

ARTICLE

The Wikipedia Republic of Literary Characters

Paula Wojcik¹, Bastian Bunzeck², Sina Zarrieß²¹ University of Vienna, ² Bielefeld University

Keywords: wikipedia, wikidata, character networks, world literature, canon studies

<https://doi.org/10.22148/001c.70251>

Journal of Cultural AnalyticsVol. 8, Issue 2, 2023

Literary characters do not receive the same attention from literary scholars as other components of literature (i.e. the narrative, theme, motif). In contrast to this lack of interest, various studies have shown that characters in particular play an important role for readers. We draw on this observation to explore a user-oriented notion of World Literature according to the collaborative encyclopedia Wikipedia. Based on its language-independent taxonomy Wikidata, we collect data from 321 Wikipedia editions on more than 7000 characters presented on more than 19000 independent character pages across the various language editions. We use this data to build a network that represents affiliations of characters to Wikipedia languages, which leads us to question some of the established presumptions towards key-concepts in World Literature studies such as the notion of major and minor, the center-periphery opposition or the canon.

1. World Literature and Wikipedia

The amount and diversity of articles, debates, and theories on world literature render it nearly impossible to identify some core characteristics or to outline a homogeneous field of study. In a recently published volume, one of the editors, Dieter Lamping, admits by the end of his introduction: "What is understood by world literature can perhaps only be described in a comparative-differentiating way" ("Was unter Weltliteratur verstanden wird, ist vielleicht nur noch komparativ-differenzierend zu beschreiben" (Lamping)). This appears to be a Solomonic solution regarding the span of world literature concepts.

Among these concepts, the most widely (and also most controversially) discussed were provided by Damrosch, Casanova and Moretti. The last two state an 'oneworldliness' of the literary system that is ruled by hegemonic powers and inequality. Underlying both is a universalizing perspective that makes world literature appear analogous to a global marketplace. It includes a notion of inequality that arises between the center and the periphery. Pascale Casanova's metaphor of the Greenwich Meridian has become emblematic of this macro perspective. Although Casanova does not use the term "world literature," her analysis provides the most concrete picture of how a world of literature ("republic of letters" in her terms) might be thought of and therefore her concept will form an essential point of reference for us.

The Greenwich Meridian measures (or: constructs) literary quality and value in Casanova's republic. This republic is a rather static formation with France, Germany, and Great Britain in the centre, Italy, Spain, Russia loosely arranged in its belt and the non-European literatures condemned to an existence in the periphery. It is an inert arrangement that developed roughly throughout the last 500 years. In the past few years, there has been no shortage of attempts to turn the globe and to shed some light on the dark side of the planet. Mappings and re-mappings of literatures at different time periods created a richness of micro perspectives on literary fields all over the planet that challenge the "Western nationalism" (Hirakawa) of world literature studies. The debates are characterized by two developments: firstly, they focus on marginalized but still big continents (Latin America (Müller et al.)) or empires (Pan Asian (Park)) rather than on languages and secondly, they rarely enter the area of reception studies.

Although reception and, more recently, fandom studies proved to be a fruitful approach in sociology, media and communication studies, there is surprisingly little work done on world literature from a reception-oriented perspective. Yet, exceptions such as Lena Henningsen's inquiry (Henningsen) into the reading habits of young Chinese during the Cultural Revolution, or research that addresses creative and collaborative formats such as fanfiction platforms (i.e. Wattpad (Pianzola et al.)) and Wikipedia (Hube et al.) provide insight not only into reading and creating patterns but also enrich and/or correct our picture of world literature's core concepts such as intercultural transfer, transculturality, or the canon. A comparison between the canon and the reading preferences reveals striking differences regarding for example the genre: Although the canonized classics are being object to fanfiction re-writings (Stemberger), their numbers are ridiculously low compared to other genres such as the detective novel, science-fiction, or fantasy. Conversely, these genres are rarely represented in anthologies of world literature. This observation aligns again with a major critique which addresses the small variety of genres considered to be of interest in world literature studies; i.e. the fundamental exclusion of oral poetry (Heath; Werberger) or children's literature. As we have discussed in an earlier article characters from popular genres are represented in Wikipedia at least as prominently as the canonized ones (Picard et al.). We think it is worthwhile to question other categories of world literature with the help of Wikipedia as well.

To examine how the "world" of world literature presents itself from a reception-oriented point of view, Wikipedia can be considered a transcultural observatory (Hube et al.). Many studies conducted on communities, societies, cultures, literatures, and languages of the world stress the multi-linguality and multi-culturality of the data that can be found in the Wikipedia universe, i.e. its various language-specific editions and links between them that can be systematically accessed via the underlying Wikipedia graph and accompanying ontologies like Wikidata (Hube et al.; Blakesley, "The Global Popularity

of William Shakespeare in 303 Wikipedias”; Blakesley, “World Literature According to Wikipedia Popularity and Book Translations”). Compared to many other existing corpora and datasets in the humanities, Wikipedia clearly stands out as a massively multi-lingual, well-structured and continuously updated collaborative resource.

Wikipedia is not only an observatory of the world’s cultural knowledge, but it also reflects its biases (Hube), e.g. gender bias, cultural bias (Voit and Paulheim), or temporal bias. We thus ask whether basic categories that structure world concepts in world literature studies are echoed in Wikipedia. In particular we focus on the center-periphery-opposition which is considered to reduce world literature to the picture of “the west and the rest,” (Hall) or an “an exclusive club of Europeans and Americans” (Hirakawa 546).

2. Why Study World Literature’s Characters?

Our focus on characters is related to a reception-oriented approach. It contrasts with the majority of works in and on literary studies where, compared to the amount of studies on narrative, research on characters is still a niche. This imbalance mirrors the conviction expressed by many scholars (from Aristotle to Henry James and Vladimir Propp) that characters are in service of the plot, rather than the other way around—a hierarchy which is questioned in recent research (Eder et al.).

As there has been a notable amount of work done on reader’s attachment to characters, we can conclude that in terms of reception this category is particularly revealing. The studies exhibit that readers share emotions with characters (Dijkstra et al.), hear their voices also while they are not reading and they even hear them comment on events that happen in their own lives (Alderson-Day et al.). They gossip about characters as if they were real (Laffer) and engage romantically with them (Liebers and Straub). A swift look on the world of media and everyday culture confirms the role of characters in other areas: Characters appear in video games, VoD-series, fanfiction, commercials, on fashion and household articles where they can be detached from their original narrative or a narrative at all.

This distribution-based autonomy has been recognized and has many names: Müller coined the term “interfiguralität” in correspondence to “intertextuality”. Others emphasize the full, or at least semi-autonomous, status by speaking of “fluctuating individuals” (Eco), “figures on loan,” (Ziolkowski) or “trans-world individuals” (Margolin 864). In this paper, we presume that the autonomous status of a character is mirrored and can be measured through Wikipedia when characters get a separate visibility through their own pages. This applies by far not only to main characters, but also to those who are on the fringes of a novel, poem, or verse epic.

By focusing on characters, we address their importance for readers and audiences. But beyond that, we also believe that the study of characters on a collaborative platform can present another picture of world literature than those we are used to. A work can be represented only once in one Wikipedia language edition, but through its character pages, the presence can be multiplied. This increment in representation indicates the impact of a work within a Wikipedia language edition and the whole Wikipedia universe.

As we will demonstrate, there are canonical as well as non-canonical works that seem to be more attractive for literature readers and Wikipedia users (including Wikipedia authors) to engage with their characters than others. This is neither due to the number or complexity of the characters, nor to the reputation or popularity of the works. Since it will not be possible to highlight all dynamics that are at work in detail, we will focus on the thesis that fan cultures develop around certain works and tend to engage extensively with their characters. This holds true for both popular works, which are distributed via potent book, film and merchandise industries (i.e. *Harry Potter*), as well as rather hermetic texts (i.e. *Ulysses*) that are only received within small and, in many cases, even specialized communities. We are interested in the impact of these fan communities on world literature in Wikipedia and we pursue the following questions: How does the relationship of center and periphery present itself when we focus on literary characters? What kind of literary canon can be identified?

3. Methods

3.1. Data

Several different and independent projects offer structured access to the knowledge and facts stored in the Wikipedias. Among them, DBpedia and Wikidata stand out as the most extensive and accessible. Both projects provide structured statements about entities inside (and in Wikidata's case also outside of) the Wikipedia, consisting of triples in the form subject-predicate-object. However, DBpedia differs from Wikidata in one crucial aspect: while the statements in Wikidata are language-independent and tied to the different Wikipedias through unique identifiers, DBpedia has localized versions attached to their respective Wikipedia. Thus, Wikidata is accessible for all language versions of the Wikipedia, while DBpedia is only available for explicitly integrated Wikipedias. For example, Hube et al. had to restrict their analysis of author representation in the Wikipedias to 125 unique language versions accessible through DBpedia (Hube et al.), whereas our Wikidata-based study incorporated data from the 321 Wikipedias that were part of the Wikipedia project at the time of data compilation. In their seminal study on bias in DBpedia, Voit and Paulheim showed how movie recommendation systems based on individual DBpedias exhibit unclear, but

significant bias patterns towards certain production countries and genres (Voit and Paulheim). Consequently, the common Wikidata knowledge graph offers a more promising alternative.

The Wikidata project was founded in October 2012 to provide a persistent, collaborative knowledge base behind the Wikipedia projects. As of March 2022, there are more than 12.000 active volunteer editors. The data structure underlying Wikidata consists of items/entities and statements made about them, serialized in the rdf-format. Over 97 million items and over 1 billion statements form the backbone of this knowledge graph (Wikidata). Additionally, statements in Wikidata are commonly associated with references for their validity. However, the crowdsourced nature of this approach promotes inconsistencies in the data scheme, some of which will be discussed shortly.

