



7

Wikipedia as a Virtual Learning Site and a Multilingual Languageing Site

Sverker Johansson and Ylva Lindberg

Wikipedia is an online user-generated free encyclopedia, the largest and most popular general reference work of any kind. It is the fifth-most visited of all websites, behind behemoths like Google and Facebook, but ahead of others like Twitter and Instagram (Alexa, 2018), and it is the only non-profit site in the top 20. It has more than 40 million articles in more than 300 different languages and has half a billion unique visitors every month.

For many people, especially outside formal learning contexts, Wikipedia will serve as the first learning site when they seek knowledge—it is typically the first search engine hit on any general topic (Marchiori & Vieno, 2018). It is also a site of intense languaging in and across numerous languages, currently available in approximately 300 different languages (Wikimedia, 2018a), with tight cross-connections and collaboration

S. Johansson (✉)
Dalarna University, Falun, Sweden
e-mail: sja@du.se

Y. Lindberg
School of Education and Communication, Jönköping University,
Jönköping, Sweden

between language versions. In many of these languages, Wikipedia is the only encyclopedia of any kind available.

A substantial body of scholarly research on Wikipedia has emerged since its foundation, as reviewed in, for example Okoli, Mehdi, Mesgari, Nielsen, and Lanamäki (2012) and Wikipedia (2018a). Much of the research concerns Wikipedia accuracy (e.g. Giles, 2005) and bias (e.g. Greenstein & Zhu, 2018) as an encyclopedia, but there are also many studies of both Wikipedia writers (e.g. Xu & Li, 2015) and readers from different perspectives. Wikipedia is also used as a tool and data source in corpus linguistics and other fields (e.g. this chapter).

The first part of this chapter will be about the culture of the Wikipedia community, and about what shapes the social interactions on Wikipedia.

The second part will be about the process of article creation and growth on Wikipedia, in particular the interaction between direct human editing and bot editing. A substantial fraction of all Wikipedia articles is created and edited by software bots.

The third part, finally, will be a comparative study of argumentation structure on different language versions of Wikipedia. The structure, openness and global scope of Wikipedia makes it feasible to extract large parallel corpora of argumentative writing from many different cultures, debating many different topics. The first stage of the project is a quantitative study of Wikipedia arguments, comparing the volume of debate surrounding the same 1000 articles on 40 different language versions of Wikipedia. With this data, argumentation is compared between languages.

The results will contribute to our knowledge about how digital literacy is related to traditional literacy and how teachers can meet new demands on writing practices, by integrating social media (Baron, 2010; Erixon, 2012).

1 Community and Culture of Wikipedia

Most visitors to Wikipedia use the site as nothing more than a reference work, a place to find information on almost any topic. They do not interact with anybody while they are there, and do not form a community within Wikipedia.

At the same time, Wikipedia is also the home of a sizeable and active cybercommunity. The people who edit Wikipedia, who make Wikipedia into what it is, do interact, and do form a cyberculture, with its own norms and patterns of behavior. Community interactions take place mainly within the Wikipedia website. Everybody knows about the millions of encyclopedia articles, but there are also a similar number of Wikipedia pages dedicated to discussions within the community. The communication within Wikipedia is all text-based and asynchronous, but users who desire “live” communication may occasionally meet also in online chat rooms or even in physical venues.

Identity on Wikipedia can take all of the forms discussed in Lessig (2006): users can be anonymous or not, and can be connected with a physical person or not (cf. Johansson & Lindberg, *forthcoming*). The cultural norm is for a physical person to have a single persistent Wikipedia identity, but the identity or characteristics of the physical person need not be disclosed online. Disclosing another user’s physical identity is regarded as a serious infraction.

The Wikipedia community is quite large, but it is difficult to put a firm number on its size; many people contribute anonymously without registering and many registered identities are dormant and unused. There are currently some 77 million registered identities (“users”) on Wikipedia (Wikimedia, 2018a) but most of them are only sporadically active; less than one in a thousand do more than a handful of edits in any given month (Wikimedia, 2018b). There is a large passive majority and a small number of very active users (Oberhaus, 2017), making the distribution of community activity highly skewed. The most active users are all bots, which do around 25% of all Wikipedia editing—just the top five bots have done more than 100 million edits between them (Wikimedia, 2018c).¹ But the bots do not form a culture; they typically do not interact with each other. Instead, the core culture builders of Wikipedia are the modest-sized core of highly active human users, numbering some tens of thousands in total. These core users are also responsible for the bulk of the content that is most frequently read (Priedhorsky et al., 2007). This

¹ Disclosure: The first author of this chapter is the creator of one of the most active Wikipedia bots, and is also an active participant in the Wikipedia community in other ways.

group of people shapes the reference information that the rest of the world relies on (Matei & Britt, 2017). A cause for concern may be that active Wikipedia editors are not representative of the population as a whole, being overwhelmingly male (Wikipedia, 2011). Single well-educated people with no children are also overrepresented (Hill & Shaw, 2013).

The editors of each separate language version of Wikipedia form separate sub-communities with a fair degree of autonomy, and culture may differ between languages. The core values are broadly shared and are largely stable over time (Heaberlin & DeDeo, 2016), but norms and patterns may differ considerably between languages. Many language sub-communities are carried by just a handful of people, too few to form much of a culture of their own, but major languages may have hundreds of people active. The largest community by far, with some thousands of active users, is the English-language one; most of the others are much smaller.

