Translating Wikis: Problems and Solutions in German-to-English Translations
William Bothe
(User:DraconicDark)

Introduction

Translation is one of the most efficient ways to create new content on Wikipedia. When creating new articles, it is important to not rely too much on automated software. In this regard, the use of straight machine translations and article creation bots has been problematic in many cases.

During Women's History Month, I have used a combination of machine translation and manual polishing to increase the number of articles on German and Austrian female politicians in the English Wikipedia. Machine translation encountered several problems, some of which are:

1. Place names versus nouns
2. Complex administrative-style language
3. Indirect speech (Konjunktiv not resolved)
4. Inversion of sentence structure (object – predicate – subject)
5. Unrecognized compound words

These problems and possible solutions are discussed here.

Problems

1. Place names vs. nouns:

Original German: Im Volksschulalter setzte sie gemeinsam mit ihrer Schwester durch, dass sie in ihrer Pfarrgemeinde Rum als Mädchen in der Kirche ministrieren durften.

Google Translate: In the school age, she and her sister went through the fact that in their parish, they were allowed to ministry rum as a girl in the church.

Published translation: When she was in elementary school, she and her sister got the church to agree that in their parish of Rum, they were allowed to act as servers in the church, even though they were girls.

Rum is a city in Tyrol, Austria. The article is not referring to the alcoholic beverage of the same name. While admittedly, the image of rum being administered as a sacrament is funny, it's not correct in any case.

2. Complex administrative-style language

Original German: Bei der parteiinternen Nominierung zum Bundestagskandidaten der CDU im Bundestagswahlkreis Flensburg – Schleswig verlor Sütterlin-Waack am 1. Oktober 2012 zunächst zwar mit knappen 307 zu 312 Stimmen gegen den amtierenden, wahlkreisfremden CDU-Landesvorsitzenden Jost de Jager, nachdem dieser jedoch eine Woche später als CDU-Landesvorsitzender und CDU-Wahlkreiskandidat wieder zurückgetreten war, wurde sie am 5. Februar 2013 mit 224 von 231 Stimmen nominiert.

Google Translate: In the party-internal nomination to Bundestag candidates of the CDU in the constituency of Flensburg - Schleswig Sütterlin-Waack lost on 1 October 2012 initially with almost 307 to 312 votes against the incumbent CDU state chairman Jost de Jager, after a week later than CDU -Landesvorsitzender and CDU constituency candidate had resigned, she was nominated on February 5, 2013 with 224 out of 231 votes. (This whole thing is a mess!)

Published translation: Sütterlin-Waack was nominated as a candidate for the federal elections on February 5, 2013 after the original candidate, Jost de Jager, had resigned. She was nominated with 224 out of 231 votes.

Needed to trim down the paragraph to make it readable, keeping only the most important details. Not every word needs to be directly translated!

3. Konjunktiv (subjunctive) not resolved

Original German: Entgegen der Legende handelt es sich dabei nicht um das Grabdenkmal Peter von Hagenbachs, das seine Wittwe an der Stätte seiner Hinrichtung habe errichten lassen.

Google Translate: Contrary to tregend, this is not the grave monument of Peter von Hagenbach, which had his widow at the site of his execution have been erected.

Published translation: Contrary to legend, this is not the grave monument of Peter von Hagenbach, which his widow is said to have erected at the site of his execution.

The German Konjunktiv (subjunctive) is a verb form that expresses opinion and hearsay. Often, words such as "is said to" have to be inserted to convey that meaning.

4. Inversion-the enemy of all computer translators!

Original German: Diese zerstörte 1793 ein Brand bei der Beschießung der Stadt.

Google Translate: This destroyed a fire in 1793 during the bombardment of the city.

Published translation: This was destroyed by a fire in 1793 during the bombardment of the city.

In German, the normal subject-predicate-object order can be inverted into object-predicate-subject. To convey the correct meaning, either the word order has to be changed, or passive voice has to be used.

5. Compound word not recognized by the dictionary

Original: Entlang der S1 gibt es auf niederrangigen Straßen flächendeckend Durchfahrverbote für Lastkraftwagen über 3,5 Tonnen um den Transitverkehr aus den Orten zu verbannen und Mautflüchtlingen vorzubeugen.

Google Translate: Along the S1, there are overland traffic bans for trucks over 3.5 tons on low-level roads to ban transit traffic from the places and prevent the fugitives from flying.

Published translation: Along the S1, there are overland traffic bans for trucks over 3.5 tons on low-level roads to ban transit traffic from local areas and prevent toll evasion.

The word "Mautflüchtlingen" is a word Google Translate does not recognize and isn't sure what to do with. The original translation "prevent the fugitives from flying" makes no sense. Later, Google Translate rendered that same section of text as "prevent toll refugees" which translates the components (*Maut*, meaning toll, and *flüchtlingen*, meaning refugees) correctly, but not the whole word.

Conclusions

The existing content translation tool helps a lot with translating articles from German to English.. However, in order to improve the user experience and make translating articles even easier, I suggest these improvements:

1. A system that recognizes and flags unusual grammatical features (for example, the aforementioned Konjunktiv (subjunctive) from German. Such a feature would warn the user that the phrase contains such a feature and suggest a better translation.

2. Support for online dictionaries (such as Word Reference or Linguee) in addition to machine translation.. Machine translation only shows one possible definition of a word, and not all of them. The way dictionary support would work is that you would highlight a word in the original article and it would then bring up a window with the dictionary's definitions.

3. If it isn't already a capability of the content translation tool, a method for restricting access to machine translation to more experienced editors is needed. Perhaps the usage of machine translation could be restricted to a subset of users who have been granted the right to use it by an admin after they have published a certain number of reviewed translations (20 or 30 maybe, the actual number would be up to the community of each Wikipedia to decide) . I say this because in English Wikipedia, when the content translation tool was first created, inexperienced editors started flooding English Wikipedia with unedited machine translations. This resulted in the content translation tool's built in machine translation being disabled for the English Wikipedia, which provides an added inconvenience for translators like myself who start with a machine translation and work from there.

References

Excerpts are from the following articles I've translated:
1. Ingrid Felipe
2. Sabine Sütterlin-Waack
3. Breisacher Stephansmünster
4. Wiener Außenring Schnellstraße

Logos and images are from Wikimedia Commons.

Acknowledgements