
| 21 : 25 | 3.0 | 0 | 0 | 0 | 0 | 1.5 |

J. F. Burrows, "Delta: a measure of stylistic difference and a guide to likely authorship", Literary and Linguistic Computing 17, pp. 267–287, 2002a.

# REPRODUCING THE APPROACH

# AN IMPLEMENTATION OF THE DELTA METHOD

- Implemented in `Python 3.4`
- Using `NLTK` library for tagging

- Algorithm is implemented in three classes
- Every `Text` is written by an `Author` of our `Database`
- These classes have methods to perform the calculations

# PROBLEMS DURING REPRODUCTION

- What does the main database consist of? PAN12
- When do the deltas indicate that there is too less difference such that further investigation is needed?

# RESULTS (I)

| | Sum of A,B (3 authors) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rank / Words** | **40** | | **60** | | **80** | | **100** | | **120** | | **150** | | **0** | | **Sum** | |
| **1st** | 10 | *0.83* | 12 | *1.00* | 12 | *1.00* | 12 | *1.00* | 12 | *1.00* | 12 | *1.00* | 10 | *0.83* | 80 | *0.95* |
| **2nd** | 2 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 1 | *0.92* | 3 | *0.99* |
| **3rd** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 1 | *1.00* | 1 | *1.00* |
| **4th-5th** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* |
| **6th-10th** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* |
| **10th-14th** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* |
| | 12 | | 12 | | 12 | | 12 | | 12 | | 12 | | 12 | | 84 | |

| | Sum of C,D (8 authors) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rank / Words** | **40** | | **60** | | **80** | | **100** | | **120** | | **150** | | **0** | | **Sum** | |
| **1st** | 12 | *0.75* | 14 | *0.88* | 14 | *0.88* | 13 | *0.81* | 13 | *0.81* | 13 | *0.81* | 14 | *0.88* | 93 | *0.83* |
| **2nd** | 3 | *0.94* | 2 | *1.00* | 1 | *0.94* | 3 | *1.00* | 2 | *0.94* | 2 | *0.94* | 2 | *1.00* | 15 | *0.96* |
| **3rd** | 1 | *1.00* | 0 | *1.00* | 1 | *1.00* | 0 | *1.00* | 1 | *1.00* | 1 | *1.00* | 0 | *1.00* | 4 | *1.00* |
| **4th-5th** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* |
| **6th-10th** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* |
| **10th-14th** | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* | 0 | *1.00* |
| | 16 | | 16 | | 16 | | 16 | | 16 | | 16 | | 16 | | 112 | |