The communities of Wikipedia are largely self-governed. The formal entity behind Wikipedia, the Wikimedia Foundation, mainly attends to technical issues and external relations and does not interfere with encyclopedia contents, which are left to the community of volunteers. The volunteers organize their communities as they see fit, and the community can be regarded as an adhocracy (Bennis, 1968; Matei & Britt, 2017), an ad hoc form of organization with an informal but nevertheless hierarchical power structure, allowing for flexibility and mobility within the organization. Wikipedia is open for everybody to edit and contribute, with all users in principle having equal say on content issues—but in reality, some users are much more equal than others. There is a fairly strong, if informal, hierarchy in the community. The hierarchy is largely meritocratic—high status grows from a long track record of constructive contributions to the encyclopedia, and to the community.

But even Wikipedia cannot be governed totally without formal powers. It is sometimes necessary to ban destructive users or delete unsuitable articles. These powers are granted to a small number of users, based on elections within the community. The people elected to these positions are typically those who are already high up in the informal hierarchy; outsiders need not apply.

2 Creation and Growth of Wikipedia Contents

The vision of the Wikipedia project is “Imagine a world in which every single human being can freely share in the sum of all knowledge” (Wikimedia, 2018d). Gathering and presenting the sum of all knowledge in a format accessible to every human being is a Sisyphean task, vastly beyond the reach of any plausible centralized organization. Wikipedia relies instead on the “wisdom of crowds” (Surowiecki, 2004), with encyclopedic knowledge being self-organized by the joint efforts of thousands of people without central control.

The Global Brain (Heylighen, [forthcoming](#)) refers to the potential for global self-organization of cognitive processes across cyberspace, leading to a collective intelligence. Wikipedia may be regarded as an early manifestation of the knowledge-organizing aspect of the global brain. This self-organization has created an encyclopedia that is an order of magnitude larger than anything published before, while maintaining an accuracy comparable to that of traditional encyclopedias (e.g. Giles, 2005).

Anybody and anything can write a Wikipedia article about anything—but anybody else can change what has been written, or even call for its deletion. Wikipedia articles are thus written in a social process with negotiations and heated arguments on dedicated discussion pages, one for each Wikipedia article. As Myers (2010) observes, English-language Wikipedians employ traditional means to support their arguments, but also Wikipedia-specific argumentative criteria, such as Neutral Point of View, No Original Research, and Verifiability. These criteria refer to cultural norms that have emerged within the Wikipedia community, which are arguably vital for maintaining the credibility of Wikipedia as an encyclopedia and source of knowledge.

Wikipedia is most obviously a large collection of articles on various topics. This is the only kind of content that most users notice. But behind the articles themselves is also a substantial body of other kinds of web pages that form an infrastructure maintaining Wikipedia as a viable encyclopedia. Articles are organized in categories, and the system of categories makes up an index of Wikipedia articles, making it easy to browse through

different articles on related topics, as well as to identify gaps where an article is missing. Categories, just like articles, can be created and edited by anybody.

There is also a special section of Wikipedia articles that are about the Wikipedia project itself. Many of the cultural norms mentioned above are written here, as well as project pages and various administrative pages (Wikipedia, 2018a, 2018b are two examples). Much of the social life and culture-building of the Wikipedia community takes place in this section.

For every Wikipedia page of every kind, there is a corresponding discussion page, where the contents of that page can be discussed and negotiated. These discussions can sometimes be voluminous, especially on controversial topics, sometimes much larger than the actual article. When viewing any Wikipedia page on a computer, the discussion page can be found in a separate tab at the top of the article. On English Wikipedia, the discussion tab is labeled “Talk”; other language versions will have a tab saying the equivalent in that language. When viewing Wikipedia on a mobile device, the discussion page is instead accessed through a button at the bottom of the page.² These discussion pages will be the focus of the next section.

An article or any other Wikipedia page can be created by anybody. If you want to create a new Wikipedia article about something that is missing in Wikipedia, you just type the article name into the search box in Wikipedia; if the article does not exist already, you will be given the option to create the article. An empty edit box will open, and article text can be typed and subsequently saved. A new article is created!

But the life expectancy of a random new article is very short. The Wikipedia community enforces rather strict standards for what kinds of topics are acceptable for articles. The main rule concerns what is called notability—is the topic sufficiently prominent, distinct, and well-known to deserve an article? Prominence must be verifiable through independent reliable sources. If the topic of a new article is not notable enough, the article will be promptly deleted. Typical cases of rapid deletion involve somebody creating an article about themselves, their small business, or

² Only visible to registered users.

their hobby band. This notability filtering is performed by the elected users mentioned in the previous section.

What is regarded as notable is itself subject to negotiation within the Wikipedia community, and may change over time. Standards of notability are determined separately on each language version of Wikipedia, but there are some common patterns. The general rule is that you cannot use Wikipedia to make something famous, it must be famous enough already; specifically, the topic must be covered by independent sources. This means that a company can never become notable through its own website and its own social media outlets; it must have attracted enough attention from business journalists to be mentioned in non-promotional texts. Standards of notability tend to be rigorous for topics that somebody may have a self-interest in promoting (businesses, artists, politicians ...) and laxer for topics that do not directly profit anybody (nature, science, geography, history...).

As noted, notability standards vary somewhat between Wikipedia languages. Which topics actually have articles written varies between languages, and is strongly affected by cultural differences (Gloor et al., 2015; Persson, 2011). Unsurprisingly, most Wikipedia languages have good coverage of topics within their own cultural sphere—all the kings of Sweden have articles on Swedish-language Wikipedia—but sparser coverage of culturally distant topics.

