

Gute Wörter für Delta: Verbesserung der Autorschaftsattribute durch autorspezifische distinktive Wörter

Dimpel, Friedrich

mail@dimpel.de
Universität Erlangen

Proisl, Thomas

thomas.proisl@fau.de
Universität Erlangen

1. Einleitung

Autorschaftsattributionsverfahren sind längst etabliert (vgl. Burrows 2002, Jannidis et al. 2015). Auf DHd-Jahrestagungen wurden Techniken vorgestellt, die die Attributionsverfahren optimieren (Büttner et al. 2016, Dimpel 2017a, 2018a/b). Die Optimierung der Verfahren ist deshalb wichtig, weil ein literaturwissenschaftliches Interesse besteht, die Frage nach der Autorschaft auch bei schwierigen Bedingungen zu klären. Gute Bedingungen, bei denen in Evaluationstests über hohe Erkennungsquoten berichtet wurde, sind dann gegeben, wenn viele Texte aus der gleichen Gattung vorliegen, wenn die Texte eine ausreichende Länge aufweisen (mindestens 5.000 Wortformen), wenn die Texte chronologisch nicht zu weit auseinander liegen und wenn die Texte einen möglichst normierten Sprachstand hinsichtlich Orthographie und Standardsprache aufweisen (Schöch 2014, Eder 2013a und 2013b).

All diese Bedingungen sind nicht erfüllt, wenn man mittelhochdeutsche Kleinepik untersuchen möchte – gerade etwa mit Blick auf die Kleinepik des Strickers wäre eine stilometrische Klärung bei einigen Texten interessant. Mittelhochdeutsche Texte folgen keiner geregelten Orthographie, es liegen sowohl mehr oder weniger normalisierte als auch nicht-normalisierte digitale Texte vor, sie sind oft dialektal geprägt und viele Texte aus dem Bereich der Kleinepik übersteigen kaum eine Anzahl von 2.000 Wortformen. Anhand der ‚Halben Birne‘ (umstrittene Autorschaft Konrads von Würzburg; 2.469 Wortformen) wurde ein Verfahren vorgestellt, wie mit Burrows’ Delta die distinktiven Wörter ermittelt werden können, die Konrads Wortschatz von anderen Autoren unterscheiden (Dimpel 2018a). Damit konnte die Erkennungsqualität so optimiert werden, dass Delta auf die ‚Halbe Birne‘ angewendet werden konnte. Nun wird

anhand eines weniger problematischen Korpus (Romane ca. 19. Jahrhundert¹) evaluiert, wie sich Erkennungsquote und False-Positives verändern, wenn man Delta nicht auf allen Most Frequent Words (MFWs) berechnet, sondern nur auf „guten“ MFWs.

2. Gute-Wörter und die Level-2-Differenz

Für Burrows’ Delta ermittelt man relative Worthäufigkeiten – hier für zwei Texte von Wolfram (Willehalm, Parzival) sowie den Tristan von Gottfried. Weiter werden Mittelwert, Standardabweichung und z -Werte (Spalten Z1, Z2, Z3) sowie die absolute Differenz der z -Werte (Spalten AbsDiff) berechnet. Der Mittelwert der absoluten z -Wert-Differenzen ergibt das Abstandsmaß Delta – hier wäre Delta für Parzival–Willehalm der Mittelwert der drittletzten Spalte.

