
Japanese Controlled Language Rules to Improve Machine Translatability of Municipal Documents

Rei Miyata rei@p.u-tokyo.ac.jp
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

Anthony Hartley a.hartley@rikkyo.ac.jp
Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo, 171-8501, Japan

Cécile Paris Cecile.Paris@csiro.au
CSIRO, Crn Vimiera and Pembroke Roads, Marsfield NSW, 2122, Australia

Midori Tatsumi tatsumi@rikkyo.ac.jp
Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo, 171-8501, Japan

Kyo Kageura kyo@p.u-tokyo.ac.jp
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

Abstract

We report on experiments to test the effectiveness of controlled language (CL) rules on texts from Japanese municipal websites. We compiled a set of rules by trial and error, systematically rewriting Japanese source texts and analysing the machine translation (MT) outputs. We then employed native English speakers with little knowledge of Japanese as human evaluators and tested the understandability and accuracy of the English machine translated text. Comparing the results of four MT systems showed that the effectiveness of CL rules varies depending on the particular MT systems. A preliminary selection of optimal rules for each system showed more than 15% increase in MT performance. We also assessed the readability of the Japanese source texts and discuss the way to make them compatible with the quality of the MT outputs.

1 Introduction

It has become increasingly important for Japanese municipalities to provide information in multiple languages. Although combining machine translation (MT) and human post-editing has increasingly proven to be an effective MT workflow, it may not be an option for the websites of municipalities, which must offer a large volume of information, such as daily announcements, procedural guidelines, emergency information and policy white papers, with the need for frequent updates. Having all these documents translated into a number of languages using human translators or post-editors is simply too expensive and time consuming. In such cases, raw machine translation output is often provided on-demand.

Translations between languages that have completely different structures such as Japanese and English are, however, still difficult compared to those between European languages or between Japanese and Korean, and in many cases the quality of raw outputs of MT is of low or even not understandable quality. When post-editing is not a standard option, and the machine translation system is used as a 'black box', controlled authoring remains the key to the improvement of translation quality. Controlled authoring involves a suite of technologies and environments that provide document templates, glossary management, grammar and style checkers. In the current study, we focus on controlled language (CL) rules as a starting point. We chose En-

glish as the target language since English is still an overwhelming choice in translating Japanese municipal texts, followed by Chinese, Korean and Portuguese (Carroll, 2010).

In our scenario, we assume that effective CL rules should (i) help to raise the quality of MT output, and (ii) not degrade the quality of the Japanese source texts. To achieve both of these requirements, we devised empirical procedures to formulate CL rules and conducted an evaluation to assess the efficacy of each rule in terms of machine translatability and Japanese readability.

The remainder of this paper is structured as follows. We describe related work in Section 2. In Section 3, we discuss how we constructed our CL rules, while Section 4 explains our experimental setup for human evaluation to assess the rules. We present our results accompanied with a discussion in Section 5. Section 6 concludes with implications for future work.

2 Related studies

Controlled (natural) language or C(N)L is ‘a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax, and/or semantics while preserving most of its natural properties’ (Kuhn, 2014, p.123). A number of English CL rule sets have been proposed to improve MT performance as well as human comprehension, and they have been actually implemented, mainly in technical documentation (e.g., Kamprath et al., 1998; Nyberg et al., 2003). Evaluation experiments on CL for MT have also been undertaken to assess machine translatability and post-editing productivity (Pym, 1990; Bernth and Gdaniec, 2001; O’Brien and Roturier, 2007; Aikawa et al., 2007), showing evidence of the effectiveness of CL.

In the case of Japanese CL, Nagao et al. (1984) devised a controlled grammar to syntactically disambiguate Japanese sentences. Yoshida and Matsuyama (1985) also advocated the need for Japanese CL in parallel with MT development and conducted pioneering work. Little practical implementation, however, resulted from their efforts. One of the reasons for this setback in Japanese CL is the difficulty in producing significant results because the machine translation task from Japanese to another major language such as English is hard, compared to the task of automatically translating between European language pairs such as English and French.

From the 1990’s to the 2000’s, a number of studies in Japanese computational linguistics have addressed the automatic rewriting (or pre-editing) and paraphrasing for MT (Kim and Ehara, 1994; Shirai et al., 1998; Inui and Fujita, 2004). However, fully automatic reformulation of natural language copes badly with highly complex expressions, which tend to require contextual information. The scope of variations in linguistic patterns was, therefore, limited.

More recently, Ogura et al. (2010) proposed ‘Simplified Technical Japanese’ (STJ) to improve MT performance. They constructed the rules by (i) identifying linguistic patterns which appeared related to MT output quality, (ii) defining putative rules, and (iii) conducting a preliminary assessment of their efficacy. They finally formulated the STJ rule set consisting of about 50 rules, while pointing out that it does not comprehensively cover the range of Japanese expression patterns. Another attempt to create Japanese CL is the on-going ‘Technical Japanese’ project, which focuses mainly on documents for business purposes. It published a ‘Patent Documents Writing Manual’, which consists of 31 rules designed to improve the clarity and translatability of patent texts (Matsuda, 2014).