If an article survives this initial notability screening, text can be added by anybody, both the original author and anybody else. The original author has no special rights; there are strong cultural norms against any possessiveness (Wikipedia, 2018b). The article topic likewise has no special rights, which is a source of conflict with both people and companies who dislike what is written on Wikipedia about them. Public relations firms that try to “improve” Wikipedia coverage of their clients are a frequent but discouraged presence on Wikipedia, and the topic of paid editing was hotly debated recently, ending with a policy document (Wikipedia, 2018c). At the same time, advice to companies on how to manage their Wikipedia presence is published in seemingly reputable scientific journals (e.g. Kaplan & Haenlein, 2014).

Articles typically grow in both size and quality over time, though the pattern of growth may vary substantially between articles. Some articles

on marginally notable topics may remain largely unchanged for years, whereas other articles grow rapidly. Zhang, Ren, and Kraut (2018) identify three typical growth patterns which they label “fast riser”, “incremental grower”, and “late bloomer”, with fairly self-explanatory names.

In total, about 1.7 billion edits have been saved on all Wikipedia versions, in 50 million different articles. This is about three edits per second, on average around the clock since the project was launched in 2001. Of these edits, about 75% are made directly by human editors, and 25% are made by autonomous software bots, editing without direct human supervision (Wikimedia, 2018c). Bots are used for a wide variety of tasks on Wikipedia; almost any task that can be automated has a bot doing it. Typical uses include making the same small fix in many similar articles, for example changing the spelling of a common name, changing links to an external source that has moved, or moving a large group of articles from one category to another. Among bot tasks requiring a bit more intelligence is the identification and linking of matching articles in different language versions, or adding suitable illustrations to articles. Bots are also used to detect and clean up vandalism. But bots typically do not add new text to existing articles; that is a task more suitable for human editors.

Creating new articles from scratch, however, is an important bot activity. Each article should contain a basic skeleton structure which is highly standardized—a brief introductory sentence telling what the article is about, an infobox to the right with basic facts about the topic, links to related topics, categorization, and links to the same article in other languages. This skeleton is easier to handle for bots than for humans, and bots do create huge numbers of articles on certain types of topics. In total, about one-third of all Wikipedia articles were originally created by bots (16.2 million out of 49.1 million total; Wikimedia, 2018e). But this global average of one-third is unevenly spread across languages—each language version decides independently whether to allow bots to create articles or not, and some have allowed it on a massive scale and others not at all. Most of the 16.2 million bot-created articles are concentrated in a handful of languages—just Cebuano, Swedish, Dutch, Waray-Waray, and Vietnamese account for 11.4 million of the total, and more than half of all languages have a negligible number (Wikimedia, 2018e). Bot article creation is also highly skewed between bots—a few hundred bots have

created articles, but the single most productive one has authored more articles than all the others together.

But bot article creation is not suitable or even feasible for all conceivable article topics. Suitable topics have the following characteristics:

- Large number of similar units, each of which is notable enough to have its own article.
- Each unit can be described with a set of standardized basic facts, the same set for all units.
- Basic facts, preferably language-independent (numbers, coordinates, scientific names, ...).
- Open data bases exist that make the basic facts available in a standardized format.

Topics that have these characteristics include species of plants and animals, geographical features (towns, mountains, ...), languages, chemical elements, astronomical objects, authors, and many others. The bulk of the existing bot-created articles is either about species or about geography.

The growth of artificial intelligence may open up new possibilities in the future. The Wikipedia movement is actively gathering structured information in a language-independent central repository (Wikidata), and is looking into the possibility of creating article text on the fly from repository information, whenever somebody searches for a non-existent article (Kaffee, 2016); a simple version is already being tested (MediaWiki, 2018).

3 Argumentation Patterns in Different Wikipedia Languages

This is a quantitative study of argumentation patterns on Wikipedia discussion pages in 40 different languages. Two questions are addressed here:

- Which languages have similar argumentation patterns?
- Which words are overrepresented in argumentative discussion texts, as compared with encyclopedic article texts on the same topics?

4 Methods and Data

The corpus used is based on a list of the 1000 most important articles that every encyclopedia should have (Wikimedia, 2013a). This list is the result of a consensus reached among Wikipedia editors. The articles in the list are organized into nine different topics: language and literature; biography; philosophy and psychology; religion; social sciences; science; technology; arts and recreation; and history and geography.

Languages were hand-selected from a machine-generated table of how well different Wikipedia languages cover the 1000-article list (Wikimedia, 2013b). All the 40 languages selected have better than 97% coverage. Among all the languages with adequate coverage, 40 languages were hand-picked based on language size and on diversity—cultural, geographical, and linguistic (Table 7.1). The final analysis was done on a subset of 966 articles present in all 40 languages.

All the 38,000+ articles (966 articles from each of the 40 languages) were accessed during the same 24-hour period (April 24–25, 2015), together with the corresponding discussion pages (including archived discussions, if any). Quantitative features were extracted from the texts using software written by the first author of this chapter. The following features were used in the analysis:

- Article size (in bytes). This is a fairly good measure of text content in articles, but does cause some bias when comparing languages using different orthographies and writing systems. In Latin scripts, the number of bytes corresponds well enough to the number of letters, but in, for example, Chinese, one character may be counted as two or more bytes.
- Discussion size (in bytes). Same measure as article size applied to discussion pages.
- Article age (number of days elapsed since article was created).
- Number of unique participants in discussion. Counted as the number of different user signatures found on a discussion page.
- Average post size (in bytes). How many bytes of text are found between consecutive signatures?
- Density of links:
 - Internal links within Wikipedia system;
 - External hyperlinks to other websites.

