# Free Software for Indigenous Languages

Kevin P. Scannell

Department of Mathematics and Computer Science
Saint Louis University
Saint Louis, Missouri, USA

19 August 2008

I am a speaker of Irish (sometimes referred to as "Irish Gaelic"), the indigenous language of Ireland. Like most other indigenous languages in the world, Irish is dying out as a community language in the few remaining areas where it is still spoken. Indeed, a study published in 2007 predicts that it will disappear entirely as a community language within the next twenty years.[1] We find ourselves in this dire situation despite the fact that the language is taught in schools throughout Ireland and receives support from the government as a result of it being enshrined in the Constitution as "the first official language" of the country (English is recognized as "a second official language").

The dynamics of the ongoing shift to English are probably familiar to all speakers of indigenous languages: English is viewed as the language of commerce, technology, higher education, and social mobility. It is the language of the Anglo-American television programs, movies, and music that young people find most attractive. Irish, on the other hand, is associated with Ireland's rural, impoverished past, with "old-fashioned" culture, and is generally viewed as "useless" in the global economy of the 21st century.[2]

My work to help reverse this language shift has focused on the domains of software and computing. Up until about seven years ago, there was virtually no software available in Irish, and this meant that when children were introduced to computers in school, it was necessarily in an English-only context. Similarly, in the business world, the lack of Irish software made it extremely difficult to maintain Irish as the language of the workplace.

For many years now the desktop computing market has been dominated by Microsoft and Apple, and so many language groups around the world have invested time and money in producing

---

[1] "Staidéar Cuimsitheach Teangeolaíoch ar Úsáid na Gaeilge sa Ghaeltacht" (Comprehensive Linguistic Study of the Usage of Irish in the Gaeltacht), prepared on behalf of the Department of Community, Rural and Gaeltacht Affairs in 2007. See `http://www.pobail.ie/en/AnGhaeltacht/LinguisticStudyoftheGaeltacht/`.

[2] See for example the results of studies conducted in the 70's and 80's concerning attitudes towards the language: Ó Riagáin, Pádraig; Mícheál Ó Gliasáin. "The Irish Language in the Republic of Ireland 1983: Preliminary Report of a National Survey". Dublin: Linguistics Institute of Ireland, 1983.

translations of these operating systems, often with unfortunate results. For example, in the 1990's a small company in Ireland worked with Apple to translate the old Macintosh operating system into Irish, but because there was no financial incentive to maintain this translation in the long run, Apple decided to drop it from later versions. In another notable case, the government of Bhutan made a huge investment (more than half a million dollars) in order to have support for their national language, Dzongkha, added to Microsoft Windows, only to have it dropped by Microsoft under pressure from the Chinese government, because of the political situation in neighboring Tibet (Dzongkha and Tibetan are closely related languages and share the same script).[3]

The central point here is that there are risks associated with translating software on behalf of a large, for-profit company. To even begin a translation, the company must agree to make the various translatable items in each program available to translators. This is in itself something non-trivial, since the translatable items are usually extracted directly from program source code, which most companies protect as proprietary intellectual property. And even if one manages to complete the translations, they must still be integrated into a finished product that can be made available to users, and this again depends on having access to the source code. It also depends on the willingness of the company to bear any additional costs associated with producing and distributing the translated software. With each new version this process repeats, and one must hope that financial or political realities don't interfere as in the examples above. In short, in such situations the language community does not control their own destiny or their own translations.

Fortunately there is an alternative. The *free software* movement was started in the 1980's by Richard Stallman, and represented a radical new perspective on how to produce and distribute software. The basic principle is that everyone should have the freedom to copy, modify, or redistribute software, as long as their redistributed versions guarantee these same freedoms to others. Because one needs to be able to see the underlying source code of a computer program to modify it, free software is sometimes called "open source" software. From the point of view of software developers, the fact that the source code is visible to the world means that free software tends to be more secure and have fewer bugs than proprietary software. For indigenous language groups, it means that we are no longer dependent on the goodwill of large companies in the senses described in the previous paragraph.

