MNH-TT: a collaborative platform for translator training

Bogdan Babych[*], Anthony Hartley[†], Kyo Kageura[‡], Martin Thomas[*], Masao Utiyama[◊]

[*]U of Leeds - UK, [†]Tokyo U of Foreign Studies, [‡]U of Tokyo, [◊]NICT – Japan

b.babych@leeds.ac.uk, thartley@tufs.ac.jp, kyo@p.u-tokyo.ac.jp,
m.thomas@leeds.ac.uk, mutiyama@nict.jp

Abstract

Increasingly, both commercial and non-commercial translation rely on highly collaborative activity. Thus, we contend, students aiming for a career in translation gain from early exposure to such a working model. On analysing a range of commercial and not-for-profit translation platforms, we identified, from our social-constructivist pedagogical perspective, a major defect. By allowing neither for preserving a trace of interactions nor for relating these interactions to the intermediate products generated during the workflow, they deny participants the chance to later reflect on them.

To scaffold the trainee experience, we therefore extended an existing platform – *Minna no Hon'yaku (Translation of/by/for All)* – with four functions. First, each participant is assigned one or more roles which map into various workflows. Second, communication between role-players is structured by a menu of dialogue acts, each act linked to an entity in the translation project. Third, a menu of revision categories is used to motivate individual edits. Fourth, these events are recorded using an extended TMX notation and can be visualised graphically via a dashboard to answer such questions as: Where within the workflow are the peaks in interaction? Do these correlate with significant modifications to the translation product?

1 Background

Increasingly, both commercial and non-commercial translation – including that undertaken by fan communities and for NPOs and NGOs – rely on highly collaborative activity (for an overview, see (O'Hagan, 2011)). Thus, we contend, students aiming for a career in translation will gain from early exposure to such a working model. Moreover, Kiraly (2000), in advocating a social constructivist approach to translator education, argues persuasively for the broader intellectual and ethical benefits of a pedagogy that inducts its participants into a community of practice. His emphasis is on authentic translation activities 'with an identifiable user and readership, with task specifications from employer or client' (Kiraly, 2000:52). As we show in the next section, volunteer translation sites offer unprecedented opportunities for trainee translators to engage in such authentic activities, subject to provisos on the quality of their contributions. Quality assurance concerns can be addressed by collaborative checking and revision, which, as Pym reminds us (2009:136), is hardly new in the translation classroom. Desjardins (2011) advocates the use of general-purpose social networking sites as a powerful driver of translator training, with their potential for involving more participants. However, our aim is to provide an online environment adapted specifically to the needs of trainee translators.

Our starting point is, therefore, to briefly review the nature of collaborative translation activities currently and the environments, tools and resources that exist to support them.

1.1 Collaborative translation

We use 'collaborative translation' as an umbrella term to describe multi-participant, distributed translation activities that rely crucially on social networks. As a term with no necessary connotations of commercial activity, it is also well suited to describing the pedagogical applications we have in mind in the present article. For a discussion of the sometimes subtle differences between this term and a number of related expressions in common usage (such as 'community translation', 'fan translation', 'user-generated translation', 'volunteer translation' and 'crowdsourcing'), see (O'Hagan, 2011).

We acknowledge two important features of collaboration on translation: motivation and infrastructure (the 'platform' on which the collaboration is built). While the principal motivation of commercial translation is financial reward, some community and volunteer translation may also attract monetary rewards. Another powerful motivator is the acquisition of kudos or status within an existing community, gained by enhancing existing resources, for example, the localization of Facebook or Twitter by members, or the translation of Adobe software documentation by users. More philanthropic motivations account for the huge mobilization of volunteer translators in response to the earthquakes that struck Haiti in 2010 and Japan in 2011. The role of the online community is clearly central, and we agree with O'Hagan in seeing 'this social dimension [as] a critical motivator' in itself (2011:14).

Other important features of such communities are decentralization and self-organisation. There is, however, as much variation in the degree to which different self-styled collaborative translation environments support these features as there is in the range of functions and tools that these same environments offer. When attempting to create a collaborative environment specifically for trainee translators, we must consider what features are necessary and appropriate for catering to their needs.

1.2 Collaborative translation platforms

While it is not easy to classify translation-related tools, services, and functionalities available online (both commercial and non-commercial), we can recognise three broad types which, of course, are not mutually exclusive: aids for translation proper, tools and services intended for translation management, tools and services intended for community building.

There are many commercial and non-commercial translation-aid tools, of which we mention only a small sample. Déja Vu and memoQ are well known commercial stand-alone or server systems, while Omega-T is an open-source translation-aid. Lingotek provides translation-aid functions online on a commercial basis. Google Translator Toolkit is free, and Minna no Hon'yaku ('Translation of/by/for All' – MNH) is available on a non-commercial basis. The functions typically offered are translation memory, terminology management, and possibly connection to machine translation (MT).

Commercial providers of project management tools include LTC (LTC Worx), ProZ and acrolinx. Globalsight is one of the best-known open-source tools for translation management. The functions typically offered cover the management of the translation workflow, documents and graphics, scheduling, translators, quality assurance, payment, quoting and invoicing.

Wiki-based systems and services such as Traduwiki and Wikitranslation provide community building platform, although they do not appear to enjoy much popularity. However, there have been several successful translation projects which have taken advantage of community-based translations. Most of these are embedded in communities which are consolidated around themes and topics, rather than independent translation activities, although the host community may have devoted time and resources to building a suitable translation support environment. Facebook and TED's 'Ideas Worth Spreading' site are probably the best-known examples. Fan subbing and fan translation (of manga, for example) also belong to this category.

