PBML

The Prague Bulletin of Mathematical Linguistics
NUMBER 108 JUNE 2017 147–158

Evaluating the Usability of a Controlled Language Authoring Assistant

Rei Miyata,[a] Anthony Hartley,[b] Kyo Kageura,[c] Cécile Paris[d]

[a] Nagoya University
[b] Rikkyo University
[c] The University of Tokyo
[d] Data61, CSIRO

Abstract

This paper presents experimental results of a usability evaluation of a controlled language (CL) authoring assistant designed to help non-professional writers create machine translatable source texts. As the author drafts the text, the system detects CL rule violations and proscribed terms. It also incorporates several support functions to facilitate rephrasing of the source. In order to assess the usability of the system, we conducted a rewriting experiment, in which we compared two groups of participants, one with the aid of the system and the other without it. The results revealed that our system helped reduce the number of CL violations by about 9% and the time to correct violations by more than 30%. The CL-applied source text resulted in higher fluency and adequacy of MT outputs. Questionnaire and interview results also implied the improved satisfaction with the task completion of those participants who used the system.

1. Introduction

In recent years, machine translation (MT) has been increasingly adopted not only for company documentation but also by public services. A number of local governments in Japan have started using MT on their websites to provide local residents with multilingual information. However, the resultant translation is often confusing given that MT between distant languages such as Japanese and English is generally difficult (Isahara, 2015), and the necessary post-editing into multiple languages is too costly. To make better use of MT in the field, one viable solution is to constrain the source into a form amenable to MT by making use of a controlled language (CL), including properly controlled terminology.

 Corresponding author: miyata@nuee.nagoya-u.ac.jp
Cite as: Rei Miyata, Anthony Hartley, Kyo Kageura, Cécile Paris. Evaluating the Usability of a Controlled Language Authoring Assistant. The Prague Bulletin of Mathematical Linguistics No. 108, 2017, pp. 147–158.
doi: 10.1515/pralin-2017-0016.

Although the effectiveness of a CL itself is evidenced by improvements in not only machine translatability but also human readability (e.g., Bernth and Gdaniec, 2001; Aikawa et al., 2007; Miyata et al., 2015), writing in accordance with a particular CL is a hard task, especially for non-professional writers, such as those who create municipality websites and documents, since it requires a command of controlled writing. Thus, in practice it is essential to support authors in checking conformity to CL guidelines and terminology, and in editing the source text appropriately. In this research project, which focuses on Japanese-to-English translation of municipal documents, we developed an interactive authoring assistant designed to help non-professional writers create machine translatable source text (ST). The key feature of our system is that it supports users' decision-making at each step in validating the source.

While much effort has been devoted to the conventional product evaluation of MT (Bojar et al., 2015), few attempts have been made to assess the *usability* of an authoring support system to maximise MT use. In this study, we conducted a usability evaluation based on the ISO standard for human-computer interaction (ISO, 2010) and related studies (e.g., Doherty and O'Brien, 2013; Sauro and Lewis, 2012). To the best of our knowledge, this is the first attempt to evaluate the usability of a CL authoring assistant intended for improving MT performance.

We discuss related work in Section 2. In Section 3, we describe our CL guidelines compiled for this study. Section 4 explains the CL authoring assistant and the implementation of the guidelines. We elaborate on our experimental set up in Section 5 and present our results accompanied by discussion in Section 6. Section 7 presents conclusions and future directions.

2. Related Work

A number of CL rule sets have been proposed with a view to improving machine translatability as well as facilitating human comprehension (Kittredge, 2003; Kuhn, 2014). Evidence of improved machine translatability and post-editing productivity has also been provided (Pym, 1990; Bernth and Gdaniec, 2001; Aikawa et al., 2007; O'Brien and Roturier, 2007). Miyata et al. (2015) revealed in an evaluation experiment comparing four MT systems that compiling optimal rules for particular MT systems yields a great improvement in MT quality, a case also mentioned by O'Brien (2003).

Terminology management also plays a central role in improving both ST consistency and MT quality. In an experiment translating technical documentation from English to French, Thicke (2011) demonstrated that simply customising an MT engine with terminology boosted post-editing productivity. She also concluded that the combination of controlling the ST via general writing guidelines and customising the MT engine with terminology further increased translation productivity, making it four times faster than human translation from scratch.

