



朝向「維基百科」語言列表之路

「ウィキペディア」言語リストへの道のり
On the Road to the List of Wikipedias

文 | 李育瑄 (德國杜賓根大學媒體研究所碩士生)

圖 | 政治大學原住民族研究中心



掃描上面的QR CODE，便可觀看「台灣原住民族語維基百科建置計畫」的成果影片。提供讀者更清楚、明瞭的族語維基百科建置歷程。

2015年起，由教育部委託政治大學原住民族研究中心（以下簡稱政大原民中心）執行的「台灣原住民族語維基百科建置計畫」目前已進行至第3年。希望藉由網路世代下社群的力量，以及維基百科人人可以參與編寫的特質，做到族語及文化知識的傳承，甚至將台灣原住民族的語言推向國際舞台。在設定各族「召集人」號召「寫作團隊」，並與本族青年所擔任的「維基助理」互相配合，舉辦工作坊、定期聚會的經營模式之下，目前16族在維基百科孵育場（Wikimedia Incubator）內的條目數量，截至今年9月底統計為止，已達8,697條條目。與去年3月份所統計的1,028條條目相比，可說是大幅度地成長。泰雅語在2016年7月、撒奇萊雅語

It has been 3 years since the Center for Aboriginal Studies (ALCD), National Chengchi University launched the Project of Taiwan Indigenous Language Wikipedia in 2015, which is entrusted by the Ministry of Education. With the strength of online communities as well as the characteristics of Wikipedia that anyone can edit almost every page, it is to be hoped that indigenous languages and their cultural knowledge could be passed down to later generations and even be seen on an international scale. By the strategy that each convener of 16 indigenous ethnic groups gathered a writing team, held workshops as well as regular meetings and cooperated with Wiki-assistants served by young indigenous individuals, the overall amount of the articles in Wikimedia Incubator has reached 8,697 by



原住民族語維基百科建置計畫工作坊會場合照。

在2017年1月，以及阿美語在2017年8月也已分別進入孵育場活躍名單之內。然而到底要到什麼時候，族人才能在網路上看到屬於自己的族語維基百科版本？要如何才能將它正式推行上線？。

活躍？不活躍？That is the question.

首先必須特別說明的是，要在維基百科上新增一個新的語言，並不像一般的網站一樣，直接向官方申請一個新的語言版本即可；而是必須在「維基孵育場（Wikimedia Incubator）」，也就是測試版的維基百科內先進行試寫與持續經營。當編寫、修改人數達到一定的活躍程度時，即有可能被列入所

the end of September, 2017. It could be said that there was a spectacular growth compared with 1,028 articles by March of last year. The project of Atayal Wikipedia in July 2016, Sakizaya Wikipedia in January 2017 and Amis Wikipedia in August 2017 have also become “active” in Wikimedia Incubator. However,

when will these languages get their own sites? How can the teams do more to push their incubating Wikis to be approved?

Active or not active? That is the question.

First of all, it is important to note that a new language edition of Wikipedia could not be established by simply submitting a request to Wikimedia Foundation. Instead, there should be some contributors who keep working in “Wikimedia Incubator” where the new language editions of Wikipedia can be arranged, written, tested and proven worthy of being hosted by the Wikimedia Foundation. When its number of users and their frequency of editing come to a certain level, a new edition is possible to be listed as “active.” Being





透過各族的「寫作團隊」與「維基助理」互相配合，一同完成族語維基百科的建置計畫。

謂的「活躍名單」。在這之後，才有機會成為維基百科全球語言版本之一。然而，進入孵育場活躍名單並不是一朝一夕即可達成，也不是免死金牌能夠永遠維持下去。由於活躍名單同時間只有8個名額，因此若無法保持每個月至少3人、每人至少10筆以上的編輯群和次數，就算進入活躍名單，也有可能隨時被其他語言擠下。

維基百科介面族語化—「單詞表」就靠你了！

除了孵育場的經營之外，不只是條目內容，在維基百科網站上所有可以看得見的功

“active” means that there is an opportunity to become one of the language editions of Wikipedia. Obviously, it is not easy to satisfy the requirements of being “active” in the test Wiki, and one might not always remain “active” as well. Because there are only 8 places for active projects at the same time, the current active projects

could be replaced by other potential ones at any time if they are unable to keep being a regularly active community. “Regularly active” means that there should be at least 3 editors who make at least 10 edits each for every month.

Localization of Wikipedia interface - check with Vocabulary List before translating!

In addition to the test wiki activity, the user interface, including every function, of Wikipedia must be translated into these languages in their contexts. The so-called “Wikipedia Interface Localization” is one of the requirements for a final approval as well.