There are several socially-oriented projects involved in promoting translation by volunteers. Translators without Borders (TWB), founded in 1993, supports translation needs by such NGOs as Doctors without Borders, Médecins du Monde, Action Against Hunger, Oxfam US and Handicap International. The Rosetta Foundation, founded in 2009, aims to 'relieve poverty, support healthcare, develop education and promote justice through access to information and knowledge across the languages of the world'. Both focus on matching translation needs with translators. Minna no Hon'yaku started in 2009 with the aim of helping translators working at NGOs and other volunteer translators working online. Its focus is on translation aids, in particular rapid access to dictionaries and term lists and integrated internet search.

An analysis of the above-mentioned and other platforms reveals four broad sets of functions to support users. Communicative functions include messaging, discussion boards, voting and editing. Organisational functions include the allocation of tasks and process tracking. Motivational functions include identification of individual translators, contribution tracking and award systems. Resource functions include accommodating shared style guides, glossaries and translation memories. Most platforms assume that users are experienced, although some offer guidance on how to tackle a translation task and prescriptions on workflow.

1.3 Gaps in support for trainee translators

As we hinted at the very beginning of Section 1, we reject the conception of university translator training as an individual exercise in transforming text from one language to another, with ephemeral, disposable outputs. In contrast, we construe translation as a collaboration between aspiring specialists in the source-language culture and subject domain, terminologists, translators and revisers. Moreover, we consider that engagement with commissioners and users of the translation motivates communicative effectiveness rather than mere linguistic correspondence. The outputs – intermediate as well as final – are, in our view, a potentially valuable resource for other translators and translator trainers.

Despite the fact that this training scenario closely resembles 'real world' translation, we concluded that the list of existing support functions identified in Section 1.2 is lacking in two

major respects, which are to do with the tracking of product and of process. First, they afford no basis for scaffolding or tracking the interactions between participants to reveal the nature of their networking. And second, they provide no means of relating these interactions to the intermediate products generated at the translation, revision and reviewing stages.

Why are these functions important for translator training? Within the framework advocated by Kiraly, Pym and ourselves, reflection – both individual and collective – on the parts played by the various participants in the success (or failure) of a collaborative project becomes an essential part of the learning process. The potential for 'self-discovery' by students through informal experimentation with different translation processes is forcefully demonstrated by Pym (2009). The tasks set by Pym (post-editing of machine translation output, completion within strict time constraints …) are intended to be thought-provoking for the individual student rather than material for rigorous scientific investigation. However, unless a trace of the decision-making process is preserved, learning relies essentially on introspection. In the case of collaborative translation, an external representation of the process and its (intermediate) products is a pre-requisite for fruitful, post-hoc discussion among the participants. Such a representation is also necessary where a teacher is expected to intervene – either as a 'consultant' as advocated by Kiraly (2000), or in a more conventional, directive mode. Once the representation exists, it can be made more widely available to observers, such as other learners or researchers into the dynamics of collaborative translation.

Thus, instead of treating student translations as ephemeral, disposable objects and eternally re-enacting dialogues about content and form with each new cohort of trainees, we attempt to enable learners to 'replay' and learn from their own experiences and those of their predecessors. This same data, we hope, can also afford translation studies researchers insights into the learning process.

To be readily exploitable, this data needs to be structured before it is preserved. At the same time, students' collaborative activities benefit from 'scaffolding', that is, support provided by teachers – or, in this case, by the environment itself – to help student learners to 'build viable roads of their own' (Kiraly, 2000:46). In the collaborative translation environment we have created – Minna no Hon'yaku for Translator Training (MNH-TT) – the pedagogical scaffolding serves also to structure and ultimately visualise the interactions between participants and the evolving collaboratively produced translation, such that the analysis of these relationships becomes tractable.

The scaffolding has four pillars. First, one or more roles (translator, reviser, …) which map into common translation workflows is assigned to each participant (Section 2.2). Second, a menu of dialogue acts (information-offer, status-check, …) serves to structure communication between role-players, each act being explicitly linked to an object in the translation domain or to a text span (Section 2.3). Third, a menu of revision categories is used to motivate and justify individual edits (Section 2.4), which are effected in a three-pane editor (Section 2.5). Fourth, all of these events are recorded and populate a searchable database defined in XML (Section 2.6). Thus, in simple terms, we preserve a record of 'who said what to whom (and in which capacities) about what at which stage' and 'who changed what and why'. Finally, we are

prototyping a dashboard where users can visualise collections or sequences of events likely to reveal patterns of behaviour on the part of the collaborators (Section 2.7).

2 Minna no Hon'yaku for Translator Training (MNH-TT)

Rather than developing a system which would incorporate all the functions already described, we decided to extend MNH (Utiyama et al., 2009a; 2009b) with the scaffolding and tracking functions we have outlined.

We believed MNH to be suitably adaptable for a number of reasons. MNH offers functions essential for translation training. On the one hand, it provides functions necessary for translation per se; thus its coverage of core functions allows users to 'graduate' easily to other, professional systems. On the other hand, it is designed for non-professionals and is relatively simple to use, which fits nicely with our immediate objective of implementing a system for trainees. Moreover, in comparison with many other translation-aid platforms, it is generous in its provision of reference resources, including online dictionaries. We have been able to confirm these beliefs in the course of translator training at Kobe Women's University and a collaborative project between the University of Leeds and Kobe City University of Foreign Studies (Clark, 2010).

From a technical-administrative perspective, MNH works in client-server mode, so user groups can avoid dealing with system management.

