

Co-creating a repository of best-practices for collaborative translation

Alain Désilets

National Research Council of Canada

Jaap van der Meer

Translation Automation Users Society

Collaborative translation has the potential for significantly changing how we translate content. However, successful deployment of this kind of approach is far from trivial, as it presents potential adopters with a rich and complex envelope of processes and technologies, whose respective impacts are still poorly understood. The present paper aims at facilitating this kind of decision making, by describing and cataloguing current best-practices in collaborative translation. More precisely, we present a collection of Design Patterns which was created collectively by a small group of practitioners, at a one-day roundtable hosted by the Translation Automation Users Society in October of 2011. This collection has been put on an open wiki site (www.collaborative-translation-patterns.com) in the hopes that other practitioners in the field will refine and augment it.

1. Introduction

Collaborative and social networking technologies, as seen on sites like Wikipedia, Facebook and Amazon Mechanical Turk, are having profound effects in many spheres of human activity. Translation is no exception, as evidenced by Facebook's use of crowdsourcing, to co-opt its loyal user base into translating the system's web interface on a volunteer basis (Ellis, 2009). Using this approach, Facebook was able to rapidly recruit 250,000 volunteers, who translated 350,000 words into 70 languages, often with very short lead time (less than two days for high density languages like French) (Baer, 2010).

Highly publicized cases like this caused somewhat of a commotion in the translation industry, as exemplified by the strong negative reaction of professional translators to a similar crowdsourcing attempt by LinkedIn, a professionally-oriented social network (Stejskal, 2009; ATA, 2009). At the time, translation crowdsourcing conjured up a picture where customers could get fast translations at rock-bottom prices, through the work of volunteer or offshore translators and the help of new technology acquired from vendors, while professional translators risked seeing their profit margins shrink drastically. However, two years after the Facebook and

LinkedIn initiatives, one cannot help but notice that the impact of translation crowdsourcing remains marginal, and that successful uses of this paradigm are still few and far between.

This does not mean that collaborative technologies cannot have a major impact on the translation industry. It simply means that the particular model used by Facebook is just one of many ways in which massive online collaboration can be leveraged for translation, and that the applicability of this particular approach may be limited to very specific contexts (for example, situations where there is a community of people with a strong emotional bond to the content being translated).

Indeed, the term collaborative translation can be used to encompass a much wider collection of approaches, such as agile translation teamware, collaborative terminology resources, translation memory sharing, online translation marketplaces, post-editing by the crowd, and of course, translation crowdsourcing (Note: all these terms will be defined more precisely below). At present, selecting and successfully deploying one of those variants is somewhat of a black art, and one must navigate through many challenging issues such as quality assurance, crowd motivation and the dangers of de-contextualization.

While there are many case studies which describe how collaborative translation was successfully implemented in specific organizations (Ambati & Vogel, 2010; Baer, 2010; Baer 2011; Bloodgood & Callison-Burch, 2010; Calvert, 2008; Ellis, 2009; Meyer, 2011; Munro, 2010; Yahaya, 2008), there is a clear need for a more concise, summative body of knowledge that captures recurrent best practices. In this paper, we advocate the building of such a compendium, in the form of a collection of design patterns (“Design pattern”, 2011), which could be written collectively by practitioners of collaborative translation.

The remainder of the paper is organized as follows. In Section 2, we provide an overview of the burgeoning field of collaborative translation. In Section 3, we provide an introduction to design patterns, and explain why they seem like a good format for capturing best practices of collaborative translation. Finally, in Section 4 we report on the outcome of a day-long workshop which was hosted by TAUS in October of 2011, with the explicit goal of generating such a collection of patterns. It brought together several leading practitioners of collaborative translation who worked together to create collaborative-translation-patterns.com, a wiki site that captures some of the most successful best-practices in the field today.

2. An overview of Collaborative Translation

2.1. The different flavors of Collaborative Translation

As mentioned in the Introduction, collaborative translation encompasses more than just Facebook-style translation crowdsourcing. Possible uses of collaborative technologies in translation include the following.

