

## 12 Collaborating on the Sum of All Knowledge Across Languages

Denny Vrandečić

Wikipedia is available in almost three hundred languages, each with independently developed content and perspectives. By extending lessons learned from Wikipedia and Wikidata toward prose and structured content, more knowledge could be shared across languages and allow each edition to focus on their unique contributions and improve their comprehensiveness and currency.

Every language edition of Wikipedia is written independently of every other language edition. A contributor may consult an existing article in another language edition when writing a new article, or they might even use the content translation tool to help with translating one article to another language, but there is nothing to ensure that articles in different language editions are aligned or kept consistent with each other. This is often regarded as a contribution to knowledge diversity since it allows every language edition to grow independently of all other language editions. So would creating a system that aligns the contents more closely with each other sacrifice that diversity?

### Differences Between Wikipedia Language Editions

Wikipedia is often described as a wonder of the modern age. There are more than fifty million articles in almost three hundred languages. The goal of allowing everyone to share in the sum of all knowledge is achieved, right?

Not yet.

The knowledge in Wikipedia is unevenly distributed.<sup>1</sup> Let's take a look at where the first twenty years of editing Wikipedia have taken us.

The number of articles varies between the different language editions of Wikipedia: English, the largest edition, has more than 5.8 million articles; Cebuano—a language spoken in the Philippines—has 5.3 million articles;

Swedish has 3.7 million articles; and German has 2.3 million articles. (Cebuano and Swedish have a large number of machine-generated articles.) In fact, the top nine languages alone hold more than half of all articles across the Wikipedia language editions—and if you take the bottom half of all Wikipedias ranked by size, together they wouldn't have 10 percent of the number of articles in the English Wikipedia.

It is not just the sheer number of articles that differ between editions but their comprehensiveness as well: the English Wikipedia article on Frankfurt has a length of 184,686 characters, a table of contents spanning eighty-seven sections and subsections, ninety-five images, tables and graphs, and ninety-two references—whereas the Hausa Wikipedia article states that it is a city in the German state of Hesse and lists its population and mayor. Hausa is a language spoken natively by forty million people and as a second language by another twenty million.

It is not always the case that the large Wikipedia language editions have more content on a topic. Although readers often consider large Wikipedias to be more comprehensive, local Wikipedias may frequently have more content on topics of local interest: the English Wikipedia knows about the Port of Călărași that it is one of the largest Romanian river ports, located at the Danube near the town of Călărași—and that's it. The Romanian Wikipedia on the other hand offers several paragraphs of content about the port.

The topics covered by the different Wikipedias also overlap less than one would initially assume. English Wikipedia has 5.8 million articles, and German has 2.2 million articles—but only 1.1 million topics are covered by both Wikipedias. A full 1.1 million topics have an article in German—but not in English. The top ten Wikipedias by activity—each of them with more than a million articles—have articles on only one hundred thousand topics in common. In total, the different language Wikipedias cover eighteen million different topics in over fifty million articles—and English only covers 31 percent of the topics.

Besides coverage, there is also the question of how up to date the different language editions are. In June 2018, San Francisco elected London Breed as its new mayor. Nine months later, in March 2019, I conducted an analysis of who the mayor of San Francisco was stated to be according to the different language versions of Wikipedia (see figure 12.1). Of the 292 language editions, a full 165 had a Wikipedia article on San Francisco. Of these, eighty-six named the mayor. The good news is that not a single Wikipedia

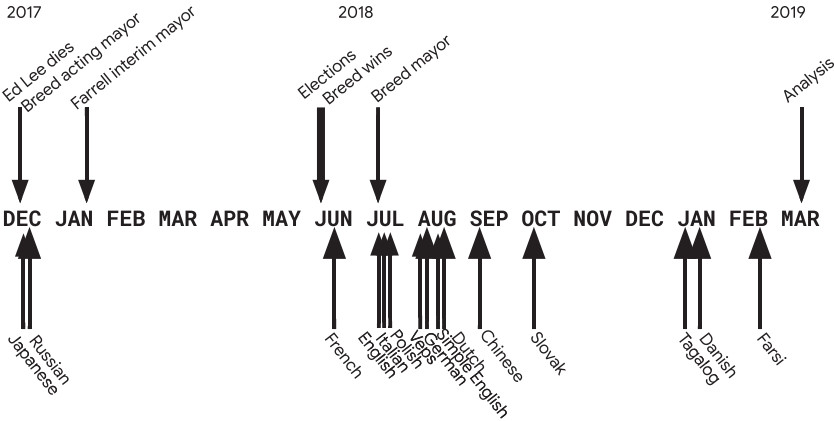


Figure 12.1

The events after the death of Ed Lee until London Breed became mayor on top. On bottom, date that a given Wikipedia was updated to list the new mayor.