Wikidata provides a SPARQL endpoint for querying the knowledge graph, either through the dedicated web interface, or automatised with libraries available for common programming languages. Due to its continually evolving nature, it is virtually impossible to query a persistent state of the complete Wikidata knowledge graph. While time-stamped dumps of the complete data are readily available, their size and structure require hardware resources beyond the reach of the average scholar. For example, a recent (2019) tutorial by Wikimedia software engineer Adam Shoreland lists 104GB of RAM and 16 CPU cores as the minimum requirement, together with a NVMe SSD RAID system, which then still needs at least one week to process and store the data dump (Shoreland). We worked around these issues by creating initial lists of our base data (Wikipedia versions and literary characters) and then querying additional information only for these lists, omitting all data points that might have been added in the meantime. Although it cannot be ruled out that changes in the additional data occurred between individual querying rounds, the limited temporal window for these changes makes substantial alterations practically impossible. Our dataset was compiled in August 2021.

As a starting point, the complete list of entities that belong to the category “literary characters” (Q3658341) was extracted, together with various additional attributes, for example “gender”. Moreover, works that feature these literary characters were queried for. This proved to be another challenge, as there are two different attributes that denote this relation, “characters” (P674) and “present in work” (P1441). Although these attributes are nominally inverse, this relationship is not implemented practically. Conversely, some character-work pairs are only coupled through one of these properties, while others use both, leading to incongruent lists derived from these attributes. In a second step, all found entities were queried for their site representations in the Wikipedias. To do so, a list of all active Wikipedia language versions was compiled from the English Wikipedia’s

overview page (Wikipedia). Because the online interface frequently timed out when extracting the linked Wikipedia pages for the entities, the Python library `qWikidata` (Kensho) was used to automatically extract batches of site links with pauses in between, to not trigger further time-out errors. For every pairing of entity (literary character or work) and Wikipedia language version, this provided either the URL for the character's page in the respective Wikipedia, or "none". The base of our co-nationality network is the resulting cross-table of characters and Wikipedia languages.

It must be noted that, generally, different Wikipedias are highly diverse in their size, data quality and overall makeup. Apart from the language versions for popular and widely spoken languages, there also exist individual Wikipedias for regional dialects (e.g. Bavarian German or Walloon French), constructed languages (like Esperanto or Volapük) and dead languages (for example, East Church Slavonic), which feature a considerably smaller amount of articles, due to their specialist nature and comparatively low numbers of speakers. Thus, while the aforementioned data collection procedure presents a straightforward and clean way of tracing characters across different languages and nations, it must be noted that the resulting data is not necessarily free of potential problems. First, some Wikipedias are not only curated by human contributors. The currently second largest Wikipedia edition, the Austronesian Cebuano Wikipedia, features mostly articles automatically generated by the software `Lsjbot`, and thus does not directly mirror the interest of this language's community in certain topics. Moreover, the linking between Wikidata entities and their corresponding Wikipedia pages proved to be inconsistent. While most links directly lead to individual pages, a minority of them simply redirect to special pages that list and describe noteworthy characters from a certain narrative universe, although in a more reduced and condensed format. In this sense, these literary characters do not have their *very own page*, but share their descriptive space with other characters. As there is no effective way to distinguish these links from truly unique pages, and because the edit history of these characters is not automatically traceable, which could show whether a unique page was later merged with others into an overview page, we do not treat these characters differently from those with unique pages. In the end, they were still deemed worthy of inclusion by the editors, and some of these overview lists offer even more detailed descriptions than other standalone sites. Lastly, also the data quality of Wikidata itself remains debatable. In a preliminary analysis of the co-nationality network, more than thirty literary characters were found to be exclusive to the Wikipedia in the constructed language Interlingua. A closer inspection of this character set revealed that it consists exclusively of minor Harry Potter characters from the extended, online-only Pottermore page. This shows that there are no clear-cut criteria for the inclusion/exclusion of literary characters in the Wikidata project. Furthermore, their articles were all created by the same Wikipedia editor, which illustrates the danger of smaller Wikipedias leaning heavily towards certain topics that most of their

active authors are interested in. Similarly, also the internal categorisation in Wikidata is sometimes spurious. For example, the Holy Bible is classified as a literary work, while its parts (e.g. the New Testament) are not. These categorisations are at the choice of individual contributors, which can have wide-ranging repercussions for a computational analysis. Some characters from sacred texts are categorised as fictional, others as real human beings or even “humans who may be fictional”. At the moment, Wikidata is still work-in-progress and in the absence of tighter control mechanisms, such inconsistencies are unavoidable. In the following, we assume that most of these potential problems affect the network only locally, and do not dominate the main patterns and tendencies regarding the importance and dynamics of characters.

3.2. Character-Wikipedia Network(s)

Using the procedure explained above, we were able to retrieve 7043 language-independent entries for literary characters in Wikidata, and 19322 character pages distributed over the aforementioned 321 language editions in Wikipedia. We represent this data as a bipartite network, i.e. a common type of network in the social sciences and the humanities for representing groups and their members (Latapy et al.). Generally, a bipartite network is a graph that consists of two sets of nodes and links (or edges) that connect nodes from these two sets. One set of nodes corresponds to the literary characters recorded in Wikidata, the other one to the individual Wikipedias. A character node is connected to a Wikipedia node if the character is represented in the national Wikipedia in terms of an independent article. The idea of our network is similar to so-called affiliation networks (Tabassum et al.) where, e.g., connections between actors affiliated with an organization or event can be studied. In the following, we will use this network to investigate the centrality and affiliation structures of characters on the one and Wikipedias on the other hand. We will analyze (i) how characters and their pages cluster across the languages of the Wikipedia world and whether there are (sets of) characters that are known world-wide or in particular areas of the globe, and (ii) how Wikipedias cluster across characters pages and whether there are sets of Wikipedias that share sets of characters.

An important goal of our study is to assess the effect of including smaller Wikipedia editions into our empirical analysis. We note that a range of previous studies that aimed at assessing the centrality of entities in the Wikipedia universe still used a rather restricted set of languages ranging between the top 15 (Hube et al.) and top 25 Wikipedias (Eom et al.). Therefore, we additionally construct a reduced version of our Character-Wikipedia network that only includes the top-15 Wikipedias according to their current global size. In the following, we will use this network to compare structures and connections between characters that arise in a highly multi-lingual network to a picture that is drawn from the dominant or central Wikipedias only. In [Table 1](#), we report some basic statistics on the

Table 1. Number of nodes and edges (links) in the Character-Wikipedia networks, and the resulting average degree for Wikipedia nodes (k_{wiki}), character nodes (k_{char}) and the full network (k)

	Full network	Top-15 network
N. of Wikipedia nodes	181	15
N. of character nodes	4623	4240
N. of edges	19322	10769
k_{wiki}	106.75	717.93
k_{char}	4.18	2.54
k	0.023	0.169

distribution of nodes and links and their average degree (i.e. the average number of links of a node) in our full and top-15 network: the number of character nodes in the the top-15 network is reduced from 4623 to 4240, i.e. only 383 characters get removed from the network when removing 166 languages and keeping only the top 15 Wikipedias. This shows that, overall, most of the characters retrieved from Wikidata and Wikipedia by our querying-based approach seem to be represented in the large Wikipedias. Consequently, the top-15 network is very dense, as indicated by the average degree k and the average degree of the Wikipedia nodes k_{wiki} .

4. Characters and their Autonomy

As has been said before, we consider the autonomy to be mirrored in Wikipedia when literary characters are having their own Wikipedia pages. To verify that characters become at least partially autonomous from their plots, we assess their autonomy by comparing the counts of Wikipedias pages associated with Wikidata entries for characters as opposed to pages associated with their original works.

Although in this article we will not analyse the page structure and content in detail, we want to give a short impression of it. On character pages, characters are being introduced with a biography, a description of their function for the plot, defined by its characteristics in more or less detail, and sometimes illustrated with a picture (from book covers, movies, video games, but also paintings or illustrations of professional, semi-professional, or amateur painters), as shown for two examples of character pages in [Figure 1](#).

Although the structure is stable in the different language editions, the content and scrupulousness of the description can vary. As described in the introduction, both the mere amount of characters from one work which have an own page and the length of the article are closely connected to reception behavior, providing information about the status of works within reader or fan communities. An analysis of the representation of literature via autonomous character-pages in comparison with the representation of the work itself displays some significant differences.