Table 7.1 Languages used in the analysis

Language	Family	Region
Arabic	Semitic	Western Asia, North Africa
Bahasa Indonesia	Austronesian	Southeast Asia
Belarusian	Slavic	Eastern Europe
Bosnian	Slavic	Eastern Europe
Bulgarian	Slavic	Eastern Europe
Catalan	Romance	Western Europe
Chinese	Sino-Tibetan	East Asia
Croatian	Slavic	Eastern Europe
Czech	Slavic	Eastern Europe
Danish	Germanic	Western Europe
Dutch	Germanic	Western Europe
English	Germanic	Western Europe
Esperanto	(artificial)	–
Estonian	Uralic	Eastern Europe
Finnish	Uralic	Western Europe
French	Romance	Western Europe
German	Germanic	Western Europe
Hebrew	Semitic	Western Asia
Hungarian	Uralic	Eastern Europe
Italian	Romance	Western Europe
Japanese	(isolate)	East Asia
Korean	(isolate)	East Asia
Latin	Romance	(Western Europe)
Latvian	Baltic	Eastern Europe
Malay	Austronesian	Southeast Asia
Malayalam	Dravidian	South Asia
Norwegian (bokmål)	Germanic	Western Europe
Persian	Indo-Iranian	Western Asia
Polish	Slavic	Eastern Europe
Russian	Slavic	Eastern Europe
Serbian	Slavic	Eastern Europe
Spanish	Romance	Western Europe, Latin America
Swedish	Germanic	Western Europe
Tamil	Dravidian	South Asia
Thai	Kra-Dai	Southeast Asia
Turkish	Turkic	Western Asia
Ukrainian	Slavic	Eastern Europe
Urdu	Indo-Iranian	South Asia
Waray-Waray	Austronesian	Southeast Asia
Vietnamese	Austroasiatic	Southeast Asia

- Word statistics—frequency of all words, counted separately for articles and discussion pages. Each graphical form is counted separately. This part of the analysis is limited to languages in which words are separated by spaces (i.e. not e.g. Chinese).

It is found that all the discussion features were dependent on article size and age. This is a confound in the language analysis, as the articles in major European languages, especially English, are systematically larger and older than the articles in other languages.

In Fig. 7.1, the relation between articles size and discussion size is shown. It is obvious that there is a strong correlation. Discussion size grows approximately quadratically with article size.

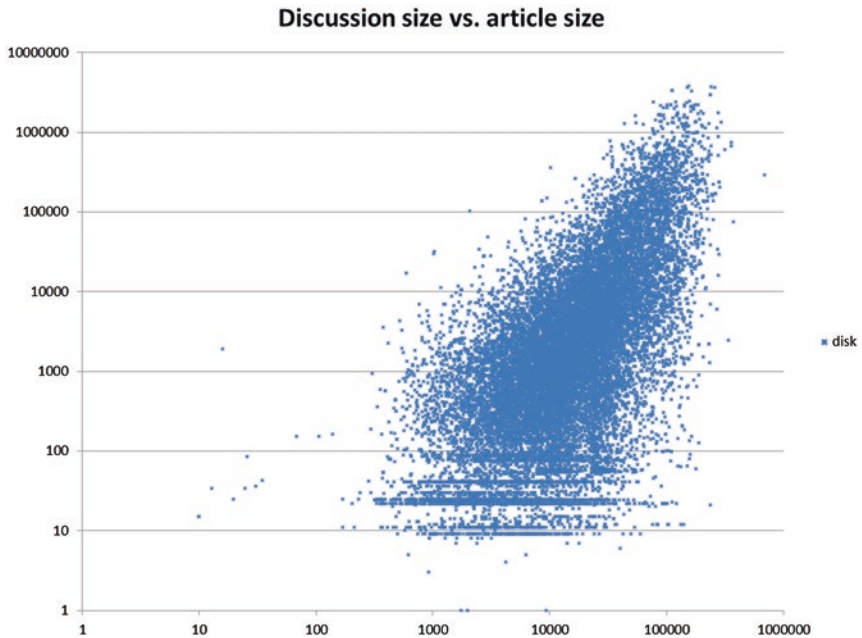


Fig. 7.1 Discussion size versus article size, pooled data from all languages. Each data point corresponds to a single article with associated discussion page. The horizontal axis shows article size in bytes and the vertical axis discussion size. Note logarithmic scale on both axes

