

# ScrumSourcing: Challenges of Collaborative Post- editing for Rugby World Cup 2019

*ScrumSourcing: retos de la postedición colaborativa para  
la Copa Mundial de Rugby 2019*

**Anthony HARTLEY\***, **Beibei HE\***, **Masao UTIYAMA\*\***, **Hitoshi ISAHARA\*\*\*** and  
**Eiichiro SUMITA\*\***

\* *Rikkyo University*, \*\* *National Institute of Information and Communications Technology*,  
\*\*\* *Toyohashi University of Technology: Japan*.

[a.hartley@rikkyo.ac.jp](mailto:a.hartley@rikkyo.ac.jp) / [beibei.h.s@gmail.com](mailto:beibei.h.s@gmail.com) / [isahara@tut.jp](mailto:isahara@tut.jp) / [eiichiro.sumita@nict.go.jp](mailto:eiichiro.sumita@nict.go.jp) /  
[mutiyama@nict.go.jp](mailto:mutiyama@nict.go.jp)

Recibido: julio 2018. Revisado: septiembre 2018. Aceptado: octubre 2018.

**Abstract:** This paper describes challenges facing the ScrumSourcing project to create a neural machine translation (NMT) service aiding interaction between Japanese- and English-speaking fans during Rugby World Cup 2019 in Japan. This is an example of «domain adaptation». The best training data for adapting NMT is large volumes of translated sentences typical of the domain. In reality, however, such parallel data for rugby does not exist. The problem is compounded by a marked asymmetry between the two languages in conventions for post-match reports; and the almost total absence of in-match commentaries in Japanese. In post-editing the NMT output to incrementally improve quality via retraining, volunteer rugby fans will play a crucial role in determining a new genre in Japanese. To avoid de-motivating the volunteers at the outset we undertake

an initial adaptation of the system using terminological data. This paper describes the compilation of this data and its effects on the quality of the systems' output.

**Key words:** collaborative translation; post-editing; NMT; domain adaptation.

**Resumen:** Este documento describe los retos a los que se enfrenta el proyecto ScrumSourcing para crear un servicio de traducción automática neuronal (NMT) que ayude a la interacción entre los aficionados de habla japonesa e inglesa durante la Copa Mundial de Rugby de 2019 en Japón. Este es un ejemplo de «adaptación al dominio». Los mejores datos de entrenamiento para adaptar la NMT son grandes volúmenes de oraciones traducidas típicas del dominio. Sin embargo, en la realidad no existen tales datos paralelos para el rugby. El problema se agrava por una marcada asimetría entre las dos lenguas en las convenciones para los informes posteriores al partido y la ausencia casi total de comentarios emitidos en directo durante el partido en japonés. En la post-edición de la producción de la NMT para mejorar de forma incremental la calidad a través del reentrenamiento, los voluntarios aficionados al rugby desempeñarán un papel crucial en la determinación de un nuevo género en japonés. Para evitar desmotivar a los voluntarios desde el principio, emprenderemos una adaptación inicial del sistema utilizando datos terminológicos. Este documento describe la compilación de estos datos y sus efectos en la calidad de la producción de los sistemas.

**Palabras clave:** traducción colaborativa; post-edición; NMT; adaptación al dominio.

## 1. INTRODUCTION

This paper describes the ScrumSourcing project, which relies on an online collaborative translation scenario to enable rugby fans to create a machine translation (MT) service to support «talking rugby» between Japanese- and English-speaking fans during Rugby World Cup 2019 (RWC2019) in Japan. A prime enabler of the project is Japan's investment of significant effort into neural MT (NMT) in preparation for the 2020 Olympics and Paralympics. This general initiative prioritises the personal safety, hassle-free travel and enjoyable tourism of visitors from overseas who speak little or no Japanese (Sumita 2017). But it is also the case that Rugby 7s was reinstated as an Olympic discipline at Rio 2016, thus ScrumSourcing can serve its own community at the Tokyo Olympics too.

The platform being created for the international rugby community consists of the publicly available TexTra<sup>1</sup> NMT engines and the VoiceTra<sup>2</sup> mobile application freely available on portable devices, both developed by the National Institute of Information

1. <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>
2. <http://voicetra.nict.go.jp/>

and Communications Technology (NICT)<sup>3</sup>. That is, VoiceTra will provide speech-to-text and text-to-speech «wrappers» around the core text-to-text MT system, enabling users to converse. The core text translation system specialised for rugby will also be freely available online to enable the translation of, for example, rugby news and rugby-related correspondence. This setting offers an opportunity for experimenting with collaborative translation by volunteer fans in order to adapt a baseline engine to a particular sports discipline, in this case rugby.

The best training data for adapting NMT is large volumes of translated sentences typical of the domain, but we have found very little parallel data to harvest. World Rugby publishes the official *Laws of the game Rugby Union* in 10 languages, including English and Japanese<sup>4</sup>. While we have included them in our data, they are of limited use for two reasons. First, the legalistic style is remote from that of match reporting. Second, many expressions that are highly frequent in match reports – for example, «feed», «clear» (Table 1), «steal», «feed» (Table 3) – are totally absent from the *Laws*. Elsewhere on its site, World Rugby has news pages in Japanese as well as English, but the volume of segment-for-segment translations is very small as a result of extensive editing, most commonly omission of segments of the original English.

In the absence of such parallel data, the role of the volunteer fans is to generate it by post-editing (i.e., revising) the translations of match reports produced by the current baseline system. The revised translations will be used as bitexts to adapt the system and incrementally improve quality over repeated post-editing and re-training cycles.

