Wiederholende Forschung in den digitalen Geisteswissenschaften

Schöch, Christof
christof.schoech@uni-wuerzburg.de
Universität Würzburg, Deutschland

Einleitung

Die Reproduzierbarkeit von Forschungsarbeiten ist in zahleichen Disziplinen ein drängendes und vieldiskutiertes Problem. Laut einer *Nature*-Umfrage nehmen 52% der befragten ForscherInnen eine “significant reproducibility crisis” wahr (Baker 2016). Metastudien aus der Psychologie (Bohannon 2015) oder den Wirtschaftswissenschaften (Camerer 2016) berichten von niedrigen Reproduzierbarkeitsquoten. Forderungen nach reproduzierbarer Forschung werden nicht nur in der Informatik (Mesirov 2010, Peng 2010) formuliert. Insbesondere für die empirisch und ggfs. quantitativ arbeitenden Teile der digitalen Geisteswissenschaften sind diese Debatten relevant (Padilla und Higgins 2016).

Hier stehen allerdings nicht die Anforderungen an wiederholbare Forschung im Fokus, sondern umgekehrt die Herausforderungen, vor denen wiederholende Forschung steht. Letztere ist in den digitalen Geisteswissenschaften in besonderem Maße aufschlussreich, stellt doch der Paradigmenwechsel von dominant hermeneutischen zu dominant empirischen Methoden in den Geisteswissenschaften die Kontinuität des wissenschaftlichen Diskurses auf eine Zerreißprobe. Die digitalen Geisteswissenschaft sind gefordert, die eigene Anschlussfähigkeit an etablierte Konzepte, Fragestellungen und Erkenntnisziele sicherzustellen. Studien, die vorhandene Arbeiten mit digitalen Mitteln wiederholen, platzieren diese Kontinuitätsfrage gewissermaßen unter einem Mikroskop. Zudem treten im praktischen Nachvollzug einer Originalstudie die (teils impliziten) Annahmen sowie die Stärken und Grenzen beider Ansätze plastisch hervor. So versprechen Wiederholungsstudien inhaltlichen ebenso wie methodischen Erkenntnisgewinn (vgl. Rockwell 2016).

Auf eine konzeptuellen und begrifflichen Klärung zum beschrieben Problemfeld der wiederholenden Forschung folgen im hier skizzierten Beitrag zwei unterschiedliche literaturwissenschaftliche Fallstudien, in denen vorhandene Forschungsbeiträge mit digitalen Daten und Methoden wiederholt worden sind.

Typen wiederholender Forschung

Für die vielfältigen Beziehungen zwischen einer bereits vorliegenden Studie und einer diese wiederholenden Studie sind in der Forschungsliteratur zahlreiche Begriffe vorgeschlagen worden, darunter insbesondere Replikation, Reproduktion und Reanalyse (Drummond 2009, Gomez und Juristo 2010). Zur konzeptuellen Klärung werden hier drei wesentlicheAspekte berücksichtigt: die Fragestellung, die Daten und die Analysemethode. Wiederholungsstudien unterscheiden sich, je nachdem ob Fragestellung, Daten und Methoden gegenüber der Originalstudie identisch oder verändert sind (Abbildung 1).

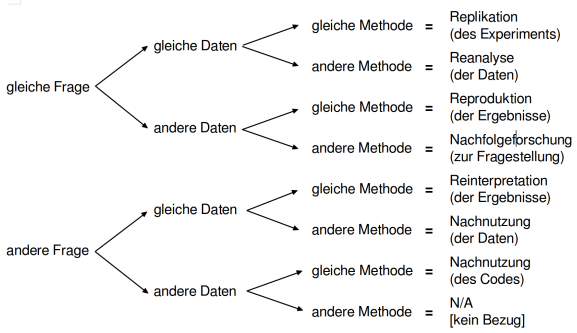


Abbildung 1: Das konzeptuelle und begriffliche Feld der wiederholenden Forschung.

