

# Die guten ins Töpfchen: Zur Anwendbarkeit von Burrows' Delta bei kurzen mittelhochdeutschen Texten nebst eines Attributionstests zu Konrads ‚Halber Birne‘

**Dimpel, Friedrich Michael**

mail@dimpel.de

FAU Erlangen-Nürnberg, Deutschland, Germanistik, und  
TU Darmstadt, Computerphilologie und Mediävistik

## Einleitung

Die Anwendbarkeit von Burrows' Delta (Burrows 2002) als Autorschaftstest für das Deutsche ist in Validierungstestreihen wiederholt eindrucksvoll demonstriert worden (Büttner et alia 2017, Eder 2013a/b, Evert et alia 2015, Evert et alia 2016); auch im Mittelhochdeutschen ist Delta anwendbar (Dimpel 2016/2018). Die Stabilität des Verfahrens wurde in Noise-Tests belegt: Wenn man etwa 12% aller Wörter durch Fremdmaterial austauscht, sinkt die Erkennungsquote kaum (Dimpel 2017a/2018).

Bei nicht-normalisierten mittelhochdeutschen Texten steigt die Erkennungsquote in einem Validierungstest von 80% auf 91%, wenn man die bei Evert et alia (2016) entwickelte Methode der Z-Wert-Begrenzung mit einem von mir zusammengestellten Normalisierungswörterbuch kombiniert (Dimpel 2017a). Kontraintuitiv ist, dass nur die Kombination dieser Optimierungsverfahren zu einer Verbesserung um 11% führt, während in diesem Setting nur der Einsatz der Z-Wert-Begrenzung zu einer minimalen Verschlechterung führt; der Einsatz nur des Normalisierungswörterbuchs führt nur zu einer Verbesserung um 5,6%. Dieser Befund wird unter dem Stichwort „Delta-Rätsel“ in einem Dariah-de-Working-Paper (Dimpel 2017b) ausführlich analysiert. Bei der Rätsel-Analyse wurde – ein Serendipitätseffekt – eine Möglichkeit entdeckt, wie man bei einem konkreten Vergleich von drei Texten die Wortformen identifizieren kann, die eine korrekte Autorschaftserkennung begünstigen oder behindern – dazu im Weiteren.

## Gute und schlechte Wortformen

Beim Delta-Test berechnet man aus den Wortfrequenzen für ein Korpus jeweils die zugehörigen Z-Werte. Beim Vergleich von zwei Wortformen aus zwei Texten wird

die Differenz der jeweiligen Z-Werte gebildet und der Betrag dieser Differenz genommen. Delta ist schließlich der Mittelwert der absoluten Z-Wert-Differenzen für alle Wortformen.

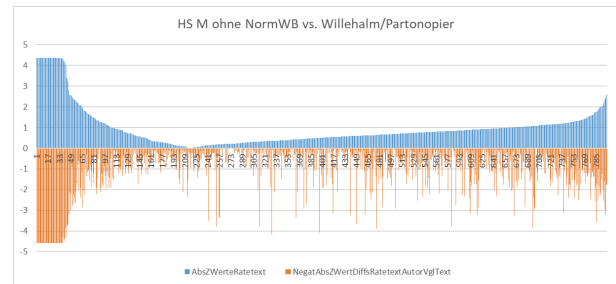


Abb. 1: Ratetext-Z-Werte (blau) sowie Z-Wert-Differenzen Ratetext–Autor-Vergleichstext (orange)

Abb. 1. zeigt oben die Z-Werte der Handschrift M von Wolframs ‚Parzival‘ in einem Test, in dem sich im Vergleichskorpus neben Wolframs ‚Willehalm‘ noch weitere 19 Distraktortexte von anderen Autoren befinden (ausführlich zum Testverfahren Dimpel 2017b). Der ‚Parzival‘ soll dem Autor-Vergleichstext (Wolframs ‚Willehalm‘) zugeordnet werden und nicht etwa Konrads ‚Partonopier‘. Im oberen linken Viertel sind positive Z-Werte blau aufgetragen und nach der Höhe der Z-Werte sortiert. Ab der Stelle, an der die blauen Balken auf 0 zurückgehen, folgt rechts der Betrag der negativen Z-Werte (blau). Unten stehen (orange) die absoluten Z-Wert-Differenzen zwischen dem Ratetext und dem Autor-Vergleichstext (Differenzen der Z-Werte von Wolframs ‚Parzival‘ und Wolframs ‚Willehalm‘).