Although the overwhelming majority of volunteer rugby fans are likely to be non-professional translators, they are highly knowledgeable about rugby and thus able to unambiguously interpret text opaque to «outsiders».

However, a challenge for fans translating live match commentaries from English to Japanese is the absence of an established corresponding genre in the target language. Essentially, they are called upon to «invent» a new style of writing influenced with certainty by the characteristics of the English source texts and – possibly – by those of the MT output.

Moreover, in the Japanese to English direction, the pool of potential volunteers is smaller, which correspondingly reduces the volume of bitext with a genuine Japanese source that we can expect to obtain. This imbalance prompts us to explore the feasibility of incorporating back-translated, «synthetic» data in the adaptation of the Japanese to English MT engine.

Finally, post-editing in itself can become a tedious activity when the quality of the «raw» MT output is poor. Such low quality MT output is inevitable when there is a big difference between the linguistic data used to train the initial general-purpose system and

3. <http://astrec.nict.go.jp/en/index.html>

4. <https://laws.worldrugby.org/>

that of the specialised target domain. Thus, to avoid de-motivating the volunteers at the outset with an excessive editing workload, it is advisable to undertake an initial adaptation of the system using terminological data and to evaluate whether the post-editing effort is within an acceptable range. This is the focal issue of the present article.

## 2. GENRE MISMATCH BETWEEN ENGLISH AND JAPANESE

In seeking to cater for fans who want to follow coverage of rugby events appearing in the hosts' or visitors' language and to engage in discussion, we turn to «ephemeral» texts which approximate to the envisaged content and style of exchanges. Such texts fall into two types: post-match reports and in-match (live) commentaries. The former typically appear in the print and online press, while the latter are delivered via specialised media platforms such as Ultimate Rugby<sup>5</sup>.

In English, both types of text are highly ambiguous to outsiders (Hartley 2017). As Table 1 illustrates, apparently familiar words such as «go», «recover», «feed», «clear» and «hit» are laden with specialised meaning: «to clear», for instance, provides a striking example of the concision achieved with respect to «Ordinary English», while «go upstairs» offers a nice example of metonymy, given the insider knowledge that the TMO has an overview of the playing area. In short, it is clear that we are dealing with a «special language» in the sense of Sager *et al.* (1980) shared by a community of «experts» and with its own conventions of terminology and syntax.

English Rugby-Speak	Ordinary English
Farrell goes high and Tipuric recovers before feeding Webb	Farrell kicks the ball high and Tipuric catches it before passing it to Webb
Daly clears	Daly catches the ball in a deep defensive position and kicks it upfield
George hits Kruis	George successfully throws the ball into the lineout to be caught by Kruis
The ref goes upstairs	The ref(eree) consults the television match official/TMO

Table 1: English examples of the special language of rugby.

As with much human activity, a typical match consists of many repetitions of a relatively small number of action types. For example, the ball is passed (backwards) from player to player (running forwards) from one side of the pitch to the other. But

5. <https://www.ultimaterugby.com/>

a writer who uses always the same form of words to describe this repeated event risks boring the reader. Hence the high frequency in our data (see Section 4.1) of synonymous expressions like «spread the ball wide», «spin the ball wide», «go through the hands» or «go wide».

This concern for variety is accompanied by a need for concision. Journalists are constrained by space, live commentators by time, thus common knowledge is left implicit. Thus, the ball being central to the game, mention of it is often pronominalized or elided altogether, hence: «Smith spins it wide» or «Jones carries ø.» Similarly, «The referee blows the whistle» is significantly less frequent than «The referee blows».

Japanese similarly uses words of the general lexicon in specialised senses, such as: «繋がる» (generally «connect», rugby «pass»), «展開» (generally «expand», rugby «spread wide») or «先発» (generally «(baseball) pitcher», rugby «member of starting lineup»).

## 2.1. Post-match reports

The reporting of rugby in Japanese in the national media is – in relative terms – rather sparse, the pre-eminent outlet being Koichi Murakami’s column for J-Sports<sup>6</sup>. Post-match reports tend to consist of a mix of interviews with players and coaches and of factual move-by-move descriptions of the game, as illustrated by Example 1. Moreover, such reports draw on a limited lexicon with little synonymy.

その後、28分にゴール前のスクラムから、SH内田がサイドを突き、リターンパスをFL長谷川に通して2つ目のトライを奪い。

[Then on 28 minutes halfback Uchida ran from a scrum in front of the goal and took a return pass from flanker Hasegawa to score the second try].

*Example 1: Japanese post-match report.*

In contrast, match reports proliferate in the national and local press of Australia, New Zealand, England, Ireland, Scotland and Wales. This means that writers compete not only on perspective but also on style. Thus, post-match reports in English often juxtapose factual descriptions with sustained figurative passages (Example 2).

Saracens, so expressive against Northampton last week, were *lost for words* in possession, *muted* by a furious rush defence and the retrieval qualities of Louw. They

6. <https://www.jsports.co.jp/press/column/writer/26/>

*found their voice* after Farrell replaced Lozowski, a move that allowed Goode to *roar*.  
[italics added].

*Example 2: English post-match report.*

This marked asymmetry between the two conventions is problematic for the fan translator. To take the earlier example of «clear», they must choose between the default but underspecified «キックを蹴る» («kick») and adding a space- and time-consuming supplementary phrase to make the player's location on the field explicit.

## 2.2. In-match commentaries

In-match commentaries, on the other hand, are almost entirely factual, providing a live account of the action, as Example 3 illustrates.

