A Case Study in Community-Driven Translation of a Fast-Changing Website

David Ellis

Facebook
dellis@facebook.com

Abstract. Facebook's translation tool allows users (translators) to click on a phrase as they browse the site, and inline see the original native string, vote on translations suggested by their peers or offer their own. We offer an innovative approach to web site internationalization that leverages a unique infrastructure and a dedicated user community to keep our interface up-to-date in translation

1 Introduction

Facebook's translation tool allows users (translators) to click on a phrase as they browse the site, and inline see the original native string, vote on translations suggested by their peers or offer their own. We offer an innovative approach to web site internationalization [1] that leverages a unique infrastructure and a dedicated user community to keep our interface up-to-date in translation. Each language community starts by translating glossary terms to encourage consistency across the site. Once phrases are translated (inline and/or in bulk mode), we hire professional linguists to ensure quality in the most commonly viewed strings.

2 Motivation

A variety of factors motivated the design of our translations process and application, but chief among them are speed, quality, cost and reach.

2.1 Speed

It is essential for our translations to keep up with new features and other changes to the interface. The traditional model is both costly and unscalable. Crowdsourcing allows us to take advantage of the size and motivation of our user community (more than 25,000 Turkish users added the translations application), often completing translations in days or weeks (as opposed to months required by professionals).

Our user base keeps growing, and with each person who signs up for Facebook but can't view it in their language, we have another potentially motivated translator. Using collaborative translations, Spanish and German were translated in

N. Aykin (Ed.): Internationalization, Design, LNCS 5623, pp. 236–244, 2009.
© Springer-Verlag Berlin Heidelberg 2009

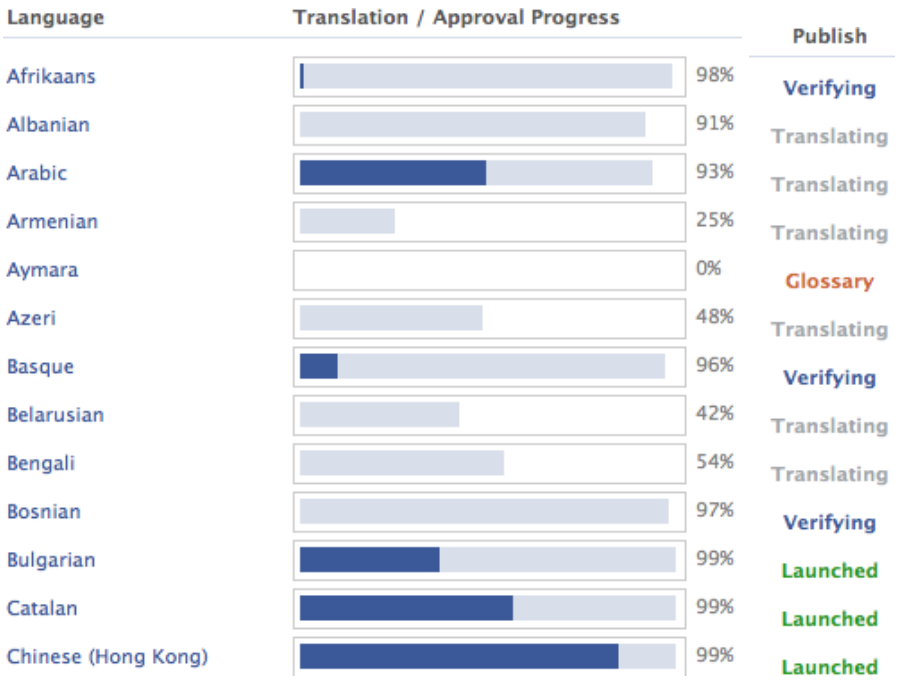

Fig. 1. Progress of a few languages, along with publication status

one week, a draft of French (ready for review) was completed in 24 hours, and we have launched over 40 languages within the first year of our localization effort. For statistics on progress of languages from A to Ch, see Fig. 1.

2.2 Quality

A translation is of high quality if it does all of the following:

1. Accurately conveys the original meaning of its source.
2. Does not sound like a translation, but a native phrase or document.
3. Results in clear and unambiguous text.

A method to achieve quality entails employing the right combination of people, process and technology. Selected people must be linguistically competent and must be experts in the relevant industry. The process should facilitate speed and quality. The best technology is one that allows the implementation of process with the least friction, which we are actively working to iteratively improve.