While recent work focused mainly on technical documents for industry and business, Tsumi et al. (2013) formulated 22 CL rules for municipal website documents drawing on existing wisdom about technical writing in Japanese. Since the study chiefly referred to writing guidelines intended for human understandability or readability, the overall efficacy of the rules with MT was not significant. Thus, there remains much room for investigating other patterns impact-

ing on MT performance within the municipal domain.

O'Brien (2006) argued persuasively for the need to tune CL rule sets to language pair and MT system. The results of evaluation experiments by (Hartley et al., 2012) also suggested there were differences between rule-based machine translation (RBMT) and statistical machine translation (SMT) systems in terms of the impact of specific CL rules on their performance, although the MT systems as such were not the focus of their investigation. In short, it is still uncertain to what extent CL rules can be effectively generalised across MT systems or how much improvement can be attained if we compile rules specifically tuned to a given system.

Practical deployment of CL requires that the readability of the source text (ST) should not be compromised in the interests of making it more tractable for MT in order to improve target text (TT) quality. Indeed, the stated aim of Technical Japanese mentioned above is to improve both readability and machine tractability (Watanabe, 2010). However, when Hartley et al. (2012); Tatsumi et al. (2013) investigated the efficacy of CL rules with respect to both the readability of the Japanese source and the quality of the MT output, they observed for some rules a 'trade-off' between ST and TT quality, which needs further exploring.

3 Controlled language rules

3.1 CL formulation protocol

Given that municipal texts should serve readers of Japanese and English equally, any CL rule we propose here should contribute to the quality of MT outputs without degrading that of source texts. We assume that comparing original source texts and more machine translatable ones rewritten by humans enables us to derive insights into how to (re)write texts amenable to MT, while still guaranteeing source text readability, as long as human authors take charge of the whole rewriting process.

To materialise our assumption above, we devised the following empirical protocol to detect linguistic or textual features potentially effective for MT performance:

1. Rewrite a source text aiming at a better quality of MT outputs.
2. Record how the text was changed and assess the quality of the outputs.
3. Repeat steps 1 and 2, until achieving satisfactory quality of the MT outputs.

Examples of the detected linguistic features are shown in Table 1.

ST	MT	Change
電力会社に連絡、使用開始手続き完了後、ブレーカーのスイッチを入れます。	You can turn on the contact, the procedures after completion of the electric power companies.	[original sentence]
電力会社に連絡 します。使用開始の 手続きが完了した後、ブレーカーのスイッチを入れます。	Will contact the electric power company. Procedures for activation is complete, you turn on the breaker.	Split sentence Add 'します' Add 'の' Expand '完了後'
電力会社に連絡してください。使用開始の手続きが完了した後に、ブレーカーのスイッチを入れてください。	Please contact the electric power company. Please turn on the breaker after the procedure of the activation is completed.	Change 'ます' to 'てください' Add particle 'に'

Table 1: Example of rewriting the source text

3.2 CL formulation for municipal documents

To formulate CL rules for Japanese municipal documents through this protocol, we first extracted 100 original Japanese sentences from municipal websites and one of the authors con-

ducted the above protocol using three MT systems, i.e., one RBMT system (TransGateway¹) and two SMT systems (Google Translate² and Minnano Jido Hon'yaku³). We then summarised the linguistic and textual features expected to have an impact on MT quality and classified them into five categories: Mood/Modal, Structural, Lexical, Textual/Orthographical, and Terminological. In this study, we focus on Structural, Lexical, and Textual/Orthographical categories. We adopted a total of 38 features (Table 2) which had not been covered in the set of 22 CL rules proposed by (Tatsumi et al., 2013) and formulated 38 CL rules to regulate these features, such as 'Avoid using multiple verbs in a sentence' (rule 1).

In this phase, we did not take the variability of the systems into account, and extracted a wide range of features which we assumed could improve MT performance. These rules are not necessarily effective for every MT system. In formulating them, we observed that a few rewriting rules intended for one MT system had no effect or even a negative effect on the others. We thus conducted a quantitative human evaluation to determine which rules were effective for which MT system(s).