2.1 Basic MNH platform

MNH-TT is based on the online translation hosting and translation aid platform Minna no Hon'yaku, which was developed by our team. Development started in 2005 and MNH became publicly accessible at http://trans-aid.jp/ in April 2009.

MNH consists of three functional parts: a translator platform; a translation-aid editor, QRedit (Figure 1); and a translation document portal.

The translator platform provides a set of functions which enable users to manage their translations, collaborate with other translators, and improve the efficiency of translating. These include:

- a blog-style registration and management of translation documents
- registration and management of terminologies
- communication functions, including messaging, bulletin board, question and answer, and translation request functions
- collaborative translation mechanisms, including a co-editing mechanism, the definition and management of translator groups and projects and the sharing of terminologies
- TM plus MT for selected languages, including Japanese<>English (due Q1 2013).

In addition, it provides functions to save successive translation versions (typically up to 10) and display a pair of any two versions side-by-side with differences highlighted.

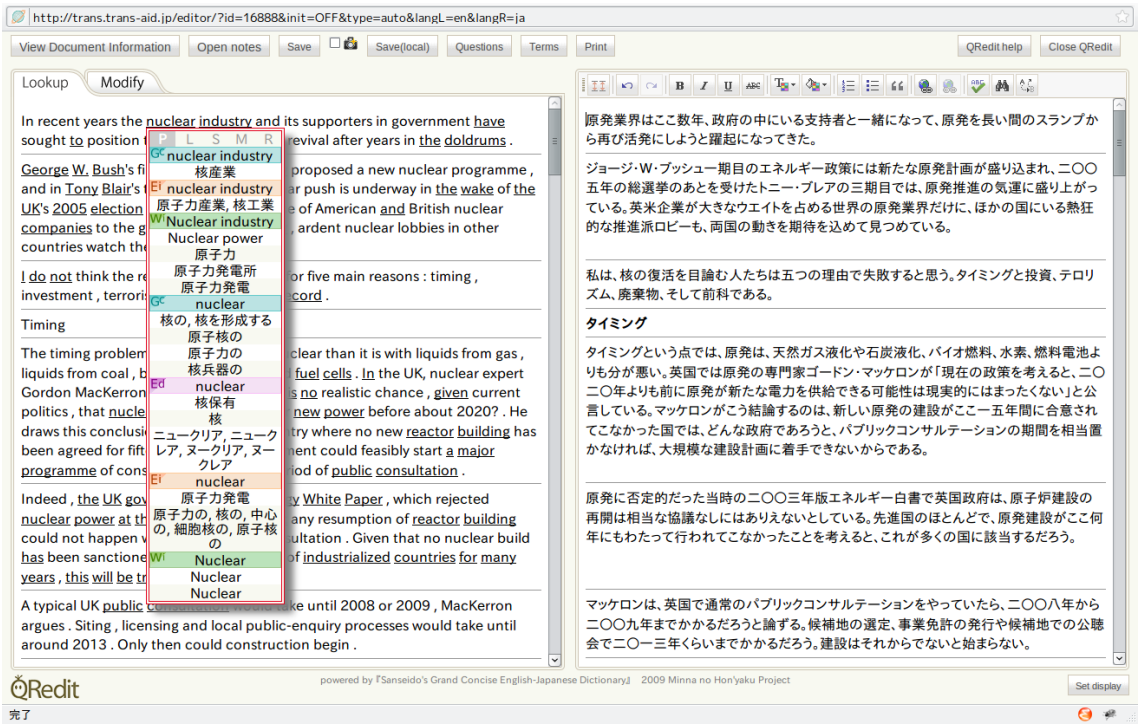

Figure 1: *Dictionary lookup in QRedit*

MNH assumes translators work voluntarily and not upon request by customers, so the management of overall workflow in commercial settings is not explicitly provided, although it can be simulated using existing MNH functions.

Translation work is carried out using the translation-aid editor QRedit, a two-pane editor which provides the following functions (Abekawa and Kageura, 2007):

- lookup of dictionaries and terminologies (idiom variants are matched to their canonical forms; multi-word units and idioms are prioritised)
- seamless connection to bilingual corpora (TMs)
- seamless connection to Wikipedia monolingual and bilingual entries
- seamless connection to Google webpage and dictionary search
- registration of terms.

These functions are triggered by mouse actions starting from the relevant words or phrases in the SL text area. Throughout these actions, the keyboard remains active in the TL text area in order to improve the efficiency of translation (Figure 1).

The three parts are modularised and basically implemented separately. QRedit is implemented using Java, the translator platform using php. The document portal is a wrapper. While the code base is not made public, modules – including QRedit – can be supplied upon request.

Anyone can sign up for MNH anonymously, make translations, and, if they wish and if copyright permits, publish them on the MNH document portal. Originally supporting only English-Japanese translation, MNH currently supports Japanese-English, English-Chinese, Chinese-English, Chinese-Japanese, Japanese-Chinese, English-Catalan and Japanese-German. Although most functions provided by MNH are language independent, MNH only supports language pairs for which dictionaries can be made available.

As of 7 November 2012, there were 2,276 registered users and 11,175 translated documents, of which 3,802 had been made public. The majority of translation work carried out using MNH is English to Japanese, with some Japanese to English and Japanese to Chinese translation.

Our team has extended and modified MNH, and two other sibling services are currently provided, i.e. Ryugakusei Net @ Minna no Hon'yaku and Kotoba no Volunteer @ Minna no Hon'yaku. The former is a crowd-based commercial translation service on which non-Japanese students work as translators; the latter is a community-site for accumulating multilingual versions of phrases important for facilitating communication in disaster situations. Details of these systems are reported by Kageura et al. (2011).