Scrum penalty to Japan. They take it quickly and are tackled 5m short.  
Nagare tries a clever inside flick, but it's knocked on.  
Japan aren't interested in kicking today.  
Nonu clears up to his 22m.

*Example 3: English in-match report.*

In their factuality, therefore, they resemble Japanese post-match reports. Moreover, their affinity to spoken language makes them much more suitable for training an MT system that is to enable fans not only to understand commentaries but also talk rugby face to face.

An apparent problem, however, is that commentaries are almost non-existent in Japanese. Example 4 is a rare instance, taken from the website of the *Guardian*.

驚きなのは初キックを得るLO姫野和樹。  
[Lock Kazuki Himeno, who gets his first cap, is a surprise.]  
しかし、姫野選手にとってLOは不慣れのポジションとなる。  
[However, lock is not a familiar position for him.]  
セットプレーで姫野の身長を生かすことが期待されている。  
[It is expected that the team takes advantage of Himeno's height in set pieces.]

*Example 4: Japanese in-match report.*

Here lies the root cause of both the imbalance in volume of data originated in English and Japanese and also the mismatch in established genres between the two languages.

### 3. COLLABORATORS AND COLLABORATION

#### 3.1. *Profile of volunteer translators*

Japan – surprisingly, perhaps – has a history of rugby dating back to the foundation of the Yokohama Football [sic] Club in January 1866<sup>7</sup>. Today the country has more registered rugby clubs and players than any other country in the world except England. Within this, it boasts an active veterans’ scene with hundreds if not thousands of players from their 40s into their 80s playing on a regular basis throughout the year. What is more, rugby remains very much a game for university students and graduates, many of whom have a good command of, at least, written English. So there appears to be a sizeable pool of knowledgeable players and ex-players from which to recruit volunteers to translate the large volumes of suitable English data available. The ScrumSourcing project is being publicized through the national network of veterans’ rugby clubs, with support from the Japan Rugby Football Union. We do not screen volunteers and anyone signing up for translation in one or other of the language directions is deemed to be competent.

However, when Harris (2017: 41) points to research showing that «over two-thirds of *Wikipedia* translators are Advanced Native Translators», his observation raises a number of questions. Are rugby match commentaries and reports as cognitively challenging as *Wikipedia* articles? What level of bilingual competence is to be expected or required of rugby fan translators? Can a sufficient number of fans be recruited, especially in the Japanese to English direction?

#### 3.2. *Monolingual post-editing*

The answers to these questions will depend to a considerable degree on whether the MT output is good enough to be post-edited by fans expert in the domain even if they lack expertise in the source language. This consideration is particularly important for Japanese to English translation, where there are relatively few potential editors with a good reading knowledge of Japanese. If the MT output is reliably unambiguous

7. <https://www.japantimes.co.jp/life/2014/03/15/lifestyle/1866-and-all-that-the-untold-early-history-of-rugby-in-japan/#.W69rOaTqaEd>

– even if disfluent – and plausibly matches the editor’s knowledge of «the world», then post-editing with minimal reference to the source text is feasible; it comes down to corrections of grammar and style.

The advantages of using «domain insiders» as post-editors of MT output have been widely demonstrated. Aikawa, Yamamoto and Isahara (2012) successfully used university students to translate the website of a Japanese university into several Asian and European languages. Koehn (2010) concluded from a study of English monolinguals post-editing translations from Arabic and Chinese that:

«a good monolingual translator has good language skills in the target language and understands the domain. In this case, this study suggests, she may be as good as a professional bilingual translator.» (Koehn 2010: 544).

Experiments from Hindi and Russian to English demonstrated that monolinguals «can successfully produce post-edited translations that convey all or most of the meaning of the original source sentence in up to 87.8% of sentences.» (Schwartz, Anderson, Gwinnup and Young 2014: 192) Thus there is evidence across a range of typologically different languages and with MT systems of varying quality that supports the feasibility of monolingual post-editing.

The above research featured the previous generation of phrase-based statistical MT (PBSMT) systems, which have since 2016 been largely replaced by NMT systems with improved output quality. Castilho and colleagues conducted experiments with professional translators comparing the post-editing of PBSMT and NMT output from English into German, Greek, Portuguese and Russian. The results showed a preference for NMT and that fewer sentences needed post-editing (Castilho *et al.* 2017). We conclude that, since NMT requires less post-editing than PBSMT and PBSMT is already good enough for monolingual post-editing, these studies provide encouraging perspectives for the ScrumSourcing project. Moreover, Jia, Carl and Wang (2019) found that even if post-editing NMT output did not reduce time spent compared with translating from scratch, it did reduce cognitive load – an important consideration for volunteers – while maintaining quality.

However, since in ScrumSourcing we are dealing with volunteers rather than paid participants in an evaluation experiment, it is essential to ensure that the quality of the output to be post-edited is not so bad as to demotivate the volunteers. No symbolic reward system (for example, stars) is in place for individual translators, but the weekly total word/character count is published.



### 3.3. Which mode of collaboration?

Jiménez-Crespo (2017) draws a useful distinction between two types of collaborative practices in translation:

«“translation crowdsourcing” [...] represents a directed and organized top-down effort leveraged by the bottom-up contribution from the crowd» [while] «“online collaborative translations” [...] refer to horizontal bottom-up efforts by self-organized communities». (Jiménez-Crespo 2017: 18-19).

