

Importance of Licensing and Distribution in Language Technology Development Planning

Boymirzaeva Zarina Usmanovna

Namangan Engineering Construction Institute

Abstract: This article discusses the practical importance of licensing and distribution in popularizing language learning technologies.

Keywords: language, language learning technologies, licensing, European methods.

The release of data itself under open source licences has been the subject of some debate in Wales as elsewhere. Large datasets are one of the core requirements to train any models in speech technology, MT and conversational AI agent applications. Finding enough appropriately licenced data is one of the biggest challenges for less-resourced communities. Any strategy for efficient and effective harvesting of data can make the difference between supporting a language or not in a software package. For example, attempts have been made with apps specifically developed to crowdsource a speech corpus from the Welsh language community (Cooper, Jones and Prys, 2019). More recent activity in crowdsourcing Welsh language speech data has focused on collaborating with and sharing efforts with Mozilla's CommonVoice initiative, since its philosophy and motives align (Prys and Jones (2018 (1)). Not all language communities are happy to lose control of their data, especially where they have had bad experiences of colonial exploitation in the past.

However, in the Welsh context, the use of crowdsourcing strategies and of permissive licensing of data has helped the development of Welsh language software by the private sector, aided in some cases by knowledge transfer partnerships between academia and industry. Even in Wales it has not always been possible to release data on open source licence, as some legacy products came with their own, previous licences. In other cases, there was the need to sell commercial software in order to fund the continuation of the work. Increasingly however, and wherever public funding was used to create the tools and resources, they were released on permissive licences such as BSD, MIT, Apache or CC-0, which permit reuse without any restrictions. This was in order to make the tools and resources attractive to enable both small and large companies to take up and use in their own products. In both cases, the private sector is less willing to take up tools and resources published under copyleft licences, such as GPL and CC-BY-SA, which stipulate that the entire utilising body of software must be released openly under the same licence.

Although Welsh, with its approximately half a million speakers is deemed to be a very small market for commercial companies, it is still large enough to support many small companies who could benefit from language technology tools and resources. These include translation companies, local media, software companies, web designers, and producers of educational games and language teaching materials. In common with the experience of many other minoritized and endangered languages in peripheral regions, there are high proportions of Welsh speakers in the rural and remote north and west of Wales, areas that are impoverished with few opportunities for well-paid employment and therefore suffer from emigration of young, talented people. Providing appropriately licensed language resources to small, local companies in these areas can therefore help make them viable and help economic as well as linguistic revitalization of these areas. The

arguments for releasing resources on permissive licences for large multinational companies are somewhat different. It can be argued that multinationals can well afford the development costs of including smaller languages amongst their multilingual offerings, and that paying for the necessary linguistic resources would be a great help to those languages.

However, in the absence of strong legislation requiring Welsh language provision, most multinationals only heed the economic argument, and if the cost of producing or procuring those resources is larger than the anticipated return on their outlay, they will not pay for their development. If, on the other hand, appropriate resources are available to them at no cost, they are then more willing to consider supporting that language amongst their offerings. The minoritized or endangered language community benefits as many of their users already use those products every day in English, Spanish, French or whatever other dominant language they speak. Additional clarity would however be welcomed in understanding the legal ramifications of different licence as there are many legal grey areas. For example, if new language or acoustic models are trained from a specific corpus, does the licence of the original corpus carry over to the new models? Or when a new MT engine is trained on a certain dataset, how does that affect the licencing of the new product? This is especially problematic when there are different licences for the two languages in a bilingual corpus, especially derived from a translation memory where the copyright of the original language text was not originally made explicit. If tools and resources are to be shared outside individual projects and institutions, then dissemination is another issue that comes to the fore. International repositories such as Metashare, Github and Docker Hub have made it easier for developers to find resources in different languages, but for the non-expert user, and anyone interested in a specific language, the plethora of different repositories can be confusing. In addition to using the international repositories therefore, a Welsh National Language Technology Portal was established as a ‘one stop shop’ or ‘brochure site’ pointing at the different resources and giving additional guidance and information on their use (Prys and Jones, 2018).

Welsh is in a fortunate position compared to many other inoritized and endangered languages. The groundwork has been laid for further development of language technologies, and the various stakeholders: academic researchers, public bodies and industry, are poised to take advantage of new opportunities offered by the Welsh Government’s Action Plan. Most importantly, the language community itself is engaged, both as consumers of digital materials and devices and as potential producers of new Welsh digital content and software. Language technologies offer a path towards economic regeneration as well as language revitalization, and while there is much hard work still to be done, current progress is encouraging.