However, writing source texts in accordance with a CL and pre-defined terminology is not an easy task. Providing writing support tools is essential, particularly for non-professional authors. A leading example of CL writing support in combina-

tion with MT is the KANTOO Controlled Language Checker (Mitamura et al., 2003; Nyberg et al., 2003), which incorporates functions to detect problems in the ST and provides diagnostic messages for interactive rewriting. Although some commercial source checking tools for Japanese have recently become available,[1] to date few practical implementations or evaluation results for Japanese CL tools have been provided.

While the performance of CL checkers has been benchmarked in terms of precision and recall of their violation detection components (Mitamura et al., 2003; Rascu, 2006; Miyata et al., 2016), to develop a workable system usability assessment is also crucially important. The relevant ISO standard defines usability as the 'extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use' (ISO, 2010). It describes these three measures as follows:

effectiveness: accuracy and completeness with which users achieve specified goals
efficiency: resources expended in relation to the accuracy and completeness with which users achieve goals
satisfaction: freedom from discomfort and positive attitudes towards the use of the product

Compared to the number of conventional product evaluations of MT by human subjective judgement or automated metrics (Bojar et al., 2015), the MT research community has published relatively few usability evaluations. Exceptions are, for example, Castilho et al. (2014) and Doherty and O'Brien (2012, 2013), who employed the three measures above to evaluate MT outputs, and Alabau et al. (2012), who conducted a user evaluation of their interactive MT systems. How usable a CL authoring system is for the end user remains an open question which needs to be addressed to enable the adoption of CL and MT in the workplace.

3. Controlled Language Guidelines

As just mentioned, CL rules become particularly effective when rule sets are tailored to specific MT systems. In our scenario, we focused on two MT systems: TransGateway,[2] a commercial rule-based MT (RBMT) system widely used in Japanese municipalities, and TexTra,[3] a freely available state-of-the-art statistical MT (SMT) system. We previously created a total of 60 Japanese CL rules and assessed the effectiveness of each rule with different MT systems (Miyata et al., 2015). Based on the evaluation results, we selected effective rules for each system and compiled two CL guidelines, henceforth **CL-R** and **CL-S**.[4]

[1] For example, Acrolinx supports several languages including Japanese. http://www.acrolinx.com/

[2] Kodensha CO., http://www.kodensha.jp

[3] NICT, https://mt-auto-minhon-mlt.ucri.jgn-x.jp

[4] R and S stand for RBMT and SMT, respectively.

No	Rule	CL-R	CL-S	Implement	
1	Try to write sentences of no more than 50 characters.	✓	✓	✓	(10)
2	Do not interrupt a sentence with a bulleted list.	✓	✓		
3	Ensure the relationship between the modifier and the modified is clear.	✓	✓		
4	Use the particle *Ga* only to mean 'but'.	✓	✓	✓	(9)
5	Do not use the preposition *Tame* to mean 'because'.	✓		✓	(10)
6	Avoid using multiple negative forms in a sentence.	✓		✓	(10)
7	Do not use *Reru/Rareru* to express the potential mood or honorifics.	✓	✓	✓	(4)
8	Avoid using words that can be interpreted in multiple ways.	✓	✓		
9	Avoid using the expression *To-iu*.	✓		✓	(8)
10	Avoid using the expression *Omowa-reru* and *Kangae-rareru*.	✓		✓	(10)
11	Avoid the single use of the form *Tari*.		✓	✓	(10)
12	Use words from a general Japanese-English dictionary.	✓	✓		
13	Avoid using compound Sahen-nouns.[5]		✓	✓	(10)
14	Ensure there are no typos or missing characters.	✓	✓		
15	Do not omit subject.	✓	✓	✓	(5)
16	Do not omit object.	✓	✓		
17	Do not use comma for connecting noun phrase enumeration.	✓	✓	✓	(7)
18	Avoid using particle *Ga* for object.	✓	✓	✓	(8)
19	Avoid using *Te-kuru/Te-iku*.		✓	✓	(10)
20	Avoid inserted adverbial clause.		✓		
21	Do not end clause with noun.	✓	✓		
22	Avoid using Sahen-noun + auxiliary verb *Desu*.	✓	✓	✓	(10)
23	Avoid using attributive use of *Shika-Nai*.	✓	✓	✓	(10)
24	Avoid using verb + *You-ni*.		✓	✓	(10)
25	Avoid using particle *Nado*.		✓	✓	(10)
26	Avoid using giving and receiving verb.	✓	✓	✓	(10)
27	Avoid using verbose word.	✓	✓	✓	(10)
28	Avoid using compound word.	✓	✓	✓	(9)
29	Do not omit parts of words in enumeration.	✓	✓	✓	(4)
30	Do not omit expression to mean 'per A'.	✓	✓	✓	(10)
31	Avoid using conjunctive particle *Te*.	✓	✓	✓	(10)
32	Avoid using particle *To* to mean 'if'.	✓		✓	(10)
33	Use Chinese Kanji characters for verb as much as possible instead of Japanese Kana characters.	✓	✓	✓	(4)
34	Avoid leaving bullet mark in texts.	✓	✓	✓	(10)
35	Avoid using machine dependent characters.	✓	✓	✓	(10)
36	Avoid using square bracket for emphasis.	✓	✓	✓	(10)
Term	Use term properly	✓	✓	✓	(10)