Der Begriff "Replikation" bezeichnet hier die exakte Wiederholung einer Studie. Die gleiche Forschungsfrage wird mit gleicher Datengrundlage und gleichen Methoden erneut bearbeitet. Ziel ist es zu prüfen, ob die gleichen Ergebnisse ermittelt werden können, was ein Hinweis auf die korrekte Durchführung der Originalstudie ist.

Der Begriff "Reproduktion" bezeichnet eine freiere Wiederholung. Die gleiche Fragestellung wird mit den gleichen Analysemethoden, aber neu erhobenen oder erweiterten Daten durchgeführt. Ziel ist es zu prüfen, ob die Analysemethode auch mit veränderten Daten die gleichen Schlussfolgerungen erlaubt, d.h. ob die Ergebnisse generalisierbar sind.

Der Begriff "Reanalyse" bezeichnet ebenfalls eine freiere Wiederholung. Hier wird die gleiche Fragestellung mit den gleichen Daten, aber einer anderen (bspw. verbesserten oder neu implementierten) Analysemethode bearbeitet. Wird die gleiche Fragestellung sowohl mit anderen Daten als auch mit anderen Methoden bearbeitet, kann man von “Nachfolgeforschung” sprechen.

Auch wenn eine veränderte Fragestellung im Fokus steht, kann ein Bezug zu einer früheren Studie bestehen. Die Bearbeitung einer veränderten Fragestellung mit den gleichen Daten und der gleichen Methode kann als “Reinterpretation” der Ergebnisse aus einer anderen Perspektive verstanden werden. Der erneute Einsatz von Daten oder Code aus einer früheren Studie für

die Bearbeitung einer neuen Fragestellung ist eine “Nachnutzung”. Kein (hier wesentlicher) Bezug besteht, wenn Fragestellung, Daten und Code gegenüber einer früheren Studie verändert wurden.

Die folgenden beiden Fallstudien beziehen sich auf sehr unterschiedliche Originalstudien, illustrieren die spezifischen Herausforderungen, die jeweils hiermit zusammenhängen und werfen ein Schlaglicht auf das Verhältnis der digitalen Geisteswissenschaften zu früherer Forschung.

Erste Fallstudie: Richeaudeau zur Satzlänge bei Georges Simenon

Die erste Fallstudie bezieht sich auf die Wiederholung einer Studie von François Richeaudeau zur Satzlänge im umfangreichen Werk des belgischen Autors Georges Simenon. Die 1982 veröffentlichte Studie ist quantitativ angelegt, wurde allerdings nicht computergestützt durchgeführt. Zentrale These ist, dass Simenons Romanwerk sich durch die Verwendung besonders kurzer Sätze auszeichne. Dies wird als ein Faktor unter anderen interpretiert, der zum weltweiten Erfolg des Autors beigetragen hat (Richeaudeau 1982).

Obwohl in diesem Fall die Textsammlung bekannt und das verwendete Verfahren quantitativ ist, kann nur in Ansätzen eine Replikation der Studie (im oben definierten Sinne) vorgenommen werden. Beispielsweise ist nicht dokumentiert, wie Satz und Wort für die Messung der Satzlänge definiert sind. Dies musste neu entschieden und implementiert werden. Die erneute Messung der Satzlängen in den 25 von Richeaudeau untersuchten Texten Simenons anhand einer einfachen, aber angemessen erscheinenden Definition von Satz und Wort ergibt um durchschnittlich 15% niedrigere Werte (siehe Abbildung 2; Details in Schöch 2016). Das scheint zunächst Richeaudeaus These sogar noch zu stärken.