References:

1. Kamolov, A. A., & Raximov, M. S. H. U. (2018). EFFICIENT ATTRACTION OF INVESTMENTS IN THE ECONOMY. *Теория и практика современной науки*, (1), 762-765.
2. Kamolov, A. A., & Aliyev, A. R. U. (2018). MARKET: ESSENCE AND FUNCTIONS. *Теория и практика современной науки*, (1), 768-770.
3. Аскарова, Д. К. (2018). Особенности воспитания в семье детей дошкольного возраста. *Молодой ученый*, (6), 161-162.
4. Аскарова, Д. К. (2017). ДЕЯТЕЛЬНОСТЬ САИДАХМАДХОДЖА СИДДИКИЙ. *NovaInfo. Ru*, 6(58), 407-409.
5. Аскарова, Д. К. (2016). НАРОДНОЕ ТВОРЧЕСТВО И ЕГО ВОСПИТАТЕЛЬНОЕ. *NovaInfo. Ru*, 3(41), 160-162.

6. Аскарова, Д. К. (2016). СОЦИАЛЬНАЯ ФУНКЦИЯ СЕМЬИ ПРИ ФОРМИРОВАНИИ ЛИЧНОСТИ РЕБЁНКА. *NovInfo. Ru*, 2(42), 209-212.
7. Аскарова, Д. К. (2019). Творческие задания на уроках математики в начальных классах и предъявляемые к ним требования. *Молодой ученый*, (9), 181-183.
8. Khodjayeva, D. S. (2020). Synonymy between dictionary units and occasionalism. *EPRA International Journal of Research and Development (IJRD)*, 5(8), 323-324.
9. Shavkatovna, K. D., & Davlatjonovich, K. E. TEACHING SLOW LEARNERS IN RUSSIAN AND ENGLISH CLASSES.
10. ХОДЖАЕВА, Д. СПОСОБЫ ВЫРАЖЕНИЯ ОБСТОЯТЕЛЬСТВЕННОЙ СЕМАНТИКИ ВО ФРАЗЕОЛОГИЗМАХ.
11. Мухамедов, У. С. (2019). ТЕХНИЧЕСКИЕ СРЕДСТВА ДЛЯ КОМПЬЮТЕРНОЙ ГРАФИКИ. *Мировая наука*, (10), 135-138.
12. Умаров, А. С. (2022). УЗЛУКСИЗ ТАЪЛИМДА ЗАМОНАВИЙ САНЪАТ МАКТАБЛАРИНИ ТАШКИЛ ЭТИШНИНГ КЛАСТЕР ТАМОЙИЛЛАРИ. *Research Focus*, 1(1), 23-28.
13. УМАРОВА, М. ЭКОНОМИКА И СОЦИУМ. *ЭКОНОМИКА*, 708-713.
14. Khodjayev, K. K. (2021). THE SPECIFICITY AND COMPLEXITY OF THE PROCESS OF LEARNING ENGLISH.
15. Abdug'afurovich, R. B. (2022). Innovation Technologies in Teaching English. *American Journal of Social and Humanitarian Research*, 3(6), 288-291.
16. Bahromjon, R. A. O. (2021). INNOVATIVE METHODS IN TEACHING FOREIGN LANGUAGES FOR STUDENTS OF NON-LANGUAGE UNIVERSITIES. *ResearchJet Journal of Analysis and Inventions*, 2(05), 53-59.
17. Razzoqov, B. (2021). SOME PROBLEMS IN LEARNING ENGLISH AND WAYS TO SOLVE THEM. *Интернаука*, (21-4), 92-93.
18. Plyosbek Ilhomjon O'g'li Tojiboev, Baxrom Abdug'afurovich Razzakov, Munojat Azamjonovna Sharofiddinova, Kamoliddin Kodirovich Khudjayev (2022) Methods of improving students' speaking completeness in teaching foreign languages in technical universities (In the example of construction, agricultural mechanization). *International Journal of Mechanical Engineering*, 3(7), 65-69.
19. Khodjayev, K. K. (2021). THE SPECIFICITY AND COMPLEXITY OF THE PROCESS OF LEARNING ENGLISH.
20. Khodjayev, K. K. (2021). FEATURES OF USING INTERACTIVE METHODS OF TEACHING ENGLISH. *Экономика и социум*, (5-1), 239-242.
21. Kodirovich, K. K. (2022). Communicative Competence and its Practical Reflection. *American Journal of Social and Humanitarian Research*, 3(6), 292–294.
22. Kodirovich, K. K. (2020) THE IMPORTANCE OF GAME METHODS IN LEARNING ENGLISH. *International Engineering Journal For Research & Development*, 5, 3.
23. Хужаев К. К. Анваров А. А., Саттаров С. Я. (2016) Classification of programs for learning English. *Молодой учёный*, 3 (107), 771-772.