Structural	
1. multiple verbs in a sentence	20. omission
2. lack of subject	21. suffix
3. lack of object	22. particle Made (まで)
4. connection	23. particle De (で)
5. particle Ga (が) for object	24. particle No (の) to mean 'by' or 'from'
6. enumeration A-Mo, B-Mo (A も、B も)	25. per A
7. Te-kuru (てくる) / Te-iku (ていく)	26. particle Te (て)
8. inserted adverbial clause	27. if particle To (と)
9. ending clause with noun	28. particle He-Ha (へは)
10. Sahen-noun ⁴ + auxiliary verb Desu (です)	29. particle Ni-Ha (には)
11. attributive use of Shika-Nai (しか-ない)	30. particle No-Ka (のか)
12. verb + You-ni (ように)	31. demonstrative pronoun (ko-so-a-do)
13. A or not	32. particle Ni (に)
14. Sahen-noun + Wo (を) + Suru (する)	Textual/Orthographical
15. Sahen-noun + Sare-ru (される)	33. Japanese Kana / Chinese Kanji
Lexical	34. bullet mark
16. particle Nado (など/等)	35. machine dependent character
17. giving and receiving verb	36. punctuation (sentence separation)
18. redundant word	37. square bracket
19. compound word	38. wave dash (~)

Table 2: A list of features to be regulated

4 Experimental setup

The aim of the evaluation was (1) to gauge how effective our CL rules are to different MT systems, and (2) to investigate whether the rules which contribute to MT performance also maintain source readability. Using texts from a municipal website and four MT systems, we assessed these two parameters through human evaluation.

¹<http://www.kodensha.jp/platform/>

²<https://translate.google.com>

³<https://mt-auto-minhon-mlt.ucri.jgn-x.jp>

⁴A Sahen-noun is a noun which can be connected to the verb Suru (する) and act as a verb.

4.1 Data

We extracted 11,075 sentences with no more than 70 characters from the Toyohashi City website⁵ in five categories (public information, Q&A, department information, news articles, and topical issues), and selected sentences violating our 38 rules – four sentences for each rule – resulting in a total of 152 Japanese-Original (JO) sentences. One of the authors rewrote all 152 sentences in accordance with each rule, so generating 152 Japanese-Rewritten (JR) sentences. We used four MT systems, two commercial RBMT systems (The Hon'yaku⁶ and TransGateway, hereafter, systems A and B) and two freely available SMT systems (Google Translate and Minnano Jido Hon'yaku, hereafter, systems C and D), without user dictionaries or any sort of customisation, to translate the 152 JO and 152 JR sentences into English. The result was 1,216 machine translated sentences: 152 sentences for each label AO (system A-Original), AR (system A-Rewritten), and so on – BO, BR, CO, CR, DO and DR.

4.2 MT quality evaluation

Our main interest in evaluating the MT quality was to assess whether or not the translation was understandable in terms of practical use of information. We followed a simple method proposed by (Tatsumi et al., 2013), which focuses on understandability at an acceptable level, disregarding grammatical and lexical errors as long as they do not impair the reader's comprehension.

Questionnaire design

In order to find out whether or not an MT output was judged understandable and, if so, whether or not the reader's understanding was in fact correct, we adopted a two-step evaluation method. In Step 1, we showed the judges an MT output without telling them that it was the result of MT, and asked them to indicate how well they understood the text, and how much effort was required to understand it, by selecting one of the following options:

- [1] *I understood fully what this sentence is saying, after reading it once.*
- [2] *I understood fully what this sentence is saying, after reading it more than once.*
- [3] *I understood partially what this sentence is saying, after reading it more than once.*
- [4] *I have no idea what this sentence is saying even after reading it more than once.*

In Step 2, the judges were shown a human translation (HT) corresponding to the MT output shown in Step 1 as an 'alternative translation'. The question asked at that point differed depending on the answer to the question in Step 1. If [1] or [2] had been selected, the judges were asked to indicate how close the meaning of the new sentence was to the first sentence (the MT output) by selecting one of the following options.

- [5] *Exactly the same meaning*
- [6] *Mostly the same meaning*
- [7] *Partly the same meaning*
- [8] *Completely different meaning*

Considering that the judge's memory from Step 1 might not last long enough to compare their understanding at Step 2, we showed the MT output again here. At the same time, in order to discourage direct comparison between the two texts, the judges were asked to compare only the general meaning, not focusing on the difference in word choice.

If [3] or [4] had been selected at Step 1, i.e., when the MT output was not understandable, it was of little use to know whether their understanding was correct or not. Instead, we needed to know if it was because of the bad quality of the MT or a problem with the content itself. The judges were shown the HT as an alternative. They were then asked to indicate how much of it

⁵<http://www.city.toyohashi.lg.jp>

⁶http://pf.toshiba-sol.co.jp/prod/hon_yaku/

they understood and how much effort was required to do so, with the same options as at Step 1, i.e., [9]=[1], [10]=[2], [11]=[3], [12]=[4]. In this case, the first sentence (the MT output) was no longer shown.

Implementation

We employed 24 judges, all adult English native speakers. They were all living in Japan, yet had little knowledge of Japanese. The judges were thus highly representative of the intended readers of the targets texts.

Each judge was asked to evaluate 152 MT sentences that corresponded to the 152 Japanese source sentences but were a mix of translations from eight sources (AO, AR, BO, BR, CO, CR, DO, DR). The evaluation was conducted online.