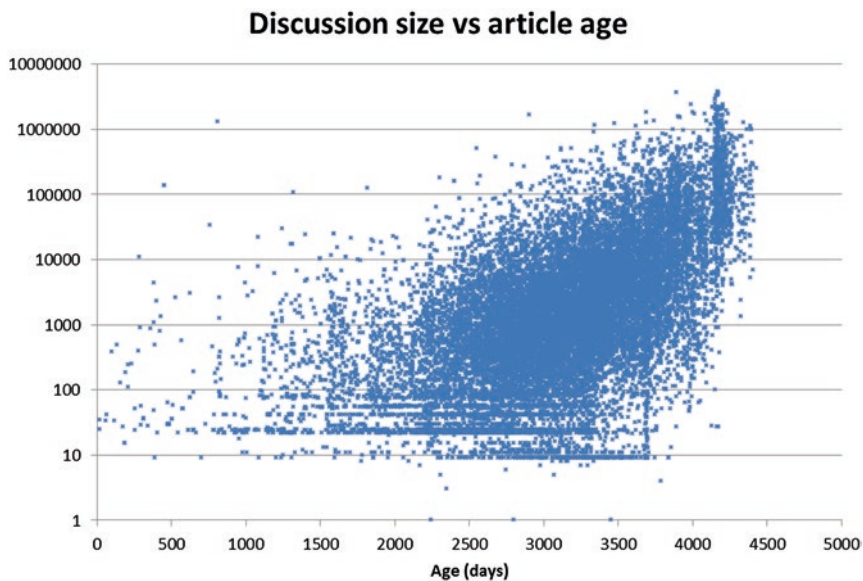


Fig. 7.2 Discussion size versus article age, pooled data from all languages. Each data point corresponds to a single article with associated discussion page. The horizontal axis shows article age in days and the vertical axis discussion size. Note logarithmic scale on vertical axis

Figure 7.2 shows the relation between discussion size and article age. Again, the correlation is obvious, but this time the growth is exponential instead.

The discussion size for each article is corrected for article size and age before the language analysis, so that the discussion size is divided by the average discussion size for all articles in the same size-age-bin.

In order to visualize which languages have similar discussion patterns, a language tree is constructed with a simple neighbor-joining algorithm (Saitou & Nei, 1987). All discussion parameters are log-normalized to zero average and unit variance before tree building.

In order to check the robustness and consistency of language-specific discussion patterns, the article sample was split in two parts, with roughly 500 articles in each. The two samples from each language were then treated as two different languages in the analysis

(“Swedish1” and “Swedish2”, etc.). If the patterns are robust, the two samples for each language should consistently cluster together in the language tree.

In the word analysis, word frequencies were compared between discussion pages and articles in the same language. For each word in each language, the statistical significance of any frequency difference between discussion and article was calculated, and a list of the most significant differences was generated.

5 Results

Normalized discussion size differs substantially between languages; the users of Hebrew are 15 times more talkative than the users of Waray-Waray. In Fig. 7.3a, b, the 40 languages are ordered by discussion size (corrected for article size and age).

There are some hints of patterns in Fig. 7.3. The top of the list is dominated by European languages, seven of the top ten are European, whereas there are more Asian languages at the bottom of the list. But it is difficult to draw any firm conclusions from this parameter alone.

The language tree in Fig. 7.4 also shows some patterns. This tree is based on the joint similarity between languages across all discussion parameters, not just discussion size. In the split-sample analysis, some but not all languages behaved in a consistent manner, with the two halves of the same language clustering together; these languages are marked in bold in Fig. 7.4.

In the lower part of Fig. 7.4, there are a few clusters that are fairly robust and that make sense culturally. There is one cluster near the bottom with East Asian languages (Chinese, Japanese, Korean), surrounded mainly by other Asian languages (Thai, Waray-Waray, Malayalam). Above it is a cluster with major European languages (Russian, French, German, Italian, English). In the center there is a large cluster containing mostly minor European languages, stretching from Czech to Hungarian in the figure, but this cluster contains also languages that are

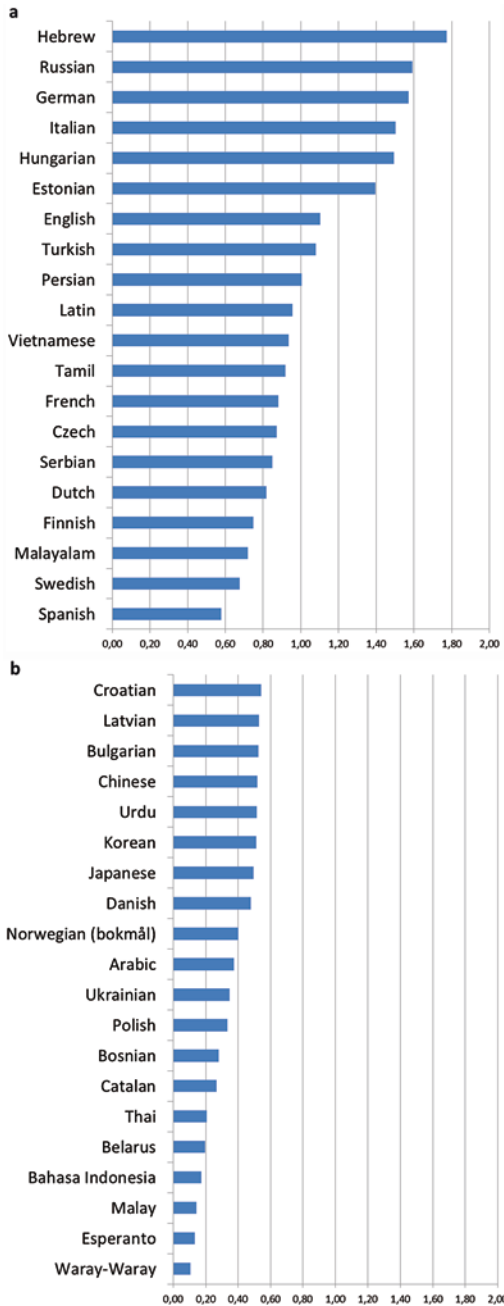


Fig. 7.3 Languages ordered by normalized discussion size (on horizontal axis: 1=global average). (a) Top half of the list. (b) Bottom half of the list