Man könnte A) den Verdacht haben, dass Wortformen bei hohen blauen Balken „gut“ sind, um einen Text von Distraktortexten zu unterscheiden, da hohe Z-Werte auf erhebliche Abweichung von den übrigen Korpusfrequenzen hindeuten. Man könnte auch B) den Verdacht haben, dass Wortformen bei hohen orangen Balken „schlecht“ für die Autorenerkennung sind: Unterschiede zwischen dem Ratetext und Autor-Vergleichstext (also Unterschiede von zwei Texten des gleichen Autors) sollten eher niedrig sein, damit die Erkennung funktioniert. Allerdings sind bei hohen blauen Balken relativ oft auch hohe orange Balken vorhanden – auch in anderen Tests (Dimpel 2017b). Dieses Diagramm erlaubt also keine Aussage darüber, welche Wortformen gut für die Autorenerkennung sind; hohe Z-Werte allein erlauben noch keine Aussage darüber, ob ein Wort hier gut geeignet ist, um einen Autor zu charakterisieren.

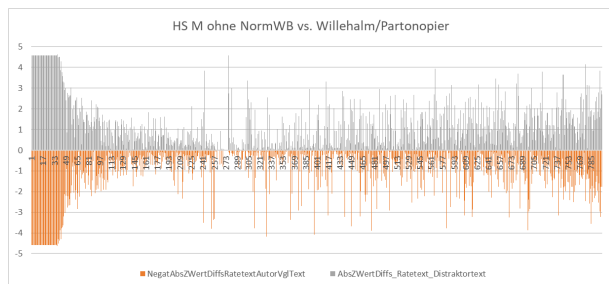


Abb. 2: Z-Wert-Differenzen Ratetext–Autor-Vergleichstext (orange: Wolframs ‚Parzival‘– Wolframs ‚Willehalm‘) und Z-Wert-Differenzen Ratetext–Distraktortext (grau: Wolframs ‚Parzival‘– Konrads ‚Partonopier‘)

Neu ist in Abb. 2 nur die obere Hälfte: Sie enthält Z-Wert-Differenzen des Ratetexts zum Distraktortext (‚Partonopier‘). Diese grauen Unterschiede sollten bei funktionierender Autorenerkennung eher groß sein; gleichzeitig sollten die orangen Unterschiede der Texte vom gleichen Autor niedriger sein als die grauen. Dort, wo die grauen Balken genauso hoch sind wie die orangen, hilft das Wort nicht bei der Autorenerkennung – dies ist bei sehr hohen positiven Z-Werten der Fall. Sind die orangen Balken höher als die grauen, stört die Wortform die Autorenerkennung: Die Differenzen zwischen Texten verschiedener Autoren müssen größer sein als die Differenzen zwischen Texten gleicher Autoren, wenn die Autorschaftserkennung funktioniert.

Die Differenz zwischen orange und grau sei ‚Level-2-Differenz‘ genannt: „Differenz aus der Z-Wert-Differenz zwischen Ratetext und Distraktortext einerseits und der Z-Wert-Differenz zwischen Ratetext und Autor-Vergleichstext andererseits“. Bei positiven Level-2-Differenzen ist eine Wortform vorteilhaft für die Autorenerkennung – mit Blick auf den einen untersuchten Distraktortext. Bei negativen Level-2-Differenzen ist die Wortform schlecht für die Autorenerkennung. Über diese Differenz kann man „gute“ und „schlechte“ Wortformen einzeln identifizieren.

## Use-Case-Szenario ‚Halbe Birne‘

Konrads Autorschaft wurde der ‚Halben Birne‘ trotz Selbstnennung im Epilog ( von *Wirzburg maister Kuonrat*) abgesprochen (Lachmann 1820, Laudan 1906, de Boor 1973, de Boor / Janota 1997; ‚Konrad‘ mit Fragezeichen bei Grubmüller 1996) – aufgrund des „obszönen“ Inhalts und sprachlicher Merkmale; anders Feistner 2000.