| F | Wh1 | Pz2 | Trist3 | Mean | StDev | Z1 Wh | Z2 Pz | Z3 Trist | Gleiche Versch. Autoren Autoren | | | |
|-------|------|------|--------|------|-------|-------|-------|----------|------------------------------------|-------------|--------|--|
| | | | | | | | | | Abs Diff | Abs Diff | L2Diff | |
| 1 der | 3,19 | 2,62 | 2,13 | 2,65 | 0,53 | 1,02 | -0,05 | -0,98 | 1,07 | 2,00 | 0,93 | |
| 2 daz | 2,02 | 1,93 | 2,77 | 2,24 | 0,46 | -0,48 | -0,67 | 1,15 | 0,19 | 1,63 | 1,44 | |
| 3 und | 1,97 | 1,01 | 3,52 | 2,17 | 1,27 | -0,15 | -0,91 | 1,07 | 0,76 | 1,22 | 0,46 | |
| 4 ir | 1,33 | 1,81 | 2,14 | 1,76 | 0,41 | -1,06 | 0,12 | 0,93 | 1,18 | 1,99 | 0,81 | |
| 5 er | 1,51 | 1,63 | 1,97 | 1,70 | 0,24 | -0,80 | -0,32 | 1,12 | 0,48 | 1,92 | 1,44 | |
| 6 in | 1,30 | 1,25 | 1,66 | 1,40 | 0,23 | -0,48 | -0,67 | 1,15 | 0,19 | 1,63 | 1,44 | |
| 7 ich | 1,29 | 1,43 | 1,30 | 1,34 | 0,08 | -0,67 | 1,15 | -0,48 | 1,82 | 0,18 | -1,63 | |
| ... | | | | | | | | | | | | |

Die guten Wörter beruhen auf den Level-2-Differenzen: Die z -Wert-Differenzen zwischen den Autortexten sollen kleiner sein als die z -Wert-Differenzen zwischen Autortexten und Distraktortexten: Man zieht die Differenzen zwischen Autor- und Distraktortexten von den Differenzen zwischen den Autortexten untereinander ab. Im Beispiel bildet man die Differenz aus der z -Wert-Differenz zwischen Wolframs ‚Willehalm‘ und Gottfrieds ‚Tristan‘ (für „der“: 2,0) und zwischen Wolframs ‚Willehalm‘ und Wolframs ‚Parzival‘ (für „der“: 1,07). Die Level-2-Differenz für „der“ beträgt 0,93. Bei einer positiven Level-2-Differenz kann ein Wort die Autortexte von den Distraktortexten unterscheiden; „ich“ mit negativer Level-2-Differenz ist hier ungeeignet. In dem Beispiel mit nur drei Texten ist dies auch anhand der Worthäufigkeitswerte und z -Werte nachvollziehbar. Sonst sind jedoch weder erhöhte Worthäufigkeitswerte noch erhöhte z -Werte allein hinreichend, um ein Wort als distinktiv zu identifizieren (Dimpel 2017b).

Formal ausgedrückt werden die Level-2-Differenzen anhand von zwei Teilkorpora ermittelt: Einem Teilkorpus $A = a_1, \dots, a_n$ von Autortexten und einem Teilkorpus $D = d_1, \dots, d_n$ von Distraktortexten anderer Autoren. Dabei ist jeder Text ein Vektor von z -Werten. Es werden für jeden Autortext a_i die absoluten z -Wert-Differenzen zu jedem Distraktortext d und zu jedem anderen Autortext a

j ermittelt, daraus die Level-2-Differenzen berechnet und anschließend gemittelt. Für die Mittelwerte der Level-2-Differenzen gilt:

$$l = \frac{\sum_{a_i \in A} \sum_{a_j \in A \setminus a_i} \sum_{d \in D} (\text{abs}(a_i - d) - \text{abs}(a_i - a_j))}{|A| |D| (|A| - 1)}$$

3. Setting und Korpus

Wir prüfen, ob mit „guten Wörtern“ Texte des Zielautors besser von denen anderer Autoren abgegrenzt werden können als mit dem üblichen Delta-Ansatz, der rein auf den n häufigsten Wörtern basiert. Wir fassen die Autorschaftsattributions als ein Zwei-Klassen-Problem auf: Für jeden Text im Ratekorpus wollen wir wissen, ob er vom Zielautor oder von einem anderen Autor ist.

Für die Berechnung der guten Wörter und zur Evaluation werden ein Ratekorpus und ein Vergleichskorpus verwendet. Im Vergleichskorpus befinden sich ein Autorvergleichstext (Text vom Zielautor), je nach Testreihe zwanzig oder mehr Distraktortexte von Autoren, die nicht im Ratekorpus vertreten sind, sowie in manchen Experimenten Texte von den übrigen im Ratekorpus vertretenen Autoren. Wenn der Delta-Abstand zwischen Ratetext und Autorvergleichstext der niedrigste ist, gilt dieser Text als dem Zielautor zugeordnet; ist der Abstand zu einem Distraktortext von einem anderen Autor am niedrigsten, gilt der Text als falsch erkannt.²Für die Evaluation werden Precision, Recall, F₁-Score und Accuracy berechnet und über alle Autoren hinweg gemittelt.