2.2 MNH-TT: roles

The first extension to MNH enables each participant to be assigned one or more roles based on the EN15038 standard for certifying the quality of translation services (Table 1). These roles map into common translation workflows.

The current set of roles allows for the configuration of workflows going from the most simple (e.g., translation by an individual of a single document belonging to any genre) to rather complex (e.g., team translation of a website into several target languages using both MT and TM). While the set of roles is more extensive than that identified by Arango-Keth and Koby (2003) as common in training settings, it does not include roles such as software-engineer, which would be required for the localization of software interfaces.

The core roles are: researcher, terminologist, translator, reviser, reviewer, proof-reader. Even if, in current pedagogical practice, all these roles tend to be held by a single individual, it is sensible to differentiate them explicitly in order to reflect both their conceptual distinctiveness and professional practice which assigns them to different individuals.

Table 1: ***Roles available in MNH-TT***

Role	Description
requester	person who commissions and signs off the translation job
project-manager	person who agrees the translation brief with the requester and manages all steps in the translation job up to and including delivery of a satisfactory product
researcher	person who identifies and collates information and documents relevant to understanding the source text and formulating an appropriate target document
terminologist	person who compiles a glossary of validated terms and expressions used in the source document and their equivalents in one or more target languages

Role	Description
translator	person who renders a target document that meets the requirements of the translation brief
reviser	person with translation competence who checks and corrects as necessary a draft target document produced by a translator
reviewer	person with domain expertise in the target language (but not necessarily in the source language)
proof-reader	person with target language expertise who checks the formal properties – language, format, completeness – of the target document
tm-manager	person who assigns the translation memory and glossary appropriate for the source document and updates the memory once the translation job is completed
mt-manager	person who configures appropriately the machine translation system, including glossary, generates a raw target document and updates the resources once the translation job is completed
mt-pre-editor	person who prepares the source document for processing by a machine translation or translation memory system by normalising format, punctuation, spelling or wording
mt-post-editor	person with translation competence who checks and corrects as necessary a raw target document produced by a machine translation system

Figure 2 shows the MNH-TT interface for assigning roles to individual team members.

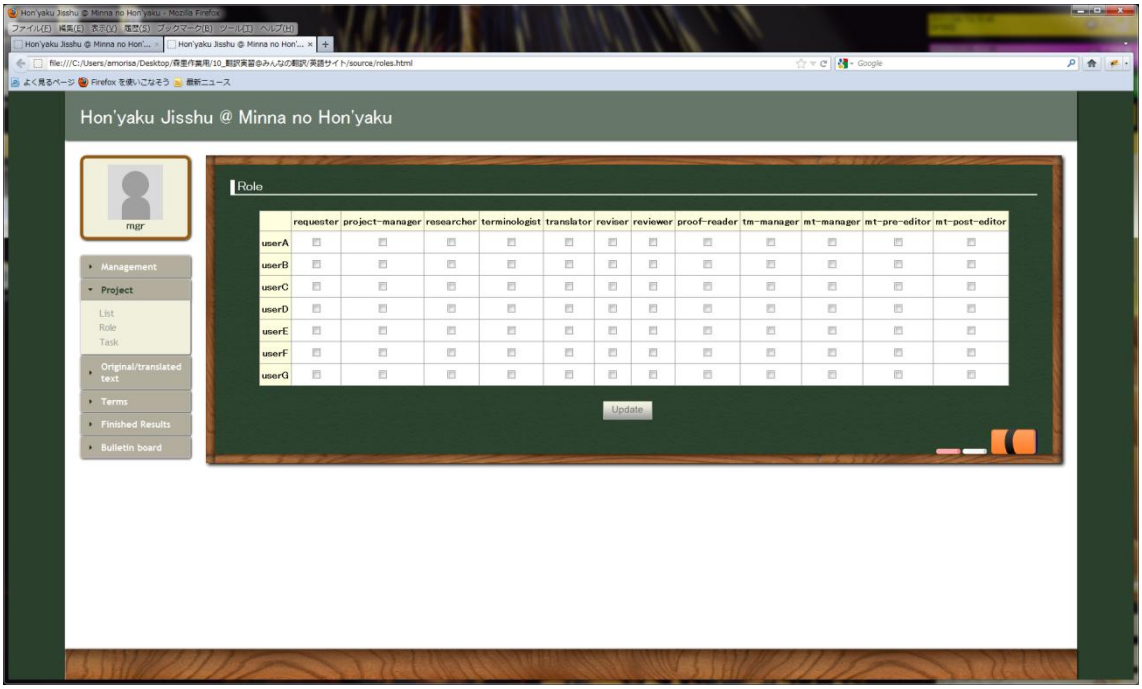

Figure 2: *Assigning roles in MNH-TT*

2.3 MNH-TT: dialogue acts

The second extension was to structure communication between role-players by a pre-defined set of dialogue acts (Table 2). These are based on research into task-oriented dialogues and structuring interactions within an intelligent tutoring system (Allen and Core, 1997; Soller, 2001). Each act can be explicitly related to an object in the translation domain or to a text span (known collectively as 'props'). This encourages trainees to respect protocols for acknowledging receipt of messages and data; it also attempts to avoid the uncertainty that sometimes arises in cross-cultural communication when, for example, requests for action or information are not explicitly signalled. The defined acts are grouped thematically.

- *information*- acts bear on the provision and discussion of information related to the substance of the translation
- *maintenance*- acts maintain the flow of dialogue and promote cohesion within the team
- *status*- acts bear on the progress of the work schedule
- *role*- acts relate to the 'staffing' of the translation team.