- **Agile translation teamware:** wiki-like systems and processes that allow multidisciplinary teams of professionals (translators, terminologists, domain experts, revisers, managers) to collaborate on large translation projects, using an agile, grassroots, parallelized process instead of the more top-down, assembly-line approach found in most translation workflow systems. Examples of this approach can be found in Beninato & De Palma, 2008; Calvert, 2008; Yahaya, 2008.
- **Collaborative terminology resources:** Wikipedia-like platforms for the creation and maintenance of large terminology resources by a crowd of translators, terminologists, domain experts, and even general members of the public. Examples of this approach include Wiktionary (Wiktionary, 2011), ProZ's Kudoz forum (Goddard, 2010) and the Urban Dictionary (Urban Dictionary, 2011).
- **Translation memory sharing:** platforms for large scale pooling and sharing of multilingual parallel corpora between organisations and individuals. Examples of this approach include the TAUS Data Association (TAUS Data Association, 2011), MyMemory (MyMemory, 2011) and Google Translator Toolkit (Google Translator Toolkit, 2011).
- **Online marketplaces for translators:** eBay-like disintermediated environments for connecting customers and translators directly, with minimal intervention by a middle man. Examples of this include ProZ.com (ProZ.com, 2011), TranslatorsCafe (TranslatorCafe, 2011) and Translated.net (Translated.net, 2011).
- **Translation crowdsourcing:** Mechanical-Turk-like systems to support the translation of content by large crowds of mostly amateurs, through an open-call process. This is by far the most talked about collaborative translation approach. It has been used successfully for translating a variety of content types, including software user interface (Ellis, 2009; Meyer, 2011), technical documentation (Meyer, 2011), transcripts of videos of an "inspirational" nature (Ted Open Translation Project, 2011; Meyer, 2011) and humanitarian aid content (Munro, 2010; Baer, 2010; Baer & Nagle, 2011; Translators without Border, 2008). It has also been used for large scale collection of linguistic data for research purposes

or machine translation training (for example, in Ambati & Vogel, 2010; Bloodgood & Callison-Burch, 2010).

- **Post-editing by the crowd:** systems allowing a large crowd of mostly amateurs to correct the output of machine translations systems, often with the aim of improving the system's accuracy. Examples of this include Asia Online's Wikipedia translation project (Asia Online, 2011), as well as features which allow anonymous users to correct the outputs produced by systems like Google Translate (Google translate, 2007) and Microsoft's Bing Translator (Microsoft Collaborative Translation Framework, 2011).

It is worth noting that the above flavors of collaborative translation are not mutually exclusive. In fact, it is quite common for an organization to leverage more than one of those approaches at the same time. For example, some translation crowdsourcing initiatives and online marketplaces also include features for sharing and collaborating on resources like terminology databases and translation memories. Post-editing by the crowd can also be used as part of a translation crowdsourcing initiative, and so on.

Also, many of these "new" approaches are in fact very similar to more conventional technologies that have existed for years, such as translation workflow systems, terminology databases and translation memories. All of those earlier technologies were already collaborative in that they allowed groups of customers, managers and translators, to coordinate their activities. In a sense, one might say that the collaborative translation revolution is not so much about introducing new technologies, as it is about using existing groupware technologies with much larger groups of people, where members of these communities know less about each other and have fewer a priori reasons for trusting each other.

2.2. Common issues in Collaborative Translation

As can be seen from the above list, collaborative translation represents a rich and complex envelope of processes and technologies. Determining which approach can be used in which context and tweaking it to meet one's goals is still somewhat of black art, and currently, trial and error is often the only way to find out.

To complicate things further, there are many poorly understood issues and open questions regarding the best way to deploy collaborative translation in specific contexts, and this comes out clearly from the proceedings of a recent workshop on that topic held at the 2010 conference of the Association for Machine Translation in the Americas (AMTA) (AMTA, 2010). Here is a sample of the most common themes we have encountered in five years of advocating collaborative translation to translators, service providers and technology vendors.

2.2.1. Business goals

Collaborative translation is not an end in itself, and it is only relevant to the extent that it can support business goals of an organisation. Different flavors of these technologies have completely different kinds of benefits, and it is important for practitioners to understand what they are likely to get from deploying them.

For example, agile translation teamware may be used as a channel for more varied and horizontal communication inside a translation team (Beninatto & De Palma, 2008; Calvert, 2008; Yahaya, 2008), in order to provide feedback loops that allow customers to communicate more rapidly with the translators who carry out the actual work, and to do so earlier in the project life-cycle. Collaboratively built terminology resources and shared translation memories may allow an organization to collect linguistic data from a wider range of contributors (for example, domain experts), and at lower costs. However, this may come at the cost of having less control over quality of the data or its specificity to one's own context and domain.

