Instead of today’s many-sliced wikis, separated by language and project, we aspire to re-establish a unified community of collaborators in the spirit of awbona. We will still respect language and cultural differences—there will still be English, German, Hebrew, Arabic, etc. —but instead of separate domains, we propose a single user experience with integrated navigation between projects and languages and the possibility of split screen views aligning related content. On a single page we can work on articles in different languages, or simultaneously edit textbook content and encyclopedia articles. Via machine translation we can facilitate conversations and collaborations spanning languages and projects, without forcing a single culture or perspective.

LanguageConverter is oriented to readers: it converts the article text unidirectionally into readable text in a consistent variant. But as soon as a user begins to edit, they are confronted with the source text in a mix of variants, as illustrated by the intermingled Cyrillic and Latin scripts in the article from Serbian Wikipedia shown below. The mixture of scripts can be a huge barrier to editing in communities where individuals are typically only fluent in a single variant.

The Parsoid team has been experimenting with a new bidirectional implementation of LanguageConverter, based on Finite State Transducers (FSTs). These allow automatic annotation of wikitext such that it can be round-tripped to its original variant losslessly. With these annotations, an Wikimedian can edit an article in their preferred consistent variant. Unedited portions of the article will round-trip to their original variant, preventing dirty diffs.

On wikis where the community has chosen to author all articles in a single variant, all text can be losslessly saved as the chosen variant, regardless of which variant the editor used.

We can make editing easier on wikis using LanguageConverter!