Table 2: ***Dialogue acts available in MNH-TT***

Dialogue Acts	Description
information-request	requester asks one or more addressees (possibly open call) for information about the substance of a translation (typically a WH question)
information-offer	provider supplies translation-related information, either in response to a request or spontaneously
information-support	supporter corroborates translation-related information or solution provided by another
information-differ	dissenter proposes contrary translation-related information or an alternative solution to that offered by another
maintenance-acknowledge	acknowledger signals receipt and/or appreciation of others' contributions to a task
maintenance-motivate	motivator provides positive feedback on others' work to support cohesion and involvement
maintenance-clarify	requester asks for clarification of an offer of services or information provided by another
maintenance-mediate	requester recommends intervention of instructor to answer a question or resolve an issue
status-check	checker asks about progress on a specified task
status-report	reporter responds to a check on progress on a specified task, or reports spontaneously
role-request	requester asks one or more addressees (possibly open call) to commit to a specified role or task (typically YES/NO question)
role-offer	offerer spontaneously volunteers services in performing a specified role or task
role-decline	decliner rejects a request or spontaneous offer to perform a specified role or task
role-accept	accepter commits to a request to perform a specified role or task, or takes up a spontaneous offer of services

Acts are represented as a menu. Each time a message is posted, the poster chooses the appropriate act type to classify the contribution. The poster may also specify their role and the object or 'prop' to which they are referring; these include: translation-brief, set-of-target-documents, research-data, glossary, tms, mt-raw-output, as well as text spans such as sentence or word. Figure 3 illustrates the bulletin board displaying a number of interactions between a project manager and volunteers signing up to play various roles in the project.

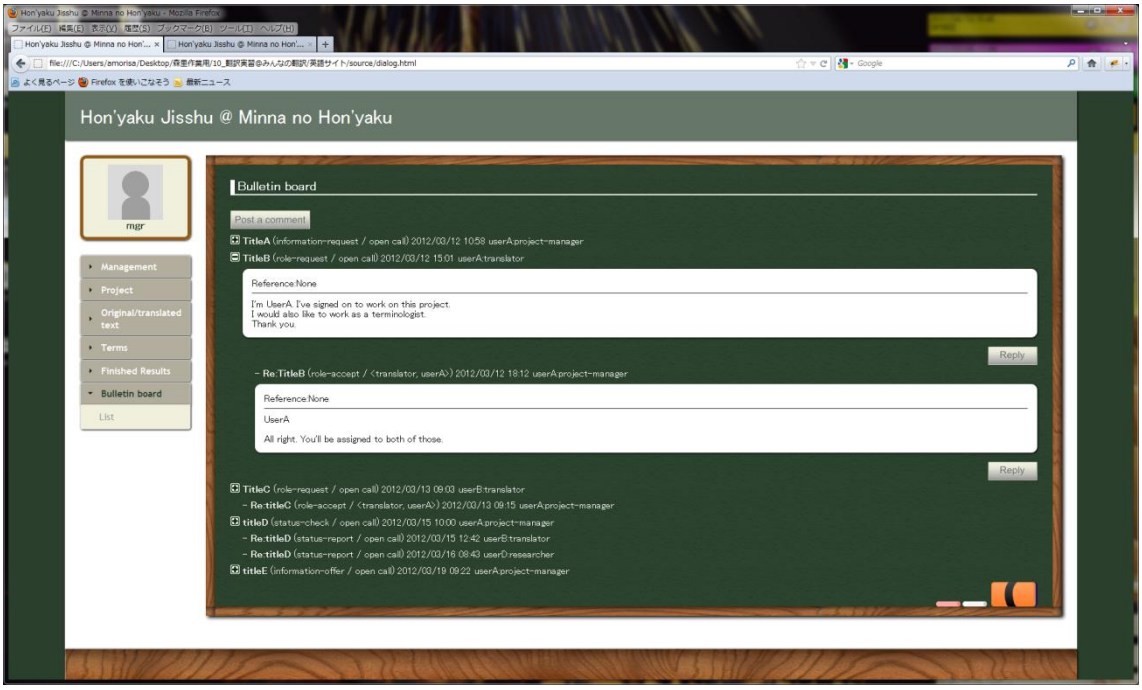

Figure 3: *Interaction in MNH-TT structured by dialogue act and referencing role*

2.4 MNH-TT: revision categorisation

The third extension provides a set of categories based on (Abekawa and Kageura, 2008; Castagnoli et al., 2006; Secara, 2005) to allow revisers to motivate and justify individual revisions (Table 3). The defined categories are grouped thematically.

- *content*- revisions bear on the perceived transfer of ideas between the source and the target document
- *lexis*- revisions bear on the choice of words and terms
- *grammar*- revisions bear on the well-formedness of the target document
- *text*- revisions relate to departures from the conventions holding for the genre of the target document, or to clumsiness, or to a lack of cohesion.

Revision categories are represented as a menu. Each time an edit is made by a reviser or reviewer, the editor chooses the appropriate revision type to classify the change. The poster is also required to specify their role (as described in Table 1).