lists a wrong mayor—but the vast majority are out of date. English switched the minute London Breed was sworn in. But sixty-two Wikipedia language editions list an out-of-date mayor—and not just the previous mayor Ed Lee, who became mayor in 2011, but also often Gavin Newsom (2004–2011), and his predecessor, Willie Brown (1996–2004). The most out-of-date entry is to be found in the Cebuano Wikipedia, who names Dianne Feinstein as the mayor of San Francisco. She had that role after the assassination of Harvey Milk and George Moscone in 1978 and remained in that position for a decade until 1988—Cebuano was more than thirty years out of date. Only twenty-four language editions had listed the current mayor, London Breed, out of the eighty-six who listed the name at all.

An even more important metric for the success of a Wikipedia are the number of contributors. English has more than thirty-one thousand active contributors—three out of seven active Wikipedians are active, with five or more edits a month, on the English Wikipedia. German, the second most active Wikipedia community, already only has 5,500 active contributors. Only eleven language editions have more than a thousand active contributors—and more than half of all Wikipedias have fewer than ten active contributors. To assume that fewer than ten active contributors can write and maintain a comprehensive encyclopedia in their spare time is optimistic at best. These numbers basically doom the mission of the

Wikimedia movement to realize a world where everyone can contribute to the sum of all knowledge.

## Enter Wikidata

Wikidata was launched in 2012 and offers a free, collaborative, multilingual secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, the other wikis of the Wikimedia movement, and for anyone else in the world.<sup>2</sup> Wikidata contains structured information in the form of simple claims, such as “San Francisco—Mayor—London Breed” qualifiers, such as “since—July 11, 2018,” and references for these claims—for example, a link to the official election results as published by the city—as shown in figure 12.2.

One of these structured claims would be on the Wikidata page about San Francisco, stating the mayor, as discussed earlier. The individual Wikipedias can then query Wikidata for the current mayor. Of the twenty-four Wikipedias that named the current mayor, eight were current because they were querying Wikidata. I hope to see that number go up. Using Wikidata more extensively can, in the long run, allow for more comprehensive, current, and accessible content while decreasing the maintenance load for contributors.<sup>3</sup>

head of government	London Breed <span>...</span>	
	start time	11 July 2018 <span>...</span>
	<span>▼</span> 2 references	
	reference URL	<a href="https://www.sfchronicle.com/politics/article/This-time-around-London-Breed-becomes-SF-mayor-13064707.php">https://www.sfchronicle.com/politics/article/This-time-around-London-Breed-becomes-SF-mayor-13064707.php</a>
	title	Big crowd at SF Civic Center sees London Breed sworn in as new mayor (English)
	stated in	San Francisco Chronicle
	author name string	Dominic Fracassa
	publication date	11 July 2018
	retrieved	3 May 2019
	reference URL	<a href="https://mayorlondonbreed.org/about/">https://mayorlondonbreed.org/about/</a>
		+ add reference
		+ add value

**Figure 12.2**

The statement in Wikidata about London Breed being mayor of San Francisco.

Wikidata was developed in the spirit of the Wikipedia's increasing drive to add structure to Wikipedia's articles. Examples of this include the introduction of infoboxes as early as 2002—a quick tabular overview of facts about the topic of the article—and categories in 2004. Over the years, the structured features became increasingly intricate: infoboxes moved to templates; templates started using more sophisticated MediaWiki functions and then later demanded the development of even more powerful MediaWiki features. To maintain the structured data, bots were created—software agents that could read content from Wikipedia or other sources and then perform automatic updates to other parts of Wikipedia. Before the introduction of Wikidata, bots keeping the language links between the different Wikipedias in sync and easily contributed 50 percent and more of all edits in many language editions.

Wikidata allowed for an outlet to many of these activities and relieved the Wikipedias of having to run bots to keep language links in sync or to run massive infobox maintenance tasks. But one lesson I learned from these activities is that I can trust the communities with mastering complex work flows spread out among community members with different capabilities: in fact, a small number of contributors working on intricate template code and developing bots can provide invaluable support to contributors who focus more on maintaining articles and contributors who write the majority of the prose. The community is very heterogeneous, and the different capabilities and backgrounds complement each other to create Wikipedia.