Die guten Wörter werden jeweils auf den alphabetisch ersten drei Texten des Autors für jeden Autor in vier Varianten berechnet. Ausgangspunkt sind jeweils die häufigsten 1.200 Wörter:

1. Als gute Wörter werden Wörter gewählt, bei denen die Mittelwerte aller Level-2-Differenzen # 0,4 sind.
2. wie A, jedoch Mittelwert # 0,2.
3. Wie A, zusätzlich müssen 3 von 3 oder 5 von 6 Autortexten zu mindestens je einem Distraktortext eine Level-2-Differenz > 1,64 aufweisen.
4. Wie C, jedoch muss die Level-2-Differenz nur > 1,2 statt 1,64 sein; zudem mit Mittelwerten # 0,4. Die zwei schlechtesten Level-2-Differenzen werden ignoriert, da ein Wort einen Text womöglich gut von vielen Distraktortexten unterscheiden kann, bei einigen wenigen Distraktortexten jedoch womöglich sehr schlecht funktioniert.³

Wir verwenden stets 400 gute Wörter, da auch kurze Texte mit 2.000 Token getestet werden sollen. Wenn eine Variante mehr als 400 gute Wörter liefert, verwenden wir die 400 Wörter mit den größten Level-2-Differenz-

Mittelwerten; wenn sie weniger als 400 gute Wörter liefert, füllen wir mit MFWs auf.

4. Experimente

Wir führen drei Testreihen durch:

1. Bootstrapping der Ratetexte. Im Vergleichskorpus werden die vollen Texte verwendet; um die Robustheit zu erhöhen, wird ein Bootstrapping-Verfahren auf die Ratetexte angewendet. Pro Ratetext werden 100 Stichproben gezogen, wobei für jede Stichprobe zufällig so viele Wörter mit Zurücklegen aus dem Ratetext gezogen werden, wie der Ratetext umfasst.
2. Subsampling der Ratetexte. Im Vergleichskorpus werden die vollen Texte verwendet, aus den Ratetexten werden jedoch zufällige Stichproben von 2.000 Wörtern gezogen, um das Verhalten auf kurzen Ratetexten zu simulieren.
3. Subsampling der Rate- und der Vergleichstexte. Sowohl aus den Rate- als auch aus den Vergleichstexten werden zufällige Stichproben von 2.000 Wörtern gezogen, um das Verhalten auf kurzen Textsorten zu simulieren.

Dadurch werden wichtige Einsatzszenarien (lange Texte, kurze Texte mit langen Vergleichstexten und kurze Texte mit kurzen Vergleichstexten) abgedeckt.

Testreihe 1: Bootstrapping der Ratetexte

Hier evaluieren wir die Methoden bei vollständigen Romanen. Einmal enthält das Vergleichskorpus Texte von allen Autoren im Ratekorpus, im zweiten Schritt enthält das Vergleichskorpus zwar einen Text des Zielautors, jedoch keinen Text der anderen Autoren im Ratekorpus.

Für das erste Experiment verwenden wir ein Vergleichskorpus, das je einen Text von allen 32 Autoren enthält. Für jeden Autor ermitteln wir die guten Wörter auf Basis des Vergleichskorpus und zwei weiteren Texten des Autors (also insgesamt drei Autortexten und 31 Distraktortexten). Die Ratekorpora für jeden Autor umfassen drei Texte des Autors und drei zufällige Texte von den übrigen Autoren.