Online translation marketplaces may allow customers to recruit the best and most qualified translators for a given project and cut down on the intermediary costs, and conversely, it may allow specialized freelancers to market their skills to a wider and larger group of customers.

In the case of translation crowdsourcing, the benefit that most readily comes to mind is cost reduction, but organizations that have used this approach are quick to point out that it may not be the most important one. Other benefits which are commonly mentioned include: community involvement, increased brand loyalty, faster turnaround time, supporting the long-tails of low-density languages and transient content (e.g., user-generated content), and production of translations that are more in tune with the particular linguistic idiosyncrasies of the actual users of the content (Baer, 2010; Meyer, 2011).

2.2.2. Quality control

All flavors of collaborative translation are more grassroots and less tightly controlled from the top than is typically found in professional translation contexts. This is true even for collaborative approaches that are aimed specifically at professionals (e.g. agile translation teamware). Therefore, a common question is how quality control can still be exercised in such decentralised environments. The answer of course depends on the context, and the level of quality that is needed to meet the customer's business goals ("fit for purpose" quality).

In some cases, customers may want in-house professional translators to revise and approve each and every translation produced by members of the group or crowd, before they are actually published (Facebook

eventually opted for this approach after some embarrassing incidents). In other cases, it may be sufficient to ask translators to go through an initial screening test and then let them loose on the content. This approach is currently used by Translators without Borders (Translators without border, 2011; De Palma, 2011) and Kiva.org (Baer & Nagle, 2011). In other cases still, we may have the crowd or group itself carry out quality control, through mechanisms like voting, mutual revision and automated reputation management. This approach is used by Facebook (Ellis, 2009), as well as by research projects that are collecting parallel corpora (Ambati & Vogel, 2010; Bloodgood & Callison-Burch, 2010).

It is worth noting that by and large, there is an assumption in professional circles that decentralized, collaborative translation will lead to lesser quality output. While this may be true in some contexts, it is far from inevitable, and there may be situations where collaborative processes can in fact lead to higher quality, through appropriate leveraging of the so-called wisdom of crowds effect (Surowiecki, 2004). Unfortunately, although there are known principles for increasing the chances that a group will collectively act smarter than its individuals (namely, diversity, independence, and aggregation), the specifics of how they can be applied in given translation contexts is still somewhat of an open question.

2.2.3. Crowd motivation

A key ingredient in any collaborative translation initiative is the presence of a compelling incentive for members of the group or crowd to participate. Even with agile translation teamware where use of the system by employees or subcontractors is mandated from above, the approach still cannot be successful without minimum buy-in from translators.

Motivation issues are most critical in crowdsourcing scenarios, and this is possibly the main reason why it has yet to become widespread. The most successful cases have been in contexts where members of the crowd are emotionally invested in the content being translated. In the case of Facebook, this emotional bond came from the social nature of the application which people use to connect with friends, relatives and acquaintances. In the case of the Haiti earthquake relief initiative (Munro, 2010), it came from the diaspora's attachment to their native country, while for Kiva and Translators without Border, it may come from a perception that the organization is pursuing worthy humanitarian goals. Finally, in a case like TEDTalks (TED Open Translation Project, 2011), it comes from the compelling, high profile nature of the talks and speakers that people are translating. Although this has not been documented in writing, some of the practitioners who participated in the TAUS roundtable and the AMTA Collaborative Translation workshop (AMTA, 2010) mentioned that for some of the volunteer translators, pride in their native tongue was also an

important motivating factor. Surprisingly enough, even for-profit companies like Adobe and Symantec have found that some of their end-users are passionate enough about their products to volunteer time for translation of content (Meyer, 2011).

All of the motivators mentioned so far are, to a certain extent, altruistic and disinterested. But that is not always the case. Researchers involved in collection of linguistic data have also reported (although not in writing) that some of the volunteers were actually second language learners who saw translation as a good way to practice a language and get feedback about their production, while being of service to the research community. In a different context where Adobe was able to co-opt third party vendors into translating highly technical content, the motivation came from the fact that this particular content was critical to that third party's niche business (Meyer, 2011). Baer and Nagle also report that, in the context of Kiva.org, some of the amateur volunteer translators from third world countries were contributing in order to establish a good track record in translation, in the hope that it would eventually allow them to make a career of it.