However, Wikidata's structured claims are of a limited expressivity: their subject always must be the topic of the page; every object of a statement must exist as its own item and thus as a page in Wikidata. If it doesn't fit in the rigid data model of Wikidata, it simply cannot be captured in Wikidata—and if it cannot be captured in Wikidata, it cannot be made accessible to the Wikipedias.<sup>4</sup>

For example, let's take a look at the following two sentences from the English Wikipedia article on Ontario, California:

To impress visitors and potential settlers with the abundance of water in Ontario, a fountain was placed at the Southern Pacific railway station. It was turned on when passenger trains were approaching and frugally turned off again after their departure.

There is no feasible way to express the content of these two sentences in Wikidata—the simple claim and qualifier structure that Wikidata supports cannot capture the subtle situation that is described here.

## An Abstract Wikipedia

I suggest that the Wikimedia movement develop an Abstract Wikipedia, a Wikipedia in which the actual textual content is being represented in a language-independent manner. This is an ambitious goal<sup>5</sup>—it requires us to push the current limits of knowledge representation,<sup>6</sup> natural language generation,<sup>7</sup> and collaborative knowledge construction<sup>8</sup> by a significant amount. An Abstract Wikipedia must allow for:

1. relations that connect more than just two participants with heterogeneous roles;
2. composition of items on the fly from values and other items;
3. the expression of knowledge about arbitrary subjects, not just the topic of the page;
4. the ordering of content, to be able to represent a narrative structure; and
5. the expression of redundant information.

Let us explore the last of these requirements: unlike the sentences of a declarative formal knowledge base, human language is usually highly redundant. Formal knowledge bases usually try to avoid redundancy, for good reason. But in a natural language text, redundancy happens frequently. One example is the following sentence:

Marie Curie is the only person who received two Nobel Prizes in two different sciences.

The sentence is redundant given a list of Nobel Prize award winners and their respective disciplines they have been awarded to—a list that basically every large Wikipedia will contain. But the content of the given sentence nevertheless appears in many of the different language articles on Marie Curie, usually in the first paragraph. So there is obviously something very interesting in this sentence, even though the knowledge expressed in this sentence is already fully contained in most of the Wikipedias it appears in. This form of redundancy is commonplace in natural language—but is usually avoided in formal knowledge bases.

The technical details of the Abstract Wikipedia proposal are presented elsewhere.<sup>9</sup> But the technical architecture is only half of the story. Much more important is the question of whether the communities can meet the challenges of this project.

Wikipedia and Wikidata have shown that the communities are capable of meeting difficult challenges—be it templates in Wikipedia, or constraints in Wikidata, the communities have proven that they can drive comprehensive policy and work-flow changes as well as the necessary technological feature development. Not everyone needs to understand the whole stack to make a feature such as templates a crucial part of Wikipedia.

The Abstract Wikipedia is an ambitious future project. I believe that this is the only way for the Wikimedia movement to achieve its goal, short of developing an artificial intelligence that will make the writing of a comprehensive encyclopedia obsolete anyway.

### **A Plea for Knowledge Diversity?**

When presenting the idea of the Abstract Wikipedia, the first question is usually, “Will this not massively reduce the knowledge diversity of Wikipedia?”<sup>10</sup> By unifying the content between the different language editions, does this not force a single point of view on all languages? Is the Abstract Wikipedia taking away the ability of minority language speakers to maintain their own encyclopedias, to have a space where, for example, indigenous speakers can foster and grow their own point of view, without being forced to unify under the Western US-dominated perspective?

I am sympathetic with the intent of this question. The goal of this question is to ensure that a rich diversity in knowledge is retained and that minority groups have spaces in which they can express themselves and keep their knowledge alive. These are, in my opinion, valuable goals.

The assumption that an Abstract Wikipedia, from which any of the individual language Wikipedias can draw content, will necessarily reduce this diversity is false. In fact, I believe that access to more knowledge and to more perspectives is crucial to achieve an effective knowledge diversity and that the currently perceived knowledge diversity in different language projects is ineffective at best and harmful at worst. In the rest of this essay I will argue why this is the case.

### **Language Does Not Align with Culture**

First, it is wrong to use language as the dimension along which to draw the demarcation line among different content if the Wikimedia movement

truly believes that different groups should be able to grow and maintain their own encyclopedias.

In case the Wikimedia movement truly believes that different groups or cultures should have their own Wikipedias, why is there only a single Wikipedia language edition for the English speakers from India, England, Scotland, Australia, the United States, and South Africa? Why is there only one Wikipedia for Brazil and Portugal, leading to much strife? Why aren't there two Wikipedias for US Democrats and Republicans?