# RESULTS (II)

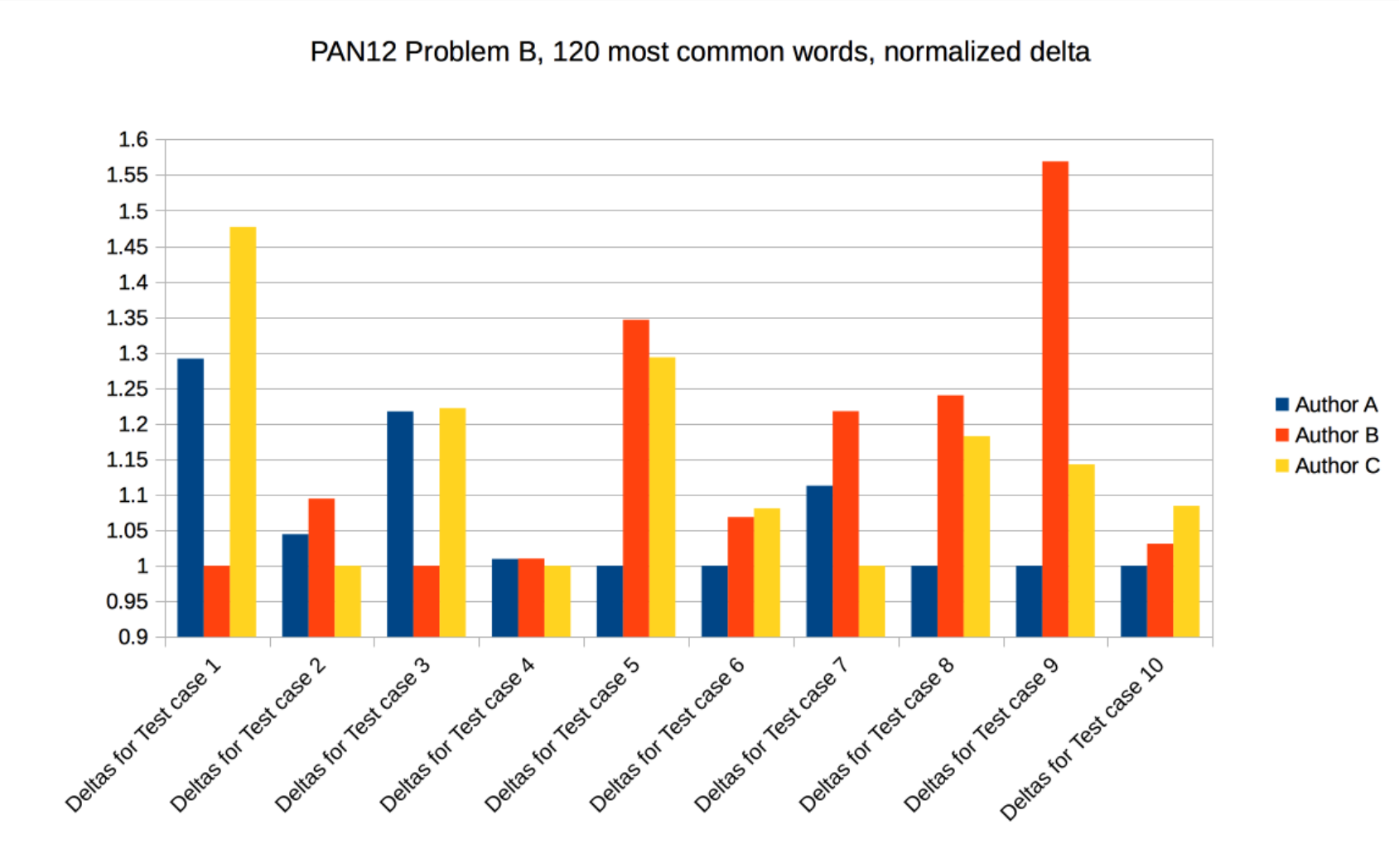
| | Sum of I,J (14 authors) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rank / Words** | **40** | | **60** | | **80** | | **100** | | **120** | | **150** | | **0** | | **Sum** | |
| **1st** | 20 | *0.71* | 24 | *0.86* | 24 | *0.86* | 24 | *0.86* | 25 | *0.89* | 25 | *0.89* | 23 | *0.82* | 165 | *0.84* |
| **2nd** | 1 | *0.75* | 0 | *0.86* | 1 | *0.89* | 2 | *0.93* | 1 | *0.93* | 1 | *0.93* | 1 | *0.86* | 7 | *0.88* |
| **3rd** | 2 | *0.82* | 1 | *0.89* | 2 | *0.96* | 0 | *0.93* | 1 | *0.96* | 1 | *0.96* | 1 | *0.89* | 8 | *0.92* |
| **4th-5th** | 4 | *0.96* | 0 | *0.89* | 0 | *0.96* | 1 | *0.96* | 0 | *0.96* | 0 | *0.96* | 1 | *0.93* | 6 | *0.95* |
| **6th-10th** | 0 | *0.96* | 2 | *0.96* | 0 | *0.96* | 0 | *0.96* | 0 | *0.96* | 0 | *0.96* | 1 | *0.96* | 3 | *0.96* |
| **10th-14th** | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 7 | *1.00* |
| | 28 | | 28 | | 28 | | 28 | | 28 | | 28 | | 28 | | 196 | |