Mittelwerte über alle 32 Autoren:

| Methode | Zielautor | | | nicht Zielautor | | | global | |
|---------|---------------|---------------|----------------|-----------------|---------------|----------------|---------------|----------------|
| | Prec. | Recall | F ₁ | Prec. | Recall | F ₁ | Acc. | F ₁ |
| MFW400 | 0,9994 | 0,7976 | 0,8633 | 0,8623 | 0,9994 | 0,9181 | 0,8985 | 0,8907 |
| MFW1200 | 0,9992 | 0,8249 | 0,8795 | 0,8831 | 0,9992 | 0,9298 | 0,9120 | 0,9046 |
| A | 0,9774 | 0,9689 | 0,9693 | 0,9753 | 0,9716 | 0,9699 | 0,9702 | 0,9696 |
| B | 0,9804 | 0,9750 | 0,9744 | 0,9804 | 0,9752 | 0,9746 | 0,9751 | 0,9745 |
| C | 0,9803 | 0,9746 | 0,9742 | 0,9800 | 0,9751 | 0,9744 | 0,9748 | 0,9743 |
| D | 0,9799 | 0,9742 | 0,9738 | 0,9796 | 0,9751 | 0,9744 | 0,9746 | 0,9741 |

Die klassischen Delta-Varianten (MFW400, MFW1200) haben in Bezug auf den Ziellautor eine hohe Precision und einen niedrigeren Recall, während sich das Verhältnis für Texte anderer Autoren umdreht: Niedrige Erkennungsquote, dafür kaum False Positives (Texte, die fälschlicherweise dem Ziellautor zugeschrieben werden).

Die Gute-Wörter-Varianten führen zu deutlichen Recall-Verbesserungen beim Ziellautor auf Kosten einer etwas geringeren Precision. Die Precision für Texte anderer Autoren steigt zu Lasten des Recalls. Mit den guten Wörtern werden also mehr Texte des Autors richtig erkannt (>97%), dafür gibt es minimal mehr False Positives (2,5%).

Mehr häufigste Wörter funktionieren besser als weniger. Alle vier Gute-Wörter-Varianten funktionieren besser als die reinen Delta-Varianten (Verbesserungen von 6–8 Punkten bei Accuracy und F_1).

Für das zweite Experiment wird lediglich das Vergleichskorpus verändert. Es enthält jetzt jeweils einen Text des Ziellautors und je einen Text von 24 zusätzlichen Autoren (die sich nicht mit den 32 getesteten Autoren überschneiden). Die wahren Autoren für drei der sechs Texte in den Ratekorpora befinden sich nicht mehr im Vergleichskorpus, was die Klassifikation der entsprechenden Texte erschwert.

| Methode | Ziellautor | | | nicht Ziellautor | | | global | |
|---------|---------------|---------------|---------------|------------------|---------------|---------------|---------------|---------------|
| | Prec. | Recall | F_1 | Prec. | Recall | F_1 | Acc. | F_1 |
| MFW400 | 0,9438 | 0,8497 | 0,8857 | 0,8982 | 0,9225 | 0,8908 | 0,8861 | 0,8782 |
| MFW1200 | 0,9453 | 0,8929 | 0,9008 | 0,9216 | 0,9280 | 0,9099 | 0,9105 | 0,9053 |
| A | 0,9111 | 0,9827 | 0,9383 | 0,9861 | 0,8751 | 0,9120 | 0,9289 | 0,9251 |
| B | 0,9112 | 0,9830 | 0,9384 | 0,9864 | 0,8750 | 0,9120 | 0,9290 | 0,9252 |
| C | 0,9131 | 0,9829 | 0,9395 | 0,9865 | 0,8781 | 0,9141 | 0,9305 | 0,9268 |
| D | 0,9134 | 0,9854 | 0,9410 | 0,9883 | 0,8777 | 0,9145 | 0,9316 | 0,9278 |

Tatsächlich bewegen sich alle Ergebnisse auf einem etwas niedrigeren Niveau als im ersten Experiment. Auch hier: Mittels guter Wörter steigt der Recall für Texte des Ziellautors deutlich, während die Precision leicht sinkt. Für Texte anderer Autoren verhält es sich umgekehrt. Auch hier führen die guten Wörter zu besseren Accuracy- und F_1 -Werten (2–5 Punkte).