4.3 Japanese readability evaluation

The Japanese source texts and their manually rewritten versions were both of understandable quality from the outset. In order to deal with the subtle differences in the sentence readability, we adopted the following evaluation method following (Hartley et al., 2012): the judges were presented with pairs of sentences JO and JR, whose ordering was randomised. Each judge was asked to evaluate each sentence of the pair on a four-point scale: *easy to read*; *fairly easy to read*; *fairly difficult to read*; *difficult to read*. We instructed the judges in advance not to focus on the minute grammatical exactness, but to judge the ease of reading.

For this task, we recruited three Japanese native speaker university graduate students as judges. Each judge evaluated 152 pairs of Japanese sentences, i.e., all 152 JO and 152 JR.

5 Results and discussions

5.1 MT quality

Overall result

Firstly, we classified the results into four categories (Table 3). The numbers in square brackets correspond to the choices in each step as described in Section 4.2. An MT output is considered useful only when either [5] or [6] was chosen at Step 2 (MT-Useful) of the evaluation task described above. Those classed [7] or [8] are the dangerous instances, as they mean that the MT output is understandable while conveying inaccurate information (MT-Inaccurate). In the case of [9] or [10], the MT output is not intelligible and is thus useless. Yet it is less dangerous than [7] and [8]. Finally, [11] and [12] mean that even the human translation was not understandable. This may suggest that the problem lies in the content which requires some domain knowledge or contextual information for full understanding.

Selected option	Category	Interpretation
[5][6]	MT-Useful	The reader understood the MT output and their understanding was correct
[7][8]	MT-Inaccurate	The reader understood the MT output but their understanding was not correct
[9][10]	MT- Unintelligible	The reader did not understand the MT output, but they understood the corresponding HT
[11][12]	HT-Unintelligible	The reader did not understand either the MT or the corresponding HT

Table 3: Result categories

Table 4 shows the percentage of judgements that fell into each category. Overall, applying the CL rules increased the percentage of [MT-Useful] by around 3–4% for three systems, A, B

and C, while system D shows a slight decrease. Although our focus is not on the analysis of the rules as a whole but on the diagnosis of the efficacy of each rule, it is worth noting that only about 30% of the MT outputs were deemed useful even after applying each CL rule.

Label	MT-Useful	MT-Inaccurate	MT-Unintelligible	HT-Unintelligible
AO	27.4%	4.6%	62.1%	5.9%
AR	30.9%	5.5%	58.8%	4.8%
BO	23.2%	5.0%	66.0%	5.7%
BR	27.2%	5.7%	63.4%	3.7%
CO	26.5%	3.9%	64.7%	4.8%
CR	30.0%	6.8%	58.3%	4.8%
DO	27.0%	6.4%	61.0%	5.7%
DR	26.3%	6.8%	60.1%	6.8%

Table 4: Overall results of MT quality

A comparison between the four MT systems shows the different effects of the CL rules on the four MT systems. While in the case of systems A and B the application of the CL rules mostly resulted in decreasing the number of [MT-Unintelligible] and increasing that of [MT-Useful], in the case of system C, it resulted in a notable increase in the number of [MT-Inaccurate]. The following example shows system C generating an [MT-Inaccurate] output after the application of rule 14.

Rule 14: avoid using Sahen-noun + Wo (を) + Suru (する)

- CO:** At the venue, the video for each of the agenda is projected on a large screen, was the description of the agenda in an easy-to-understand local residents.
- CR:** At the venue, the video for each of the agenda is projected on a large screen, I explained the agenda in an easy-to-understand local residents.
- HT:** At the venue, images related to each topic were projected on a large screen and local residents explained the topics in an easy to understand manner.

We changed only ‘説明-を-しました’ (‘gave an explanation’) into ‘説明-しました’ (‘explained’) in the source. System C then incorrectly (and unexpectedly) inferred the subject ‘I’ though the true subject ‘地域住民’ (‘local residents’) is present in the source, and generated an understandable but misleading output. This kind of unpredictable change not directly related to the feature in question tends to occur with SMT systems.

Importantly, the score of around 5% in the [HT-Unintelligible] category shows that even human translated sentences are sometimes not understandable, implying a fundamental difficulty with evaluating at the sentence-level. We had instructed the human translator to translate the source sentences without adding explanations or suppressing information to make them comparable with the MT outputs. Moreover, we did not show the judges the context of each sentence, so the occasional failure of judges to understand the human translation even though it was grammatically correct could be due to a lack of knowledge of the Japanese municipal domain.