Our intention is that ScrumSourcing will come to rest under the second heading: fans will be able to select their own texts for translation, with a view to sharing the information with the international rugby community – plausibly a growing need as international ties develop between Japanese and non-Japanese clubs and increasing numbers of international world-class players pursue their careers in Japan.

Yet in the early stages the facilitators are soliciting translations of material most likely to boost the initial adaptation in a top-down mode characteristic of translation crowdsourcing. This material consists primarily of post-match reports in Japanese (Section 2.1), in-match commentaries in English (Section 2.2). Most of these texts relate to current tournaments – 6 Nations, Pacific Nations Cup, Super Rugby, International Friendlies. The material also includes «saturated» concordance lines in both languages (Section 4.4) extracted from corpora spanning two years or more.

### 3.4. Post-editing interface

The NICT TexTra interface allows volunteers to sign up to the ScrumSourcing group and to access the texts to be post-edited. These are uploaded by the ScrumSourcing team, generally in MS Word or plain text format. Reports acquired from Ultimate Rugby require some pre-processing. It is also possible to provide simply a URL and have TexTra display the translated version as an HTML page. Volunteers can also add their own texts or URLs for translation and editing by the community. 150 shows an extract of a Japanese text segmented into sentences and translated by the current version of the adapted MT engine. The post-editor can select a target sentence and modify the translation. The editing history can be viewed, as shown in 150.



Figure 1: Post-editing interface.

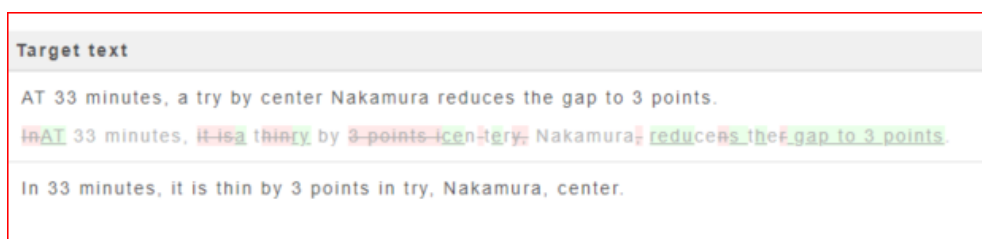


Figure 2: Post-editing history.

## 4. INITIAL ADAPTATION OF THE MT ENGINES

The chosen translation engines – English to Japanese and Japanese to English – are NMT systems produced by NICT (Utiyama 2017). They adopt an approach first implemented by Luong and Manning (2015) which is specifically geared to two scenarios: adapting a model trained on one domain to another domain, and achieving competitive performance even with little training data, so-called *low-resource translation*. Thus it is well suited to the goals and resources of the ScrumSourcing project.

The best training data for NMT is large volumes of translated sentences («bitexts» or bilingual texts) typical of the target translation domain. In general, 10,000 to 1,000,000 parallel sentences are needed to get an adapted NMT engine that is better than the general NMT engine. It is also possible to specify whether the sentences edited by volunteers in the translation editor will be used in adaptation. In our case the translations produced by fans will indeed be used for this purpose as they become available.

However, the NICT engine also provides for initial domain adaptation by means of two kinds of glossaries – bilingual and monolingual – which we extract from corpus data. In the remainder of this Section we describe the process of creating these glossaries.

## 4.1. Corpus creation and term extraction

We used the Sketch Engine<sup>8</sup> software designed for lexicographers to harvest from the web comparable English and Japanese corpora of texts in the rugby domain, totalling over 3.4 million words in each language. Additionally, we have a (currently) 5.5 million word English corpus of in-match commentaries with the permission of Ultimate Rugby, who provided an API for downloading data. Sketch Engine offers in-built functions for identifying a wide range of distributional patterns, including collocations and thesaurus relations, as well as for obtaining and filtering concordance lines.

Sketch Engine allows us, then, to substantiate the intuitions we have when looking at the raw data and the trends we reported informally in Section 2, while revealing unexpected yet typical patterns of usage. For example, presents a visualisation (at medium density) of the verb «maul» and the words with a similar distribution in the 3.4 million word rugby corpus and the 15.7 billion word English Web 2015 corpus. NMT essentially learns associations between words in two languages and these images show how starkly they differ between general and specialised usage. Moreover, in the rugby corpus «maul» occurs on average 736 times per 1 million words (736 pmw), while in the general corpus it occurs only once per million words (1 pmw).

Sketch Engine also offers powerful concordancing to show contrasting instances of use, as can be seen in and (Section 4.2), which we return to in Section 4.3.

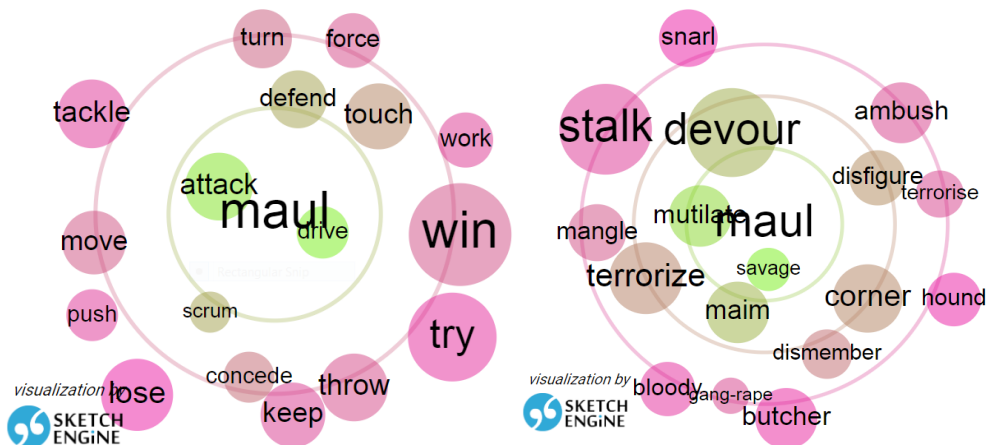


Figure 3: Sketch Engine visualisation of verb «maul» in rugby (left) and general corpora (right).

