

Expert interview

Case: Wikipedia
Interviewee (IE): Amir Aharoni
Interviewer (IR): Anja Ebersbach
Date: 2020-10-23, 11:00am
Location: Skype

	[Welcoming]
1	IR: The last thing we talked about in July was about rewarding people on translation, do you remember? #00:01:21#
2	IE: Yeah, I remember, vaguely. I'm pretty sure that in some Wikipedias people give something, a barnstar, wiki love, awards, you know, simple things, not real money and prizes. #00:01:28#
3	IR: Do you know somebody who studies this field? #00:02:14#
4	IE: Who studies this field? I'm not sure, I don't think I can give you any name. I wish somebody did a study of a lot of Wikipedias at once because I am pretty sure, there are several dozens of Wikipedias that give awards like that, but it's difficult to find a list of that, because it's not saved in a uniform structured way. You would have to go Wikipedia after Wikipedia and check if they have such a thing. #00:02:39#
5	IR: Maybe more in the field of science, somebody who wants to know it and makes kind of a documentation about that. I'm thinking about someone like Dariusz, but he's more in the general topic. #00:03:28#
6	IE: Yeah, probably, maybe, I'm not sure. #00:03:51#
7	IR: Let's skip to the next question, last question about in the human dimension. It's about the conflicts that might appear in this multilingual area, can you think about, or do you know about conflicts which are specially about languages, translation or anything like that? #00:03:58#
8	IE: I have to mention this, in lot of places, repeatedly, because lots of people misunderstand this. So, lots of people, everywhere, not everybody but a lot of people, it even came up in formal research, lots of people confuse translation with machine translation. Like people hear translation and they immediately think that it's machine translation, which is wrong. #00:04:46#
9	IR: Really? #00:05:31#
10	IE: Lots of people. Everywhere, from all countries and languages and wiki projects, everywhere. Not everybody, some people understand this. To me it's totally obvious that translation is what people do and machine translation is what machines do when people ask them, and translation is what humans do, actually writing the translation and lots of people misunderstand this, even like smart, experienced, otherwise knowledgeable people, they hear translation and immediately a red light goes on in their head and they think: "Oh, translation is a

bad thing, because it makes mistakes, and Google translation is bad.” But I was not talking about Google translation, I was talking about human translating things. It’s just incredible how many people confuse this. #00:05:32#

11 IR: It’s funny, I’ve never experienced that. #00:06:32#

12 IE: It’s extremely common, and informally I experience this a lot when talking about translation. Like when I do such things, I do this more informally and intuitively and about a year ago, our team in the Foundation hired a design researcher who, I think, has a PhD in linguistics or something, and he does research with very nice organization, like research preparations and very detailed written reports. He found the same thing in his formal research. He interviewed lots of people, maybe he published this already, if he didn’t publish this already, he will soon. He found the same thing, that lots of people from lots of languages think that the translation is the same as machine translation, and people don’t understand this, this is the number one confusion everywhere, in English, in Indonesian, in Russian, in Bengali, in all the languages. It’s very easy to blame Google for this who, for years and years showed people this button in Google Chrome, “do you want to translate this page?”, and it’s a horrible button. Not because the machine translation makes mistakes but because of the way it’s written, it’s not like YOU translate this page, it’s a machine, it’s Google’s algorithm make an automatic translation, which may contain mistakes. But now lots of people identify the word „translate“ with this, so it’s pretty horrible, I don’t know how we can reclaim the word „translate“ and bring it back to its original real meaning. It’s a real design challenge, where we may end up changing some things in the design of our translation tools, maybe use some other word instead of the word „translation“, maybe use some buttons differently, I don’t know. By now we have to probably address this by using different terminology, different branding, different buttons, I don’t know, it’s a big design question. So, when talking to people, or for example, content translation is enabled in all the Wikipedias since 2015 and you can translate from any language to any language, in my terminology right, you can enable content, you can get into content translation, start translating an article from Ukrainian to Vietnamese. But then, some people say “content translation does not work in Zulu”. It does work in Zulu. If there is a Wikipedia in this language, it works in that language. But if machine translation doesn’t appear, they think that content translation doesn’t work, but no, content translation works everywhere, it just means that you will have to type the translation by yourself. But that’s what we mean by the word „translate“. Even if there is machine translation, you still have to check every single word and correct all the mistakes that machine translation makes. It’s still your responsibility, whether machine translation works or not, but people say content translation doesn’t work. So that’s the biggest misunderstanding. Some Wikipedia administrators - again in several languages - see that an article was created by content translation. And content translation, when it creates an article, it puts in the edit summary: created by translating the page such and such and the title. And just from seeing this edit summary, they assume that it’s a machine translation. They don’t even read the article, they think that this is a machine translation and then I have to explain that maybe if you read the article and you see typical machine translation mistakes, then yes, then it’s a machine translation, then you should either correct the mistakes or delete the article, but some people

think that it's always a machine translation. This is the number one most common misunderstanding. #00:06:35#