Table 1. CL rules and implementation (with precision scores)

Guideline **CL-R** comprises 30 rules while **CL-S** comprises 31 rules. The total number of distinct rules is 36, with 25 rules belonging to the both guidelines (Table 1).

For each CL rule, we provided a description and example rewrites, to enable authors to fully understand and apply the rule while drafting or revising.

[5] A Sahen-noun is a noun which can be connected to the verb *Suru* and act as a verb.

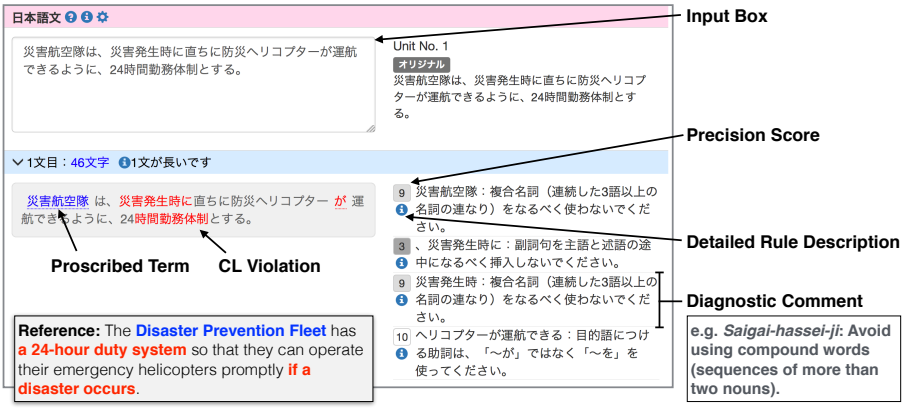

Figure 1. User interface

4. Controlled Language Authoring Assistant

4.1. Concept

The aim of our system is to help users create controlled STs. We designed a real-time, interactive system to check texts for conformity to CL rules and terminology during drafting or revision. Whenever a user enters input violating any of the operative CL rules or a term registered in a proscribed term list, the system alerts and supports the user in amending it.

Given that our target users—non-professional writers—tend to be unaccustomed to the principle of controlled writing and unfamiliar with writing tools, we need to provide support explanations and instructions. We therefore implemented several functions to assist the author's decision-making at each step of (re)writing, i.e., *detection*, *suggestion* and *correction*. Hitherto, CL 'checkers' have been deployed in two settings: post-hoc revision or rewriting (of legacy documents, for example) and assistance with 'drafting-from-scratch' (the more productive workflow). Our tool is designed to fit both scenarios.

4.2. Interface and Function

Figure 1 shows the system interface. The use scenario is as follows.

1. Users enter Japanese text in the **input box**.
2. The system automatically analyses each sentence and displays any detected **CL violations** in red and **proscribed terms** in blue (*detection*).
3. Users modify the problematic segments based on the **diagnostic comments**, referring to **detailed rule descriptions**, if needed.
4. For particular highlighted segments, the function offers alternative expressions displayed by clicking the segments (*suggestion*).
5. If the author clicks a suggestion, the segment in the input box is automatically replaced (*correction*).