Table 3: ***Revision categories available in MNH-TT***

Revision Categories	Description
content-omission	content present in the source document is wrongly omitted from the target document
content-addition	content not present in the source document is wrongly added to the target document
content-distortion	content present in the source document is mis-represented in the target document
content-sd-intrusion	target document contains elements of the source document left untranslated in error, or translated too literally
content-tl-intrusion	target document contains elements of the source document translated in error rather than preserved, or translated too freely
content-indecision	target document contains alternative choices left unresolved by the translator
lexis-incorrect-term	item is a non-term, incorrect, inconsistent with the glossary or inconsistent within the target document
lexis-inappropriate-collocation	item is not a usual collocate of a neighbour it governs or is governed by
grammar-syntax	syntactically incorrect construction
grammar-preposition/particle	incorrect preposition or (Japanese) particle
grammar-inflection	incorrect inflection or agreement for tense, aspect, number, case or gender
grammar-spelling	incorrect spelling
grammar-punctuation	incorrect punctuation
text-sd-inconsistent-register	lexical choice or phrasing is inconsistent with the text type of the source document
text-td-inappropriate-register	lexical choice or phrasing is inappropriate for the intended text type of the target document
text-awkward-style	phrasing is clumsy, tautologous or unnecessarily verbose
text-cohesion	inappropriate use or non-use of anaphoric expressions, or wrong ordering of given and new elements of information

2.5 MNH-TT: two- and three-pane editors

Draft translation, revision and review are all carried out using a translation-aid editor QRedit-T, which is an extended version of the two-pane QRedit (Figure 1) specially adapted for translator training. Like QRedit, it works as a two pane editor at the translation and revision stages, but with an enhanced function that allows the reviser to specify an arbitrary text span (within a paragraph) and open a small popup window with a pull-down menu listing the available revision categories.

In a standard setting for translator training, it is important for translators to keep track of the revisions made to their draft translations by revisers and reviewers. This motivated the creation of the QRedit-T three-pane editor (Figure 4), which displays in parallel the SL text, the current draft and the previous draft before revision (or, for that matter, any pair of different

translation versions). On request by the user, the editor also highlights revisions and displays the reason (error category) for a specific revision.

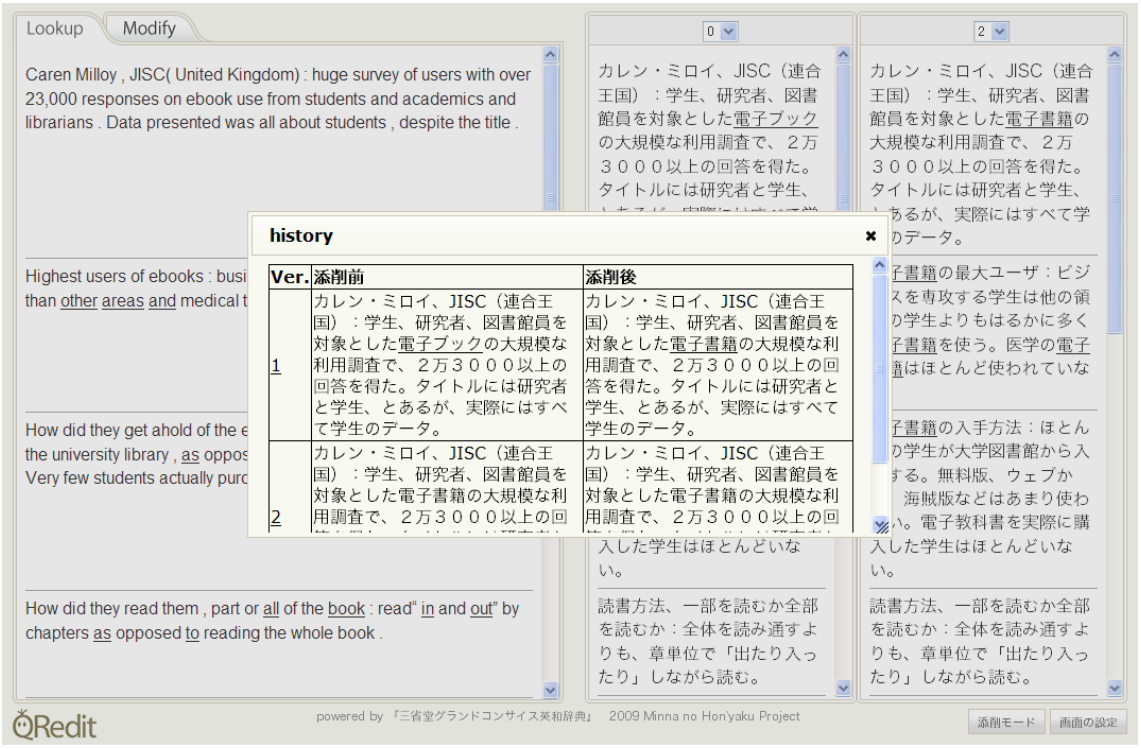

Figure 4: *Tracking the revision history in QRedit-T's three-pane editor*

2.6 MNH-TT: XML notation for recording interactions

To support these functions, we have adapted and extended the XML Translation Memory eXchange (TMX) standard notation for managing source and target document pairs. Figure 5 illustrates a key XML element of this extension, <qc:diff>, which serves to record, for a given version of the target side of a translation unit, the attributes relevant to the modification.

The revision of the sentence "添削**される**文章です。" as "添削**された**文章です。" is recorded as:

```
<qc:diff>
        qc:lang="ja"
        qc:range_from="3"
        qc:range_to="5"
        qc:user="yoko"
        qc:role="reviser"
        qc:reason="grammar-syntax">
        <qc:modified>された/qc:modified>
        <qc:comment>完了形に修正しました</qc:comment>
</qc:diff>
```

Figure 5: *Example of extended TMX notation*

2.7 MNH-TT: dashboard

A prototype dashboard has been created for the interrogation of data generated by the collaborative activity captured by the system. The dashboard is intended for three categories of user:

- the trainees themselves, who might gain insight into their own contributions to the collaboration
- their instructors, who seek to identify effective ways to intervene in the collaboration to pedagogical ends
- researchers interested specifically in collaborative translation or translator training or, more generally, in social constructivist learning scenarios or social interaction.

