

Expert interview

Case: Wikipedia
Interviewee (IE): Amir Aharoni
Interviewer (IR): Anja Ebersbach
Date: 2020-03-31, 12:00pm
Location: Skype

[Welcoming]

- 1 IR: We stopped in the content section, and I think we talked about the forms of translation, and I wanted to ask you, if you know about bots which translate. Are there bots? #00:02:07#
- 2 IE: Not that I know and I don't think that it should be done, and I'm pretty sure it's not done anywhere. No, I don't think so. It's impossible to translate text. You can generate text from databases. There are several wikis in which there are bots that generated articles from databases about towns in Italy or towns in France, stuff like that, asteroids is another example, but that is it. You can generate text, you cannot translate with a bot without human intervention and expect any kind of quality. It's a very bad idea to do that. I don't think that any wiki allows such a thing. #00:02:28#
- 3 IR: I think there are already discussions about automatic translation as well, where people just... they don't know the target language, they just put some sections into DeepL or something like that and put it into the other Wikipedia. #00:03:39#
- 4 IE: Discussions where? #00:04:03#
- 5 IR: I think it's on the mailing list, about people who use automatic translation tools very often. #00:04:06#
- 6 IE: I haven't seen such a thing and it's highly unlikely that such a thing will ever happen. What is the point in doing this? You cannot publish these things as an article, because if you publish it as an article, it's static text and it's bad, and it lost its connection to the original. If you ever want to do a machine translation, you can do machine translation. Google translate isn't going anywhere, as far as I know. In an instant you can do it, there is no point in publishing. #00:04:22#
- 7 IR: That is what I mean, maybe it's the wrong wording. People use those tools just to translate and they don't have the competence to evaluate the result, so... don't you think that this very often happens? #00:05:08#
- 8 IE: That people use Google translate to read an article, this definitely happens a lot. #00:05:33#
- 9 IR: That is what I mean. #00:05:40#
- 10 IE: We cannot know how often, but it definitely happens a lot. There was actually some research about this, I don't know if it was published, the design research team in the Wikimedia Foundation conducted research about this

recently, I don't know if they published it, I should ask them... but they ran a survey in some languages, I don't remember which ones, Arabic was one of them, and probably a few others. Like how do people perceive machine translation of Wikipedia articles. I don't remember, it's a pretty long document. I have to see if they published it, I don't think it's internal, I think they published it somewhere, but I don't know where exactly, I can try to find it. #00:05:46#

11 IR: That would be great. What about your Content Translation tool? It supports machine translation as well. #00:06:40#

12 IE: Yes, for some languages, for most languages that are supported by Google translate or Yandex translate, and a few other engines. We enabled this. We don't have German and English as target languages. #00:06:40#

13 IR: Really? #00:07:24#

14 IE: Yeah, it's very silly, but convincing the editing communities in English and German Wikipedia is a challenge. We'll do it eventually. I know for sure that there are people who want this, but there are also people who are very loud, and complain about everything, and don't want anything to change, so convincing them is a bit of a challenge. Anyway, we will definitely do this someday, I don't know when, maybe next month, maybe next year we will do it, probably not next month. Sometime during the next year or so, we will enable this for German and English, but yes, we have this in Content Translation. When you click on a paragraph, when you are translating to a language which is supported, it automatically fills the paragraph with machine translation and the translator, the human, is expected to fix this, to fix all the mistakes the machine translation makes, and machine translation always makes mistakes. For languages for which this is not enabled, this includes English and German, and a lot of other languages, for a lot of small languages there is no machine translation. No one ever built machine translation for these languages. For these languages, you can just type manually. You can translate from other languages into English, but you have to type everything by yourself. #00:07:25#

15 IR: I don't understand why it is not possible to translate into English and German, isn't it kind of a generic thing to put this machine translation tool into your tool and then it translates from every language to every language? #00:09:00#