24. Mamatova, G. (2021). Improving methodology of game technology in the teaching of lexical materials in English lessons. *Science and Education*, 2(12), 571-573.
25. Mamatova, G. INGLIZ TILI DARSLARIDA SO'Z BOYLIGINI OSHIRISHDA INTERFAOL O'YINLARDAN FOYDALANISH. JOURNAL OF NEW CENTURY INNOVATIONS, 3 (1), 115-120
26. Mamatova, G. GAMES AS A TOOL TO TEACH VOCABULARY Ta'limda raqamli texnologiyalarni tadbiiq etishning zamonaviy tendensiyalari va rivojlanish omillari, 5(1), 57-59.
27. Эрнazarова, Ё. (2016). ШАХС КАСБИЙ ФАОЛИЯТИДА АХЛОҚИЙ ФАЗИЛАТЛАРНИНГ ЎРНИ. ВЕСТНИК КАРАКАЛПАКСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА ИМЕНИ БЕРДАХА, 33(4), 52-53.
28. Ollaberganovna, E. Y. (2022). Improving the Professional Culture of Civil Servants as an Important Factor in the Development of Civil Society. *American Journal of Social and Humanitarian Research*, 3(8), 153-160.
29. Ernazarova, Y. O. (2016). PEDAGOGICAL ASPECTS OF FORMATION OF MORAL-AESTHETIC CULTURE OF THE PROFESSIONAL ACTIVITY OF STUDENT. *Theoretical & Applied Science*, (11), 143-146.
30. Ollaberganovna, E. Y. (2022). Socio-Philosophical Essence of the Professional Culture of the Government Officer. *Central Asian Journal of Literature, Philosophy and Culture*, 3(10), 88-96.
31. Эрнazarова Ё. (2016). Фуқароларнинг ўзини ўзи бошқариш органларининг ёшларни касбга йўналтириш соҳасидаги фаолияти (хуқуқий асосларнинг ривожланишига оид). Фуқаролик жамияти, 3(47), 40-43.
32. Rutkauskas, A. V., & Ergashev, A. (2012). Small business in Uzbekistan: situation, problems and modernization possibilities'. In *7th International Scientific Conference on Business and Management, Vilnius, Lithuania*.
33. Эргашев, А. М. (2017). Ўзбекистонда кичик бизнес ва оилавий тadbirkorликни молиявий институтлар томонидан қўллаб-қувватланиши. *Иқтисодиёт ва таълим*, 8(6), 106.
34. Ergashev, A. EXPERIENCE OF FOREIGN COUNTRIES AND UZBEKISTAN IN DEVELOPMENT OF SMALL BUSINESS.
35. Эргашев, А. М. (2016). Аҳоли фаровонлигини таъминлашда оилавий тadbirkorликнинг ўрни ва аҳамияти. Тежамкорликнинг концептуал асослари ва унинг ижтимоий-иқтисодий шарт-шароитлари. 2(174), 254.
36. Toshkhujayeva, S. (2021). LINGUAPOETIC RESEARCH OF BELLE-LETTER–DESCRIPTIVE MEANS. *World Bulletin of Social Sciences*, 4(11), 47-51.
37. ТОШХУЖАЕВА, Ш., & РАСУЛОВА, О. (2021). ЛИНГВОПОЭТИЧЕСКИЕ ВОЗМОЖНОСТИ ПЕРЕНОСНОГО ЗНАЧЕНИЯ СЛОВ. *CENTRAL ASIAN JOURNAL OF LITERATURE, PHILOSOPHY AND CULTURE*, 2(11), 1-3.
38. Тошхужаева, Ш. Г. (2016). Лингвопоэтическое исследование художественной литературы–описательные средства. *Молодий вчений*, (1), 382-386.
39. Тошхужаева, Ш. Г. (2016). Использование метафор в работах Эркина Азама. In *The Chicago Journals in Liberal Arts* (pp. 76-79).