Generally applicable CL rules

To diagnose the effectiveness of each CL rule for different MT systems in detail, we focused on the [MT-Useful] cases, since an increase in [MT-Useful] mostly corresponds to an decrease in [MT-Unintelligible], except for system C, which showed a significant rise in [MT-Inaccurate] together with [MT-Useful]. We counted the judgements that fell in this category for each rule and calculated the improvement (or degradation) scores as a percentage, emphasising the improvements in bold (Table 5).

No	A	B	C	D	No	A	B	C	D
1	-25.0	8.3	-8.3	-16.7	20	16.7	16.7	8.3	0.0
2	25.0	25.0	25.0	-16.7	21	8.3	0.0	16.7	-33.3
3	8.3	33.3	0.0	-8.3	22	25.0	-16.7	16.7	-25.0
4	0.0	16.7	8.3	16.7	23	0.0	16.7	-8.3	-8.3
5	0.0	8.3	-8.3	-16.7	24	16.7	-8.3	0.0	-8.3
6	8.3	-16.7	0.0	-16.7	25	8.3	16.7	16.7	8.3
7	-16.7	0.0	8.3	25.0	26	-16.7	8.3	0.0	0.0
8	33.3	-16.7	8.3	16.7	27	-16.7	16.7	-8.3	-25.0
9	-8.3	0.0	0.0	25.0	28	25.0	-16.7	16.7	25.0
10	16.7	16.7	25.0	16.7	29	0.0	25.0	-8.3	-33.3
11	-8.3	0.0	25.0	8.3	30	-8.3	16.7	-16.7	-25.0
12	8.3	-8.3	8.3	16.7	31	0.0	0.0	16.7	0.0
13	16.7	25.0	8.3	8.3	32	0.0	0.0	-33.3	0.0
14	-8.3	-25.0	8.3	0.0	33	-25.0	8.3	-25.0	-16.7
15	-16.7	8.3	-8.3	25.0	34	8.3	-25.0	25.0	0.0
16	0.0	-16.7	8.3	8.3	35	16.7	25.0	33.3	25.0
17	0.0	25.0	0.0	-8.3	36	33.3	-8.3	-8.3	-25.0
18	0.0	0.0	8.3	8.3	37	16.7	16.7	0.0	16.7
19	-16.7	8.3	-8.3	16.7	38	8.3	-33.3	-16.7	-8.3

Table 5: Improvement in [MT-Useful] category

We can see that four rules – 10, 13, 25 and 35 – have a positive effect on all four MT systems. Some examples of translations of the original sentences and their rewrites are listed below.

Rule 10: avoid using Sahen-noun + Desu (です)

AO: An admission ticket is the 10:00 a.m. sales start on Fri., April 22.

AR: An admission ticket starts sale at 10:00 a.m. on Fri., April 22.

BO: An admission ticket is sales starting on Friday, April 22 at 10:00am.

BR: An admission ticket begins to sell it at 10:00am on Friday, April 22.

CO: Admission ticket is sales start at April 22 (gold) 10 am.

CR: Tickets will start April 22 (Friday) at 10 am selling.

DO: Admission ticket is April 22 (Kim) 10 a.m. launch.

DR: Tickets will be on sale at 10 a.m. on April 22 (Kim).

HT: Ticket sales will start at on Friday, April 22 at 10:00 AM.

In this case, we rewrote Sahen-noun + Desu construction ‘開始-です’ into Sahen-noun + Suru construction ‘開始-します’, which resulted in more natural expressions in the MT outputs.

Rules 2, 4, 8, 12, 20, 28 and 37 show a positive effect on three systems and can also be regarded as generally applicable rules. We provide below examples of the application of one of these rules.

Rule 2: avoid omitting subject

AO: A home and the community are places where a child spends much time daily, and study that it is various in a life.

AR: A home and the community are places where a child spends much time daily, and a child studies that it is various in a life.

- BO:** A house and an area are the place where a child spends much time daily, and various things will be learned in the life.
- BR:** A house and an area are the place where a child spends much time daily, and a child will learn various things in the life.
- CO:** Home and regions, children are routinely spend place a lot of time, you will learn a variety of things in life.
- CR:** Home and regions, children are routinely spend place a lot of time, children will learn a variety of things in life.
- HT:** Homes and communities are places where children spend a lot of time every day, and where they learn many things about life.

In Japanese writing, subjects tend to be omitted. Humans normally have no problems inferring the subjects from the context. In this case, for example, ‘children’ (子ども) or ‘they’ (彼ら) can be inferred in the latter clause. MT systems, however, often have difficulties in dealing with null-subject expressions. This is evidenced in AO, BO and CO above: system A did not insert a subject, but this caused a disagreement between the subject ‘a child’ and the verb ‘study’; system B adopted a passive construction ‘will be learned’; system C wrongly inserted ‘you’ as a subject. We can see that inserting ‘子ども’ as a subject in the original Japanese sentence enhanced the performance of the three systems (see AR, BR and CR).