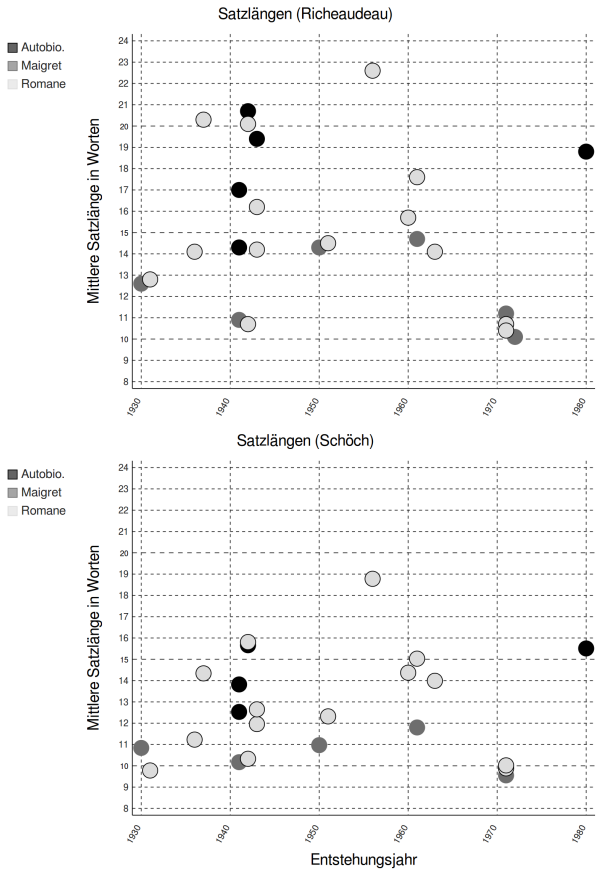


Abbildung 2: Von Richeaudeau (oben) und in der Wiederholungsstudie (unten) erhobene Satzlängen unter Verwendung der gleichen Texte.

Allerdings wird deutlich, dass 25 Werken nicht ausreichen, um Richeaudeaus weiterführende Thesen einer Entwicklung Simenons’ Stils über die Zeit (hin zu zunehmend kürzeren Sätzen in den Romanen) sowie in Abhängigkeit der von ihm praktizierten Gattungen (längere Sätze in den autobiographischen Schriften als in den Romanen) zu prüfen. Erst mit deutlich mehr Werken (hier 127 Texte) und mit Hilfe eines statistischen Signifikanztests, kann die erste dieser Thesen geprüft und widerlegt werden, die zweite dieser Thesen dagegen klar bestätigt werden (Abbildung 3).

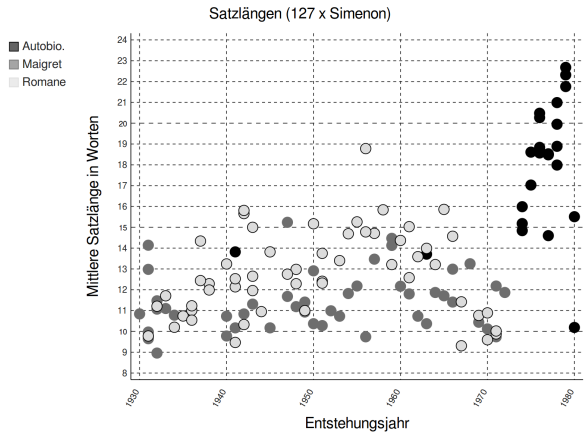


Abbildung 3: Satzlängen für 127 Werke Simenons in drei Gattungen: autobiographische Werke (schwarz), Maigret-Romane (grau), psychologische Romane (weiß). Statistisch massiv signifikanter Unterschied zwischen Romanen und autobiographischen Werken.

Zudem verfügt Richeaudeau als Vergleichsmaßstab nur über Zahlen aus einer Einzelstudie zu Marcel Proust, im Vergleich zu dessen langen Sätzen Simenons Sätze kurz erscheinen müssen. Der Vergleich mit 195 französischen Romanen, die wie Simenons Werke zwischen 1930 und 1980 erschienen sind, zeigt hingegen, dass es zwar einige wenige Romanciers gibt, die deutlich längere Sätze verwenden als Simenon, dieser aber keinesfalls ungewöhnlich kurze Sätze verwendet (Abbildung 4).