Two of the most prominent free software projects are the web browser Mozilla Firefox (`http://www.mozilla.com/firefox/`) and OpenOffice.org (`http://www.openoffice.org/`), a free alternative to Microsoft Office. Both programs work on Apple, Microsoft, and Linux computers. Firefox offers many advantages in terms of speed, security, and customizability over browsers like Apple's Safari and Microsoft's Internet Explorer and has steadily been increasing in market share since it was first released about four years ago. OpenOffice.org contains a full-featured editor, spreadsheet software, and presentation software. Because it is available free of charge, it has been widely adopted in countries where the commercial analogues (Word, Excel, Powerpoint) are

---

[3]Poser, Bill, "Microsoft outlaws Dzongkha". Language Log. 2 November 2005. (`http://itre.cis.upenn.edu/~myl/languagelog/archives/002619.html`) Retrieved 22 October 2008.

prohibitively expensive. But don't take my word for it – follow the links above and try them out yourself, for free!

Translating these programs is very easy. In fact, thanks to a wonderful web-based program called "Pootle" you don't even need to have them installed on your computer to contribute translations – all you need is a simple web browser. For example, if you visit the web site `http://pootle.translate.org.za/projects/fftb/`, you'll see translations of Firefox into several languages – in this case, the eleven official languages of South Africa. You can navigate the files that need translation as you would navigate any other web site. As a demonstration, click one of the languages, "Swati" for example. Then you'll see a number of folders: "browser" "dom" "editor" etc. Clicking "netwerk" (sic) will display a file named "necko.properties.po" Clicking this will open the file to be translated, showing the original English messages on the left-hand side and the target language translations on the right-hand side. This particular file happens to be complete, but once you have registered with the site and started a project for your own language, you'll be able to enter translations directly into this window (details below). There is a corresponding Pootle web site for translating OpenOffice.org, which hosts a significantly greater number of languages (31 in all): `http://pootle.sunvirtuallab.com/`.

Pootle is not the only tool available for translating these programs. For example, in some countries where Internet access is intermittent or costly, a web-based solution is not ideal, and so similar tools have been developed that can run on the translator's desktop, and the resulting translations can be uploaded periodically to a server. Every team finds the tools that work the best for their local situation, and the results speak for themselves: Firefox 3.0 was officially released in 48 languages, with another 22 languages available as unofficial versions contributed by the community. In all, **98 languages** have translation projects underway, including some seriously endangered languages such as Irish and Upper Sorbian (less than 20,000 daily speakers each). I'd like to see many more indigenous languages and endangered languages added to this list!

It is worth looking at some of the obstacles to creating fully-translated versions of these programs. First of all, Firefox and OpenOffice.org are relatively large programs and contain thousands of messages and menu items needing translation. I've been translating free software into Irish for about eight years now, primarily in collaboration with my friend Séamus Ó Ciardhuáin from the Limerick Institute of Technology, and we're still only about 75% of the way toward our goal of a complete Irish desktop system (web browsers, office software, games, email handlers, multimedia software, etc.). At the same time, taking the long view, we've made tremendous progress. And even though quite a few other people have contributed translations over years (at least a dozen), the vast majority of the translation work has been done by Séamus and myself. This is not to discount the important contributions others have made, but simply to illustrate just how much can be accomplished by one or two dedicated individuals (working on a volunteer basis and entirely in our spare time).

Creating terminology can also be serious obstacle when undertaking a new translation project into an indigenous language: how does one say "download" or "web browser" in Irish, or Cherokee

for that matter? As difficult as terminology creation can be at times, to me it is actually one of the truly enjoyable parts of the work; it's a chance to be creative and use native elements of your language in new and interesting ways. For example, *comhad* is an Irish word meaning "a place for holding or keeping something" according to the old dictionaries; it came to be used for (paper) files years ago, and by extension for computer files more recently. Now Irish also has a suffix *-lann* meaning literally "house" or "floor" that combines as follows: *leabhar* (book) + *-lann* gives *leabharlann* (library), *amharc* (looking) + *-lann* gives *amharclann* (theatre), etc. These two elements are combined to give the new word *comhadlann* for a directory on a computer. Similarly, the old word *ríomh* meaning "calculating" or "counting" is the root of the modern word *ríomhaire* for computer, and combines in many ways to give computing terminology, e.g. *ríomhlann* for "computer lab" as above, or *ríomhphost* for "e-mail".