MT-dependent CL rules

As Table 5 clearly demonstrates, it is also important to note that the effectiveness of each CL rule is variable. For example, rule 11 had a positive effect on the output of systems C and D. We look at it in more detail below.

Rule 11: avoid using attributive use of Shika-Nai (しかない)

- AO/AR:** Although it is a plant of a greenhouse, please look at this flower that makes a flower bloom only at this time once.
- BO/BR:** It’s a plant in a greenhouse, but please see this flower which makes a flower bloom only at this time once by all means.
- CO:** Although it is greenhouse plants, please come visit once this flower only at this time does not bloom.
- CR:** Although it is greenhouse plants, please come visit once the flowers bloom this time only to flower.
- DO:** Is a plant of the greenhouse, you take the time to peruse this flower only during this period not bloom.
- DR:** Is a plant of the greenhouse, you take the time to peruse the flowers that bloom only in this time of the year.
- HT:** Among the greenhouse plants, please be sure to take a look at this flower, which only blooms during this time of year.

In this case, we rewrote the attributive expression Shika-Nai (しかない) into another attributive particle Dake (だけ). For the RBMT systems A and B, both attributive patterns were linguistically processed in the same manner, using the adverb ‘only’, and this rule shows no improvement. On the other hand, for the SMT systems C and D, Shika-Nai (しかない) is dealt with as a kind of negative construction, which leads to an unnecessary negation in the output. Thus, regulating it is effective in improving MT quality.

While this particular rule triggered differing reactions in the RBMT versus SMT systems, we could not discern a regular correlation between system architectures and the effectiveness of CL rules. Instead, the results showed the idiosyncrasy of each system.

It should also be added that, in our experimental design to examine the effectiveness of each rule, we rewrote only that part of the sentence which was directly related to the applicable rule. This caused two major issues in our results. First, applying a single CL rule did not necessarily address the quality of the entire sentence, and thus did not contribute to as great an increase in [MT–Useful] as we had expected. In particular, there are cases where rewriting the source text did not change the MT outputs at all, as shown in the examples of rule 11 above. Even worse, rule 32 produced no improvement in any of the four MT systems. Second, in some cases a CL rule successfully brought a local improvement in the quality of the translation, yet other critical mistranslations (which, we observed, often stemmed from technical terms and proper nouns) overrode its positive effect and thus led to an overall low grade. This suggests that piece-by-piece modifications are not always enough to improve MT output quality.

Optimal rule set for each system

Table 6 shows how much improvement we could see if we select effective CL rules for each MT system. We preliminarily selected those rules which produced an increase in [MT–Useful] according to Table 5, i.e., 18 rules for system A, 19 for B, 19 for C and 16 for D, and summarised the results as in Table 4.

For all systems, [MT–Useful] category increases by more than 15 % with no or little increase in the [MT–Inaccurate] category. This result clearly indicates the necessity of tailoring the selection of rules to a particular MT system. In addition, further improvement can be expected if all applicable optimal rules are applied to a given sentence.

Label	MT–Useful	MT–Inaccurate	MT–Unintelligible	HT–Unintelligible
AO	20.6%	3.9%	68.9%	6.6%
AR	37.3%	3.9%	55.7%	3.1%
BO	20.6%	4.4%	70.6%	4.4%
BR	38.2%	3.5%	54.4%	3.9%
CO	21.1%	4.8%	70.2%	3.9%
CR	36.4%	5.3%	53.1%	5.3%
DO	17.7%	7.8%	69.3%	5.2%
DR	34.4%	6.8%	53.1%	5.7%

Table 6: Overall results of MT quality (after optimal rules were selected)

5.2 Japanese readability

The judgements of readability can be separated into two levels: acceptable and unacceptable. We defined the first two options of the question – *easy to read* and *fairly easy to read* – as acceptable and the other two options – *fairly difficult to read* and *difficult to read* – as unacceptable, on the assumption that the gap between the former and the latter is significant from the point of reading ease by humans.

Columns JO and JR in Table 7 show the percentages of judgements categorised as acceptable, and column JR-JO shows the improvement or deterioration. The higher the score in JR-JO, the greater the reading ease of JR compared to JO. Given our requirement that the CL should not degrade source text readability, figures greater than or equal to 0.0 (%) in JR-JO are highlighted in bold.

