



Evaluation of GloWbE as a Tool for Big Data Corpus Linguistics

Authors: Philipp Reichenbach
Submitted: 18. February 2019
Published: 22. February 2019
Volume: 6
Issue: 2
Languages: English
Keywords: Big Data, Corpus Linguistics, GloWbE, Digital Divide, Mark Davies, Robert Fuchs, Christian Mair, Digital Humanities
Categories: Artificial Intelligence, Modeling and Simulation
DOI: 10.17160/josha.6.2.538

Abstract:

Researchers have been interested in media and their impact on spoken and written language prior to the rise of the internet and Web 2.0. Plato already discussed the relationship between written and spoken language more than 400 BC. The chances that came with the beginning of digitalization are technical possibilities of detection, analysis, and evaluation of language. Academic disciplines such as linguistics and digital humanities are dedicated not only to dealing with digitalization and language but also to exploring their interrelationship, opportunities, and occurring (linguistic) phenomena. This paper critically questions Big Data practices in corpus linguistics research. I examine the Big Data corpus GloWbE (Corpus of Global Web-based English) as a linguistic research tool by highlighting advantages, emphasizing critical aspects, and presenting constructive approaches and concepts in the context of this paper. I make some suggestions on how researchers can deal with the mentioned issues of ethics, geographical boundaries and non-native speakers, technological challenges, access and participation, and various authors. Finally, I then summarize the results as chances and challenges in working with Big Data corpora as linguistic research tools and propose topics for further research. As always. I am happy about

JOSHA

josha.org

Journal of Science,
Humanities and Arts

JOSHA is a service that helps scholars, researchers, and students discover, use, and build upon a wide range of content



Table of Contents

1. Introduction	1
2. Definitions	1
2.1. Big Data	2
2.2. Corpus Linguistics	3
2.3. The GloWbE Corpus	4
3. GloWbE as a Tool for Big Data Analysis	5
3.1. Ethical Justification	5
3.2. Geographical Boundaries and Non-Native Speakers	6
3.3. Global Technology	7
3.4. Digital Divide	8
3.5. Various Authors	9
4. Conclusion	10
5. References	12
6. Philipp Reichenbach	14



1. Introduction

Researchers have been interested in media and their impact on spoken and written language prior to the rise of the internet and Web 2.0. Plato already discussed the relationship between written and spoken language more than 400 BC (Heitsch 1993: 62). Since Gutenberg's invention of the letterpress in the 15th century, discourse concerning the influence of new media on language has been controversial. McLuhan, for instance, asks if media such as letters change our sensory relationships and transform mental processes (2011: 30). Clearly, academic interest in the correlation between language and new media is nothing new. The chances that came with the beginning of digitalization are technical possibilities of detection, analysis, and evaluation of language. Academic disciplines such as linguistics and digital humanities are dedicated not only to dealing with digitalization and language but also to exploring their interrelationship, opportunities, and occurring (linguistic) phenomena (Russel 2017: 50).

This paper critically questions Big Data practices in corpus linguistics research. The aim is to examine the Big Data corpus GloWbE (Corpus of Global Web-based English) (Davies 2013) as a linguistic research tool by highlighting advantages, emphasizing critical aspects, and presenting constructive approaches and concepts in the context of this paper. These endeavors are accompanied by answering this question: To what extent do researchers have to engage with ethical and technical challenges when dealing with Big Data corpora such as GloWbE? The discussion precedes a definition of Big Data and descriptions of corpus linguistics and GloWbE. GloWbE is then critically examined in terms of ethical justification, the impact of non-native speakers, technological challenges, internet access and users, and the production of content. Finally, I then summarize the results as chances and challenges in working with Big Data corpora as linguistic research tools and propose topics for further research.

2. Definitions

In the following sections, the terms Big Data, corpus linguistics, and GloWbE are briefly defined as they are the theoretical basis for the analysis in chapter 3.



2.1.1. Big Data

Formerly known as Cloud Computing, Big Data has become a buzzword and is no longer limited to various disciplines of science but has already entered the mass media and public discourse (Hofstetter 2014: 87).

Creating a uniform and universally valid definition of Big Data proves to be difficult, given the diversity of the phenomenon. On the one hand, the phenomenon is controversially discussed, on the other hand, Big Data is evaluated, classified, and redefined differently depending on relevance and priority. For example, a researcher with an economical interest deals with different aspects (e.g. increasing efficiency) than someone with an interest in humanities research (e.g. ethical questions).