For some languages there are strict controls over who may coin new terminology, while for other languages there are no restrictions at all, and terminology evolves "organically" over time. Irish lands somewhere between these two extremes; we have an official Terminology Committee that creates terminology and maintains an online database (`http://focal.ie/`) that we consult quite often. In instances where we need a translation that is not in the database, we coin the necessary terms ourselves, and in some cases have passed these on to the Committee for consideration as "official" terms. In rare instances we have been unhappy with an existing official term (usually for technical reasons) and have used our own preferred translation. For languages where there is no official control over terminology, it is important to involve the community as widely as possible at an early stage in order to select terms that will be easily understandable by the target audience, and that are correct, both technically and linguistically.

The logistics of translation work vary greatly from language to language, depending on the tools that are chosen for translation, and on the size of the team. In reality, most "teams" are like Irish and have only one or two active translators. In our case, time is split between work on new translations (existing packages that do not yet have Irish versions) and maintenance of translations that were completed earlier. Maintenance takes up a surprising amount of time – free software programs tend to upgrade to new versions more often than proprietary programs precisely because the source code is visible and it's easier to fix bugs and make improvements. For translators, this means the programs can sometimes seem like "moving targets"; as soon as one version is complete, work begins on the next version. The good news is that the vast majority of translatable items remain unchanged version to version, and the changes tend to be small: e.g. differences in punctuation, capitalization, phrasing, or terminology. To keep up with these incremental changes, Séamus and I have instituted a system we call "Patch Tuesdays" – on the second Tuesday of each month we gather together and translate any new or changed messages from our existing packages. This way, when the time comes to prepare a new version for release, the process is relatively painless.

Some translators fear they do not have the technical expertise to translate software. Indeed, every translator will at some point come across a technical-sounding error message that seems to make no sense. This is where the free software community comes in. There are mailing lists for

each of the major free software projects dedicated entirely to discussions of issues that arise during the translation process. When you encounter a confusing message, it is simply a matter of sending a quick email to the appropriate list and you're certain to get a response within a few hours, if not minutes, sometimes from a fellow translator who has encountered the same difficulty already, and other times from the program developers themselves. There is a real feeling of solidarity among the translators working on free software, particularly so among those of us working on smaller languages.

What to translate first? There are literally thousands of free software projects out there to choose from. The graphics program "TuxPaint" (`http://www.tuxpaint.org/`) is an appealing choice for many reasons. First, it is aimed at young children, who make up the most important demographic in language revitalization efforts. Second, in practical terms, it is quite easy to translate: there are only about 240 translatable items and many of these are comprised of one or two words, e.g. names of colors. Also, it can be translated using Pootle like the programs mentioned above; see `http://translate.sourceforge.net/wiki/decathlon/tuxpaint#procedure_for_translating_tux_paint` for detailed instructions on getting started.

For groups wishing to be a bit more ambitious, I would recommend Firefox as a first project. It is probably the highest-profile free software project out there and it is a wonderful piece of software for end-users, children included. Also, there is a tremendous support infrastructure in place to help new translation teams, and the main menus are small enough and non-technical enough that it's easy to get a decent partially-translated version up and running quickly.

There are many other possibilities. Below is a small selection of popular free software packages; they can all be translated using Pootle. A comprehensive directory of thousands of free software packages can be found at `http://directory.fsf.org/`.

- TuxPaint. Drawing program for children: `http://www.tuxpaint.org/`

- Mozilla Firefox. Web browser: `http://www.mozilla.com/firefox/`

- OpenOffice.org. Free office suite: `http://www.openoffice.org/`

- Mozilla Thunderbird. Email handler: `http://www.mozilla.com/thunderbird/`

- Audacity. Sound editor: `http://audacity.sourceforge.net/`

- WordPress. Blogging system: `http://wordpress.org/`

- Abiword. WYSIWYG editor: `http://www.abisource.com/`

- VLC. Multimedia player: `http://www.videolan.org/`

- Art of Illusion. Package for 3D modeling: `http://www.artofillusion.org/`

- Plone. Content management system: `http://plone.org/`

- Yudit. Text editor with excellent Unicode support: `http://www.yudit.org/`

For me personally, this work has provided the chance to work together with other Irish speakers to create something that has a real impact on people's daily lives, and something which belongs to the community as a whole. It's also opened my eyes to how much indigenous language groups can achieve by working together; we are, after all, facing the same obstacles. If you're interested in learning more, I'm willing to help – just send me an email at `kscanne@gmail.com`.