16 IE: Of course, it's possible. Technically all we have to do to enable this is to change one line in the configuration file and enable it for English and another line for German. However, if we do this, then what very often happens in big Wikipedia editing communities, definitely English, definitely German, occasionally Russian, when people notice there is some change in the software, they start complaining very loudly, we didn't ask for this, you didn't notify us, this is a very bad change, you must revert immediately, otherwise, we blackout the whole site, stuff like that. Communicating with these people is challenging, so if we enable this, if we make this change - technically it's very easy - we can just do this, but if we did this, these people would complain and scream. Before we do this, we have to be sure they will not scream at us. It requires a bit of community management work first. But we will do it eventually. #00:09:26#

17 IR: That is interesting, so the English and the Germans are the most complicated ones? #00:10:43#

18 IE: In terms of community management yes, I'm not a community manager myself, we used to call this community liaison, now we call it community relations specialist. So we have several people in the Wikimedia Foundation, whose job it is to talk to people in the various editing communities, and listen to them and pass messages between the Foundation and the communities and make sure that everybody is happy. We will eventually do this for English and for German, we had some challenges doing it for Italian, we eventually did it for Italian, same for Dutch, took us some time to get this enabled in Dutch, but now it's enabled. As you may notice, the bigger, better developed languages, they are also more suspicious of any software changes, because they are very big, they don't like moving parts, but eventually it will happen for German and English. It's very often like that, lots of new features get enabled first on smaller languages and then later on the bigger languages. It's possible to use Content Translation there, it's just... you have to type everything by yourself and you cannot use machine translation. The funniest thing is that somebody made a gadget that enables machine translation using Google, and you can use that gadget, and it was not made by the Foundation, it was made by a volunteer, and that will work, and nobody is complaining about that, but if the Foundation did that.... #00:10:52#

19 IR: I know, it's complicated. Ok, what do you think, in which direction do most translations flow? #00:12:47#

20 IE: I don't have to think, I can check my statistics. I can just send you the statistics page. The top source language for translation, from which most people are translating, is Ukrainian, of course (that was a joke, it's not Ukrainian, it's English). The second I think it's Spanish, but let me check. For Content Translation, you can have very precise numbers. But I'm pretty sure that the numbers are very similar for people who don't use Content Translation and simply translate manually. So English is the top source language for translation with more than 400.000 translated articles from English, the second is Russian and the third is Spanish, and then you have French, German, Catalan, Ukrainian, Italian, Chinese, Portuguese and so on. I can just send you... #00:12:47#

21 IR: So generally speaking, it's from the big Wikipedias to the small ones? #00:14:32#

22 IE: Yeah, and now I can give you top target languages. They are: Spanish, French, Catalan, Vietnamese, Ukrainian - I was not totally joking about Ukrainian, they are very active in translation or all kinds of things. So, Spanish, French, Catalan, Vietnamese, Ukrainian, Arabic, Portuguese, Russian, Chinese, Italian, Persian, Hebrew, English as a target language, Tamil, Greek, Galician and so on. #00:14:39#

23 IR: It's not from the big ones to the small ones, but mainly from the big to the big, isn't it? #00:15:10#

24 IE: Well... yeah and it's not so surprising, because the bigger are active in translation. There are some relatively small communities which happen to be

very active in translation. Bashkir, if you heard about that language, it is spoken in Russia, there is a very active Wikipedia editing community in that language. They use Content Translation a lot, they translated more than 10.000 articles, that is a very good number for such a small language. There are something like two million people speaking that language, and they translated more than 10.000 articles, most of them from Russian, and it is above Indonesian, and Japanese and Turkish and Romanian, which are much bigger languages. #00:15:19#

25 IR: But I could also understand a logic where smaller communities transfer their knowledge to the bigger ones. Imagine there is special knowledge about geographic things. #00:16:16#

26 IE: This is a very good thing to do, and some people do it, but not as much as I would love it to be. #00:16:44#

27 IR: Maybe it's also a question of relevance? #00:16:57#

28 IE: Not necessarily. I can check the numbers for deletion. Articles that were created using Content Translation are much less likely to be deleted than articles created from scratch. Out of articles created from scratch in the English Wikipedia every day, something like 40% or 50% of these are deleted. #00:17:01#