Big Data is often used as a collective term for large amounts of data and generally it can be said that the focus is usually on the unstructured nature and the associated difficulty of evaluating them using conventional or manual methods of data analysis. The underlying digital technologies are a technical and social part of Big Data (Reichert 2014: 9). Transactional (digital) user data obtained from web tracking, mobile phone monitoring, and/or sensor acquisition form the basis for new approaches in business, politics, and research. The aim is usually to identify developments in economic markets at an early stage and, ideally, to react before the competition does. In this respect, "the Social Web has become the most important data source" (ibid.: 10).

Regardless of the field of interest, the systematic evaluation of Big Data is often used to show correlations, draw comparisons or disparities, and to predict future developments or behavior at an early stage (Reichert 2014: 12). In corpus linguistics, this refers to the description of languages, their elements and structures as well as their usage, as explained in more detail in the following section. Considering democratic participation and open access, Big Data is the "most appealing source of knowledge at the beginning of the 21st century" (Loureiro-Porto 2017: 449) which makes tools such as GloWbE attractive for linguistic research.

boyd and Crawford describe Big Data "as a cultural, technological and scholarly phenomenon" (2012: 663) basing the argument on the interplay of the following three aspects:

- (1) Technology: maximizing computing power and algorithmic accuracy to gather, analyze, link, and compare large data sets



- (2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims
- (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy (ibid.: 663).

The definition of the two researchers emphasizes three components of Big Data. Firstly, the technological aspect illustrates the amount of work required of machines which, secondly, enables researchers to analyze huge data sets for the particular interest. Thirdly, mythology is not directly related to machines' work but describes the widespread credibility given to numbers and statistics by humans. What distinguishes this definition from most others is this last component of mythology. As we will see later in this paper this aspect is interesting in relation to linguistic research using Big Data corpora.

2.1.2. Corpus Linguistics

In order to understand what corpus linguistics means, it is necessary to clarify what a corpus is.

A corpus is a collection of written or spoken utterances. The data of the corpus are typically digitized, i.e. stored on computers and machine-readable. The components of the corpus consist of the data, and possibly metadata describing these data, and linguistic annotations associated with them (Lemnitzer & Zinsmeister 2015: 13 *my own translation*).

The collection of utterances is often the result of careful planning. Nowadays, corpus data are available in a machine-readable form. The value of a corpus increases if the primary data are provided with descriptive data. These metadata can provide information about the background such as the authors of texts. This must be distinguished from annotations which do not describe entire texts or coherent utterance sequences like metadata but rather mark and classify individual units, for example the categorization of words as a part of speech (Lemnitzer & Zinsmeister 2015: 13). Transferring this onto the concept of corpus linguistics, the corpus comprises a collection of natural texts and may serve various purposes such as foreign language teaching, language development, language documentation, lexicography, and machine language processing (ibid.: 15). Consequently, formal languages such as programming languages are not considered. It should be noted that the one condition is for the texts to be digitalized. One of the strengths of corpus linguistics is that not only the structure of language but also its usage can be investigated on the basis of data (ibid.: 15).



As corpus linguistics is subject to the basic principles of scientific work the results must be verifiable or reproducible for other researchers. In addition to this, the research method must comply with scientific standards. One great advantage of corpus linguistics is that researchers can work with already produced authentic linguistic material and naturally occurring speech instead of relying on their own or other native speakers' intuition. Thus, linguists also escape the observer's paradox.

2.1.3. The GloWbE Corpus

GloWbE, launched by Mark Davies in 2013, is a Big Data corpus used for the purpose of linguistic analysis. In terms of size, the GloWbE corpus consists of 1.9 billion words that are taken from 1.8 million webpages from 20 different English-speaking countries (Davies & Fuchs 2015b: 1). Davies and Fuchs indicate that GloWbE features approximately 60% of informal language taken from blogs and the other 40% "from a wider range of other genres and text types" (ibid.: 2). Google Ngram Viewer assisted the researchers in tracking word use across time (Russel 2017: 50).