4.3. Rule Implementation

To implement the CL violation detection function, we created surface part-of-speech pattern matching rules using the Japanese morphological analyser MeCab.[6] We then conducted a benchmark evaluation to calculate the precision and recall of the detection performance of each rule based on a previous study (Miyata et al., 2016). If the precision was below 0.4, we chose not to implement the rule. If the precision was above 0.4, we mapped it to a 10-point scale (**precision score**), which informs users how reliable the detection presented by the system is (shown in Figure 1). The rightmost column in Table 1 shows the 28 implemented rules with their precision scores. The terminology check function can similarly be implemented by simple string matching rules and integrated into the system. What is needed is to create a list of synsets of preferred and proscribed terms (Warburton, 2014).

5. Experimental Setup

Based on the ISO definition of usability introduced in Section 2, our questions for the system evaluation are: (1) Does the system help reduce CL violations and improve MT quality? (**effectiveness**); (2) Does the system help reduce time spent on controlled writing? (**efficiency**); (3) Is the system easy for non-professional writers to use and favourably accepted? (**satisfaction**)

To assess these three aspects, we designed a rewriting task in which two groups of participants were asked to amend Japanese source sentences violating CL rules and terminology, respectively with and without the aid of the system. Thus, we emulate the post-hoc revision setting. We (1-a) counted the number of corrected violations, (1-b) evaluated the MT quality, (2) measured the time taken to correct violations, and (3) gauged subjective satisfaction.

5.1. Task Design

Data: To count the number of corrected violations (**1-a. effectiveness**), we prepared a manually annotated dataset. We used sentence data extracted from Japanese municipal websites and selected 30 sentences to ensure that the dataset contained at least one violation of each of the 36 rules. Additionally, we artificially modified two proper nouns from the municipal domain into proscribed forms. The final dataset consisted of 67 violations of **CL-R** and 76 violations of **CL-S**, including two terminology violations of each.

Condition: For each of the two CL guidelines, **CL-R** and **CL-S**, one group of participants rewrites sentences with the sole aid of a print copy of the guideline and a term list[7] without access to the system's support functions (**control**), while the other

[6] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://taku910.github.io/mecab/

[7] All participants were given the same term list, which enumerates 100 Japanese municipal terms including some proscribed forms. It was artificially created by the authors for the purpose of the evaluation.

1	Overall, I am satisfied with the ease of completing the tasks in this scenario.
2	Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.
3	Overall, I am satisfied with the support information (online-line help, messages, documentation) when completing the tasks.

Table 2. After-Scenario Questionnaire (ASQ)

1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

Table 3. System Usability Scale (SUS)

group can use the full assistance of the system (**treatment**). Thus, four conditions were prepared: (1) **C**ontrol group with CL-**R** (**CR**); (2) **C**ontrol group with CL-**S** (**CS**); (3) **T**reatment group with CL-**R** (**TR**); (4) **T**reatment group with CL-**S** (**TS**).

Procedure: Each participant is presented with a sentence in the input box of the system (see Figure 1) and is asked to amend any segments that violate CL rules or terminology, while maintaining the meaning of the source. All functions of the authoring assistant are disabled for the control group. As soon as the correction is completed, the resulting sentence is automatically saved and the participant proceeds to the next sentence. The system also records the elapsed time of the task for each sentence (**2. efficiency**).

Post-task questionnaire: To investigate **3. satisfaction** with the task and the system, we employed two standardised questionnaires widely used in usability studies: After-Scenario Questionnaire (ASQ) (Sauro and Lewis, 2012) and System Usability Scale (SUS) (Brooke, 1996). To evaluate how satisfied users were with the task, we used an ASQ with three questions on a seven-point Likert scale from '1: strongly disagree' to '7: strongly agree' (Table 2).[8] To evaluate the usability of the system itself, we used a SUS with ten questions on a five-point Likert scale from '1: strongly disagree' to '5: strongly agree' (Table 3).[9] Odd-numbered questions are worded positively, while even-numbered questions are worded negatively.

MT evaluation: To evaluate the resultant MT outputs (**1-b. effectiveness**), we conducted the traditional human evaluation. An evaluator judges each MT output in

[8]Since the questionnaire is originally in English, we translated it into Japanese. We also changed 'online help, messages, documentation' to 'documentation' in the third question for the control group as we did not provide them with any online help or messages.