Die stilometrische Analyse ist in mehrfacher Hinsicht eine Herausforderung: Eine gattungsübergreifende Attribution ist mangels anderer Vergleichstexte nötig (nach Schöch 2014 wäre eine Gattungsmischung möglichst zu meiden). In Konrads Oeuvre herrscht eine Vielfalt an Themen, Frivoles wie in der ‚Halben Birne‘ ist eher

selten – auch im einzigen anderen Märentext Konrads: im ‚Herzmäre‘ bleibt die Liebe unerfüllt, es kommt zum doppelten Minnetod. Zudem ist die ‚Halbe Birne‘ recht kurz: sehr gute Quoten erreicht Delta ab 5.000 Wortformen in einer Bag-of-Words (vgl. Abb. 3 sowie Eder 2013a und Eder 2013b).

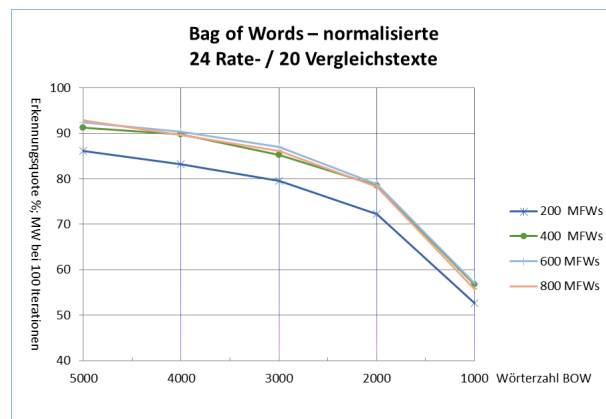


Abb. 3: zum Setting vgl. Dimpel 2018.

Die ‚Halbe Birne‘ enthält jedoch nur 2.469 Wortformen. Wenn man nun die ‚Birne‘ gegen ein Konrad-Korpus testet, kann man entweder die Wörter mit hoher Level-2-Differenz, die einer Erkennung von Konrad entgegenstehen, aus der Liste der untersuchten Most-Frequent-Words (MFWs) streichen. Oder man kann eine Positivliste mit „guten“ Wörtern bevorzugt verwenden – Wörter mit hoher positiver Level-2-Differenz.

Vorab wird das Verfahren validiert: In einer Ermittlungsgruppe (vier Konrad-Texte) werden „gute“ und „schlechte“ Wörter identifiziert.<sup>1</sup> In einer Kontrollgruppe (vier andere Konrad-Texte) zeigt sich, dass die Erkennungsquote durch dieses Verfahren bei Bag-of-Words mit 2.000 Wortformen steigt – beim bevorzugten Verwenden „guter“ Wörtern stärker als beim Aussortieren der „schlechten“. Danach werden alle acht Konrad-Texte erneut zur Bildung der Listen der „guten“ und „schlechten“ Wortformen herangezogen. Als geeignete Parameter haben sich gezeigt:

„Gute Wörter“: Level-2-Differenzen  $> +2,31$  in 6 von 7 Ermittlungsgruppen-Ratetexten, 304 items

„Schlechte Wörter“: Level-2-Differenzen  $< -1,2$  in 2 von 7 Ermittlungsgruppen-Ratetexten, 174 items

Attributionstest 1: Halbe Birne und Herzmäre im Vergleichskorpus vs. Konrad-Ratekorpus

Erkennungsquote in %. Parameter: Bag-of-Words mit 2.000 Wortformen und mit 100 Iterationen, 400 MFWs, mit Normalisierungswörterbuch, Pronomina-Entfernung

	Halbe Birne	Herzmäre
Ohne Wortliste	28,6%	4,5%
Negativliste: Schlechte Wörter	35,8%	10,3%
Positivliste: Gute Wörter	83,8%	42,5%