13 IR: And they don't like it, that's the problem, isn't it? #00:11:48#

14 IE: They are right not to like machine translation when it's actually a machine translation. When it's actually a bad machine translation that makes mistakes, it's good that they don't like it, because nobody should do this. You don't need any tools, in theory you can run a bot that translates all the articles in the English Wikipedia to Vietnamese by running them through Google translate, but that would be completely pointless and harmful. You wouldn't use machine translation for that, you will need Content translation for that. Content translation is for making the corrections. If you don't make the corrections then you are not using that correctly. But that's the most common misunderstanding. Other than that, if you remove this huge issue, there are no big conflicts actually, it's much less likely that an article that was translated will be deleted or that anybody will complain that this article doesn't have notability or something like that. #00:11:51#

15 IR: And they don't talk about for example the wrong term, the wrong translation for a sentence or anything like that? #00:13:00#

16 IE: Yeah, this can happen, of course, but it can happen with anything. It's just not such a big deal. That can happen. There are so many kinds of mistakes that people could make writing Wikipedia articles. #00:13:09#

17 IR: And do you know, how they solve this problem? Do they say, we base this upon a certain dictionary or who decides which is the right translation then? #00:13:33#

18 IE: This doesn't necessarily happen with translation, you could write an article from scratch, not by translating it from a Wikipedia article in another language, and still, you would have to use some terms and I don't know every time it's something else. Some people will just use the term that they know, some people will translate the foreign term, without working too hard to adapt it to the local language. It happens in some languages that people make up a neologism sometimes this even sticks in the language. It's very different in every language. Some languages are famous for being more open to neologisms, like Croatian is a famous example, Hebrew is semi-open to neologisms, Lithuanian, Icelandic, these are like famous examples of languages with a lot of neologisms. Some languages don't mind adapting foreign terms. Russian adapts a lot of foreign terms without a problem, Hindi adapts a lot of foreign terms. Although there are some people who disagree with this and some people want more puristic sanskritized classic Hindi. I heard that there are some arguments about this in the Hindi Wikipedia, whether to use more modern and more English foreign terms or to translate terms into something more official. I don't know Hindi very well, but I heard that this happens in Hindi and several more languages of India. But it's not necessarily a matter of translation because questions of terminology can come up even when you are just writing, not necessarily when you are just translating. In English, amusingly enough, this does not come up because English is a source for so much

	terminology. So, in English this does not come up but in lot of other languages it does. #00:13:52#
19	IR: But in English it must be the way that they have to find an English word for the foreign word? #00:16:35#
20	IE: I can kind of imagine this happening, but I think if something like this happens in English Wikipedia, a more likely argument will not be about which word to use but whether to write about this at all, whether this is notable for the English Wikipedia. This is the kind of thing that I hear about much more in the context of the English Wikipedia. Like some social, cultural, religious phenomena that happens in India or Kenya or whatever, is this phenomenon described in reliable sources, is this phenomenon notable and should we write about this at all. To be notable in the English Wikipedia, this thing has to be described in external, reliable sources and if these sources are in English, then you use the term that appears in these sources. So, there is not supposed to be an argument about that. If it's described under several different names, if there are several reliable sources but different reliable sources use different names, then you may have an argument about which name to use as the title of article maybe, because there can only be one title, but it's not that common in English. #00:17:04#
21	IR: Okay, then let's turn to technology, this is your expertise, isn't it? #00:19:06#
22	IE: Maybe. #00:19:34#
23	IR: Okay, so if you look at the management of all Wikipedias concerning language specific technical features, if you look at the architecture, what does every Wikipedia have in common? Concerning language management and where is it localized and where is the base, so to say? #00:19:35#
24	IE: What exactly do you mean by language management? #00:20:07#
25	IR: For example, if you create a new Wikipedia, then what is the core that every Wikipedia gets? As you said, every Wikipedia has content translation, hasn't it? #00:20:12#
26	IE: Yeah, content translation is a MediaWiki extension and in the configuration of MediaWiki as it is used on Wikimedia site, the content translation extension is enabled by default. #00:20:41#
27	IR: And is this the only one? #00:21:04#
28	IE: No, there are several dozens of other extensions that are enabled in all the languages. #00:21:08#
29	IR: That's what I mean. Only the most important things. Concerning the language. #00:21:14#
30	IE: Okay, so let's see. So, let's start with MediaWiki itself. So, MediaWiki, core MediaWiki has something like 4.000 translatable messages, and they are translated in the same way on Translatewiki and the messages for extensions are also translated on Translatewiki. Some extensions are specific for language, so Content Translation is one and Universal Language Selector. Universal Language Selector has several functions, it's for selecting the fonts, for selecting the user