Atayal Wikipedia project was once out of the “Active List” by reason of the exact lack of user

能、版面說明與使用按鍵等，也必須是全族語的環境。此項「維基百科介面族語化」的工作，是族語版維基百科正式上線的必要條件之一。以首先進入活躍名單、定期舉辦寫作聚會，亦擁有不少條目數量的泰雅語來說，便曾一度以介面翻譯不足為理由，掉出活躍名單之外。據估算，在維基翻譯網上（translatewiki.net，為進行介面翻譯工作的網站）一共有67,448項待完成的翻譯項目，而正式上線所要求的最低門檻約為8,250項。其中短則包含單一字詞，例如「登入」、「搜尋」或「偏好設定」等；長則達數句以上的艱澀程式語言術語。由於翻譯網中涉及大量網路新興詞彙，對此，政大原民中心採用「單詞表」的解決方式，也就是事先將翻譯網中常見的專業詞彙整理成一份對照表，優先進行翻譯。之後利用電腦操作，將已翻譯成族語的單詞「代入」必須整句翻譯的句子中，再由族語譯者確認、梳理句子的文法。此項作法是為了確保8,250個工作項目裡，使用的單字達到前後一致的穩定效果，避免網站上「一義多字」的狀況發生。此項介面在地化的工作，可說是本計畫中最大的挑戰。目前除了孵育場的經營，16族皆以介面翻譯作為主要的工作目標，並已舉辦兩場專門的工作坊，邀請各族譯者針對翻譯內容互相討論、交流。其中泰雅語截至今年9月底，已翻譯完成4,107項；撒奇萊雅語5,570項，皆已接近正式上線所須的最低門檻。然而，新時代的網路用語不一定能夠在族語中

interface translation, although it has a great number of articles and its community meets together regularly to flesh out the project. It is estimated that there are 67,448 items waiting to be done on Translatewiki (translatewiki.net) where people can provide their contribution to interface translation. The requirement for a final approval is that at least 8,250 items shall be translated. Some of them are short such as “login,” “search” or “preferences settings,” while the others are quite longer and might be composed of several difficult terms. As it involves a wide range of new terms in the fields of computer and internet which doesn't exist in the indigenous languages, ALCD uses the “Vocabulary List” as a solution. That is to say, the common terms on Translatewiki would be put in order on a Vocabulary table and translated into different indigenous languages first. With the help of using computer software, the original terms in the phrases would be replaced by the translated indigenous terms which are included in the Vocabulary List. Then, the translators could translate the rest of them and confirm the correctness of the phrases as a whole. This is to ensure that all translated terms, especially the newly created indigenous terms, would be the same and to avoid creating lots of terms with only one definition.

The localization requirement is the biggest challenge to the project of Wikipedia. Besides keeping contributing in the test Wiki, all 16 indigenous language teams are focusing on the work





就撒奇萊雅語言代碼之問題與在地族人討論，經過確認，應是 SIL 及民族語言網刊載錯誤。

找到相對應的詞彙，因此單一字詞最後翻譯出來的結果，往往需要許多字數才足以說明其中的含意。若欲精簡用語、採用「音譯」的方式，又頗失族語真正的靈魂，可說是一項既耗時，又必須兼容傳統與新世代的大工程。

語言代碼「正名」—最後殺出來的魔王

在維基百科孵育場上申請測試版的專案計畫，需要擁有一組語言代碼，用以辨識和歸納條目所屬單位。在維基百科上，其語言委員會（Wikimedia Foundation Language Committee）依據的是由「美國國際語言暑期學院（SIL International）」於2007年2月5日出版的「ISO 639-3國際語種代號」，以及「民族語言網（Ethnologue）」上公佈的資訊標準。每個語言代碼由三個英文字母所組成，例如漢語為zho，泰雅語為tay，阿美語為ami。然而，就在撒奇萊雅語於今年下半

of interface translation at present. Two workshops have been organized and many translators were there to exchange their experiences and discuss with each other. By the end of September, 2017, it is really exciting that Atayal and Sakizaya projects have translated over 4,107 and 5,570 items, both approaching the minimum number of the requirement. But before the approval, there are many obstacles to overcome. Sometimes it is hard to find corresponding terms which exist in indigenous languages for the new generations' way of speaking. Therefore, the translation of a single term might be very long. In other words, it needs many terms to express the meaning. Moreover, adopting loanwords is one of the solutions to streamline the terms, but it seems to lose the soul of the language in some way. So, it could be said, it is a time-consuming task as well as a great process of harmonizing the traditions with new generations.