We also have a style guide and glossary for each language, and in-tool warnings for inconsistent punctuation (e.g., translating "Click here:" as "¡Haga clic aquì!"), for leaving out glossary terms (e.g., translating "A friend's profile has the following components..." without using the translated glossary term "amie" for friend), and for submitting identical translations.

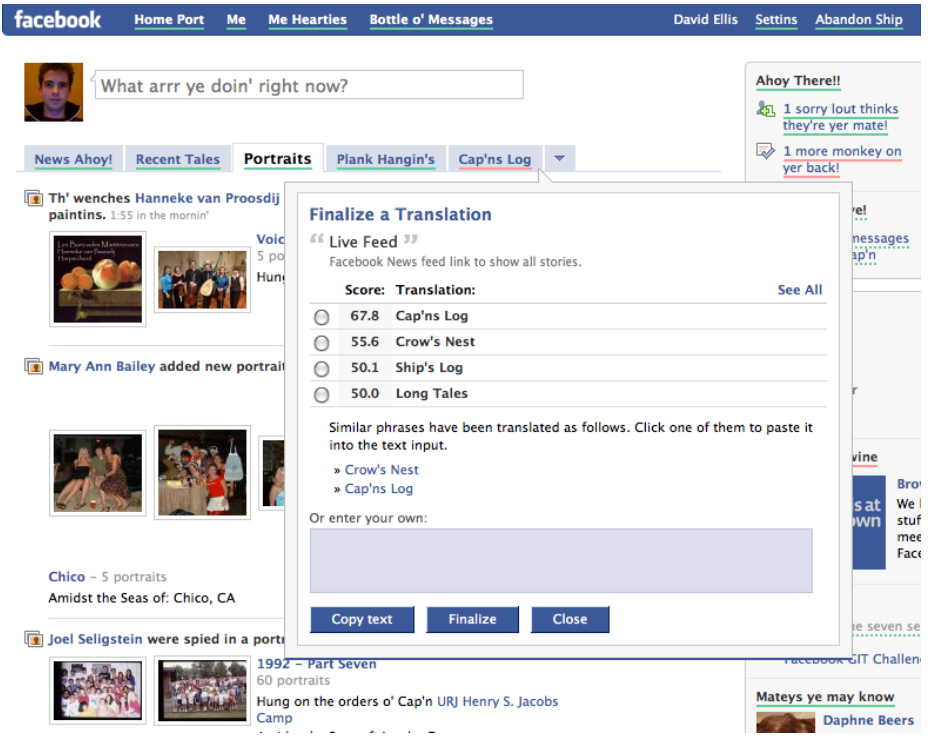

Fig. 2. Inline approval mode, in the English (Pirate) locale

2.3 Cost

Investment in crowd-sourcing technology offers some of the savings associated with "free" translations. However, much of that is spent because the majority of community translations undergo extensive quality assurance effort by professional agencies. Our cost advantages are achieved through unsupported languages, process automation and the ability to prioritize text.

Unsupported Languages. Welsh is our first unsupported language (chosen largely due to the relatively low visibility and small adoption), but there are many more in the long tail that will go without professional quality assurance. We also offer translation into "English (Pirate)", which has launched in beta through the efforts of over 20,000 volunteer translators who have submitted and voted on nearly 40,000 translations.

Automation. The automation extends even to the quality assurance, which takes place in our tool. See a screenshot of its welcome page in the Estonian locale in Fig. 3. We have begun to hire full-time, onsite language managers for popular languages (including French, Spanish, Chinese, Japanese, and German).

Prioritization. Rather than compile a word-processing document or spreadsheet containing a batch of new phrases to be translated, we ask translators