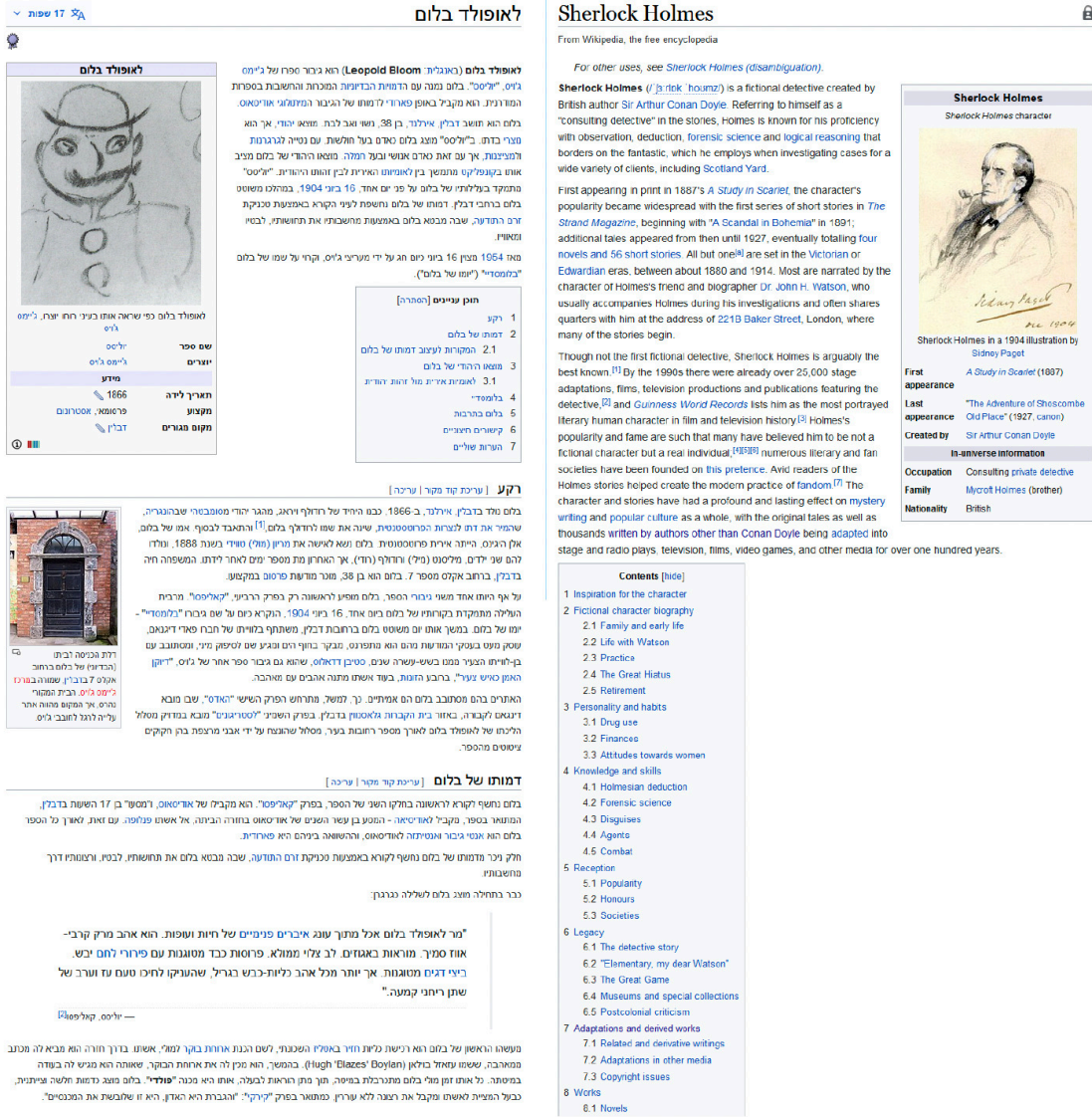


Figure 1. Two (fragments from) character pages from different Wikipedias: Leopold Bloom in the Hebrew Wikipedia (left) and Sherlock Holmes in the English Wikipedia (right)

Table 2a shows the characters in our data that have the most Wikipedia pages associated with their “top work” which is the literary work that they are associated with in Wikidata and that has most Wikipedia pages across languages. As many of the top characters are from *Harry Potter* and appear in the same top work, we removed characters appearing in the same top work to show a more diverse overview of our data in Table 2a. We contrast the character ranking with a ranking of the works that have most Wikipedia pages, shown in Table 2b, where we removed works featuring the same top character to show a more diverse overview of our data (i.e., by this means, we mostly exclude sequels of *Harry Potter*). As we see in these two tables, the ranking of works clearly diverges from the ranking via character-pages: the *Qu’ran* which is leading the first ranking is replaced by *Harry Potter*; the epics *Illiad* and *Mahabharata* or the stories from *1001 Nights* make room for Irish, German, US-American, French, or Chinese classics, fantasy, sci-fi novels.

Table 2a. Top 30 characters according to their individual character page count, with titles of their associated top work and its number of Wikipedia pages.

Character	N_{wiki}	Top work	$N_{wiki-work}$
Sherlock Holmes	95	The Great Mouse Detective	92
Superman	91	Space Jam: A New Legacy	88
Santa Claus	90	Up on the House Top	28
Gilgamesh	79	Sumerian King List	96
Harry Potter	76	Harry Potter and the Goblet of Fire	130
Tarzan	70	Tarzan and the Lion Man	30
Winnie the Pooh	69	Disney's House of Mouse	37
Pandora	67	The Paradise of Children	1
Arjuna	63	Mahabharata	124
Peter Pan	59	Pan	39
Gandalf	56	The Hobbit	118
Maitreya	55	Journey to the West	62
Wolverine	55	Ultimate Spider-Man	67
Pinocchio	54	The Adventures of Pinocchio	69
Hercule Poirot	54	Five Little Pigs	57
Pippi Longstocking	52	Pippi Goes on Board	28
Count Dracula	48	Bram Stoker's Dracula	74
Scheherazade	46	The Thousand-and-Second Tale of Scheherazade	110
Golem	44	none	0
Miss Marple	44	Sleeping Murder	42
Faust	43	The Tragical History of Doctor Faustus	53
Conan the Barbarian	43	Conan the Barbarian	44
Hannibal Lecter	39	The Silence of the Lambs	75
Lancelot	39	King Arthur	58
Robert Langdon	39	The Da Vinci Code	78

We removed characters appearing in the same top work as a higher ranked character (mainly characters from Harry Potter, e.g., Hermione Granger, Ron Weasley).

We see a divergence between characters that have reached some degree of autonomy from their original literary work, and well-known literary characters that are still very closely associated with their original. The top three most autonomous characters (Sherlock Holmes, Superman, Santa Claus) are actually more widely represented as independent characters, than in terms of their works. A result that may seem a little bit funny at the first glance is that Sherlock Holmes' most widely represented work is the comic book *The Great Mouse Detective* - or Santa Clause is most prominently featured in *Space Jam* - a comic book inspired by a movie of the same title. At a second glance however, this result confirms the autonomous status because it highlights the intermedia circulation of characters and their emancipation from the literary origin (Picard et al.).

The various characters from the Harry Potter universe have independent articles in many Wikipedias, but the works are even more widely spread. Pandora is classified as a literary character in many languages, but only has one article about the original text. The character of Golem has 44

Table 2b. Top 30 works according to page count, with their associated top characters and its number of Wikipedia pages.

Work	N_{wiki}	Top character	$N_{wiki-char}$
Qu'ran	169	Talut	6
Harry Potter and the Goblet of Fire	130	Harry Potter	76
Mahabharata	124	Arjuna	63
Star Wars: Thrawn	122	Grand Admiral Thrawn	20
Pottermore	120	Artemisia Lufkin	1
Voldemort: Origins of the Heir	120	Lord Voldemort	66
The Hobbit	118	Gandalf	56
Aladdin and the Wonderful Lamp	112	Aladdin	18
Ali Baba and the Forty Thieves	111	Ali Baba	29
The Thousand-and-Second Tale of Scheherazade	110	Scheherazade	46
The History of Cardenio	109	Cardenio	2
Don Quixote	109	Sancho Panza	24
Ramayana	100	Shabari	14
Alice's Adventures in Wonderland	97	Alice in Wonderland	30
Sumerian King List	96	Gilgamesh	79
The Divine Comedy	94	Ciacco	8
The Great Mouse Detective	92	Sherlock Holmes	95
The Godfather II	89	Vito Corleone	36
Space Jam: A New Legacy	88	Superman	91
Star Trek: Deep Space Nine	85	Jean-Luc Picard	30
Star Trek: Starship Creator	85	Spock	36
Star Trek	84	William Riker	22
Never Say Never Again	84	James Bond	30
Dream of the Red Chamber	83	Wang Xifeng	5
Snow White	83	Snow White	1

We removed works with the same top characters as a higher ranked works (mainly sequels of Harry Potter)

independent pages but no work connected to its character. These differences in representation between work and character initially indicate that the study of characters can present a different picture of world literature. Against their background, we focus one of the key-concepts of world literature studies: the opposition of center and periphery.

5. Wikipedia: Global Center and Periphery

Prior works that have analyzed world-wide cultural knowledge through Wikipedia commonly focus on the largest Wikipedia editions, disregarding the long tail of medium-sized and small Wikipedias (Eom and Shepelyansky; Eom et al.; Hube et al.). This substantial reduction of the Wikipedia universe and its multi-culturality has, presumably, mostly practical reasons: for accurately identifying, e.g., entities and pages of particular types or analyzing the articles, studies often need language-specific tools, which are not yet readily available (or not easy to access) for most of the world's smaller languages. As already mentioned, we have considered 321 Wikipedias for our study. In this way, we were able to include many small languages and dialects. The two maps in [Figure 2](#) show how much the picture shifts as a