8. <https://the.sketchengine.co.uk/>

## 4.2. Bilingual glossaries for translation

Sketch Engine performs automatic extraction of single and multi-word terms by using statistical measures to compare weighted frequencies of occurrence in the specialised corpus and a reference corpus, such as English Web 2015. For Japanese, Sketch Engine's output is very noisy, so we use instead the open source TermExtract package<sup>9</sup>. Both applications sort their output by the relative salience of the items in the specialised corpus.

In this way we were able to align not only obviously technical terms («scrum»/«スクラム», «lineout»/«ラインアウト») – which pose in fact less of a challenge in translation – but also highly frequent expressions with specialized meanings, such as «big hit» (strong tackle) or «quick hands» (rapid passing) or «run it» (run carrying the ball).

Make sure **to run it** on an image with multiple faces!  
I premise **to run it** as a XBMC hub or a PLEX run.  
There aren't enough exceptional people in the economy **to run it**  
The former race winner decided **to run it** again.

*Example 5: Phrase «to run it» in the English Web 2015 corpus (1 occurrence pmw).*

Pienaar takes the ball from the back of the ruck and tries **to run it** from his own 22.  
Stormers kick the ball in behind the Blue defence, but the chase is poor and the Blues choose **to run it**.  
They choose **to run it**.

*Example 6: Phrase «to run it» in the 3.4 mw rugby corpus (132 occurrences pmw).*

These are incrementally incorporated in a bilingual glossary (currently over 300 entries) that coerces the adapted MT engine into producing the desired translation, for example, «maul» (rather than «mall») from «モール» and «ruck» (rather than «rack») from «ラック». Further bilingual glossaries specify the translations of the proper names of players, coaches, officials, stadia and clubs (currently 1,100 entries). Within the NMT process terms are replaced with special tags which remain unchanged during translation and are replaced back in a post-processing step, an approach demonstrated by Crego

9. A web implementation with an English interface is available on the «Gensen Web» site: [http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb\\_eng.html](http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb_eng.html)

et al. (2016). This approach may sometimes have negative consequences (Section 5.2).

### 4.3. Monolingual glossaries for pre-translation

Monolingual glossaries contain expressions in the source language which, for a general-domain MT engine, are highly ambiguous, having multiple possible interpretations. For English this is often due to the elliptical features of the texts already mentioned, as illustrated by the expression «to run it» in 152.

The solution is to provide on the target side of the glossary an alternative «normalised» expression whose translation will be interpretable – in this case «to run with the ball». The system substitutes this into the source text in an internal pre-editing step prior to translation. Further examples are given in Table 2. This option only works for expressions with no or very little variation.

Idiomatic English	Normalised English
Beale <i>has a go</i>	Beale <i>runs with the ball</i>
A <i>big hit</i> from Tokula on Sopoaga	A <i>strong tackle</i> from Tokula on Sopoaga
Little bit of <i>handbags</i> going on	Little bit of <i>fighting</i> going on
Blues <i>look to</i> maul	Blues <i>decide to</i> maul
Tuilagi <i>breaks the line</i>	Tuilagi <i>breaks through the line</i>

Table 2: Monolingual (English-English) glossary examples.

### 4.4. «Saturated» concordance lines for translation

At the initial translation crowdsourcing stage (see Section 3.3) Sketch Engine can be used to speed up the identification of frequent, unknown and otherwise problematic expressions for translation and post-editing in batch. These concordance lines, which are «saturated» with rugby idioms, can be prioritised over other sentences which occur in reports and commentaries but which are closer to general language. Examples are shown in Table 3 and Table 4.

EN Term	JA Term	Source text
gather ~ kick	キックキヤッチ する	Mackenzie gathers the kick from Tillous-Borde but his pass is forward however.

EN Term	JA Term	Source text
		Mike Brown gathers a kick through inside his 22 and runs it back to the 10 metre line.
		His kick is gathered by CJ Stander who takes the ball into contact inside his 22.
steal ~ lineout	ラインアウトを取られました	England steal a lineout and the set piece is a real worry for the visitors.
		Australia steal the lineout then get a penalty as Duran goes off his feet at a ruck.
		Lavanini steals the Georgian lineout.
look to ~	カウンターアタックする	Canada again look to move it wide.
		NZ win a turnover after a good counter ruck and look to counter attack.
		HK win the lineout in their 22 and look to run it out.
feed ~ scrum	スクラムにボールを投入する	Time is up as Care prepares to feed the scrum.
		Smith knocks it on the Blues will feed the next scrum.
		Van Zyl feeds the first scrum of the match inside his own half.

Table 3: Saturated concordance lines in English for translation and post-editing.