Im Attributionstest 1 wird die ‚Halbe Birne‘ als Autor-Vergleichstext verwendet, als Ratetexte werden die acht Konrad-Texte sowie das ‚Herzmäre‘ verwendet; im ‚Herzmäre‘-Test bleibt es bei acht Konrad-Ratetexten; das ‚Herzmäre‘ ist Autor-Vergleichstext. Hier erreicht das ‚Herzmäre‘ nur 4,5%, ein schlechter Wert, obwohl hier die Autorschaft nicht infrage gestellt wurde. Dagegen liegt die Erkennungsquote bei der ‚Halben Birne‘ auch ohne zusätzliche Wortlisten bereits über dem Zufallswert: Wenn ein Konrad-Text aus dem Ratekorpus nun nicht einem der 20 Texte von anderen Autoren zuordnet wird, sondern der ‚Halben Birne‘, dann stehen die Chancen dafür 1 zu 21. Wenn es also auf den Zufall zurückzuführen wäre, dass ein Text dem richtigen Autor zugeordnet wird, dann müsste die Erkennungsquote bei 5% liegen – so beim ‚Herzmäre‘. 83,8% bei der ‚Halben Birne‘ sind ein ordentlicher Wert, wenn man bedenkt, dass nur kurze Bag-of-Words mit 2.000 Wortformen getestet werden können und dass gattungsübergreifend getestet wird.

Beim Attributionstest 1 befand sich die ‚Halbe Birne‘ im Vergleichskorpus. Im Ratekorpus waren inklusive ‚Herzmäre‘ 9 Konrad-Texte. Nun werden umgekehrt ‚Halbe Birne‘ bzw. ‚Herzmäre‘ als Ratetexte verwendet. Ins Vergleichskorpus gebe ich zu den 20 Distraktortexten in separaten Tests jeweils einen Konrad-Text als Autor-Vergleichstext ins Vergleichskorpus.

Attributionstest 2:

Autor-Vergleichstext	Ratetext Halbe Birne	Ratetext Herzmäre
Alexius	99,0	89,0
Engelhard	69,0	97,0
Turnier von Nantes	99,0	3,0
Pantaleon	94,0	88,0
Partonopier	86,0	89,0
Schwanritter	97,0	87,0
Silvester	95,0	100,0
Trojan. Krieg	94,0	71,0
Herzmäre	2,0	

Im Attributionstest 2 übersteigen die meisten Werte 86%. Es gibt lediglich zwei deutliche Ausreißer, an denen jeweils das ‚Herzmäre‘ beteiligt ist. Dieses Minneleid-und-Minnetod-Märe fügt sich nicht zur politischen Propagandadichtung ‚Turnier von Nantes‘. Auch zur ‚Halben Birne‘ passt das ‚Herzmäre‘ nicht: Dort geht es um eine Dame, die einen Ritter abweist, weil er beim Birnenverzehr keine Tischmanieren an den Tag legt. Die Dame schläft mit einem vermeintlich taubstummen Hofnarren, der sich jedoch später als der abgewiesene Birnen-Ritter entpuppt. Interessante Fehlattraktionen (etwa ‚Birne‘ zu ‚Häselein‘ statt zum ‚Herzmäre‘) werden im Vortrag vorgestellt.

## Ein kleiner Schritt für die Attribution der ‚Halben Birne‘ an Konrad

Als Katharina Zeppezauer-Wachauer (Salzburg) mir einige Mären aus der Mittelhochdeutschen Begriffsdatenbank überlassen hat (vielen Dank dafür!), hat sie notiert: „Vielleicht können Sie ja wirklich, wie Edith Feistner gefordert hat, ‚Konrad seine Birne wiedergeben!‘“ Auch wenn die Zahlen in beiden Attributionstests trotz der geringen Textlänge und trotz der Gattungsproblematik überraschend eindeutig sind, möchte ich bei einer vorsichtigen Interpretation bleiben. Zwar ist die Wahrscheinlichkeit sehr gering, dass die gefundene Nähe der ‚Halben Birne‘ zum Konrad-Korpus auf dem Zufall beruht. Allerdings wären ‚Kontrollpeilungen‘ (Eibl 2013) wünschenswert: Eine Attribution sollte nicht auf einem einzelnen Test mit einer Methode erfolgen, wünschenswert wären Bestätigungen mit anderen Methoden. Immerhin aber geht es hier nicht um eine blinde Attribution, sondern lediglich um Widerspruch gegen eine Athetese der Forschung. Eine Attribution stünde in Einklang mit Konrads Selbstnennung in fünf von sieben überlieferten Textzeugen.

Zudem würde ich den Test gerne mit einem größeren Mären-Korpus wiederholen, in dem idealerweise längere Texte wären und mehr Texte, die näher an Konrads Schaffenszeit liegen. Dass die Birne nicht zu Kaufringer clustert, könnte auch dem zeitlichen Abstand geschuldet sein, der durch gemeinsame groteske oder frivole Inhaltselemente nicht überlagert wird.