Finally, there will be cases where money is the only realistic incentive, and an open question is the extent to which this type of situation will predominate or not in collaborative contexts (especially crowdsourcing). Another open question which is relevant in that type of situation, is how to determine a level of compensation that is high enough to motivate members of the crowd to participate and do a good job, but not so high that it interferes with intrinsic motivators (Mason & Watts, 2009; Rogstadius et al., 2011) or attracts people who are out to game the system (for example, by entering random text, or raw machine translation taken from free systems like Google or Bing).

2.2.4. Role of professionals

Not surprisingly, a pressing question with collaborative translation is the role of professionals in this brave new world. While some forms of collaborative translation (agile teamware and online marketplaces in particular) are designed for professionals, there are many flavors of translation crowdsourcing that seem to de-emphasize their role. But that need not be the case. For example, one might use a small "crowd" of paid professionals working in parallel on a large project, as a way to dramatically decrease lead time, while ensuring professional level quality (Beninatto & De Palma, 2008). Even in cases where the crowd consists mainly of unpaid (or low-paid) non-professional translators, professionals could still be used to revise and vet the final result. In that sense, crowdsourcing might allow professionals to delegate simple routine parts of the translation to the crowd, and focus their talent on more challenging aspects such as terminology, style and fluidity (Orr Priebe, 2009).

Crowdsourcing may also open new types of jobs for professionals, for example, community management and coaching. Finally, given that a frequent goal of crowdsourcing is to allow the translation of content that otherwise would not have been translated at all, and that some part of this work will have to be revised or facilitated by professionals, it may be that crowdsourcing will not so much decrease demand for professionals, as it will change the nature of their interventions.

2.2.5. Parallelism and de-contextualization

Most implementations of collaborative translation involve some level of parallelism, in other words, splitting the work into smaller chunks, and dispatching them to different members of a community. Parallelism can help by dramatically decreasing lead time, or by making it possible to recruit volunteers who might otherwise not be willing to commit to the translation of complete documents. The granularity of chunks may vary widely across applications, ranging from complete sections or documents, down to single sentences.

In spite of its advantages, parallelism presents some dangers in that it de-contextualizes translation, which in turn can affect quality and consistency of the end result. It can also decrease job satisfaction, by making it harder for the translator to see how his work is contributing to a meaningful complete picture. Fortunately, there are ways to offset these drawbacks, for example, revising documents in a more global, document wide fashion, or implementing back channels like chat rooms and discussion forums to allow translators to socialize and get a sense of the big picture (Munro, 2011). However, such measures can probably never completely resolve the tension between parallelism and de-contextualization.

3. Using design patterns to capture collaborative translation best practices

As can be seen from the previous section, collaborative translation is a complex space of possibilities, and it can be challenging to choose and fine-tune appropriate processes and tools to meet specific needs. While there are several case studies that describe how specific collaborative translation approaches have been successfully implemented in the context of specific organizations, sifting through those can be time consuming for a practitioner wanting to emulate the success of others. There is a clear need for a more compact, well-validated and easy to consult compendium of best practices in that area.

In this section, we argue that such a compendium could be written in the form of a collection of design patterns.

3.1. About design patterns

A design pattern is a formal way of documenting a common solution to a common problem in a particular field of expertise. The idea was introduced by Christopher Alexander in the field of architecture (“Pattern (architecture)”, 2011) and was subsequently adopted by other disciplines, including computer science (“Design pattern (computer science)”, 2011), user interface design (“Interaction design pattern”, 2011) and education (“Pedagogical patterns”, 2010). Patterns do not exist in isolation, and they are typically organized into collections of interlinked patterns for a given field. Such collections are referred to as pattern languages.

Pattern languages are a very flexible way of describing the solution space for complex domains. Instead of having to subscribe to all-encompassing, one-size fits all solution, practitioners can cherry pick the patterns that best represent their particular context, then adapt and recombine them in original ways to fit a given situation. Another advantage of patterns is that they provide practitioners with a common vocabulary for sharing and talking about solutions to problems in their domain. In software design for example, it is quite common for developers to refer to complex designs using standard pattern names like *Singleton*, *Observer* or *Adapter* (“Design pattern (computer science)”, 2011), without having to explain them from scratch.

Another advantage of design patterns is that they can evolve with time, as we learn more and more about their respective applicability in different contexts. Indeed, it is quite common for design patterns to be stored on wiki sites which are open to modification by the public. In fact, the world’s very first wiki site was deployed by Ward Cunningham for the very purpose of creating a collection of design patterns in the field of software engineering (“WikiWikiWeb”, 2011). Given that collaborative translation is still in its infancy, this ability for patterns to evolve seems like a highly desirable attribute.