Testreihe 2: Subsampling der Ratetexte

In dieser Testreihe wollen wir prüfen, wie gut die einzelnen Methoden für kurze Rate- und lange (vollständige) Vergleichstexte funktionieren. Dazu ziehen wir für jeden Ratetext 100 Stichproben à 2.000 Wörter; der übrige Versuchsaufbau ist identisch zu Testreihe 1.

Test 2a) Alle Autoren aus den Ratekorpora sind im Vergleichskorpus vertreten:

| Methode | Ziellautor | | | nicht Ziellautor | | | global | |
|---------|---------------|---------------|---------------|------------------|---------------|---------------|---------------|---------------|
| | Prec. | Recall | F_1 | Prec. | Recall | F_1 | Acc. | F_1 |
| MFW400 | 0,9823 | 0,6993 | 0,7936 | 0,7888 | 0,9835 | 0,8685 | 0,8414 | 0,8310 |
| MFW1200 | 0,9816 | 0,6704 | 0,7555 | 0,7816 | 0,9781 | 0,8573 | 0,8243 | 0,8064 |
| A | 0,9025 | 0,9365 | 0,9081 | 0,9460 | 0,8694 | 0,8886 | 0,9029 | 0,8983 |
| B | 0,8955 | 0,9368 | 0,9032 | 0,9479 | 0,8569 | 0,8777 | 0,8968 | 0,8905 |
| C | 0,8935 | 0,9358 | 0,9019 | 0,9463 | 0,8544 | 0,8750 | 0,8951 | 0,8884 |
| D | 0,9035 | 0,9368 | 0,9081 | 0,9472 | 0,8685 | 0,8851 | 0,9027 | 0,8966 |

Die Ergebnisse sind auf den 2.000-Wort-Samples spürbar schlechter. Das Gesamtbild ist jedoch dasselbe: Die guten Wörter verbessern die Klassifikation und führen zu deutlich höheren Recall-Werten für den Ziellautor bei etwas niedrigerer Precision.

Test 2B) Die Nicht-Ziellautoren sind nicht im Vergleichskorpus vertreten:

| Methode | Ziellautor | | | nicht Ziellautor | | | global | |
|---------|---------------|---------------|---------------|------------------|---------------|---------------|---------------|---------------|
| | Prec. | Rec. | F_1 | Prec. | Rec. | F_1 | Acc. | F_1 |
| MFW400 | 0,9345 | 0,7601 | 0,8177 | 0,8197 | 0,9358 | 0,8648 | 0,8480 | 0,8413 |
| MFW1200 | 0,9303 | 0,7546 | 0,8003 | 0,8261 | 0,9116 | 0,8435 | 0,8331 | 0,8219 |
| A | 0,7844 | 0,9670 | 0,8565 | 0,9605 | 0,6806 | 0,7570 | 0,8238 | 0,8067 |
| B | 0,7823 | 0,9682 | 0,8560 | 0,9618 | 0,6783 | 0,7562 | 0,8233 | 0,8061 |
| C | 0,7860 | 0,9692 | 0,8586 | 0,9627 | 0,6829 | 0,7589 | 0,8260 | 0,8087 |
| D | 0,7906 | 0,9670 | 0,8605 | 0,9615 | 0,6925 | 0,7678 | 0,8297 | 0,8141 |

Auch hier verbessern die guten Wörter die Erkennung der Ziellautortexte (Erkennungsquote >96%). Größere Einbußen gibt es bei der Erkennung von Texten, die nicht vom Ziellautor stammen, so dass die guten Wörter zwar die Erkennung von Texten des Ziellautors enorm verbessern (Reduktion der Fehlerrate >80%), die Ergebnisse insgesamt aber leicht schlechter sind.