As a whole, 23 out of 38 the CL rules improved or at least retained the quality of the source text. In particular, rules 2, 18 and 26 were effective with more than 40% improvement in readability. According to rule 18 (avoid using redundant word), for instance, we deleted redundant expressions such as ‘ものとする’ and ‘こととする’. After this rewrite human evaluators judged JR to be more readable than JO. These kinds of periphrastic expressions are commonly used in

No	JO	JR	JR-JO	No	JO	JR	JR-JO
1	91.7	75.0	-16.7	20	75.0	50.0	-25.0
2	50.0	91.7	41.7	21	100.0	50.0	-50.0
3	58.3	91.7	33.3	22	100.0	66.7	-33.3
4	83.3	58.3	-25.0	23	91.7	75.0	-16.7
5	58.3	91.7	33.3	24	66.7	91.7	25.0
6	91.7	50.0	-41.7	25	66.7	91.7	25.0
7	83.3	100.0	16.7	26	50.0	91.7	41.7
8	75.0	75.0	0.0	27	83.3	91.7	8.3
9	75.0	58.3	-16.7	28	66.7	66.7	0.0
10	58.3	91.7	33.3	29	66.7	83.3	16.7
11	83.3	75.0	-8.3	30	91.7	83.3	-8.3
12	83.3	91.7	8.3	31	58.3	58.3	0.0
13	91.7	75.0	-16.7	32	100.0	66.7	-33.3
14	75.0	83.3	8.3	33	75.0	100.0	25.0
15	75.0	83.3	8.3	34	83.3	100.0	16.7
16	66.7	66.7	0.0	35	83.3	66.7	-16.7
17	66.7	100.0	33.3	36	75.0	91.7	16.7
18	58.3	100.0	41.7	37	66.7	66.7	0.0
19	83.3	75.0	-8.3	38	100.0	41.7	-58.3

Table 7: Improvement in Japanese readability

municipal documents in Japan.

In contrast, rules 6, 21 and 38 resulted in more than 40% degradation in readability. Avoid using suffix (rule 21) and wave dash ‘~’ (rule 38) introduce redundancy by replacing the feature with a longer sequence of words, such as replacing ‘午後 1 時 ~ 4 時’ (‘1:00–4:00 PM’) with ‘午後 1 時から 4 時まで’ (‘from 1:00 to 4:00 PM’).

More detailed analysis revealed there are some rules for which the evaluations of their effectiveness differed depending on the more specific features of sentences. For example, rewriting ‘記念品代相当分’ (‘an appropriate amount of money toward the commemorative item’) as ‘記念品代に相当する分’ (‘an appropriate amount of money corresponding to the commemorative item’) according to rule 19 (avoid using compound noun) improves readability, while rewriting ‘市民提供資料’ (‘materials provided by residents’) as ‘市民が提供した資料’ (‘materials that residents provided’) based on the same rule degrades readability.

5.3 Compatibility of machine translatability and Japanese readability

Comparing the results of the MT quality (Table 5) and Japanese readability (Table 7), we now discuss the compatibility of the two requirements. Focusing on the generally applicable rules, we see that rules 2, 8, 10, 12, 25, 28 and 37 improved or retained Japanese readability. In particular, rule 2 and 10 greatly improved both machine translatability and Japanese readability. There are, however, some rules which are effective for MT quality in general but have an adverse effect on human readability, such as rules 4, 13, 20 and 35. Rule 20, for instance, produced a better MT output for systems A, B and C, but degraded the readability of the source text. We give an example below.

Rule 20: avoid omission

JO: 月・水・金曜日 の午前 9 時から午後 4 時まで開設しており、3 月末まで開設しています。

JR: 月曜日・水曜日・金曜日 の午前 9 時から午後 4 時まで開設しており、3 月末まで開設しています。

BO: It's established from a month and 9:00am of water and Friday to 4:00pm and it's established until the end of March.

BR: It's established from 9:00am of Monday, Wednesday and Friday to 4:00pm and it's established until the end of March.

HT: It will be open from 9:00 AM to 4:00 PM on Mondays, Wednesdays, and Fridays until the end of March.

System B failed to recognise '月・水・金曜日' as an elliptic expression and literally translated '月' into 'month', '水' into 'water', and '金曜日' into 'Friday' (BO). Restoring an omitted element '曜日' according to rule 20 helped the system deal correctly with this expression (BR). In contrast, human evaluators preferred JO to JR in terms of readability. This is no doubt because too much complementation made the sentence longer and hindered reading. The point here is that we need to find a way to meet both requirements of the MT quality and human readability. In the case of rule 20, for instance, we should keep JO for human readers, but employ pre-processing to produce JR for MT purposes.

6 Conclusion and future work

We have proposed an empirical protocol for formulating CL rules with a view to improving the machine translatability of Japanese municipal documents. Focusing on Japanese to English translation and using 100 sentences from municipal websites, we derived 38 CL rules different from the 22 rules that had previously been formulated on the basis of collective wisdom about technical writing.

We assessed the efficacy of the rules on municipal documents, with respect to both MT quality and source text readability. We identified a total of 11 rules which are effective for at least three MT systems. Since previous studies could identify few 'general' rules (Hartley et al., 2012; Tatsumi et al., 2013), this result encourages us to pursue our CL formulation protocol. In addition, we consider the protocol described in Section 3.1 to be generalisable to other language pairs and text domains, even though the CL rules in this study are formulated particularly for Japanese-to-English translation of municipal documents.