29 IR: That is a lot. #00:17:56#

30 IE: That is a lot because lots of people want to create articles on the English Wikipedia, and lots of them don't know that they cannot create such an article or they intentionally try to vandalize, or occasionally it also happens that the English Wikipedia administrators delete articles that should not have been deleted, but that definitely happens, unfortunately. Out of articles created using Content Translation the number of deleted articles is smaller than 20%. An article created using Content Translation is much less likely to be deleted than an article created from scratch. In other languages even lower than 20%, much lower, usually it's lower than 10%. #00:17:58#

31 IR: What is your opinion, why is it this way? Because the ones that delete articles have a kind of quality criterion that an article already exists in one of the Wikipedias? #00:19:01#

32 IE: By itself, it's usually not a criterion. That if an article exists in another Wikipedia, then it must be notable, by itself it's not a criterion. Nevertheless, it probably means if it exists in another Wikipedia, then quite likely it's notable at least for the Wikipedia in that language, so this already raises the chances that it may be notable for another language. That sounds like the most sensible thing, also Content Translation usually creates articles with some formatting, it's usually not perfect formatting, very often you have to fix some things after you publish the first version. You usually have to fix some templates, you usually have to fix something, but already the first version of an article looks like an article, you know, like a wiki article, with links with images and so on. If you just post a page with nothing but text, no images, no links and no infoboxes and no footnotes, this will immediately light some warning signs for a lot of administrators, that there could be something wrong about this article. Which is

sad, because lots of people could write something good, but they just don't know how to use the formatting with all these things, like the links and the footnotes and so on, but the text is otherwise good. But lots of administrators say, oh no, this is not good for Wikipedia, this must be deleted or moved to drafts or something like that. And Content Translation immediately automatically creates formatted articles, which makes this published page look like an article.
#00:19:23#

33 IR: What do you think, or maybe you know... you said 20% per day are deleted.
#00:21:18#

34 IE: That is an estimation, but this is more or less the number. #00:21:29#

35 IR: And do you know why, what are the reasons to delete a translated article?
#00:21:33#

36 IE: Let me check what is the number now for English. So either the administrator who deleted it, decided that it is not notable for the English Wikipedia, even though it's notable for another language, that's one possibility. Another possibility is that somebody wrote a very bad translation, like they were typing the translation, but they don't know English well enough, and what they wrote is not good English, and they deleted it, that's another possibility, and I think that is about it. What other reasons there could be for deletion? Maybe if the Content Translation software really screwed up with the automatic formatting and published wiki syntax which is so bad that it's easier to write from scratch than to fix the wiki text, this is imaginable, probably not very common, but conceivable.
#00:21:40#

37 IR: Ok. #00:22:57#

38 IE: But most likely it's either notability or somebody wrote very bad English, this is possible. #00:23:00#

39 IR: Ok. #00:23:08#

40 IE: And now, by the way, that I look at the statistics, the number of deleted articles in English is really, really low recently. It used to be, as I said, just as an example: there was a week in the summer of 2019, where there were 80 articles published in English, and 24 out of them were deleted. It's very different from week to week, in another week there were 256 articles published in English, that is a lot, and 12 articles were deleted on the same week, so it's very different every week. There are better weeks, there are worse weeks, but in any case the percent of deletion in English is pretty low, much lower than is it for articles that were created manually from scratch. That is the main outtake, and for languages other than English it's usually even lower. #00:23:10#

41 IR: But you talk about deletion in general now? #00:24:27#

42 IE: That is our easiest way to measure the quality, so if the article is deleted by an administrator, then something, somewhere had to be wrong with the quality, or the administrator is very wrong, but that is less likely. #00:24:33#