Notably, the importance of the size of a corpus is evaluated differently by different researchers. Davies and Fuchs point out advantages of GloWbE over small-sized corpora like the International Corpus of English (ICE), British National Corpus (BNC) or Corpus of Contemporary American English (COCA). GloWbE allows users to carry out comparisons of the different sections of the corpus. This has, in contrast to other corpora like BNC or COCA, the advantage of allowing comparisons of different national varieties (Davies & Fuchs 2015b: 5). Corpora like ICE may "not provide enough data for in-depth research on lexical variation, morphological variation, variation with medium- and lower-frequency syntactic constructions, or differences in word meaning between dialects" (ibid.: 2). As a consequence, Davies and Fuchs felt "the need to create a very large corpus of World Englishes, which would be available to a wide range of researchers" (ibid.: 2). According to Loureiro-Porto, however, the relevance of corpus size has widely been overestimated as the corpus size is only significant when examining particularly infrequent elements (2017: 452).

Whereas ICE is split into neat genres, the genres of GloWbE are controversial (Mair 2015: 30). Thus, ICE corpora represent specific genres in the respective varieties and GloWbE is a better tool in order to determine to which variety a given word, phrase, or construction belongs (Loreiro-Porto



2017: 461). Therefore, depending on the research question, either corpus will offer fruitful findings. GloWbE cannot simply replace smaller and tidier corpora such as ICE (Peters 2015: 42) but it can be considered as an enlargement and as another “tool” in the “toolbox” of researchers studying World Englishes (Mukherjee 2015: 34).

3. GloWbE as a Tool for Big Data Analysis

Big Data corpora, including GloWbE, undoubtedly offer great opportunities for linguistic research. However, the following sections discuss the challenges researchers may encounter when working with GloWbE as a tool for linguistic analysis. It needs to be stressed that the following questions are not intended to discourage researchers working with Big Data corpora but rather serve as constructive criticism for research practices with great possibilities and as suggestions for improvement.

3.1.1. Ethical Justification

The blurring of lines between private and public data in times of digitalization and questions linked to data protection are not only relevant for politics, economy, and media representatives, but also for researchers. Androutsopoulos emphasizes the privacy of informants and points out that researchers should reflect ethical questions about what data to use for their studies (2013: 247). Privacy means that the user has control over what is accessible to the world wide web of what he/she produces/publishes on the web (Rössler 2001: 24). It further implies that the user can decide who has a say in the possible interventions that are relevant to him/her (ibid.: 24). According to Rössler, it is crucial to have control over who has access to one's personal online data to generate privacy. On some social websites such as Facebook this can partly be controlled through privacy settings. Computer-mediated communication (CMC) researchers often divide various online environments into private, semi-public and public data (Androutsopoulos 2013: 247). In the course of this paper only public data are relevant because the GloWbE corpus is compiled with open access data.

Androutsopoulos notes the importance of anonymity of the informants for research ethics: “It is common sense among CMC researchers that we need to protect the anonymity of our informants by not directly disclosing their offline identities and avoiding any cues that may lead to their identification” (2013: 247). Furthermore, it is questionable whether using accessible data is ethical.



boyd and Crawford underline that “[j]ust because content is publicly accessible does not mean that it was meant to be consumed by just anyone” (2012: 672). It needs to be considered here that the accessible data is not only used for research purpose but also for the purpose of publication.

In the case of GloWbE and its several million authors it is impossible to ask these authors for permission each time or even once. How do researchers deal with this issue? In the case of publication one option would be to use the “Context function”. It gives the researcher access to the original web page (Davies & Fuchs 2015b: 7). Web pages usually contain an impressum through which the data owners can be contacted and asked for their permission. One problem that needs to be mentioned is that the KWIC function is the only feature for such a background check but the stored links often do not work or are outdated.

In cases where the used data set is too large to ask everyone separately or if the original source is untraceable researchers would probably have to decide individually which data can ethically be used and which cannot. This, however, opens ground to further questions such as who is in the position to decide what data is acceptable to be used and published. Unfortunately, the scope of this paper does not allow to discuss this issue in further details.

3.1.2. Geographical Boundaries and Non-Native Speakers

Nelson claims “that we know little or nothing about the authors of the webpages” (2015: 39) especially of weblogs which make up 60% of GloWbE’s content. In their reply, Davies and Fuchs refer to the provided URLs for each of the web pages referring to the original web pages in the corpus which enable users to identify the country of origin of the author (2015a: 46). However, the URL may not always indicate the author’s first language. Migrants who are usually non-native speakers of the majority language of the country they reside in, use the internet to the same extent as native speakers. They also participate in virtual worlds and produce online content. In a globalized world with the internet as a global network of communication the blurry boundaries between varieties of English and their speakers are not identical with the sharp lines between national web domains and their texts (Mukherjee 2015: 36). Nelson points out that “the domain name alone does not provide foolproof evidence of the origin of any text” (2015: 39). In other words, web users may also publish



on websites associated with countries they do not live in or on websites associates with a language that is not their first.