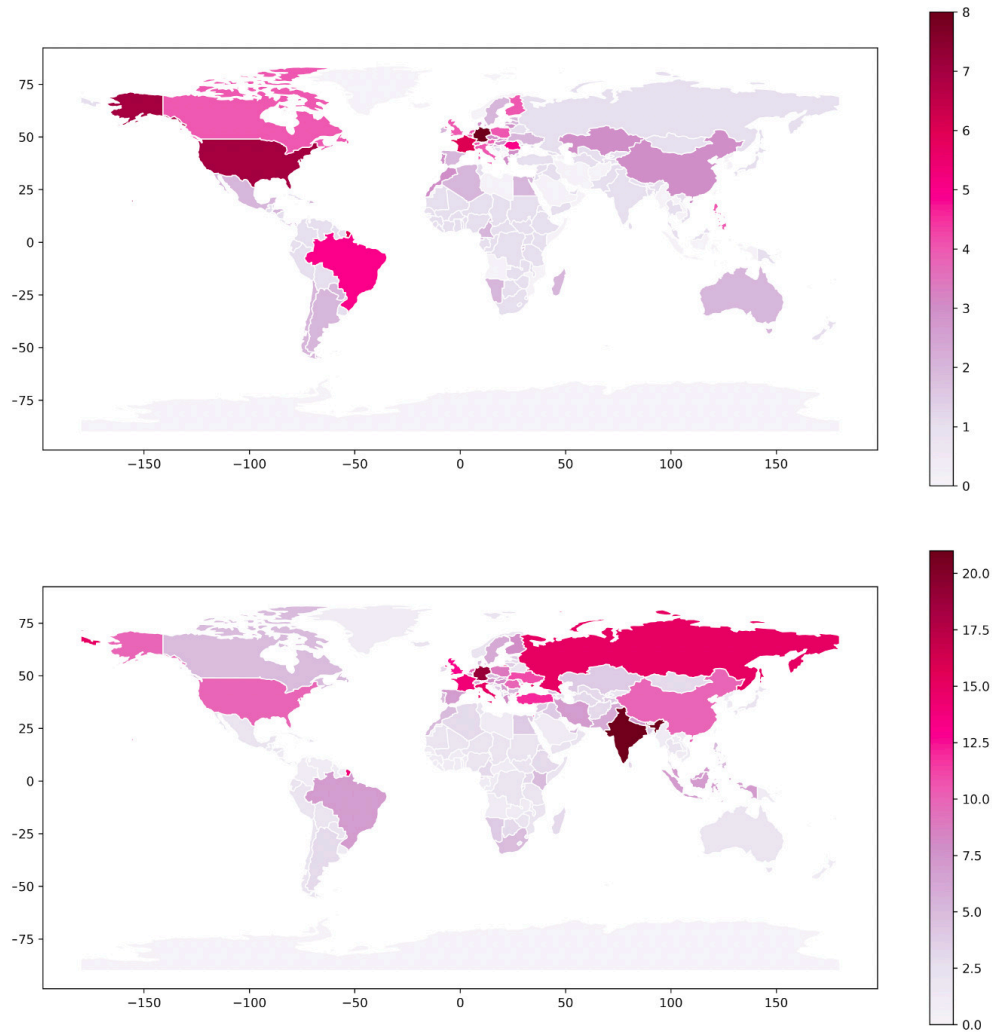


Figure 2. Number of Wikipedias (bottom = all; top = top 15 Wikipedias according to page number) that contain at least one literary character page, mapped to all languages spoken in the respective country.

result. We have geographically linked the Wikipedia languages back to nations where they are predominantly spoken, according to the territory-language information compiled by the Unicode Common Locale Data Repository and the geographic data provided by GeoPandas (Jordahl et al.). In particular, nations in which a large number of languages are being spoken, such as India, appear peripheral if we only consider the largest Wikipedias. The Top15-Wikipedias map presents a *République des lettres mondiale* that is largely coherent with the one as sketched by Pascale Casanova: The “old world” and the USA constitute a Greenwich meridian. A little bit more surprising may be that, next to it, there are Brazil, Canada, and China. The mapping of all languages shows a different meridian. While Europe is still well represented, we certainly cannot speak of “the West and the rest” anymore: India, Russia and Turkey are equally central. However, on both maps we observe the digital divide between economically more and less developed regions that has been discussed for Wikipedia (Börner and Kopf).

The issue of centre and periphery in world literature studies is commonly addressed via the classification of “major” and “minor” literatures. As Galin Tihanov summarizes, “minor” is usually understood in two different ways: Firstly, as “a potential for social and political energy that originates in the writing of a minority within a dominant majority” and secondly as “derivative, deprived of originality when measured by the yardstick of ‘mainstream literatures’” (Tihanov 212). While the first definition has become acquainted through Deleuze’s and Guattari’s *Essay on Kafka* and the “littérature mineure”, the second reaches back in European history, according to Tihanov to the 18th century. How present it is in European thinking shows Franco Moretti, who apologizes in his article *More Conjectures* for previously having made the separation between ‘original’ and ‘derivative’ (Moretti, “More Conjectures”). However, when we consider “center” and “periphery” in terms of representation in Wikipedia, neither of these definitions is suitable.

A different definition of center can be derived from our network. As the graph in [Figure 3](#) shows, the majority of languages share a large character pool, and the major editions also have their own character clusters, which appear here like satellites. Before we move on to the characters, we stick to the geographical aspect and ask which language editions are really central? [Table 3](#) lists the top 30 Wikipedias featuring the highest number of character pages in our data, along with the cumulative percentage of character pages they contribute to the overall data sets. As can be seen in [Table 3](#), ten Wikipedia editions cover 91 percent of all character pages. These 10 Wikipedias form the “center” of our Wikipedia-republic of world literature.

Compared to the overall ranking of Wikipedia editions, our ranking according to literary characters differs in some aspects, as shown by the global size ranks of our top 30 Wikipedias in [Table 3](#). The English edition holds the stable pole position and some of the editions (Italian, Russian, French, Spanish and Dutch) are represented in both cases. Wikipedias that are entirely or significantly created by the Lsjbot and among the top ten in the overall ranking (Cebuano 2, Swedish 4) are not in our character-based ranking. In fact, Cebuano has only four articles about characters. Surprisingly the German edition is not among our top ten (position 3 in the overall ranking). Conversely, editions that are not among the top ten in the overall ranking (Polish, Japanese, Mandarin, Portuguese) appear on top in our character-based ranking. The plot in [Figure 4](#) illustrates the relationship between the global size ranking of the 181 Wikipedias and the ranking according to their number of character pages: while there is, unsurprisingly, an overall trend for smaller Wikipedias to contain less character pages, there are many outliers and exceptions in both directions, i.e. bigger Wikipedias featuring rather few and smaller Wikipedias featuring rather many characters.

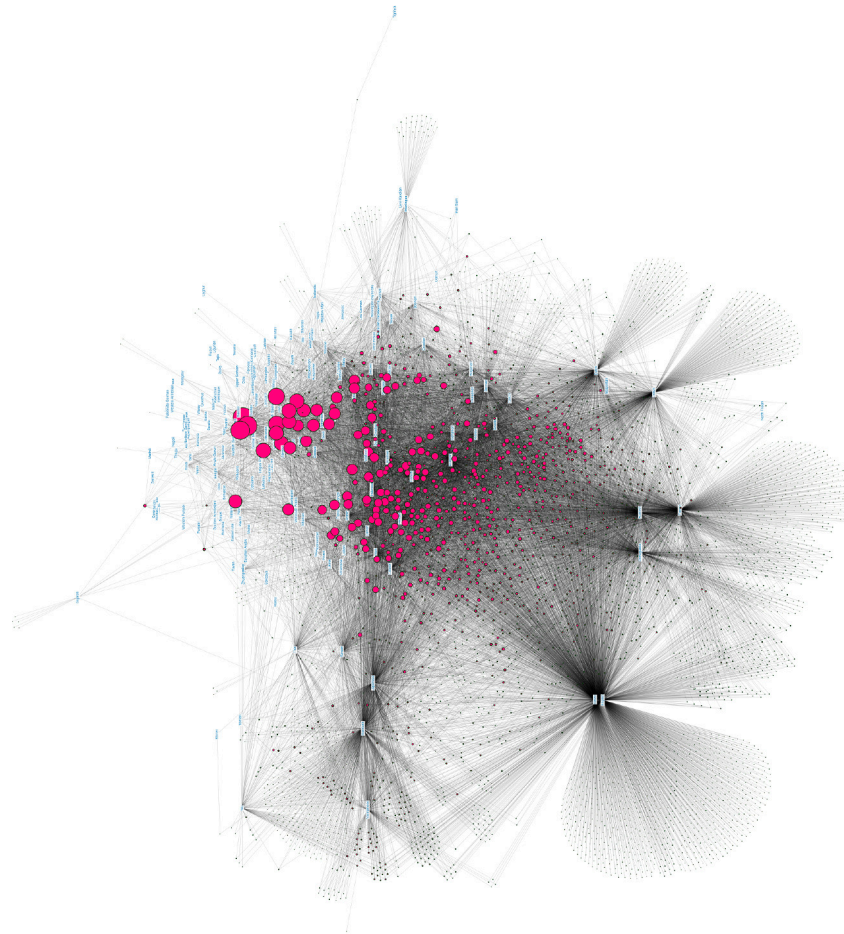


Figure 3. Visualisation of character network with nodes scaled according to number of links (available at <https://clause-bielefeld.github.io/wikipedia-character-network/> and persistently stored at <https://doi.org/10.7910/DVN/2VA5FQ>)

Table 3. Top 30 Wikipedias ranked by number of character pages in our data, with cumulative percentage of characters, the current rank of the Wikipedia in terms of its global size, and its betweenness centrality (Faust) in the full network

	Language	N _{character}	Cum.%	Global size rank	Centrality	Code
1	English	1901	0.41	1	0.31	en
2	Italian	1461	0.56	9	0.24	it
3	French	1254	0.65	5	0.17	fr
4	Spanish	983	0.71	8	0.10	es
5	Mandarin	865	0.76	15	0.11	zh
6	Polish	770	0.81	11	0.10	pl
7	Russian	767	0.84	7	0.07	ru
8	Japanese	756	0.87	12	0.07	ja
9	Dutch	747	0.89	6	0.07	nl
10	Portuguese	628	0.91	18	0.05	pt
11	German	474	0.92	3	0.03	de
12	Swedish	448	0.93	4	0.02	sv
13	Ukrainian	372	0.93	17	0.01	uk
14	Finnish	368	0.94	25	0.03	fi
15	Turkish	355	0.95	28	0.02	tr
16	Czech	340	0.96	27	0.02	cs
17	Catalan	337	0.96	20	0.02	ca
18	Korean	325	0.97	23	0.01	ko
19	Vietnamese	317	0.97	13	0.01	vi
20	Indonesian	307	0.97	22	0.01	id
21	Hebrew	276	0.97	37	0.01	he
22	Arabic	262	0.97	16	0.01	ar
23	Persian	247	0.97	19	0.00	fa
24	Hungarian	240	0.97	26	0.01	hu
25	Romanian	236	0.97	32	0.00	ro
26	Bulgarian	236	0.97	39	0.01	bg
27	Danish	218	0.98	40	0.01	da
28	Thai	208	0.98	58	0.00	th
29	Croatian	190	0.98	47	0.00	hr
30	Estonian	171	0.98	44	0.01	et

Pascale Casanova defines the “literary prestige” of a nation or region in terms of a “professional ‘milieu,’” which includes “literary institutions, academies, juries, critics, reviews, schools of literature” and she follows Priscilla Clark Ferguson by adding some parameters, such as book publications and sales, time spent reading, number of publishers and bookstores (Casanova 15). Since Wikipedia is a collaborative effort, it is reasonable to assume that the number of Wikipedia entries about literature in general and characters in particular is also an indicator of the ‘literariness’ of a nation. In this perspective, we see overlaps, but also differences: English and French still form the center but for example Germany, which is the third essential nation of Casanova’s republic, is located beyond our top ten-centre in the proximate periphery (rank 11).