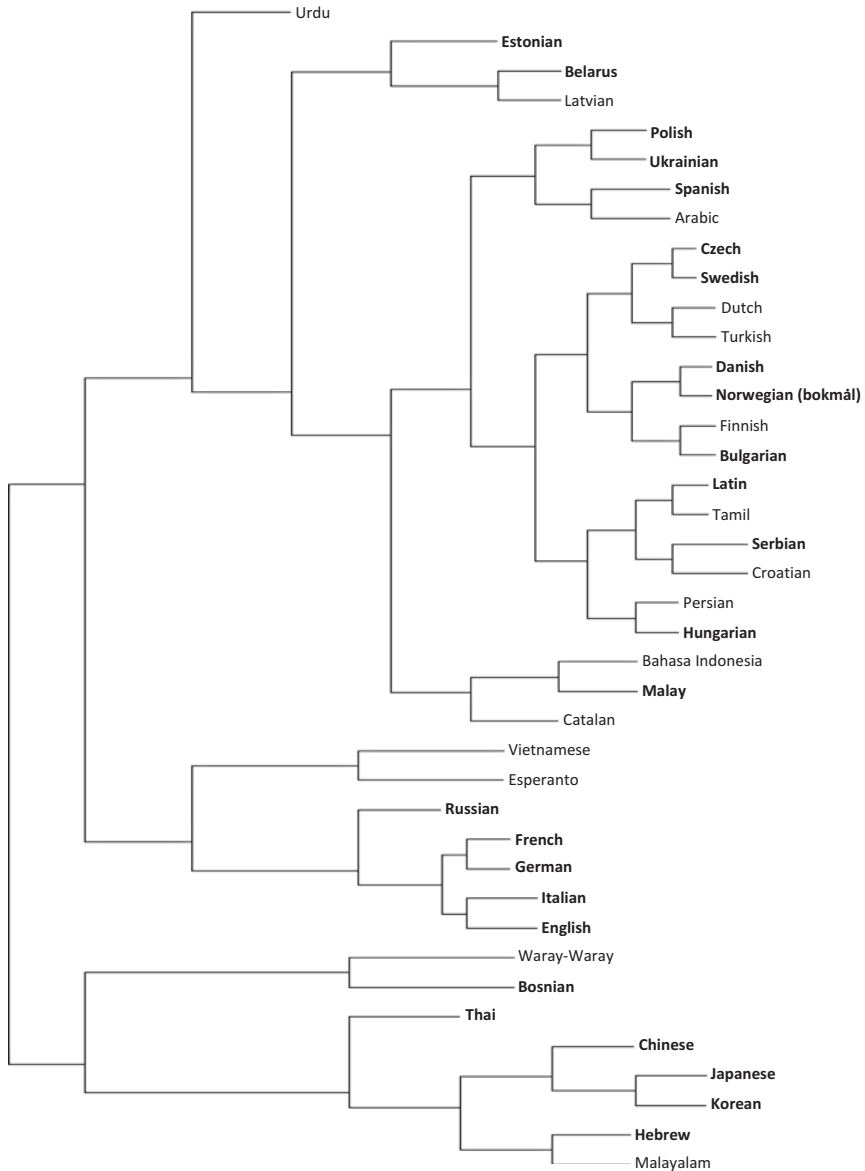


Fig. 7.4 Tree built through neighbor-joining of languages according to the distance between their normalized discussion parameters. Language names in **bold** are consistently placed when the article sample is split in half; the placement of non-bold languages should be treated with caution

culturally distant from Europe, such as Tamil and Persian. Near the top of the tree is what might be labeled a Baltic cluster, with Estonian, Latvian, and Belarus, three languages spoken in areas that largely share a common history, even though the languages are not closely related linguistically.

There are also smaller groupings that join languages that are culturally close. Having Danish and Norwegian together is unsurprising, as are the groupings of Serbian with Croatian and Malay with Indonesian. In all three cases, these languages are mutually understandable and are regarded as separate languages mainly for political reasons. But other groupings are more surprising. Why does Swedish cluster with Czech, rather than with its cultural siblings Danish and Norwegian? Dutch/Turkish, Latin/Tamil, and Hebrew/Malayalam are other language pairs in the tree with no obvious commonalities.

Selected results from the word frequency analysis are shown in Table 7.2 on the next page. The 12 words that are most significantly overrepresented in argumentative discussion page text, compared with the corresponding article texts, are shown for each of 12 selected languages.

Some patterns in the list are trivial, for example the presence of words meaning “article” in the various languages—the discussions are about articles and have frequent reason to refer to articles, whereas self-reference to “article” is naturally rare in the main article texts.

Other differences are more interesting, such as the much higher frequency of negations in the discussions. Article texts typically just describe what is, not what isn't, leading to a dearth of negations in the article genre. Discussion pages, in contrast, contain a fair fraction of exchanges of the type: “Is!”, “Is not!”, “Is too!”, and so on, leading to a high frequency of negations.

The use of pronouns is also significantly different between genres, with especially first- and second-person forms being much more frequent in discussion pages. This is also perhaps unsurprising, as there is rarely any reason to refer to either me or you in article texts.