The conclusion is that the Wikimedia movement does not believe that language is the right dimension to split knowledge—it is a historical decision, driven by convenience. The core Wikipedia policies, vision, and mission are all geared toward enabling access to the sum of all knowledge to every single reader, no matter what their language, and not toward capturing all knowledge and then subdividing it for consumption based on the languages the reader is comfortable in.

The split along languages leads to the problem that it is much easier for a small language community to go “off the rails”—either to become heavily biased as a whole or to adopt rules and processes which are problematic. The fact that the larger communities have different rules, processes, and outcomes can be beneficial for Wikipedia as a whole since they can experiment with different rules and approaches. But this does not seem to hold true when the communities drop under a certain size and activity level, when there are not enough eyeballs to avoid the development of bad outcomes and traditions. For one example, the article about skirts in the Bavarian Wikipedia features three upskirt pictures—one porn actress, an anime screenshot, and a video showing a drawing of a woman with a skirt getting continuously shorter. The article became like this within a day or two of its creation and, even though it has been edited by a dozen different accounts since then, has remained like this over the last seven years. (This describes the state of the article as of this writing—I hope that with the publication of this chapter, the article will finally be cleaned up).

### **A Look on Some South Slavic Language Wikipedias**

Second, a natural experiment is going on where contributors that are more separated by politics than language differences have separate Wikipedias. There exist individual Wikipedia language editions for Croatian, Serbian,



Bosnian, and Serbo-Croatian. Linguistically, the differences among the dialects of Croatian are often larger than the differences between standard Croatian and standard Serbian. Particularly, the existence of the Serbo-Croatian Wikipedia poses interesting questions about these delineations.

The Croatian Wikipedia has turned to a point of view that has been described as problematic. Certain events and Croat actors during the 1990s independence wars or the 1940s fascist puppet state might be represented more favorably than in most other Wikipedias.<sup>11</sup>

Here are two observations based on my work on south Slavic language Wikipedias.

First, claiming that a more fascist-friendly point of view within a Wikipedia increases the knowledge diversity across all Wikipedias might be technically true but is practically insufficient. Being able to benefit from this diversity requires the reader not only to be comfortable reading several different languages but also to engage deeply enough and spend the time and interest to actually read the article in different languages, which is mostly a profoundly boring exercise since a lot of the content will be overlapping. Finding the juicy differences is anything but easy, especially considering that most readers are reading Wikipedia from mobile devices and are just looking to satisfy a quick information need from a source whose curation they trust.

Most readers will only read a single language version of an article, and thus any diversity that exists across different language editions is practically lost. The sheer existence of this diversity might even be counterproductive as one may argue that the communities should not spend resources on reflecting the true diversity of a topic within each individual language. This would cement the practical uselessness of the knowledge diversity across languages.

Second, many of the same contributors that write the articles with a certain point of view in the Croatian Wikipedia also contribute on the English Wikipedia on the articles about the same topics—but there they suddenly are forced and able to compromise and incorporate a much wider variety of points of view. One might hope the contributors would take the more diverse points of view and migrate them back to their home Wikipedias—but that is often not the case. If contributors harbor a certain point of view (and who doesn't?), it often leads to a situation where they push that point of view as much as they can get away with in each of the projects.

It has to be noted that the most blatant digressions from a neutral point of view in Wikipedias like the Croatian Wikipedia will not be found in the

most central articles but in the large periphery of articles surrounding these central articles, which are much harder to keep an eye on.

### **Abstract Wikipedia and Knowledge Diversity**

The Abstract Wikipedia proposal does not require any of the individual language editions to use it. Each language community can decide for each article whether to fall back on the Abstract Wikipedia or whether to create their own article in their language. And even that decision can be more fine-grained: for an individual article, a contributor can decide to incorporate sections or paragraphs from the Abstract Wikipedia.

This allows the individual Wikipedia communities the luxury to entirely concentrate on the differences that are relevant to them. I distinctly remember the situation when I started the Croatian Wikipedia: it felt like I had the burden to first write an article about every country in the world before I could write the articles I cared about, such as my mother's home village—because how could anyone defend a general purpose encyclopedia that might not even have an article on Nigeria, a country with a population of a hundred million, but an article on Donji Humac, a village with a population of 157? Wouldn't you first need an article on all of the chemical elements that make up the world before you can write about a local food?