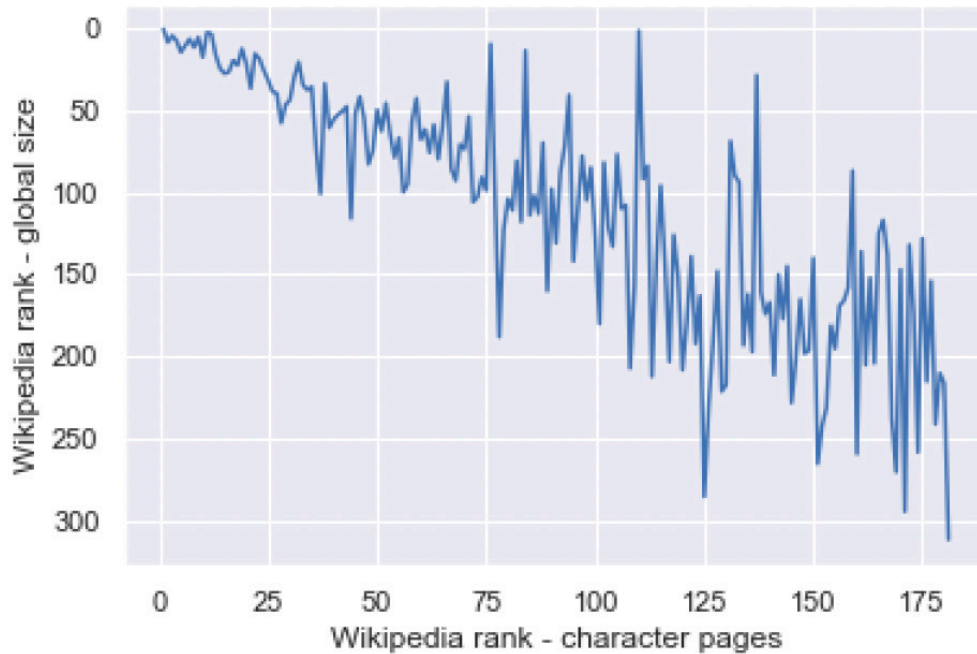


Figure 4. Illustration of the poor correlation between Wikipedias ranked by their number of character pages (x-axis) and Wikipedias ranked by their global size (y-axis)

Based on the network, we can score the importance or centrality of characters in the Wikipedia universe by means of different metrics. The picture in our centre changes slightly when we look at the “betweenness centrality,” (Faust) a metric that is based on the number of shortest paths that pass through a node. As can be seen in [Table 3](#), the English edition is the most central but there are some shifts in the ranking: The Mandarin Wikipedia edition has less pages than the Spanish, but is more central, the Finnish Wikipedia has less pages than the Swedish and Ukrainian, but is more central.

As the English edition covers not even half of all character pages in Wikipedia (41 percent), other Wikipedias in our centre make a substantial contribution to the variety of characters. For each of the 181 Wikipedias in our network, we compute the portion of its character pages that also have a page in the English Wikipedia. [Figure 5](#) illustrates the general relationship between Wikipedias ranked by their number of characters and their overlap with the English Wikipedia’s characters (i.e. most of the small Wikipedias overlap completely with the English Wikipedia). Among the Wikipedias that have the lowest overlap, we find some that belong to our top ten (Italian, Dutch, French, Polish, Mandarin). Casanova considers the center to be aesthetically and culturally largely homogeneous because it is constituted by a few languages that, moreover, share a literary tradition. With regard to the cultures represented by the language editions in our center we can observe relatively little homogeneity.

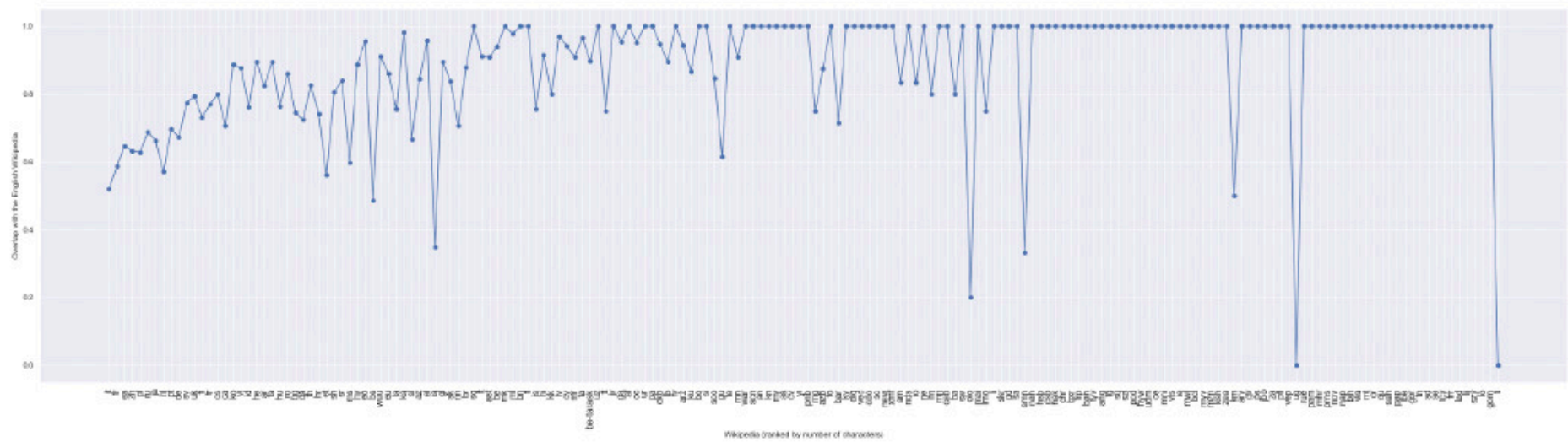


Figure 5. Wikipedias ranked by number of characters and their overlap with the English Wikipedia's character pages

There are, of course, many other opportunities to engage with the issue of center (i.e. translations, search interests) but the Wikidata analysis presented here unveil some interesting patterns. Although there is a set of ‘core’ languages which provide most of the character representations, this core is not identical with the triad of German, English, and the Romance languages that still dominates comparative and thus world literature studies. Our results also raise the question how strongly the euro-centric idea of the conversion of a (and one) national literature and a (and one) national language still influences our picture of what and how world literature is.

6. Character Canon – One or Many?

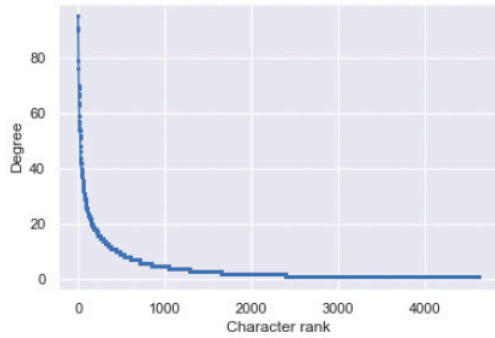
Which characters form the canon of the Wikipedia universe? Our detailed analysis of the large variety of Wikipedia’s language editions allows us to determine the canon in a fine-grained manner. [Figure 6](#) shows the degree distribution of character nodes in our two bipartite networks (also see [Figure 3](#) and [Table 1](#)). Since the top-15 networks features only 15 Wikipedias, characters fall into 15 discrete groups ranging from degree 15 (i.e. characters linked to all of the top 15 Wikipedias) and degree 1 (i.e. characters linked to only one of the top 15 Wikipedias). The full network features a much more fine-grained ranking, as the characters in our dataset seem to be adopted to rather varying degrees in the smaller languages. Hence, broadly speaking, the most important characters also appear in the so-called minor languages and their distribution in these minor languages determines their centrality. In [Table 4](#), we illustrate the ranking of characters that emerges from the top-15 network (left columns in [Table 4](#)) and the full network). Already the 4th group (characters with a degree of 12) in the top-15 network includes 39 characters and the largest one includes 2256 characters, which makes a differentiated analysis difficult. In contrast, the full network with character pages covering a large variety of languages show a much more differentiated ranking of individual characters (in [Table 4](#), we do not top characters appearing in the same top work as we did in our analysis of autonomy in [Table 2a](#)).

The right columns showing the ranking in the full network in [Table 4](#) confirms the diagnosis of “Europe and the Other” as - except for Gilgamesh, Arjuna, and Maitreya - it represents mainly British and US-American characters. Remarkably, the non-European characters belong to canonical works (*Gilgamesh*, *Mahabharata*) or have a religious background (Maitreya) - in the common understanding of the canon only they can be considered as “valuable” or “sacred”. From the so-called Western tradition we meet only Pandora, all other characters are non-canonical, but can be considered classics of the respective genre (detective novel, superhero comic, adventure novel). Characters from the Harry Potter universe are significantly over-represented, which is consistent with our thesis that fan communities have an impact on the representation of literature in Wikipedia.

