Title: Knowledge gap or added value? Self-related content in two lesser resourced languages

Author: dr. Ziko van Dijk, Klexikon/ZUM.
Keywords: Wikipedia, Heinz Kloss, Western Frisian, Luxembourgish, lesser resourced language

Extended abstract:
The Wikipedia language versions differ from each other: in size, but also in the content they provide [3, 8]. When the Wikipedia community of a language version produces content, does it try to represent all topics of the world equally? Or does it focus on topics that are particularly related to the language in question?
Some languages are considered lesser resourced languages. Accordingly, Wikipedia in such a language also tends to have limited resources. Should these resources be used for a local focus? Or is that an undesired self-restraint creating a knowledge gap itself?

„Eigenbezug" according to Kloss
The linguist Heinz Kloss assumes that a small language produces nonfiction texts primarily in certain domains and at certain levels. The three levels here are: 1. the knowledge that everybody learns at school, 2. knowledge for the job or other specialized domains, 3. the university and research level.
The three domains are: 1. 'self-related' [eigenbezogenen] topics from the own sphere of life of the linguistic community", for example its language, literature, local history, its native trades et cetera, 2. the humanities, 3. science and technology.
According to his assumption, literature is produced in the small language on all three levels, but only if the subject belongs to the first domain ("Eigenbezug", "self-related" topics). In the second domain one reaches only levels 1 and 2, and in the third domain only the lowest level [7].
The two languages selected here are West Germanic languages from Western Europe. Both Frisian and Luxembourgish are rather "small" languages with a few hundred thousand speakers each. In many domains of society, they are not dominant, but have to compete with larger languages. In Fryslân (Friesland, NL) this is Dutch, in Luxembourg, the other two official languages are French and German [1, 4, 5].

Wikipedia on „Eigenbezug"
Wikipedia in Frisian says on the page "What Wikipedia is not" that it is not an "encyclopedia of Fryslân". It is supposed to be an encyclopedia about the whole world. The Wikipedia in Luxembourgish has "Eigenbezug" in the rules for notability.
I have asked several contributors of Frisian Wikipedia and Luxembourgish Wikipedia. While none of them are in favor of a mandatory focus, there are differences. Some contributors argue that the language version should prioritize these topics, while others believe that an encyclopedia should cover all areas.

Content in comparison
I have evaluated and compared a number of categories and articles, with or without "Eigenbezug". Here are just some examples:
Wikipedia in Frisian has 11 articles on dialects of Frisian. I looked at the size and quality of these articles and compared them to Dutch and English Wikipedia. The articles of Frisian Wikipedia perform best, Dutch Wikipedia significantly worse and English Wikipedia has only three short articles on 3 of the 11 dialects.
The Frisian municipality Harlingen has eight church buildings. Each of these have an article in Frisian and a corresponding article in Dutch. The content is not exactly the same, but similar in detail and scope.
Wikipedia in Luxembourghish has a category on novels in Luxembourgish. None of the 40 articles has a counterpart in another language.
If an article mentions sources at all, the source is mostly in a dominant language. Sources in the small language are quite rare. This is especially true for specialist or scientific literature.

Readers
As seen, the Wikipedias in Frisian and Luxembourgish have content that can compete to some extent with the content of other Wikipedias. Consulting Wikipedia in the small language may be a rational choice indeed. Several contributors believe that readers do come to the small language version for content with "Eigenbezug". Others are afraid that readers don't know about their language version at all.
There are official Wikipedia statistics regarding the number of page views. However, I have often the impression that they are flawed. For example, in March 2023, Luxembourgish Wikipedia has been accessed

78.000 times from Luxembourg, but also 21.000 times from India. It is hard to believe that there is a significant Luxembourgish speaking group in India.

Articles with "Eigenbezug" do not necessarily generate many page views. One example is the 40 articles about novels in Luxembourgish. Only two articles had an average of at least one page view per day.

However, generating a lot of page views does not have to be the main goal for the own language. It can also be a value in itself that a language version covers a topic completely, for example, all architectural monuments in the region.

Discussion and conclusion

A Wikipedia language version has limited resources. These should be used primarily for content that is most important for the readers. However, this assumes (without proof) that these readers are indeed interested in topics with "Eigenbezug".

One contributor said that most speakers of Luxembourgish speak several languages anyway. They find other topics in other language versions. This may be a suitable strategy in Luxembourg, because Luxembourgish and German are mutually intelligible, but not in other countries.

A "negative" view sees "Eigenbezug" as an undesirable self-restraint, that runs counter to values such as multiculturalism. As one interviewee said, an enforced restriction does not fit to an open encyclopedia.

Wikipedia in English grew because of the wiki cycle: content is found by potential readers via a search machine, some readers become contributors, more contributors create more content, etc. Small languages experience the problem that search machines often ignore them. Supporters of the language versions need to find alternatives [2].

In this research, I used an approach by Kloss. There is clearly a focus on "Eigenbezug" in the two Wikipedias. However, it was difficult to see with the same clarity the three different levels. One has to keep in mind that Wikipedia is supposed to be a general reference work. Accordingly, one would have to expect only level 1 and level 2 anyway.

To examine more nuanced differences between levels, one would perhaps need to examine a larger Wikipedia. For example, Gordejeva et al. criticize the Wikipedias in English, Russian and German for being too difficult to read. To understand articles about deseases, one needs a college degree [6].

These "large" language versions have partially become specialist and scientific encyclopedias. Conversely, one might say that the Wikipedias in Frisian and Luxembourgish perhaps come closer to the ideal of Wikipedia, as an encyclopedia "for everyone".

References:

[1] Guillem Belmar (2019). Attitudes and language use of (potential) new speakers of a minoritized language: The case of adults learning West Frisian in formal courses. Sustainable Multilingualism, 15, pp. 70–88; pp. 71/72.

[2] Ziko van Dijk (2021): Wikis und die Wikipedia verstehen. Eine Einführung. transcript: Bielefeld, pp. 49-51.

[3] Dwaipayan Roy, Sumit Bhatia, Prateek Jain (2021): Information asymmetry in Wikipedia across different languages: A statistical analysis. In: Journal of the Association for Information Science and Technology. Volume 73, issue 3, pp. 347-361; pp. 347-349.

[4] Fernand Fehlen, Andreas Heinz (2016): Die Luxemburger Mehrsprachigkeit. Ergebnisse einer Volkszählung. transcript: Bielefeld, p. 175/176.

[5] Peter Gilles, Sebastian Seela, Heinz Sieburg, Melanie Wagner: Sprachen und Identitäten. In: IPSE – Identités Politiques Sociétés Espaces (ed.): Doing Identity in Luxemburg. Subjektive Aneignungen – institutionelle Zuschreibungen – sozio-kulturelle Milieus. transcript: Bielefeld, 2010, pp. 63-104; p. 63.

[6] Jelizaveta Gordejeva, Richard Zowalla, Monika Pobiruchin, Martin Wiesner (2022): Readability of English, German, and Russian Disease-Related Wikipedia pages: Automated Computational Analysis. In: Journal of Medical Internet Research, 2022. Volume 24, issue 5, p. 19.

[7] Heinz Kloss (1976): Die Entwicklung neuer germanischer Kultursprachen seit 1800. 2nd edition. Pädagogischer Verlag Schwann: Düsseldorf (1967), pp. 43-49.

[8] Miriam Redi, Martin Gerlach, Isaac Johnson, Jonathan Morgan, Leila Zia (2020): A Taxonomy of knowledge gaps. https://arxiv.org/pdf/2008.12314.pdf.