43 IR: But you don't talk about the translated articles only, but in general?
#00:24:49#

- 44 IE: Well, in general, yeah, but since I'm in the team that developed Content Translation, we care about articles that were created using Content Translation. #00:24:56#
- 45 IR: Ok, so, with the low number you meant the translated articles, not in general. #00:25:04#
- 46 IE: Not in general, no. This is the crazy part, it's a quality metric that we adapted for our team, for our product. I used to be product manager for Content Translation, and this was one of the vital sign metrics, how many of the articles that are published using Content Translation were later deleted, and we have checked this every month for five years. #00:25:11#
- 47 IR: That is really great #00:25:49#
- 48 IE: And the crazy part is that the VisualEditor team, for example, they don't measure this, they don't care how many articles were deleted, that were created using VisualEditor, but we do. I think there should be some project manager who checks for such things, but I don't think anybody in Wikimedia Foundation checks this. I think there are some people on the wikis, maybe some administrators, who maybe check for this, I'm not sure, it's just a guess, maybe nobody, I don't know. I think that somebody should take a look at this, at least occasionally. #00:25:50#
- 49 IR: And if you talk about quality management, how are translations evaluated? Do the people look only at the text and say: the text is ok, and the facts are there, and citations are there, or do they look at the translation as well? Do they know where it comes from and they look at the quality of the translation as well? #00:26:32#
- 50 IE: I can only guess, I can give you a very naive answer and it's so naive that you should consider it wrong, ok? I guess that they don't compare this to the source article, usually. They probably just judge it by how well it fits to the usual style and quality measurement of the Wikipedia in their language. This is my general guess. Maybe, if they have any suspicions about something being incorrect or incomplete, or something like that, they possibly check the article in the source language, if they know the source language - not everybody knows the source language - and maybe they check the references, or they check whether something was translated correctly, something like that. Why do I strongly suspect this answer is not entirely correct? Because this is a very naive answer. I do have administrator privileges, administrator rights in English and in Hebrew, I don't use them very often to delete articles. Occasionally I do, but very rarely. There are lots of other Wikipedia administrators, in other languages, who are much busier than I am in quality control and checking that new articles are really good. So you should ask them, you should not ask me. And different people in different languages will probably tell you different things, from what I just told you. Even different administrators in the same language will quite possibly tell you different things. Another very strange thing, it's really strange to me, it means that we kind of partly failed and we need to fix Content Translation in this regard, I heard from some administrators in some languages, not a lot, but I heard from some people, and even though it's not a lot of people I still take this

seriously and want to improve the situation. What they told us is that they think articles created using Content Translation are something else, they are not like usual articles, so they don't even check them, and that is really wrong, because the whole point of Content Translation is that it's supposed to help editors create the first version of an article, and once the first version is created, it's just a usual article, like any other and should be judged according to the same quality metrics, the same content policies, style guides, whatever they have. I have heard from certain people in certain languages, if they see it was created using Content Translation they don't check it at all. Either because they think it's very good, like everything that is created using Content Translation is by definition very good, which it isn't, or contrariwise, because they think that it's bad. For me this would be all the more reason to check this and if you think that it's bad delete it or something, but apparently not all the administrators understand this, which is a bit of a failure on my part, because every time I speak about Content Translation, I say that once the first version is created, it's an article like all others. There is nothing special about it, it just helps you create the first revision. And, that is according to all the same style guides, all the same policies, all the same formatting rules, whatever. But apparently some people don't totally understand this, so this could be a curious thing to research, but to do something proper, you would have to ask lots of administrators in lots of different languages. It's challenging but very interesting. #00:27:17#

51 IR: Yes, but maybe the low number of deleted articles is just the consequence of this behavior, because if they don't want to touch it... in German we say "durchwinken", they just let it pass. #00:31:42#

52 IE: If you could sent me that word, I would love to [inaudible] #00:32:07#

53 IR: Ok. #00:32:11#

54 IE: What what is, last time, Macht... #00:32:12#

55 IR: Machtgefälle. #00:32:16#

56 IE: I looked that up in a dictionary, so if you could also send me this word I would be happy to check it, since it's very interesting. Yes, it's a possibility, and it's not a very good thing, it's a bit of as programmers say, it's a bit of a smell, right? #00:32:17#

57 IR: Yes. #00:32:36#

58 IE: This could be a problem, maybe they don't delete it because they think they don't even have to look at it. I don't think that it's pervasive, but it probably happens in some languages and with some administrators. Even though I don't think it's pervasive, I do think that we should address it somehow, someday. I don't know how, but we should, because I want this to be perfectly integrated with Wikipedia. I know that for lots of languages, it's very well integrated, lots of people use Content Translation and administrators know how it works, and they know how to find translated articles, and check them for quality, and in some languages they don't understand it so well, maybe we could solve it with documentation, maybe we could solve it by publishing some kind of a video, that

explains it, I don't know, but we will have to address it someday, somehow.
#00:32:27#