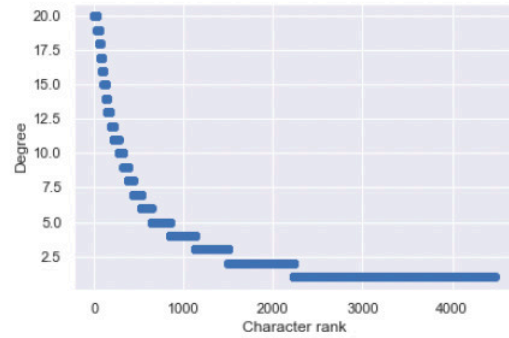
Table 4. Characters ranked by their degrees k (number of Wikipedias) in the full and the Top-15 network, bold names on the left appear in the top 15 on the right, n_{char} is the number of characters with the same degree k

Rank	Top-15 Network			Full Network		
	k	Characters	n_{char}	k	Characters	n_{char}
1	15	Sherlock Holmes	1	95	Sherlock Holmes	1
2	14	Tarzan, Superman, Santa Claus, Pandora	4	91	Superman	1
3	13	Wolverine, Winnie the Pooh, Gilgamesh, Peter Pan, Sinbad the Sailor, Scheherazade, Robert Langdon, Count Dracula, Arsène Lupin	10	90	Santa Claus	1
4	12	Maitreya, Arjuna, Ron Weasley, Harry Potter, Hermione Granger, Albus Dumbledore, Lord Voldemort, Severus Snape, Gandalf, Zorro, Terminator, Sun Wukong, Slender Man, Sauron, Samantabhadra, ...	39	79	Gilgamesh	1
5	11	Rubeus Hagrid, Zwarte Piet, William Riker, Vito Corleone, Uhura, Tuvok, Till Eulenspiegel, Spawn, Smaug, Shelob, Seven of Nine, Samwise Gamgee, ...	49	76	Harry Potter	1
6	10	Zefram Cochrane, Worf, Wesley Crusher, Vaiśravaṇa, Ungoliant, Trip Tucker, Tom Paris, Tinker Bell, Sweeney Todd, Susan Pevensie, Spock, Snegurochka, Sirius Bla...	43	70	Tarzan	1
7	9	Zhu Bajie, White Rabbit, Victor Frankenstein, Vesper Lynd, Travis Mayweather, Tom Sawyer, The Hatter, Tasha Yar, T-1000, Sinterklaas, Sif, Reepicheep, Quasimod...	58	69	Winnie the Pooh, Ron Weasley, Hermione Granger	3
8	8	Winnetou, White Witch, Wendy Darling, Wedge Antilles, Watcher in the Water, Utnapishtim, Tom Hagen, Tigger, The Shadow, Sybill Trelawney, Shylock, ...	69	67	Pandora	1
9	7	Village idiot, Zuko, Zaphod Beeblebrox, Xuan Wu, Winston Smith, Violet Baudelaire, Tsathoggua, Trillian, Susan Calvin, Stephen Maturin, Sonny Corleone, Solomo...	89	66	Lord Voldemort	1
10	6	Wyatt Halliwell, Willy Wonka, William of Baskerville, Vladimir Harkonnen, Urizen, Trinity, Tom Thumb, Tom Ripley, Superintendent Battle, ...	113	64	Albus Dumbledore	1
11	5	Éponine, Zhou Cang, Xuanzang, Wu Song, Woland, White Dragon Horse, Wellington Yueh, Wang Zhi, Walter Mitty, Virūḍhaka, Virūpākṣa, Van Veeteren, ...	158	63	Arjuna	1
12	4	Zou Yuan, Zou Run, Zhu Wu, Zhu Tong, Zhu Gui, Zhou Tong, Zheng Tianshou, Zhao Min, Zhao Cen, Zhang Wuji, Zhang Shun, Zhang Qing, Zhang Heng, ...	329	59	Severus Snape, Peter Pan	2
13	3	Þorbjörg Lítillólva, Zuvembie, Zhu Fu, Zhou Zhiruo, Zankou, Yugo Amaryl, Yig, Yan Zheng, Yan Ming, Xue Baochai, Ximen Qing, Xie Xun, Xiahou De, ...	350	57	Rubeus Hagrid	1
14	2	Æschere, Zhenyuan Daxian, Zheng Biao, Zhang Guifang, Zhang Cuishan, Zerbino, Yūgao, Yuuko Sakaki, Yutaka Seto, Yuri Bohun, Yunzhongzi, ...	672	56	Gandalf	1
15	1	captain Vrungel, Ōmyōbu, Ōmiya, Ōmi no kimi, Ōigimi, Úrsula Iguarán, Étienne Lousteau, ...	2256	55	Wolverine, Maitreya	2

The result in [Table 4](#) is interesting with regard to the assumption that “age” can provide information about the cultural capital of literatures and literary nations; as Pascale Casanova put it: “The age [...] testifies to wealth” (Casanova 14). In our canon at least 5 characters are actually ‘old’ regarding the date of origin (Gilgamesh, Arjuna) or mythological and thus somehow ageless (Santa Claus, Maitreya, Pandora). In comparison, Sherlock Holmes, Superman, Wolverine, or Tarzan seem young - but even they are of a respectable age compared to younger representatives of their respective genres and can be considered as classics.



(a) Degree distribution of character nodes in the full network



(b) Degree distribution of character nodes in the top-15 network

Figure 6. Degree distribution for character nodes in the full and the top-15 network

Table 5. Top-15 characters in terms of their betweenness centrality (Faust) in the full network

Rank _{N-wiki}	Character	N_{wiki}	Centrality
12	Arjuna	63	0.006367
0	Sherlock Holmes	95	0.005796
2	Santa Claus	90	0.005779
1	Superman	91	0.005598
3	Gilgamesh	79	0.005233
4	Harry Potter	76	0.004269
7	Winnie the Pooh	69	0.004160
5	Tarzan	70	0.004118
6	Hermione Granger	69	0.003862
9	Pandora	67	0.003694
8	Ron Weasley	69	0.003455
14	Severus Snape	59	0.003413
17	Maitreya	55	0.003159
13	Peter Pan	59	0.003026
10	Lord Voldemort	66	0.003008

Another possibility to determine the canonicity of characters is to compute their centrality as nodes that connect different languages and Wikipedias in the network. In [Table 5](#), we show the top-15 characters according to their betweenness centrality (Faust) (calculated based on the number of shortest paths that pass through the node): here, Arjuna who is not among the top-10 in terms of the number of Wikipedia pages becomes the most central character, which indicates that Arjuna directly connects parts of the network (Wikipedias) that would be less directly linked, if the Arjuna node would be missing from the network.

Metaphorically speaking, in our Wikipedia republic of literary characters Arjuna's function is comparable to the one Paris performs in Casanova's *Republique*. It is the center of a dense network, that simplifies or enables connections between characters from different literary languages just like

Table 6. Characters that occur once among the top-15 Wikipedias (according to global size) and more than 4 times in the smaller Wikipedias

Character	N_{wiki}	Wikipedias	$N_{top15wiki}$
Shabari	14	en-gu-hi-id-jv-mai-ml-mr-ne-pa-pnb-sa-ta-te	1
Nagini	8	bs-da-et-hr-hu-id-ms-nl	1
Byomkesh Bakshi	8	en-bn-gu-he-hi-id-mr-te	1
Oliver Wood	7	fa-ia-id-ja-ms-sl-uk	1
Feluda	7	en-as-bn-eo-gu-ml-mr	1
Albus Severus Potter	7	ar-es-gl-he-sl-tr-uk	1
Bamsı Beyrek	5	en-ar-az-azb-tr	1
Fluffy	5	ar-nl-pt-sh-tr	1
Jason Grace	5	fi-he-pt-tr-zh	1
Friday	5	en-cs-eo-et-sk	1
Carl Hamilton	5	da-fi-nn-pt-sv	1
Karabas Barabas	5	bg-hy-lt-ru-uk	1
Yeshua Ha-Notsri	5	ar-bg-ru-tt-uk	1
Walburga Black	5	bs-fa-id-ja-ms	1

Paris brought together writers that were political dissidents and expats. Although not as popular as Sherlock Holmes or Scheherazade, Arjuna seems to provide a link between the global West and East, North and South.

The analysis of the cumulative character distribution across Wikipedias ([Table 3](#)) and their overlap with the English Wikipedia ([Figure 5](#)) already showed a strong tendency for the smaller Wikipedias to include central characters from the top-15 Wikipedias rather than the national ones. We now further refine this picture and deliberately search for characters that do not appear at all or only once in the top-15 and more than 5 times in the smaller Wikipedias. [Table 6](#) shows the result: there are only 14 characters in the full network that have this property, and all of them appear once in one of the big Wikipedias.

This mirrors the distribution of power between the small and big languages in the Wikipedia republic. However, it is surprising that characters from English language works like those from the *Harry Potter* or the *Percy Jackson* universe are not all represented in the English (or some even not in one of the Western) Wikipedia(s): Oliver Wood and Walburga Black=Japanese, Nagini and Fluffy=Dutch, Albus Severus Potter=Spanish, Jason Grace=Mandarin. In fact, the only character from an English language work that is represented in the English edition is Friday from *Robinson Crusoe*. Hence, when we measure the characters' global status in numbers of character representations, we, again, observe that the small languages make an important contribution to the overall network.

At the same time, hitherto hidden characters come to light that have an overarching impact within language clusters: Shabari from the *Ramayana* as well as the Bengali PIs Byomkesh Bakshi and Feluda within an Indian-

Indonesian cluster, the fairy tale character of Bamsı Beyrek within a Turkic languages cluster, Tolstoy's Karabas Barabas and Bulgakov's Yeshua Ha-Nosri predominantly in a Slavic languages cluster, and the Swedish Carl Hamilton in a Scandinavian languages cluster. These characters are important with respect to particular regions and they point to a plurality and variety of canons that emerges when we chose a micro instead of the macro perspective that still dominates the conceptualization of world literature.