Although the format and structure of a pattern is open-ended and can be tailored to the needs of particular domains, there are a number of characteristics which have been found useful across disciplines. These are described below.

Clear Name: Each pattern has a short name (2-5 words) which clearly communicates the pattern’s essence to the reader. Good names are important, because they provide the basic vocabulary that practitioners can use to discuss the solution space for problems in their domain. Coming up with a good name is often the most challenging part of pattern writing, and

the inability to do so is often a symptom that the author hasn't yet grasped the actual core of the solution he is trying to describe.

Context: Each pattern must describe the context in which it applies. For example, is this pattern applicable to collaborative translation at large, or is it only relevant in volunteer-based translation crowdsourcing situations? Without proper context, the reader cannot easily determine the applicability of the pattern to his current situation.

Problem Description: The pattern must describe the exact nature of the problem that it solves. A common approach is to describe the problem in terms of tensions between opposing forces. Something along the lines of: “*On the one hand, one would want X, but on the other hand, one would also want Y, and the two are partly incompatible for reasons A, B and C*”.

Solution: The pattern must describe the solution to the problem. Again, this is often framed in terms of a way to balance the opposing forces mentioned in the *Problem Description* section, so as to reach a sustainable equilibrium. The solution should be general enough to be applied in very different situations within its context, but still specific enough to give constructive guidance to the practitioner.

Links to related patterns: Patterns generally do not exist in isolation, and relate to other patterns in a given domain. These links can be made explicit through inline references in different sections of the pattern, but also in a separate *Related Patterns* section.

Real-life examples: Good patterns provide references to real-life examples where it has been shown to work. Such examples are important because they provide a sense of how well-trying the pattern actually is.

Figure 1 provides an example of a pattern for the domain of collaborative translation.

4. The Collaborative Translation Patterns Repository

In order to kick-start the creation of a collection of patterns for collaborative translation, TAUS organized a one-day workshop which was held at Localization World 2011 in Santa Clara, on October 10th, 2011. The workshop brought together 12 practitioners, which included seasoned users of collaborative translation from organizations like Adobe, Symantec, Kiva and Worldwide Lexicon. The list of participants also included practitioners who did not have hands-on experience with these approaches, but were seriously considering them. After short presentations by the experienced users, participants brainstormed a list of best-practices that were mentioned in one or more of them. Each best-practice was given a clear, communicative name, as well as a short description of what it entails. We then collaboratively tried to organize those practices into groups of related themes.

Publish Contributions Rapidly

Context
This pattern is useful for motivating contributors in any collaborative translation context, but it is particularly useful in translation crowdsourcing scenarios.

Problem description
Contributors are often motivated by a desire to have a positive impact on the community they are participating in. However, they cannot achieve this sense of being useful, if their contributions do not become available to the rest of the community in a reasonable amount of time.

Solution
Therefore, minimize the delay between the moment when a member of the community contributes to the site, and the moment where it becomes publicly available to the rest of the community.
Ideally, the contribution should become visible to the rest of the community as soon as the user clicks on the *Save* button. This "ideal" may not always be achievable, for example in situations where some level of quality control must be done before publication. But even in those situations, you may want to consider a Publish then Revise approach rather than the more conventional revise then publish.

Links to related patterns

- Point System is another way for a contributor to get a sense of how useful he has been to the community.
- Campaign Progress Gauge is another practice which allows members of the community to see the positive impact of their actions. The main difference is that it operates more at a community/project level rather than at an individual/contribution level.

Real-life examples

- At Facebook, translations become available in a matter of hours.
- In the context of software localization by the crowd, Adobe makes a conscious effort to wrap the community's translations into every new releases of the product.

Figure 1: Example of a design pattern for collaborative translation. Words that are underlined are references to other patterns in the same domain.

4.1. Initial set of best practices

This exercise resulted in an initial set of 53 best-practices, organized into 6 themes. They are listed below. We do not have sufficient room in this paper to discuss each practice in detail, but the name is usually sufficient to provide the gist of what it pertains to. Interested readers can get more detail about specific practices at www.collaborative-translation-patterns.com.