[9]We also translated this into Japanese.

	Treatment (with system)			Control (without system)		
	TR	TS	Mean	CR	CS	Mean
Corrected violation (num.)	55.0	62.7	58.8	49.7	55.7	52.7
Missed violation (num.)	11.0	13.3	12.2	16.3	20.3	18.3
Correction rate (%)	83.3	82.5	82.9	75.3	73.3	74.3

Table 4. Effectiveness for each condition

terms of *fluency* from '5: Flawless English' to '1: Incomprehensible' and *adequacy* from '5: All' (of the meaning correctly correctly expressed) to '1: None'. The rewritten versions of the ST were translated by the intended MT systems, TransGateway (RBMT) or TexTra (SMT). As baseline and oracle outputs, we also translated the original ST and two sets of fully CL-compliant STs that we rewrote according to **CL-R** and **CL-S**.

5.2. Implementation

We recruited 12 university students, all of them native speakers of Japanese and regularly writing Japanese texts on computers, but none engaged in professional writing activity, such as technical writing or translation. They can thus be regarded as typical of our target end-users, i.e., non-professional writers. Three participants were randomly placed in each of the four conditions.

We first gave participants brief instructions for the rewriting task, then asked them to read through the CL guideline and the term list. In a preliminary session, each participant rewrote five example sentences to get used to the task and the system. In the task proper, each participant rewrote all 30 sentences, the order of which was randomised. Since this task imposes a heavy cognitive load on participants, we divided the 30 sentences into three sets, each of 10 sentences, and let participants take a short rest between the sets. After the main task, we asked them to answer ASQ and SUS,[10] and conducted a follow-up interview based on the responses.

For the MT evaluation task, we employed three native English speakers, who are engaged in Japanese-to-English translation, to evaluate all versions of the MT outputs.

6. Results and Discussions

6.1. Effectiveness

Table 4 shows the result of the effectiveness measures. Correction rate indicates the percentage of violations correctly amended throughout the task. On average, the treatment group achieved about a 9% higher correction rate than the control group, which an independent t-test found to be a significant difference ($t = -2.878, df = 10, p = .016$).

Detailed analysis of the results revealed that the correction rate for four rules—12, 14, 16 and 29 in Table 1—of the treatment group is lower than that of control group. We also noted that three of these four rules—12, 14 and 16—are not yet implemented. This implies that users tend to rely on the system and overlook any violations the

[10]Participants assigned to the control group answered only ASQ.

	RBMT				SMT			
	TR	CR	Original	Oracle	TS	CS	Original	Oracle
Fluency	2.74	2.72	2.36	2.87	2.58	2.57	2.32	2.70
Adequacy	3.20	3.19	2.80	3.46	2.89	2.80	2.62	3.34

Table 5. Result of MT quality evaluation

	Treatment (with system)			Control (without system)		
	TR	TS	Mean	CR	CS	Mean
Total time (sec.)	2405	2206	2306	3744	2844	3294
Time per sentence (sec.)	80.2	73.5	76.9	124.8	94.8	109.8
Time per correction (sec.)	43.7	35.2	39.5	75.3	51.1	63.2

Table 6. Efficiency for each condition

system does not detect. It is worthwhile pointing out that rules 12 and 14 can be implemented by utilising existing dictionaries and spell checkers, while rule 16 can be implemented by integrating deeper language tools such as parsers and chunkers, a task for future work.

Table 5 summarises the human evaluation results of fluency and adequacy by the MT systems. Comparing the control and treatment groups, we can see the fluency and adequacy for the MT outputs by the treatment group, **TR** and **TS**, are almost equal to or slightly higher than those by the control group, **CR** and **CS**. More notable is that the rewritten versions of ST, regardless of the help by the system, showed much higher MT quality than the original ST, which demonstrated our selected CL rule sets were indeed effective in improving machine translatability.

The oracle STs in which CL violations were corrected as much as possible, not surprisingly, exhibited the best MT quality. The adequacy scores achieved 3.46 for RBMT and 3.34 for SMT, well surpassing the score of '3: Much of the meaning correctly expressed'. The oracle scores can be regarded as the upper bound of the MT quality when our CL is properly applied. To achieve this point, further support for writers is needed.