The Abstract Wikipedia frees a language edition from this burden and allows each community to entirely focus on the parts they care about most—and to simply import the articles from the common source for the topics that are less within their focus. It allows the community to make these decisions. As the communities grow and shift, they can revisit these decisions at any time and adapt them.

At the same time, the Abstract Wikipedia makes these differences more visible since they become explicit. Right now there is no easy way to say whether the fact that Dianne Feinstein is listed as the mayor of San Francisco in the Cebuano Wikipedia is due to cultural particularities of the Cebuano language communities or not. Are the different population numbers of Frankfurt in the different language editions intentional expressions of knowledge diversity? With an Abstract Wikipedia, the individual communities could explicitly choose which articles to create and maintain on their own, and at the same time remove a lot of unintentional differences.

By making these decisions more explicit, it becomes possible to imagine an effective workflow that observes these intentional differences and that sets up a path to integrate them into the common article in the Abstract Wikipedia. Right now, there are 166 different language versions of the article on the chemical element helium—it is basically impossible for a single person to go through all of them and find the content that is intentionally different between them. With an Abstract Wikipedia, which contains commonly shared knowledge, contributors, researchers, and readers can actually take a look at those articles that intentionally have content replacing or adding to the text of the commonly shared one, assess these differences, and see if contributors should integrate the differences in the shared article.

The differences in content may be reflecting difference in policies, particularly in policies of notability and reliability. Whereas on first glance it might seem that the Abstract Wikipedia might require unified notability and reliability requirements across all Wikipedias, this is not the case: due to the fact that local Wikipedias can overlay and suppress content from the Abstract Wikipedia, they can adjust the content displayed on their local Wikipedia based on their own rules. And the increased visibility of such decisions will lead to more easily identified biases and hopefully also to updated rules to reduce said bias.

### **A New Incentive Infrastructure**

The Abstract Wikipedia will evolve the incentive infrastructure of Wikipedia.

Presently, many underrepresented languages are spoken in areas that are multilingual. Often another language spoken in this area is regarded as a high-prestige language and is thus the language of education and literature, whereas the underrepresented language is a low-prestige language. So even though the low-prestige language might have more speakers, the most likely recruits for the Wikipedia communities—people with education who can afford internet access and have enough spare time—will be able to contribute in either of the two languages.

In which language should I contribute? If I write the article about my mother's home town in Croatian, I make it accessible to a few million people. If I write the article about my mother's home town in English, it becomes accessible to more than a hundred times as many people! The

work might be the same, but the perceived benefit is orders of magnitude higher: the question becomes, do I teach the world about a local tradition, or do I tell my own people about their tradition? The world is bigger and thus more likely to react, creating a positive feedback loop.

This cannibalizes the communities for local languages by diverting them to the English Wikipedia, which is perceived as the global knowledge community (or to other high-prestige languages, such as Russian or French). This is also reflected in a lot of articles in the press and in academic works about Wikipedia, where the English Wikipedia is being understood as *the* Wikipedia. Whereas it is known that Wikipedia exists in many other languages, journalists and researchers are, often unintentionally, regarding the English Wikipedia as the One True Wikipedia.

Another strong impediment to recruiting contributors to smaller Wikipedia communities is rarely explicitly called out. It is pretty clear that, given the current architecture, these Wikipedias are doomed in achieving their mission. As discussed above, more than half of all Wikipedia language editions have fewer than ten active contributors—and writing a comprehensive, up-to-date Wikipedia is not an achievable goal with so few people writing in their free time. The translation tools offered by the Wikimedia Foundation can considerably help within certain circumstances<sup>12</sup>—but for most of the Wikipedia languages, automatic translation models don't even exist and thus cannot help the languages which would need it the most.

With the Abstract Wikipedia, though, the goal of providing a comprehensive and current encyclopedia in almost any language becomes much more tangible. Instead of taking on the task of creating and maintaining the entire content, only the grammatical and lexical knowledge of a given language needs to be created. This is a far smaller task. Further, this grammatical and lexical knowledge is comparably static—it does not change as much as the encyclopedic content of Wikipedia, thus turning a task that is huge and ongoing into one where the content will grow and be maintained without the need of too much maintenance by the individual language communities.