Testreihe 3: Subsampling der Rate- und der Vergleichstexte

Nunmehr werden die guten Wörter mit zwei Ratetexten und einem Autorvergleichstext berechnet; wenn jedoch mehr als sechs Texte pro Autor vorliegen, werden entsprechend mehr Ratetexte verwendet. Da diese drei (oder mehr) Texte nicht wie oben reihum als Rate- und Autorvergleichstext verwendet werden, ist die Qualität der Wortlisten etwas schlechter. Im Evaluationstest werden nicht nur im Ratekorpora kleine Bag-of-Words mit 2.000 Wortformen verwendet, sondern auch im Vergleichskorpus, wodurch die Erkennungsquoten deutlich schlechter werden. Die Parameter für die Author-Recall-Ermittlung sind: 250 Stichproben bei 400 MFWs, im Vergleichskorpus befinden sich neben dem Autorvergleichstext 24 Texte von anderen Autoren. Die False-Positives werden hier mit je zwei Texten von

allen anderen 31 Autoren als Ratetexte in 32 Tests gegen das Vergleichskorpus ermittelt, in dem neben 24 Texten von anderen Autoren reihum alle 32 Autoren mit einem Autorvergleichstext vertreten waren (Parameter: Bag-of-Words mit 2.000 Wortformen, 400 MFWs, 100 Stichproben).

| | Author Recall | Other Recall 1. Ratetextautor im Vergleichskorpus vorhanden | Other Recall 2. Ratetextautor im Vergleichskorpus nicht vorhanden |
|----------------------|---------------|--|--|
| MFW400 (Baseline) | 0,48 | 0,89 | 0,89 |
| A | 0,55 | 0,91 | 0,86 |
| B | 0,56 | 0,92 | 0,89 |
| C | 0,69 | 0,77 | 0,67 |
| D | 0,57 | 0,88 | 0,85 |

Auch hier übertrifft der Gewinn bei der Verbesserung der Erkennungsquote die Verschlechterung bei den False-Positives. Wenn sich der fragliche Autor im Vergleichskorpus befindet, ergibt sich sogar bei den False-Positives bei A) und B) eine leichte Verbesserung. Bei C) ist die Verschlechterung bei den False-Positives problematisch.

5. Fazit

Die Gute-Wörter-Verfahren führen zu besseren Erkennungsquoten; in geringem Umfang treten mehr False-Positives auf, doch sind die positiven Effekte größer als die negativen. Die detaillierten Zahlen, für die hier kein Platz verfügbar ist, zeigen, dass die False-Positives je nach Autor erheblich schwanken. In einem Anwendungsfall kann vorab getestet werden, ob es bei dem fraglichen Autor zu erhöhten False-Positives kommt – um zu prüfen, ob das Verfahren bei diesem Autor gut anwendbar ist. Künftig wollen wir untersuchen, ob ein mittels Maschinellern erstelltes autorspezifisches Vokabular zu weiteren Verbesserungen führt.

Fußnoten

1. Verwendet werden alle Romane aus dem Textgrid-Repository bzw. von gutenber.org, bei denen mindestens sechs Texte eines Autors vorliegen. Aussortiert wurden Novellensammlungen etc. sowie dialektal geprägte Texte. Es verbleiben 280 Texte von 32 Autoren.
2. Eine ausführliche Darstellung des Evaluationsverfahrens findet sich in Dimpel 2017b.
3. Diese Parameter haben sich in Prätests als vielversprechend erwiesen. Mit Blick auf die Rechenzeit muss es zunächst bei vier Varianten bleiben – für die False-Positives-Experimente in Testreihe 3 war pro Wert eine mehrstündige Rechenzeit nötig.

Bibliographie

Büttner, Andreas / Dimpel, Friedrich Michael / Evert, Stefan / Jannidis, Fotis / Pielström, Steffen / Proisl, Thomas / Reger, Isabella / Schöch, Christof / Vitt, Thorsten (2016): „Delta“ in der stilometrischen Autorschafts Attribution, in: *Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma*. Konferenzabstracts zur DHd-Tagung 2016 in Leipzig, <http://dhd2016.de/>, S. 61–74

Büttner, Andreas / Dimpel, Friedrich Michael / Evert, Stefan / Jannidis, Fotis / Pielström, Steffen / Proisl, Thomas / Reger, Isabella / Schöch, Christof / Vitt, Thorsten (2017): „„Delta“ in der stilometrischen Autorschafts Attribution“, in: *Zeitschrift für digitale Geisteswissenschaft* 2017. DOI: 10.17175/2017_006.