6.2. Efficiency

Table 6 shows the results for the efficiency measures. Time per correction indicates the average time taken to correct one violation. We can observe that the treatment group corrected violations 30% faster than the control group. An independent t-test also found a significant difference in scores between the two groups ($t = 2.826, df = 10, p = .018$). This result demonstrates that our system greatly enhanced the efficiency of checking for and correcting violations.

6.3. Satisfaction

Finally, we look at the results of the two usability questionnaires and the follow-up interviews.

ASQ (satisfaction with the task) revealed no statistically significant difference between the control group and the treatment group, nonetheless we can see that for all

	Treatment (with system)							Control (without system)						
Q.	TR1	TR2	TR3	TS1	TS2	TS3	Mean	CR1	CR2	CR3	CS1	CS2	CS3	Mean
1	5	3	6	4	3	6	4.5	4	2	3	5	6	3	3.8
2	5	4	6	5	6	4	5.0	3	2	3	5	6	5	4.0
3	6	5	7	4	5	5	5.3	6	6	4	5	5	3	4.8

Table 7. Result of Questionnaire ASQ (satisfaction with the task)

Q.	TR1	TR2	TR3	TS1	TS2	TS3	Mean
1	4	3	5	3	5	4	4.0
2	1	2	1	2	2	2	1.7
3	5	3	5	2	4	4	3.8
4	3	4	1	4	4	3	3.2
5	2	4	5	2	5	4	3.7
6	4	2	1	2	1	2	2.0
7	4	4	5	2	4	4	3.8
8	1	2	1	3	2	2	1.8
9	3	3	5	3	5	4	3.8
10	4	3	1	4	4	2	3.0
Score	70	68	100	54	80	78	75.0

Table 8. Result of Questionnaire SUS (satisfaction with the system)

three questions the mean scores of the treatment group are higher than those of the control group. It is also evident that, while there are only two negative answers (i.e., Likert scale of 1–3) from the treatment group, from the control group there are seven. This suggests that participants assisted by the system were generally satisfied with the task completion.

The SUS results (satisfaction with the system) pertain only to the treatment group. To calculate an overall SUS score ranging from 0 to 100 in 2-point increments, we inverted the scale of even-numbered questions from 1–5 to 5–1, and then doubled the sum of all the scores (see the bottom row of Table 8). The higher the score, the more usable the system was judged to be. The mean score is 75.0, which is reasonably high.

It is important to note that most participants agreed with Question 4 ('I think that I would need the support of a technical person to be able to use this system') and Question 10 ('I needed to learn a lot of things before I could get going with this system') with scores of 3–5. Both questions relate to the 'learnability' of the system. The follow-up interview results also revealed that some participants were unable to use the various support functions, such as suggestions of alternative expressions. Moreover, we found that some participants failed to correct proscribed terms highlighted in blue, even though they recognised them, simply because they forgot what the blue highlighting indicated. To make the system more effective, we need to provide more detailed user instructions and further simplify the interface.

7. Conclusions and Future Work

We have presented an experiment to assess the usability of a CL authoring assistant developed to support non-professional writers in checking conformity to CL

rules and terminology. Based on the ISO definition of usability, we assessed three aspects: effectiveness, efficiency and user satisfaction. Comparing two groups of participants—respectively, with and without the help of the system—we reached the following conclusions:

- The system helped reduce rule violations by about 9% (**effectiveness**).
- The system helped reduce the time taken to correct violations by more than 30% (**efficiency**).
- Participants were generally satisfied with the system, although some did not find the functions and interface easy to learn (**satisfaction**).

Our system now implements optimal CL rule sets individually tailored to two MT systems and the STs written in accordance with the rule sets proved to greatly improve machine translatability. The usability evaluation demonstrates that the system significantly enhances the efficiency of CL authoring by non-professional writers. This opens the promising prospect of practical joint deployment of CL and MT in real world scenarios.

The MT evaluation results of the oracle ST suggested that there is still room for improvement in MT quality. In future research, we plan to utilise existing language resources and tools to implement the remaining CL rules and so further assist authors in eliminating CL violations. We will also improve the interface and user documentation so that users take effective advantage of the full range of available functions.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 16J11185 and by the Research Grant Program of KDDI Foundation, Japan.