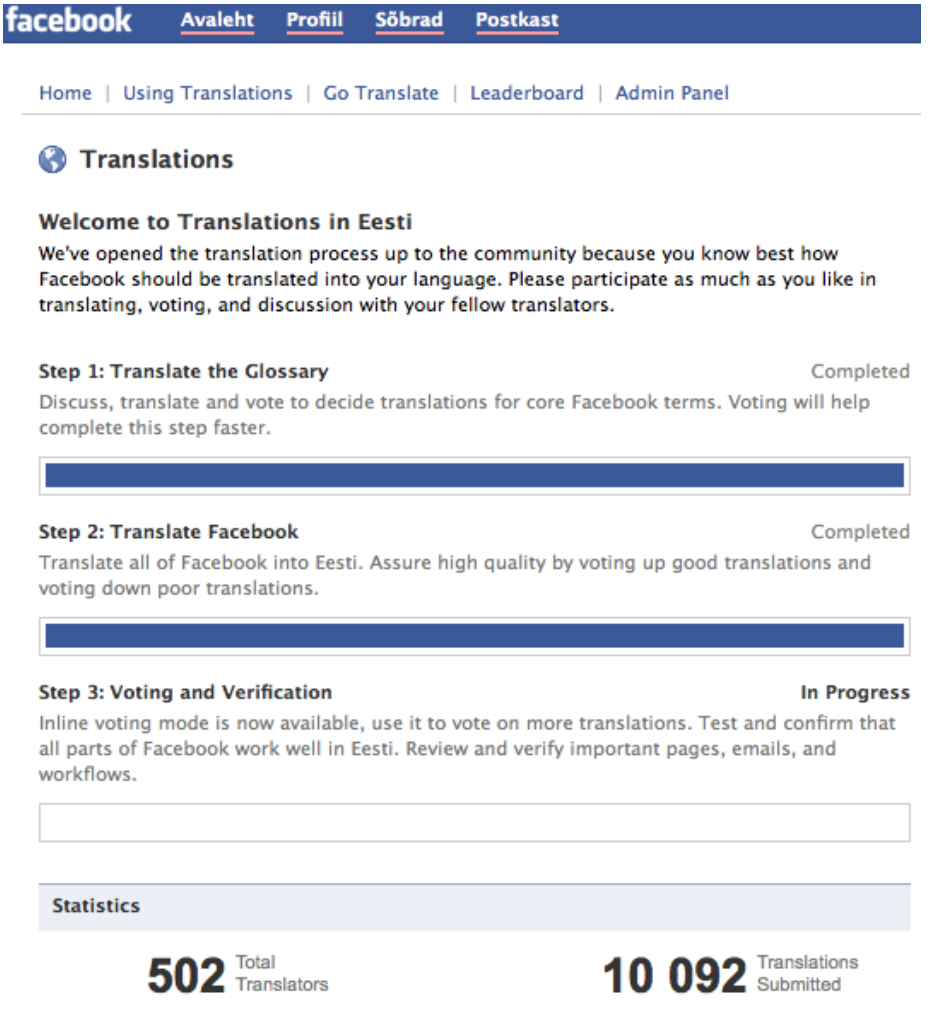

Fig. 3. The translations application, in the Estonian locale

(normal users and contracted professionals) to browse the site, a process which will naturally surface high-visibility untranslated strings. But most users rarely see some parts of the site: for instance, the sign-up flow. Since these are critical to growth and new user experience, we prioritize them in the bulk interface (as in Fig. 4).

2.4 Reach

Our crowd-sourced translations process allows us to extend the reach of Facebook to all internet users, including speakers of commonly ignored languages. We also give application developers access to the same technology and process as used

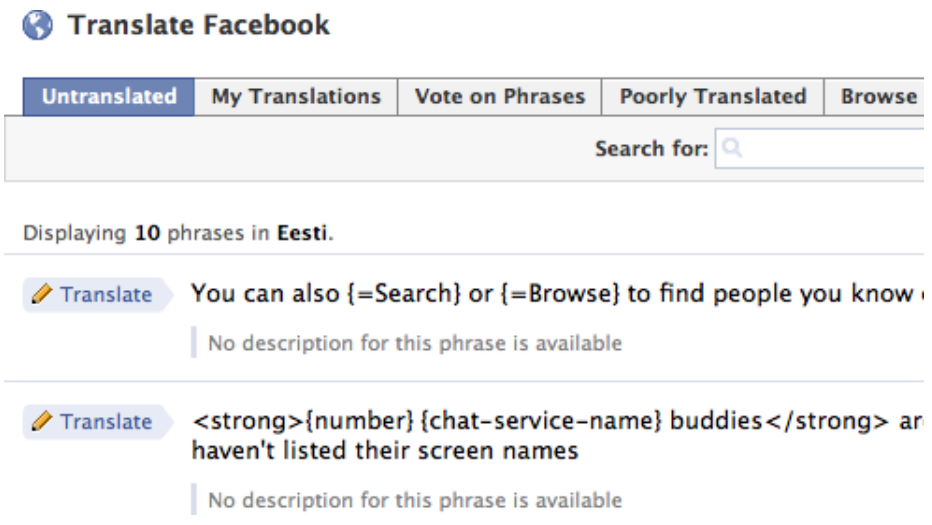

Fig. 4. The bulk translation interface, which allows translators to translate or vote on strings, filtering by a specific page or group of pages