Interestingly, the effectiveness of CL rules was not shown to align with architectural differences between RBMT and SMT. This implies that we need to tune CL rule sets at the level of particular MT systems rather than at the level of MT types. In addition, a preliminary selection of optimal rules for each system achieved a greater than 15% increase in the [MT-Useful] category.

On the other hand, the results of the Japanese readability assessment showed that about two thirds of the CL rules improved or at least maintained source text readability. To achieve both machine translatability and human readability, it is important to serve different texts to humans and machines. We identified that degradations in readability for humans often correlate with redundancy generated by the rules. Thus an effective solution would be, for instance, to unpack the elliptic expressions and insert linguistic elements such as subjects only for MT. Moreover, this pre-processing for MT can be automated to some extent by employing existing pre-editing methods (e.g., Shirai et al., 1998), which can reduce the cost of implementing CL rules. In a future study, we will assess the total productivity of controlled authoring and translation in combination with post-editing.

Acknowledgement

This work was supported by the Research Grant Program of KDDI Foundation, Japan. The MT system J-SERVER Professional TransGateway V3 was offered by Kodensha Co.

References

- Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., and Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. In *Proceedings of the Machine Translation Summit XI*, pages 1–7, Copenhagen, Denmark.
- Berth, A. and Gdaniec, C. (2001). MTranslatability. *Machine Translation*, 16(3):175–218.
- Carroll, T. (2010). Local government websites in Japan: international, multicultural, multilingual? *Japanese Studies*, 30(3):373–392.
- Hartley, A., Tatsumi, M., Isahara, H., Kageura, K., , and Miyata, R. (2012). Readability and translatability judgments for ‘Controlled Japanese’. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, pages 237–244, Trento, IT.
- Inui, K. and Fujita, A. (2004). A survey on paraphrase generation and recognition. *Journal of Natural Language Processing*, 11(5):151–198. (in Japanese).
- Kamprath, C., Adolphson, E., Mitamura, T., and Nyberg, E. (1998). Controlled language for multilingual document production: Experience with Caterpillar Technical English. In *Proceedings of the 2nd International Workshop on Controlled Language Applications*, pages 51–61, Pittsburgh, PA.
- Kim, Y.-B. and Ehara, T. (1994). An automatic sentence breaking and subject supplement method for J / E machine translation. *IPSJ Journal*, 35(6):1018–1028.
- Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1):121–170.
- Matsuda, S. (2014). Efforts for Technical Japanese : Focusing mainly on the ‘Patent Documents Writing Manual’. *Journal of Information Processing and Management*, 57(6):387–394. (in Japanese).
- Nagao, M., Tanaka, N., and Tsujii, J. (1984). Support system for writing texts based on controlled grammar. *IPSJ SIG Technical Reports*, 1984(27):33–40. (in Japanese).
- Nyberg, E., Mitamura, T., and Huijsen, W.-O. (2003). Controlled language for authoring and translation. In Somers, H., editor, *Computers and the Translator*, pages 245–281. Benjamins, Amsterdam.
- O’Brien, S. (2006). Controlled language and post-editing. *Multilingual*, 17(7):17–19.
- O’Brien, S. and Roturier, J. (2007). How portable are controlled language rules? In *Proceedings of the Machine Translation Summit XI*, pages 345–352, Copenhagen, DK.
- Ogura, E., Kudo, M., and Yanagi, H. (2010). Simplified Technical Japanese: Writing translation-ready Japanese documents. *IPSJ SIG Technical Reports*, 2010-DD-78(5):1–8. (in Japanese).
- Pym, P. (1990). Pre-editing and the use of simplified writing for MT. In Mayorcas, P., editor, *Translating and the Computer 10*, pages 80–95. Aslib, London.
- Shirai, S., Ikehara, S., Yokoo, A., and Ooyama, Y. (1998). Automatic rewriting method for internal expressions in Japanese to English MT and its effects. In *Proceedings of the 2nd International Workshop on Controlled Language Applications*, pages 62–75, Pittsburgh, PA.
- Tatsumi, M., Miyata, R., Hartley, A., Kageura, K., and Isahara, H. (2013). Towards acceptable quality machine translation without post-editing for municipal websites: An evaluation of Japanese controlled language rules. In *Machine Translation Summit XIV: QTLA LaunchPad Workshop on Human-Centric Machine Translation and Evaluation*, Nice, France. (Slide available: <http://reimiyata.com/materials/mtsummit2013-ws.pdf>).

- Watanabe, T. (2010). Outline of the 'Technical Japanese' project: Activity for acceleration of patent technological information utilization. *Journal of Information Processing and Management*, 53(9):480–491. (in Japanese).
- Yoshida, S. and Matsuyama, A. (1985). Standardizing Japanese: Standardizing dependency relations and transformation rules. *IPSJ SIG Technical Reports*, 1985(31):1–6. (in Japanese).