| | Sum of all | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rank / Words** | **40** | | **60** | | **80** | | **100** | | **120** | | **150** | | **0** | | **Sum** | |
| **1st** | 42 | *0.75* | 50 | *0.89* | 50 | *0.89* | 49 | *0.88* | 50 | *0.89* | 50 | *0.89* | 47 | *0.84* | 338 | *0.86* |
| **2nd** | 6 | *0.86* | 2 | *0.93* | 2 | *0.93* | 5 | *0.96* | 3 | *0.95* | 3 | *0.95* | 4 | *0.91* | 25 | *0.93* |
| **3rd** | 3 | *0.91* | 1 | *0.95* | 3 | *0.98* | 0 | *0.96* | 2 | *0.98* | 2 | *0.98* | 2 | *0.95* | 13 | *0.96* |
| **4th-5th** | 4 | *0.98* | 0 | *0.95* | 0 | *0.98* | 1 | *0.98* | 0 | *0.98* | 0 | *0.98* | 1 | *0.96* | 6 | *0.97* |
| **6th-10th** | 0 | *0.98* | 2 | *0.98* | 0 | *0.98* | 0 | *0.98* | 0 | *0.98* | 0 | *0.98* | 1 | *0.98* | 3 | *0.98* |
| **10th-14th** | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 1 | *1.00* | 7 | *1.00* |
| | 56 | | 56 | | 56 | | 56 | | 56 | | 56 | | 56 | | 392 | |

# PROBLEMS DURING REPRODUCTION

- What does the main database consist of? **PAN12**
- When do the deltas indicate that there is too less difference such that further investigation is needed?

# LET'S HAVE A CLOSER LOOK...




- test cases 4, 6, 8 and 10 are not of authors from the database
- with a threshold at 1.10, we have a success rate of 8/10

# AN IDEA TO SOLVE THE OPEN-CLASS PROBLEMS

- choose a reasonable threshold $x$
- normalize all deltas with respect to the minimum delta value, i.e.

$$\delta_i = \frac{\Delta_s(a_i)}{\Delta_{\min}}$$

- if there is no $i$ with $\delta_i \in [1, x)$ then output $a_i$
- otherwise further investigation is needed (output **none**)

# RESULTS OF THE OPEN-CLASS PROBLEMS (I)

| | | | | number of correct cases (i.e. correct author or correct none) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Problem B | | | 3 authors (2 training files each), 6/10 test cases | | | | | | | | #test | 10 | |
| Threshold / W | 40 | | 60 | | 80 | | 100 | | 120 | | 150 | | 0 | | |
| 1.01 | 5 | *0.50* | 6 | *0.60* | 8 | *0.80* | 6 | *0.60* | 7 | *0.70* | 7 | *0.70* | 5 | *0.50* | 0.63 |
| 1.025 | 5 | *0.50* | 8 | *0.80* | 8 | *0.80* | 7 | *0.70* | 7 | *0.70* | 8 | *0.80* | 5 | *0.50* | 0.69 |
| 1.05 | 5 | *0.50* | 8 | *0.80* | 8 | *0.80* | 8 | *0.80* | 8 | *0.80* | 9 | *0.90* | 5 | *0.50* | 0.73 |
| 1.1 | 7 | *0.70* | 8 | *0.80* | 8 | *0.80* | 7 | *0.70* | 8 | *0.80* | 7 | 0.70 | 6 | *0.60* | 0.73 |
| 1.2 | 7 | *0.70* | 8 | *0.80* | 7 | *0.70* | 5 | *0.50* | 6 | *0.60* | 6 | *0.60* | 5 | *0.50* | 0.63 |
| | | *0.58* | | *0.8* | | *0.8* | | *0.7* | | *0.7* | | *0.7* | | 0.52 | |