on the rest of our site [3]. Note that with Facebook Connect, those applications don't even need to be on Facebook, and can allow a user to carry his identity, friends and locale (language preferences) to any domain, where the translated interface is generated using our process.

3 Framework

We offer a rich interface with simple but powerful controls, an example of which can be seen in Fig. 5. Many phrases on the site necessarily involve tokens like "{name}" that are replaced with values that depend on the context in which they appear (e.g., which user is logged in). Translating tokenized phrases is problematic because the values (words) need to be inflected in some languages. We have two systems in place to enable correct translations, requiring minimal effort from translators.

3.1 Dynamic Explosion

This technique allows us to split strings on language-specific variations based on translator feedback. For example, a Hebrew translator indicates that in the phrase "name wrote on your wall", the verb conjugation depends on the gender of the subject. Translators can then submit (and vote on) translations for each case: where the actor is male, female, or unspecified.

In Arabic, there are different inflections for singular, dual and plural, so in the phrase "number hours ago", the value of the number affects the translation.

Fig. 5. A close-up of the inline translation dialog (from Fig. 2) that includes results of voting on existing translations, and reveals similar-phrase translations (to ease translators' burden and encourage consistency)

Translators can easily see and modify each of these translations (as in Fig. 6), and the appropriate variant is shown to Arabic users (in this case, in their newsfeeds).

3.2 Phonological Rules

Orthographic or phonological rules [4] can affect the spelling of words, and are applied automatically when tokens are substituted with their values. For example, Turkish inflection rules affect any token in possessive, dative or accusative case, such that there are 12 different forms for each. We allow translators to use a proto-form that will be adjusted to match the token when displayed.

Specifically,

“{name1} wrote on {name2}’s wall”

is translated as follows:

“{name1} {name2}’(n)in duvarına yazdı”

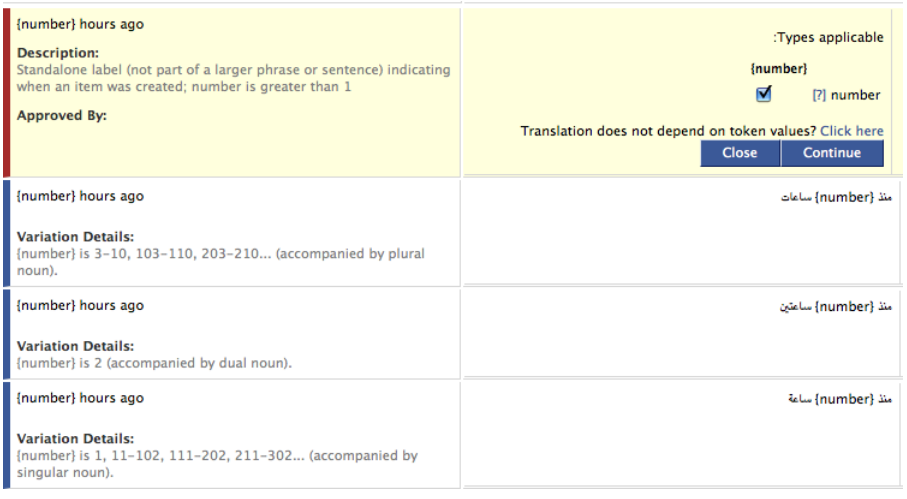

Fig. 6. This interface, for separation (explosion) of a native string into related strings, allows translators to specify how phrases depend on variations in token values

If {name2} is "Malmö", it will be displayed as "...Malmö'nün..."
But if {name2} is "Barış", it will be displayed as "...Barış'ın..."
Without this framework for handling dynamic content in a linguistically sound manner, many users would see the site as having worse than a 3-year-old's comprehension of their language.