4.1.1. Planning and scoping

This theme contains practices that come to play before a collaborative translation community or project is actually started. It currently includes the following practices:

- Align Stakeholder Expectations
- Early Clarification of Translator Expectations
- Backup Plan
- Project Check Points
- Appoint Initial Community Manager

4.1.2. Community motivation

This theme contains practices that help with the recruitment, retention, and motivation of members of the translator community. It currently includes the following practices:

- Publish Contributions Rapidly
- Campaign Progress Gauge
- Contributor Recognition
- Leader Board
- Official Certificate
- Point System
- Offer Double Points
- Give Unique Branded Products
- Contribution Loop
- Contributor of the Month
- Grant Special Access Rights
- Rewards for Contribution
- Campaign

4.1.3. Quality

This theme contains practices which help ensure that translations produced by the community are of an appropriate level of quality. It currently includes the following practices:

- Publish then Revise
- Content-Specific Testing
- Entry Exam
- Peer Review
- Automatic Reputation Management
- Random Spot-Checking
- Revision Crowdsourcing
- Users as Translators
- Voting
- Transparent Quality Level

4.1.4. Contributor career path

This theme contains practices which allow people in the community to grow into different roles and participate meaningfully and to the best of their ability and availability. It currently includes the following practices:

- Flexible Contributor Career Path
- Lurker to Contributor Transition
- Anonymous Translation
- Find the Leaders
- Support Variable Levels of Involvement
- Community Manager
- Content Prioritizer

4.1.5. Right-sizing

This theme contains practices which are used for splitting large communities or big chunks of work into smaller and more manageable ones. It currently includes the following practices:

- Contributor-Appropriate Chunk Size
- Community-Appropriate Project Size
- Break Up Crowd Into Teams
- Require Minimal Involvement Level
- Keep the Crowd Small
- Volunteer Team Leaders

4.1.6. Tools and processes

This theme contains practices that specify important characteristics of the online tools used by the community, as well as the processes and workflows that they implement. It currently includes the following practices:

- Hint at Content Priority
- First In, First Out
- Task Self Selection
- Layered Fallbacks
- Official Linguistic Resources
- Automatic Equivalents Suggestions
- Provide Context
- In-Place Translation
- On-the-Side Translation
- Community Forum
- Analytics for Content Priorization
- Simplicity First

4.2. Interesting trends

Looking at the set of initial best-practices mentioned in Section 4.1, we can see a number of interesting trends.

Firstly, it is worth noting that most of the practices relate to the context of translation crowdsourcing. This is not surprising, given that this is the flavor of collaborative translation that seems to have captured the imagination of more people in the field. However, it does point to the need for more exploration of other collaborative modalities in translation.

Secondly, most of the practices are not specific to translation per se. To be sure, some practices like *Users as Translators* (the idea that end users of a particular piece of software or web site may be uniquely qualified to translate its user interface) are only applicable in a translation context. But other practices, like *Voting* (inviting members of the crowd to vote on the quality of content produced by other members of the crowd), are useful for any kind of crowdsourcing effort, and the majority of the patterns we list above seem to fall in that category. This raises the following question: “*Are the best practices for translation crowdsourcing essentially the same as those for crowdsourcing in general, or are there some unique problems in translation that call for unique solutions?*”. The same question could be asked for other flavors of collaborative translation. For example, are the best-practices needed in the context of building collaborative terminology resources different from the ones which have been used for some years to build collaborative knowledge sources like Wikipedia and Wiktionary?

One point which is not apparent from the list of best-practice, but which came out clearly in the discussion at the roundtable, is that *Community Motivation* issues seem the most difficult ones to resolve in a translation crowdsourcing context. In contrast, there was a sense that *Quality Control* issues tend to resolve themselves, provided that enough of the “right” people can be enticed to participate and that you provide them with lightweight tools and processes by which they can spot and fix errors.

Another interesting point which is not directly apparent from the above list is that different kinds of organizations seem to use different best-practices. For example, when it comes to vetting translators, Adobe and Symantec (two software vendors who use crowdsourcing for translating material related to their products) employ similar, fairly open practices. In contrast, Kiva and Translators without Borders (two not-for-profit humanitarian organizations) use a more closed approach that requires contributors to pass an *Entry Exam*. Liz Nagle of Kiva explains this difference by the fact that translation is core to their operation. Indeed, without translation Kiva cannot achieve its mission of facilitating microloans, because loan applications are usually written in the native tongue of the applicant (often small density languages) which usually differs from the language of their donor population (mostly English).

The fact that different types of organizations need different kinds of practices is a strong argument in favor of a patterns-based approach, because it allows practitioners to cherry-pick and adapt those practices that seem most applicable to their situation.