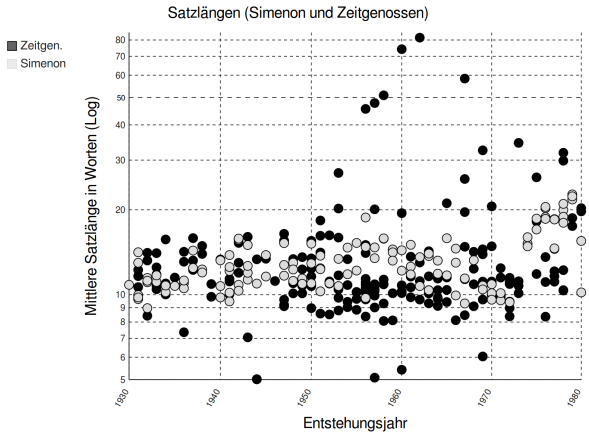


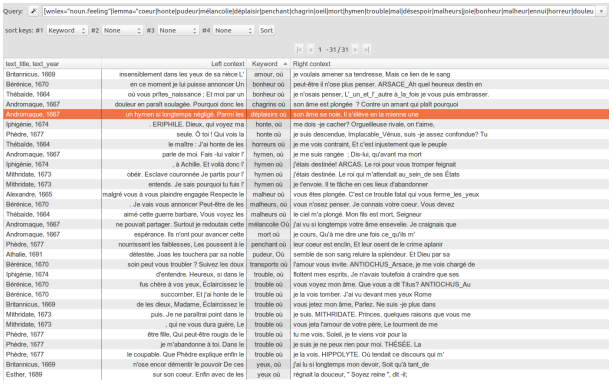
Abbildung 4: Satzlänge bei Georges Simenon (weiß) und in 195 zeitgenössischen Romanen (schwarz). Kein statistisch signifikanter Unterschied.

Abschließend kann festgehalten werden, dass hier weniger eine methodische Kluft überwunden werden musste, als vielmehr mangelnde Dokumentation des eingesetzten Verfahrens eine Herausforderung darstellt. Anders in der folgenden Fallstudie.

Zweite Fallstudie: Spitzers Stilanalyse des Werks Jean Racines

Die zweite Fallstudie bezieht sich auf die Wiederholung einer bis heute viel beachteten Stilanalyse, die der Romanist Leo Spitzer 1928 über den französischen Dramatiker Jean Racine vorgelegt hat. Spitzer verfolgt die These, dass in Racines Tragödien ein stilistischer "Dämpfungseffekt" (als Autorenstil) aufgezeigt werden kann. Offen lässt Spitzer, inwiefern dieser Effekt zugleich paradigmatisch für die Klassik (als Epochenstil) ist. Spitzer unterscheidet rund 50 stilistische Phänomene, die zum “Nüchtern-Gedämpften, Verstandesmäßig-Kühlen, fast Formelhaften” in Racines Stils beitragen. Er beschreibt sie nuancenreich und illustriert sie mit zahlreichen Beispielen. Zur Veranschaulichung seien hier nur einige Definitionen Spitzers zitiert: “die Personifizierung von Abstrakta”, “konturverwischende Plurale” oder “das entgrenzende *où*” (Spitzer 1928).