Table 7.2 Words that are overrepresented in discussions compared with articles, in selected languages. Frequent word types are color-coded, with negations shown in blue, first-person forms in yellow, and second-person forms in peach

Swedish	Catalan	Danish	German	English	Esperanto	Spanish	Finnish	French	Indonesian	Korean	Arab
det	no	er	ich	you	mi	que	kello	pas	saya	있습니다	و
inte	que	jeg	nicht	article	ne	no	ei	je	artikel	화	لا
jag	l'article	ikke	Artikel	it	estas	es	etä	que	tidak	제가	ان
att	y	det	ja	not	pri	articulo	ole	en	Indonesia	일	ان
om	si	at	ist	this	artikolo	lo	on	en	ada	합니다	المقالة
är	he	du	mir	that	ke	me	se	ce	itu	금	المغل
artikeln	article	så	mal	is	vi	si	voisi	vous	saja	것	الله
du	Colom	artiklen	hier	think	sed	pero	olisi	l'article	bukan	토	هنا
så	ho	om	es	we	ĉu	eso	sen	à	istilah	월	تعم
man	hi	skal	habe	should	ni	hay	artikkeli	c'est	sudah	것은	الى
här	lo	har	man	your	oni	creo	artikkelin	de	kalau	목	هو
ju	em	artikel	du	don't	Esperanto	porque	nyt	me	karena	없습니다	ما

6 Conclusions

Wikipedia is interesting to study from many perspectives, and its huge influence as a learning site makes it important to explore and understand in depth. This knowledge source is not fixed, but dynamic, and in constant movement, though, regulated by a global community of Wikipedians. In addition, sub-cultures in different languages are operating according to locally set rules, for example, regarding which articles are notable, and to which extent bots are permitted in article construction. Some rules are not negotiable, such as Neutral Point of View, No Original Research, and Verifiability, but also the fundamental policy of non-possessiveness. Wikipedia is one rare digital forum where individual owner and authorship is annulled. Recognition is not acquired through fame, but rather through participative contributions to a range of articles and Wikipedia activities. Thus, the Wikipedia structure represents a frame of creative freedom and constraints, where participants engaged with the project strive toward collaborative knowledge construction based on “objectivity” and “facts”. Ongoing intrusive human practices show that these fundamental pillars constantly are threatened and in need of defense by the cybercommunity active on this social forum. Insights into languaging practices on Wikipedia offer an overview of the interplay between critical and digital literacy competences, as well as linguistic skills acquired, in order to fully take part in digitally and humanly generated processes of consistent knowledge building with access for all.

A portion of the world population gets a large fraction of their information from Wikipedia. The modest-sized community that produces Wikipedia texts thus has an influence far beyond its numbers, and the structure, culture, processes, and composition of the community are relevant to understand what information will be presented, and who the initiator is of the distributed information. For example, English-language articles are circulating worldwide to a much larger extent than articles in peripheral languages. Just as knowledge construction and language use on Wikipedia reflect cultural hegemonies on a global level, Wikipedia is also a congenial tool to diffuse knowledge about languages and cultures from global peripheries. The forum allows for and often requires inter-

connectedness and communication between and across languages and cultures that seldom meet in other publishing landscapes, for example the book market.

Just as Wikipedia has the potential to be a forceful tool in language and literacy learning, it is an important data source, and can as well be used for research, as a multilingual corpus, with easily accessible quasi-parallel texts in a very wide selection of languages. In this chapter results from a comparison of the argumentative genre between languages have been presented. Discussion pages and Wikipedia articles are largely written by the same authors with the same tools in the same context, and comparing the two thus removes many confounding factors, leaving only the genre difference between discussions and articles. The choice of words in argumentative texts shows similarities between languages, but macro-patterns of argumentation differ. The clusters of languages presented in the results indicate surprising connections between languages that culturally are conceived of as remote from one another. These findings need in-depth qualitative studies, in order to further understand the meaning of the connections, and to explore how Wikipedia languaging practices can be seen as a counterforce to monolingual and monocultural norms in the world.

References

- Alexa. (2018). The top 500 sites on the web. Retrieved March 21, 2018, from <https://www.alexa.com/topsites>
- Baron, S. (2010). *Always On*. Oxford: Oxford University Press.
- Bennis, W. (1968). *The Temporary Society*. New York: Harper & Row.
- Erixon, P.-O. (2012). Svenskämnet i ett nytt medieekologiskt sammanhang. In G. Skar & M. Tengberg (Eds.), *Svenskämnet i går, i dag, i morgon* (pp. 178–193). Stockholm: Svenskläraryörelsen/Natur & Kultur.
- Giles, J. (2005). Internet Encyclopaedias Go Head To Head. *Nature*, 438(7070), 900–901.
- Gloor, P. A., de Boer, P., Lo, W., Wagner, S., Nemoto, K., & Fuehres, H. (2015). Cultural Anthropology Through the Lens of Wikipedia – A Comparison of Historical Leadership Networks in the English, Chinese, Japanese, and German Wikipedia. In *Proceedings COINs15*. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1502/1502.05256.pdf>