JA Term	EN Term	Source text
ボール出し	pass the ball out of breakdowns	FW第一列を軸に8人のまとまりで、まずはフィジールのボール出しを乱したい。
		特に得点チャンスの局面でセットプレーから安定したボール出しができるかどうかは勝敗のカギを握る。
		ここで圧力をかけ、ボール出しを乱すことができれば優位に立てるといふことだ。
展開	go wide	BKの展開力ではアルゼンチンよりも上と考えたほうがよく、ジョージア戦から、さらなる組織ディフェンスの向上が求められる。

		しかし、蓋を開けて見れば、オールブラックスに常に接点でプレッシャーを受けて、ボールを素早く展開できなかった。
		左に右に大きく展開して、レメキが抜け出し、最後は、ラファエレが左中間にトライ。

Table 4: Saturated concordance lines in Japanese for translation and post-editing.

In principle, as the system learns from such examples the need for pre-translation will disappear.

#### 4.5. Back-translation and translationese

The imbalance in volumes of source data available in Japanese and English data (Section 2) and also the smaller number of competent translators available for work into English (Section 3.2) is likely to result in an imbalance in the available training data. Given, however, that the Japanese source texts tend to be simpler and more repetitious in their use of terms, this could be offset if monolingual post-editing from Japanese became a real possibility. Nevertheless, the paucity of in-match live commentaries in Japanese is very likely to make recourse to back-translation necessary.

Traditionally understood, back-translation takes as its source a target text which is then translated back into the language from which it was first translated. The possible distortions introduced by this process are well known. Furthermore, the fact that translated texts have properties that distinguish them from «originals» is equally well established. In contemporary translation studies «translationese» is problematized by Frawley (1984) and first empirically substantiated on corpora by Gellerstam (1986) and Baker (1993). Teich (2003) proposes a methodology for investigating translationese from a systemic-functional linguistics perspective. More recently, translation studies scholars furnish strong support for the translationese hypothesis in showing that machine learning techniques for text categorization can outperform professional translators in discriminating between translated and non-translated texts (Baroni and Bernardini 2005). Vered *et al.* (2013) provide a detailed picture of those linguistic features which characterise translated text across a range of languages.

In the field of MT, Lembersky *et al.* (2013) confirm that phrase-based statistical MT (PBSMT) engines trained on parallel corpora translated in the opposite direction perform less well. However, they further demonstrate that a «mixture model» trained on texts translated both in the «right» and in the «wrong» directions brings a significant improvement in translation quality.

Within the current NMT paradigm, Poncelos *et al.* (2018) report surprisingly good results with this approach. In one experiment with only 1 million sentences of

training data the «synthetic-only» system (trained exclusively on back-translated data) «astonishingly» outperformed an «authentic-only» system trained on the same amount of forward-translated data (Poncelos *et al.*, 2018, 10). Most recently, Hoang *et al.* (2018) propose:

«*iterative backtranslation*, where back-translated data is used to build better translation systems in forward and backward directions, which in turn is used to re-back-translate monolingual data. This process can be “iterated” several times». (Hoang *et al.* 2018: 18).

Iterative back-translation is shown to produce better results than simple back-translation and it is this iterative approach that has been adopted by NICT for training its NMT engines (Marie *et al.* 2018).

In short, back-translation creates «synthetic» data by machine translating into the intended source language the «authentic» data available in the intended target language. In the case of ScrumSourcing, that primarily means translating English commentaries into Japanese and then using that synthetic parallel data to adapt the Japanese into English NMT engine.

## 5. QUALITY EVALUATION

### 5.1. Frameworks

Broadly speaking, the two most widely invoked parameters for human evaluations of the quality of a translated product have long been adequacy and fluency (White 2003). Adequacy entails a measure of how much information derivable from the source text can be correctly derived from the translation. Making this judgement requires comparison either with the source text (using bilingual evaluators) or with a «reference translation» validated by a human translator (using monolingual evaluators). Fluency entails a measure of the well-formedness and naturalness of the translated text in its own right. Analysis schemes have been developed to sub-categorise errors of both adequacy and fluency, linking them to linguistic features of the source and target texts and languages. Operational definitions of adequacy and fluency can be found on the websites of TAUS<sup>10</sup> and the QT21 project<sup>11</sup> responsible for devising the Multidimensional Quality Metrics (MQM). There is a vast literature on translation quality assessment, of which Moorkens *et al.* (2018) offer a timely discussion encompassing both human and machine translation.

10. <https://www.taus.net/>

11. <http://www.qt21.eu/>



Given the prohibitive time and cost of human evaluations, MT developers rely heavily on automated metrics. Of these the best known is BLEU (Papineni *et al.* 2002), which basically measures the similarity of a machine-translated sentence to one or more human translations of that sentence. While it has been long and widely criticised for, among other limitations, failing to provide any grounded measure of quality (Babych *et al.* 2005; Callison-Burch *et al.* 2006), it remains in widespread use on the generally-held assumption that an improved BLEU score on a given set of test data reliably indicates an improvement in translation quality.

At this stage in the ScrumSourcing project it is premature to conduct formal evaluations of translation quality. Moreover, in the absence of a test set of source and human translated sentences kept separate from the training data, it is unlikely that we will make use of BLEU at all. Rather, we rely in the long term on the assumption that the «injection» of additional human-translated training data will incrementally improve performance. Thus Section 5.2 presents only indicative observations of some of the gains and losses induced by initially adapting the engines with bilingual glossaries.

## 5.2. Observations on output quality

Neural machine translation aims to directly model the probability of translating a source *sentence* to a target *sentence* as a result of training on a large collection of (authentic and synthetic) parallel sentences. It is only to be expected, then, that coercing the system to use terms imposed by a glossary that is external to the model (Section 4.2) is likely to perturb the output of the model. That is indeed what we can witness in places. For instance, in Table 6 example 1 the erroneous «Li» stems from an entry in the names glossary «リ」/'Li» which has matched a single character within «サントリー」/'Suntory». In example 4 in the same table, the correct translation of «DF» as «defender» has perturbed the translation of the rest of the sentence.