Für die Reproduktionsstudie stehen die gleichen Texte zur Verfügung, die auch Spitzer verwendet hat, allerdings in digitaler Form und anderen Textausgaben folgend. Spitzers stilistische Phänomene wurden in Form komplexer Suchabfragen, die mit Hilfe der “Corpus Query Processing”-Sprache CQP (http://cwb.sourceforge.net/files/CQP_Tutorial/) formuliert wurden, im Textanalyse-Tool TXM (http://www.textometrie.fr) nachmodelliert und quantifiziert (siehe Abbildung 5). Auch mit Hilfe aufwändiger Annotationen der Texte (morpho-syntaktische sowie semantische Annotation mit WordNet) gelang dies mit zufriedenstellender Genauigkeit nur für 30 der rund 50 von Spitzer analysierten stilistischen Phänomene.



text_title, text_year	Left context	Keyword	Right context
Britannicus, 1669	insensiblement dans les yeux de sa nièce L'	amour, où	je voulais amener sa tendresse, Mais ce lien de le sang
Bérénice, 1670	en ce moment je lui puisse annoncer Un	bonheur où	peut-être il t'ose plus penser. ARSACE_Ah que heureux destin en
Thébaïde, 1664	où vous prîtes_naissance ; Et moi par un	bonheur où	je n'osais penser. L'_un_et_l'_autre à_la_fois je vous puis embrasser.
Andromaque, 1667	douleur en paraît soulagée. Pourquoi donc les	chagrins où	son âme est plongée ? Contre un amant qui plaît pourquoi
Andromaque, 1667	un hymen si longtemps négligé ? Parmi les	déplaisirs où	son âme se noie, Il s'élève en la mienne une
Iphigénie, 1674	ERIPHILE. Dieux, qui voyez ma	honte, où	me dois- je cacher? Orgueilleuse rivale, on l'ame,
Phèdre, 1677	seule. Ô toi ! Qui vois la	honte où	je suis descendue, Implacable_Vénus, suis -je assez confondue? Tu
Thébaïde, 1664	le maître : J'ai honte de les	horreurs où	je me vois contraint, Et c'est injustement que le peuple
Andromaque, 1667	parle de moi. Fais -lui valoir l'	hymen, où	je me suis rangée . Dis-lui, qu'avant ma mort
Iphigénie, 1674	, à Achille. Et voilà donc l'	hymen où	j'étais destinée! ARCAS. Le roi pour vous tromper feignait
Mithridate, 1673	obéir. Esclave couronnée Je partis pour l'	hymen où	j'étais destinée. Le roi qui m'attendait au_sein_de ses États
Mithridate, 1673	entends. Je sais pourquoi tu fuis l'	hymen où	je t'envoie. Il te fâche en ces lieux d'abandonner
Alexandre, 1665	malgré vous à vous plaindre engagée Respecte le	malheur où	vous êtes plongée. C'est ce trouble fatal qui vous ferme_les_yeux
Bérénice, 1670	. Je vais vous annoncer Peut-être de les	malheurs, où	vous n'osez penser. Je connais votre coeur. Vous devez
Thébaïde, 1664	aimé cette guerre barbare, Vous voyez les	malheurs où	le ciel m'a plongé. Mon fils est mort, Seigneur
Andromaque, 1667	ne pouvait partager. Surtout je redoutais cette	mélancolie Où	j'ai vu si longtemps votre âme ensevelie. Je craignais que
Andromaque, 1667	espérance. Ils n'ont pour avancer cette	mort où	je cours, Qu'à me dire une fois ce_qu'ils m'
Phèdre, 1677	nourrissent les faiblesses, Les poussent à le	penchant où	leur coeur est enclin, Et leur osent de le crime aplanir
Athalie, 1691	détestée. Joas les touchera par sa noble	pudeur, Où	semble de son sang reluire la splendeur. Et Dieu par sa
Bérénice, 1670	soin peut vous troubler ? Suivez les doux	transports où	l'amour vous invite. ANTIOCHUS_Arsace, je me vois chargé de
Iphigénie, 1674	d'entendre. Heureux, si dans le	trouble, où	flottent mes esprits, Je n'avais toutefois à craindre que ses
Bérénice, 1670	fus chère à vos yeux, Éclaircissez le	trouble où	vous voyez mon âme. Que vous a dit Titus? ANTIOCHUS_Au
Bérénice, 1670	succomber, Et j'ai honte de le	trouble où	je la vais tomber. J'ai vu devant mes yeux Rome
Britannicus, 1669	de les dieux, Madame, Éclaircissez le	trouble où	vous jetez mon âme, Parlez. Ne suis -je plus dans
Mithridate, 1673	puis. Je ne paraîtrai point dans le	trouble où	je suis. MITHRIDATE. Princes, quelques raisons que vous me
Mithridate, 1673	, qui ne vous dura guère, Le	trouble où	vous jeta l'amour de votre père, Le tourment de me
Phèdre, 1677	être fille, Qui peut-être rougis de le	trouble où	tu me vois. Soleil, je te viens voir pour la
Phèdre, 1677	je m'abandonne à toi. Dans le	trouble où	je suis je ne peux rien pour moi. THÉSÉE. La
Phèdre, 1677	le coupable. Que Phèdre explique enfin le	trouble où	je la vois. HIPPOLYTE. Où tendait ce discours qui m'
Britannicus, 1669	n'ose encor démentir le pouvoir De ces	yeux, où	j'ai lu si longtemps mon devoir, Soit qu'à tant_de
Esther, 1689	sur son coeur. Enfin avec de les	yeux où	régnait la douceur, " Soyez reine ", dit -il;