interface language, for selecting typing support in several languages. A lot of languages don't need any special typing support, but some languages do, but it's installed in all languages in any case. Visual editor, it's installed everywhere. Visual editor actually has some language features, which are not very famous, but they are language features. It has the special characters bar, which is kind of thing for language support, I don't use it very much but some people think it's very useful and important, so whatever. #00:21:24#

31 IR: What does it do? #00:23:38#

32 IE: At the top of the Visual editor, there is a button in the tool bar, which looks like the letter Omega, the Greek letter, and if you click it, then it shows you a very long list of special characters. Even very simple things, if you usually type in English and you want to type a German name with an umlaut, it doesn't appear on the English keyboard, so English Wikipedians often use thing to insert the special characters. For you it's not special, because you write in German all the time, but for the English Wikipedians it's important. #00:23:42#

33 IR: This is a very important point, that's good. #00:24:33#

34 IE: Yeah, when Visual editor was first deployed, it didn't have this special characters button, it did exist in the Wiki syntax editor for a long time, and then when Visual editor was deployed, people missed it and people asked for this. So, this was one of the features that the Visual editor team added, by community request. So, it's important for some people, so I guess you could call it a language feature. Then there are some things that are special, and that are only needed in some languages. So, there is the language converter, which is not used for English, even though in theory it could be used for English, like for some differences between British and Indian and American and Canadian English, but currently it's not done like that, but the language converter is most famously used for Serbian, to convert automatically between Cyrillic and Latin Serbian and also for Chinese, also for several other languages, I think for Uzbek and Kurdish and a few others. But most famously it's for Serbian and Chinese, so they automatically convert and simplify the traditional Chinese and Latin Serbian and Cyrillic Serbian and then Latin Uzbek and Cyrillic Uzbek and a few other languages too. So, it's currently irrelevant for English, there are occasionally proposals to make such a thing also for American and British English, and I think it would be a pretty good thing, but the English Wikipedia community would probably discuss it to death, like seven years and then it will decide to scrap the project because they are arguing about everything and not actually doing anything. That's the problem with English Wikipedia. #00:24:36#

35 IR: Okay, are there other important extensions you can think of? #00:26:36#

36 IE: Yes. So, I started from the messages, in addition to the messages, there are also some other things that are translatable. So, there is a list of what's called magic words. For example, when you insert an image, then you do, it's so funny to say this out loud, so you do square bracket, square bracket, file, colon, file name.jpg, pipe, and then you can say thumb, right? So, thumb is the English magic word, the English parameter name, but this is translatable, so to actually translate this, you don't do this is Translatewiki, you have to do this in a PHP file

in Core Media Wiki, but this is translatable, and that's translated to a lot of languages. Some languages just write thumb, as in English, but for some languages you can use something else, thumb works everywhere, like the English, by default works everywhere but you can use the translated name. There are several dozens of these things. #00:26:48#

37 IR: But then if you install a new Wikipedia, it's a lot of work, isn't it? You have to wait for the Translatewiki things and then you have to look into all those small corners, where something like the magic words are still there, and you have to translate it in a very exotic language maybe? #00:28:12#