| | | | | number of correct cases (i.e. correct author or correct none) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Problem D | | | 8 authors (2 training files each), 8/17 test cases | | | | | | | | #test | 17 | |
| Threshold / W | 40 | | 60 | | 80 | | 100 | | 120 | | 150 | | 0 | | |
| 1.01 | 6 | *0.35* | 8 | *0.47* | 9 | *0.53* | 7 | *0.41* | 8 | *0.47* | 6 | *0.35* | 10 | *0.59* | 0.45 |
| 1.025 | 7 | *0.41* | 9 | *0.53* | 9 | *0.53* | 9 | *0.53* | 7 | *0.41* | 7 | *0.41* | 11 | *0.65* | 0.50 |
| 1.05 | 8 | *0.47* | 9 | *0.53* | 10 | *0.59* | 9 | *0.53* | 9 | *0.53* | 9 | *0.53* | 10 | *0.59* | 0.54 |
| 1.1 | 8 | *0.47* | 8 | *0.47* | 11 | *0.65* | 9 | *0.53* | 12 | *0.71* | 14 | *0.82* | 9 | *0.53* | 0.60 |
| 1.2 | 10 | *0.59* | 9 | *0.53* | 12 | *0.71* | 10 | *0.59* | 11 | *0.65* | 11 | *0.65* | 9 | *0.53* | 0.61 |
| | | *0.46* | | *0.5* | | *0.6* | | *0.5* | | *0.6* | | *0.6* | | 0.58 | |

# Results of the open-class problems (II)

| | | | | number of correct cases (i.e. correct author or correct none) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Problem J | | | 14 authors (2 training files each), 14/16 test case | | | | | | | | #test | 16 | |
| **Threshold / V** | **40** | | **60** | | **80** | | **100** | | **120** | | **150** | | **0** | | |
| **1.01** | 11 | *0.69* | 13 | *0.81* | 13 | *0.81* | 13 | *0.81* | 14 | *0.88* | 14 | *0.88* | 14 | *0.88* | 0.82 |
| **1.025** | 10 | *0.63* | 13 | *0.81* | 13 | *0.81* | 14 | *0.88* | 14 | *0.88* | 14 | *0.88* | 12 | *0.75* | 0.80 |
| **1.05** | 9 | *0.56* | 12 | *0.75* | 14 | *0.88* | 13 | *0.81* | 14 | *0.88* | 13 | *0.81* | 7 | *0.44* | 0.73 |
| **1.1** | 9 | *0.56* | 11 | *0.69* | 14 | *0.88* | 14 | *0.88* | 14 | *0.88* | 12 | *0.75* | 3 | *0.19* | 0.69 |
| **1.2** | 10 | *0.63* | 10 | *0.63* | 10 | *0.63* | 9 | *0.56* | 10 | *0.63* | 10 | *0.63* | 2 | *0.13* | 0.54 |
| | | *0.61* | | *0.7* | | *0.8* | | *0.8* | | *0.8* | | *0.8* | | 0.48 | |
| | #15 correct: 40/1.2, 60/1.2, 80/1.05, 80/1.1, 80/1.2, 100/1.1, 100/1.2, | | | | | | | | | | | | | | |
| | 120/all, 150/1.05, 150/1.1, 150/1.2, 0/all | | | | | | | | | | | | | | |
| | #16 correct: 40/1.2, 60/1.05, 60/1.1, 60/1.2, 80/1.025, 80/1.05, 80/1.1, 80/1.2 | | | | | | | | | | | | | | |
| | 100/1.025, 100/1.05, 100/1.1, 100/1.2, 120/all, 150/all, 0/all | | | | | | | | | | | | | | |

# CONCLUSION

## Regarding the Delta method and the tests with PAN12 data

- Delta works good to reduce large sets of possible authors
- Sometimes Delta has no clue

## Regarding Burrow's paper, i.e. the reproduction

- It was not possible to reproduce Burrow's example because of missing information (How did he form his database?)
- It was necessary to find a way to deal with open-class problems
- It can be confirmed that Delta is useful for reducing the set of possible authors