Abbildung 5: Keyword-in-Context Ansicht der Suchergebnisse einer CQP-Abfrage in TXM für das “entgrenzende où”.

Der über Spitzers Studie hinausgehende Vergleich der Häufigkeiten der Phänomene bei Racine einerseits, in einem Vergleichskorpus zeitgenössischer französischer Verstragödien andererseits, zeigt, dass überhaupt nur drei der 30 Phänomene bei Racine in statistisch signifikanter

Weise überrepräsentiert sind (Abbildung 6). Trotz Spitzers Fokus auf Racine handelt es sich bei den von ihm identifizierten “gedämpften Stil” also gerade nicht um einen für Racine distinktiven Autorenstil, sondern um einen weit verbreitete Epochenstil.

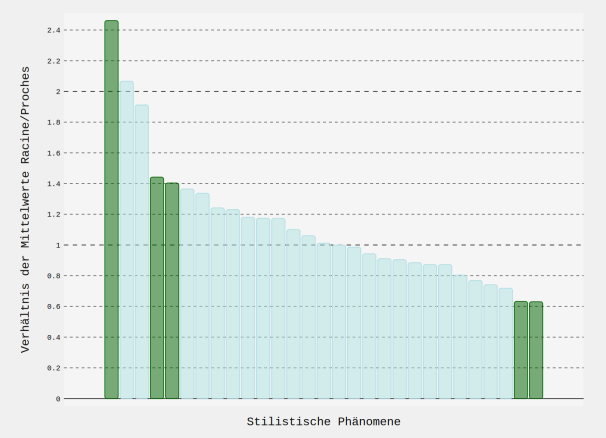


Abbildung 6: Die Häufigkeiten 30 stilistischer Phänomene im Vergleich der Tragödien Racines und 49 zeitgenössischer Tragödien. Werte > 1: bei Racine überrepräsentiert; Werte < 1: bei Racine unterrepräsentiert. Statistisch signifikant abweichende Phänomene sind hervorgehoben.

Bei der Racine-Studie liegt die wesentliche Herausforderung in der algorithmischen Modellierung stilistischer Phänomene, für deren Definition Spitzer subtile semantische Unterscheidungen und kontextuelle Informationen einsetzt, wie sie der algorithmischen Analyse derzeit nur unvollständig zugänglich sind.

Schlussfolgerungen

Beiden Fallstudien zeigen, dass Forschungsarbeiten, die verwendete Daten und Code nicht publizieren, sich nicht für eine Replikation, Reproduktion oder Reanalyse im oben definierten, engen Sinne eignen. Zu Vieles bleibt implizit, wenn das methodische Vorgehen nicht als Solches detailliert dokumentiert wurde. Das gilt auch dann, wenn Fragestellung und Methode prinzipiell einer datengestützten Wiederholung entgegenkommen (wie bei der Satzlängen-Studie).