38 IE: Yeah, so translating the magic words is actually relatively small thing because there are just a few dozens of these and people usually do it very quickly, they just get a list and translate it in no more than a day, usually takes just an hour. When you create a new Wikipedia, there is a list of about 600 messages in TranslateWiki that you have to do. For some languages it takes many months, for some languages it takes just a few days, it just depends on how motivated and enthusiastic and productive the volunteer translator is. I saw languages that completed these 600 messages in less than a week and I saw languages for which it takes years. It just depends on the volunteer translators, how fast they are. I've been looking at this for years all the time, in lot of languages and I do what I can to make this easier, like for example I sometimes find, in this list of 600 messages, I sometimes find the things that are not very useful or that are unnecessarily difficult to translate, so I either simplify them or I just remove them from this initial list of most important messages. After hearing the same complains about translation difficulties from several languages, I wrote some documentation that answers the common questions, like how to translate the magic words, how to translate some common Wiki syntax and these things. There still is an assumption that if you are translating in Translatewiki, that you are already experienced in writing in the Wikipedia and by now, this assumption is obviously wrong. It was probably right back in 2007, but by now it's definitely wrong because a lot of people want to start a new Wikipedia in their language and they don't have any experience writing in the English Wikipedia or in the Russian Wikipedia, in 2020 this assumption is definitely wrong. So, I sometimes need to explain some basic wiki syntax, because you need some basic wiki syntax like links and stuff like that, but it's even the most simple things like the links and pipes and stuff like that. So, I wrote some super basic documentation for translators, maybe like two years ago, so whenever a new translator joins, I send them this documentation, people usually find it helpful. But yeah, that is not that difficult, whenever I find some objectively difficult things, I try to fix them, but there are also some things, they just depend on how motivated you are about getting your language translated. I really try to get rid of assumptions, lots of people don't, but I really do, about what technical knowledge do people have to have. I try not to assume that people know HTML, I try not to assume that people know wiki syntax. I do have to assume that people know their language and are able to type in their language and that they have some available time to volunteer to do this, that's not something that I can solve. But whatever technical hurdles I can remove, I try to remove. #00:28:34#

- 39 IR: Who does these translations for the new Wikipedia in the Translatewiki, is it the community? How does this project, the new Wikipedia as technical project, how does it work? #00:34:02#
- 40 IE: So, the order is more or less like this: the volunteer who wants to create a new Wikipedia makes a request in Meta: I want to create a Wikipedia in the Madurese language - this is one of the languages of Indonesia, recently there was a new request for that, for example. And then, I add support for this language in Translatewiki, I mean usually I do this, but there are several administrators in Translatewiki, there is Raimond and a few others. I often do this, but it doesn't have to be me. And then the volunteers have to start writing articles in the Incubator and to make translations in Translatewiki and once the translation is completed and the activity in the Incubator is reasonable, then they ping the language community, of which I also happen to be a member, and they say, can you please check the Incubator, if the Incubator is fine, can you please approve it and convert it to main. If the language community finds that the activity in the Incubator is fine, then the language community tries to find the external expert who knows this language and the external expert checks that the content in the Incubator is good. And if the content is good then somebody from the language community creates a task in Phabricator that says: create a new Wikipedia in such and such language. And then some people who have access to the Wikimedia servers do the configuration that actually creates a new wiki and once the actual wiki is created, there are some people who have a permission to export the articles from the Incubator to the new domain. And that's when it is complete, that's the process. #00:34:38#
- 41 IR: Okay, and is this a lot of work or does this depend on the language? #00:36:46#
- 42 IE: It's more or less the same work for every language. There is usually nothing special, the process is the same. I wouldn't say that it's simple. It's more complicated than it should be, and I don't know, did I send you this talk that I gave at the Celtic knot conference a few months ago? #00:36:55#
- 43 IR: No. You told me about the Celtic know conference but not about a special... #00:37:21#
- 44 IE: Okay, so, my talk is recorded, it's 20 minutes and I talk about these things there, so I can send you. #00:37:33#
- 45 IR: Yes, if you could send it to me, it would be great. #00:37:40#
- 46 IE: In the talk I describe the current process and what are the problems with it and I have some suggestions for improving it. #00:37:46#
- 47 IR: So far, we have talked about the past. Are there changes planned, concerning this process, do you have new projects to improve anything? #00:37:59#
- 48 IE: Okay, I have a proposal. I haven't really started any formal process to change this, but I really should start it, but my idea is basically to move the creation of wiki from the end of the process to the beginning of the process. To skip the whole incubator thing, so that you create new articles not in the incubator but you would create a wiki and then write the articles there. #00:38:14#

49 IR: That's great. #00:38:57#

50 IE: I think so, there should be something that encourages people to actually write the articles and actually translate. I am thinking of ways to make it somehow limited, somehow downgraded. For example not to use the name Wikipedia, like to use some other name as long as you don't complete the translation of the basic messages in Translatewiki, you cannot call this website a Wikipedia, you can use all the features, but you cannot use the Wikipedia brand or maybe you cannot have administrator privileges or something like that. There must be something that encourages people, it's like a silly gamification but from my experience of several years there are something like people just abandoned this and then it's just dead weight and it's not great. #00:38:59#

[Goodbye]