While many of the smaller Wikipedias do not feature their individual characters, still a significant amount is represented in only one or two editions and appear in our network like satellites of individual language editions. They can be seen as analogous to national canons and we can ask how national these national canons are. In [Table 7](#), we provide an overview of these “singleton” characters where we were able to retrieve a country of origin from the Wikidata entry of their associated work. This overview suggest that only the larger Wikipedia languages have their “own” character clusters. The overall picture is culturally diverse (i.e. the Turkish cluster is totally non-national) and only a few editions (the Finnish, Mandarin, and Latin) tend toward something that we can label as a ‘literary nationalism.’ However, there is a clear global divide because Western language editions produce primarily Western character pages (with the exception of the English edition). Comparable to this, non-Western Wikipedia editions such as the Japanese or Chinese have mainly pages of non-Western characters, adding again mainly US-characters (the Japanese has many Chinese, the Malaysian only Japanese, the Mandarin besides very many Chinese, some Japanese and so on).

We get an interesting picture of cultural relations when we look at the larger clusters shared by two and three Wikipedia language editions respectively, and describe them qualitatively. Some of these partnerships seem to have very homologous fan communities not only with regard to genre but even to a certain writer or work. For example, the French-Portuguese cluster contains almost exclusively characters from works by Honoré de Balzac (76 in total, only exception is a character from fantasy series *Spiderwick Chronicles*). The Catalan-French-Russian cluster shares only 13 characters, but all of them are taken from Marcel Proust's *À la recherche du temps perdu*. The preferences in the other clusters are more along the lines of the fantasy and thriller genre. In the Dutch-Spanish cluster 18 of 24 character pages are dedicated to the Japanese dystopia thriller *Battle Royal* while in the Dutch-Italian 11 of 15 concern characters from the fantasy novel series *Wheel of Time*. The 23 characters in the French-Italian cluster are mixed but homogeneous with regard to the genres (fantasy, science-fiction, children's literature).

The outcome in the single- and shared-character clusters contrast the impact ascribed to the category of national literature. This category not only determines disciplines in literary studies, it also structures both universities and the global book market. Paradoxically, even within concepts that indicate

Table 7. Wikipedias and their number of singleton characters in the full network, with distribution of countries for singletons

Wiki	Singletons	Countries for singletons
en	600	United States of America (268), United Kingdom (160), France (73), Japan (31), China (12), Australia (9), Canada (9)
it	418	United States of America (200), Italy (129), Roman Empire (29), United Kingdom (19), Japan (10), Republic of Ireland (6), Hungary (6), Germany (5)
fr	238	France (127), United States of America (68), United Kingdom (25), Canada (4), Russia (2), Spain (2), Roman Empire (2), Germany (2), Hungary (1)
pl	220	Poland (121), United States of America (51), France (21), Hungary (10), Italy (5), United Kingdom (4), Germany (3), Canada (1)
zh	195	China (135), Hong Kong (34), United States of America (7), Taiwan (6)
es	151	United States of America (69), United Kingdom (21), Spain (9), Argentina (9), Chile (9), Japan (6), Switzerland (6), Republic of Ireland (5), Venezuela (3), France (3)
nl	105	United Kingdom (41), United States of America (29), France (25), Netherlands (3), Germany (3), Japan (3), Sweden (1)
ru	100	Poland (37), Russia (42), United States of America (9), Hungary (6), France (4), Uruguay (1), United Kingdom (1)
ja	87	Japan (53), United States of America (22), United Kingdom (8), China (4)
fi	71	Finland (60), United States of America (10), Canada (1)
pt	50	Portugal (18), United States of America (15), United Kingdom (12), Brazil (4), Germany (1)
de	46	Germany (28), France (8), United Kingdom (6), Netherlands (1), United States of America (1), Roman Empire (1), Austria (1)
et	37	Estonia (19), Russia (17), Sweden (1)
sv	31	Sweden (23), United States of America (3), France (2), Finland (1), England (1), Italy (1)
da	26	United Kingdom (14), United States of America (8), Sweden (2), Denmark (1), Norway (1)
cs	25	United States of America (17), Czech Republic (4), Canada (2), Japan (1), United Kingdom (1)
ca	18	Spain (16), France (1), United States of America (1)
tr	13	United States of America (6), United Kingdom (5), France (1), Russia (1)
ia	13	United Kingdom (11), United States of America (2)
ar	10	Egypt (5), Near East (2), Japan (1), Syria (1), United Kingdom (1)
la	9	Roman Empire (9)
he	7	United States of America (7)
nn	6	United Kingdom (4), United States of America (2)
gu	5	India (5)

a transnational status of literature, such as European or world literature, national identity is deeply anchored (Zeman). From the point of view of the Wikipedia the impact category appears diminished. There obviously is no such thing as a literary nationalism that would guide the character representation.

To get an idea of the distribution of characters in a comparison between works that belong to the so-called central languages and those that are generally considered marginal (though not necessarily small), we take a closer look at the distribution structure. For this purpose we have examined four works in more detail. *Harry Potter* (261 autonomous character pages) and *Ulysses* (76 autonomous character pages) are our examples of a popular and a canonical text respectively (if one can make the distinction at all) from a central literary language. The Chinese novel *Water Margin* (154 autonomous character pages) is our non-European canonical text. As a popular text of a rather marginal literary nation we have chosen the Polish

novel series *The Witcher*. We deliberately decided against an obvious example like Japanese entertainment literature (fantasy or manga), because it has globally such a big impact that we hardly can speak of marginalization.

From James Joyce's canonical *Ulysses*, only the main characters Leopold (19) and Molly Bloom (7) as well as Stephen Dedalus (17) are represented more broadly. However, the distribution is restricted to the bigger and Western Wikipedias with the exception of Mandarin and Arabic. This again aligns with our observation that a proper representation depends on the small language editions. Also the distribution of characters from the canonical work *Water Margin* is geographically restricted and represents pages (i.e. of the character of Wu Yong, 11) only for Asian language editions with the sole exception of English. The distribution of the main characters from Andrzej Sapkowski's *The Witcher* saga (Geralt of Riva 29, Cirilla 11, Yennefer of Vengerberg 10) covers geographically Western and Eastern Europe with small language editions such as Bulgarian, Macedonian, or Belarussian. As we have expected the most diverse representations with regard to the variety of Wikipedia editions we find for the characters from the *Harry Potter* universe, such as Harry Potter (86) himself but also Albus Dumbledore (81) or Hermione Granger (76). However, there is a significant amount of characters (156) that are represented only 1 to 4 times. Again we can observe that not all of them are represented in the English-language Wikipedia edition (for example the Interlingua edition has a whole cluster of minor characters from the *Harry Potter* series and the character of Amos Diggory is represented only in the Dutch and Japanese editions etc).

In terms of a world literature that measures value in distribution and circulation, we need to define *Harry Potter* as the most canonical work in the Wikipedia universe. But the comparison reveals that the practices of distribution do not differ much for works from genres which are usually considered to be canonical and non-canonical, central or marginalized, Western or Non-Western literary nations: They are usually geographically-clustered (exception in *HP*) but not nationally restricted, and they tend to engage extensively with a work's characters. We observe what Steffen Martus defined as a non-selective attention ("selektionslose Aufmerksamkeit"): a valuation and representation practice obsessed with detail "that can find even the deficient interesting because it is historically 'significant'." (Martus 471) It is suitable for defining the practice in Wikipedia, since we observe not only a multiplicity of Wikipedia representations of individual, often marginal, characters (or at least an awareness that they should be represented, as evidenced in the Wikidata structure), but also specific forms of representation. These include, for example, an often extensive and detailed biography and sometimes genealogy, or self-made portraits of the characters. Sites such as Wikiquote assemble quotations whose selection is not based on any (at least identifiable) particular significance for the work or its reception but rather reflects the affection that the creators of the site show for the

character. In summary, it became evident that these practices do not differ in terms of genre or geographical or language-specific cultural areas and this, in turn, allows us to question what is considered to be canonical or non-canonical in world literature.

7. Conclusion

We would like to sum up our results with regard to the question what picture we get of world literature when we consider the reception-oriented perspective of Wikipedia and look at literary characters instead of works as its currency.

Firstly, our analysis confirms the importance of characters to readers and Wikipedia users. Nationally and transnationally organised fan communities participate in the Wikipedia republic of literary characters. They strengthen the autonomy and distribution of characters. What we can see from our initial analysis is that canonical literature and classics of non-canonized genres (detective novel, fantasy and science fiction) are prominently represented. Furthermore, there is a tendency towards a non-selective attention as an engagement with main and minor characters. Further work can reveal more about the forms of engagement by focusing on the information that are given about the characters, the quality of the page, the practices (i.e. tendency of non-selectivity also considering character-related information, use of pictures etc), and the role of translations and intermediality. Secondly, our analysis shows that the Wikipedia republic of literary characters presents a different picture of some core concepts of world literature studies. For example, the impact of small languages is bigger than commonly discussed. It provides a corrected picture of what can be considered a center and periphery and a 'meridian' that is more diverse than the one of the three nations presented by Casanova. Our network shows a transcultural entanglement of languages via literary characters and also a closer look at the individual national clusters negates a fixation on national literatures. However, we still observe a divide between the Western World and Asia as well as an almost entire absence of characters from the African continent (which mirrors the objectively-given digital divide).

World literature is usually measured in terms of distribution, influence, value, quality, an assumed universality, or importance. Instead of questioning these categories, we chose to redefine them methodologically. For example, we can understand importance from a supportive perspective. Then, the significance of characters is based on the support of the smaller languages. When we consider influence in terms of connecting rather than gate-keeping, we see that the non-Western character of Arjuna is central for a Wikipedia republic of literary characters. When we confront what is commonly considered to be valuable with distribution, we witness a bigger genre diversity. And when we replace the universalising macro-perspective of a 'oneworldliness' by a micro-perspective and take a look behind the curtain of the big languages, a

variety of transnational characters indicates the existence of particular canons. Finally, if we switch to a reception-oriented perspective, we witness value-creating practices that foreground transnationally-oriented non-selective attention rather than nationalism. The results introduced in this article are not meant to be Charles Sanders Peirce's famous 'final opinion' on characters in the Wikipedia universe but invite further exploration in relation to other statistics, i.e. translations, estimated sales, search interests, representation in fan fiction, or on social media.