4 Conclusion

Over 90 languages are in active translation using our tool. We have recently launched (in beta) languages with right-to-left scripts, like Hebrew and Arabic, where the layout must match the directionality of the text (see 7). Since we began this effort, our international growth has skyrocketed [5], and user experience (as measured by activity) per locale has improved with the quality of translations. Application developers on Facebook Platform can take advantage of the same tight integration and feedback loop to get their interface translated by interested users [2].

4.1 Challenges

We have to keep the community interested and engaged throughout the initial translation and maintenance phases. It is also important that we have enough available contributors to complete the translation, especially when working with "unsupported" languages. Ensuring acceptable quality in all languages is very important and difficult.

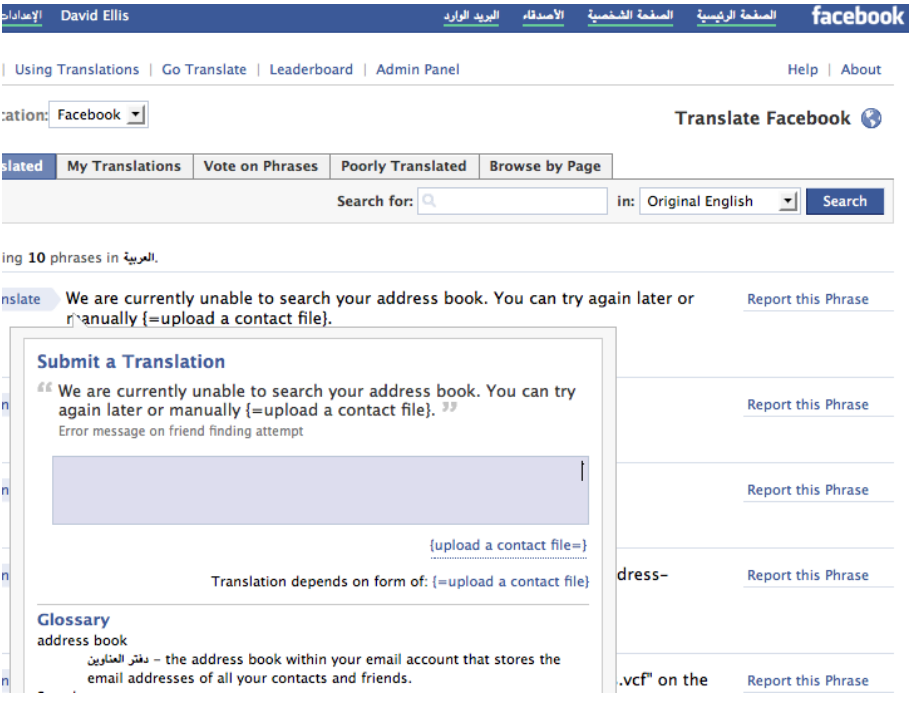

Fig. 7. Bulk translation interface in Arabic, where layout is appropriately shifted (e.g., our logo is on the right of the header)

Language re-use could be useful, particularly in cases of related languages or dialects (e.g., Spanish from Spain versus Latin America, Catalá, Portuguese (Brazil and Portugal)), but how can we allow them to share translations, and streamline the process of finer-grained localization? We have experimented with shallow machine translation in these cases, but the quality isn't yet good enough.

From an engineer's perspective, technical issues, including access, scalability, control, and security are also essential for the tool to be useful. From a linguist's perspective, the highly dynamic nature of our text, along with the complexity of some languages, must be properly handled to ensure high quality.

Acknowledgments

This technology has been developed with support from i18n team (engineers, language managers and others) at Facebook, and all our international users.

References

1. Aykin, N.M.: Internationalization and localization of the web sites. In: Bullinger, H.-J., Ziegler, J. (eds.) HCI (1), pp. 1218–1222. Lawrence Erlbaum, Mahwah (1999)
2. Bunyan, K.: Tutorial: Translating your applications using Facebooks crowd-sourced translation service (2008)

3. Facebook. Translating platform applications: Facebook developers wiki. Website (2009),
http://wiki.developers.facebook.com/index.php/Translating_Platform_Applications
4. Kaplan, R.M., Martin Kay, T.: Regular models of phonological rule systems. Computational Linguistics 20, 331–378 (1994)
5. Sargent, B.B.: Community translation lifts facebook to top of social networking world (2008)