Language Code Name Correction - the final boss!

To launch a project in Wikimedia Incubator, it needs to have a language code to identify and index the pages of its own. Wikimedia Foundation language committee (LangCom) follows the standard of "ISO 639-3," which was published by SIL International on February 5, 2007, and "Ethnologue," an annual reference publication in print and online that provides statistics and other information regarding the living languages in the world. In the scale of ISO 639-3, each code consists of three

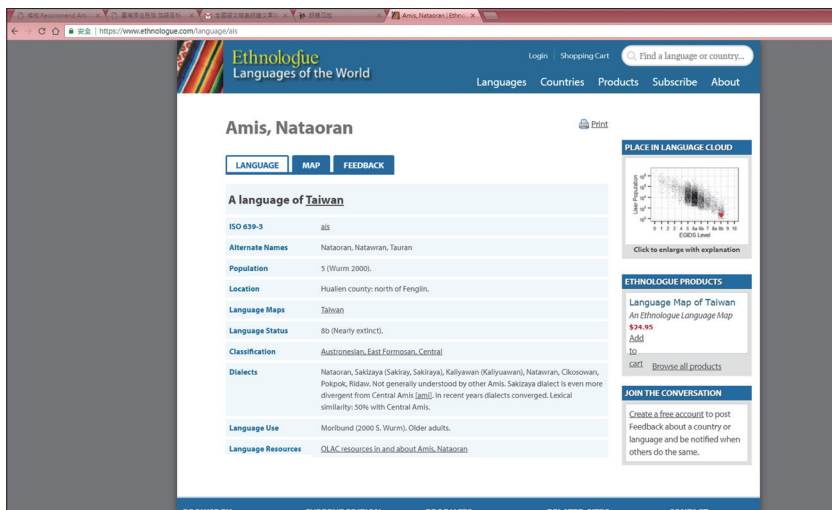
These have been approved and/or created:

- Bashkir Wikibooks
- Hindi Wikivoyage
- Fiji Hindi Wiktionary

These are active and might get their own site soon:

- Amis Wikipedia
- Atayal Wikipedia
- Darja (Algerian Arabic) Wikipedia
- Gorontalo Wikipedia
- Ingush Wikipedia
- Lingua Franca Nova Wikipedia
- Sakizaya Wikipedia
- Shan Wikipedia

孵育場活躍名單只有8個名額，若不能保持每月上線與編輯的活躍度，可能隨時被其他語言擠下。



目前使用語言代碼ais的撒奇萊雅語，被委員會成員認為是那荳蘭阿美語底下的一種語言變體，不足以單獨成為一個維基語言版本。

年度，當每日編輯群人數來到5至7位的活躍狀況，並與維基百科語言委員會討論正式上線的進程時，發現撒奇萊雅語目前使用的語言代碼ais存在一個棘手的問題。由於此語言代碼於SIL和民族語言網上所顯示的正式語言名稱為「那荳蘭阿美語」，被委員會成員認為「撒奇萊雅語為那荳蘭阿美語底下的一種語言變體，不足以單獨成為一個維基語言版本。」然而經過確認，應是SIL及民族語言網刊載錯誤。依照原民會之族語認證考試標準，阿美語的5個語言別為海岸阿美語、

letters for identifying languages, for example, zho represents Chinese, tay represents Atayal, and ami represents Amis language.

It is in the second half of this year, when the number of users and the activity in Sakizaya's project come to a certain level, a tricky problem was then found. Because the language code "ais," which is currently used to represent Sakizaya edition, is shown as "Nataoran language" for its official name on the website of Ethnologue, and Sakizaya is characterized

as a highly divergent dialect of Nataoran, the LangCom considers that Sakizaya should be the only ethnic group using "ais" in the Wikipedia project. However, these are apparently the publishing errors. According to the information for language tests provided by the Council of Indigenous People (CIP) in Taiwan, the five dialects of Amis are Pasawalian Pangcah (Central Amis), Siwkolan 'Amis, Farangaw Amis, Palidaw 'Amis and 'Amisay a Pangcah (Nanshi Amis). In other words, Nataoran is not a dialect of Amis language. Besides, Kaliyawan, Cikosowan, Pokpok, Ridaw are even listed