Zugleich zeigt sich, dass stärker von der Originalstudie abweichende Nachfolgestudien es erst erlauben, eine umfangreichere Datengrundlage zu verwenden und/ oder verbesserte Analysemethoden einzusetzen, wodurch sich die Aussagekraft der Analysen gegenüber der Originalstudie deutlich erhöht. Solche Studien sind zudem hilfreich, um die Anschlussfähigkeit aktueller empirischer und ggfs. quantitativer Methoden in den digitalen Geisteswissenschaften an frühere Forschung zu erproben. Und erst das bewußte, kontrollierte Abweichen von der Originalstudie macht wesentliche Grundannahmen und Erkenntnisinteressen sowohl der Originalstudie als auch der Wiederholungsstudie bewußt, beispielsweise die jeweils unterschiedlichen Stilbegriffe.

Die beiden Fallstudien hinterfragen auch die oben eingeführte binäre Opposition zwischen “identischen” und “veränderten” Fragestellungen, Daten und Methoden im Kontext solcher Wiederholungsstudien. Dennschon der Wechsel von gedruckten Texten zu digitalen Textdaten, selbst bei identischer Korpuszusammenstellung, führt zwar zu vergleichbaren, keinesfalls aber identischen Daten und verlangt auch veränderte Methoden.

Schließlich liegt nahe, dass gerade wiederholende Studien auch selbst dem Anspruch an Reproduzierbarkeit gerecht werden sollten. In diesem Sinne sind zugrundliegende Texte, Metadaten, Code und (teils interaktive) Abbildungen der hier dargestellten Wiederholungsstudien verfügbar, soweit es urheberrechtliche Einschänkungen möglich machen. Siehe https://github.com/cligs/projects (Ordner “2016/ simenon” und “2016/racine”), DOI: http://doi.org/10.5281/ zenodo.163223 .

Förderhinweis

Die vorliegende Arbeit wurde im Rahmen der Nachwuchsgruppe "Computergestützte literarische Gattungsstilistik" (CLiGS) erstellt, die vom BMBF gefördert wird (FKZ 01UG1508).

Bibliographie

Baker, Monya (2016): „Is there a reproducibility crisis?“, in: *Nature* 533: 452–454.

Bohannon, John (2015): „Many psychology papers fail replication test“, in: *Science Magazine* 349.6251: 910–911.

Camerer, Colin F. et al. (2016): „Evaluating replicability of laboratory experiments in economics“, in: *Science Magazine* 351.6280: 1433–1436.

Drummond, Chris (2009): „Replicability is not Reproducibility: Nor is it Good Science“, in: *Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML*.

Gomez, Omar S. / Juristo, Natalia / Vegas, Sira (2010): „Replication, Reproduction and Re-analysis: Three ways for verifying experimental findings“, in: *RESER ’2010*.

Padilla, Thomas / Higgins, Devin (2016): „Data Praxis in the Digital Humanities: Use, Production, Access“, in: *DH2016: Conference Abstracts* 644–646 http://dh2016.adho.org/abstracts/150 .

Peng, Roger D. (2011): „Reproducible Research in Computational Science“, in: *Science Magazine* 334: 1226–1227.

Richeaudeau, François (1982): „Simenon: une écriture pas si simple qu'on le penserait“, in: *Communication et langages* 53: 11–32 10.3406/colan.1982.1484.

Schöch, Christof (2016): „Does Short Sell Better? Belgian Author George Simenon's use of sentence length", in: *The Dragonfly's Gaze* https://dragonfly.hypotheses.org/922 / http://dragonfly.hypotheses.org/1005 .

Spitzer, Leo ([1928]): „Die klassische Dämpfung in Racines Stil", in: *Romanische Stil- und Literaturstudien* I. Marburg: Elwert (1931) 135–268.