A large number of humans regularly travel around the globe leaving traces in both the analogue as well as in the digital space as digital footprints. As a consequence of Mukherjee and Nelson's claims, digital footprints and data of any kind left behind by migrants can influence and falsify research results as soon as they end up in corpora such as GloWbE if they are not recognized or tagged correctly (Mukherjee 2015: 36; Nelson 2015: 39). A German traveler, for example, might create a blog in Australia with the Australian domain .au. A researcher who only has access to the data cannot know that this person is not a native Australian English speaker. In other words, we can use the domain name or visit the web page as a starting point but not rely on it alone to indicate the author's first language or background. A further example is India. The country is listed in GloWbE as an English-speaking country. In India more than 20 languages are spoken. Even if internet users come from the same country they can have different first languages. The risk remains that researchers working with GloWbE draw wrong conclusions because the aspect of non-native speakers is not and perhaps cannot be precisely classified. Researchers should be aware of this problem, examine their research data accordingly, and consider in which cases which corpus is suitable; depending on the research interest a small and structured corpus could be a good or even a better choice. As a consequence, new linguistic phenomena should be verified using different sources, for example ICE, in order to not base new discoveries on wrong assumptions.

3.1.3. Global Technology

boyd and Crawford point out that we attribute particularly high credibility, accuracy, and objectivity to large data sets and their underlying technology is often associated with the widespread belief that knowledge that was previously impossible can be generated (2012: 673). The use of technology for one's own research usually presupposes trusting the software and its functionality. However, how well does a researcher know the software he/she works with and can we always presuppose the tool functions flawlessly? It is not very likely that all researchers study a software in detail in advance and know how its system works. Arguably, this may not be necessary, however, one



problem with not knowing the software used for research well could be, for example, that strengths and weaknesses cannot be identified and reflected at an early stage.

One issue that researchers encounter working with GloWbE is the large amount of data which entails not only duplicates such as mirrored newspaper articles (Davies & Fuchs 2015a: 46f.), but also incorrect tagging (Mair 2015: 30). Davies and Fuchs point out that they are aware of the first and that they followed several approaches to eliminate duplication. The researchers indicate that “in a corpus this size, it would of course be impossible to identify and remove all duplicates” (2015a: 47). They also understand that “the more informal and nonstandard the language [...] is, the less reliable the tagging will become” (ibid.: 47). Considering the amount of data, the authors take useful measures to tidy the corpus as much as possible but can obviously not ensure that only speakers of the respective variety are presented in GloWbE (ibid.: 46). In other words, wrong tagging and duplication cannot be completely prevented. Therefore, using data from GloWbE does not guarantee precise results and researchers should invest time in verifying tagging and removing duplicates from their data set. Hence, major problems arise which cannot be fully resolved in this paper: How media literate or technology-affined do researchers have to be in the digital age in order to conduct accurate web-based research projects?

3.1.4. Digital Divide

Questions of internet access and participation can be traced back to an issue called digital divide. Researchers discuss the digital divide on various levels. In principle, the digital divide refers to unequal global access to information and communication technologies that lead to inequalities in opportunities, access to information, democracy, and mobility. Linguists point out that there is a digital and linguistic divide between internet users in richer and poorer countries which is not only reflected in varying degrees of internet access but also in the number of linguistic resources available for the supply of services in the respective national languages (Leppänen & Peuronen 2012: 4). In the context of Big Data, Manovich mentions three classes of people: “those who create data (both consciously and [unconsciously] by leaving digital footprints), those who have the means to collect it, and those who have [the] expertise to analyze it” (2011: 10). boyd and Crawford point out that there are not only questions about access and linguistic knowledge but also about



computational skills. Among the researchers who have computational skills, the majority is still male and has access to data and thus influences the research (2012: 674). This imbalance of gender and access can also be observed in the economy and society: Companies skilled in manipulating large data control the economy and web users with highly developed computational skills have access to information, knowledge, and education. boyd and Crawford name this new digital divide “the Big Data rich and the Big Data poor” (2012: 674). In other words, privileged users create data and privileged researchers have access to these data. They stand in contrast to less privileged and less skilled web users.

Ultimately, what does the digital divide and its access issue mean to GloWbE? If we assume that internet access and participation are unequal, we must also assume that some groups are underrepresented or not represented at all. Researchers working with Big Data corpora like GloWbE should be aware of this. It confirms the statement that GloWbE should be regarded as one of many, but not as a tool replacing other linguistic research tools (Davies & Fuchs 2015b: 26). Even though digital Big Data corpora provide great advantages, some linguistic phenomena may still be examined more accurately in a manually collected small corpus or in the field.