5. Conclusion

Collaborative translation has the potential for significantly changing how we translate content. However, it presents potential adopters with a complex envelope of possible approaches which can be hard to navigate. There is a clear need for a concise body of knowledge that summarises current best practices in that domain.

We believe that a collection of design patterns is a good way to achieve that, and we have presented a first attempt at generating such a repository. The result of this effort now resides on a wiki site which is open for editing and commenting by people in the community (collaborative-translation-patterns.com).

While this site is a good start, it raises some interesting questions. For one thing, most of the practices that have been documented so far in this repository are not that different from best-practices which are being used for crowdsourcing in other domains. It is worth asking whether there is anything particular to translation which requires our field to come up with its own set of best-practices, or do we simply need to learn more about practices for crowdsourcing in general? Also, the repository as it currently

stands, focuses mainly on translation crowdsourcing, which is only one of many possible flavors of collaborative translation. It would be interesting to try and better document practices for other flavors as well.

It is our hope that this site will continue to grow and be improved, as we learn more and more about good ways to implement collaborative translation applications that work and are acceptable to all parties involved, and we encourage the reader to contribute to it.

Acknowledgements

The authors are indebted to Rahzeb Choudhoury (TAUS) who helped organize the roundtable which served as the starting point for the collaborative-translation-patterns.com repository. Also, to the participants of the roundtable who shared their own experience and best-practices with us: Dirk Meyer (Adobe), Jason Rickard (Symantec), Liz Nagle (Kiva), Terena Bell (In Every Language), Renato Benninato, David Canek (MemSource), Christopher Klapp (Firma8), Brian McConnell (Worldwide Lexicon) and Laura van Nigtevegt (Spil Games).

References

- Ambati, V., Vogel, S. (2010). Can crowds build parallel corpora for machine translation systems? *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's Mechanical Turk*, Los Angeles, California, June 2010.
- AMTA (2010). *Collaborative translation: technology, crowdsourcing, and the translator perspective*. AMTA 2010, Oct 31, 2010, Denver, Colorado, USA. Retrieved from <http://www.wiki-translation.com/AMTA+2010+Workshop>.
- Asia Online. (2011, October 17). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Asia_Online&oldid=456007085#Portal_Initiatives
- ATA (2011). Press release, Head of largest professional translators' organization blasts LinkedIn CEO for "thoroughly unprofessional practices". June 30, 2009. Retrieved from http://www.atanet.org/pressroom/linkedin_2009.pdf
- Baer, N. (2010). Crowdsourcing: Outrage or opportunity? *Translorial, Journal of the Northern California Translators Association*, February 2010.
- Baer, N. (2010). *Trends in crowdsourcing: Case studies from not-for-profit and for-profit organisations*. ATA 2010, Oct 27-30, 2010, Denver, Colorado, USA. Retrieved from http://www.wiki-translation.com/tiki-download_wiki_attachment.php?attId=62.
- Baer, N., & Nagle, L. (2011). *Scaling and strengthening a closed crowd: Lessons from Kiva's evolving model*. TAUS Roundtable on Collaborative Translation, Santa

- Clara, Oct 10, 2011. Retrieved from http://collaborative-translation-patterns.com/tiki-download_file.php?fileId=2
- Beninato, R., & De Palma, D. (2008). Collaborative translation. *Multilingual*, 2008 Resource directory & index 2008. Retrieved from <http://www.scribd.com/doc/4069269/Structuring-Collaborative-Translation-20-Less-Delivery-Time-Better-Quality>)
- Bloodgood, M., & Callison-Burch, C. (2010). Using Mechanical Turk to build machine translation evaluation set. *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's Mechanical Turk*, Los Angeles, California, June 2010.
- Calvert, D. (2008). Wiki behind the firewall – Microscale online collaboration in a translation agency. *ASLIB Translating and the Computer* 30, Nov 27-28, 2008, London, UK.
- De Palma, D. (2011). Translators without borders extends its reach. *Common Sense Advisory Blogs*. Retrieved from <http://www.common senseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=691&moduleId=391>
- Design pattern (computer science). (2011, October 17). In *Wikipedia, The Free Encyclopedia*. Retrieved from [http://en.wikipedia.org/w/index.php?title=Design_pattern_\(computer_science\)&oldid=456093478](http://en.wikipedia.org/w/index.php?title=Design_pattern_(computer_science)&oldid=456093478)
- Ellis, D. (2009). A case study in community-driven translation of a fast-changing website. In *Internationalization, Design and Global Development Lecture Notes in Computer Science*, 2009, Volume 5623/2009, 236-244, DOI: 10.1007/978-3-642-02767-3_26.
- Goddard, P. (2010). The power of ProZ. *ITI Bulletin*, JanFeb 2010.
- Google translate. (2007, December 1). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Google_translate&oldid=175043259#Features_and_limitations
- Google Translator Toolkit. (2011, August 30). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Google_Translator_Toolkit&oldid=447439495
- Interaction design pattern. (2011, July 16). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Interaction_design_pattern&oldid=439853212.
- Kudoz, (2011, October 17). *Kudoz term questions – Translators helping translators*. Retrieved from <http://www.proz.com/kudoz/>.
- Mason, W., & Watts, D. J. (2009). *Financial incentives and the “performance of crowds”*. KDD-HCOMP '09, Paris, France, June 28, 2009.
- Meyer, D. (2011). *Community translation: Leveraging communities and crowds*. TAUS Roundtable on Collaborative Translation, Santa Clara, California, USA, Oct 10, 2011. Retrieved from http://collaborative-translation-patterns.com/tiki-download_file.php?fileId=2.
- Microsoft Collaborative Translation Framework (2011, October 17). *Collaborative translations: Announcing the next version of Microsoft Translator technology – V2 APIs and widget*. Retrieved from <http://blogs.msdn.com/b/>