Yes, the Abstract Wikipedia will require more and different capabilities from a community that has yet to be found, and the challenges will be both novel and big. But the communities of the many Wikimedia projects have repeatedly shown that they can meet complex challenges with ingenious combinations of processes and technological advancements.<sup>13</sup> Wikipedia and Wikidata have both demonstrated the ability to draw on technologically

rather simple canvases and create extraordinary rich and complex masterpieces that stand the test of time. The Abstract Wikipedia aims to challenge the communities once again, and the promise this time is nothing else but to finally be able to reap the ultimate goal: to allow every one, no matter what their native language is, to share in the sum of all knowledge.

**Acknowledgments:** Thanks to the valuable suggestions on improving the article to Jamie Taylor, Daniel Russell, Joseph Reagle, Stephen LaPorte, and Jake Orlowitz.

### Notes

1. Patti Bao, Brent J. Hecht, Samuel Carton, Mahmood Quaderi, Michael S. Horn, and Darren Gergle, "Omnipedia: Bridging the Wikipedia Language Gap," in *CHI '12: Proceedings of the Conference on Human Factors in Computing Systems*, ed. Joseph A. Konstan, Ed H. Chi, and Kristina Höök (New York: ACM, 2012), 1075–1084.
2. Denny Vrandečić and Markus Krötzsch, "Wikidata: A Free Collaborative Knowledgebase," *Communications of the ACM* 57, no. 10 (October 2014): 78–85, <https://doi.org/10.1145/2629489>.
3. Lucie-Aimée Kaffee, Hady ElSahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon S. Hare, and Elena Simperl, "Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for Article Placeholders," in *Proceedings of the 15th European Semantic Web Conference (ESWC 2018)*, ed. Aldo Gangemi, Roberto Navigli, Marie-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam (Berlin: Springer, 2018), 319–334; Lucie-Aimée Kaffee, Hady ElSahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon S. Hare, and Elena Simperl, "Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, ed. Marilyn Walker, Heng Ji, and Amanda Stent (Stroudsburg, PA: ACL, 2018), 640–645.
4. Denny Vrandečić, "Restricting the World," *Wikimedia Deutschland (blog)*, February 22, 2013, <https://blog.wikimedia.de/2013/02/22/restricting-the-world/>.
5. Umberto Eco, *The Search for the Perfect Language (The Making of Europe)*, trans. James Fentress (Oxford, UK: Blackwell, 1995).
6. Anna Wierzbicka, *Semantics: Primes and Universals* (Oxford, UK: Oxford University Press, 1996); Thomas Hoffmann and Graeme Trousdale, "Construction Grammar: Introduction," in *The Oxford Handbook of Construction Grammar*, ed. Thomas

Hoffmann and Graeme Trousdale (Oxford, UK: Oxford University Press, 2013), 1–14.

7. Aarne Ranta, *Grammatical Framework: Programming with Multilingual Grammars* (Stanford, CA: CSLI Publications, 2011).

8. Kaarel Kaljurand and Tobias Kuhn, “A Multilingual Semantic Wiki Based on Attempto Controlled English and Grammatical Framework,” in *Proceedings of the 10th European Semantic Web Conference (ESWC 2013)*, ed. Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph (Berlin: Springer, 2013), 427–441; Wikidata, s.v. “Wikidata: Lexicographical data,” accessed June 1, 2019, [https://www.wikidata.org/wiki/Wikidata:Lexicographical\\_data](https://www.wikidata.org/wiki/Wikidata:Lexicographical_data).

9. Denny Vrandečić, “Towards a Multilingual Wikipedia,” in *Proceedings of the 31st International Workshop on Description Logics (DL 2018)*, ed. Magdalena Ortiz and Thomas Schneider (Aachen: Ceur-WS, 2018).

10. Mark Graham, “The Problem with Wikidata,” *The Atlantic*, April 6, 2012, <https://www.theatlantic.com/technology/archive/2012/04/the-problem-with-wikidata/255564/>.

11. Sven Milekić, “Croatian-Language Wikipedia: When the Extreme Right Rewrites History,” *Osservatorio Balcani e Caucaso*, September 27, 2018, <https://www.balcanicaucaso.org/eng/Areas/Croatia/Croatian-language-Wikipedia-when-the-extreme-right-rewrites-history-190081>.

12. Ellery Wulczyn, Robert West, Leila Zia, and Jure Leskovec, “Growing Wikipedia Across Languages via Recommendation,” in *Proceedings of the 25th International World-Wide Web Conference (WWW 2016)*, ed. Jaqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao (New York: ACM, 2016), 975–985.

13. Mathias Schindler and Denny Vrandečić, “Introducing New Features to Wikipedia: Case Studies for Web Science,” *IEEE Intelligent Systems* 26, no. 1 (January-February 2011): 56–61.