While it is not possible to foresee all of the types of question that this range of users might wish to ask, it is possible to characterize some scenarios. These range from rather specific queries to questions intended to reveal more general themes.

At the more specific end of this cline, the system is able to retrieve the reason for the revision of a specific part of the translation (as recorded by revision category) and any interaction linked to the revision. The ability to cut a vertical slice through the data allows users to ask more general questions, such as: Do particular dialogue act types typically accompany revisions of type X? Which dialogue types are associated with which role? Which roles are pivotal in communication among participants? Do the results of such analysis correlate with expectations about the effectiveness of the collaboration? Do they offer potential for pedagogic intervention?

Equally, the logging of temporal data, in relation both to communication between participants and modifications of the evolving translation, allows for diachronic or process-oriented investigations. For example, we might ask: Where within the project workflow are the peaks in interaction? Do these correlate with significant modifications to the translation product? In turn, the answers to such questions might inform the optimization of workflows in both training and production environments, for example by motivating the allocation of roles to individuals or by modifying the volume, timing and nature of interventions from specific role-holders, such as the project-manager.

In order to make tractable the interpretation of data and results of queries, we build on work by those engaged in the visualisation of social networks (e.g. Viégas and Donath, 1999; Mutton, 2004), thus providing graphic representations of relationships and interactions at each stage in the life of the project.

Graphs have been used to describe social networks, at least since Moreno's work in the 1930s (Freeman, 2000). One typical form of visualization of social interaction is by means of a two dimensional graph, in which nodes represent participants and edges represent interactions. In our case, participants can be generalized in accordance with their roles; edges, which may or may not be directed, can be weighted or iteratively re-drawn to represent frequency of

interactions. Equally, the presentation of nodes can be used to reflect the relative salience of a given role at a given stage in the workflow. Other properties of interactions, such as the dialogue act they instantiate or their position in terms of overall sequence, can be represented visually, by colour-coding or labelling, as illustrated by Figure 6.

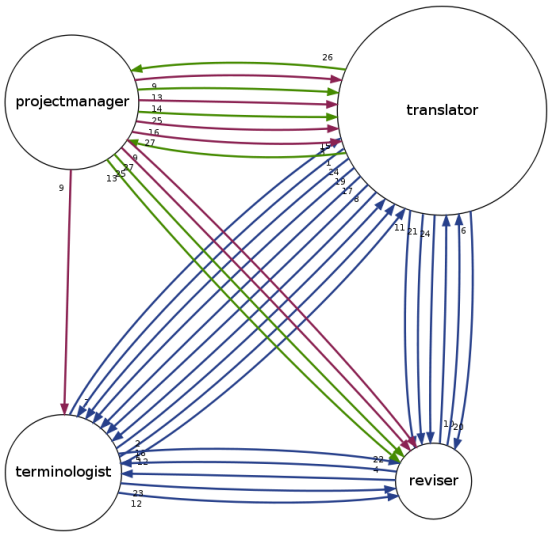

Figure 6: *Visualisation of interactions between project members during the translation stage*

Following Mutton's approach to the modelling of internet relay chat as social networks (2004), the user could watch interaction evolve through the lifetime of the translation project. The principle of temporal decay, which sees a reduction of all weightings with each re-drawing of the graph, means that less active nodes and edges fade from the picture. Such a representation offers a very straightforward means of identifying pivotal roles in the group interaction, i.e., by giving a clear indication of those nodes from which many edges emanate.

An alternative visualization is suggested by Viégas and Donath (1999), which they call a 'conversation landscape'. Such a representation foregrounds the temporal development of interaction with each discourse move being indicated by a mark on a time line. Here we adapt Viégas and Donath's model, representing major workflow stages along the x axis (as columns progressing left-to-right), with individual moves in each stage being represented on the y axis (bottom-to-top in each column). Again, these are coded for role and dialogue act, as illustrated by Figure 7.

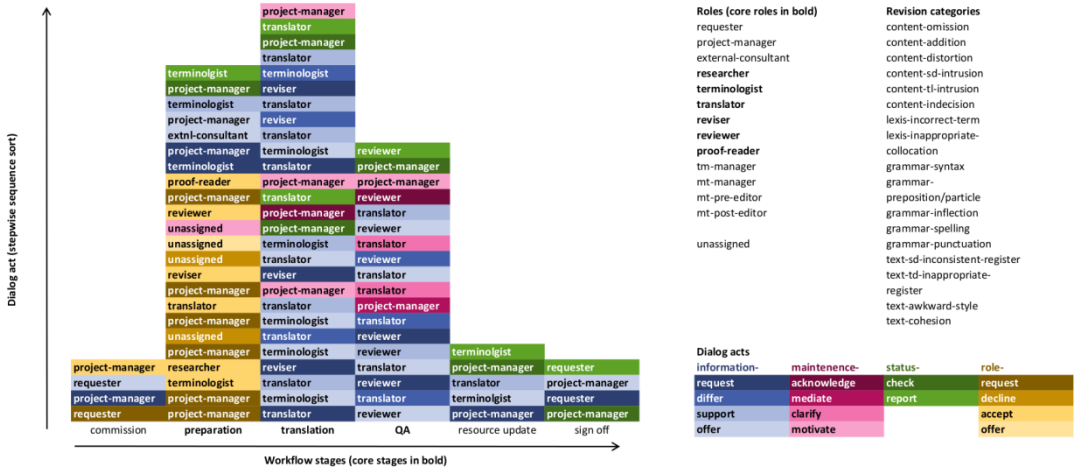

Figure 7: *Visualisation of temporal development of interactions between project members*

Thus, the primary novel contribution we make is the retrievable linkage between social interaction among participants in the collaborative translation and the evolving translation artefact.