59 IR: And how can you see that an article is created with Content Translation? Is there a tag? #00:33:36#

60 IE: Yes, there is a tag in the first revision of the article. #00:33:44#

61 IR: Ah, ok. And it links to the source article? #00:33:49#

62 IE: Yes, the link to the source article is automatically added to the edit summary.
#00:33:55#

63 IR: Ok. #00:34:02#

64 IE: It's generally a useful information, and it's also a legal requirement, license requires it. Funny thing, by the way, in a small number of languages, definitely in German, maybe also in Romanian, when an article is translated, somebody imports all the old revisions of the source article into Wikipedia. In German, if an article was translated using Content Translation, all the versions of the source article, like in English, or whatever language from which you translate it, will be imported to the German Wikipedia, because some people in the German Wikipedia believe that this a legal requirement. As far as I know, it's not a legal requirement, but for some reason, people do this, and I think it's a total waste of time and electricity and storage space, but for some reason people do it, definitely in German, maybe also in Romanian and maybe in some other languages. Totally pointless, but people do it anyway, I don't know why. I read the terms of use, and I read the license and it's not required anywhere. License specifically says that you have to give a hyperlink in edit summary, and it was checked by lawyers, like real lawyers, I am not a lawyer, but real lawyers told me a link in the edit summary is enough, but nevertheless, people in the German Wikipedia think that it's not enough. I have no idea why. And if you talk about this to anybody, they will say: of course, it's illegal without it and the license requires it. Ask them: which section in the license requires it? Ask them to point you to specific sentence, I don't think they will be able to show you this section and if they do, please tell me, because I tried it several times, and they were never able to show me which part of the license requires it. #00:34:03#

65 IR: Ok. Strange. An urban legend or something like that. #00:36:19#

66 IE: That is what I think also. #00:36:28#

67 IR: I understand it for the Germans, they love something like that, but the Romanian, I don't understand. #00:36:31#

68 IE: I think I saw it also in Romanian, maybe it's not always done in Romanians, maybe only sometimes, I don't know, but I'm pretty sure that I saw it at least a few times in Romanian, maybe some other languages, too. But in German, it's big and consistent. #00:36:39#

69 IR: Ok. Now, next question...

70 IE: This creates a really silly situation. Just for software testing, I created several translations in the German Wikipedia, and I didn't even publish them to the

article space, but the previous versions were imported anyway, and even more than that I see that I have a lot of contributions in the German Wikipedia, because people translate articles to which I contributed something and my revisions were imported into German, so technically I have lots of contributions to the German Wikipedia, even though I didn't write them in German, I don't know German very much, it's my contributions from English that were imported, so it's a silly weird thing. #00:37:00#

71 IR: Ok. How are media translated, or are media translated? #00:37:53#

72 IE: Mostly manually, if at all. I know that in some languages there are talk pages like community discussion pages, where volunteers who can use graphics software, get files in different languages and translate them. Usually, there is a translator, who knows the languages, and the translator writes the text and then somebody who can use graphics software just deletes all the text in the source language and replaces it with the translation. It's all very manual, and that is just how it is. I know that, for example, I'm very bad at graphics software, I don't do this ever, but somebody, for example asked me once to translate, this was really fun thing to do, to translate a chart, something like a chart of how the metallurgical plant does work. Like how iron works, how do they make real iron from iron ore, and the original chart was in Ukrainian, which makes a lot of sense, because there are lots of iron plants in Ukraine. So somebody asked me to translate it from Ukrainian to Hebrew, and I know a bit of Ukrainian, and I know Hebrew, so I translated from Ukrainian to Hebrew, and now there is such a map of this whole iron manufacturing plant, it's translated to Hebrew, so I translated the strings and somebody else made the actual image. #00:38:03#

73 IR: Ok. #00:39:54#

74 IE: So, it's a very manual process, and if the Wikipedia in a new language is lucky enough to have somebody who knows how to use graphics software and can volunteer some of their time, then your language is lucky, and if you don't have such a volunteer, then your language is not lucky. I know that there is a certain tool, somewhere, which can automatically extract strings from SVG files, and translate them in a kind of a structured way and output a SVG file. Because in SVG you can store the strings as actual strings, rather than graphics, you just need to replace tags in SVG. I know that it exists, I don't know how often it is actually used. #00:39:55#