3.1.5. Various Authors

As already mentioned in section 3.2 we know little or nothing about the author’s first language or background (Nelson 2015: 39). Nowadays, anyone can easily create a blog or a webpage, even a computer bot programmed to do so. Actually, a web user without technical skills would be able to do so with the help of an online tutorial or a short introduction. Linked to this is the question of scalability. boyd discusses the question of what contents appear on the internet (2011: 47f.). According to her, a great part of what appears in networked publics are the wrong kinds of contents in the sense of insignificant or irrelevant data of “those seeking broad attention” (ibid.: 48). It becomes problematic for research purposes when a large part of the data is irrelevant or misleading. How do we then know who the authors behind the data fed into GloWbE are? There cannot be certainty about the background of the author without doing a source verification. Depending on the research interest this might not even be necessary; however, it can make a difference whether the



data are produced by artificial intelligence or a human being. It may also have an impact on the study whether a page and its content are created by a journalist, a company, or a student.

Again, the KWIC function with the provided URL helps. This function proves to be a very decisive and indispensable function of the GloWbE corpus. As described in the previous sections, it can be used to see the original source in case the URL is valid. Ideally, the author should then be found as well as further background information. This procedure is of course very time-consuming. Depending on the research interest and data volume it may be difficult to implement.

4. Conclusion

This paper deals with the GloWbE corpus and critical questions researchers using the tool should ask themselves when working with Big Data corpora. I made some suggestions on how researchers can deal with the mentioned issues of ethics, geographical boundaries and non-native speakers, technological challenges, access and participation, and various authors. Researchers working with GloWbE face the following advantages and challenges:

The interface of GloWbE brings out several strengths. To mention a few, it allows users to carry out comparisons of the different sections of the corpus to be able to compare different national varieties. Additionally, GloWbE is a good tool to determine to which variety a given word, phrase, or construction belongs. Its large size allows in-depth research on lexical or morphological variation, variation with medium- and lower-frequency syntactic constructions, or differences in word meaning between dialects. The latter in particular may be a more useful strategy applied in GloWbE than in other, smaller corpora. They might simply display too little data for rare structures and words to appear.

Five main challenges are briefly summarized hereafter: Firstly, researchers often have to decide for themselves which data can be used ethically, and which cannot. Possibilities include asking the authors for permission by using the KWIC function to get to the original source or keeping the authors anonymous. Secondly, non-native speakers who are categorized in 20 different countries that may not be associated with their first language may pose a further challenge when analyzing data on GloWbE. In this case, it would be possible to use a smaller structured corpus. Thirdly, the fact that many humans trust in technology without knowing much about it may even be a social



and not only an academic issue. For this purpose, further research work towards media literacy, data literacy, or digital literacy would have to be carried out in academic environments. Fourthly, the digital divide is a phenomenon that can pose a challenge in the digital space due to the lack of representativeness of groups. Depending on the subject matter of the research a structured corpus might be a better alternative. Last but not least, missing background information about the author can theoretically be obtained via the KWIC function. However, the procedure cannot promise success and is very time-consuming. One hundred percent source verification cannot be achieved.

Even in this short overview, GloWbE has proven to be a valuable and complementary tool for corpus linguistics research on linguistic phenomena. Nonetheless, researchers should be aware of certain ethical and technical challenges. At some point it will be interesting to pursue further research such as open questions of who is in the position to decide what data is acceptable to be used and published, how media literate researchers need to be in web-based research, the compilation of clean data in the creation of a Big Data corpus, and technically more efficient possibilities for source verification.