Overall, the baseline engine is better than the adapted engine for about 17% of segments. However, an encouraging 47% of sentences output by the adapted Japanese to English engine appear to require no post-editing even at this early stage. We have not yet assessed the effort required to post-edit the other 53% of segments to an acceptable level.

	JA	NICT-NMT-Standard	NICT-RWC-20181029
1	驚きなのは初キヤップを得るLO姫野和樹。	The LO Kazuki Himeno is the first to get the first cap.	It is surprising that lock Kazuki Himeno gets the first cap.
2	セットプレーで4姫野の身長を生かすことが期待されている。	It is expected that the Himeno will be used in a set play.	It is expected that the height of Himeno will be utilized in set piece.
3	今季トップリーグで結果を残している松田力也がSOで先発。	Rikiya Matsuda, who left the results in this season's top league, started with SO.	Rikiya Matsuda, which has left the results in Top League this season, started in stand-off.
4	CTBは立川とラファエレ2人のエース。	The CTB has 2 aces, Tachikawa and Rafalore.	Center has 2 aces, Tatekawa and Lafaele.
5	その経験をスクラムとラインアウトの面で生かすだろうと予測できる。	We can expect that the experience will take advantage of scrum and lineout.	It can be predicted that the experience will be utilized in both scrum and lineout.

Table 5: Example improvements over the baseline engine.

	JA	NICT-NMT-Standard	NICT-RWC-20181029
1	No8のショーン・マクマホンが来季からサントリーへ加入する予定。	Sean MacMahon of No. 8 will join Suntory from the coming season.	Sean MacMahon of No. 8 will join Santo Li from the coming season.
2	両チームの選手がピッチに入場してきた!	The players of both teams entered the pitch!	Both side players have entered the pitch!
3	ワラビーズの2トライ目。	This is the second try of Wallabies.	This is the second try of Brackens.
4	ヘンリー・スペイトが日本のDFを突破し、抜け出す。	Henry Spiekite breaks out of Japan's DF and breaks out.	Henry Stapet broke out of the Japanese defender and broke out of the country.
5	21分: 先発のLOヘルウヴェに代えてヴァインピー・ファンデルヴァルトを投入。	21 minutes : Instead of starting LO cheluve, we put a vinpyo vanderveldt in place.	21 minutes : Wimpie Van Der Walt instead of lock Uwe Helu, the starting pitcher of throw-in.

Table 6: Example degradations below the baseline engine.

Although research points to the feasibility of using community members to evaluate translation quality and fitness for purpose (for example, Mitchell, O'Brien and Roturier

2014; Jiménez-Crespo 2018), conducting such an exercise is beyond the resources of the ScrumSourcing project.

## 6. CONCLUSIONS

We have outlined an initiative to create an NMT system capable of supporting the Japanese-English «rugby-speaking» community, highlighting major differences in the styles and genres that exist in the two languages. We have shown how tools like Sketch Engine can be harnessed to identify terms and other expressions characteristic of the rugby domain. These can be used to quickly adapt an initial version of the desired system to avoid demotivating volunteers at the outset. We have illustrated the occasional disadvantages of coercing the translation of specific terms by means of a bilingual glossary external to the neural, sentence-based translation model. In the medium term, we are focusing the effort on «translation crowdsourcing» where we solicit translations of particular commentaries or even concordance lines featuring highly frequent and highly idiomatic expressions. Fans will nonetheless be free to use the system for «online collaborative translations» of texts of their choice. Once our evaluations show that the system has matured, this unsolicited use will prevail. Post-edited texts will continue to be automatically used to further adapt the engines. We envisage that this practice will have a profound effect on the nature of in-match rugby commentaries in Japanese.

## 7. REFERENCES

- AIKAWA, Takako, Kentaro YAMAMOTO and Hitoshi ISAHARA. 2012. «The Impact of Crowdsourcing Post-editing with the Collaborative Translation Framework». In *JapTAL 2012, LNAI 7614*, ed. by Hitoshi Isahara and Kyoko Kanzaki, 1-10. Berlin/Heidelberg: Springer-Verlag.
- BABYCH, Bogdan, Anthony HARTLEY and Debbie ELLIOTT. 2005. «Estimating the predictive power of n-gram MT evaluation metrics across languages and text types». *Proceedings of MT Summit X*. Phuket, Thailand, September 12-16.
- BAKER, Mona. 1993. «Corpus linguistics and translation studies: Implications and applications». In *Text and Technology: In Honour of John Sinclair*, ed. by Mona Baker, Gill Francis, and Elena Tognini-Bonelli. Amsterdam: Benjamins, 233-250.
- BARONI, Marco and Silvia BERNARDINI. 2006. «A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text». *Literary and Linguistic Computing*, 21(3): 259-274, September 2006. <https://doi.org/10.1093/lc/fqj039>
- CALLISON-BURCH, C., M. OSBORNE, and P. KOEHN. 2006. «Re-evaluating the Role of BLEU in Machine Translation Research». *Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics: EACL 2006*, 249-256.