75 IR: What about the images at Commons? The name of the files, they are not translated? #00:40:51#

76 IE: The name of the file is not translated, the names of the categories are not translated, they are all in English, which is very sad. There is this Structured Data, which is a relatively new thing. The Structured Data on Commons which uses wiki base to store descriptions. Even before Structured Data you had templates on files, in which you could translate the descriptions to different languages. Templates are kind of structured, they are not perfectly structured. They are easy to use for people who know wikitext very well, they are not as easy to use for software that wants to parse this information, so Structured Data on Commons is supposed to eventually, some day in the far future, replace these

templates with something that is properly structured, similar to Wikidata, with really tightly structured translations, but it's just the beginning of that process. The software more or less exists, and it's deployed and people occasionally use it, but it's just the beginning of that, we will have to wait for a few years to see how well it will work. The fact that categories are not translated, that really sucks, categories are all in English, which is really sad. File names, they can be in any language, but there is only the file name. Eventually I hope that file names will not be very important, I hope that someday Structured Data will be more important for finding files than the file name. File name should eventually be just the technical way to insert the file into the page, but for things like searching and sorting, categorizing, stuff like that, that is what Structured Data is supposed to be for. #00:41:00#

- 77 IR: But if you talk about the categories, is it possible to have several category systems within one MediaWiki? #00:43:03#
- 78 IE: I'm pretty sure it's impossible. Because categories are simply stored by name, they are not like Wikidata items, where you have a generic number which is translated to different languages. A category is a name, a classic MediaWiki category is a name, and on Commons this name is in English, which is really unfortunate. #00:43:15#
- 79 IR: How could you solve this? To have categories in different languages? For example, if you have category "animal", and you cannot use just another category and say: this is "Tier" in German and put the image to the two categories? #00:43:45#
- 80 IE: It's actually possible, you can create a category named "Tier", but the problem is... so, ok, why don't you also create a category called "животное" in Russian, and a hundred other languages. #00:44:14#
- 81 IR: Yeah, ok #00:44:27#
- 82 IE: That would not be so great. A better way to do this, would be to change core MediaWiki, so that a category would be something like a generic number, which would have labels, similarly to Wikidata, but this is really not likely to ever happen, it's much more likely that someday, but not like next year. If it happens, it will take at least five years, if not more, until somebody totally gets rid of categories on Commons, and it will probably take even more than five years. But probably Structured Data will address this problem much sooner than the category system will change. The category system is pretty outdated. It's fairly okay for single language Wikis, like the German Wikipedia, so we have categories in German, that is good enough. But if you want a multilingual wiki, like Commons, that is not good enough, because you can have categories only in one language, but the wiki is multilingual, that is not great, really really not great. It's also not very good for collaboration between wikis, because each wiki has its own category tree, and it would make a lot more sense if the category trees were somehow synchronized, between all the languages, and currently they are not. That is what we have, so we have both Wikidata, which does some kind of ontology and categorization, and we also have categories, and it's like two

parallel systems. It's not the worst problem, not the most important problem to solve, but it is a problem. #00:44:28#

83 IR: And when you talk about structured language, you talk about SemanticMediaWiki, or is it another tool? #00:46:31#

84 IE: No, on Wikimedia, wikis don't use Semantic MediaWiki, we do have WikiBase, so WikiBase is the software that runs on wikidata.org, and the same software is also used on Commons now. The name of the project is „Structured Data on Commons“, it's a bit of a clumsy name, but it is fairly clear. It's not a very good brand name, but... #00:46:39#

85 IR: And it's already started, this project? #00:47:08#

86 IE: Yeah, yeah, it's already deployed #00:47:10#

87 IR: Ah, I didn't know that. #00:47:13#

88 IE: It's already deployed, if you go to any image page on Commons, let me just find something quickly... under the file itself, you will see that there are tabs that say: file information and Structured data, and both of these tabs are for a part of this project, and after this, after these two tabs, the old style Commons information, which is just done with MediaWiki templates. But the thing between the file and the templates, that is Structured Data. In that Structured Data you can have something like “depicts”, this is one of the most common things used there, like: this image depicts a person, this image depicts Angela Merkel, or whatever. And then the value of „depicts“ is a Wikidata item. #00:47:14#