5. References

- Androutsopoulos, Jannis. 2013. Online Data Collection. In *Data Collection in Sociolinguistics: Methods and Applications*. London and New York: Routledge.
- boyd, danah & Kate Crawford. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15 (5): 662–679. URL: <http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>, accessed 19 June 2018.
- boyd, danah. 2011. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *Identity, Community, and Culture on Social Network Sites*. 39-58. New York: Routledge.
- Davies, Mark. 2013. *The Corpus of Global Web-Based English*. URL: <https://corpus.byu.edu/glowbe/>, accessed 19 June 2018.
- Davies, Mark & Robert Fuchs. 2015a. A Reply. *English World-Wide* 36 (1): 45–47. URL: <https://www.benjamins.com/#catalog/journals/eww.36.1.03dav/fulltext>, accessed 20 June 2018.
- Davies, Mark & Robert Fuchs. 2015b. Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-Based English Corpus (GloWbE). *English World-Wide* 36 (1): 1–28. URL: <https://www.benjamins.com/#catalog/journals/eww.36.1.01dav/fulltext>, accessed 19 June 2018.
- Heitsch, Ernst. 1993. *Platon, Phaidros*. Göttingen: Vandenhoeck und Ruprecht.
- Hofstetter, Yvonne. 2014. *Sie Wissen Alles: Wie Intelligente Maschinen in unser Leben Eindringen und Warum Wir für unsere Freiheit Kämpfen Müssen*. München: C. Bertelsmann.
- Lemnitzer, Lothar & Heike Zinsmeister. 2015. *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.



- Leppänen, Sirpa & Saija Peuronen. 2012. Multilingualism on the Internet. *The Routledge Handbook of Multilingualism*. London and New York: Routledge.
- Loureiro-Porto, Lucía. 2017. ICE vs GloWbE: Big Data and Corpus Compilation. *World Englishes* 36 (3): 448–470. URL: <http://onlinelibrary.wiley.com/doi/10.1111/weng.12281/epdf>, accessed 22 June 2018.
- Mair, Christian. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1): 29–33. URL: <https://www.benjamins.com/#catalog/journals/eww.36.1.02mai/fulltext>, accessed 21 June 2018.
- Manovich, Lev. 2011. Trending: The Promises and the Challenges of Big Social Data. In *Debates in the Digital Humanities*. The University of Minnesota Press. URL <http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data>, accessed 13 July 2018.
- McLuhan, Marshall. 2011. *Die Gutenberg-Galaxis. Die Entstehung des Typographischen Menschen*. Mit einem Vorwort von Richard Cavell. Hamburg: Gingko Press.
- Mukherjee, Joybrato. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1): 34–37. URL: <https://www.benjamins.com/#catalog/journals/eww.36.1.02muk/fulltext>, accessed 20 June 2018.
- Nelson, Gerald. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1): 38–40. URL: <https://www.benjamins.com/catalog/eww.36.1.02nel/fulltext>, accessed 20 June 2018.
- Peters, Pam. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1): 41–44. URL: <https://www.benjamins.com/#catalog/journals/eww.36.1.02pet/fulltext>, accessed 03 July 2018.
- Reichert, Ramón. 2014. *Big Data: Analysen zum Digitalen Wandel von Wissen, Macht und Ökonomie*. Bielefeld: Transcript Digitale Gesellschaft.
- Rössler, Beate. 2001. *Der Wert des Privaten*. Frankfurt am Main: Suhrkamp.



Russell, John. 2017. Supporting Digital Humanities. The Basics. *Online Searcher* 41 (2): 49-52. URL <http://www.infoday.com/OnlineSearcher/Articles/Features/Supporting-Digital-Humanities-The-Basics-116789.shtml>, accessed 20 June 2018.

6. Philipp Reichenbach



Duale Ausbildung zum Veranstaltungskaufmann am Theater Freiburg und mehrjährige Tätigkeit im o.g. Beruf; Studium der Medienkulturwissenschaft (B.A.), Bildungswissenschaft & Bildungsmanagement (B.A.) sowie der Medienkulturforschung (M.A.) an der Albert-Ludwigs-Universität Freiburg. Doppelte Stipendienförderung im Bachelorstudium durch das Deutschlandstipendium (Vergabe durch das Bundesministerium für Bildung und Forschung) sowie die Dr.-Leo-Ricker-Stiftung der Stiftungsverwaltung Freiburg für Studierende mit außergewöhnlichen Leistungen; aktuelle Förderung im zweiten Jahr des Masters durch das Deutschlandstipendium, Stipendium für begabte und leistungsstarke Studierende; Tätigkeit als Hilfswissenschaftlicher Mitarbeiter im Bereich Studienkoordination, Veranstaltungsmanagement und Geschäftsführung am Institut für Medienkulturwissenschaft der Universität Freiburg; Interessen in Studium und Forschung: Privatheit- und Öffentlichkeitdiskurse, Big Data, Medienbildung, Algorithmen-Ethik, Künstliche Intelligenzen, Technikforschung, Mensch-Maschine-Interaktionen, Medienwirkung, soziale Sicherheits- und Angstdiskurse in medialen Kontexten.