- translation/archive/2010/03/15/collaborative-translations-announcing-the-next-version-of-microsoft-translator-technology-v2-apis-and-widget.aspx.
- Munro, R. (2010). *Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge*. In *Proceedings of Collaborative Translation: technology, crowdsourcing, and the translator perspective*, AMTA 2010, Oct 31, 2010, Denver, Colorado, USA. Retrieved from http://www.wiki-translation.com/tiki-download_wiki_attachment.php?attId=63)
- MyMemory. (2011, June 19). In *Wikipedia, The Free Encyclopedia*. Retrieved from <http://en.wikipedia.org/w/index.php?title=MyMemory&oldid=435180033>.
- Orr Priebe, S. (2009). *Tom Sawyer – A crowdsourcing pioneer?* TCWorld, November 2009. <http://www.tcworld.info/tcworld/outsourcing/article/tom-sawyer-a-crowdsourcing-pioneer/>
- Pattern (architecture). (2011, May 21). In *Wikipedia, The Free Encyclopedia*. Retrieved from [http://en.wikipedia.org/w/index.php?title=Pattern_\(architecture\)&oldid=430249377](http://en.wikipedia.org/w/index.php?title=Pattern_(architecture)&oldid=430249377).
- Pedagogical patterns. (2010, December 16). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Pedagogical_patterns&oldid=402717235
- ProZ.com. (2011, September 5). In *Wikipedia, The Free Encyclopedia*. Retrieved from <http://en.wikipedia.org/w/index.php?title=ProZ.com&oldid=448648429>.
- Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. *Proceedings Of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, July 17-21, 2011.
- Stejskal, J. (2009). Ask not what the crowd can do for you, but what you can do for the crowd. *The ATA Chronicle*, August, 2009.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Random House, NYC, NY, USA.
- Translated.net (2011, October 17). *Fast and easy professional translation services*. Retrieved from <http://www.translated.net/>.
- TranslatorCafe (2011, October 17). *TranslatorCafe.com – A place for translators, interpreters and translation agencies*. Retrieved from <http://www.translatorscafe.com/>.
- Translators without borders (2008, December 20). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Translators_without_borders&oldid=259138738
- Urban Dictionary (2011, October 17). *Urban dictionary*. Retrieved from <http://www.urbandictionary.com/>.
- TAUS Data Association (2011, October 17). *TAUS Data Association*. Retrieved from <http://www.tausdata.org/>.
- TED Open Translation Project (2011, October 17). *TED open translation project*. Retrieved from <http://www.ted.com/index.php/OpenTranslationProject/>.
- Wiktionary (2011, October 17). *A multilingual free encyclopedia*. Retrieved from <http://www.wiktionary.org/>.

- WikiWikiWeb. (2011, August 16). In *Wikipedia, The Free Encyclopedia*. Retrieved from <http://en.wikipedia.org/w/index.php?title=WikiWikiWeb&oldid=445126672>
- Yahaya, F. (2008). Managing complex translation projects through virtual spaces: A case study. *ASLIB Translating and the Computer 30*, Nov 27-28, 2008, London, UK.