3 Conclusions

We have motivated and described the design of extensions of the Minna no Hon'yaku collaborative translation platform, already widely used by volunteer translators, in order to realize Minna no Hon'yaku for Translator training, or MNH-TT. These extensions aim specifically to support trainees collaborating on the translation of authentic materials by structuring their roles and tasks in the overall enterprise, their verbal contributions to advancing the team effort, and their suggestions for revisions of the evolving draft translation. These four pillars of the scaffolding – roles, dialogue acts, revision categorization and extended TMX notation – have been implemented. A dashboard for visualizing the interactions within the environment and their relation to the evolving translation is at the prototype stage.

We are about to trial MNH-TT with selected groups of translation students in Japan, UK and Germany. Once we start to accumulate data through use, we will explore further our ideas for implementing the dashboard. Once its robustness is proven, MNH-TT will be made publicly available.

The explicitly provided information we have described in this paper forms only part of the available research data. The 'implicit' logs of dictionary lookup and information seeking activity are also available for mining. In the longer term, we hope to go beyond the modeling of trainee translator behavior to offering a pro-active mechanism for promoting collaboration and self-learning. Trainees could be alerted to likely problematic sections of the source text or

of the draft translation; lookup could be automated, or additional resources suggested; consultation with specific role-holders could be suggested.

Beyond the domain of human translation, accumulated aligned data consisting of human draft and revised texts or MT output and human reference texts are potentially rich resources for improving SMT systems by providing better training data. Equally, they can be exploited to build automatic post-editors, either on a purely statistical model following the approach pioneered by Dugast et al. (2007) or on the rule-based model described by Parton et al. (2012).

In short, we believe MNH-TT will be a powerful tool for trainees themselves, for their instructors and for translation researchers more widely.

Acknowledgments

We acknowledge the generous support of the Great Britain Sasakawa Foundation at the initial stages of this research.

References

Abekawa, T. and K Kageura (2007) A translation aid system with a stratified lookup interface. *ACL 2007 Demo and Poster Sessions*. Prague, Czech Republic.

Abekawa, T. and K Kageura (2008) Constructing a corpus that indicates patterns of modification between draft and final translations by human translators. *Procs. LREC 2008*. Marrakech, Morocco.

Abekawa, T, M Utiyama, E Sumita and K Kageura (2010) Community-based construction of draft and final translation corpus through a translation hosting site Minna no Hon'yaku (MNH). *Procs. LREC 2010.* Valetta, Malta.

Allen, J and M Core (1997) *DAMSL: dialog act markup in several layers.* Rochester, University of Rochester. http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/.

Arango-Keeth, F and G Koby (2003) Assessing assessment: Translator training evaluation and the needs of industry quality assessment. In B Baer and G Koby (eds) *Beyond the Ivory Tower: Rethinking translation pedagogy*. Amsterdam, Benjamins. 117-134.

Castagnoli S., D Ciobanu, N Kübler, K Kunz and A Volanschi (2006) Designing a Learner Translator Corpus for Training Purposes. *Procs. TALC2006*. Paris, France.

Clark, L (2010) Collaborative translation: Japanese-English. *MRes Dissertation*. University of Leeds, Centre for Translation Studies.

Desjardins, R (2011) 'Facebook me!': Arguing in favour of using social networking as a tool for translator training. *Linguistica Antwerpensia 10*. 175-194.

Dugast, L, J Senellart and P Koehn (2007) Statistical post-editing on SYSTRAN's Rule-Based Translation System. *Procs. 2nd ACL Workshop on Statistical Machine Translation*. Prague, Czech Republic. 220-223.

Freeman, L (2000) Visualizing social networks. *Journal of Social Stucture 1* (1).

Kageura, K, T Abekawa, M Utiyama, M Sagara and E Sumita (2011) Has translation gone online and collaborative? An experience from Minna no Hon'yaku. *Linguistica Antwerpensia 10*. 47-71.

Kiraly, D (2000) *A social constructivist approach to translator education.* Manchester, St Jerome.

Mutton, P (2004) Inferring and Visualizing Social Networks on Internet Relay Chat. *IEEE Eighth International Conference on Information Visualization.*

O'Hagan, M (2011) Introduction: Community translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond. *Linguistica Antwerpensia* 10. 11-23.

Parton, K, N Habash, K McKeown, G Iglesias and A de Gispert (2012) Can Automatic Post-Editing make MT more Meaningful? *Procs. EAMT-2012.* Trento, Italy.

Pym, A (2009) Using process studies in translator training: Self-discovery through lousy experiments. In S Göpferich, F Alves and I Mees (eds) *Methodology, Technology and Innovation in Translation Process Research*. Copenhagen, Samfundslitteratur. 135-156.

Secară, A (2005) Translation evaluation: a state of the art survey. *Procs. eCoLoRe/MeLLANGE Workshop.* University of Leeds, UK. http://ecolore.leeds.ac.uk/downloads/workshop/.

Soller, A (2001) Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education.* 12(1) 40-62.

Utiyama, M, T Abekawa, E Sumita and K Kageura (2009a) Hosting volunteer translators. *Procs. MT Summit XII*. Ottawa, Canada.

Utiyama, M, T Abekawa, E Sumita and K Kageura (2009b) Minna no Hon'yaku: A website for hosting, archiving, and promoting translations. *Procs. Translating and the Computer 31*. London, UK.

Viégas, F and J Donath (1999) Chat Circles. *Procs. ACM Conference on Computer-Human Interaction (CHI)*.