Wichtig ist mir auch das Verfahren: Bislang ist eine Feature-Eliminierung oder Feature-Selektion häufig auf dem Weg des maschinellen Lernens erfolgt (Büttner et alia 2016) – mit dem Nachteil, dass der Weg der Kategorisierung teilweise im Dunklen bleibt. Ermittelt man „gute“ oder „schlechte“ Wörter via Level-2-Differenzen, so ist transparent, wie man zu den Parametern kommt und wie auf dieser Basis die weiteren Berechnungen erfolgen.

## Fußnoten

1. Im Vergleichskorpus verwende ich hier und für die folgenden Attributionstests 7 Romane und 13 Mären: Barlaam, Daniel, Lanzelet, Meleranz, Parzival, Tristan, Wigalois; Frauentreue, Haeslein, Heidun\_B, JvFreiberg\_Raedlein, Kaufringer\_Moerderin, Kaufringer\_Rache, Kaufringer\_listige\_Frauen, Pyramus, Rosenpluet\_Pfarrer, Schlegel, Schueler\_Paris, StudentenAbenteuer\_A, Zwickauer\_Moenches\_Not.

## Bibliographie

- Büttner, Andreas / Dimpel, Friedrich Michael / Evert, Stefan / Jannidis, Fotis / Pielström, Steffen / Proisl, Thomas / Reger, Isabella / Schöch, Christof / Vitt, Thorsten** (forthcoming 2017): „Delta“ in der stilometrischen Autorschaftsattributions“, in: *ZfdG*.
- 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/lc/17.3.267.
- De Boor, Helmut** (1967): „Die Chronologie der Werke Konrads von Würzburg, insbesondere die Stellung des Turniers von Nantes“, in: *PBB* 89: 210–269.
- De Boor, Helmut / Janota, Johannes** (1997): *Geschichte der deutschen Literatur von den Anfängen bis zur Gegenwart. Band III /1. Die deutsche Literatur im späten Mittelalter: Epik, Lyrik, Didaktik, geistliche und historische Dichtung: 1250–1350*, 5., neubearb. Aufl. von Johannes Janota. München.
- 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): „Autorschaftsattributions 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** (forthcoming 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* n.n.
- Dimpel, Friedrich Michael** (forthcoming 2018): „Stabile Autorschaft trotz handschriftlicher Varianz? Die Erfolgsquote von Burrows' Delta bei nicht-normalisierten mittelhochdeutschen Texten optimieren“ (in Begutachtung, n.n.).
- Eder, Maciej** (2013a): „Mind Your Corpus: systematic errors in authorship attribution“, in: *Literary and Linguistic Computing* 28:603–614. 10.1093/lc/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/lc/fqt066.
- Eibl, Karl** (2013): „Ist Literaturwissenschaft als Erfahrungswissenschaft möglich? Mit einigen Anmerkungen zur Wissenschaftsphilosophie des Wiener Kreises“, in: *Philip Ajouri [u. a.]* (Hrsg.): *Empirie in der Literaturwissenschaft, Münster* (Poetogenesis. Studien zur empirischen Anthropologie der Literatur 8): 19–45.
- 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.
- Edith Feistner** (2000): „Kulinarische Begegnungen. Konrad von Würzburg und ‚Die halbe Birne‘“, in: *Dorothea Klein et al. (Hrsg.): Vom Mittelalter zur Neuzeit. FS Horst Brunner*. Wiesbaden: 291–304. Grubmüller, Klaus (1996): *Novellistik des Mittelalters. Märendichtung*. Frankfurt/Main 1996 (Bibliothek des Mittelalters 23).
- 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\\_emi/JANNIDIS\\_Fotis\\_Improving\\_Burrows\\_Delta\\_An\\_empirical\\_.html](http://dh2015.org/abstracts/xml/JANNIDIS_Fotis_Improving_Burrows_Delta_An_emi/JANNIDIS_Fotis_Improving_Burrows_Delta_An_empirical_.html).
- Lachmann, Karl** (1820): *Auswahl aus den Hochdeutschen Dichtern des dreizehnten Jahrhunderts*, Berlin.
- Laudan, Hans** (1908): „Die Halbe Birne“ nicht von Konrad von Würzburg“, in: *ZfdA* 50: 158–166.
- 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.