Data Repository: <https://doi.org/10.7910/DVN/2VA5FQ>

Peer reviewer: Mads Rosendahl Thomsen (Aarhus University)

Submitted: May 01, 2022 EDT, Accepted: May 22, 2023 EDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

WORKS CITED

- Alderson-Day, Ben, et al. "Uncharted Features and Dynamics of Reading: Voices, Characters, and Crossing of Experiences." *Consciousness and Cognition*, vol. 49, Mar. 2017, pp. 98–109, <https://doi.org/10.1016/j.concog.2017.01.003>.
- Blakesley, Jacob. "The Global Popularity of William Shakespeare in 303 Wikipedias." *Memoria Di Shakespeare. A Journal of Shakespearean Studies*, no. 5, Dec. 2018, pp. 149–71, <https://doi.org/10.13133/2283-8759/14509>.
- . "World Literature According to Wikipedia Popularity and Book Translations: The Case of Modern Italian Poets." *Comparative Critical Studies*, vol. 17, no. 3, Oct. 2020, pp. 433–58, <https://doi.org/10.3366/ccs.2020.0373>.
- Börner, Ingo, and Susanne Kopf. *Wikipedia Studies: sprach- und literaturwissenschaftliche Zugänge zu Wikipedia*. Edited by Ingo Börner et al., 2018, pp. 212–27.
- Casanova, Pascale. *The World Republic of Letters*. Translated by Malcolm DeBevoise, Harvard University Press, 2007.
- Damrosch, David. *What Is World Literature?* Princeton University Press, 2003.
- Dijkstra, Katinka, et al. "Character and Reader Emotions in Literary Texts." *Poetics*, vol. 23, no. 1–2, Jan. 1995, pp. 139–57, [https://doi.org/10.1016/0304-422x\(94\)00009-u](https://doi.org/10.1016/0304-422x(94)00009-u).
- Eco, Umberto. "On the Ontology of Fictional Characters: A Semiotic Approach." *Sign Systems Studies*, vol. 37, no. 1/2, Dec. 2009, pp. 82–98, <https://doi.org/10.12697/sss.2009.37.1-2.04>.
- Eder, Jens, et al. "Characters in Fictional Worlds: An Introduction." *Characters in Fictional Worlds*, vol. 3, Nov. 2010, pp. 3–66, <https://doi.org/10.1515/9783110232424.1.3>.
- Eom, Young-Ho, et al. "Interactions of Cultures and Top People of Wikipedia from Ranking of 24 Language Editions." *Plos One*, vol. 10, no. 3, Mar. 2015, p. e0114825, <https://doi.org/10.1371/journal.pone.0114825>.
- Eom, Young-Ho, and Dima L. Shepelyansky. "Highlighting Entanglement of Cultures via Ranking of Multilingual Wikipedia Articles." *Plos One*, vol. 8, no. 10, Oct. 2013, <https://doi.org/10.1371/journal.pone.0074554>.
- Faust, Katherine. "Centrality in Affiliation Networks." *Social Networks*, vol. 19, no. 2, Apr. 1997, pp. 157–91, [https://doi.org/10.1016/s0378-8733\(96\)00300-0](https://doi.org/10.1016/s0378-8733(96)00300-0).
- Hall, Stuart. "The West and the Rest: Discourse and Power." *Essential Essays, Volume 2*, Duke University Press, 2018, <https://doi.org/10.1215/9781478002710-009>.
- Heath, Stephen. "Politics of Genre." *Debating World Literature*, edited by Christopher Prendergast, Verso Books, 2004, pp. 163–74.
- Henningsen, Lena. "Kosmopolitisches lesen. Weltliteratur während der chinesischen kulturellen Revolution." *Vergleichende Weltliteraturen / Comparative World Literatures*, edited by Dieter Lamping and Galin Tihanov, J.B. Metzler, 2019, pp. 235–51, https://doi.org/10.1007/978-3-476-04925-4_16.
- Hirakawa, Sukehiro. *Japan's Love-Hate Relationship with the West*. Brill, 2005, <https://doi.org/10.1163/9789004213821>.
- Hube, Christoph. "Bias in Wikipedia." *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, ACM Press, 2017, <https://doi.org/10.1145/3041021.3053375>.
- . "World Literature According to Wikipedia: Introduction to a DBpedia-Based Framework." *arXiv*, Jan. 2017, <https://arxiv.org/abs/1701.00991>.

- Jordahl, Kelsey, et al. *Geopandas/geopandas: v0.10.2*. Zenodo, 2021, <https://doi.org/10.5281/ZENODO.5573592>.
- Kensho. *Qwikidata v0.4.1*. 2022, <https://github.com/kensho-technologies/qwikidata>.
- Laffer, Alexander. “When Readers Talk about Characters as If They Were Real, How Do They Talk about Them? Empathy and Gossip in Reading Group Discourse.” *Poetics*, vol. 85, Apr. 2021, p. 101503, <https://doi.org/10.1016/j.poetic.2020.101503>.
- Lamping, Dieter. “Einleitung.” *Vergleichende Weltliteraturen / Comparative World Literatures*, 2019, pp. 1–9, https://doi.org/10.1007/978-3-476-04925-4_1.
- Latapy, Matthieu, et al. “Basic Notions for the Analysis of Large Two-Mode Networks.” *Social Networks*, vol. 30, no. 1, Jan. 2008, pp. 31–48, <https://doi.org/10.1016/j.socnet.2007.04.006>.
- Liebers, Nicole, and Romina Straub. “Fantastic Relationships and Where to Find Them: Fantasy and Its Impact on Romantic Parasocial Phenomena with Media Characters.” *Poetics*, vol. 83, Dec. 2020, p. 101481, <https://doi.org/10.1016/j.poetic.2020.101481>.
- Margolin, Uri. “Individuals in Narrative Worlds: An Ontological Perspective.” *Poetics Today*, vol. 11, no. 4, 1990, pp. 843–71, <https://doi.org/10.2307/1773080>.
- Martus, Steffen. *Werkpolitik: Zur Literaturgeschichte Kritischer Kommunikation Vom 17. Bis Ins 20. Jahrhundert Mit Studien Zu Klopstock, Tieck, Goethe Und George*. Walter de Gruyter, 2007, <https://doi.org/10.1515/9783110201772>.
- Moretti, Franco. “Conjectures on World Literature.” *New Left Review*, no. 1, 2000, pp. 54–68. ---. “More Conjectures.” *New Left Review*, no. 20, 2003, pp. 73–81.
- Müller, Gesine, et al., editors. *Re-Mapping World Literature: Writing, Book Markets and Epistemologies between Latin America and the Global South*. De Gruyter, 2018, <https://doi.org/10.1515/9783110549577>.
- Müller, Wolfgang G. “Interfiguralität. A Study on the Interdependence of Literary Figures.” *Intertextuality*, edited by Heinrich F. Plett, De Gruyter, 1991, pp. 101–21.
- Park, Sowon S. “The Pan-Asian Empire and World Literatures.” *CLCWeb: Comparative Literature and Culture*, vol. 15, no. 5, Dec. 2013, <https://doi.org/10.7771/1481-4374.2348>.
- Pianzola, Federico, et al. “Wattpad as a Resource for Literary Studies. Quantitative and Qualitative Examples of the Importance of Digital Social Reading and Readers’ Comments in the Margins.” *PLOS ONE*, vol. 15, no. 1, Jan. 2020, p. e0226708, <https://doi.org/10.1371/journal.pone.0226708>.
- Picard, Sophie, et al. “Zirkulation und Wertschöpfung am Beispiel literarischer Figuren.” *Der Wert der Literarischen Zirkulation / The Value of Literary Circulation*, edited by Michael Gamper et al., J.B. Metzler, 2022, pp. 431–53, https://doi.org/10.1007/978-3-662-65544-3_26.
- Shoreland, Adam. *Your Own Wikidata Query Service, with No Limits*. Oct. 2019, <https://addshore.com/2019/10/your-own-wikidata-query-service-with-no-limits/>.
- Stemberger, Martina. *Homer Meets Harry Potter: Fanfiction Zwischen Klassik Und Populärkultur*. Attempto Verlag, 2021.
- Tabassum, Shazia, et al. “Social Network Analysis: An Overview.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 5, 2018, p. e1256.
- Tihanov, Galin. *Do ‘Minor Literatures’ Still Exist?* Wuhan Guoyang Union Culture & Education Company Forum for World Literature Studies, 2015.
- Voit, Michael Matthias, and Heiko Paulheim. “Bias in Knowledge Graphs - An Empirical Study with Movie Recommendation and Different Language Editions of DBpedia.” *Schloss Dagstuhl - Leibniz-Zentrum Für Informatik*, vol. 93, 2021, p. 14:1-14:13, <https://doi.org/10.4230/OASICS.LDK.2021.14>.

- Werberger, Annette. "Weltliteratur Und Folklore." *Vergleichende Weltliteraturen / Comparative World Literatures*, 2019, pp. 323–41. Crossref, https://doi.org/10.1007/978-3-476-04925-4_22.
- Wikidata. *Wikidata Statistics*. 2022, <https://www.scribbr.com/chicago-style/website-citations/>.
- Wikipedia. *List of Wikipedias*. 2022, https://en.wikipedia.org/wiki/List_of_Wikipedias.
- Zeman, Mirna. "Literaturmoden und weltliteratur. Hypes um die gusla." *Weltliteratur in der Longue Durée*, 2021, pp. 135–60, https://doi.org/10.30965/9783846770597_008.
- Ziolkowski, Theodore. "Figures on Loan: The Boundries of Literature and Life." *Varieties of Literary Thematics*, 1983, pp. 123–51.