Burrows, John (2002): „‘Delta’: A Measure of Stylistic Difference and a Guide to Likely Authorship“, in: *Literary and Linguistic Computing* 17/3: 267–87. 10.1093/llc/17.3.267.

Dimpel, Friedrich Michael (2016): „Burrows’ Delta im Mittelalter: Wilde Graphien und metrische Analysedaten“, in: *Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma*. Konferenzabstracts zur DHd-Tagung 2016 in Leipzig, <http://dhd2016.de/>: 65–70.

Dimpel, Friedrich Michael (2017a): „Autorschafts Attribution bei nicht-normalisiertem Mittelhochdeutsch. Bessere Erkennungsquoten durch ein Normalisierungswörterbuch“, in **Stolz, Michael (Hrsg.):** *Konferenzabstracts DHd 2017 Bern. Digitale Nachhaltigkeit*. Bern: 100–103. <http://www.dhd2017.ch/programm>.

Dimpel, Friedrich Michael (2017b): „Ein Delta-Rätsel: Nicht-normalisierte mittelhochdeutsche Texte, Z-Wert-Begrenzung und ein Normalisierungswörterbuch. Oder: Auf welche Wörter kommt es bei Delta an?“, in: *Dariah-de-Working Papers* 25, 2017, resolver.sub.uni-goettingen.de/purl/, <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2017-5-1>.

Dimpel, Friedrich Michael (2018a): *Die guten ins Töpfchen: Zur Anwendbarkeit von Burrows’ Delta bei kurzen mittelhochdeutschen Texten nebst eines Attributionstests zu Konrads ‚Halber Birne‘*, in: **Georg Vogeler (Hg.):** *Kritik der digitalen Vernunft. Konferenzabstracts DHd 2018*, Köln 2018, <http://dhd2018.uni-koeln.de>, S. 168–173.

Dimpel, Friedrich Michael (2018b): „Stabile Autorschaft trotz handschriftlicher Varianz? Die Erfolgsquote von Burrows’ Delta bei nicht-normalisierten mittelhochdeutschen Texten optimieren“, in: *ZfdA* 147, 2018, S. 341–363.

Eder, Maciej (2013a): „Mind Your Corpus: systematic errors in authorship attribution“, in: *Literary and Linguistic Computing* 28:603–614. 10.1093/llc/fqt039.

Eder, Maciej (2013b): „Does size matter? Authorship attribution, small samples, big problem“, in: *Literary and Linguistic Computing* Advanced Access 29:1-16. 10.1093/llc/fqt066.

Evert, Stefan / Proisl, Thomas / Jannidis, Fotis / Pielström, Steffen / Schöch, Christof / Vitt, Thorsten (2015): „Towards a better understanding of Burrows's Delta in literary authorship attribution“, in: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. Denver, CO: Association for Computational Linguistics: 79–88. 10.5281/zenodo.18177. <http://www.aclweb.org/anthology/W/W15/W15-0709.pdf> [Abruf 20.8.2015].

Evert, Stefan / Proisl, Thomas / Jannidis, Fotis / Pielström, Steffen / Reger, Isabella/ Schöch, Christof / Vitt, Thorsten (2016): „Burrows Delta verstehen“, in: *Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma*. Konferenzabstracts zur DHd-Tagung 2016 in Leipzig, <http://dhd2016.de/>: 61–65.

Jannidis, Fotis / Lauer, Gerhard (2014): „Burrows's Delta and Its Use in German Literary History“ in: **Erlin, Matt / Tatlock, Lynne (eds.):** *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. New York: 29–54.

Jannidis, Fotis / Pielström, Steffen / Schöch, Christof / Vitt, Thorsten (2015): „Improving Burrows' Delta – An Empirical Evaluation of Text Distance Measures“, in: *Digital Humanities Conference 2015*, Sydney. http://dh2015.org/abstracts/xml/JANNIDIS_Fotis_Improving_Burrows_Delta__An_empirical_.html .

Schöch, Christof (2014): „Corneille, Molière et les autres. Stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik“ in: **Christof Schöch und Lars Schneider (Hrsg.):** *Literaturwissenschaft im digitalen Medienwandel*, Berlin (Philologie im Netz, Beiheft 7): 130–157.