- CASTILHO, Sheila, Joss MOORKENS, Federico GASPARI, Rico SENNRICH, Vilemini SOSONI, Panayota GEORGAKOPOULOU, Pintu LOHAR, Andy Way, Antonio Valerio MICELI BARONE and Maria GIALAMA. 2017. «A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators». *Proceedings of MT Summit XVI, vol. 1: Research Track*, 116–131, Nagoya, Japan, September 18-22, 2017.
- FRAWLEY, William. 1984. «Prolegomenon to a Theory of Translation». In *Translation: Literary, Linguistic and Philosophical Perspectives*, ed. by William Frawley. Newark: University of Delaware Press, 159-175.
- GELLERSTAM, M. 1986. «Translationese in Swedish Novels Translated from English». In *Translation Studies in Scandinavia*, ed. by L. Wollin and H. Lindquist. Lund: CWK Gleerup, 88-95.
- CREGO, Josep et al. 2016. «SYSTRAN's Pure Neural Machine Translation Systems». Accessed December 15, 2018. arXiv:1610.05540.
- HARRIS, Brian. 2017. «Unprofessional translation: A blog-based overview». In *Non-professional Interpreting and Translation*, ed. by Rachele Antonini, Letizia Cirillo, Linda Rossato and Ira Torresi. Amsterdam/Philadelphia: John Benjamins.
- HARTLEY, Anthony. 2017. «Will MT blow for full time on ノーサイド (no side)?». *AAMT Journal*, 65: 1-3.
- HOANG, Vu Cong Duy, Philipp KOEHN, Gholamreza HAFFARI and Trevor COHN. 2018. «Iterative back-translation for neural machine translation». *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, 18-24. Melbourne, Australia, July 20, 2018.
- JIA, Yanfang, Michael Carl and Xiangling Wang. 2019. «How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study». *The Journal of Specialised Translation*, 31: 61-86, January 2019.
- JIMÉNEZ-CRESPO, Miguel A. 2017. *Crowdsourcing and Online Collaborative Translations*. Amsterdam/Philadelphia: John Benjamins.
- JIMÉNEZ-CRESPO, Miguel A. 2018. «Crowdsourcing and Translation Quality: Novel Approaches in the Language Industry and Translation Studies». In *Translation Quality Assessment*, ed. by Joss Moorkens, Sheila Castilho, Federico Gaspari, Stephen Doherty. Berlin: Springer. 69-93.
- KOEHN, Philipp. 2010. «Enabling Monolingual Translators: Post-editing vs. Options». *Proceedings Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, 537–545, Los Angeles, California, June 2010.
- LEMBERSKY, Gennadi, Noam ORDAN and Shuly WINTNER. 2013. «Improving statistical machine translation by adapting translation models to translationese». *Computational Linguistics*, 39 (4): 999-1023, December 2013. [http://dx.doi.org/10.1162/COLI\\_a\\_00159](http://dx.doi.org/10.1162/COLI_a_00159)
- LUONG, Minh-Thang and Christopher D. MANNING. 2015. «Stanford Neural Machine Translation Systems for Spoken Language Domains». *International Workshop on Spoken Language Translation*, Da Nang, Vietnam, December 2015.
- MARIE, Benjamin, Rui WANG, Atsushi FUJITA, Masao UTIYAMA and Eiichiro SUMITA. 2018. «NICT's Neural and Statistical Machine Translation Systems for the WMT18 News Translation Task». *Proceedings of the Third Conference on Machine Translation (WMT)*, Volume 2: Shared Task Papers, 449-455. Brussels, Belgium, October 31-November 1 2018.
- MITCHELL, Linda, Sharon O'BRIEN and Johann ROTURIER. 2014. «Quality evaluation in community post-editing». *Machine Translation*, 28: 237-262, November 2014.

- MOORKENS, Joss, Sheila CASTILHO, Federico GASPARI and Stephen DOHERTY (eds). 2018. *Translation Quality Assessment*. Berlin: Springer.
- PAPINENI, K., S. Roukos, T. WARD and W. JING ZHU. 2002. «Bleu: a method for automatic evaluation of machine translation». *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Saarbruecken, Germany, July 2002.
- PONCELAS, Alberto, Dimitar SHTERIONOV, Andy WAY, Gideon MAILLETTE DE BUY WENNIGERAND and Peyman PASSBAN. 2018. «Investigating Backtranslation in Neural Machine Translation». arXiv:1804.06189v1.
- SAGER, Juan, David DUNGWORTH and Peter McDONALD. 1980. *English Special Languages: Principles and practice in science and technology*. Wiesbaden: Brandstetter.
- SCHWARTZ, Lane, T. ANDERSON, J. GWINNUP and K. M. YOUNG. 2014. «Machine translation and monolingual postediting: The AFRL WMT-14 system». *Proceedings of the Ninth Workshop on Statistical Machine Translation*, 186-194, Baltimore, Maryland.
- SUMITA, Eiichiro. 2017. «Social innovation based on speech-to-speech translation technology targeting the 2020 Tokyo Olympic/Paralympic Games». *MT Summit XVI, Invited Talk*, Nagoya, Japan, September 18-22, 2017.
- TEICH, Elke. 2003. *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: Mouton de Gruyter.
- UTIYAMA, Masao. 2017. «Recipe for High Quality Machine Translation». *MT Summit XVI JTF Workshop: Machine Translation acceptance among the Language Industry*, Nagoya, Japan, September 22, 2017.
- VOLANSKY, Vered, Noam ORDAN and Shuly WINTNER. 2015. «On the features of translationese». *Digital Scholarship in the Humanities*, 30 (1): 98-118, April 2015.
- WHITE, John. 2003. «How to evaluate machine translation». In *Computers and Translation: A translator's guide*, ed. by Harold Somers. Amsterdam/Philadelphia: John Benjamins. 211-244.