S IR: Great. #00:48:43#

90 IE: If you have a photograph with five people, you can say: depicts, and then you can add Wikidata items for each of these people. #00:48:44#

91 IR: So, there is hope for Wiki Commons? #00:48:55#

92 IE: Yes, yes, that is a pretty nice development. So Wikibase is the infrastructure for that, and upon Wikibase there is another extension which was developed by Wikimedia Foundation primarily. There is a team in Wikimedia Foundation called Structured Data, and that is what they do. They already deployed this part, and they also have plans to do more things in this area, so this is just the beginning as far as I know. So yeah, that is pretty good, so now even though the file name is just in English, you can have this Structured Data which shows what this photo depicts, like people, cities, whatever, and Wikidata items, they can be translated into any language, so that is pretty good, the labels can be translated into any language. #00:49:00#

93 IR: Yeah, so finally all those great images will be found at the platform, which is the biggest problem so far. #00:49:57#

94 IE: Yes, a surprise thing that happened two or three years ago, some millionaire heard about this problem and donated a few millions of dollars to Wikimedia Foundation to solve it, so Wikimedia Foundation started a team, and that is it. Now it works and it's still going on, but it's already started. #00:50:05#

95	IR: A millionaire... #00:50:32#
96	IE: It's a true story. I'm not making anything up. #00:50:34#
97	IR: That's nice. Ok, let's go to the organization. You sent me a very helpful link about the incubator, and I know about the procedure of a new Wikipedia how it is now, but I have not found material about the procedure before 2006, when there was a cut I recognized, somehow... no? #00:50:38#
98	IE: Uh, I don't know, but you are asking a very interesting question, and I will have to ask other people. I don't know. I know that it was more or less anarchy, the Wikimedia Foundation had very few employees. The site configuration, like the server management was done by a very small number of Wikimedia Foundation employees and some volunteers, and there were no processes, almost no rules or procedures around that, and it was just programmers and people who know how to run things on the servers, and they would just do things. When somebody had an idea, let's create a Wikipedia in a new language, they would just ask the server administrators, and they would just do it, which caused several problems. One of the problems is that several wikis have an invalid language code. By invalid I mean non-standard language code. Sometime in 2006 or so, people decided that from now on only valid ISO 639 codes will be used. So wikis that were created before that use non-standard codes, which is very inconvenient, but that is how it is. Also, there were some hoaxes, some people created wikis in languages that don't actually exist, and they were subsequently deleted. So to prevent such hoaxes, the language committee was created, which is a group of volunteers who approve new languages. And also the incubator, where people write the first pages, to prove that they are serious. That is more or less what I know, because that was before my time. If you want to know more I can find some names of some people who were around back then, maybe they will be able to tell you more than that. #00:51:11#
99	IR: So if you wanted to create a Wikipedia, a ticket was made and then there it was. #00:53:50#
100	IE: Pretty much. #00:54:05#
101	IR: Ok. #00:54:08#
102	IE: There were advantages and disadvantages in that. The biggest disadvantage is that there is no good way to prove that you are serious. Back then nobody even checked that your language is real. Most of them are ok, most of them are real languages, but nobody checked that you are in any way serious, and that you actually plan to write something and not just get it created and then abandon it, which happened to quite a lot of languages, several dozens of them are completely abandoned and nobody writes there. This is really sad. I hope to fix it someday, but currently I don't have any good tools or resources to do that. The advantage was that you got your own real site, if you have your own complete wiki, then you have all the extensions and you can have your own templates, and you have your proper search and you are connected to Wikidata and you can use Content Translation. With incubator you don't get any of that. There is a very inconvenient way to create templates on incubator, it's possible, but very inconvenient. There is no way to use Wikidata or Content Translation, which

really sucks, even though it would be really useful on incubator to use Content Translation to create your first few articles using translation. #00:54:09#

103 IR: Ok. We are through the time already, again. #00:56:05#

104 IE: Ok. #00:56:08#

[Goodbye]