- Greenstein, S., & Zhu, F. (2018). Do Experts or Crowd-based Models Produce More Bias? Evidence from Encyclopaedia Britannica and Wikipedia. *MIS Quarterly*, 42, 945–959. Retrieved from <https://www.hbs.edu/faculty/Pages/item.aspx?num=52723>
- Heaberlin, B., & DeDeo, S. (2016). The Evolution of Wikipedia's Norm Network. *Future Internet*, 8, 14. <https://doi.org/10.3390/fi8020014>
- Heylighen, F. (forthcoming). The Global Brain: Extending Symbolic Cognition Across a Planetary ICT Network. In A. Lock, C. Sinha, & N. Gontier (Eds.), *Oxford Handbook on Human Symbolic Evolution*. Oxford: Oxford University Press.
- Hill, B. M., & Shaw, A. (2013). The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. *PLoS One*, 8(6), e65782. Retrieved October 29, 2018, from <https://doi.org/10.1371/journal.pone.0065782>
- Johansson, S., & Lindberg, Y. (forthcoming). Cyberculture. In A. Lock, C. Sinha, & N. Gontier (Eds.), *Oxford Handbook on Human Symbolic Evolution*. Oxford: Oxford University Press.
- Kaffee, L.-A. (2016). *Generating Article Placeholders from Wikidata for Wikipedia: Increasing Access to Free and Open Knowledge*. B.Sc. Thesis, HTW Berlin – University of Applied Sciences. Retrieved from https://upload.wikimedia.org/wikipedia/commons/9/99/Generating_Article_Placeholders_from_Wikidata_for_Wikipedia_-_Increasing_Access_to_Free_and_Open_Knowledge.pdf
- Kaplan, A., & Haenlein, M. (2014). Collaborative Projects (Social Media Application): About Wikipedia, the Free Encyclopedia. *Business Horizons*, 57, 617–626.
- Lessig, L. (2006). *Code 2.0: Code and Other Laws of Cyberspace*. New York: Basic Books.
- Marchiori, M., & Vieno, E. B. (2018). *The Battle for Information: Exposing Wikipedia*. 2018 IEEE 16th Int. Conf. on Dependable, Automatic & Secure Comp. Retrieved from <https://ieeexplore.ieee.org/document/8512003>
- Matei, S. A., & Britt, B. C. (2017). *Structural Differentiation in Social Media – Adhocracy, Entropy, and the “1 % Effect”*. Berlin: Springer International.
- MediaWiki. (2018). *Extension: ArticlePlaceholder*. Retrieved March 31, 2019, from <https://www.mediawiki.org/wiki/Extension:ArticlePlaceholder>
- Myers, G. (2010). *The Discourse of Blogs and Wikis*. London: Continuum.
- Oberhaus, D. (2017). *Nearly All of Wikipedia Is Written by Just 1 Percent of Its Editors*. Retrieved October 15, 2018, from https://motherboard.vice.com/en_us/article/7x47bb/wikipedia-editors-elite-diversity-foundation

- Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F. Å., & Lanamäki, A. (2012). *The People's Encyclopedia Under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia*. Retrieved from SSRN <https://ssrn.com/abstract=2021326> or <https://doi.org/10.2139/ssrn.2021326>
- Persson, R. S. (2011). Ability Climates in Europe as Socially Represented Notability. *High Ability Studies*, 22(1), 79–101. <https://doi.org/10.1080/13598139.2011.576086>
- Priedhorsky, R., Chen, J., Lam, S. (Tony), Panciera, K., Terveen, L., & Austin, S. (2007). Creating, Destroying, and Restoring Value in Wikipedia. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work. Conference on Supporting Group Work* (pp. 259–268). ACM Press. <https://doi.org/10.1145/1316624.1316663>.
- Saitou, N., & Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, 4(4, July), 406–425.
- Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Anchor Books.
- Wikimedia. (2013a). *List of Articles Every Wikipedia Should Have*. Retrieved January 30, 2013, from https://meta.wikimedia.org/w/index.php?title=List_of_articles_every_Wikipedia_should_have&oldid=5154330
- Wikimedia. (2013b). *List of Wikipedias by Sample of Articles*. Retrieved January 30, 2013, from https://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias_by_sample_of_articles&oldid=4972742
- Wikimedia. (2018a). *List of Wikipedias*. Retrieved October 15, 2018, from https://meta.wikimedia.org/wiki/List_of_Wikipedias
- Wikimedia. (2018b). *Wikipedia Statistics – Active Wikipedians*. Retrieved October 15, 2018, from <https://stats.wikimedia.org/EN/TablesWikipediansEditsGt5.htm>
- Wikimedia. (2018c). *Wikipedia Statistics – Bot Editing Activity*. Retrieved October 15, 2018, from <https://stats.wikimedia.org/EN/BotActivityMatrixEdits.htm>
- Wikimedia. (2018d). *Wikimedia Foundation Vision*. Retrieved October 15, 2018, from <https://wikimediafoundation.org/about/vision/>
- Wikimedia. (2018e). *Wikipedia Statistics – Bot Article Creations*. Retrieved January 1, 2019, from <https://stats.wikimedia.org/EN/BotActivityMatrixCreates.htm>
- Wikipedia. (2011). *Wikipedia Editors Study*. Retrieved April 13, 2019, from https://upload.wikimedia.org/wikipedia/commons/7/76/Editor_Survey_Report_-_April_2011.pdf

- Wikipedia. (2018a). *Wikipedia: Academic Studies of Wikipedia*. Retrieved December 21, 2018, from https://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia
- Wikipedia. (2018b). *Wikipedia: Ownership of Content*. Retrieved December 30, 2018, from https://en.wikipedia.org/w/index.php?title=Wikipedia:Ownership_of_content&oldid=875873290
- Wikipedia. (2018c). *Wikipedia: Paid-contribution Disclosure*. Retrieved December 30, 2018, from https://en.wikipedia.org/w/index.php?title=Wikipedia:Paid-contribution_disclosure&oldid=858438932
- Xu, B., & Li, D. (2015). An Empirical Study of the Motivation for Content Contribution and Community Participation in Wikipedia. *Information & Management*, 52, 275–286.
- Zhang, H., Ren, Y., & Kraut, R. E. (2018). Mining and Predicting Temporal Patterns in the Quality Evolution of Wikipedia Articles. *Academy of Management Proceedings*, 2018(1). <https://doi.org/10.5465/AMBPP.2018.13746abstract>