泰雅族的「寫作團隊召集人」與「維基助理」一起討論撰寫條目的問題。

秀姑巒阿美語、馬蘭阿美語、恆春阿美語與南勢阿美語，其中並不包含那荳蘭阿美語。加上SIL及民族語言網上所公佈的「那荳蘭阿美語」底下，甚至還有嘉里阿美語、七腳川阿美語、簿簿阿美語，以及里漏阿美語。很明顯的，除了撒奇萊雅語之外，其他應僅為阿美族部落名稱，非語言名。並且，在這些部落裡所使用的語言別實為南勢阿美語。從撒奇萊雅族族群歷史的角度來推測，SIL國際組織所公佈的資訊應是當初日本學者分類錯誤所致。撒奇萊雅族族人於1878年「加禮宛事件」中遭大舉屠殺，倖存者擔心被追殺殆盡，從此隱姓埋名躲藏於阿美族部落之中，日治時代日本政府便將其列為阿美族的一支。只是未想到，從1990年起，經過17年的正名運動，終於在2007年1月17日由中華民國政府官方承認為台灣第13個原住民族的撒奇萊雅族，在維基百科孵育計畫中，又再一

as the dialects of Nataoran on Ethnologue. Clearly, except for Sakizaya, the others are only the names of Amis settlements. In fact, the Amis in those settlements all speak 'Amisay a Pangcah (Nanshi Amis).

From the historical perspective of Sakizaya, the improper information published by SIL and Ethnologue seems to be the result by the misclassification of Japanese scholars in the period that Taiwan was under Japanese rule. During "Takobowan Incident" (also known as "Galeewan Incident" or "Kalyawan Battle") in 1876, many Sakizaya people were killed, while the survivors hid into Nataoran Amis and were ruled by the latter. For this reason, Sakizaya people and their language were wrongly categorized as one variety of Amis by Japanese government. It was not expected that Sakizaya, who had been officially recognized on January 17, 2007 as the 13th indigenous people of Taiwan after 17 years of Name Correction Movement since 1990, had to see their historical traces in the Wikipedia project again.

After discussing with the LangCom to uncover the story, the members of the committee suggested that Sakizaya project team should ask SIL for a new language code or reassigning "ais" to Sakizaya. No matter which way they go, it requires much time to put themselves into the annual review process, in which there should be many interactions and they have to provide references and papers as the evidential information. If Sakizaya project succeeds in the "Name Correction" in the 2018 series requests on SIL as well as keeps working in the

次看見了族群的歷史痕跡。經過向語言委員會說明撒奇萊雅族的族群歷史後，委員會成員建議，在撒奇萊雅族維基百科正式上線前，應事先向SIL申請一個新的語言代碼，或向SIL提出變更要求，讓代碼ais重新導向，賦予撒奇萊雅族專屬使用。而不管哪一種方式，都需經過SIL國際組織每一申請年度的審查流程，並提出相關文獻資料作為佐證。倘若撒奇萊雅族成功在2018年向SIL「正名」，並持續經營維基孵育場、完成介面翻譯等要素，最快將有可能於2019年初正式上線。

族語維基百科—語言和文化保存的曙光

那麼泰雅語、阿美語以及其他族語的孵育狀況呢？根據維基語言委員會的建議，除了增加條目數量之外，首先應將現有的條目內容進行擴充。確實，若此計畫目標是為了保存族語、進而推廣族語學習與其智識文化，朝向內容豐富化是一條必要走的道路。回顧這3年，每一個族群在維基百科孵育中所遇到的困難「不同世代對於電腦網路操作的熟悉度、族語書寫能力的差異，以及稀少的志願者定期經營」，似乎都體現了整個族群目前的問題與現況。也因此，期盼透過維基百科共同編修的特質，能夠吸引更多人參與，開啟世代的對話，做到傳承的效果。也期待台灣第一個原住民族語維基百科能在不遠的未來正式上線。◆

test Wiki and complies with the localization requirement, the participants may see their own site soon in the early 2019.

Wikipedia of indigenous language edition- to keep the language and culture alive

How about the situation of other projects such as Atayal and Amis? The LangCom has pointed out that, instead of focusing on creating new articles, the contributors should begin to turn their attention to enrich the existing articles. Indeed, if the purpose of an indigenous Wikipedia is to keep the language alive and promote language learning as well as the unique culture, it is necessary to gather abundant elements for the contents in the projects.

The encountered challenges for the 3 years, which are the knowledges of digital network for different generations, the differences in writing skills in indigenous languages, and the difficulty to find volunteers to take part in the projects, seem to tell the current situations of each group outside the Wiki's world as well. Consequently, it is to be hoped that Wikipedia projects with the open editing characteristics could attract more individuals to participate in the communication of generations, so that their knowledge would be passed down. Furthermore, it is exciting to witness the first Wikipedia edition of Formosan indigenous languages to be online soon.◆