Bibliography

Aikawa, Takako, Lee Schwartz, Ronit King, Monica Corston-Oliver, and Carmen Lozano. Impact of Controlled Language on Translation Quality and Post-Editing in a Statistical Machine Translation Environment. In *Proc. of MT Summit*, pages 1–7, 2007.

Alabau, Vicent, Luis A. Leiva, Daniel Ortiz-Martínez, and Francisco Casacuberta. User Evaluation of Interactive Machine Translation Systems. In *Proc. of EAMT*, pages 20–23, 2012.

Bernth, Arendse and Claudia Gdaniec. MTranslatability. *Machine Translation*, 16(3):175–218, 2001.

Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi. Findings of the 2015 Workshop on Statistical Machine Translation. In *Proc. of WMT*, pages 1–46, 2015.

Brooke, John. SUS: A Quick and Dirty Usability Scale. In Jordan, P. W., B. Thomas, B. A. Weerdmeester, and A. L. McClelland, editors, *Usability Evaluation in Industry*, pages 189–194. Taylor and Francis, London, 1996.

Castilho, Sheila, Sharon O'Brien, Fabio Alves, and Morgan O'Brien. Does Post-editing Increase Usability? A Study with Brazilian Portuguese as Target Language. In *Proc. of EAMT*, pages 183–190, 2014.

Doherty, Stephen and Sharon O'Brien. A User-Based Usability Assessment of Raw Machine Translated Technical Instructions. In *Proc. of AMTA*, 2012.

Doherty, Stephen and Sharon O'Brien. Assessing the Usability of Raw Machine Translated Output: A User-Centered Study Using Eye Tracking. *International Journal of Human Computer Interaction*, 30(1):40–51, 2013.

Isahara, Hitoshi. Translation Technology in Japan. In Chan, Sin-Wai, editor, *Routledge Encyclopedia of Translation Technology*, pages 315–326. Routledge, New York, 2015.

ISO. ISO 9241-210:2010 Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems, 2010.

Kittredge, Richard. Sublanguages and Controlled Languages. In Mitkov, Ruslan, editor, *Oxford Handbook of Computational Linguistics*, pages 430–437. Oxford University Press, Oxford, 2003.

Kuhn, Tobias. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, 40(1):121–170, 2014.

Mitamura, Teruko, Kathryn Baker, Eric Nyberg, and David Svoboda. Diagnostics for Interactive Controlled Language Checking. In *Proc. of EAMT/CLAW*, pages 237–244, 2003.

Miyata, Rei, Anthony Hartley, Cécile Paris, Midori Tatsumi, and Kyo Kageura. Japanese Controlled Language Rules to Improve Machine Translatability of Municipal Documents. In *Proc. of MT Summit*, pages 90–103, 2015.

Miyata, Rei, Anthony Hartley, Cécile Paris, and Kyo Kageura. Evaluating and Implementing a Controlled Language Checker. In *Proc. of CLAW*, pages 30–35, 2016.

Nyberg, Eric, Teruko Mitamura, and Willem-Olaf Huijsen. Controlled Language for Authoring and Translation. In Somers, Harold, editor, *Computers and Translation: A Translator's Guide*, pages 245–281. John Benjamins, Amsterdam, 2003.

O'Brien, Sharon. Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets. In *Proc. of EAMT/CLAW*, pages 105–114, 2003.

O'Brien, Sharon and Johann Roturier. How Portable are Controlled Language Rules? In *Proc. of MT Summit*, pages 345–352, 2007.

Pym, Peter. Pre-editing and the Use of Simplified Writing for MT. In Mayorcas, Pamela, editor, *Translating and the Computer 10*, pages 80–95. Aslib, London, 1990.

Rascu, Ecaterina. A Controlled Language Approach to Text Optimization in Technical Documentation. In *Proc. of KONVENS*, pages 107–114, 2006.

Sauro, Jeff and James R. Lewis. *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann, Burlington, 2012.

Thicke, Lori. Improving MT results: A study. *Multilingual*, January/February:37–40, 2011.

Warburton, Kara. Developing Lexical Resources for Controlled Authoring Purposes. In *Proc. of LREC Workshop: Controlled Natural Language Simplifying Language Use*, pages 90–103, 2014.

Address for correspondence:
Rei Miyata
miyata@nuee.nagoya-u.ac.jp
Graduate School of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan