The background of the cover is a detailed architectural drawing in white lines on a dark green background. It features various geometric shapes, including rectangles, circles, and arcs, some of which are filled with patterns of small hexagons or dots. The drawing appears to be a technical sketch of a building or a complex structure, with lines radiating from various points and intersecting to form a complex network of shapes.

Routledge Studies in Translation Technology

CONTROLLED DOCUMENT AUTHORING IN A MACHINE TRANSLATION AGE

Rei Miyata

Controlled Document Authoring in a Machine Translation Age

This book explains the concept, framework, implementation, and evaluation of controlled document authoring in this age of translation technologies. Machine translation (MT) is routinely used in many situations, by companies, governments, and individuals. Despite recent advances, MT tools are still known to be imperfect, sometimes producing critical errors. To enhance the performance of MT, researchers and language practitioners have developed controlled languages that impose restrictions on the form or length of the source-language text. However, a fundamental, persisting problem is that both current MT systems and controlled languages deal only with the sentence as the unit of processing. To be effective, controlled languages must be contextualised at the document level, consequently enabling MT to generate outputs appropriate for their functional context within the target document. With a specific focus on Japanese municipal documents, this book establishes a framework for controlled document authoring by integrating various research strands including document formalisation, controlled language, and terminology management. It then presents the development and evaluation of an authoring support system, MuTUAL, that is designed to help non-professional writers create well-organised documents that are both readable and translatable. The book provides useful insights for researchers and practitioners interested in translation technology, technical writing, and natural language processing applications.

Rei Miyata, PhD, is an Assistant Professor at the Graduate School of Engineering, Nagoya University, Japan.

Routledge Studies in Translation Technology

Series Editor: Chan Sin-wai

This cutting-edge research series examines translation technology and explores the relationships between human beings and machines in translating the written and spoken word. The series welcomes authored monographs and edited collections.

The Future of Translation Technology

Towards a World without Babel

Chan Sin-wai

The Human Factor in Machine Translation

Edited by Chan Sin-wai

Controlled Document Authoring in a Machine Translation Age

Rei Miyata

For more information on this series, please visit

www.routledge.com/Routledge-Studies-in-Translation-Technology/book-series/RSITT

Controlled Document Authoring in a Machine Translation Age

Rei Miyata

ROUTLEDGE

Routledge
Taylor & Francis Group

LONDON AND NEW YORK

First published 2021
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2021 Rei Miyata

The right of Rei Miyata to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

With the exception of the Preface, Chapter 1, and the Bibliography, no part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

The Preface, Chapter 1, and the Bibliography of this book are available for free in PDF format as Open Access from the individual product page at www.routledge.com. They have been made available under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 license

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Miyata, Rei, author.

Title: Controlled document authoring in a machine translation age / Rei Miyata.

Description: London ; New York : Routledge, 2020. |

Series: Routledge studies in translation technology |

Includes bibliographical references and index.

Identifiers: LCCN 2020020455

Subjects: LCSH: Machine translating. | Computational linguistics. |

Natural language processing (Computer science)

Classification: LCC P308 .M59 2020 | DDC 418/.020285—dc23

LC record available at <https://lcn.loc.gov/2020020455>

ISBN: 978-0-367-50019-1 (hbk)

ISBN: 978-1-003-04852-7 (ebk)

Typeset in Times New Roman
by Newgen Publishing UK

Contents

<i>List of figures</i>	vii
<i>List of tables</i>	viii
<i>List of abbreviations</i>	x
<i>Preface</i>	xi
<i>Acknowledgements</i>	xiii
PART I	
Research background	1
1 Introduction	3
2 Related work	12
PART II	
Controlled document authoring	57
3 Document formalisation	59
4 Controlled language	68
5 CL contextualisation	91
6 Terminology management	101
PART III	
MuTUAL: An authoring support system	125
7 System development	127
8 Evaluation of CL violation detection component	138
9 System usability evaluation	144

PART IV

Conclusion **167**

10 Research findings and outlook 169

Appendices 177

Bibliography 199

Index 215

Figures

1.1 Personal seal registration procedure (excerpted from the website of Shinjuku City)	5
1.2 Chapter organisation with research questions	10
2.1 Functional elements of research papers (Kando, 1997, p.3)	18
2.2 Machine translation work flow with human intervention	23
3.1 Functional elements of municipal procedural documents	62
3.2 Example of the functional elements in the procedure for personal seal registration (excerpted from a document in CLAIR)	63
3.3 Example of the functional elements in the procedure for personal seal registration (excerpted from a document in Hamamatsu)	63
3.4 Analysis of an existing municipal document using our DITA framework	65
3.5 Model example of the seal registration procedure	66
4.1 MT quality questionnaire – Step 2 (when [1] or [2] is selected in Step 1)	74
4.2 Source-readability questionnaire	83
5.1 Approaches to resolve incompatibilities between CL_{st} and CL_{tt} for Steps (* undesirable sentence)	93
5.2 Revised flow of pre- and post-translation processing for Steps (* undesirable sentence)	99
6.1 Term registration platform	105
6.2 The frequency spectrum of terms in municipal corpus (m : frequency class; $V(m, N)$: number of types with frequency m)	106
6.3 Growth curve of terminologies in municipal corpus	119
7.1 Modules of MuTUAL	128
7.2 Task topic template	130
7.3 CL authoring assistant, MT and back translation (BT)	131
7.4 Terminology check function	132
7.5 CL rule selection modal	133
7.6 CL guideline modal	134
7.7 Similar text search	136
9.1 User interface	147

Tables

2.1	Seven information types defined in Information Mapping (Horn, 1989, pp.110–111)	15
2.2	Controlled writing processes and support mechanisms	50
3.1	Examples of hierarchical levels of websites	60
3.2	The number of documents in each category of municipal-life information (the number of procedural documents is in parenthesis)	61
3.3	Specialisation of the DITA Task topic to municipal procedures	64
4.1	A list of technical writing based CL rules (CL-T)	70
4.2	Example of source-text rewriting	71
4.3	A list of rewriting trial based CL rules (CL-R)	72
4.4	Result categories	75
4.5	Overall results of MT quality (CL-T)	76
4.6	Overall results of MT quality (CL-R)	76
4.7	Improvement in [MT-Useful] category (CL-T)	78
4.8	Improvement in [MT-Useful] category (CL-R)	78
4.9	Overall results after optimal rules were selected (CL-T)	81
4.10	Overall results after optimal rules were selected (CL-R)	82
4.11	Improvement in Japanese readability (CL-T)	83
4.12	Improvement in Japanese readability (CL-R)	84
4.13	Selected optimal rules for two MT systems (common rules are shown in bold)	87
5.1	Linguistic specification	92
5.2	The number of MT outputs that realise the desired linguistic forms in the English translation before and after pre-translation processing	95
5.3	Example MT outputs before and after applying Rule 1 (inserted segment in brackets)	96
5.4	Example of remaining disconformity to CL _{tt}	96
5.5	Example MT outputs before and after applying Rule 2 ‘ <i>shiro</i> ’ (transformed segment in brackets)	97
5.6	Example MT outputs before and after applying Rule 2 ‘ <i>shinasai</i> ’ (transformed segment in brackets)	98

5.7	Example MT outputs before and after applying Rule 2 ‘ <i>shitekudasai</i> ’ (transformed segment in brackets)	98
6.1	Basic statistics of extracted sentences	103
6.2	The number of terms extracted and their occurrences in the corpus	106
6.3	The 20 most frequent terms in the corpus (before controlling)	107
6.4	Example of extracted bilingual term pairs	108
6.5	Criteria for defining preferred and proscribed terms	112
6.6	Examination of term variations	113
6.7	The basic statistics of controlled terminology	113
6.8	The 20 most frequent controlled terms occurred in the corpus	114
6.9	Population types $E[S]$ and coverage CR	118
6.10	Growth rate	120
6.11	Shift in the coverage ratio (%)	120
8.1	Results of the benchmark evaluation	140
9.1	Quantitative aspects to be measured	145
9.2	CL rules and implementation (with confidence scores)	146
9.3	After-scenario questionnaire (ASQ)	148
9.4	System usability scale (SUS)	149
9.5	Effectiveness for each condition	151
9.6	Correction rate for each rule (* implemented rule)	152
9.7	Result of MT quality evaluation (system B)	153
9.8	Result of MT quality evaluation (system D)	153
9.9	Example of ST and MT (system D) of different conditions	154
9.10	Result of ST quality evaluation	155
9.11	Time efficiency	155
9.12	Detailed edit log	156
9.13	Edit distance	157
9.14	Result of questionnaire ASQ (satisfaction with the task)	158
9.15	Result of questionnaire SUS (satisfaction with the system)	159
9.16	Text similarity between participants	161
9.17	Example of high-text-similarity sentence	162
9.18	Example of low-text-similarity sentence	163
9.19	Task time transition (time per sentence, in seconds)	164
9.20	Task time transition (time per character, in seconds)	164

Abbreviations

AEM	Automatic Evaluation Metric
API	Application Programming Interface
ASQ	After-Scenario Questionnaire
ATE	Automatic Term Extraction
BT	Back Translation
CL	Controlled Language
CLAIR	Council of Local Authorities for International Relations
DITA	Darwin Information Typing Architecture
EBMT	Example-Based Machine Translation
ESP	English for Specific Purposes
HT	Human Translation
IMRD	Introduction, Methods, Results and Discussion
ISO	International Organization for Standardization
MT	Machine Translation
NLG	Natural Language Generation
NLP	Natural Language Processing
NMT	Neural Machine Translation
OOV	Out-Of-Vocabulary
PE	Post-Edit
POS	Part-Of-Speech
RBMT	Rule-Based Machine Translation
SL	Source Language
SMT	Statistical Machine Translation
ST	Source Text
SUS	System Usability Scale
TL	Target Language
TM	Translation Memory
TT	Target Text

Preface

Every change in the paradigm of machine translation (MT) architectures has lifted the potential of MT technologies, evoking expectations, or dreams, of ‘human parity’ in general-purpose MT. Recent years have witnessed a significant improvement in MT performance due to the advent of neural MT, which makes use of a huge volume of text data to train a deep neural network model to generate target text from source text in a single, unified process. An increasing number of companies, governments and individuals have started to use MT tools to meet their various needs, many of which might not have emerged without the development of MT. Technologies have been expanding the sphere of translation market more than ever before.

However, greatly improved quality of MT—specifically surface fluency of neural MT output—sometimes induces in users a false perception of its reliability. Indeed, fatal misuse of MT has been seen, for example, in the multilingual dissemination of disaster information in Japan. This is attributable to the users’ literacy in MT, not the technologies themselves. MT users, including clients, translation companies and translators, do not necessarily have sufficient knowledge of what MT can and cannot do. Furthermore, MT developers and researchers may not be able to explain what the essential difference between MT and human translation is. More broadly, there is a lack of consensus amongst the various actors on what translation is in the first place. In these circumstances, facing the superficial resemblance of MT output and human translation, we are now in the position to reconsider the proper place of machine and human in translation.

With this fundamental motivation in the background, this book aims to bridge two gaps. The first one is the gap between *document* and *language*. In a nutshell, human translators deal with documents as basic units of translation, while current MT systems chiefly process decontextualised sentences or expressions from language A to language B. Even when human translators handle an individual sentence, phrase or word, this is premised on the existence of the document of which it forms part, and they refer to the document explicitly or implicitly for their decision-making at each step of translation. On the other hand, although context-aware MT has been much researched recently, commercially and publicly

available MT systems are still sentence-based and do not explicitly take into account the document properties in their translation models.

Here, to address this problem, we employ the notion of controlled language (CL) and extend it to a document level. CLs are artificial languages that are developed by restricting the lexicon, grammar and style of a given natural language, and can be used to reduce the ambiguities and complexity of source text, eventually leading to the enhanced performance of MT. As previous CLs are also sentence-based, we incorporate document-level text properties into the CL rules, which opens a way to contextual MT without directly modifying the MT engines.

The second gap is between *authoring* and *translation*. In this global age, various kinds of information are to be distributed in multiple languages. MT tools have boosted the speed of document multilingualisation, but the output is not always of usable quality and thus needs to be revised by human workers, namely, post-editors. If the authoring process is optimised to the subsequent translation process, we can envisage an enhanced overall productivity. Again, CL is one of the viable devices for bridging the gap between authoring and translation. We pursue both human readability of source texts and their machine translatability by defining human- and machine-oriented CL rules.

In summary, this book proposes to establish *controlled document authoring* that enables contextual MT. To implement controlled document authoring, we orchestrated a wide variety of frameworks and methods covering document formalisation, CL, technical writing, and terminology management. From the practical viewpoint, controlled document authoring is difficult to deploy by non-professional writers. Therefore, we designed, implemented and evaluated an integrated authoring support system to help them write well-structured source documents that are not only easy to read but also easily translated by MT systems.

This book is based on my PhD thesis submitted in 2017, when statistical and rule-based MT systems were widely used in practical situations. While neural MT has become a dominant paradigm in the past few years, previous types of MT are still used in companies and governments, and more importantly, the fundamental problem of the sentence-based MT architecture is yet to be solved. Although some of the technologies used in this work are unavoidably outdated, the idea and frameworks of controlled document authoring and contextual MT continue to be valid in a rapidly changing age of translation technologies.

Acknowledgements

The core work described in this book was conducted at the University of Tokyo with the help of many people. First and foremost, I would like to express my sincere gratitude to Emeritus Professor Anthony Hartley of the University of Leeds. This book is built on his pioneering idea of document-level controlled languages. He led me into the machine translation (MT) research field when I was an undergraduate student, and since then he has continued to guide me in shaping the work described in the book. Every piece of advice he gave me regarding my research was eye-opening and inspiring.

I am extremely grateful to Professor Kyo Kageura of the University of Tokyo for his continuous support and guidance. He gave me many essential comments from a broad perspective, which was the driving force for me to step into a wide range of research fields. It was not only his clear-cut expert advice, but also our ‘casual conversations’ which enabled me to develop and foster my research questions. Since 2019, I have been involved in a research project on translation process and technologies (JSPS KAKENHI Grant Number 19H05660), which he has led as a principal researcher. This project has further helped me to reconsider my view on translation.

I am also very thankful to Dr Cécile Paris of the Commonwealth Scientific and Industrial Research Organisation (CSIRO). During her stay at the University of Tokyo in 2014 and 2015, we intensively discussed the document-level framework of the research. Our mutual collaboration over the years evolved into the authoring system, MuTUAL. Professor Akira Nemoto of Keio University and Professor Yuko Yoshida of University of Tsukuba offered me insightful, sometimes critical, comments specifically from the perspective of Library and Information Science, which helped me to reconsider and establish the standpoint of my research.

I am obliged to members of the National Institute of Information and Communications Technology (NICT), where I stayed for about one year conducting research on MT applications. Dr Eiichiro Sumita and Dr Masao Utiyama gave me a chance to work at NICT. During my internship period, I was able to develop the core part of the authoring system. Dr Atsushi Fujita spared much time to discuss our research topic together. He always gave me clear, detailed advice based on his vast range of knowledge and experience.

Professor Hitoshi Isahara of Toyohashi University of Technology allowed me to be involved in the SCOPE project, to which the study of controlled language is considerably indebted. In this project I worked closely with Dr Midori Tatsumi. We exchanged our views and ideas frankly.

My acknowledgements also go to Professor Kayoko Takeda of Rikkyo University and Professor Masaru Yamada of Kansai University, who introduced me to the field of translation studies and provided me with practical suggestions for my research. Further integration of the viewpoints of translation studies into MT application research is what I would like to pursue in the future.

This book was completed at Nagoya University, with which I have been affiliated since 2017. Professor Satoshi Sato of Nagoya University has provided me with a great research environment. His broad view of the interrelation between document and language has helped me to verbalise my idea of controlled document authoring.

I would like to express my appreciation to two reviewers, who kindly took the time to review the proposal and draft chapters of this book and gave me the detailed feedback and comments. Although I could not integrate all of them in the final draft, I believe that they helped improve the content as well as the writing style of this book.

This book is based on my PhD thesis entitled *Controlled Authoring for Document Multilingualisation Using Machine Translation*, submitted to the University of Tokyo in 2017. The following published work has been integrated into this book:

- Rei Miyata, Anthony Hartley, Cécile Paris, Midori Tatsumi and Kyo Kageura (2015). Japanese Controlled Language Rules to Improve Machine Translatability of Municipal Documents. In *Proceedings of the Machine Translation Summit XV*, Miami, Florida, pp. 90–103.
—Chapter 2 (Section 2.3) and Chapter 4.
- Rei Miyata, Anthony Hartley, Kyo Kageura and Cécile Paris (2016). “Garbage Let’s Take Away”: Producing Understandable and Translatable Government Documents: A Case Study from Japan. *Social Media for Government Services*, Springer, pp. 367–393. Adapted by permission from Springer Nature Customer Service Centre GmbH: Springer.
—Chapter 1, Chapter 2 (Section 2.1.1.1, 2.3 and 2.5.1), Chapter 3 (including Figure 3.4 and Figure 3.5) and Chapter 5 (Section 5.1 and Section 5.2).
- Rei Miyata, Anthony Hartley, Kyo Kageura, Cécile Paris, Masao Utiyama and Eiichiro Sumita (2016). MuTUAL: A Controlled Authoring Support System Enabling Contextual Machine Translation. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING), System Demonstrations*, Osaka, Japan, pp. 35–39.
—Chapter 2 (Section 2.1), Chapter 3, Chapter 5 and Chapter 7.

- Rei Miyata and Kyo Kageura (2016). Constructing and Evaluating Controlled Bilingual Terminologies. In *Proceedings the 5th International Workshop on Computational Terminology (CompuTerm)*, Osaka, Japan, pp. 83–93.
—Chapter 2 (Section 2.4) and Chapter 6.
- Rei Miyata, Anthony Hartley, Cécile Paris and Kyo Kageura (2016). Evaluating and Implementing a Controlled Language Checker. In *Proceedings of the 6th International Workshop on Controlled Language Applications (CLAW)*, Portorož, Slovenia, pp. 30–35.
—Chapter 7 and Chapter 8.
- Rei Miyata, Anthony Hartley, Kyo Kageura and Cécile Paris (2017). Evaluating the Usability of a Controlled Language Authoring Assistant. *The Prague Bulletin of Mathematical Linguistics*, No. 108, pp. 147–158.
—Chapter 9.
- Rei Miyata and Kyo Kageura (2018). Building Controlled Bilingual Terminologies for the Municipal Domain and Evaluating Them Using a Coverage Estimation Approach. *Terminology*, Vol. 24, No. 2, pp. 149–180. Adapted by permission from John Benjamins Publishing Company.
—Chapter 2 (Section 2.4) and Chapter 6.

The following material was redrawn as Figure 2.1 in Chapter 2 with the kind permission of Professor Noriko Kando of National Institute of Informatics (NII):

Figure 1 ‘The Categories’ (p. 3): from Noriko Kando (1997). Text-Level Structure of Research Articles: Implications for Text-Based Information Processing Systems. In *Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research*, Aberdeen, Scotland, pp. 1–14.

The original PhD work was partly supported by JSPS KAKENHI Grant Number 16J11185, the Program of KDDI Foundation, the Strategic Information and Communication R&D Promotion Programme of the Ministry of Internal Affairs and Communications (SCOPE), and the Research Grant of Tokyo Institute of Technology. The MT system used in the study was offered by Kodensha Co. In the process of turning my PhD thesis into this book, I re-examined what I had done in view of what I am working on as part of JSPS KAKENHI Grant Number 19H05660 and 19K20628. Thus, this book is also partly supported by the grants.

I would like to deeply thank my parents, relatives, and friends for supporting me throughout the long period of education. Finally, many thanks to my partner Asuka, who always helps and encourages me, and Ei, who brings great happiness to my life.

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part I

Research background

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

1 Introduction

1.1 Background

In this digital age, we have witnessed an increasing proliferation of information that is digitally created and disseminated online. In conjunction with this, rapid advances in translation technologies, such as machine translation (MT), have promoted the multilingualisation of digital text. Not only companies but also governments have increasingly adopted commercially or freely available MT systems to create translations in multiple languages to reach a wide audience. End users themselves can take advantage of online MT services to obtain information communicated in languages that they cannot understand.

In the context of Japanese municipalities, which is the main focus of this book, a variety of information is published online regarding not only regional events and tourism, but also certain procedures that must be complied with when living in the municipalities (e.g. registering residency with the local city hall; sorting and recycling garbage; taking action in the case of emergencies). In general, such texts are produced in the official language(s) of the country—in our case, Japanese. There are, however, many foreign residents who do not have the Japanese language skills necessary to understand official documents written in Japanese. Although some of the larger municipalities provide human translations of their websites into various other languages spoken by local communities, the target languages are limited, usually to English alone. Moreover, the scope of the translated versions is often much more restricted than the original Japanese documents since, as Carroll (2010, p.386) points out, ‘to expect local governments with limited resources to translate their entire websites into one or more foreign languages would be unrealistic’. In most small municipalities, resources are so scarce that they cannot even provide English translations. Under such circumstances, municipalities typically rely on MT tools, or else the residents themselves rely on MT, such as Google Translate, to grasp the meaning of texts.

Several issues arise. The original Japanese documents are typically created by non-professional writers, since local councils often do not have the financial resources to hire trained authors. Despite recent advances, MT tools are still known to be imperfect. The documents are often embedded in HTML (since they are web pages), which may further complicate automatic translation by

4 *Research background*

fragmenting sentences. As a result, the translated texts are often misinterpreted or not understandable by their intended audience.

1.2 **Problems**

We identified that residents might encounter several difficulties when they read governmental or municipal documents, at three different levels. We outline each of them below.

1.2.1 *Document-level issues*

In Japanese municipal documents, we often find cases in which individual sentences make sense, but directions provided by the document as a whole are confusing. Figure 1.1 provides an example of this.¹ The figure explains the process for registering personal seals in Shinjuku City, Tokyo, which is one of the largest municipalities in Japan and provides human-translated information in multiple languages.

When we read this document, we may be at a loss, wondering whether we are eligible for seal registration or not, since the eligibility conditions for registering a personal seal are stated only at the bottom of the document. Furthermore, the conditions are not expressed in a clear manner, for example: ‘Those Who Are Not Eligible for a Residence Records’. Similarly, the requirements for re-registering a seal are vague: ‘If You Move Out of Shinjuku City’; ‘If You Leave Japan’. Moreover, the sub-section ‘Personal Seals Registration Certificate’, which is a distinct and separate task from seal registration, offers no explanation as to why we would need to obtain such a certificate and whether it is required or optional, although this is alluded to at the end of ‘Personal Seal (*Inkan*)’.

1.2.2 *Sentence-level issues*

Suppose you have not yet gained a sufficient operational command of Japanese to understand instructions when they are provided in Japanese only. While some information on municipal websites is provided in languages such as English, Chinese, Korean and Portuguese, these translations are usually created in part by MT. Many municipalities provide information vital for completing necessary administrative procedures or tasks in daily life in Japanese only. In such circumstances, having recourse to free, online and thus readily available MT may seem an attractive option. However, while the quality of MT output is reasonable for many language pairs in many practical situations, MT systems does not always produce satisfactory results. Occasionally, you will encounter MT outputs similar to the following:

- (a) Garbage let’s take away.
- (b) From July 2013, you will not be able to use only the specified garbage bag.

Seal Registration

■ Personal Seal (*Inkan*)

In Japan, personal seals are used as a symbol of agreement or approval, like a signature, to verify official documents, such as contracts.

You can order a personal seal for your name at a stamp engraving outlet and register the imprint at the City Office. When necessary, you can request a personal seal registration certificate that certifies that the personal seal is registered.

■ Personal Seal That Cannot Be Registered

- Stamps with letters that do not combine to form part of your full name, last name, or first name as registered in your residence record
- [...]
- Stamps that are inappropriate for registration (for example, stamps without an outer rim, cracked stamps, ready-made stamps, ring stamps, etc.)

■ Personal Seal Registration Procedures

Please bring the personal seal you wish to register along with your valid residence card [...] and complete the application procedures in person. [...] Some registration restrictions apply, such as those on age (must be 15 years of age or older).

[...] bring the following items to the service counter where you filed your application:

- The response sheet
- [...]

When registration has been completed, you will be issued a personal seal registration card. [...]

■ Personal Seals Registration Certificate

To apply for a certificate, please complete application procedures [...] and show your personal seal registration card. You will be issued a personal seal registration certificate, which certifies that your personal seal has been registered. [...]

■ When Notification Is Necessary (for Personal Seal Registration)

(a) If you lose your personal seal [...] → Notification of Discontinuation [...]

■ If You Move Out of Shinjuku City (Personal Seal Registration)

If you have completed personal seal registration but are moving out of Shinjuku City, [...].

■ If You Leave Japan (Personal Seal Registration)

If a person with personal seal registration leaves Japan, the personal seal registration becomes invalid and is deleted

Even if you move back to the same address, you must complete personal seal registration again.

■ Those Who Are Not Eligible for a Residence Record

Anyone who is not eligible for a residence record [...] cannot register a seal.

Figure 1.1 Personal seal registration procedure (excerpted from the website of Shinjuku City)

The Japanese input for (a) is ‘ごみは持ち帰ろう/*Gomi wa mochikaero*’, a sensible translation of which is ‘Please take your garbage home’. The MT error stems from differences between the Japanese and English languages when expressing public requests. The Japanese input for (b) is ‘2013年7月からは、指定ごみ袋しか使えません/*2013-nen 7-gatsu kara wa, shitei-gomibukuro sika*

tsukae-masen', which is correctly translated as 'From July 2013 you can use only the specified garbage bags'. In this case, the MT system has mis-translated the Japanese construction 'しか...ない/*shika ... nai*', which is somewhat akin to a double negation.

While, as a native or non-native speaker of English, you may be able to guess what (a) means, you may equally be misled into interpreting this notice as a call for community volunteers to clean up garbage in, for instance, a nearby park. In the case of (b), the meaning of the MT output is the exact opposite of the Japanese original. You are therefore at risk of completely misinterpreting the message or, even if you suspect its true intended meaning, being left in a state of uncertainty. Such misunderstandings and doubts can pose significant problems for you when living as a non Japanese-speaking resident in Japan.

Given that MT systems are being and will continue to be widely used in this domain, along with the cost of having all municipal information translated solely by human translators, improving the quality of MT output to a reliable level is an urgent task.

From the perspective of MT technologies, MT systems dealing with Japanese as source language (SL) or target language (TL) are state-of-the-art level. To date, Japanese natural language processing (NLP) researchers have invested considerable energy in MT research and at times have been world leaders in developing this technology, such as the paradigm proposed for example-based machine translation (EBMT) (Nagao, 1984; Sato and Nagao, 1990). Moreover, there are many commercially and freely available Japanese MT systems developed in Japan which are based on different architectures and technologies: rule-based machine translation (RBMT), which uses (manually constructed) rules to transform SL into TL; EBMT, which uses analogical reasoning based on translated examples; statistical machine translation (SMT), which relies on statistical learning from large aligned bilingual text corpora; and neural machine translation (NMT), which also uses large bilingual corpora to build a neural network model that consumes source text (ST) and generates target text (TT) in an 'end-to-end' manner. Although we have recently witnessed the great improvement of MT since the advent of NMT, current MT systems still face difficulties in dealing with certain types of linguistic patterns, such as long complex sentences. Alternative—or complementary—approaches to improving the performance of MT include imposing restrictions on the form and/or length of SL texts, using controlled language (CL). If we can diagnose what MT can and cannot do and embed MT within the overall framework of information flow, we anticipate being able to use MT for producing reliable outputs.

1.2.3 Terminology issues

Finally, terminology issues cannot be ignored in terms of accurate and consistent understanding of both ST and TT (Wright and Budin, 2001; Warburton, 2015b). We sometimes observe that several different terms refer to the same concept within a website or across websites of municipalities. For example, in a

municipal website, ‘印鑑証明書/*inkan-shomei-sho*’ and ‘印鑑登録証明書/*inkan-toroku-shomei-sho*’ are used in source Japanese side and their English translations are respectively ‘personal seal proof certificate’ and ‘seal registration certificate’. These variations may be confusing for those who have not sufficient domain knowledge about Japanese municipal procedures.

Moreover, from the point of view of MT technologies, terminology influences the output quality. One example MT output awkwardly translated is ‘Burned trash’. The Japanese input is ‘燃やすごみ/*moyasu gomi*’, a proper translation of which is ‘Combustibles’. The particular MT system fails to capture the term ‘燃やすごみ’, and erroneously processed the verb ‘燃やす/*moyasu*’ (burn) as a past participle. The problem of technical terms can be addressed by maintaining terminology for MT dictionaries.

1.3 Solution scenario

We thus have different, but related, problems. We might face both ill-organised document structure and poor MT output, which may well aggravate the situation as readers have no reliable context to aid them in ‘guessing’ the actions being described. This observation led us to realise the necessity of pursuing a unified solution to the overall issue of multilingualisation of municipal procedural information, namely, introducing *controlled authoring*, where the control is applied consistently and seamlessly at both the sentence- and document-level.

Controlled authoring is ‘the process of applying a set of predefined style, grammar, punctuation rules and approved terminology to content (documentation or software) during its development’ (Ó Broin, 2009, p.12). According to the ISO standard, controlled authoring is defined as an ‘authoring that uses limited vocabulary and textual complexity to produce clear documents’ (ISO, 2012). Though these two definitions focus on linguistic and terminological control of authoring processes, we notice that they presuppose the existence of ‘documentation’ or ‘documents’. Hence, we can reasonably extend the idea of controlled authoring to cover document-level control, and thus we propose the notion of *controlled document authoring*, that is, authoring that uses formalised document structure, limited grammar, lexicon and style, and approved terminology to produce well-structured documents that are clear and consistent.

As we will elaborate in later chapters, we find that an integrated approach of embedding controlled sentences within a well-designed document structure has clear benefits in further improving MT output. Take, for instance, ‘文書を印刷する/*Bunsho o insatsu-suru*’, which may naturally appear as a task title or as a step in a procedure. A given MT system may translate this as ‘To print the document’, which is appropriate wording for a title but not for a step in a process, where ‘Print the document’ (imperative) is needed. If we know the functional element in which a Japanese expression occurs, we can exploit this knowledge to pre-process expressions where necessary by transforming them so that the MT system is coerced into producing a contextually appropriate English translation. Since the pre-processing would be an internal operation which does not change

the Japanese text seen by readers, the readability of the ST would not be degraded by TL-oriented writing rules designed to improve MT output quality, which can happen (e.g. Hartley et al., 2012).

While the idea of contextual MT, i.e. the unification of document elements, CL and MT, has been already proposed (Bernth, 2006; Hartley, 2010), its feasibility and applicability have not yet been fully investigated. We address this research challenge with a special focus on the task of translating municipal procedural documents, which is the most significant contribution of this book.

1.4 Research questions

The main research question to be answered in this book is as follows:

RQ Can controlled document authoring help non-professional writers to create well-structured source texts that are machine-translatable and human-readable?

This research question can be divided into two aspects: framework and application. From the point of view of framework, we specify our questions as follows:

- RQ-F1** Can municipal documents be well formalised?
- RQ-F2** To what extent can a Japanese CL improve the quality of source text (ST) and target text (TT)? (**sentence-level CL**)
- RQ-F3** Can the combination of controlled language and document structure further improve the TT quality without degrading ST quality? (**document-level CL**)
- RQ-F4** Can municipal terms be comprehensively captured and well controlled?

To answer **RQ-F1**, we employ an existing document standard used for technical documentation. If we can properly formalise the municipal documents based on this standard, we can conclude that they are well formalised. To answer **RQ-F2**, we formulate CL rules specifically intended for the municipal domain and evaluate their effectiveness in terms of ST readability and MT output quality. To answer **RQ-F3**, we contextualise CL rules into the municipal document structure and diagnose the MT outputs. To answer **RQ-F4**, we construct Japanese–English bilingual controlled terminologies and evaluate them in terms of coverage and quality of control.

Based on the results obtained through answering the questions **RQ-F1** to **RQ-F4**, we propose and implement an authoring support system, **MuTUAL**, which is designed to help non-professional municipal writers create controlled documents that are both machine-translatable and human-readable. The core module of the system is the controlled authoring assistant, which automatically checks conformity to the CL and controlled terminology when users are drafting and rewriting ST. Hence, in terms of application, this book asks following questions:

- RQ-A1** How accurately does the system detect CL rule violations in text?
RQ-A2 Is our system usable for non-professional writers?
RQ-A3 Does the use of the system help improve the quality of ST and TT?

To answer **RQ-A1**, we implement CL violation detection rules and benchmark their detection performance using a test dataset. To answer **RQ-A2** and **RQ-A3**, we conducted a usability evaluation to see whether our proposed system can improve the user's writing performance and output text quality.

1.5 Scope

The ultimate goal of this research project is to provide an integrated authoring environment that makes use of off-the-shelf MT systems to enable writers to create and publish documents and their multilingual equivalents in the Japanese municipal domain. Tackling all possible varieties of texts available on municipal websites and also multiple target languages is not a realistic goal at this stage. As such, this book focuses on the task of Japanese-to-English translation of municipal documents regarding daily life as a starting point, with a specific focus on procedural documents.

According to Oda (2010, p.22) and OpenUM Project (2011, p.9), Japanese municipal websites typically feature the following kinds of information:

1. Legal information (including constitution, law, government ordinance and ministry ordinance)
2. Official information (including municipal bylaw, regulation and notice)
3. Public-related information
 - 3-1. Information for residents (**municipal-life information**)²
 - 3-2. Information for business operators
 - 3-3. Information for tourists
 - 3-4. Policy information of administrations

Municipal-life information usually pertains to content which is directly related to citizens' daily life, and there is a growing need for multilingualisation of this content. In particular, procedural documents are of most importance as they enable residents to avail of municipal services (such as child allowance) and carry out necessary municipal procedures (such as tax payment). From a methodological standpoint, we assume procedural documents are well-suited for document formalisation, and we can make use of existing document structures developed in the field of technical writing and business documentation. Therefore, as a pilot study, we investigate municipal procedural documents, which offers a point of reference for future work.

The reasons why we choose English as the target language are as follows:

- English is still an overwhelmingly popular choice when translating Japanese municipal texts, followed by Chinese, Korean and Portuguese (Carroll, 2010).

- MT systems often use English as a pivot language. For example, Google Translate appears to produce Japanese-to-Vietnamese translation by first translating Japanese into English and then translating English into Vietnamese.³ Thus, improving English-language MT output quality leads to secondary improvements in MT output quality for many other languages.

It is worth noting in advance that the framework and system environment that we propose in this book are applicable to other text domains/types and language pairs.

1.6 Chapter organisation

This book consists of four parts, divided into a total of ten chapters (see Figure 1.2).

Part I, Chapters 1 and 2, explains the background to our research. Chapter 2 summarises existing research on document formalisation, MT and its practical implementation, CLs, terminology management and authoring environments.

Figure 1.2 Chapter organisation with research questions

Part II, Chapters 3–6, presents our research on controlled authoring at different textual levels. In Chapter 3, as a document-level study, we present (with examples in English) an analysis of procedural texts from Japanese municipalities, and show how a standard document structure can be specialised to cover these texts. In Chapter 4, as a sentence-level study, we design and evaluate Japanese controlled-language rules for improved machine-translatability and source readability, focusing on texts featuring municipal-life information. Chapter 5 details how to combine document structure with controlled language, which is the most innovative idea proposed by this study, and suggests mechanisms to further improve text quality. In Chapter 6, as a terminology-level study, we manually construct and evaluate Japanese–English controlled terminologies for the municipal domain.

Part III, Chapters 7–9, proposes an authoring environment, MuTUAL, that exploits the framework established in the previous chapters to support the writing of municipal procedures such that they are easily translatable by automated tools. Chapter 7 demonstrates the concept, module organisation and intended use scenario of MuTUAL, and describes the implementation of each module. We first evaluate the precision and recall performance of a subcomponent of the system in Chapter 8. Based on the result, we conduct a user study to evaluate the usability of the core authoring module in Chapter 9.

Part IV, Chapter 10, summarises research findings and concludes the book. We discuss the results of the previous chapters, pointing out the major contributions and limitations of the study, and sketch out our future plans towards a practical implementation of the system in real-world scenarios.

Notes

- 1 Shinjuku City, Seal Registration, www.foreign.city.shinjuku.lg.jp/en/todoke/todoke_6/
- 2 In this book, we use ‘municipal-life information’ instead of ‘information for residents’.
- 3 This is not announced by Google, but we can reasonably infer that Google Translate uses English as a pivot language from the fact that a Japanese-to-Vietnamese MT output is almost the same as the Japanese-to-English-to-Vietnamese MT output.

2 Related work

In this book, we propose to solve the problem of producing machine-translatable documents through an integrated approach of embedding controlled sentences within a well-designed document structure. This chapter reviews existing work related to conceptual and technical elements of this book. Section 2.1 describes the existing frameworks of document formalisation and research regarding descriptive document analysis; Section 2.2 gives an overview of machine translation (MT) and its applications; Section 2.3 focuses on the study of controlled language (CL) for improved machine-translatability; Section 2.4 is devoted to terminology management, with a specific focus on controlled terminology development; Section 2.5 summarises existing system environments for controlled authoring.

2.1 Document formalisation

Documents are not mere collections of pieces of information. They are organised—preferably, well-organised—to achieve specified communicative goals, such as informing municipal residents of how to conduct certain procedures. In order to formalise document models, there are, broadly speaking, two perspectives to be considered: content (*What contents should be included in the document?*) and structure (*In what order the contents should be arranged in the document?*). According to the terminology in the field of natural language generation (NLG), the former is the task of *content determination* and the latter is the task of *document structuring* (Reiter and Dale, 2000). Similarly, Mossop (2014, p.77) distinguished two types of text structure, *conceptual structure* and *physical structure*, examples of which are as follows:

An example of the former [conceptual structure] would be an argument structure: presentation of problem, tentative solution, arguments for, arguments against, conclusion. An example of physical structure would be the parts of an article: title, summary, section head, sequence of paragraphs, inserted table, next section head, and so on.

Although it is true that the ‘argument structure’ pertains to both the content and structure, we can loosely associate the ‘conceptual structure’ with content and

the ‘physical structure’ with the structure. While the content and structure of a document are closely interrelated and it is sometimes difficult to approach them as distinct and separate phenomena, it is convenient to distinguish between the two in order to analyse and model documents.

In Section 2.1.1, we first describe existing frameworks for document formalisation, specifically focusing on the DITA (Darwin Information Typing Architecture) standard (OASIS, 2010). In Section 2.1.2, we summarise methods for descriptively analysing documents (texts) which can be employed to reveal the content and structure of documents in a given domain. Finally, in Section 2.1.3, we mention several studies that evaluate document quality.

2.1.1 Existing framework for document formalisation

2.1.1.1 DITA

DITA is an XML architecture for authoring and publishing technical information. DITA supports topic-based authoring, which helps writers compose modularised information covering not more than one concept, one procedure or one unit of referential information (Bellamy et al., 2012). DITA was first developed at IBM and donated to the OASIS (Organization for the Advancement of Structured Information Standards) in 2004 (Day et al., 2005). DITA version 1.0 was approved by the OASIS DITA TC in 2005, and the latest version (1.3) was approved in 2015.

DITA has two basic components: *topic* and *map*. Topic is a self-contained informational unit. Map is a mechanism for creating different deliverables from a single source, in other words, organising multiple topics as a document according to the output medium and document purpose. The focus of this book is, however, chiefly on the topic, although map is a useful mechanism for compiling and publishing documents both for websites and for print documents. Note that a single topic can be regarded as a document since it is self-contained and has a certain communicative goal.

A topic is composed of *functional elements*, i.e. elements which play certain communicative roles within a document. It guides authors as to what kind of content should be included and in what way this content should be organised. At the highest level, the functional elements of a topic are predefined as follows (OASIS, 2010):

- **Title** ‘contains the subject of the topic’.
- **Short description** is ‘used both in topic content (as the first paragraph), in generated summaries that include the topic, and in links to the topic’.
- **Prolog** (prologue) is ‘the container for topic metadata, such as change history, audience, product, and so on’.
- **Body** ‘contains the topic content: paragraphs, lists, sections and other content that the information type permits’.
- **Related** links ‘connect to other topics’.

DITA defines by default three basic topic types. The ‘Concept’ topic answers the question ‘What is it?’ and is used to provide conceptual information. The ‘Task’ topic answers the question ‘How to do it?’ and is used to describe a step-by-step procedure. The ‘Reference’ topic is used to present reference information which guides readers to other related documents or websites. These topics are defined by specialising the **Body** element of the generic topic. The Task topic, for example, contains the following functional elements to cover the necessary information for describing certain tasks (OASIS, 2010):

- **Prereq** (prerequisite) ‘describes information that the user needs to know or do before starting the immediate task’.
- **Context** ‘provides background information for the task’.
- **Steps** ‘provides the main content of the task topic. A task consists of a series of steps that accomplish the task’.
- **Result** ‘describes the expected outcome for the task as a whole’.
- **Example** ‘provides an example that illustrates or supports the task’.
- **Postreq** (postrequisite) ‘describes steps or tasks that the user should do after the successful completion of the current task’.

However, the functional elements of the Task topic as defined in DITA are still too coarse-grained to afford specific guidance to authors of municipal procedural documents. It is still uncertain as to what kind of content is specifically to be included in each functional element. It becomes necessary to instantiate each element and specialise certain elements to match the needs of municipal procedures.

2.1.1.2 *Other document frameworks*

DITA is not the only framework that is used to formulate well-organised documents. Another similar framework that is also widely used is information mapping (Horn, 1989, 1998). A core methodology of information mapping is structured writing, in which a basic unit of information consisting one or more sentences and/or graphical structures (*information block*) is defined and multiple information blocks are organised into an *information map* (Horn, 1989).¹ Each information block is associated to one of the seven *information types* described in Table 2.1.² We notice several similarities between DITA and information mapping:

- Both view a document as a collection of modularised chunks of self-contained information.
- Each chunk of information has a certain communicative purpose, such as describing a procedure.

Table 2.1 Seven information types defined in Information Mapping (Horn, 1989, pp.110–111)

<i>Information types</i>	<i>Definition</i>
Procedure:	a set of sequential steps that one person or entity performs to obtain a specified outcome. This includes the decisions that need to be made and the action that must be carried out as a result of those decisions.
Process:	a series of events or phases which take place over time and usually have an identifiable purpose or result.
Structure:	a physical object or something that can be divided into parts or has boundaries.
Concept:	a group or class of objects, conditions, events, ideas, responses or relations that <ul style="list-style-type: none"> • all have one or more attributes in common • are different from one another in some other respect • are all designated by a common name
Fact:	a statement of data without supporting information that is asserted with certainty.
Classification:	the division of specimens or things into categories using one or more sorting factors.
Principle:	a statement that <ol style="list-style-type: none"> 1. tells what should or should not be done such as <ul style="list-style-type: none"> • rules • policies or guidelines • warnings or cautions 2. seems to be true in light of the evidence such as <ul style="list-style-type: none"> • generalisation • theorems 3. is unprovable but implied by other statements, such as <ul style="list-style-type: none"> • assumptions • axioms • postulates.

More broadly, the field of technical writing has devised and implemented a number of mechanisms to create quality documents for academic and business purposes. For example, according to Rubens (2001, p.7), procedural documents usually have the following functional elements:

- ‘Overviews of product operations;
- Modules presenting sequences that support one or more activities;
- Some conceptual models, analogies, and examples;
- Topic and section links;
- Macro-level retrieval, location, and navigation aids’.

While overlap can be noticed between these elements and the DITA elements in the Task topic, in this book we adopt the DITA since it provides finer-grained (though not fully sufficient) functional elements by default, and we can use it as a basis for further specification of the document elements.

2.1.2 Descriptive analysis of document

The previous section summarised existing frameworks that guide authors in creating well-organised documents. The remaining issue is that they are general-purpose frameworks and insufficient for formalising content and structure in a certain domain (class of documents), such as municipal procedural documents. Here we review two similar descriptive methods that can be employed to reveal fine-grained functional elements specific to a given domain, namely genre analysis and functional structure analysis.

2.1.2.1 Genre analysis

It is first useful to confirm the definition and perspective of *genre*. Bhatia (2004, p.23) defined genre as follows:

Genre essentially refers to language use in a conventionalized communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to stable structural forms by imposing constraints on the use of lexico-grammatical as well as discoursal resources.

Biber and Conrad (2009, p.16) differentiated the perspective of genre from that of *register* and *style* as follows:

In the genre perspective, the focus is on the linguistic characteristics that are used to structure *complete texts*, while in both the register perspective and the style perspective, the focus is on the pervasive linguistic characteristics of representative text excerpts from the variety. [emphasis ours]

The most important point is that we can deal with textual features in light of genre only under the condition that there exist complete texts (what we call ‘documents’ in this book). In other words, textual features are always identified in relation to the rhetorical/functional roles they play in the complete texts. From the perspective of genre, linguistic characteristics often refer to ‘specialised expressions’, ‘rhetorical organisation’ and ‘formatting’, and these usually ‘occur only once in a text’ (Biber and Conrad, 2009, p.16). For example, ‘specialised expressions’ would be ‘Dear Mr –’, a name of recipient in a business letter, while ‘rhetorical organisation’ would be IMRD (Introduction, Methods, Results and Discussion) structure of research articles. Here, the genre perspective is particularly effective in investigating functional elements in documents.

A wide range of studies have been undertaken to investigate the rhetorical structure of a certain class of documents in the field of English for Specific Purposes (ESP). Genre analysis is conducted descriptively: (1) collect sufficient amount of (pieces of) documents³ from specific domains, such as research articles and news reports, and (2) identify and annotate smaller functional/rhetorical chunks of documents. One widely known genre analysis was made by Swales (1990, 2004), who specifically analysed textual conventions of research articles. For instance, Swales (1990, p.141) pointed out that introduction sections of research articles typically consist of three rhetorical movements (simply called *moves*) and proposed Create a Research Space (CARS) model:

Move 1 Establishing a territory

- Step 1** Claiming centrality
- Step 2** Making topic generalisation(s)
- Step 3** Reviewing items of previous research

Move 2 Establishing a niche

- Step 1A** Counter-claiming
- Step 1B** Indicating a gap
- Step 1C** Question-raising
- Step 1D** Continuing a tradition

Move 3 Occupying the niche

- Step 1A** Outlining purposes
- Step 1B** Announcing present research
- Step 2** Announcing principal findings
- Step 3** Indicating RA [research article] structure

This method, also known as *move analysis*, has been particularly influential in the field of ESP. A number of researchers have conducted a genre (move) analysis targeting various classes of texts, including not only research articles (Brett, 1994; Bunton, 2005; Cross and Oppenheim, 2006; Maswana et al., 2015; Tessuto, 2015) but also company audit reports (Flowerdew and Wan, 2010) and customer reviews of products (Skalicky, 2013).

2.1.2.2 Functional structure analysis

Similar attempts to reveal the functional elements in a class of documents were conducted by Kando (1997, 1999). She analysed 40 writing manuals and 127 Japanese research papers collected from four disciplines: medicine, physics, economics and Japanese literature. Functional elements (components) typically observed in research papers were categorised and arranged hierarchically. Figure 2.1 shows the full set of functional elements.⁴ The most specific—the third or fourth—level of elements are assigned to each sentence. Each element ‘represents the role or function that the part of text plays in the whole text, the

Figure 2.1 Functional elements of research papers (Kando, 1997, p.3)

relationship between a part of text and the whole text, and between parts in the text and other parts in other text’ (Kando, 1997, p.2).

This analysis shares the same viewpoint as the genre analysis above, that is, to grasp the functional roles of subdivisions of text in relation to the larger

subdivisions or the complete text. However, we can point out differences between the Swales's move analysis (**Swales**) and Kando's structure analysis (**Kando**):

- **Swales** defined elements (moves) under IMRD structure, while **Kando** defined elements without explicitly using IMRD.⁵
- The granularity of elements by **Kando** is finer than that by **Swales**.
- **Swales** is oriented for writing (or teaching how to write papers), while **Kando** is oriented for information retrieval.⁶

In order to formalise municipal procedures, we consider that it is first useful to comprehensively identify fine-grained functional elements in documents as **Kando** did. Then, for the purpose of document authoring, identified elements should be rearranged according to DITA elements.

2.1.3 Evaluation of document quality

Evaluation of document quality is a crucial step before documents are actually distributed and used. To date, a number of methods have been implemented by researchers to assess document quality (see De Jong and Schellens (2000) for survey of evaluation methodologies). To evaluate instructional (procedural) texts, in addition to a diagnostic evaluation by technical reviewers (Carey et al., 2014), reader-focused, task-based methods are beneficial in gauging whether or not the documents actually help readers achieve their intended goals (Schriver, 1997; Reiter et al., 2001; Colineau et al., 2002).

For example, Schriver (1997) conducted a user study to evaluate instruction guides of a VCR and a stereo system. Three types of guides are prepared: (1) original instruction guides, (2) initial revisions, and (3) final revisions. Participants were presented one of the guides and asked to perform a series of tasks, such as setting the timer (VCR) or recording a sequence of songs (stereo), and to think aloud while conducting the tasks. The results indicated that participants who used the final revisions completed the given tasks significantly more accurately and quickly than those who used the original instruction guides.

Reiter et al. (2001) and Colineau et al. (2002) conducted evaluation studies to assess whether documents generated by their NLG systems really help readers achieve their goals in real-world scenarios. Reiter et al. (2001) evaluated the **STOP** system, which generates personalised health information letters for smoking-cessation, by a randomised controlled clinical trial. They collected a pool of 2553 participants who were all smokers and filled out a questionnaire which asks about the habit of smoking. These smokers were randomly split into three groups: (1) *Tailored* group who received the personalised letters generated by the system, (2) *Non-tailored* group who received a fixed (non-tailored) letter generated by the system, and (3) *No-letter* group who received only a thank-you letter. The evaluation assessed how often participants who read tailored letters actually stopped smoking as compared to those who read generic letters. Colineau et al. (2002), on the other hand, evaluated the **Isolde** system, which helps

writers create procedural instructions, by a task-based method. They compared the effectiveness of the instructional texts generated by the system with that of professionally authored instructions by measuring the user's performance (accuracy and time) in completing a specific task.

Finally, with regards to our research, the fact that the DITA is already adopted widely suggests its effectiveness in real circumstances. Given that the DITA elements are sufficiently specialised, it is reasonable to assume that authoring in accordance with DITA ensures that the created documents cover necessary contents and are well-structured, which eventually helps readers achieve their goals. However, DITA itself does not ensure linguistic quality written in documents. Therefore, in this book, we will not delve into task-based document evaluation as such, but focus more on linguistic quality assessment.⁷

2.2 Machine translation

Machine translation (MT) is a computerised system 'responsible for the production of translations from one natural language into another, with or without human assistance' (Hutchins and Somers, 1992, p.3).⁸ Since the advent of the computer more than half a century ago, much effort has been devoted to develop and improve MT systems. Several breakthroughs have been achieved and MT has become increasingly adopted in many fields. Japanese municipalities too have started using MT on their websites to provide information in multiple languages for residents with diverse linguistic backgrounds.

Though there exists an extensive range of research on developing and improving MT systems, in this section we mainly focus on the 'practical' aspect of MT deployment with our scenario in mind. We first give an overview of the basic architectures of MT, mentioning the difficulties involved in achieving fully automatic, high-quality MT. We then look at the different approaches towards practical use of MT, specifically how to employ MT combined with human intervention. Finally, we summarise existing methodologies for MT evaluation.

2.2.1 Overview

2.2.1.1 Architecture

MT architectures can be broadly classified into the following four categories:

- Rule-based MT (RBMT)
- Statistical MT (SMT)
- Example-based MT (EBMT)
- Neural MT (NMT)

The most traditional of the four is RBMT (Hutchins and Somers, 1992). According to the degree of depth of linguistic analysis, we can further specify the three sub-categories within RBMT: (1) direct approach, (2) transfer approach and

(3) interlingua approach. In the direct MT system, source input sentences are morphologically analysed and each source word is directly translated word-by-word through bilingual dictionary look-up. This primitive approach is rarely adopted nowadays. In contrast to the direct approach, the other two approaches are categorised as ‘indirect approaches’. They analyse the syntactic structure of the ST and create an intermediate representation of it. The transfer MT system first parses the ST to create a language-dependant intermediate representation in the SL, then transfers it to a counterpart intermediate representation in the TL. Finally, it generates a target sentence from the transferred representation. The interlingua MT system executes a more in-depth analysis of the ST to create an abstract representation which is (ideally) language-independent and acts as a pivot. TT can be generated from this interlingua representation. This approach is particularly advantageous for multilingualisation as there is no need to devise transfer components for each pair of languages. In practice, however, it is difficult to create language-independent representations. Thus, the transfer approach is often adopted in RBMT. For major language pairs, including Japanese to/from English, RBMT is still used in commercial MT systems.

While RBMT is based on hand-crafted bilingual dictionary and transfer rules, the remaining three types of MT exploit bilingual parallel text corpora through one of several strategies. Thus, they are generally called ‘corpus-based’ methods (Hutchins, 2015). Since the end of the 1980’s, researchers’ focus has shifted from rule-based methods to corpus-based methods.

One of the three main forms of corpus-based MT is SMT (Koehn, 2009), which was originally developed by Brown et al. (1988, 1990, 1993) and became the dominant paradigm in the field of MT research. In this approach, statistical models are learned from existing corpora. A ‘translation model’, which assigns translation probabilities to a given ST segment, is learned from a bilingual parallel corpus. A ‘language model’, which determines the most probable sequence of target words, is learned from a monolingual corpus. One of the most notable advantages of SMT compared to RBMT is that the MT does not require a deep knowledge of source and target languages. On the other hand, the availability of large bilingual parallel corpora is indispensable to develop reliable statistical models. In the Japanese municipal domain, few well-maintained bilingual resources are available, which is a significant issue in building domain-specific SMT engines.

Another corpus-based approach is EBMT (Nagao, 1984; Sato and Nagao, 1990; Carl and Way, 2003), the underlying idea of which is ‘translation by analogy’, i.e. simulating human translators’ behaviour of finding and recalling analogous examples previously translated. Though it is difficult to clearly distinguish EBMT from both RBMT and SMT, Hutchins (2005b, p.203) neatly summarises the essence of EBMT as follows:

In the case of EBMT, the core bilingual process is the matching of SL fragments (from an input text) against SL fragments, and the retrieval of equivalent TL fragments (as potential partial translations).

More recently, NMT (Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014; Bahdanau et al., 2015) has drawn attention from the MT research community. NMT also utilises bilingual corpora, but the basic architecture is different from that of SMT. Specifically, while SMT typically comprises several sub-components that need to be trained separately, NMT attempts to build a single neural network that consumes ST and generates TT. As such, NMT can be characterised as an ‘end-to-end’ approach. Though NMT is currently dominating the MT research field, it has several drawbacks, including a lack of reliability, especially when dealing with unknown words (Jean et al., 2015), and the computational cost, which hinders its practical deployment. While the end-to-end approach of NMT is one of its strengths, the ‘black box’ nature of this architecture makes it more difficult to systematically cope with particular MT errors.

As of 2016, many systems publicly and commercially available are RBMT and SMT, including hybridisations of both.⁹ In the following sections, keeping in mind the availability of MT systems to Japanese municipalities, we focus on RBMT and SMT.

2.2.1.2 *Difficulties*

Due to the improved algorithm and computational environments, the performance of MT is continuously improving. However, fully automatic, high-quality MT is still too ambitious a goal.

The fundamental difficulty for any MT, as well as human translators, is ambiguity. Two levels of ambiguity are often highlighted: structural ambiguity and lexical ambiguity. Structural ambiguity arises when the underlying structure of a sentence can be interpreted in more than one way. A typical example of such ambiguity is the sentence ‘Flying planes can be dangerous’, which can be interpreted in two ways: (1) ‘It can be dangerous to fly planes’; (2) ‘Planes which are flying can be dangerous’ (Hutchins and Somers, 1992, p.88). Lexical ambiguity, on the other hand, arises when a word can be interpreted in more than one way, which typically involves polysemy and homonymy. Unlike human translators, MT is not good at making use of subject matter knowledge and/or real-world knowledge in order to resolve these kinds of ambiguity, as it is difficult to codify or teach such knowledge to MT.

It is also widely acknowledged that MT between distant language pairs, such as Japanese and English, is more difficult than for closer language pairs, such as English and French (Isahara, 2015). Recently, to lessen the barrier of structural differences between languages, several methods have been proposed within the SMT research community, such as pre-ordering (or reordering) (Collins et al., 2005; Isozaki et al., 2010b; Hoshino et al., 2015). However, translating between distant language pairs remains an issue for commercial MT systems.

At the lexical/terminological level, out-of-vocabulary (OOV) words degrade MT performance. If a word is not registered in the MT dictionary or does not occur in the training data, the MT has to either infer its translation or simply reproduce the original source word in the TT.

2.2.2 Practical use of MT

The performance of MT is still imperfect due to numerous difficulties, including those described above. Thus, it is essential to incorporate human intervention in the translation process in order to ensure accurate, serviceable translations.

Figure 2.2 illustrates three basic approaches for the practical implementation of MT at each step in the translation work flow: (1) source input control, (2) MT engine customisation and (3) target output control. We describe each approach below.

2.2.2.1 Source input control

Controlling the ST to make it amenable to particular MT systems is one well-tried method to improve MT output quality. The main goal at this stage is to reduce ST ambiguity before MT is used. We first introduce two related concepts: *controlled language (CL)* and *pre-editing* (Kittredge, 2003). A CL is a version of a natural language whose lexicon, grammar and style have been restricted with a view to improving machine-translatability as well as human comprehensibility. Pre-editing, on the other hand, is the post-hoc process of rewriting an original ST so that it is more amenable to MT.

In many pre-editing scenarios, a CL is deployed to guide human or automatic pre-editors to amend CL rule violations in an ST. CLs are also deployed in writing, i.e. controlled writing, in which human writers draft STs from scratch following a certain CL guideline. CL is the main focus of this book and will be detailed in full in Section 2.3.

We will briefly outline existing pre-editing strategies which do not involve CLs or pre-defined rewriting rules. Several human-in-the-loop protocols were devised to improve machine-translatability of STs. Resnik et al. (2010) proposed a ‘targeted paraphrasing’ protocol: (1) monolingual target speakers identify putative error spans in MT outputs; (2) the spans are projected back to source sentences; and (3) monolingual SL speakers paraphrase the marked spans in the ST to generate alternative sentences which are once again machine-translated. Uchimoto

Figure 2.2 Machine translation work flow with human intervention

et al. (2006) used back translation¹⁰ as a means to spot non-machine-translatable spans in STs, which are subsequently served to humans to be rewritten. Similar attempts were also made by Miyabe et al. (2009). Mirkin et al. (2013) devised an interactive tool for monolingual authors to choose the most appropriate rewriting from a number of suggestions offered by the system. The generation of rewriting candidates is based on text simplification (TS) techniques (Feng, 2008; Specia, 2010), which assumes that simpler sentences are more likely to be machine-translatable. In contrast to human-intervention approaches, there have been several ideas which propose building a fully automatic pre-editor by training monolingual SMTs (Sun et al., 2010; Nanjo et al., 2012). Though the automatic pre-editing approaches have not yet shown a great increase in MT output quality, there is still room for further improvement.

We should quickly note that these non-CL-guided pre-editing approaches tend to overlook the ST quality, which is important when STs are also intended for consumption by human readers.

2.2.2.2 *Customisation*

Optimising general-purpose MT systems for a particular domain is also essential in achieving higher-quality MT. Currently, many of the available MT software packages allow for user customisation.

As for RBMT, the most common way to customise MT engines is to update dictionaries. This involves developing bilingual terminological resources and registering them in the system's user dictionary. Vasconcellos (2001) presented terminology management systems with different types of RBMT, i.e. direct systems, transfer systems and interlingual systems. She pointed out that building dictionaries with a high specificity of linguistic knowledge is not an easy task and, as such, some systems offer 'easy dictionary-building in exchange for less linguistic specificity' (Vasconcellos, 2001, p.699).

For SMT, there are two main ways to customise off-the-shelf MT engines. The first is the 'retraining' of statistical models by using domain-specific parallel data that users own. For example, TexTra,¹¹ a freely available SMT system, provides the means to customise MT systems by retraining general-purpose MT engines. However, retraining data usually requires a large parallel corpus, which is difficult to obtain in some domains and/or language pairs.

Terminology integration is also effective in improving SMT performance. Langlais and Carl (2004) identified that degradation of SMT outputs is mainly attributed to a high proportion of OOV words and poor translation of terminological units. They then integrated terminological resources into a general-purpose SMT engine to translate domain-specific texts from English to French. The evaluation results showed a significantly reduced word error rate. In an experiment translating technical documentation from English to French, Thicke (2011) demonstrated that simply customising an SMT engine with terminology boosted human 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.

In this study, our focus is on terminology resources that are suited to both RBMT and SMT. A review of terminology management with a focus on resource development will be presented in Section 2.4.

2.2.2.3 *Target output control*

Target output control, specifically post-editing (PE), is a process which aims to raise MT output quality to an acceptable level. PE is especially important when the MT outputs are disseminated for consumption by the public. Combining MT with PE was shown to be more productive than human translation from scratch. Plitt and Masselot (2010) conducted an experiment to measure the productivity of SMT post-editing as compared to traditional human translation from English to French, Italian, German and Spanish translation settings. Twelve participants were involved in translating from scratch and post-editing SMT outputs. The result demonstrated that PE plus SMT allows translators to increase the translation throughput, i.e. words processed per hour, by 74% on average.

While some attempts have been made to develop automatic PE (Knight and Chander, 1994; Simard et al., 2007), in practice, PE involves TL human post-editors, who are often bilingual, to revise the raw MT outputs with reference to the STs.¹² If STs are to be translated into multiple languages, PE must also be conducted by human post-editors for each language.

It should be noted that these three approaches—source input control, customisation, target output control—are tied to each other, and it is important to find an optimal solution in an integrated manner. For example, as Allen (2003, p.298) perceptively stated, ‘controlled language writing enhances and speeds up the translation and post-editing process’.

In the case of Japanese municipalities, however, PE is not a feasible option due to budget restrictions which limit their ability to hire multiple post-editors for different languages. Thus, in this study, we take as our starting point the upper and middle processes of translation, i.e. CL with MT customisation.

2.2.3 *Evaluation of MT*

Finally, this section examines existing literature on MT evaluation. MT can be evaluated using various criteria and different perspectives. Hutchins and Somers (1992, p.161–163) illustrated several stages of MT evaluation: ‘prototype evaluation’, ‘development evaluation’, ‘operational evaluation’, ‘translator evaluation’ and ‘recipient evaluation’, noting that ‘the testing of the linguistic quality of the output’ is common to all stages. We review here two conventional evaluation methods to assess the (linguistic) quality of ‘raw’ MT outputs: human evaluation and automatic evaluation.

2.2.3.1 Human evaluation

One of the earliest methods of human evaluation of MT was reported by the Automatic Language Processing Advisory Committee (ALPAC) (Carroll, 1966). In this experiment, Russian to English MT outputs were evaluated in terms of intelligibility (of the translation) and informativeness (of the translation relative to the original). Human evaluators, in this case university students, were asked to evaluate each sentence based on two scales: intelligibility on a nine-point scale and informativeness on a ten-point scale.

In the 1990s the Advanced Research Projects Agency (ARPA) developed three evaluation criteria (White and O'Connell, 1994, p.136):

Fluency: well-formedness of the MT outputs

Adequacy: extent to which meaning expressed in expert translations is also expressed in the MT outputs

Comprehension: amount of information correctly conveyed to readers

Recently, similar to the above-mentioned criteria, fluency and adequacy have become common criteria for human evaluation. According to LDC (2005), which provides a detailed description of how to conduct human evaluation, fluency refers to 'the degree to which the target is well formed according to the rules of Standard Written English'. Human evaluators refer only to the MT output and rate the fluency on a five-point scale:

- 5: Flawless English
- 4: Good English
- 3: Non-native English
- 2: Disfluent English
- 1: Incomprehensible

On the other hand, adequacy refers to 'the degree to which information present in the original is also communicated in the translation' (LDC, 2005). For adequacy evaluation, gold-standard human translation is also provided as a proxy for the original ST. Human evaluators refer to both the human translation and MT output, and rate the adequacy on a five-point scale, which indicates the extent to which the meaning expressed in the human translation(s) is also expressed in the MT output:

- 5: All
- 4: Most
- 3: Much
- 2: Little
- 1: None

Human evaluation is still widely employed to assess MT output quality and we can identify its limitations as follows:

- Human evaluations are subjective and can vary between individuals.
- Large-scale human evaluation is too costly in terms of time and finances.
- Judgement scores themselves do not identify the problem in the MT outputs.

To measure the reliability of human judgements, inter-rater agreement scores such as the Kappa coefficient are calculated (Cohen, 1960; Artstein and Poesio, 2008). Though the interpretation of Kappa depends on the nature of the task, according to Landis and Koch (1977), a value less than 0 means *no agreement*, 0–0.20 means *slight*, 0.21–0.40 means *fair*, 0.41–0.60 means *moderate*, 0.61–0.80 means *substantial*, and 0.81–1 means *almost perfect*. Callison-Burch et al. (2007) reported the Kappa coefficients of human-evaluation tasks conducted in an MT evaluation workshop: 0.250 for fluency and 0.226 for adequacy. The relatively low reliability of the result led them to abandon the idea of using fluency and adequacy evaluation and to instead adopt segment ranking evaluation in which human evaluators are asked to rank several MT outputs from worst to best (e.g. Callison-Burch et al., 2008; Bojar et al., 2015).

2.2.3.2 Automatic evaluation

To address the issues of the subjective nature of human judgement and the cost of large-scale evaluation, a number of automatic evaluation metrics (AEMs) were contrived and widely used in MT evaluation tasks. The basic idea of AEM is that the closer the MT output is to gold-standard human translation(s), the better it is.

The de facto standard AEM in MT research is BLEU (Bilingual Evaluation Understudy) (Papineni et al., 2002), which calculates the n-gram¹³ overlap between the MT output string and human translation(s). Other common metrics, such as NIST (Doddington, 2002), METEOR (Banerjee and Lavie, 2005) and TER (Snover et al., 2006), are also based on n-gram similarity or the edit distance of text strings. However, the above-mentioned AEMs rarely take into account the word order and hence do not work well for language pairs with differing word order, such as Japanese and English. Isozaki et al. (2010a) proposed an alternative AEM specifically for such distant language pairs, called RIBES, which is based on rank correlation coefficients that capture the global word order. The meta-evaluation of AEMs, i.e. calculating the correlation between AEMs and human judgements, revealed that the proposed AEM outperformed conventional methods such as BLEU and NIST for the Japanese-to-English translation data.

The most significant advantages of AEM are that it can be implemented rapidly if human reference translations are available, and that it produces evaluation scores consistently. These factors significantly facilitate MT development process. However, the following shortcomings should be acknowledged:

- There is doubt as to whether or not MT quality can really be evaluated by simply comparing MT output text strings and human reference translations.

- AEM is suitable for corpus-wise evaluation, but not for sentence-wise evaluation (Specia et al., 2010).
- Preparing human translations is still costly in terms of both time and finances.
- Interpreting the meaning of the scores is difficult (*What does a 1.0 increase in BLEU score mean?*).

Interpretation of evaluation results is often overlooked. Achieving high scores is not the final goal. What is important is to assess whether the MT outputs are really useful in certain MT use-case scenarios. If not, it is then necessary to diagnose the problems in order to achieve the required MT quality. AEM scores themselves does not tell us what the specific problems are with the MT outputs. This is also the case with conventional human evaluation. Hence, it is essential to diagnose the outputs in detail. There are several typologies and taxonomies which are useful for MT results (errors) analysis. Vilar et al. (2006) proposed the error typology, which consists of five major categories: ‘Missing Words’, ‘Word Order’, ‘Incorrect Words’, ‘Unknown Words’ and ‘Punctuation’ errors. Comparing several error taxonomies, Costa et al. (2015) developed more comprehensive taxonomy with the following broad categories: ‘Orthography’, ‘Lexis’, ‘Grammar’, ‘Semantic’ and ‘Discourse Style’.

More practically, as mentioned in Section 2.2.2.3, increasing attempts have been undertaken to assess post-editing productivity and/or efforts (e.g. Plitt and Masselot, 2010; Tatsumi, 2010; Koehn and Hermann, 2014; Koponen, 2016), which in turn enables us to directly judge the cost-effectiveness of MT deployment in the translation work flow.

2.3 Controlled language

In this section, we delve further into the concept of controlled language (CL), which was introduced in Section 2.2.2.1. A wide-ranging survey of the field includes the following short definition: ‘A controlled natural language 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). Since we are not concerned here with artificial, formal languages, for the sake of brevity we focus exclusively on controlled languages (CLs). We first give an overview of CLs specifically intended for MT in Section 2.3.1, followed by Japanese based CLs in Section 2.3.2. In Section 2.3.4, we summarise important parameters that should be taken into account when considering CLs for particular use scenarios. Referring to the parameters, we then review previous research on CLs in terms of formulation (Section 2.3.5), deployment (Section 2.3.6) and evaluation (Section 2.3.7).

2.3.1 CL for MT

CLs can be categorised according to the problem they have been constructed to address: ‘to improve communication among humans [...]; to improve [...] automatic translation; and to provide a natural and intuitive representation for formal notations’ (Kuhn, 2014, p.125). This book is not concerned with the third category. Rather, the aim of this book is to improve both monolingual communication and automatic translation, and thereby improve multilingual communication. In our case the constructed language is based on Japanese and the target language of translation is, in the first instance, English.

CLs can be further categorised according to whether they support communication among specialists or with ‘lay’ readers. Many CLs are designed for a specific domain—for example, AECMA (1995) for aircraft maintenance, and Caterpillar Technical English (Kamprath et al., 1998) for engineering—and are characterised by a closed lexicon. SMART Controlled English (Smart, 2006), while imposing unchanging syntactic restrictions, can accommodate different lexicons so it can be used in a range of domains, although each is assumed to be specialised. In contrast, PLAIN (2011) is designed to make official US government documents easier to understand for the general public. The application of CL in this book has the same goal of promoting ‘lay’ understanding in a multilingual society.

With respect to language properties, Kuhn (2014, pp.128–132) identifies four largely independent dimensions for categorising CLs, each divided into five distinct classes, 1–5. First, precision (P) captures the degree to which meaning can be directly retrieved from the textual form. Languages where every sentence is vague to some degree (like natural languages) are categorised as *imprecise languages* (P1). Our aim is to construct a *less imprecise language* (P2) with lower ambiguity and context-dependency. Second, expressiveness (E) describes ‘the range of propositions that a certain language is able to express’. Our CL needs to be on a par with PLAIN (2011), classified as a *language with maximal expressiveness* (E5); the same goal holds for the English translated text. Third, naturalness (N) describes the CL’s proximity to a natural language in terms of readability and understandability. Again, our goal with the Japanese CL is to achieve the categorisation (N5) of PLAIN—‘complete texts and documents can be written in a natural style, with a natural text flow’. However, we may have to accept that, without editing, the English translated document may appear as (N4) where, although single sentences may flow naturally, the text as a whole may not. Fourth, simplicity (S) measures the degree of simplicity or complexity of an exact and comprehensive description of the language. While natural Japanese, like any natural language, is classified as *very complex* (S1), we are attempting to eliminate many of the complex structures by imposing restrictions on it, while taking its description (knowledge, from the author’s perspective) for granted. Crucially, achieving greater simplicity (S2) requires implementing a tool that can identify violations of these restrictions and, ideally, propose legitimate alternatives (see Section 2.5.2).

CLs intended for communication and translation consist of a lexical component and a syntactic component. The former ideally respects the principles of ‘one term–one meaning’ and ‘one meaning–one term’, to eliminate ambiguity and synonymy. Thus, for example, the AECMA standard (AECMA, 1995) allows the use of ‘free’ only as an adjective meaning ‘moves easily’ (e.g. Ensure the fasteners are free.); its use as a verb is proscribed in favour of ‘release’ (e.g. Release the fasteners.). The syntactic component typically comprises between 30 and 60 rules, stated, depending on the particular CL, to varying degrees of specificity; thus it is not always possible to identify shared rules (O’Brien, 2003). It is possible, however, to distinguish a common underlying ambition, which is to eliminate syntactic complexity and ambiguity by restricting sentence length and complexity. As such, sentences are limited to 20 words and the maximum number of clauses is normally two, with constraints on the connectives that are used. Local dependencies are flagged by devices such as obligatory use of the complementiser ‘that’ and of relative pronouns, e.g. ‘which’. This synthesis of rules from Perkins Approved Clear English (Pym, 1990, pp.85–86) typifies many English-based CLs designed for domain-specific communication:

1. Keep sentences short.
2. Omit redundant words.
3. Order the parts of the sentence logically.
4. Do not change construction in mid sentence.
5. Take care with the logic of ‘and’ and ‘or’.
6. Avoid elliptical constructions.
7. Do not omit conjunctions or relatives.
8. Adhere to the PACE dictionary.
9. Avoid strings of nouns.
10. Do not use ‘ing’ unless the word appears thus in the PACE dictionary.

2.3.2 Japanese-based CL

In the case of Japanese CLs, Nagao et al. (1984) devised a controlled grammar to syntactically disambiguate Japanese sentences. Other pioneering work on Japanese CL was conducted by Yoshida and Matsuyama (1985). Although these researchers advocated the need for a Japanese CL in parallel with MT development, little practical implementation resulted from their efforts. This has been largely due to the difficulties involved in producing high-quality output, as the MT task from Japanese to another major language (e.g. English) was significantly more difficult compared to the task of automatically translating between European language pairs, such as English and French.

From the 1990s to the 2000s, research in computational linguistics focused on automatic rewriting (pre-editing) and paraphrasing (Shirai et al., 1998; Yoshimi, 2001; Sato et al., 2003; Inui and Fujita, 2004). However, this work on the automatic processing of natural language could not deal with highly complex or difficult expressions, and the scope of variations of 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 to be 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 entire range of Japanese expression patterns.

Hartley et al. (2012) provided two sets of CL guidelines for technical documents in Japanese: 20 guidelines intended for consumer user manuals and 10 guidelines intended for internal company documents. Below are the guidelines for the latter:

1. Do not use single-byte Katakana characters.
2. Do not use symbols in sentences.
3. Do not use nakaguro (bullet) as a delimiter.
4. Avoid using inappropriate Kanji characters.
5. Avoid creating long noun strings.
6. Do not use ‘perform’ to create a *sa*-verb.¹⁴
7. Avoid topicalisation.
8. Do not connect sentences to make a long sentence.
9. Do not interrupt a sentence with a bulleted list.

Hartley et al. (2012) also investigated the efficacy of CLs with respect to both the readability of the Japanese ST and the quality of the English MT outputs.

Another attempt to create Japanese CL is the on-going ‘Technical Japanese’ project (Watanabe, 2010), which focuses mainly on documents for business purposes. The project published Patent Documents Writing Manual, which consists of 31 rules designed to improve the clarity and translatability of Japanese-language patent texts (Japio, 2013; Matsuda, 2014).

While recent work focused mainly on technical documents for industry and business, there remains much room for investigating other patterns impacting MT performance within the municipal domain.

2.3.3 Document-level CL

The fundamental problem with the current formulation of CL rule sets is that they are almost exclusively specified at the level of the sentence rather than at the level of the document (Hartley and Paris, 2001). According to O’Brien (2003), CL rules can be classified in the following main categories:

1. Lexical
2. Syntactic
3. Textual
 - 3-1. Text Structure
 - 3-2. Pragmatic

O'Brien (2003) further divided them into sub-categories. Rules in the *lexical* and *syntactic* categories are defined at sentence rather than document-level, such as rules governing spelling, synonymy, modifier usage and coordination. While some rules in the *textual* category pertain to the document-level, most of the rules are sentence-internal, such as rules governing sentence length (in *text structure*) and use of metaphor (in *pragmatic*). There exist a few exceptions, such as rules that specify the number of sentences in a paragraph (in *text structure*) and rules that specify verb forms for specific text purposes, e.g. imperative form to signify an instruction (in *pragmatic*).

The need for document-aware CLs has not been ignored; Bernth (2006, p.2) implemented document-level CL rules:

ESG [English Slot Grammar] and EEA [EasyEnglishAnalyzer] already make use of document structure tags, so it was obvious to expand the use of these to help get the document structure, which is important because the decision on which checks to apply depends on which part of the document is being checked.

For example, items in bulleted lists should be consistently written in the same manner, e.g. 'complements of the lead-in sentence, complete sentences, or noun phrases', and can be further specified: 'If the list element is a phrase that continues the lead-in sentence, then the lead-in should end with a colon; if the lead-in is a complete sentence, then it should be terminated with a period' (Bernth, 2006, pp.4–5). Though only several exemplar rules were provided and the full specifications of the entire rule set is not available, this seminal work encourages us to further explore the document-level CLs.

We argue that previously proposed document-level CLs have a limitation: the notion of document element is itself very coarse-grained. The AECMA standard, for example, recognises only three document elements: procedures (instructions for use or action), descriptions and warnings/cautions. There are some cases of linkage between syntactic rules and document elements, e.g. the constraint of 20 words per sentence is relaxed to 25 in descriptions, while warnings are required to begin with a simple, clear command. However, such linkages are not sufficiently specified in the existing CL rules. What is crucially required here is a fine-grained document model to allow rules to be tied to functional elements in order to provide context-sensitive guidance to both authors and to MT systems for translation decision-making. Moreover, generalising the principles of CL across languages requires a language-independent model of document structure that has been lacking hitherto. Such a model is necessarily functionally oriented; DITA, an XML standard for defining documents in terms of their functional constituents described in Section 2.1, affords just such a framework (Hartley, 2010). We address this topic in Chapters 3 and 5.

2.3.4 *Parameters to be considered*

To formulate, deploy and evaluate CLs in certain use scenarios, the following parameters should be taken into account:

- Goal/objective
- Target language
- MT system
- Text domain

Firstly, as previously mentioned, CLs can be categorised depending on their intended goals such as improved human communication and/or improved (machine) translation. Though we can loosely distinguish human-oriented CL and machine/MT-oriented CL, the exact distinction between them is not relevant in our case since both types of CL share a common principle, i.e. to reduce ambiguity and complexity of texts. Indeed, the stated aim of Technical Japanese, mentioned above, is to improve both readability and machine tractability (Watanabe, 2010). However, the compatibility of the both requirements has not been examined in much detail. Reuther (2003) collected a total of 70 writing rules in the German technical writing domain and evaluated the rules with respect to readability and translatability. She concluded that ‘readability rules are a subset of translatability rules’ (Reuther, 2003, p.131). Though this investigation did not focus on machine-translatability and did not conduct an evaluation of the text quality, we can observe an overlap between the CLs for different purposes. Hartley et al. (2012), on the other hand, conducted quantitative human-evaluation experiments to examine the effectiveness of the Japanese CL rules in terms of both readability of the Japanese source and quality of English MT outputs. The result showed ‘no intersection between those rules that boost [machine] translatability and those that enhance readability’ (Hartley et al., 2012, p.243). This suggests that readability does not necessarily guarantee machine-translatability, and vice versa, which is not in accord with what Reuther (2003) concluded. Thus, the compatibility of both requirements should be validated when a CL is developed.

As for the second and third parameters to be considered, O’Brien (2006a) argued persuasively for the need to tune CL rule sets to particular language pairs and MT systems. In order to identify common rules shared by different sets of CL rules, O’Brien (2003) compared a total of eight existing CL rule sets:

- AECMA Simplified English (SE)
- Attempto Controlled English
- Alcatel’s COGRAM
- IBM’s Easy English
- GM’s CASL
- Océ’s Controlled English

- Sun Microsystem's Controlled English
- Avaya's Controlled English

She concluded that only one rule, which recommends short sentences, is shared between all CLs, and that seven rules are common to half or more of the CLs. One of the reasons for the lack of commonality was stated as follows (O'Brien, 2003, p.111):

If source text is destined to be translated by a specific MT system for specific language pairs, then the rules will reflect the inherent weaknesses of the MT system and the known transfer problems between specific language pairs.

The results of evaluation experiments by Hartley et al. (2012) also suggested that 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. This provides the insight that it is important to formulate not only CL rules generally applicable to all MT systems, such as 'avoid long sentences', but also rules specifically tuned to a particular system.

Finally, with respect to text domain, most of the existing CL rule sets were developed for industry (e.g. Pym, 1990; AECMA, 1995; Kamprath et al., 1998; Bernth and Gdaniec, 2001; ASD, 2017), and thus tended to focus on technical documents. We should also be aware of the applicability of CL rules across different domains. Recently, the ACCEPT project has promoted the development of pre-editing rules specifically intended for user-generated content (UGC), i.e. information posted by users in online communities (Seretan et al., 2014b; Gulati et al., 2015). The language usage observed in UGC is characterised by an informal style, abbreviations, slang, proper nouns, and irregular grammar and spelling (Nguyen and Rosé, 2011; Seretan et al., 2014a). The English pre-editing rules 'have been created by adapting the existing Acrolinx rule set for the general domain to our target domains' (Seretan et al., 2014a, p.1795). From the list of rules published online (ACCEPT, 2013a), we notice several rules might be specifically intended for the UGC domain, such as 'capitalize the first letter of a word that begins a sentence' and 'avoid duplicate punctuation marks'. The range of linguistic and stylistic patterns of texts can vary depending on the domain, in other words, the CLs proposed in a certain domain may not sufficiently cover the textual phenomena in another domain.

2.3.5 Formulation of CL

The lack of overlap between different CL rule sets (O'Brien, 2003) leads us to the idea of developing CL rule sets that are tailored to specific use scenarios. There are essentially two approaches to formulate CL rule sets: (1) obtain existing rules and (2) create rules from scratch. The two approaches are complementary, not mutually exclusive.

Reinventing previously proposed rules is not efficient; if existing rules are available, it is preferable to make use of them. To date, a large number of English-based CLs have been developed as outlined by Kuhn (2014) in a comprehensive survey. '[D]ue to confidentiality reasons, some results of studies relating to CL were not published fully; some of the CL rules used also suffered the same fate and have never been made completely public' (Doherty, 2012, p.30), but we can still obtain many rule sets for English. Roturier (2006), for example, consulted nine sources of previously developed English CL rules, reorganised the rules, and finally derived 54 individual, unique rules from them. There have been, however, few CL rules proposed for Japanese; exceptions are listed below:

- Patent Documents Writing Manual (Japio, 2013),¹⁵ the aim of which is to improve human-readability and machine-translatability. It consists of 31 rules focusing on patent documents. All rules are publicly available.
- Simplified Technical Japanese (Ogura et al., 2010), which is intended for improved machine-translatability, focusing on technical documents such as manuals. The full set of rules is not publicly available.
- Easy Japanese (*Yasashii Nihongo*),¹⁶ which was originally proposed for writing information regarding disasters such that it is easy for non-native speakers of Japanese to understand. It defines 12 Japanese writing guidelines regulating the use of difficult words and structures. All rules are publicly available.

Although there are significantly fewer CLs available for Japanese than for English, examining the field of technical writing opens an opportunity to collect Japanese writing rules. Though the technical writing rules may not originally be intended for subsequent MT use, it is reasonable to assume that they can be used as MT-oriented CL rules since the basic principle of technical writing is the same as that of CLs, i.e. reducing the ambiguity and complexity of text. Fortunately, many books about Japanese technical writing have been published, such as JTCA (2011), which we can refer to.

Creating CL rules from scratch is also an important step in formulating an effective and comprehensive rule set optimised to particular purposes, language pairs, MT systems, and text domains. This approach is particularly needed for languages other than English, due to the lack of existing rules that we can utilise. Ogura et al. (2010) created Japanese CL rules by spotting MT errors occurring in their in-house translated texts and identifying Japanese linguistic patterns that appear to be responsible for the MT errors. The Patent Documents Writing Manual was also created based on an analysis of MT errors (Japio, 2013). Kim et al. (2007, p.415) analysed 'the most frequent and fatal translation errors' of their MT system for scientific papers and defined CL rules for Korean. French CL rules used in the ACCEPT project were created from scratch (Gerlach et al., 2013; ACCEPT, 2013a), but the detailed procedure of their CL creation was not reported. In summary, analysing MT errors is a well-tried approach, but the process of creating rules has yet to be detailed and formalised. A systematic CL creation procedure

would give us insight into the intrinsic status of the whole rule set, i.e. whether the CL rule set sufficiently covers the linguistic/textual phenomena that needs to be covered. Thus, there is a need to formalise the rule-creation procedure for CLs.

2.3.6 *Deployment of CL*

Writing or rewriting in accordance with a CL is not an easy task. It is ‘an acquired skill for technical writers’ (Kittredge, 2003, p.443). We can first distinguish two difficulties when CL rules are deployed by human writers and pre-editors.

The first difficulty lies in understanding and accepting CLs, as Mitamura and Nyberg (2001, p.3) outlined as follows:

For successful deployment of controlled language, the authors must accept the notion of controlled language and be willing to receive appropriate training. When authors become accustomed to writing texts in their own style over many years, it may be difficult for them to change their writing style.

Some CL rules may contain linguistic terms that are unfamiliar to writers and, furthermore, some restrictions defined by CLs run counter to writers’ daily language use. It is therefore necessary to provide users with sufficient information to facilitate correct understanding of CLs. For instance, in a book concerning writing guidelines for Global English by Kohl (2008), each rule is presented together with a description of the rule and reference rewriting. For example the rule ‘Keep phrasal verbs together’ (Kohl, 2008, p.38) features the short description ‘Whenever possible, keep the parts of a phrasal verb together’ and the following rewriting example:

Turn the zoom tool **off** by clicking the circle tool.
⇒ **Turn off** the zoom tool by clicking the circle tool.

It also outlines several reasons for following this guideline, such as ‘separated phrasal verbs confuse those non-native speakers who are not accustomed to them’ and ‘this practice is better for machine translation’. In another example, Kohl (2008) provides explanations of linguistic terms, such as ‘determiner’ and ‘tense’. Such detailed information regarding rules helps human writers understand and accept them.

The second difficulty lies in the application of the CL rules. Even if human writers fully understand the rules as such, they may fail to conform to CL rules. When the number of rules is large, it is difficult to properly handle them particularly for non-technical writers. To overcome this, writers should first be trained for controlled writing. Nyberg et al. (2003, p.275) stated as follows:

It seems that authors who receive comprehensive training and who use CL on a daily basis achieve the best results and highest productivity. It is also important that these well-trained authors act as mentors during the training of other authors new to CL.

Though training incurs substantial initial costs, it is an essential step towards effective deployment of CLs. Another way to address the difficulty in applying CLs is a mechanical support, specifically a CL checker. Even professional writers sometimes overlook rule violations in texts. We review literature on the support mechanisms and tools for controlled writing in Section 2.5.2.

2.3.7 Evaluation of CL

Both collected and created CL rules were evaluated in a number of ways. We summarise how CLs can be evaluated using different but related criteria:

Machine-translatability:	Do CLs improve MT quality?
Post-editability:	Do CLs improve post-editing productivity and improve post-edited TT quality?
Source readability/ understandability:	Do CLs improve ST quality?
Human-translatability:	Do CLs improve human translation productivity and improve the translated TT quality?
Usability/feasibility:	Are CLs easy to use for writers or pre-editors?

MT-oriented CL rules were evaluated in terms of machine-translatability (e.g. Bernth, 1999; Bernth and Gdaniec, 2001; Roturier, 2004; Hartley et al., 2012; Seretan et al., 2014a). Bernth (1999) introduced their CL tool, EasyEnglishAnalyzer, and reported on the effectiveness of the implemented rules. When their rules are applied to a technical document, the number of ‘useful’ translations, as judged by a native speaker of the target language, increased dramatically from about 68% to 93%. To assess MT output quality, Roturier (2004) employed the following four rating categories to classify MT results: Excellent, Good, Medium and Poor. The results of a human-evaluation experiment showed that, when a commercial MT system is used, the CL rule set at least doubled the number of Excellent ratings, meaning that MT has attained satisfactory quality and there is no need to modify it. Hartley et al. (2012) conducted a human-evaluation experiment to assess whether CL rules improve the quality of MT outputs. They employed a pairwise comparison method, in which human evaluators compare the MT outputs of CL-noncompliant ST and CL-compliant ST in terms of TT readability. To assess machine-translatability, we can employ not only human-evaluation methods as outlined above, but also automatic evaluation metrics (AEMs) (see also Section 2.2.3). Roturier et al. (2012) explored the possibility of using AEM to assess the impact of source reformulations on MT output quality. Correlating AEM scores and human judgements suggested that AEM is generally consistent with human judgements.

In addition to machine-translatability, post-editability was also evaluated. Evidence of reduced post-editing costs when a CL is employed is provided (Pym, 1990; O’Brien, 2006b; Aikawa et al., 2007; Thicke, 2011). Aikawa et al. (2007), for example, investigated the relationships among CLs, quality of MT outputs and post-editing effort. For quantifying post-editing effort, the character-based edit

distance between MT outputs and their post-edited version was calculated. The result supported the hypothesis that CLs both improve MT quality and reduce post-editing effort.

To evaluate human-oriented CLs, source readability or understandability is an essential criterion. Although CLs are said to be effective in improving source readability (Reuther, 2003; Spaggiari et al., 2003), comparatively less effort has been devoted to providing empirical evidence of the influence of CLs on readability. Improved comprehension by readers of the source documents themselves is demonstrated by authors such as Shubert et al. (1995), Cadwell (2008) and O'Brien (2010). The hypothesis of O'Brien (2010, p.1) is that 'texts written according to CL rules will be "easier to read" than those that have been written without such controls'. When evaluating readability, she defines it 'as being the property of a text which contributes to or detracts from reading ease, as measured by eye movements' (O'Brien, 2010, p.2). Hartley et al. (2012), in contrast, used human-subjective judgements and measured the change in readability after CL was applied. Human evaluators were presented pairs of CL-applied sentences and CL-nonapplied sentences in a random order, and were asked to rate the readability of each sentence on a four-point scale: easy to read, fairly easy to read, fairly difficult to read and difficult to read. The result revealed some trade-off between source readability and machine-translatability. The compatibility of ST and TT quality requires further investigation.

There exists a limited amount of literature reporting the results of experiments on human translatability. Spyridakis et al. (1997) measured to what extent the use of Simplified English can improve quality and ease of translation for native speakers of Spanish, Chinese and Japanese. The results revealed that Simplified English texts were translated significantly better than non-Simplified English texts by Spanish speakers. Though human translatability is beyond the scope of this book, it will be necessary to take it into consideration if we further integrate human translation in the multilingualisation work flow.

While the aforementioned evaluation studies assessed the *effectiveness* of CLs, i.e. to what extent CLs improve the ST/TT quality or post-editing/human-translation productivity, *usage* of CLs can be evaluated in light of usability or feasibility. When CLs are deployed in the work place, it is useful to assess how easily CL rules can be used and how well CLs are accepted by users (writers and pre-editors). This aspect of CLs is inextricably linked to CL assistant tools and will be dealt with in Section 2.5.2.

Finally, it should be noted that CLs can be evaluated both at the level of *rule set* and *individual rules*. To assess the total efficacy of a CL, its whole rule set should be applied at the same time and be evaluated. However, as Nyberg et al. (2003, p.257) stated, 'it is unclear what the contribution of each individual writing rule is to the overall effect of the CL' and '[s]ome writing rules may do more harm than good'. To assess the impact of individual CL rules on MT quality, Roturier (2006) first compiled a test suite which consists of segments violating each of the 54 rules, and prepared pre-CL and post-CL segments to be translated. Based on the human-evaluation results to assess the comprehensibility of French and German

MT outputs, he finally identified the 28 most effective rules. As such, we can say that the steps to evaluate a newly formulated CL are (1) evaluate individual rules, (2) select optimal rules to compile a CL rule set and (3) evaluate the set as a whole.

2.4 Terminology management

Terminology management is defined as ‘any deliberate manipulation of terminological information’ (Wright and Budin, 1997, p.1). Though this definition encompasses a wide range of activities and practical applications, in this section we focus mainly on terminology management in the context of controlled authoring and MT. The importance of well-managed terminology for both controlled authoring (Schmidt-Wigger, 1999; Daille, 2005; Møller and Christoffersen, 2006) and MT (L’Homme, 1994; Vasconcellos, 2001; Hutchins, 2005a; Reynolds, 2015) has long been acknowledged.

To elucidate our standpoint, we first review the notion of ‘terminology’ and ‘term’ in Section 2.4.1 and the main methods of terminology management in Section 2.4.2. We then summarise previous research on controlled terminology construction in Section 2.4.3 and on terminology evaluation in Section 2.4.4.

2.4.1 Terminology and term

According to Sager (1990, p.3), the word ‘terminology’ has three possible meanings:

1. The set of practices and methods used for the collection, description and presentation of terms;
2. A theory, i.e. the set of premises, arguments and conclusions required for explaining the relationships between concepts and terms which are fundamental for a coherent activity under 1; or
3. A vocabulary of a special subject field.

In this book, we use the word in the sense of the third definition. As for the definition of ‘term’, Sager (1990, p.19) detailed it as follows:

The items which are characterised by special reference within a discipline are the ‘terms’ of that discipline, and collectively they form its ‘terminology’

Reviewing existing definitions of ‘terms’, Kageura (2012, p.9) defined them as ‘lexical units *used in a more or less specialised way* in a domain’. These definitions focus on the functional status of terms in a special subject field, domain or discipline.

On the other hand, ISO (2012) defines a ‘term’ as a ‘word, or several words, that denote a concept’, noting that:

In terminology theory, terms denote concepts in specific subject fields, and words from the general lexicon are not considered to be terms. In a TDC

[terminological data collections], however, words from the general lexicon are sometimes recorded in terminological entries, where they are still referred to as “terms”.

Here we notice two points: (1) ‘terms’ denote concepts and (2) general lexicon can be regarded as ‘terms’ in practice. Regarding the second point, Fischer (2010, p.30) noted that translators ‘tend to consider terms in the broader sense, wishing to include everything which makes their work easier into a terminological database’.

In line with the the above observation, in this book we define ‘terms’ as functional lexical units which refer to domain-specific concepts, while taking as wide a definition of ‘terms’ as possible with the practical purpose of translation and authoring in mind.

2.4.2 Methods of terminology management

The terminology management approach can be divided into three categories (ISO, 2012):

Descriptive terminology: ‘approach for managing terminology that documents the way that terms are used in contexts without indicating preferred usage’

Prescriptive terminology: ‘approach for managing terminology that indicates preferred usage’

Normative terminology: ‘approach for managing terminology that is used in standards work or governmental regulation’

Warburton (2015b, p.650) stated that prescriptive terminology ‘is concerned about consistency and quality of terminology and therefore it “prescribes” terms to use and terms to avoid in cases of synonymy’. She also noted that the prescriptive approach ‘is common in institutional terminology management’. It is also worth mentioning that ‘prescriptive’ approach inevitably encompasses the ‘descriptive’ approach as an initial step.

Since there exists no government initiative to standardise terminology in our research context (Japanese municipal domain), we aim to construct controlled terminologies which indicate preferred usage of municipal terms, with a view towards controlled authoring and MT. Hence, we adopt both a descriptive and prescriptive approach for terminology management.

2.4.3 Controlled terminology construction

There are multiple kinds of resources we can utilise for the purpose of domain-specific terminology construction. Dillinger (2001) used the following five categories to summarise possible sources for MT dictionary development: human specialist knowledge, parallel corpora, human-readable dictionaries, machine-readable dictionaries and web-available dictionaries. If existing dictionaries are both easily available and portable, it is preferable to use them. If domain

experts are available, we can make use of human knowledge, but it is, of course, time and labour intensive. At the very least, it is possible to extract terminology from existing textual resources, such as corpora.

The advantages of using corpora (running text) as terminological resources were advocated by Sager (1990, p.132) as follows:

Terminology extracted from running text or discourse offers a greater guarantee of thematic completeness and coherence. All relevant textual variants will be covered and suitable contexts which demonstrate the linguistic behaviour of terms can be selected. Running text also dates the term.

In this review we focus on corpora as resources for terminology construction, since there are no easily available, domain-specific dictionaries in our target domain (Japanese municipalities). Furthermore, it is difficult to find human resources specialising in terminology from this domain, but running text is easily available in the form of, for example, website text.

Broadly speaking, the prescriptive process of controlled terminology construction can be divided into the following three phases:

1. Corpus compilation
2. Term extraction
3. Term variation management

The first two phases pertain to descriptive methods, while the final phase can be regarded as prescriptive. Each phase is outlined in the following subsections.

2.4.3.1 *Corpus compilation*

The initial step towards terminology construction is to compile a corpus. A well-designed corpus enables us to ensure the quality and sufficiency of constructed terminology as a whole. Ideally, we would include whole running texts from our target domain in the corpus. This is, however, almost impossible and we thus seek to *sample* text that is representative of a well-balanced range of texts in the domain. To develop a corpus as a source for term extraction, we need to take the concept of *representativeness*¹⁷ into account. According to Biber (1993, p.243), representativeness refers to ‘the extent to which a sample includes the full range of variability in a population’. Representativeness is also described in terms of text attributes, including ‘how a text was produced and was (intended to be) received, how the authors organized the text in terms of their own goals, when and where, and what language(s) were used in the production of the texts’ (Ahmad and Rogers, 2001, p.734).

The size of a corpus is one of the important aspects to consider. Cabré et al. (2012) stated as follows:

Ideally, the corpus to be analyzed should be large enough to be considered representative of the domain. Unfortunately, there is no precise mathematical formula to determine what should be the size of a corpus to guarantee the sample's representativeness [...] As a general rule, corpus linguists say *Big data is better data*; thus, the corpus should contain as many documents as possible, because the bigger it is, the more terminological units it will contain and the more reliable our conclusions will be.

Sager (2001, p.763) also remarked that corpus linguistics can assist in '[d]etermining what is a sufficiently large body for reliable term extraction', noting '[s]tatistical means can be used to decide when the addition of more text does not produce any new terms'. The sufficiency of the corpus is closely tied to the sufficiency of the terminology. We can paraphrase the remark above as a scenario in which 'the addition of more *terms* does not produce any new terms'. We will return to this issue in Section 2.4.4 from the perspective of terminology evaluation.

A corpus involving more than one language is called a *multilingual* corpus, which includes a *bilingual* corpus in which only two languages are involved. A multilingual corpus can be further classified into two categories, a *comparable corpus* and a *parallel corpus*. Though there is some confusion regarding their exact definitions, in this book we employ definitions used by McEnery and Hardie (2012, p.20), summarised as follows:

Comparable corpus: 'corpus containing components that are collected using the same sampling method, e.g. the *same proportions* of the texts of the *same genres* in the *same domains* in a range of *different language* in the *same sampling period*'

Parallel corpus: 'corpus that contains native language (L1) source texts and their (L2) translations'

In our study we focus on a parallel corpus, since Japanese source texts and their corresponding English translations are freely available from municipal websites. McEnery and Hardie (2012) also highlighted that in order to make a parallel corpus useful, it is essential to align the source texts and their corresponding translations. A properly aligned corpus facilitates not only the human term extraction but also the mechanical extraction of terms.

2.4.3.2 *Term extraction*

Once the textual resources are built, the next phase is to extract terms. From running text, we need to distinguish terms from general words based on their 'termhood'. Kageura and Umino (1996) defined 'termhood' as 'the degree to which a stable lexical unit is related to some domain-specific concepts'.

Traditionally, term extraction is conducted manually; human workers—preferably terminologists, linguists and domain experts—scan a compiled set of resources

and extract terms to be registered in the terminology database. The advantage of manual extraction is that, provided the workers are properly trained and the extracted terms are properly validated, terms in the corpus can be comprehensively and accurately captured. However it is necessary to be aware that it can sometimes be difficult to assess the termhood of a particular expression, as Frantzi et al. (2000, p.7) pointed out ‘[t]here exists a lack of formal or precise rules which would help us to decide between a term and a non-term’ and even ‘[d]omain experts (who are not linguists or terminologists) do not always agree on termhood’.

Manual term extraction is time-consuming and labour-intensive (Sager, 1990; Fulford, 2001), which impedes frequent updating of terminological resources. One solution to this problem is automatic term extraction (ATE), also known as automatic term recognition (ATR), which is a well-established discipline in the NLP and terminology research fields. There have been many attempts to automatically extract terms from corpora and web resources (see Kageura and Umino (1996); Foo (2012) for the survey). ATE has long been used to tackle not only monolingual term extraction (Damerau, 1990; Daille et al., 1996; Frantzi et al., 1998), but also bilingual term extraction (Kupiec, 1993; Gaussier, 1998). The most common approach for bilingual ATE consists of two steps: (1) identify term candidates monolingually and then (2) align the extracted terms of source and target languages (e.g. Daille et al., 1994; Špela Vintar, 2010; Haque et al., 2014). Commercial tools, such as Sketch Engine (Kilgarrieff et al., 2004; Baisa et al., 2015)¹⁸ and SDL MultiTerm,¹⁹ are also available. Focusing on the Japanese–English language pair, Tsuji and Kageura (2004) proposed a language-dependent method to extract low-frequency translation pairs from bilingual corpora. The basic idea was to utilise transliteration in addition to word frequency in order to extract low-frequency translation pairs which are, in many cases, loan-word pairs that can be captured by transliteration patterns. As Tsuji and Kageura (2004, p.28) remarked, ‘more attention should be paid to language-pair-dependent knowledge’, further investigation into language-dependent features will be needed towards effective use of ATE methods.

In practice, the outputs of such ATE methods are not directly included in term collections, but rather are taken as ‘term candidates’ to be verified by human experts within the domain. What needs to be taken into account is not just the precision and recall ability of the ATE, but rather its total efficiency when the human validation process of term candidates is also accounted for. In order to minimise the human effort needed to construct standardised term banks, Foo and Merkel (2010), amongst others, provided a suite of tools for domain experts and linguists to validate automatically aligned term candidates and to control the term variations.

Another option for efficiently collecting terms which merits consideration is collaborative terminology development. Recent advances in web technology have made it easier to provide online collaborative platforms to gain participation from a crowd. Major translation memories, translation-aid tools and localisation platforms often provide terminology management functions which enable terminology sharing. Collaborative approaches contribute not only to efficient

terminology work, but also to the improvement of resultant terminology, as Karsch (2015, p.299) rightly pointed out:

A single terminologist might find good candidates by him or herself. But if the terminologist gets suggestions from a crowd or even better from a community of SMEs [subject matter experts], the ultimate decision for the term might be more suitable and thus longer lasting.

One of the most important points when engaging a crowd is to ensure that suitable roles are assigned to different classes of collaborators. Désilets et al. (2009), for example, implemented wiki-based software for supporting collaborative development of multilingual terminology. The software identifies four groups of users with different permission levels: *anonymous*, *registered*, *editors* and *admin*. *Anonymous* users can search and view the content but cannot modify or add comments to it, while *admin* have permissions to add, modify and delete content, and to change site settings for all users. In regards to crowd resources, Désilets et al. (2009, p.6) observed that ‘it is generally a good idea to give known and trusted contributors more permissions than say, anonymous contributors’. Karsch (2015) identified the steps involved in the terminology tasks, and identified which steps crowd participants should be allowed to conduct. For example, releasing term entries is a task that is not suitable for completion by a crowd, while reviewing term candidates might be assigned to a crowd.

With regard to the task of term extraction, such work is traditionally trusted to terminologists or domain experts. If we employ lay people to extract terms, validating these terms is a crucial step in ensuring the final quality of the terminology.

2.4.3.3 *Term variation management*

The principle of univocity, ‘one term—one meaning’, is a requirement of controlled authoring. In running texts, however, we encounter a number of variant forms of the same concept (referent) in ST and TT. Term variation management is thus required to ensure consistent use of terminology not only by authors, but also by translators and MT. Daille et al. (1996, p.4) provided a definition for term variation as follows: ‘a variant of a term is an utterance which is semantically and conceptually related to an original term’. The extent of term variation is estimated at between 15% and 35%, depending on the domain and text type (Daille, 2005). From the point of view of controlled authoring, it is necessary to construct synsets of a preferred (approved) term and proscribed (prohibited) terms (Møller and Christoffersen, 2006; Warburton, 2014). To maintain consistency of terminology use at the TT when MT is employed, we need to prescribe authorised translations. In this book, we call these well-managed terminologies ‘controlled terminologies’.

To manage term variations and construct controlled terminologies, it is first useful to categorise descriptively the range of term variations. Term variation typologies were proposed in relation to particular applications and languages.

Daille (2005) summarised four different typologies of variations established for (1) information retrieval (IR), (2) text indexing, (3) terminology watch and (4) controlled terminology. Daille (2005, p.189) noted that ‘different definitions of terminological variation are related to the foreseen application’, and, in the case of controlled terminology, ‘the typology seems more relative to the studied corpus and lacks generality’. Møller and Christoffersen (2006) focused on the application of a controlled language lexicon for Danish and identified the following categories of variation: (1) synonyms and homonyms, (2) spelling variants, (3) syntactic variants and (4) compression of terms (abbreviations, acronyms, codes and head words).

As for language pairs, to date substantial studies on term variations have been undertaken for English (Daille et al., 1996) and French (Daille, 2003), while a smaller body of work has been done regarding other languages. In the case of Japanese term variations, Yoshikane et al. (2003) observed actual occurrences of term variations and defined the following variants: modification, decompounding/compounding, coordination, *sahen*-noun-verb variations, noun-noun variations, *na*-adjective-noun variations, and noun-adjective variations. We can easily recognise that some identified patterns are analogous to those in English, such as modification and coordination, while some are unique to the Japanese language such as *sahen*-noun-verb variations. For the purpose of IR-related applications, Yoshikane et al. (2003) focused on texts from scientific articles, particularly titles and abstracts for conference presentations. Thus, there is room for investigating other text domains and applications.

Moreover, from a contrastive viewpoint, fewer studies investigated term variations bilingually (Daille et al., 1994; Carl et al., 2004). Carl et al. (2004) examined English-French bilingual parallel texts and created generalised variation patterns based on Daille (1996) and Jacquemin (2001). The intended application of these patterns is in the area of controlled terminology, and they are as follows (provided with English examples):

Omission:	<i>inclined groove</i> → <i>groove</i>
Insertion:	<i>prone position</i> → <i>prone supported position</i>
Permutation:	<i>c3a1 sniper rifle</i> → <i>sniper rifle c3a1</i> <i>pocket of the shoulder</i> → <i>shoulder pocket</i>
Coordination:	<i>visual acuity</i> → <i>visual ability and acuity</i>
Synonym:	<i>spotting telescope</i> → <i>spotting scope</i>
Writing and	<i>hand stop</i> → <i>handstop</i>
derivation:	<i>re-insert</i> → <i>reinsert</i>

The typologies above supply the point of departure for our study, i.e. investigating Japanese-English term variations for the purpose of building controlled terminology.

The next step in term variation management is to control terminology. The question that must be asked is how to prescriptively define preferred and proscribed terms from variant forms identified. However, the criteria for selecting preferred terms do not appear to be explicitly described in previous literature which is readily available. Here are different but related criteria from separate domain which Warburton (2015a, pp.381–384) presented as term selection criteria (in addition to the classical termhood criterion) in a commercial setting:

1. Frequency of occurrences;
2. Embeddedness;
3. Visibility;
4. Translation difficulty.

Firstly, *frequency of occurrences* is an informative criterion by which to judge the order of preference of particular terms, as ‘frequency can override classical semantic criteria for determining termhood in a commercial setting’ (Warburton, 2015a, p.382). The second criterion *embeddedness* refers to ‘term’s productivity in forming longer multiword terms’ (Warburton, 2015a, p.382). She illustrated the following examples in which the term ‘sustainable development’ is contained (Warburton, 2015a, pp.382–383) (emphasis ours):

1. ecologically **sustainable development** economics;
2. environmentally **sustainable development**;
3. **sustainable development** strategy;
4. **sustainable development** policy planning;
5. World Summit on **Sustainable Development**.

We can reasonably assume that this core term will form other longer multiword terms, and thus registering this term (and its translation) will help to maintain consistent use of numerous other terms which contain this core term. The third criterion *visibility* refers to ‘how prominently a term appears in company materials’, and ‘[t]he more visible a term is, the more important it is to get it right’ (Warburton, 2015a, p.383). Finally, with regard to *translation difficulty*, ‘[t]erms that are difficult to translate are obvious candidates for a term base that is used by translators’ (Warburton, 2015a, p.383).

Cerrella Bauer (2015, p.336) also presented the following criteria for selecting terms to be included in a terminology collection:

1. The frequency of the use in external sources;
2. The opinion of internal domain specialists;
3. The occurrence in established internal and/or external sources;
4. Internal language guidelines;
5. Observance of established standards or;
6. A weighted combination of two or more of these criteria

While the criteria presented by Warburton (2015a) are based on the textual (corpus) evidence, the criteria by Cerrella Bauer (2015) rely more on the external materials and human expert knowledge.

For our task of defining preferred and proscribed terms in the Japanese municipal domain, it is difficult to consult domain specialists or established sources and standards. Nevertheless, we should build clear criteria to assess the term variations, adopting several of the above mentioned criteria.

2.4.4 Evaluation of terminology

Finally, we examine evaluation of constructed terminology. Many attempts have been made to conduct extrinsic evaluation of terminological resources. Quantitative evidence for improved MT output quality and post-editing productivity was provided (e.g. Langlais and Carl (2004); Thicke (2011); see Section 2.2.2.2 for detail). Itagaki et al. (2007) proposed a method to automatically validate terminology consistency in translated texts. They devised a consistency index to check whether extracted compound nouns are translated consistently both in and across texts. This method, in turn, can be utilised for the evaluation of controlled terminologies, i.e. assessing whether and to what extent the consistency of terms can improve if a controlled terminology is provided. While previous studies focused on the translated texts, i.e. MT quality, post-editing productivity and translation consistency, we can use a similar method to evaluate the controlled terminology in terms of the ST quality, controlled authoring productivity and ST consistency.

Compared to the extrinsic effectiveness of terminology, the intrinsic status of terminology such as coverage, i.e. how much of the potential terminology in the given domain is covered by the current terminology, has not been examined to a significant extent. The methodological difficulty involved in validating coverage is due to the fact that the population size of terminology with which to compare is rarely available.²⁰ Sager (2001, p.763), however, indicated that statistical means ‘can be used to decide when the addition of more text does not produce any new terms’. We can tackle this issue by employing a statistical method proposed for inspecting the current status of a corpus (Kageura and Kikui, 2006). The basic evaluation procedure was (1) to estimate the population size of lexical items in the given domain by extrapolating the current size of lexical items in the corpus and (2) to compare the current status in comparison with the population one. Though Kageura and Kikui (2006) applied this method to assess the lexical sufficiency of *corpus* to the travel domain, we assume it is applicable to our task of evaluating the sufficiency of *terminology* in the municipal domain (see Chapter 6).

2.5 Support system for authoring and translation

The practical goal of this research is to create an integrated authoring and multilingualisation system that makes use of customised MT systems. In this section, we review studies on the application of controlled authoring. In

Section 2.5.1, we present a summary of research on document generation, which is the most formal kind of controlled authoring application. In Section 2.5.2, we review existing systems and literature on controlled authoring assistant tools, with a specific focus on sentence-level CL checkers. Finally, we describe the system-evaluation methodologies in Section 2.5.3.

2.5.1 Document generation system

A number of research projects explored the feasibility of generating multilingual instructional text from an underlying conceptual model of the task to be performed (e.g. Hartley and Paris, 1997; Kruijff et al., 2000; Bouayad-Agha et al., 2002; Power et al., 2003; Biller et al., 2005; Paris et al., 2005, 2010). While the output could in principle be constrained to conform to CL rules, since these systems used rule-based text generation, in practice this was not the case. Moreover, they required a full ontological model of the target domain, which is not practicable for the relatively diffuse scope of Japanese municipal documents.

We can refer to a related, thought separate, body of work relating to the broader context of Document Automation (DA). Commercial DA systems exist, e.g. HotDocs,²¹ Exari,²² LogicNets²³ and Arbortext.²⁴ They have been used in the field of technical documentation to produce model-specific product documentation and decision-support systems, as well as in the legal profession to automate the production of custom-built legal documents (e.g. standardised agreements).

These systems provide an environment in which one can specify a document template to produce a set of documents with some degree of variability, thus enabling personalisation. These tools provide mail-merge-like features which are extended with conditional inclusion/exclusion of coarse-grained text units, generally of the order of sections then paragraphs or, less frequently, sentences. DA systems make use of a document structure, which is relevant to our topic, but do not explicitly address issues of multilinguality or translation.

Other authoring environments of interest include those designed specifically for producing personalised documents (Marco et al., 2008; Colineau et al., 2012, 2013). Like commercial DA systems, HealthDoc, which generates tailored health-education documents, enables the authoring of a ‘master document’ which includes conditions that specify when to include various elements in the document (e.g. depending on the intended user). Colineau et al. (2012, 2013) presented a system pertinent to our work in terms of the domain of its application: the generation of public administration documents. All the environments described above make use of a document structure. They open the possibility of eventually having an environment which enables not only the production of understandable multilingual texts in the public administration domain, but also their personalisation so that readers obtain only those parts which are relevant to them. In the domain of public administration, Colineau and her colleagues showed that such personalisation is effective in terms of information-seeking and also considered desirable by citizens.

2.5.2 *Controlled authoring support*

2.5.2.1 *Integrated environment*

To date, a number of authoring support tools and MT systems have been proposed, and some are available both commercially and publicly. Nevertheless, there are not many environments that unify document-/sentence-level authoring and translation functions in an integrated manner. Specifically, the idea of combining document formalisation and CL has been overlooked. While several document authoring software packages, such as FrameMaker,²⁵ XmetaL²⁶ and Oxygen,²⁷ support DITA frameworks and enable authors to compile well-structured content, how to write and translate text in a controlled manner is beyond the scope of these tools. Likewise, while controlled sentence-writing tools, such as Acrolinx²⁸ and MAXit Checker²⁹ (as we will detail in the next section), are also available, they do not fully exploit functional document elements. There is thus a clear need to establish an integrated environment that implements well-structured document templates, CL rules associated with document elements and customised MT systems.

Indeed, it is possible to combine these existing tools to a limited extent; for example, Acrolinx supports FrameMaker. Rather than simply not being feasible, the practical limitations of employing commercial tools are attributable to cost constraints. The aforementioned commercial tools are generally so expensive that Japanese municipalities can simply not afford them within their limited budget. Furthermore, these commercial tools are chiefly intended for technical writers and thus too complicated for laypeople to learn to use quickly. For document-level authoring support, it is sufficient to provide domain-specific document templates to be followed, whose functional document elements are defined in advance. For sentence-level authoring support, there is a need for language- and terminology-checking functions with a simple user interface. Here we emphasise the fact that writing in accordance with a CL and a controlled terminology is more difficult than complying with pre-defined document templates, as the number of CL rules and terms that should be followed is much larger than that of document elements defined in a narrow domain. In the next section, we further detail the sentence-level CL authoring assistant.

2.5.2.2 *CL authoring assistant*

As composing STs in accordance with a CL (including a controlled terminology) is not an easy task, particularly for non-professional writers, offering software support for applying a CL is essential for its practical implementation. At the extreme end, one possible solution is the fully automatic rewriting of source texts, which was explored not only for English (Mitamura and Nyberg, 2001) but also for Japanese (Shirai et al., 1998). However, rewriting without human intervention is a difficult task³⁰ and could induce other, even additional, errors in terms of grammar and style.

A more moderate solution is human-machine interactive writing. When the number of rules to be consulted is large and the application of both broadly and

Table 2.2 Controlled writing processes and support mechanisms

<i>Writing process</i>	<i>Support mechanism</i>
Notice violations of rules Find alternatives	Detect violations Suggest alternatives Provide examples
Decide the best one	Rank alternatives Provide information for decision-making
Rewrite text	Correct text

narrowly defined rules is difficult, supporting human decision-making in authoring becomes essential. CL writing is divided into four processes, of which the support mechanisms can be defined as in Table 2.2.

Neither the idea nor implementation of CL authoring-assistant tools, known as CL checkers, is new (e.g. Adriaens and Schreurs, 1992; Hoard et al., 1992; Bernth, 1997). Commercial tools are currently available to check conformity to general or company-specific writing standards, such as the aforementioned Acrolinx and MAXit. A leading example of a CL writing assistant linked to MT is the KANTOO Controlled Language Checker (Nyberg and Mitamura, 2000; Mitamura et al., 2003; Nyberg et al., 2003), which was designed for producing multilingual documentation in a range of technical domains. The system detects problems occurring in input source texts and provides diagnostic information for authors. Another recent example is the ACCEPT portal,³¹ which provides an integrated web-based environment mainly targeting user-generated community content (see also Section 2.3.4). The predominant advantage of the portal is that it seamlessly combines pre-editing, SMT and post-editing modules to enhance the total productivity of the translation work flow (Seretan et al., 2014b; ACCEPT, 2013c; Roturier et al., 2013). The pre-editing module is implemented by using Acrolinx IQ engine, a rule-based shallow language checker which can detect spelling, grammar, style and terminology errors in text (Bredenkamp et al., 2000; Gerlach et al., 2013). If a pre-editor clicks a CL-violated segment highlighted in the ST, the system suggests a list of alternative expressions and the segment can be automatically corrected by clicking on the most appropriate suggestion. Gerlach et al. (2013) demonstrated that the pre-editing ST have significantly improved MT quality and reduced post-editing time by almost half.³² Several other CL writing assistants have also been proposed for languages other than English, such as Greek (Karkaletsis et al., 2001) and German (Rascu, 2006). Although pioneering research by Nagao et al. (1984) proposed a tool to disambiguate the construction of Japanese sentences, and a small number of commercial text-checking tools for Japanese have recently become available,³³ to date there exist few examples of successful practical implementation of and evaluation results for Japanese CL tools.

CL authoring tools are, for the most part, language-dependent. Though checking functions for some general rules, such as ‘keep sentences short’, and terminology can be implemented in the same manner across different languages, to deal with CL rules specifically categorised for particular languages at structural and grammatical levels, it is necessary to formulate the linguistic rules from scratch to capture the patterns. Moreover, Japanese is so distant from English in structure that English based CL-checking rules are unlikely to be portable to Japanese. Fortunately, the NLP research community in Japan has established a wide variety of language tools; morphological analysers such as McCab³⁴ and JUMAN³⁵ and parsers such as CaboCha³⁶ and KNP.³⁷ We can make use of such resources to develop Japanese CL tools.

Although the existing CL tools are often designed for post-hoc revision or rewriting, for our purposes we envision a more productive work flow of document authoring; authors get feedback from the system if their input violates the CL rule and/or proscribed terms when composing text. Assistance with a ‘drafting-from-scratch’ scenario requires a real-time interactive system, something which previous research on CL tools has failed to address.

2.5.3 Evaluation of authoring system

2.5.3.1 Precision and recall

The performance of CL checkers has been benchmarked in terms of precision and recall of rule violation detection (Adriaens and Macken, 1995; Mitamura et al., 2003; Rascu, 2006). According to Nyberg et al. (2003, p.258), precision and recall are defined as follows:

Precision is the proportion of the number of correctly flagged errors to the total number of errors flagged; recall is the proportion of the number of correctly flagged errors to the total number of errors actually occurring

Mitamura et al. (2003) tested their implemented CL diagnostic component by using a test set which consists of 1302 sentences. The system generated diagnostic messages for 569 sentences (44% of 1302 sentences). Although Mitamura et al. (2003) found out that 35 of 569 sentences received incorrect diagnostic messages (93.8% precision³⁸), recall was not investigated. Calculating recall is more difficult as it is necessary to manually annotate all errors that the system should have detected in the test set.

Ideally we would aim for perfect precision and recall. This is, however, almost unattainable since certain CL rules deal with complicated linguistic or stylistic patterns which require deep analysis of sentence structures and/or semantics. Thus, the balance between precision and recall is the point at issue for the practical deployment of CL checking tools. Generally speaking, there is a trade-off between precision and recall: precision increases at the expense of recall, or vice versa.

Previous literature tends to emphasise the need for high precision. Adriaens and Macken (1995, p.128), for instance, remarked as follows:

[...] it is exactly the precision rate that should be improved as much as possible. Spurious errors are at least misleading, often irritating (especially if there are many of them) and in the worst case they lead to a total rejection of the tool (when the user is sure that the errors are spurious). Missed errors are not so bad (from a user's point of view): mostly, the user will never know there were missed errors at all.

Though it is reasonable to assume that high precision is an essential factor for user-acceptance of the tools, we should also pay attention to the issue of low recall. When considering text quality, it should be noted that CL violations which would significantly degrade MT quality need to be identified as much as possible, in spite of the risk of false detections.

It is worth pointing out that recent advances in computer interfaces enable us to find ways to 'alleviate' the issue of low precision. Even if the system mistakenly flags a CL-compliant segment as a rule violation, we can present it to users in a non-intrusive manner. The question to be asked is as follows: To what extent can users tolerate false alerts generated by the system? To clarify acceptable precision levels, a detailed usability assessment is required (see next section).

2.5.3.2 *Usability*

Usability evaluation is an integral part of developing and deploying any workplace system. According to Nielsen (2012), usability is 'a quality attribute that assesses how easy user interfaces are to use' and is defined by five quality attributes as follows:

- Learnability:** How easy is it for users to accomplish basic tasks the first time they encounter the design?
- Efficiency:** Once users have learned the design, how quickly can they perform tasks?
- Memorability:** When users return to the design after a period of not using it, how easily can they reestablish proficiency?
- Errors:** How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- Satisfaction:** How pleasant is it to use the design?

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 defines these three attributes 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

Here we can see that the both Nielsen (2012) and ISO (2010) have 'efficiency' and 'satisfaction' attributes in common, though there are differences in their descriptions of these attributes. On the other hand, differences can be noticed between them: Nielsen (2012) highlights the user's familiarity with system (see 'learnability' and 'memorability'), while ISO (2010) is more concerned with the user's accomplishments when the system is used (see 'effectiveness').³⁹

To assess system usability, a number of methods have been proposed and implemented. Though an extensive survey on usability studies is beyond the scope of this book,⁴⁰ we focus on work related to NLP (MT and CL) research. Compared to the number of conventional product evaluations of MT by human-subjective judgement or automated metrics as described in Section 2.2.3, the MT research community has published relatively few usability evaluations.

Exceptions include, for example, Alabau et al. (2012), who conducted a user evaluation of their interactive MT systems, comparing the proposed system against a baseline system with limited functionalities. Fifteen university students were recruited as participants and three criteria by ISO (2010) were measured: for effectiveness, the BLEU scores of resultant MT outputs were calculated; for efficiency, the time taken to complete translation was gauged; for satisfaction, a standardised questionnaire, system usability scale (SUS) (Brooke, 1996) was employed. Though the results of the experiment were not conclusive, displaying no statistically significant differences, based on the results, they further improved the interface of the system and conducted an additional usability experiment.

In contrast to the MT system evaluation, Doherty and O'Brien (2012, 2013) and Castilho et al. (2014) evaluated the usability of MT outputs. For example, Doherty and O'Brien (2013) adhered to the ISO definition of usability, and examined the differences in usability between English STs and the raw MT outputs (in Spanish, French, German and Japanese). They implemented an experiment in which a user conducts a series of tasks referring to instructional texts written in their native language and then evaluated the text according to the following four variables: (1) user satisfaction, (2) goal completion, (3) total task time and (4) efficiency. The user satisfaction was measured by post-task questionnaires. The goal completion was gauged by counting the number of tasks that are correctly completed. The total task time was the overall duration taken to complete the whole set of tasks. The efficiency was calculated by dividing the goal completion accuracy (the ratio of successfully completed tasks against the all tasks) by the total task time.

As for the evaluation of CL checkers, Hoard et al. (1992) conducted a pilot user test of a CL checker, which implements AECMA Simplified English rules, to determine whether the checker is indeed helpful in rewriting existing manuals. A 12-sentence passage with 28 rule violations was selected for the evaluation. Twenty writers were separated into two groups of ten and each group was asked to rewrite the text in accordance with the Simplified English rules. One group, Group A, used only the Simplified English, while the other group, Group B, used the Simplified English and an error report from the CL checker. The focus of this evaluation was only on the conformity to CL rules (in the ISO definition, effectiveness), and no time limit was imposed.⁴¹ After the rewriting task, the number of both violations remaining and violations newly introduced was counted. The results showed that, on average, Group A ended up with 12.3 of 28 violations (6.8 remaining and 5.5 newly introduced), while Group B had 5.6 violations (2.7 remaining and 2.9 newly introduced), suggesting their CL checker made a positive impact. Though useful insights can be derived regarding to how to evaluate the controlled authoring system from this seminal work, we observe several limitations.

- The CL checker itself was not used, but the error report from the checker was presented, which might not accurately reflect the probable use scenario of the CL checker.
- The study evaluated only the degree of conformity to the CL—one aspect of effectiveness measure, and checked neither for efficiency nor satisfaction.
- The scale of the experiments was too small for the results to be conclusive. The researchers noted that ‘no statistical significance should be attached to the results of this simple experiment’ (Hoard et al., 1992, p.291).

In short, how usable a controlled authoring system is for the end user remains an open question, which needs to be addressed.

Notes

- 1 ‘Information block’ and ‘information map’ respectively correspond to ‘topic’ and ‘map’ defined in DITA.
- 2 This table is excerpted from Horn (1989, pp.110–111).
- 3 In general, less than 100 documents (complete chunks of text) are collected.
- 4 This figure is redrawn from Kando (1997, p.3).
- 5 This is because **Kando** dealt with wider research disciplines including Japanese literature, which often does not follow IMRD.
- 6 Elements defined by **Kando** do not necessarily guide writers to create research articles.
- 7 This does not mean it is of little use to adopt task-based evaluation methods. Rather, it is important to conduct task-based evaluation after the issue of low text quality is solved.
- 8 While other terms such as ‘automatic translation’ and ‘mechanical translation’ are also used to refer to these systems, in this paper we will use the term ‘machine translation’ and its initialism ‘MT’ for the sake of consistency. We also note that the term ‘machine translation’ is used to denote not only machine translation *systems*, but also the *process* and *output*.

- 9 As of 2020, while NMT became popular in public and commercial settings as NMT systems such as Google Translate have been freely available, many institutions including Japanese municipalities still adopt RBMT and SMT.
- 10 Back translation (BT) is a process of translating target texts which have been translated back to the original language texts.
- 11 National Institute of Information and Communications Technology (NICT), <https://mt-auto-minhon-mlt.ucrj.jgn-x.jp>
- 12 Note that Schwartz (2014) attempted to explore the feasibility of employing monolingual domain experts as post-editors.
- 13 An n -gram is a continuing sequence of n letters/morphemes/words of a given sequence of text.
- 14 *Sa*-verb is a combination of a *Sa*hen-noun (such as *dokusho* or ‘reading’) and a verb (such as *suru* or ‘do’).
- 15 Japio, <https://www.tech-jpn.jp/tokkyo-writing-manual/>
- 16 Hirosaki University sociolinguistics laboratory, <http://human.cc.hirosaki-u.ac.jp/kokugo/tagengoenglish.html>
- 17 See Leech (2007) for the critical discussion of the notion and definition of representativeness.
- 18 Sketch Engine, www.sketchengine.co.uk
- 19 SDL, MultiTerm, www.translationzone.com/products/multiterm-extract/
- 20 If it is available, we no longer need to *evaluate* such a gold standard.
- 21 HotDocs, www.hotdocs.com
- 22 Exari, www.exari.com
- 23 LogicNets, <http://logicnets.com>
- 24 Arbortext, www.ptc.com/service-lifecycle-management/arbortext
- 25 Adobe, www.adobe.com/products/framemaker.html
- 26 JustSystems, <http://xmetal.com>
- 27 Syncro Soft, www.oxygenxml.com
- 28 Acrolinx, www.acrolinx.com
- 29 SMART Communications, www.smartny.com/smart_starterkit.html
- 30 Just consider, for example, disambiguating sentence structure or word sense, and inferring omitted elements.
- 31 The ACCEPT Project, www.accept.unige.ch/index.html
- 32 Detailed evaluation results are available in ACCEPT (2013b).
- 33 For example, Acrolinx caters for several languages including Japanese.
- 34 MeCab, <http://taku910.github.io/mecab/>
- 35 JUMAN, <http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>
- 36 CaboCha, <http://taku910.github.io/cabocha/>
- 37 KNP, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>
- 38 It is calculated by $534/569 = 0.938$.
- 39 Also note that we can assume that the ‘errors’ attribute in the definition by Nielsen (2012) partially captures the ‘effectiveness’ attribute in the definition by ISO (2010).
- 40 See also Nielsen (1993) for an overview of the usability study and Sauro and Lewis (2012) for quantitative methodologies for user research.
- 41 Note they reported that the average amount of time spent on the task was the same for both groups.

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part II

Controlled document authoring

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

3 Document formalisation

In this chapter, we address the research question **RQ-F1**: Can municipal documents be well formalised? The practical purpose of document formalisation is twofold: (1) to provide a document template that helps authors formulate well-structured documents and (2) to provide a basis that enables contextual translation using MT systems. This chapter focuses on document formalisation. The mechanisms for the contextual translation will be detailed in Chapter 5.

There are few standardised document structures available in the municipal domain. In the field of business communication, in contrast, much effort has been made to formalise the structures of technical documents, such as manuals. One of the most widely acknowledged frameworks is the Darwin Information Typing Architecture (DITA) (OASIS, 2010), which we adopt as a point of departure. As we pointed out in Section 2.1.1.1, the functional elements of the Task topic as defined in DITA are too coarse-grained to be used for properly organising municipal procedural documents and specifying detailed linguistic patterns for each element.

The goal of this chapter is to formalise a document structure for municipal procedures by further specialising the document elements provided by the DITA.¹ In Section 3.1, we elaborate on the analysis of functional document elements. We then map the specified elements to a document structure defined in the DITA framework to standardise the structure of municipal procedural documents in Section 3.2. Finally, we summarise the results in Section 3.3.

3.1 Analysis of functional document elements

3.1.1 Documents to be analysed

To investigate the range of functional document elements as broadly as possible, we surveyed websites that provide comprehensible municipal-life information and its human-translated versions in multiple languages, and decided to focus on the following three websites:

CLAIR² Multilingual Living Information provided by the Council of Local Authorities for International Relations (CLAIR).

Table 3.1 Examples of hierarchical levels of websites

Source	Hierarchical levels
CLAIR	Other notifications (1st) > Personal seals (2nd) > Personal seal registrations (3rd)
Shinjuku	Notifications and Procedures to Be Completed at the City Office (1st) > Seal Registration (2nd)
Hamamatsu	Information for Daily Life (1st) > Personal Seal (2nd)

Shinjuku ³	Living Information provided by Shinjuku City, Tokyo, which is one of the largest municipalities in Japan.
Hamamatsu ⁴	CANAL HAMAMATSU provided by Hamamatsu City, Shizuoka Prefecture.

Note that the purpose of **CLAIR** is to offer municipal-life information independent from particular municipalities, while **Shinjuku** and **Hamamatsu** provide information for residents in specific areas, namely Shinjuku City and Hamamatsu City. It is worth noting that **CLAIR** provides more detailed information and has a more complex hierarchical section structure for whole documents than **Shinjuku** and **Hamamatsu**.

In this investigation, we regard the following hierarchical section of the source as *one* document: third level for **CLAIR**; second level for **Shinjuku** and **Hamamatsu** (highlighted in bold in Table 3.1). From our three sources, we identified a total of 373 Japanese documents, of which 123 deal with municipal procedures. Table 3.2 summarises the number of documents under the first level sections of each source.

3.1.2 *Functional elements of municipal procedures*

We ourselves comprehensively extracted the functional elements from the 123 procedural documents and categorised them based on genre analysis (Swales, 1990; Biber and Conrad, 2009) and functional structure analysis (Kando, 1997). The analysis was conducted manually in a bottom-up format as follows:

1. For each sentence (or smaller segment), assign an indicative label which represents what the sentence/segment is about.
2. Combine similar labels and assign a broader label, while revising previously defined labels.
3. Formulate a hierarchical structure of these labels.

Figure 3.1 shows the formulated structure of functional elements we identified. It forms a three-level hierarchy, with the first level comprising four main categories: Categories A–C pertain to municipal procedures and, in particular,

Table 3.2 The number of documents in each category of municipal-life information (the number of procedural documents is in parenthesis)

<i>CLAIR</i>	#	<i>Shinjuku</i>	#	<i>Hamamatsu</i>	#
Status of residence	11(7)	Emergencies	4(0)	Living in Japan	3(3)
A New Residency Management System	8(3)	Preparing for disasters	4(0)	Information for daily life	7(3)
Marriage / divorce	5(3)	Notifications and procedures	7(7)	Medical & health	5(3)
Other notifications	6(4)	Taxes, medical treatment and health insurance	10(4)	Emergency calls	2(0)
Work, technical intern training and training	15(5)	Welfare	10(8)	Disaster prevention	13(0)
Medical	13(1)	Employment	9(4)	Education	7(2)
Pension	4(2)	Childbirth, child-raising and education	11(7)	Labor	4(1)
Childbirth / childcare	19(10)	Living	9(1)	Taxes	2(2)
Other welfare	13(9)	Leisure	7(0)	Rubbish & recycling	13(1)
Education	21(2)	Useful information	10(2)	Child care	13(4)
Japanese language education	3(0)	Consultation / inquiries	6(1)	Learning Japanese	1(0)
Tax	9(6)	Main public facilities	10(0)	Transport	7(2)
Housing / moving	18(1)			Pension	4(1)
Transportation	17(8)			Social welfare	4(4)
Other daily life issues	15(1)				
Emergencies / disasters	14(1)				
Total:	191(63)		97(34)		85(26)

Figure 3.1 Functional elements of municipal procedural documents

Category B includes the core information about procedures, while Category D includes general reference information about municipalities.

Figures 3.2 and 3.3 illustrate examples of the functional elements assigned to actual municipal text. These are documents about personal seal registration excerpted from **CLAIR** and **Hamamatsu**. For each of the textual segments (table cells/list items), an appropriate label is assigned.

Necessary documents	Where to submit application / enquiries	When	Fee
1. The seal which is to be registered 2. An ID confirmation document, issued by a public organisation, with a facial photograph: <ul style="list-style-type: none"> • Resident Card or special permanent resident certificate • Driver's license etc. 	The administrative office of the municipality	Day of application	Registration: Free Issue of personal seal registration certificate: Charge is applied (this varies depending on the municipal administrative office, but is in the region of 300 yen)
B12. Items to Bring	B31. Office	C2. Duration	B41. Charge

Figure 3.2 Example of the functional elements in the procedure for personal seal registration (excerpted from a document in **CLAIR**)

<div>■ Procedures for Personal Seal Registration</div> <div>→ Where</div> <div>* Foreign residents should go the Ward Municipal Services Division of their local ward office.</div> <div>▷ Municipal Services Div.</div>	<div></div> <div></div> <div></div> <div></div>	<div>B31. Office</div>
<div>→ Who</div> <div>In person (see below for details on using a proxy)</div>	<div></div> <div></div> <div></div> <div></div>	<div>B21. Principal (Applicant)</div>
<div>→ Forms to be Submitted</div> <div>Personal Seal Registration Application (<i>inkan toroku shinseisho</i>)</div>	<div></div> <div></div> <div></div> <div></div>	<div>B11. Forms</div>
<div>→ Items to Bring</div> <div><div>• Seal to be registered</div><div>• Personal identification (e.g. residence card or special permanent resident certificate)</div></div>	<div></div> <div></div> <div></div> <div></div>	<div>B12. Items to Bring</div>
<div>→ Cost</div> <div>Free (a fee of 350 yen is charged for re-registration)</div>	<div></div> <div></div> <div></div> <div></div>	<div>B41. Charge</div>
<div>→ Processing Time</div> <div>Same day</div>	<div></div> <div></div> <div></div> <div></div>	<div>C2. Duration</div>

Figure 3.3 Example of the functional elements in the procedure for personal seal registration (excerpted from a document in **Hamamatsu**)

3.2 DITA specialisation

3.2.1 Mapping of document elements

As we saw in Section 2.1.1.1, the functional elements of the Task topic as defined in DITA are still too coarse-grained to sufficiently organise municipal procedures

and specify detailed linguistic patterns for each element. However, as DITA allows for ‘specialisation’, we can instantiate each of the elements DITA defines by default.

We first mapped the functional elements identified in the previous section to the DITA elements:

Prereq	A2, A3, A4, A5
Context	A1, A6, D1
Steps	B1, B2, B3, B4, B5
Result	C1, C2, C31, C32, C33
Example	[N/A]
Postreq	C34, C4, D2, D3

Note that **Example** is specified in DITA by default to support the entire procedure, but we rarely found this element in the municipal procedural documents and decided to omit it at this stage. Based on the mapping results, we further modified the labels and created fine-grained sub-elements of DITA (the right column of Table 3.3).

Prereq can be specialised in line with three types of conditions for municipal procedures: **Personal condition**, **Event condition** and **Item condition**. **Personal condition** defines conditions for the social status of applicants such as ‘those who are 15 years of age or older’ and ‘those who do not have Japanese nationality’. **Event condition** defines conditions related to life events such as ‘when you enter Japan’ and ‘when you come into Shinjuku-city from other municipalities’. **Item condition** defines conditions for physical objects such as ‘Stamps with

Table 3.3 Specialisation of the DITA Task topic to municipal procedures

<i>DITA element in Body (default)</i>		<i>Specified functional element</i>
Prereq	information the user needs to know before starting	Personal condition Event condition Item condition
Context	background information	Explanation (Summary, Purpose, Expiration of validity, Penalty, Related concept)
Steps	main content: a series of steps	Necessary items to bring Place to go Form(s) to complete Payment
Result	expected outcome	Result (Period for procedure, Items to be issued, Contact from local government)
Postreq	steps to do after completion of current task	Guidance to other procedures

imprints that are smaller than an 8-mm square or larger than a 25-mm square', which is the constraint on personal seal size when registering them. **Context** is background information helping readers understand and facilitating smooth conduct of procedures. It provides explanations such as 'It is necessary to register your seal to use it as a registered seal' (**Purpose**) or 'Living in Japan beyond the permitted period of stay is illegal and violators will be deported' (**Penalty**).

Steps, an essential element for properly carrying out municipal procedures, can be specialised into four main functional elements shown in Table 3.3. This element directly specifies readers' actions (e.g. 'bring', 'go', 'submit' and 'pay').

Result is not necessarily required for procedural tasks, but it is useful for readers to know in advance the expected results, such as 'When registration has been completed, you will be issued a personal seal registration card'.

Finally, for **Postreq**, in some cases municipal procedures refer to other related procedures. For instance, documents explaining 'Resident Registration' tend to contain within them the procedure for obtaining a copy of a Residence Record. It is of practical use to provide **Guidance to other procedures** to help readers know related procedures to be conducted as necessary.

3.2.2 Application of specialised DITA

Using this DITA framework helps us to diagnose existing documents and decide how to reorganise them. Figure 3.4 is an excerpt from the 'Seal Registration' document from **Shinjuku**, here annotated in accordance with our DITA elements to highlight several problems.

- **Steps** are scattered in the text.
- **Results** are also scattered in the text.
- **Forms to submit** of **Steps** is missing.
- **Prereq (Personal condition)** abruptly appears in the middle of the text.

■ Personal Seal Registration Procedures

Please bring the personal seal you wish to register along with your [Steps] Items to bring valid residence card or special permanent resident certificate (foreign resident registration card) to the Family and Resident Registration [Steps] Place to go Division of the City Office or local Branch Office, and complete the application procedures in person. Registration can be completed the [Result] Period for procedure same day. Some registration restrictions apply, such as those on age [Prereq] Personal Condition (must be 15 years of age or older).

When registration has been completed, you will be issued a [Result] Items to be issued personal seal registration card. The handling fee for personal seal [Steps] Payment registration is ¥50.

Figure 3.4 Analysis of an existing municipal document using our DITA framework

Personal Seal Registration

Figure 3.5 Model example of the seal registration procedure

Our ultimate objective in applying DITA is to offer document models that non-professional authors can use with ease to create municipal documents from scratch. Figure 3.5 shows a sample Task topic of ‘Seal Registration’ procedure we created as a reformulation of the information in Table 1.1 (Section 1.2)

following the analysis of documents from **CLAIR**, **Shinjuku** and **Hamamatsu**. Since our aim here is to illustrate the advantages of a well-structured document, this sample text is created by ‘recycling’ text spans that instantiate our specialised DITA elements in the publications of these three municipalities. Sentence-level issues of clarity and simplicity are left unaddressed here but are a major focus in Chapters 4–5.

3.3 Summary

To answer the research question **RQ-F1** (Can municipal documents be well formalised?), we comprehensively extracted functional document elements in municipal procedures and created a document model based on DITA, an existing document standard which is widely adopted in technical documentation. After analysing 123 municipal procedural documents, we formulated a typology of functional document elements and mapped the elements into the DITA Task topic structure. The resultant specialised DITA structure demonstrated that it comprehensively accommodates the functional elements of municipal procedures. We can conclude that municipal procedural documents were well formalised.

The formalised document model can be used as (1) a document template that guides writers in terms of what contents should be included and in what way these contents should be arranged to create municipal procedures (see Chapter 7) and (2) a basis for contextual machine translation in combination with controlled language (see Chapters 5 and 7).

As previous studies on genre analysis have not focused on municipal documents, this pilot study provides a reference point for future related studies. Future work should include the analysis of other types of documents, such as the Concept topic and Reference topic of DITA, in order to cover full range of municipal documents.

Notes

- 1 Note that although procedural documents include ‘concept’, ‘task’ and ‘reference’ information (see also Section 2.1.1.1), as a starting point we focus on the Task topic, which forms the core of procedural documents.
- 2 CLAIR (Council of Local Authorities for International Relations) Multilingual Living Information. www.clair.or.jp/tagengo/
- 3 Shinjuku City, Living Information. www.city.shinjuku.lg.jp/foreign/english/index.html
- 4 Hamamatsu City, Canal Hamamatsu. www.city.hamamatsu.shizuoka.jp/hamaeng/

4 Controlled language

This chapter addresses the sentence-level issue of municipal documents, focusing on the concept of controlled language (CL). To answer the research question **RQ-F2** (To what extent can a Japanese CL improve the quality of ST and TT?), we construct Japanese CL rules and conduct human-evaluation experiments to test the effectiveness of the rules on texts from Japanese municipal websites. With practical deployment of CL in mind, we further formulate CL guidelines that are easy to use for non-professional writers in municipal departments. The created CL rules will be implemented in our CL authoring assistant in Chapter 7.

The objectives of the research outlined in this chapter are as follows:

1. To create CL rules for municipal documents.
2. To evaluate the effectiveness of each CL rule in terms of machine-translatability and source readability.
3. To compile CL guidelines that are easy to use for non-professional human writers.

As we stated in Chapter 2, while several Japanese-based CL rules have been proposed in recent years, in actuality few of them focus on or are suitable for municipal documents. A fundamental problem in the field of CL study is that a formalised method by which to create CL rules has yet to be established. Therefore, in this chapter, we propose a protocol to identify the linguistic/textual features by systematically rewriting Japanese STs and analysing the MT outputs. Using this protocol, we formulate Japanese CL rules that are specifically intended for texts in the municipal domain.

Given that municipal documents should serve readers of the original Japanese and English translation equally, we assume that effective CL rules should help to raise the quality of MT output while not degrading the quality of the Japanese ST. To examine the compatibility of these two requirements, we conduct quantitative human evaluation to measure the effectiveness of our CL in terms of both ST and TT quality. As O'Brien (2006a) argued, there is a need to adjust CL rules in accordance with the language pairs and MT systems being used. As such, another point to be clarified is to what extent our CL rules are effective across different

MT systems. To investigate the applicability of CL rules, in this study we compare the results of four major MT systems commercially and publicly available.

From a practical point of view, non-professional municipal writers should be able to use our CL effectively and with little difficulty. Drafting or rewriting texts in accordance with CL rules is a difficult task even for professional writers. In preparation for our software support for human writing in Chapter 7, we compile CL guidelines which provide a detailed description and example rewrites for each rule.

In Section 4.1, we describe how we constructed our CL rules. Section 4.2 explains our experimental setup for human evaluation and presents our results accompanied with a discussion of them. Based on the results, we formulate CL guidelines intended for particular MT systems in Section 4.3. We summarise the findings and contributions of the study in Section 4.4.

4.1 Formulation of CL rules

Our premise is that any CL rule we propose should contribute to the quality of MT outputs without degrading that of the ST.

We first compiled 22 rules based on knowledge regarding technical writing (Section 4.1.1). To investigate more linguistic/textual patterns impacting on MT within the municipal domain, we devised a protocol in which human writers systematically rewrite STs aiming for improved MT quality. Through this protocol, we created an additional 38 rules (Section 4.1.2).

4.1.1 Technical writing based CL rules

We provisionally formulated 47 CL rules, collected mainly from books and documents about general and technical writing guidelines in Japanese. We then conducted a preliminary test on the effectiveness of CL rules by rewriting two illustrative sentences that violate each rule, machine-translating the original and rewritten versions, and comparing the output quality. After this test, a total of 22 CL rules expected to improve MT quality were adopted (Table 4.1). We denote these rules T01–T22, or collectively **CL-T**.¹ It is reasonable to assume that these rules generally have a positive impact on source readability as they were originally intended for human ST readers.

4.1.2 Rewriting trial based CL rules

We assume that comparing the original STs and the more machine-translatable ones rewritten by humans enables us to derive insights into how to (re)write texts amenable to MT.

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

Table 4.1 A list of technical writing based CL rules (**CL-T**)

T01.	Try to write sentences of no more than 70 characters. In no case use more than 100 characters.
T02.	Do not interrupt a sentence with a bulleted list.
T03.	Do not use parentheses to embed a sentence or long expression in a surrounding sentence.
T04.	Ensure the relationship between the subject and the predicate is clear.
T05.	Ensure the relationship between the modifier and the modified is clear.
T06.	Use the particle <i>ga</i> (が) only to mean ‘but’.
T07.	Do not use the preposition <i>tame</i> (ため) to mean ‘because’.
T08.	To express ‘from’, use the particle <i>kara</i> (から). Use particle <i>yor</i> (より) only in comparisons.
T09.	Avoid using multiple negative forms in a sentence.
T10.	Do not use <i>reru/rareru</i> (れる/られる) to express the potential mood or honorifics.
T11.	Avoid using words that can be interpreted in multiple ways.
T12.	Avoid using the colloquial expression <i>ni-narimasu</i> (になります).
T13.	Avoid using the expression <i>to-<i>iu</i></i> (という).
T14.	Avoid using the expressions <i>youna</i> (ような), <i>koto</i> (こと) and <i>mono</i> (もの).
T15.	Do not double-up on words with the same meaning in a single sentence.
T16.	Avoid using the expressions <i>omowa-reru</i> (思われる) and <i>kangae-rareru</i> (考えられる).
T17.	Avoid using the verb <i>okonau</i> (行う) with <i>sahen</i> -nouns.*
T18.	Avoid the single use of the form <i>tari</i> (たり).
T19.	When listing items, make sure they are syntactically parallel.
T20.	Use words from a general Japanese–English dictionary.
T21.	Avoid using compound <i>sahen</i> -nouns.
T22.	Ensure there are no typographical errors or missing characters.

*A *sahen*-noun is a noun which can be connected to the verb *suru* (する) and act as a verb.

1. Rewrite a source text aiming at improved-quality MT output.
2. Record how the text was changed and assess the quality of the output.
3. Repeat steps 1 and 2, until MT output of a satisfactory quality is achieved.

Examples of the detected linguistic features are shown in Table 4.2.

To formulate CL rules for Japanese municipal documents through this protocol, we first extracted 100 original Japanese sentences from municipal websites and conducted the above protocol using three MT systems, i.e. one RBMT system (TransGateway²) and two SMT systems (Google Translate³ and TexTra⁴). We then summarised the linguistic and textual features expected to have a negative impact on MT quality and classified them into five categories: Mood/Modal, Structural, Lexical, Textual/Orthographical and Terminological. We focus on Structural, Lexical and Textual/Orthographical categories, as Mood/Modal features are mainly dealt with in Chapter 5 and the issue of terminology is addressed extensively in Chapter 6. We finally adopted a total of 38 features which had not been covered in **CL-T** and formulated 38 CL rules to regulate these features

Table 4.2 Example of source-text rewriting

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。(Denryoku gaisha ni renraku, shiyo-kaishi-tetsuzuki kanryo-go , bureka no suicchi o iremasu.)	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。(Denryoku gaisha ni renraku shimasu . Shiyo-kaishi no tetsuzuki ga kanryo-shita ato , bureka no suicchi o iremasu.)	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。(Denryoku gaisha ni renraku shite-kudasai . Shiyo-kaishi no tetsuzuki ga kanryo-shita ato ni , bureka no suicchi o irete-kudasai .)	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 4.3). We denote these rules R01–R38, or collectively **CL-R**.⁵ Unlike the technical writing based **CL-T**, these rules are originally MT-oriented.

In this phase of our research, we did not take the variability of the MT systems into account, and extracted a wide range of features. These rules are thus 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 others. In addition, though ST readability was still guaranteed as long as human authors take charge of the entire rewriting process, **CL-R** contains rules which might be perceived as unnatural for the Japanese language. We thus decided to conduct a quantitative human evaluation to determine which rules are effective for which MT system(s) and which rules have a negative impact on source readability.

4.2 Evaluation

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

Table 4.3 A list of rewriting trial based CL rules (**CL-R**)**Structural**

- R01. Avoid using multiple verbs in a sentence.
- R02. Do not omit subjects.
- R03. Do not omit objects.
- R04. Do not use commas for connecting noun phrase enumeration.
- R05. Avoid using particle *ga* (が) for object.
- R06. Avoid using enumeration A-*mo*, B-*mo* (Aも、Bも).
- R07. Avoid using *te-kuru* (てくる) / *te-iku* (ていく).
- R08. Avoid inserted adverbial clauses.
- R09. Do not end clause with nouns.
- R10. Avoid using *sahen*-noun + auxiliary verb *desu* (です).
- R11. Avoid using attributive use of *shika-nai* (しかない).
- R12. Avoid using verb + *you-ni* (ように).
- R13. Avoid using *ka-douka* (かどうか).
- R14. Avoid using *sahen*-noun + *o* (を) + *suru* (する).
- R15. Avoid using *sahen*-noun + honorific *sareru* (される).

Lexical

- R16. Avoid using particle *nado* (など/等).
- R17. Avoid using giving and receiving verbs.
- R18. Avoid using verbose wording.
- R19. Avoid using compound words.
- R20. Do not omit parts of words in enumeration.
- R21. Avoid using suffixes as much as possible.
- R22. Avoid using particle *made* (まで).
- R23. Avoid using particle *de* (で).
- R24. Avoid using particle *no* (の) to mean ‘by’ or ‘from’.
- R25. Do not omit expressions that mean ‘per A’.
- R26. Avoid using conjunctive particle *te* (て).
- R27. Avoid using ‘if’ particle *to* (と).
- R28. Avoid using particle *e-wa* (へは).
- R29. Avoid using particle *ni-wa* (には).
- R30. Avoid using particle *no-ka* (のか).
- R31. Avoid using demonstrative pronouns (*ko*-/*so*-/*a*-/*do*-).
- R32. Avoid using particle *ni* (に) if it can be replaced with particle *to* (と).

Textual/Orthographical

- R33. Use Chinese kanji characters for verbs as much as possible instead of Japanese kana characters.
- R34. Avoid leaving bullet marks in texts.
- R35. Avoid using machine-dependent characters.
- R36. Use punctuations properly to separate sentences.
- R37. Avoid using square brackets for emphasis.
- R38. Avoid using wave dashes (～).

We present MT quality evaluation, followed by source-readability evaluation. We then discuss the compatibility of the both parameters.

4.2.1 MT quality evaluation

4.2.1.1 Experimental setup

DATA

We extracted 13727 sentences from the Toyohashi City website in five categories (public information, Q&A, department information, news articles and topical issues),⁶ and selected sentences violating our rules—four to ten sentences for each rule—resulting in a total of 272 Japanese-original (JO) sentences (120 for **CL-T** and 152 for **CL-R**). Note that for rules other than T01 and T02, sentences not exceeding 70 characters were selected. We rewrote all JO sentences in accordance with each rule, thus generating 272 Japanese-rewritten (JR) sentences. We used four major MT systems that offer Japanese-to-English translation: two commercial RBMT systems (The Hon'yaku⁷ and TransGateway, hereafter, system A and system B) and two freely available SMT systems (Google Translate and TexTra, hereafter, system C and system D). Without user dictionaries or any sort of customisation, the four systems translated the 272 JO and 272 JR sentences into English. The result was 2176 machine-translated sentences: 272 sentences for each label AO (system A–Original), AR (system A–Rewritten) and so on—BO, BR, CO, CR, DO and DR.

QUESTIONNAIRE DESIGN

Our main interest in evaluating the MT quality is to assess whether or not the translation is understandable in terms of the practical information it aims to communicate. We devised a simple method, which focuses on understandability at an acceptable level, disregarding grammatical and lexical errors as long as they do not impair the reader's comprehension.

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.*

Unit.38

Step 2

This is an alternative translation.

Read the sentence, and indicate how close the meaning of this sentence is to your understanding of the translation presented in the previous page.

Please compare the general meaning, not focusing on the difference in the word choice.

Homes and communities are places where children spend a lot of time every day, and where they learn many things about life.

- ☐ Exactly the same meaning
- ☐ Mostly the same meaning
- ☐ Partly the same meaning
- ☐ Completely different meaning

Complete

If you have trouble remembering your understanding from the previous sentence, you can refer to the following translation. Please, however, avoid direct comparison between the two texts; focus on the difference in your understanding between the two.

Families and local communities, children are in place to routinely spend a lot of time, You will learn various in daily life.

Figure 4.1 MT quality questionnaire – Step 2 (when [1] or [2] is selected in Step 1)

In Step 2, the judges were shown a human translation (HT) corresponding to the MT output shown in Step 1. 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 (see also Figure 4.1):

- [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 at this point. At the same time, in order to discourage direct comparison between the two texts, the judges were asked to compare only the overall meaning, not focusing on the differences 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

Table 4.4 Result categories

<i>Selected option</i>	<i>Category</i>	<i>Interpretation</i>
[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

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 human translation as an alternative. They were then asked to indicate how much of it 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 adult English speakers, both native and non-native, as judges. Each judge was asked to evaluate MT sentences that corresponded to the Japanese source sentences but were a mix of translations from eight sources (AO, AR, BO, BR, CO, CR, DO and DR). We assigned three judges to each MT output.

4.2.1.2 Overall result of MT quality

Firstly, we classified the results into four categories (Table 4.4). The numbers in square brackets correspond to the choices in each step as described in Section 4.2.1.1. 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-specific knowledge or contextual information to be understood fully.

Table 4.5 and 4.6 show the percentage of judgements that fell into each category. Overall, applying the CL rules increased the percentage of [MT–Useful] by around 3–6% for all systems except for system D, which shows a slight decrease in the result of **CL-R**.

Table 4.5 Overall results of MT quality (CL-T)

Label	MT-Useful	MT-Inaccurate	MT-Unintelligible	HT-Unintelligible
AO	25.6%	3.3%	66.7%	4.4%
AR	30.0%	6.1%	58.9%	5.0%
BO	8.3%	2.8%	75.0%	13.9%
BR	13.3%	3.1%	73.6%	10.0%
CO	24.4%	5.3%	65.3%	5.0%
CR	30.3%	6.1%	58.6%	5.0%
DO	13.9%	4.2%	69.4%	12.5%
DR	15.6%	4.4%	66.4%	13.6%

Table 4.6 Overall results of MT quality (CL-R)

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%

A comparison between the four MT systems shows the different effects of the CL rules on the four MT systems. While in the most of the systems the application of CL rules decreased the number of [MT-Unintelligible] and increased the number of [MT-Useful], system A for **CL-T** and system C for **CL-R** resulted in a notable increase in the number of [MT-Inaccurate]. The following example shows an [MT-Inaccurate] output generated by system C after the application of rule R14.

Rule R14: Avoid using *sahen-noun* + *o* (を) + *suru* (する).

JO: 会場では、大型スクリーンにそれぞれの議題に関する映像が映され、地域住民が分かりやすく議題の説明をしました。

(*Kaijo dewa, ogata skurin ni soezore no gidai ni kansuru eizo ga utsusare, chiki jumin ga wakariyasuku gidai no setsume o shimasita.*)

JR: 会場では、大型スクリーンにそれぞれの議題に関する映像が映され、地域住民が分かりやすく議題を説明しました。

(*Kaijo dewa, ogata skurin ni soezore no gidai ni kansuru eizo ga utsusare, chiki jumin ga wakariyasuku gidai o setsumei-shimasita.*)

- 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/*setsumei o shimashita*’ (gave an explanation) into ‘説明しました/*setsumei-shimashita*’ (explained) in the source. System C then incorrectly (and unexpectedly) inferred the subject ‘I’ though the true subject ‘地域住民/*chiki jumin*’ (local residents) is present in the source, and generated an understandable but misleading output.

It should be noted that a score of around 5–14% in the [HT–Unintelligible] category shows that even human-translated sentences are sometimes not understandable, implying a fundamental difficulty when 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.

- JR:** 警戒宣言発令後、次の各号に掲げる列車の運転取扱いを実施する。

(*Keikai-sengen hatsurei-go, tsugi no kakugo ni kakageru ressha no unten-toriatsukai o jisshi-suru.*)

- HT:** After an official warning is issued, conduct operation and handling for the trains listed below.

For example, this human translation of JR sentence was evaluated as unintelligible by 5 of 12 judges. Though Japanese speakers can probably guess that this sentence is about how railway companies should take measures when an emergency event (such as an earthquake) occurs, it is difficult for non-native speakers of Japanese to grasp its meaning without context.

4.2.1.3 Generally effective CL rules

To diagnose the effectiveness of each CL rule for different MT systems in detail, we focus on the [MT–Useful] cases. 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 4.7 and 4.8).

We can see that six rules—T02, T17, R10, R13, R25 and R35—have a positive effect on all four MT systems. Some examples of translations of the original sentences and their rewrites are listed below.

Table 4.7 Improvement in [MT-Useful] category (CL-T)

No.	A	B	C	D	No.	A	B	C	D
T01	13.3	0.0	0.0	10.0	T12	-8.3	-16.7	8.3	-25.0
T02	25.0	8.3	25.0	8.3	T13	-8.3	8.3	-16.7	16.7
T03	-8.3	0.0	41.7	8.3	T14	-11.1	0.0	5.6	-5.6
T04	10.0	0.0	-6.7	-3.3	T15	8.3	0.0	-16.7	8.3
T05	6.7	-3.3	6.7	3.3	T16	16.7	8.3	16.7	-8.3
T06	25.0	0.0	0.0	25.0	T17	11.1	5.6	27.8	5.6
T07	16.7	25.0	0.0	-8.3	T18	-8.3	0.0	25.0	16.7
T08	0.0	0.0	8.3	-8.3	T19	16.7	8.3	8.3	-33.3
T09	-16.7	16.7	0.0	-8.3	T20	27.8	16.7	-5.6	0.0
T10	-16.7	20.0	20.0	10.0	T21	11.1	-5.6	-11.1	-5.6
T11	-8.3	16.7	8.3	16.7	T22	-8.3	8.3	-8.3	0.0

Table 4.8 Improvement in [MT-Useful] category (CL-R)

No.	A	B	C	D	No.	A	B	C	D
R01	-25.0	8.3	-8.3	-16.7	R20	16.7	16.7	8.3	0.0
R02	25.0	25.0	25.0	-16.7	R21	8.3	0.0	16.7	-33.3
R03	8.3	33.3	0.0	-8.3	R22	25.0	-16.7	16.7	-25.0
R04	0.0	16.7	8.3	16.7	R23	0.0	16.7	-8.3	-8.3
R05	0.0	8.3	-8.3	-16.7	R24	16.7	-8.3	0.0	-8.3
R06	8.3	-16.7	0.0	-16.7	R25	8.3	16.7	16.7	8.3
R07	-16.7	0.0	8.3	25.0	R26	-16.7	8.3	0.0	0.0
R08	33.3	-16.7	8.3	16.7	R27	-16.7	16.7	-8.3	-25.0
R09	-8.3	0.0	0.0	25.0	R28	25.0	-16.7	16.7	25.0
R10	16.7	16.7	25.0	16.7	R29	0.0	25.0	-8.3	-33.3
R11	-8.3	0.0	25.0	8.3	R30	-8.3	16.7	-16.7	-25.0
R12	8.3	-8.3	8.3	16.7	R31	0.0	0.0	16.7	0.0
R13	16.7	25.0	8.3	8.3	R32	0.0	0.0	-33.3	0.0
R14	-8.3	-25.0	8.3	0.0	R33	-25.0	8.3	-25.0	-16.7
R15	-16.7	8.3	-8.3	25.0	R34	8.3	-25.0	25.0	0.0
R16	0.0	-16.7	8.3	8.3	R35	16.7	25.0	33.3	25.0
R17	0.0	25.0	0.0	-8.3	R36	33.3	-8.3	-8.3	-25.0
R18	0.0	0.0	8.3	8.3	R37	16.7	16.7	0.0	16.7
R19	-16.7	8.3	-8.3	16.7	R38	8.3	-33.3	-16.7	-8.3

Rule R10: Avoid using *sahen-noun* + *desu* (です).

JO: 入場券は4月22日（金）午前10時販売開始です。

(*Nyujoken wa 4-gatsu 22-nichi (kin) gozen 10-ji hanbai-kaishi-desu.*)

JR: 入場券は4月22日（金）午前10時に販売を開始します。

(*Nyujoken wa 4-gatsu 22-nichi (kin) gozen 10-ji ni hanbai o kaishi-shimasu.*)

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/kaishi-desu’ (it is a start) into *sahen*-noun + *suru* construction ‘開始します/kaishi-shimasu’ (to start), which resulted in more natural expressions in the MT outputs.

Moreover, 12 rules—T05, T10, T11, T16, T19, R02, R04, R08, R12, R20, R28 and R37—show a positive effect on three of the systems and can also be regarded as generally effective rules. We provide below examples of one of these rules:

Rule T11: Avoid using words that can be interpreted in multiple ways.

- JO:** またリサイクルのための費用(リサイクル料金)がかかってきます。
 (Mata risaikuru no tameno hiyo (risaikuru ryokin) ga kakatte-kimasu.)
JR: またリサイクルのための費用(リサイクル料金)を支払う必要があります。
 (Mata risaikuru no tameno hiyo (risaikuru ryokin) o shiharau hitsuyo ga arimasu.)
BO: It's requiring a cost for recycling (the recycling charge).
BR: It's necessary to pay a cost for recycling (the recycling charge).
CO: Incoming (recycling fee) for the cost of recycling.
CR: Must be paid (recycling fee) for the cost of recycling.
DO: And the cost for recycling (recycling fees).
DR: You will have to pay the cost for recycling (recycling fees).
HT: In addition, it will be necessary to pay the recycling cost (recycling fees).

In this case, we avoided using the polysemous verb ‘かかる/kakaru’ (put, take, cost) and used ‘支払う/shiharau’ (pay) instead, which resulted in better outputs from three of the MT systems.

4.2.1.4 MT-dependent CL rules

As Tables 4.7 and 4.8 clearly demonstrate, the effectiveness of each CL rule varies depending on MT systems. For example, rule R11 had a positive effect on the output of systems C and D. This is looked at in more detail below.

Rule R11: Avoid using attributive use of *shika-nai* (しか-ない).

- JO:** 温室の植物ですが、この時期にしか花を咲かせないこの花をぜひ一度ご覧ください。
(*Onshitsu no shokubutsu-desu ga, kono jiki ni shika hana o sakase-nai kono hana o zehi ichido goran-kudasai.*)
- JR:** 温室の植物ですが、この時期だけに花を咲かせるこの花をぜひ一度ご覧ください。
(*Onshitsu no shokubutsu-desu ga, kono jiki dake ni hana o sakaseru kono hana o zehi ichido goran-kudasai.*)
- 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 ‘しか-ない/*shika-nai*’ into another attributive particle ‘だけ/*dake*’. For the RMBT 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* is dealt with as a kind of negative construction, which leads to an unnecessary negation of the output. Thus, regulating this expression 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. The example below demonstrates the different effects of the rule on the two SMT systems.

Rule T13: Avoid using the expression *to-iu* (という).

- JO:** 1ヶ月単位で入院費が安くなるという制度があると聞きました。
(*1-kagetsu-tan'i de nyuin-hi ga yasuku-naru to-iu seido ga aru to kiki-masita.*)

Table 4.9 Overall results after optimal rules were selected (CL-T)

Label	MT-Useful	MT-Inaccurate	MT-Unintelligible	HT-Unintelligible
AO	19.9%	1.4%	74.5%	4.2%
AR	34.3%	3.2%	57.9%	4.6%
BO	6.8%	3.7%	79.6%	9.9%
BR	20.4%	3.7%	68.5%	7.4%
CO	24.5%	7.8%	63.5%	4.2%
CR	40.6%	5.2%	50.0%	4.2%
DO	12.5%	4.2%	71.4%	12.0%
DR	22.9%	4.7%	59.4%	13.0%

- JR:** 1ヶ月単位で入院費が安くなる制度があると聞きました。
(*1-kagetsu-tan'i de nyuin-hi ga yasuku-naru seido ga aru to kiki-masita.*)
- CO:** I heard that there is a system that will be cheaper in hospital charges per month.
- CR:** I heard that there is a system that will be lower hospital charges per month.
- DO:** I heard that there is a system called the hospitalisation expenses will be cheaper by 1 months.
- DR:** I've heard that hospitalisation expenses will be cheaper by 1 months.
- HT:** We have heard there is a system wherein hospitalisation costs are made cheaper in 1-month units.

In this case, according to this rule, we omitted ‘*という/to-iiu*’ (called) in the source sentence. System C did not make a significant structural change between CO and CR, while system D successfully omitted the problematic segment ‘called’ in DO, which made DR more understandable, although it unnecessarily omitted ‘there is a system’ in DR.

4.2.1.5 Optimal rules for each system

Tables 4.9 and 4.10 show how much improvement can be achieved if we select effective CL rules for each MT system. We preliminarily selected those rules which produced an increase in [MT-Useful] according to Tables 4.7 and 4.8 and summarised the results as in Tables 4.5 and 4.6.

For all systems, [MT-Useful] category increases by about 10–15% 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.

Table 4.10 Overall results after optimal rules were selected (CL-R)

<i>Label</i>	<i>MT–Useful</i>	<i>MT–Inaccurate</i>	<i>MT–Unintelligible</i>	<i>HT–Unintelligible</i>
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%

4.2.2 *Source-readability evaluation*

4.2.2.1 *Experimental setup*

QUESTIONNAIRE DESIGN

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 sentence readability, we adopted the following evaluation method in line with Hartley et al. (2012) (see Figure 4.2). 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 overall ease of reading.

IMPLEMENTATION

We recruited university students who are native speakers of Japanese as judges. We assigned three judges for each sentence.

4.2.2.2 *Results and discussions*

The readability judgements can be separated into two categories: 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 Tables 4.11 and 4.12 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 ST

Q.23

A	また、各学校、施設においては、次のことを定め、職員に徹底するよう指導するものとする。
B	また、各学校、施設においては、次のことを定め、職員に徹底するよう指導する。

Aの読みやすさについて、4つの選択肢の中から最もあてはまると思うものを1つ選んでください。

- ☐ 読みやすい
- ☐ どちらかといえば読みやすい
- ☒ どちらかといえば読みにくい
- ☐ 読みにくい

Bの読みやすさについて、4つの選択肢の中から最もあてはまると思うものを1つ選んでください。

- ☐ 読みやすい
- ☒ どちらかといえば読みやすい
- ☐ どちらかといえば読みにくい
- ☐ 読みにくい

完了

Figure 4.2 Source-readability questionnaire

Table 4.11 Improvement in Japanese readability (CL-T)

No.	JO	JR	JR-JO	No.	JO	JR	JR-JO
T01	46.7	90.0	43.3	T12	83.3	91.7	8.3
T02	41.7	100.0	58.3	T13	83.3	66.7	-16.7
T03	66.7	66.7	0.0	T14	83.3	100.0	16.7
T04	50.0	96.7	46.7	T15	75.0	100.0	25.0
T05	53.3	86.7	33.3	T16	58.3	91.7	33.3
T06	58.3	100.0	41.7	T17	88.9	88.9	0.0
T07	100.0	91.7	-8.3	T18	83.3	66.7	-16.7
T08	91.7	91.7	0.0	T19	75.0	75.0	0.0
T09	66.7	75.0	8.3	T20	33.3	88.9	55.6
T10	60.0	90.0	30.0	T21	66.7	94.4	27.8
T11	83.3	83.3	0.0	T22	41.7	100.0	58.3

readability, figures greater than or equal to 0.0 (%) in JR-JO are highlighted in bold.

As a whole, 19 out of 22 rules of **CL-T** and 23 out of 38 rules of **CL-R** improved or at least maintained the quality of the source text. In particular, rules T01, T02, T04, T06, T20, T22, R02, R18 and R26 were effective with more

Table 4.12 Improvement in Japanese readability (**CL-R**)

No.	JO	JR	JR-JO	No.	JO	JR	JR-JO
R01	91.7	75.0	-16.7	R20	75.0	50.0	-25.0
R02	50.0	91.7	41.7	R21	100.0	50.0	-50.0
R03	58.3	91.7	33.3	R22	100.0	66.7	-33.3
R04	83.3	58.3	-25.0	R23	91.7	75.0	-16.7
R05	58.3	91.7	33.3	R24	66.7	91.7	25.0
R06	91.7	50.0	-41.7	R25	66.7	91.7	25.0
R07	83.3	100.0	16.7	R26	50.0	91.7	41.7
R08	75.0	75.0	0.0	R27	83.3	91.7	8.3
R09	75.0	58.3	-16.7	R28	66.7	66.7	0.0
R10	58.3	91.7	33.3	R29	66.7	83.3	16.7
R11	83.3	75.0	-8.3	R30	91.7	83.3	-8.3
R12	83.3	91.7	8.3	R31	58.3	58.3	0.0
R13	91.7	75.0	-16.7	R32	100.0	66.7	-33.3
R14	75.0	83.3	8.3	R33	75.0	100.0	25.0
R15	75.0	83.3	8.3	R34	83.3	100.0	16.7
R16	66.7	66.7	0.0	R35	83.3	66.7	-16.7
R17	66.7	100.0	33.3	R36	75.0	91.7	16.7
R18	58.3	100.0	41.7	R37	66.7	66.7	0.0
R19	83.3	75.0	-8.3	R38	100.0	41.7	-58.3

than a 40% improvement in readability. We can observe that, as a whole, rules of **CL-T** are more effective for source readability than those of **CL-R**. This can be explained by the fact that **CL-T** is based on technical writing guidelines that were originally designed to improve the human-readability of source texts, while **CL-R** is originally intended for machine-translatability.

At the level of sentence structure, rules T01, T02 and T04 help to reduce the complexity of sentences, which can lead to improved readability. Moving the location of the predicate before a bulleted list according to rule T02, for instance, enables readers to grasp the meaning of the whole sentence at an early stage and results in reduced cognitive effort to parse sentence structure. Below is an example of the rule.

Rule T02: Do not interrupt a sentence with a bulleted list.

- JO:** 不法投棄については、現在、
1. 監視パトロールの実施
2. 監視カメラ・警告看板の設置
にて、抑止を図っております。
(*Fuho-toki ni tsuite wa, genzai,*
1. *Kanshi-patororu no jishshi*
2. *Kanshi-kamera · keikoku-kanban no secchi*
nite, boshi o hakatte ori-masu.)
- JR:** 不法投棄については、現在、以下の方法にて抑止を図っております。

1. 監視パトロールの実施
 2. 監視カメラ・警告看板の設置
- (*Fuho-toki ni tsuite wa, genzai, ika no hoho nite boshi o hakatte ori-masu.*
 1. *Kanshi-patororu no jisshi*
 2. *Kanshi-kamera · keikoku-kanban no secchi*)

HT: We currently strive to suppress illegal dumping by the following methods.

1. Implementing monitoring patrols
2. Setting up monitoring cameras/and warning placards

Lexical-level simplification is also effective. According to rule R18 (Avoid using verbose wording), for instance, we deleted periphrastic expressions such as ‘ものとする/*mono to suru*’ and ‘こととする/*koto to suru*’, which do not have concrete meaning, but are commonly used in municipal documents in Japan. After this rewrite human evaluators judged JR to be more readable than JO.

In contrast, rules R06, R21 and R38 resulted in more than 40% degradation in readability. Avoidance of suffixes (rule R21) and wave dashes ‘～’ (rule R38) would introduce redundancy by replacing the feature with a longer sequence of words, such as replacing ‘午後1時～4時/*gogo 1-ji ~ 4-ji*’ (1:00–4:00 PM) with ‘午後1時から4時まで/*gogo 1-ji kara 4-ji made*’ (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/*kinenhin-dai-soto-bun*’ (an appropriate amount of money toward the commemorative item) as ‘記念品代に相当する分/*kinenhin-dai ni soto suru bun*’ (an appropriate amount of money corresponding to the commemorative item) according to rule R19 (Avoid using compound nouns) improves readability, while rewriting ‘市民提供資料/*shimin-teikyo-shiryō*’ (materials provided by residents) as ‘市民が提供した資料/*shimin ga teikyo shita shiryō*’ (materials that residents provided) based on the same rule degrades readability. The former compound noun consists of five noun morphemes including two suffixes. We suspect that even human readers find it difficult to parse and regulating it improved the readability. On the other hand, the latter consists of three noun morphemes and is easy for humans to grasp. The avoidance of the compound noun might induce the observed verbosity, leading to readability degradation.

To be more effective, we thus need to further investigate the variance within the same rule and define more strict specifications for each rule.

4.2.3 Compatibility of machine-translatability and source readability

Comparing the results of MT quality (Tables 4.7 and 4.8) and source readability (Tables 4.11 and 4.12), we now discuss the compatibility of the two requirements. Focusing on the generally effective rules, i.e. the rules which are effective for at least three MT systems, we see that 14 rules—T02, T05, T10, T11, T16, T17, T19, R02, R08, R10, R12, R25, R28 and R37—improved or retained source

Japanese readability. In particular, rules T02, T05, T10, T16, R02 and R10 greatly improved both machine-translatability and Japanese readability. This result strongly encourages us to further deploy these rules when composing documents for the municipal domain.

There are, however, some rules which are effective for MT quality in general but have an adverse effect on human-readability, such as rules R04, R13, R20 and R35. Rule R20, 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 R20: Do not omit parts of words in enumeration.

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。

(*Getsu・sui・kin-yobi no gozen 9-ji kara gogo 4-ji made kaisetsu shite-ori, 3-gatsu matsu made kaisetsu shite-imasu.*)

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。

(*Getsu-yobi・sui-yobi・kin-yobi no gozen 9-ji kara gogo 4-ji made kaisetsu shite-ori, 3-gatsu matsu made kaisetsu shite-imasu.*)

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/*getsu・sui・kin-yobi*’ (Monday, Wednesday and Friday) as an elliptic expression and literally translated ‘月/*getsu*’ into ‘month’, ‘水/*sui*’ into ‘water’ and ‘金曜日/*kin-yobi*’ into ‘Friday’ (BO). Restoring an omitted element ‘曜日/*yobi*’ according to rule R20 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 due to the fact that excessive complementation made the sentence longer and hindered reading.

The main finding here is that we need to find a way to meet both the requirement of MT and ST quality. To achieve both machine-translatability and human-readability, it is important to serve different texts to humans and machines. In the case of rule R20, an effective solution would be, for instance, to unpack the elliptic expressions and insert linguistic elements 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

Table 4.13 Selected optimal rules for two MT systems (common rules are shown in bold)

<i>MT system</i>	<i>Selected rules</i>
System B	T01, T02, T05, T06 , T07, T09, T10, T11 , T13, T16, T20, T22, R02, R03, R04, R05, R09, R10, R11, R17, R18, R19, R20, R25, R26, R27, R33, R34, R35, R37
System D	T01, T02, T05, T06, T10, T11 , T18, T20, T21, T22, R02, R03, R04, R05 , R07, R08, R09, R10, R11 , R12, R16, R17, R18, R19, R20, R25, R26, R33, R34, R35, R37

cost of implementing CL rules. We detail the idea of background pre-translation processing to resolve incompatibility of ST and MT quality in Chapter 5.

4.3 CL guidelines

In this section, we compile sets of effective rules to form CL guidelines which are used with positive results by human writers. The results of the previous human evaluation revealed that if we select optimal rules for a particular MT system, we can obtain a significant improvement in machine-translatability. With a practical deployment of our CL rules with Japanese municipalities in mind, we focused at this stage on two MT systems: system B (TransGateway), the most used MT system for Japanese municipal websites, and system D (TexTra), a freely available state-of-the-art SMT system especially intended for Japanese as a source language. It is also crucially important to provide detailed descriptions of rules to enable non-professional writers to fully understand and apply the CL rule set while drafting or revising. In this study, we use the term ‘guideline’ to refer to a set of selected CL rules together with rule descriptions and example rewrites. We compiled two guidelines for systems B and D, henceforth, guidelines B and D.

4.3.1 Selection of CL rules

Based on the evaluation results, we selected rules which met the requirements of machine-translatability. Guideline B comprises 30 rules while guideline D comprises 31 rules. The total number of distinct rules is 36, with 25 rules belonging to the both guidelines (Table 4.13).

Though some rules have positive effects on MT quality but degrade the readability of the ST, in accordance with the previous evaluation results, we decided to postpone tackling this problem of occasional incompatibility between MT and ST quality to the later stage of CL deployment. This is why we retained rules T07, T13, T18, R04, R09, R19, R20 and R35, which might degrade ST readability.

It should also be noted that, after the human evaluation of each rule, we made modifications to two of the rules as shown below:

- Rule T01 was changed to ‘Try to write sentences of no more than 50 characters’ since we observed the limitation of 70 characters is too large to improve quality for Japanese-to-English MT systems as they exist now.
- Rule R15 was merged into T10 since T10 can cover R15.

4.3.2 Detailed description of CL rules

As Tables 4.1 and 4.3 show, some rules are defined broadly (e.g. rule R04, R18 and R19). To remedy this, we first provided more detailed specifications for each rule. For instance, for rule R19 (Avoid using compound words) we decided to define ‘compound word’ as ‘a sequence of more than two nouns (or noun-equivalents)’, given that a compound word consists, at minimum, of a sequence of two nouns. Two-noun compounds are as frequent in technical Japanese as they are in English and proscribing them would severely degrade both expressivity and naturalness. Moreover, they are easily represented in MT systems that accept lexicons and are generally well-captured in the n-gram models of SMT systems.

Most of our CL rules are defined in a proscriptive manner such as ‘avoid using —’ or ‘do not —’, which only specifies what kind of linguistic/textual patterns should be regulated. However, this is not always sufficient for human writers to create CL-compliant sentences or to amend CL-violated segments. For the effective use of CL rules, we provide, for each rule, more detailed descriptions of the rule and at least one example for rewriting.

For instance, let us first look at the following description for rule T10 (Do not use *reru/rareru* to express the potential mood or honorifics).

Description: Use *reru/rareru* only for the passive voice. In order to avoid ambiguity, do not use *reru/rareru* to express the potential mood or honorifics.

This short description tells in what way we should use this particular expression *reru/rareru* in a prescriptive manner as ‘Use —’, as well as in a proscriptive manner. This facilitates the proper understanding of the usage of this rule.

However, the above description of rule T10 still does not specify how to amend violations. Even though human writers can correctly identify CL-violated expressions in texts, they will not necessarily be able to think of alternative expressions that conform to CL. We thus provided two examples to show how to replace *reru/rareru* for both potential mood and honorifics:

Use a direct style instead of honorifics.

登録された方のみ使用できます。

Toroku sareta kata nomi shiyo deki-masu.

⇒ 登録した方のみ使用できます。

Toroku shita kata nomi shiyo deki-masu.

(Only registered users can use it.)

Use a *dekiru* form instead of a potential mood form.

当施設は50人まで受け入れられます。

To-shisetsu wa 50-nin made ukeire-rare-masu.

⇒ 当施設は50人まで受け入れることができます。

To-shisetsu wa 50-nin made ukeireru koto ga deki-masu.

(The facility can accommodate up to 50 people.)

We argue that these examples, which were artificially created by the authors, are succinct enough to be easily grasped by human writers and can be generalised to other cases that would appear in actual texts, while we are aware of the difficulty of comprehensively defining ways to rewrite CL violations.

4.4 Summary

In this chapter, in order to answer the research question **RQ-F2** (To what extent can a Japanese CL improve the quality of ST and TT?), we constructed Japanese CL rules intended for municipal texts and evaluated their effectiveness. We first constructed CL rules using two different ways: (1) collecting information about technical writing and (2) conducting a trial-and-error experiment with source rewriting. We then assessed the effectiveness of the individual rules on municipal documents, with respect to both MT quality and ST readability. This enabled us to identify a total of 18 rules which are effective for at least three MT systems. 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 10% increase in the [MT–Useful] category. On the other hand, the results of the Japanese-source-readability assessment showed that about two thirds of the CL rules improved or at least maintained ST readability. We also observed that degradations in readability for humans often correlate with redundancy generated by the rules. To achieve both machine-translatability and human-readability, it is important to serve different texts to humans and machines. Overall, our CL rules have proved to be effective in improving MT output quality and ST readability, provided we choose optimal rules. We also compiled the two CL guidelines tailored to the particular MT systems and provided detailed descriptions of the rules so that non-professional writers can properly understand and make use of our CL rules. We are now in a position to implement our CL rules in our controlled authoring support system, which we will deal with in Chapter 7.

The original contribution of this study to the field of CL can be summarised as follows:

- We devised an empirical protocol to formulate CL rules, which is applicable to other language pairs, domains and MT systems.
- We revealed that the effectiveness of each CL rule for machine-translatability greatly depends on the MT systems to be used, and also that selecting optimal CL rules is essential to boost the MT performance.
- We examined the compatibility of machine-translatability and source readability.

In future research, we will further investigate the linguistic patterns impacting machine-translatability using our CL formulation protocol. We also plan to examine what linguistic features can be attributed to the degradation of source readability.

Notes

- 1 **T** stands for **T**echnical writing.
- 2 Kodensha, www.kodensha.jp/platform/
- 3 Google, <https://translate.google.com>
- 4 NICT, <https://mt-auto-minhon-mlt.ucri.jgn-x.jp>
- 5 **R** stands for **R**ewriting trial.
- 6 Toyohashi City, www.city.toyohashi.lg.jp
- 7 The Hon'yaku, http://pf.toshiba-sol.co.jp/prod/hon_yaku/

5 CL contextualisation

In this chapter, we create context-dependent controlled language (CL) rules that are linked to particular document elements, which were formalised in Chapter 3. While previous chapters separately dealt with the three components of controlled authoring, namely document formalisation, CL and terminology management, this chapter introduces an integrated approach that combines the document elements and CL rules. To answer the research question **RQ-F3** (Can the combination of controlled language and document structure further improve the TT quality without degrading ST quality?), we define context-dependent CL rules in combination with internal pre-translation processing rules, apply them to four MT systems, and diagnose the translation outputs.

In Section 5.1, we illustrate the linguistic specifications defined within particular document elements in both SL (Japanese) and TL (English). Then, from these specifications we derive context-dependent CL rules. In Section 5.2, we introduce mechanisms to resolve any conflict between the source-side and target-side CL rules in order to achieve the desired linguistic realisations in both languages while using MT. In Section 5.3, we examine the feasibility of pre-translation processing which internally modifies the ST into one amenable to a chosen MT system. Finally, in Section 5.4, we summarise our findings.

5.1 Context-dependent CL

Most of the CL rules previously proposed are context-independent rules. For example, the CL rules ‘Use punctuation properly’ and ‘Avoid compound words as much as possible’, are applicable to almost all document elements. Whilst we dealt with these general rules in Chapter 4, the focus of this chapter is on context-dependent local rules. We describe how to construct context-dependent CLs using the specialised DITA framework we formalised in Chapter 3.

The DITA structure provides a language-independent functional framework. Authors who want to create certain municipal procedures can identify what should be included and how to organise them in a well-structured manner (as in Table 3.3). However, we have yet to establish how to compose and translate each element. In order to instantiate the elements as texts in multiple languages we need to define linguistic specifications, or desired linguistic patterns, for each element

Table 5.1 Linguistic specification

<i>DITA element</i>	<i>Japanese</i>	<i>English</i>
Title	noun phrase	noun phrase with the first letters of words capitalised
Event condition	conditional clause ‘とき/ <i>toki</i> ’ or ‘場合/ <i>baai</i> ’	conditional clause with subject ‘when you’ or ‘if you’
Steps	polite speech style with declarative form ‘します/ <i>shimasu</i> ’	imperative form ‘do’
Context, Result, Postreq	polite speech style ‘です/ <i>desu</i> ’ or ‘ます/ <i>masu</i> ’	[N/A]

with reference to technical writing guidelines, such as JTCA (2011), and actual municipal procedures.

Table 5.1 shows the Japanese and English specifications we defined (also refer to Sections 2.1.1.1 and 3.2.1 for detailed descriptions of DITA elements). For example, **Title** should be written using noun phrases in Japanese and English such as ‘印鑑登録/*inkan-toroku*’ (Seal Registration). Note that the rule to capitalise the first letters of words in titles and headings is applicable to English only. On the other hand, **Event condition** requires a conditional clause, such as ‘日本に来たとき/*nihon ni kita toki*’ (when you arrive in Japan). We also assigned a rather strict pattern for the **Steps** element, polite speech style with declarative form ‘します/*shimasu*’ in Japanese and imperative form ‘do’ in English, such as ‘以下の書類を持参します/*ika no shorui o jisan-shimasu*’ (Bring the following documents), while the constraint is relaxed in **Context, Results** and **Postreq**.

From the linguistic specifications, here we can easily derive CL writing rules for both ST and TT, henceforth CL_{st} and CL_{tt}. One example of CL_{st} is ‘Use conditional clause *toki* (とき) or *baai* (場合) in **Event condition** element’.

5.2 Resolving incompatibilities

Since in our scenario STs are also intended for human readers, we must ensure that our CL improves TT quality without degrading ST quality. One significant issue we face here is that the CL_{st}-compliant ST segment is not always machine-translated as a CL_{tt}-compliant TT. Conversely, a ST whose MT output complies CL_{tt} might not be suitable as a text for source readers. For instance, an MT system translates CL_{st}-compliant Japanese text ‘日本に来たとき/*nihon ni kita toki*’ in **Event condition** as ‘When I came to Japan’, wrongly inferring ‘I’ as the subject. To produce a better result which conforms to CL_{tt}, such as ‘When you came to Japan’, we need to insert the subject ‘あなたが/*anata ga*’ (you) in the source sentence. Here, we acknowledge that inserting subjects in Japanese

CLst: Use polite speech style with declarative form ‘shimasu’: **CLtt: Use imperative form ‘do’**

Figure 5.1 Approaches to resolve incompatibilities between CL_{st} and CL_{tt} for Steps
(* undesirable sentence)

sentences tends to make them sound somewhat unnatural for proficient Japanese speakers.

We thus need to introduce a mechanism to ‘internally reconcile’ the CL_{st} and CL_{tt} in cases of conflict. There are three approaches available to us (in isolation or combination) in order to achieve this goal: (1) MT tuning, (2) pre-translation processing and (3) post-translation processing.¹ We explain each approach using Figure 5.1.

As for (1) MT tuning, DITA elements inform the decision-making of an MT system. For instance, the polite speech style with declarative form ‘*shimasu*’ in Japanese should be translated into English using imperative form ‘do’ in the **Steps** element. Rule-based MT (RBMT) systems can be customised by registering external glossaries, but fine-grained adjustments, such as changes in modality, are difficult to realise. Furthermore, tuning the internal transfer rules is not a realistic option, as most of the off-the-shelf MT systems do not allow access to their internal engines. On the other hand, some statistical MT (SMT) systems offer means to retrain their models by adding training corpora and/or glossaries, which may be a feasible option if we can prepare a large amount of ‘controlled’ parallel texts.

(2) Pre-translation processing is an internal process which further rewrites CL_{st}-compliant sentences into ones specifically amenable to a particular MT system, such as rewriting ST1 into ST2. The advantage of this approach is that if an ST is sufficiently controlled, it is easy to systematically transform it into an internal form to be served to MT. The disadvantage is, however, that it is not necessarily guaranteed to realise desired linguistic forms at TT, since outputs of MT, especially SMT, are unpredictable. For example, as the figure above shows, the imperative form in Japanese ‘*shiro*’ is supposed to be translated into the imperative form ‘do’ in English, but we can not be overconfident that this will be the case. The unpredictability of MT results is examined in the next section.

Finally, (3) post-translation processing corrects a segment in accordance with the target language specifications, such as modifying **TT1** (the MT output of the CL_{st}-compliant sentence **ST1**) into **TT2**. This approach is advantageous in that if we can create comprehensive transformation rules, desirable target-side linguistic realisation can reliably be expected. For example, if we only want to capitalise the first letters of words, this approach is the best as it is sufficient to create simple surface transformation rules.² The difficulty is that, as is the case with pre-translation processing, we must tackle the unpredictability of MT. It should also be noted that if we translate STs into multiple languages, we need to create transformation rules for each of the TLs, resulting in increased development costs. To conclude this section, we emphasise the following points which generally pertain to all the approaches above:

- These fine-grained adjustments which aim to maintain the compatibility of CL_{st} and CL_{tt} are enabled by the employment of specialised document elements.
- **ST1** is served to SL (Japanese) readers and **TT2** to TL (English) readers. In other words, **ST2** and **TT1** are internal forms that are not offered to readers.
- Each approach has advantages and disadvantages, so we need to choose suitable one(s) with application scenarios in mind, combining multiple approaches as necessary.

5.3 Preliminary investigation on pre-translation processing

The previous section introduced mechanisms to resolve the incompatibilities between CL_{st} and CL_{tt} when an MT is employed. Hereafter, we concentrate on the second type of adjustment, pre-translation processing, since it can be naturally extended from the source input control based on the CL. Before implementing the pre-translation processing into our system, it is useful to investigate to what extent our proposed technique achieves its intended result. In this section, we report on the result of a preliminary diagnostic evaluation of the pre-translation processing in a Japanese-to-English translation setting.

5.3.1 Setup

Based on the specification described in Table 5.1, we created the following two transformation rules.³

- Rule 1** Insert a subject ‘あなたが/*anata ga*’ (you) in the conditional clause with ‘とき/*toki*’ or ‘場合/*baai*’ in the **Event condition** element to achieve a conditional clause with subject ‘when you’ or ‘if you’.
- Rule 2** Change the polite speech style with declarative form ‘します/*shimasu*’ into the imperative style (a) ‘しろ/*shiro*’, (b) ‘しなさい/*shinasai*’ or (c) ‘してください/*shitekudasai*’ in the **Steps** element to achieve imperative form ‘do’.

As for Rule 2, there are several forms to express an imperative modal in Japanese. In order to select the optimal one for the particular MT systems to be used, we compare three different imperative forms: (a) ‘*shiro*’, (b) ‘*shinasai*’ and (c) ‘*shitekudasai*’.

We prepared 20 example sentences from the Japanese municipal-life information corpus for each transformation rule, and then manually applied the pre-translation processing to them.⁴ We then translated both sentences before and after applying the processing using four MT systems: two RBMT systems, The Hon’yaku (system A) and TransGateway (system B), and two SMT systems: Google Translate (system C) and TexTra (system D). We analysed whether the MT results successfully achieved the TT linguistic specification as defined above.

5.3.2 Results and analysis

Table 5.2 shows, out of 20 MT outputs, the number of cases that correctly realised the desired linguistic forms in the target language before and after the pre-translation processing was applied.

5.3.2.1 Rule 1

Regarding Rule 1, the transformed versions successfully realised the form ‘when you’ or ‘if you’ in almost all the MT outputs (20, 19, 18 and 18 for systems A–D respectively), while the original versions realised them in only a few outputs except for system C, which realised the desired linguistic form in 13 cases even without pre-translation processing.

We present an example in Table 5.3. Before the subject ‘*anata ga*’ (you) is inserted, system A and B adopted a passive voice, ‘is lost’ and ‘was lost’, which is grammatically correct but not desirable in terms of CL_{TT}. System D failed to process the omission of the subject and produced an ungrammatical sentence, while system C incidentally inferred the intended subject as ‘you’. We observed that system C, an SMT system, tended to predict subjects, and sometimes inferred subjects other than ‘you’ such as ‘Please refer to the procedure to enter the national

Table 5.2 The number of MT outputs that realise the desired linguistic forms in the English translation before and after pre-translation processing

Rule	System A		System B		System C		System D	
	Before	After	Before	After	Before	After	Before	After
1 (<i>anata ga</i>)	3	20	1	19	13	18	4	18
2 (<i>shiro</i>)		19		20		5		5
(<i>shinasai</i>)	1	19	0	20	4	8	2	7
(<i>shitekudasai</i>)		0		0		1		1

Table 5.3 Example MT outputs before and after applying Rule 1 (inserted segment in brackets)

	ST	[あなたが]外国人登録証明書を紛失したとき [Anata ga] gaikokujin-toroku-shomeisho o funshitsu-shita toki
	HT	If you have lost an alien registration card
A	before after	When a Certificate of Alien Registration is lost When you lose a Certificate of Alien Registration
B	before after	When a foreign resident registration card was lost. When you lost a foreign resident registration card.
C	before after	When you lose your alien registration certificate When you have lost your alien registration certificate
D	before after	When the lost certificate of alien registration When you lose the alien registration certificate

Table 5.4 Example of remaining disconformity to CL_{tt}

Transformed ST	[あなたが]印鑑をなくしたり、登録の必要がなくなったとき [Anata ga] inkan o nakushi-tari, toroku no hitsuyo ga nakunatta toki
HT	If you lose your personal seal or no longer need to your stamp to be registered
MT (system C)	Or lost you seal, when the need for registration is no longer

pension when **he** left the company’ and ‘Such as when **I** got salary from two or more locations’.

Focusing on the cases in which the subject is included in the ST, on the other hand, we can observe that all MT systems produced the desired form ‘When you’ in almost all TT. This strongly suggests that Rule 1 is particularly effective in achieving CL_{tt}-compliant TT. Table 5.4 shows, however, the one of the few cases that did not conform to CL_{tt} even after the processing. In this case, system C failed to recognise that the subject ‘*anata ga*’ is applied to two verbs in the ST, and thus generated an undesirable construction. We may be able to improve these cases by inserting the subject repeatedly for each verb.

5.3.2.2 *Rule 2*

Firstly, we observe the effectiveness of the sub-rule which transforms ‘*shimasu*’ into ‘*shiro*’. The transformed versions successfully produced the imperative form ‘do’ in 19, 20, 5 and 5 MT outputs of systems A–D, respectively, while the original versions realised it in 1, 0, 4 and 2 outputs, respectively. We can see significantly different ‘reactions’ between RBMT and SMT; the RBMT systems systematically reflected the ST transformations as we expected, while the SMT systems showed irregular reactions to the changes in STs. As Table 5.5

Table 5.5 Example MT outputs before and after applying Rule 2 ‘*shiro*’ (transformed segment in brackets)

		ST	毎年3月15日までに、前年中の所得を区役所税務課に申告します[→しろ]。 <i>Maitoshi 3-gatsu 15-nichi made ni, zennen-chu no shotoku o kuyakusho-zeimuka ni shinkoku shimasu</i> [→ <i>shiro</i>]
		HT	Report the income during last year to the ward office tax matter department by March 15 every year.
A	before		The income in the previous year is notified to a ward office revenue department by March 15 every year.
	after		Notify the income in the previous year to a ward office revenue department by March 15 every year.
B	before		The income during last year is reported to the ward office tax matter department by March 15 every year.
	after		Report the income during last year to the ward office tax matter department by March 15 every year.
C	before		Until the year March 15, it will declare the income in the previous year to ward office Tax Affairs Section.
	after		Until the year March 15, white declare the income in the previous year to ward office Tax Affairs Section.
D	before		By the end of March 15 each year, the Office’s Tax Department income earned in the previous year.
	after		By the end of March 15 each year, tax or income earned in the previous year to the Office’s Tax Department.

demonstrated, systems A and B realised imperative forms such as ‘Notify’ and ‘Report’ after the reformation of the ST, whereas systems C and D did not.⁵

These tendencies also apply to the next sub-rule which transforms ‘*shimasu*’ into ‘*shinasai*’; the RBMT systems successfully realised the imperative form, but SMT systems did not, though the the number of cases that realised CL_{tt}-compliant TT after the transformation increased slightly compared to the first sub-rule (from 5 to 8 for system C and from 5 to 7 for system D). Table 5.6 shows that systems A, B and C realised imperative forms ‘Pay’, ‘Pay’ and ‘Make’ respectively, while system D produced a passive imperative form ‘Be paid’, which is incorrect.

Finally we look at the third sub-rule which transforms ‘*shimasu*’ into ‘*shitekudasai*’. According to Table 5.2, at first glance this rule did not appear to have positive—but even negative—effects on the realisation of the target linguistic specification, as there were no or very few cases where desired imperative forms, such as ‘do’ were correctly produced.

However, detailed observation of the MT outputs revealed an important point. Table 5.7 shows that, after pre-translation processing, all MT outputs achieved polite-style imperative form ‘please do’; ‘Please accept’ (system A) and ‘Please receive’ (systems B–D). We found that the number of MT outputs that realised this form is 19, 20, 17 and 15 for systems A–D. The main finding here is that even SMT systems are generally consistent in producing the typical construction ‘please

Table 5.6 Example MT outputs before and after applying Rule 2 ‘*shinasai*’ (transformed segment in brackets)

	ST	その納付書により納付します[→しなさい]。 <i>Sono nofusho ni-yori nofu-shimasu</i> [→ <i>shinasai</i>]
	HT	Pay as indicated on the slip.
A	before	It pays by the statement of payment.
	after	Pay by the statement of payment.
B	before	It’s paid by the statement of payment.
	after	Pay by the statement of payment.
C	before	You pay by the payment notice.
	after	Make payment by the payment notice.
D	before	Paid by the payment.
	after	Be paid by the payment.

Table 5.7 Example MT outputs before and after applying Rule 2 ‘*shitekudasai*’ (transformed segment in brackets)

	ST	払込証明書を受け取ります[→ってください]。 <i>furikomi-shomeisho o uketorimasu</i> [→ <i>tte-kudasai</i>]
	HT	Receive a payment certificate
A	before	A payment certificate is received.
	after	Please accept a payment certificate.
B	before	A payment certificate is received.
	after	Please receive a payment certificate.
C	before	You will receive a payment certificate.
	after	Please receive the payment certificate.
D	before	You receive a certificate of payment.
	after	Please receive a certificate of payment.

do’, which enables us to further rewrite MT outputs into CL_{tt}-compliant TT in a systematic manner by omitting ‘please’ and capitalising the first letter of verb.

At this stage it is more effective to incorporate post-translation processing to further modify MT outputs, particularly when SMT is adopted. Here, we propose an integrated flow of pre- and post-translation processing to effectively meet the requirements of both CL_{st} and CL_{tt} in **Steps** elements (Figure 5.2). **ST1** is internally transformed into **ST2**, with the polite-style imperative ‘*shitekudasai*’. MT first translates **ST2** as **TT1** in the form ‘please do’, which is then transformed into **TT2**, thus complying with the CL_{tt}, with the use of imperative form ‘do’. We also argue that this processing can be fully automated by surface pattern-matching modification.

CLst: Use polite speech style with declarative form ‘shimasu’

CLtt: Use imperative form ‘do’

Figure 5.2 Revised flow of pre- and post-translation processing for Steps (* undesirable sentence)

5.4 Summary

This chapter addressed the research question **RQ-F3** (Can the combination of controlled language and document structure further improve the TT quality without degrading ST quality?). To answer the question, we first created CL rules that are linked to the functional document elements defined in Chapter 3. To resolve the incompatibility between source- and target-side CL rules, we investigated the possibility of pre-translation processing, which transforms a CL-compliant ST into a machine-translatable internal form. We implemented two transformation rules and diagnosed the four different MT results to see whether pre-translation processing is really effective in the realisation of desired MT outputs. According to the results, we can conclude that overall the pre-translation processing is particularly effective for the RBMT systems. If we use the SMT systems, we need to face the issue of unpredictable results. A detailed analysis of results suggested that if we additionally introduce post-translation processing, we can expect desired results in TT even when using SMT.

The implementation of the pre- and post-processing into our authoring system will be addressed in Chapter 7. The transformation rules presented in this chapter can be easily implemented based on the morphological information of the ST.

In future work, we will formulate more document-element-dependent CL rules by investigating the linguistic features peculiar to each document element. We also intend to explore the MT tuning approach by retraining statistical models of SMT using controlled bilingual corpus tailored to each document element.

Notes

- 1 Though the pre-translation (post-translation) processing is a mode of pre-editing (post-editing), in this book we use the term to denote internal processing specifically

employed to realise a CL_{tt}-compliant sentence from its CL_{st}-compliant counterpart using MT.

- 2 Note that it is difficult to achieve capitalisation of English letters in the TT if only using pre-translation processing on the Japanese ST.
- 3 As we pointed out in the previous section, the linguistic specification ‘noun phrase with the first letters of words capitalised in **Title** element’ can be easily realised by simple post-translation processing, so we decided not to address it here.
- 4 Details of the corpus are described in Section 6.1.1
- 5 We also found that system D tends to recognise ‘*shiro*’ wrongly as a noun ‘白’ (white) such as ‘white declare’ in the example.

6 Terminology management

This chapter tackles the terminology issue of municipal documents. To answer the research question **RQ-F4** (Can municipal terms be comprehensively captured and well controlled?), we construct controlled terminologies for the municipal domain from scratch and evaluate them. As mentioned in Section 2.4, well-managed terminology is essential for authoring and (machine) translation. Our controlled terminologies are thus multi-purpose—for controlled authoring and for MT, which will be incorporated into our authoring support system in Chapter 7.

In relation to controlled language, general vocabulary can be controlled to facilitate the exact usage and interpretation of lexical items. For instance, AECMA (1995) provides a controlled vocabulary consisting of approximately 2300 headings. Rather than embarking on vocabulary control for a wider range of lexical items, however, we focus here on domain-specific terms, because (1) we can reasonably assume that municipal writers can easily assimilate the controlled usage of terminology as they are already familiar with the municipal domain, and (2) the control of technical terminology, among the wider range of lexical items, is expected to contribute much to the quality of ST and TT (Wright and Budin, 2001; Warburton, 2015b).

The practical goal of terminology management is twofold: (1) to enable consistent use of terminology by human authors in the form of a terminology checker; and (2) to improve the MT quality in the form of an MT dictionary. To that end, it is required to construct a list of synsets of preferred and proscribed terms for the source language (Warburton, 2014), and to define the authorised translations for each of the synsets.

For example, we can find variant forms of the same referent such as ‘印鑑登録証明書/*inkan-toroku-shomeisho*’ and ‘印鑑証明書/*inkan-shomeisho*’. As the former might be a preferred or standard term in the municipal domain, we can define the latter as a proscribed or prohibited term. In the target-language texts, on the other hand, we also encounter various translations that correspond to the source terms such as ‘seal registration certificate’, ‘personal seals registration certificate’ and ‘personal seal proof certificate’. To maintain the consistency of the terminology use on the target side, we need to prescribe authorised translations.

The problem here is that there are few (bilingual controlled) terminologies or dictionaries maintained in the municipal domain.¹ It is difficult to give a positive proof that some terms should be authorised and some should not.

Therefore, we decided to take descriptive and prescriptive approaches: we first observe a wide range of term occurrences in municipal documents and then specify which terms should be preferred, also referring to existing dictionaries. More specifically, the construction of our terminology proceeds in the following steps: (1) manually collect Japanese and English municipal terms from municipal websites; (2) gather variant forms of terms and create a typology of term variation; (3) define preferred and proscribed forms.

Our terminologies should be not only accurate but also comprehensive. Although a number of automatic term extraction (ATE) methods have been proposed (see Section 2.4.3.2), we adopt manual extraction of terms from the corpus at step (1), because the ATE methods are still imperfect at this time and, more importantly, most of them are inclined to precision (accuracy) rather than to recall (comprehensiveness) (Zhang et al., 2008). In a practical sense, our target domain is narrowed down to municipal-life information, and it is feasible to resort to human effort to do the terminology work.

To validate the sufficiency of our constructed terminology in the domain, we examine the coverage of the constructed terminologies. Using techniques for predicting the vocabulary growth (Baayen, 2001; Evert and Baroni, 2007), we estimate to what extent our terminologies cover the potential range of terms and concepts in the domain.

Section 6.1 describes how we extract bilingual terms from the municipal parallel corpus and control the variations based on the typology of terminology variations. In Section 6.2, we validate the coverage of our constructed terminologies, comparing them with the estimated population size of terms and concepts in the municipal domain. Finally, Section 6.3 summarises the findings and presents directions for further compilation and improvement of terminologies.

6.1 Terminology construction

6.1.1 *Municipal parallel corpus*

Since there are no bilingual parallel corpora freely available in the municipal domain, we first constructed one by extracting texts from municipal websites. At this stage, focusing on the municipal-life information, we continue to use the three websites as sources: **CLAIR**, **Shinjuku** and **Hamamatsu** (see also Section 3.1.1). The **CLAIR** website provides general-purpose life information independent to particular municipalities, while **Shinjuku** and **Hamamatsu** websites provide life information pertaining to the particular municipalities. Moreover, it should be noted that the source texts of **Hamamatsu** are written in *Yasashii Nihongo* (Easy Japanese), the lexicon and grammar of which are simplified in order to make texts easier to read for non-native speakers of the language. We can suppose that the three websites cover a wide range of contents and linguistic phenomena.

Table 6.1 Basic statistics of extracted sentences

	<i>Japanese</i>			<i>English</i>			<i>Aligned</i>		
	<i>Token</i>	<i>Type</i>	<i>TTR</i>	<i>Token</i>	<i>Type</i>	<i>TTR</i>	<i>Token</i>	<i>Type</i>	<i>TTR</i>
CLAIR	6385	5236	0.820	6016	5000	0.831	6006	5123	0.853
Shinjuku	4938	3322	0.673	4926	3317	0.673	4926	3456	0.702
Hamamatsu	5418	4838	0.893	4561	4058	0.890	4459	4101	0.920
All	16741	13216	0.789	15503	12268	0.791	15391	12620	0.820

To create a parallel corpus of Japanese and English, we (1) manually extracted all Japanese and English sentences from the websites, including body texts, headings and texts in tables, and (2) aligned the sentences between the languages. The basic statistics of the extracted sentences are shown in Table 6.1.

At the alignment step, for some Japanese sentences, there were no corresponding English sentences, and vice versa. That is why the total number of Japanese sentences (16741), English sentences (15503) and aligned sentence pairs (15391) are different. In the following, we focus on the 15391 aligned sentence pairs to construct bilingual terminologies.

TTR in Table 6.1 stands for type/token ratio, which is calculated by dividing the types (the total number of distinct items) by the tokens (the total number of items). The TTR of all aligned sentence pairs is 0.820. This means that a substantial proportion of the aligned sentence pairs (tokens) are repeatedly used in the corpus. For instance, the aligned pair of a Japanese source sentence and its English translation ‘詳しくはお問い合わせください/*Kuwashiku wa otoiawase kudasai*’ (Please inquire for more information) are observed three times in the corpus.²

6.1.2 Manual extraction of terms

6.1.2.1 Terms to be captured

We focus on noun phrases (including single words) as terms. Our aim is not to create a standard dictionary of the municipal domain, but, more practically, to provide terminologies useful for authoring and (machine) translation. As Fischer (2010, p.30) pointed out that translators ‘tend to consider terms in the broader sense, wishing to include everything which makes their work easier into a terminological database’, we thus decided to collect terms as widely as possible. The range of the terms to be captured is defined as below.

1. Technical terms and proper nouns, for example:

- 外国人登録証明書/*gaikokujin-toroku-shomeisho* (alien registration card)
- JR西日本/*JR-nishi-nihon* (JR West Japan)

2. More general words that refer to municipal services and activities, for example:

- 収入印紙/*shunyu-inshi* (stamp)
- 外交活動/*gaiko-katsudo* (diplomatic activity)

Legal and medical terms such as ‘最低賃金法/*saitei-chingin-ho*’ (Minimum Wage Law) and ‘問診票/*monshin-hyo*’ (medical questionnaire) are frequently used in the municipal domain. We also regard these terms as municipal terms.

At this extraction phase, we normalised the following four types of variations:

1. Expand expressions coordinating multiple terms.
combustible and non-combustible garbage
⇒ (1) combustible garbage, (2) non-combustible garbage
2. Expand expressions wrapped in parenthesis.
認可保育所（園）/*ninka hoiku-sho (-en)* (accredited nursery school)
⇒ (1) 認可保育所/*ninka hoiku-sho*, (2) 認可保育園/*ninka hoiku-en*
3. Convert plural forms to singular forms (only for English).
Mother-Child Health Handbooks
⇒ Mother-Child Health Handbook
4. Convert upper cases in headings or heads of sentences to lower cases (only for English).
Seal Registration
⇒ seal registration

6.1.2.2 *Platform*

In order to facilitate the manual extraction of terms, we developed a web-based platform to collect terms from parallel sentences.

Figure 6.1 depicts the interface in which a pair of paralleled sentences (called an *unit*) is presented. This system enables us to capture the span of a term by clicking the starting word and the ending word.³ At the bottom of the screen, terms that have been previously registered are also shown. Registration of pairs of bilingual terms identical to the ones already identified is not allowed. These mechanisms support human decision-making and prevent duplicate registration, leading to improved efficiency of extraction.

Another important feature of this platform is that it is designed to facilitate collaborative term extraction and validation. As soon as a user adds a comment to each unit and/or to each term, other users can refer to the comments and a task manager can promptly respond to the comment if necessary. The status of work progress as well as extracted terms can be checked online at any time, which helps conduct the task smoothly.

Back

Previous Unit

Next Unit

Comment for this unit

Save

Unit. SC_0214_ja_en

登録完了後、印鑑登録証（カード）を交付します。

When registration has been completed , you will be issued a personal seal registration card .

印鑑登録証

personal

Clear

Register

Comment for this term

ID	Japanese	English	Comment
	【別ユニットの用語】		
✓4004	ハンコ	seal	一般語？
✓2377	印鑑	seal	
✓2375	印鑑登録証	personal seal registration certificate	
✓2372	印鑑登録	personal seal registration system	
✓2367	印鑑	personal seal	
✓2274	捺印	seal	一般語？
✓1916	認印	personal seal	
✓1855	印鑑登録	Seal Registration	
✓1854	印鑑登録証	Certificate of Seal Registration	
✓1852	印鑑登録	Personal Seal Registration	
✓1434	印鑑登録	personal seal registration	
✓1119	認印	a personal seal	

Figure 6.1 Term registration platform

6.1.2.3 Term extraction and validation

Ideally, the term extraction task should be conducted by experts of the municipal domain, but as Yasui (2009) pointed out, there is a shortage of skilled municipal writers and it is unrealistic to hire such experts. Compared to the legal or medical domains, however, the municipal domain, especially life information, is more familiar to the general public. We thus employed four university students and asked them to manually extract bilingual term candidates using the term registration platform. They were native speakers of Japanese and have a reasonably good command of English, sufficient to correctly identify the translated terms.

The identification of terms is difficult even for experts (Frantzi et al., 2000). To alleviate the individual differences of term identification and ensure comprehensibility, we instructed the students to capture the terms as widely as possible. We also frequently checked their extracted terms and comments, and, if necessary, gave them detailed directions for the task. Finally, we validated all the terms they extracted to improve the accuracy of the terms.

6.1.2.4 Results

A total of 3741 bilingual term pairs were collected from 15391 aligned sentence pairs. To present and analyse the extracted terms quantitatively, hereafter we use the following symbols based on Baayen (2001):

Table 6.2 The number of terms extracted and their occurrences in the corpus

	$V(N)$	N	$N/V(N)$
Japanese	3012	15313	5.084
English	3465	15708	4.533

Figure 6.2 The frequency spectrum of terms in municipal corpus (m : frequency class; $V(m, N)$: number of types with frequency m)

- $V(N)$: number of distinct terms (number of types).
- N : number of term occurrences in the corpus (number of tokens).
- m : index for frequency class (m is an integer value).
- $V(m, N)$: number of types that occur m times in the corpus.

Table 6.2 gives the statistics of terms extracted from the corpus. Figure 6.2 shows the frequency spectrum, or grouped frequency distribution, of terms in the municipal corpus for Japanese and English. The horizontal axis stands for the frequency class m , while the vertical axis stands for $V(m, N)$. Both axes take logarithmic scales.

Although according to Table 6.2 the mean frequencies $N/V(N)$ are 4.5 for Japanese and 5.1 for English, Figure 6.2 clearly indicates that most of the terms rarely appear, while a small number of terms appear very frequently. Indeed, the number of terms observed only once, i.e. $V(1, N)$, is 1227 for Japanese and 1126 for English (depicted as the circles in the very top left of Figure 6.2).

Importantly, we can see that the distributions form roughly straight lines from the top left to the bottom right, which means that Zipf’s second law approximately holds (Zipf, 1935, 1949; Kageura, 2012). Zipf’s second law is as follows:

$$V(m) = \frac{a}{m^b}, \tag{6.1}$$

where a and b are parameters. Though we will not delve into the details of Zipf's law, also known as the power law, we note that it holds for a variety of phenomena, including not only natural language utterances (Baroni, 2009; Moreno-Sánchez et al., 2016) but also scientific productivity (Lotka, 1926), city population size (Jiang et al., 2015) and website visits (Adamic, 2002).

According to Table 6.3, which shows the 20 most frequent terms, the most frequent terms are observed 369 times for Japanese and 245 times for English. We also notice that rather general terms are listed in the high places, such as ‘問い合わせ/*toiawase*’ (contact) and ‘外国人/*gaikokujin*’ (foreign resident). It is also notable that two city names, ‘新宿区’ (Shinjuku-ku/Shinjuku City) and ‘浜松市’ (Hamamatsu), hold high places in the list, which reflects the fact that the source corpus was constructed by extracting texts from websites of Shinjuku City and Hamamatsu City.

Finally, we look at the issue of term variation, which in principle should be avoided in terms of consistent authoring and translation. Table 6.4 lists some examples of extracted bilingual term pairs, all of which refer to the same concept. While there are four term types for Japanese—‘健康診査/*kenko shinsa*’, ‘健康診断/*kenko shindan*’, ‘検査/*kensa*’ and ‘健診/*ken-shin*’, there are seven for English—‘health medical examination’, ‘health check-up’, ‘medical check-up’,

Table 6.3 The 20 most frequent terms in the corpus (before controlling)

Rank	Japanese	#token	English	#token
1	新宿区/ <i>shinjuku-ku</i>	369	Shinjuku-ku	245
2	区/ <i>ku</i>	249	Shinjuku City	179
3	問い合わせ/ <i>toiawase</i>	237	procedure	177
4	手続/ <i>tetsuzuki</i>	191	contact	169
5	外国人/ <i>gaikokujin</i>	189	application	162
6	在留資格/ <i>zairyu-shikaku</i>	168	city	160
7	申請/ <i>shinsei</i>	157	inquiry	147
8	浜松市/ <i>hamamatsu-shi</i>	128	fee	147
9	市区町村の役所/ <i>shi-ku-cho-son no yakusho</i>	120	status of residence	131
10	届出/ <i>todokede</i>	110	junior high school	114
11	区役所/ <i>kuyakusho</i>	101	notification	113
12	保険料/ <i>hokenryo</i>	99	foreign resident	109
13	問合せ先/ <i>tooiawase-saki</i>	87	income	104
14	日本人/ <i>nihonjin</i>	86	parent	101
15	無料/ <i>muryo</i>	85	Hamamatsu	95
16	窓口/ <i>madoguchi</i>	83	municipality	95
17	所得/ <i>shotoku</i>	80	birth	93
18	在留カード/ <i>zairyu-kado</i>	80	foreigner	91
19	提出/ <i>teishutsu</i>	80	payment	90
20	在留期間/ <i>zairyu-kan</i>	77	foreign	87

Table 6.4 Example of extracted bilingual term pairs

No	Japanese	English
1	健康診査/ <i>kenko shinsa</i>	health medical examination
2	健康診査/ <i>kenko shinsa</i>	health check-up
3	健康診査/ <i>kenko shinsa</i>	medical check-up
4	健康診査/ <i>kenko shinsa</i>	health checkup
5	健康診査/ <i>kenko shinsa</i>	check-up
6	健康診断/ <i>kenko shindan</i>	health check
7	検査/ <i>kensa</i>	check-up
8	健診/ <i>ken-shin</i>	physical check-up
9	健診/ <i>ken-shin</i>	health checkup
10	健診/ <i>ken-shin</i>	check-up

‘health checkup’ ‘check-up’, ‘health check’ and ‘physical check-up’. It is required to control these variations.

Also notable is the different status of Japanese and English terminologies. Table 6.2 shows that the number of distinct Japanese terms is 3012, while that of English terms is 3465, suggesting that in general the translated English terms are more varied than the Japanese source terms. This can be explained by the general tendency of greater inconsistencies in the translated terms, i.e. ‘terminology inconsistencies often increase in frequency in the translated version compared to the original, due to the fact that there can be several ways to translate a given term or expressions’ (Warburton, 2015b, p.649). It is also pointed out that an important factor leading to the terminology inconsistencies is as follows: ‘When a document or a collection of documents is divided into smaller parts which are translated by several translators, terminology in target language will be more inconsistent than when only one translator is involved’. We can reasonably assume that several translators took charge of translating the municipal texts (terms) we deal with here, as the organisations in charge (CLAIR, Shinjuku City and Hamamatsu City) are different. Besides, the unavailability of bilingual municipal terminologies they can consult might aggravate the problem of terminology inconsistencies.

The next task is to control the terminologies. We first comprehensively collect variations in the terminologies and create a typology for both Japanese and English term variations (Section 6.1.3). We then control the variations based on the typology and other evidence (Section 6.1.4).

6.1.3 *Typology of term variation*

Creating a typology as such is not the final goal of this study, but it helps us consider how to control the term variations and how to cope with *possible* variations we might encounter. Based on the typologies previously established for Japanese (Yoshikane et al., 2003) and for English (Daille et al., 1996; Jacquemin,

2001), we formulated our typologies for the Japanese and English term variations in the municipal corpus.

The range of the terminology variations is dependent on the foreseen applications (Daille, 2005). In this study, from the point of view of controlled authoring and (machine) translation, we cover a wide range of variations, including not only morphological and syntactic variations, but also synonym and orthographic variants.

Investigating all the term pairs extracted from the corpus, we identified 374 Japanese term variations (12.4% of 3012 Japanese term types) and 1258 English term variations (36.3% of 3465 English term types). In a bottom-up manner, we created a typology that fits for both languages. It consists of five basic categories: (A) syntax/morphology, (B) synonym, (C) orthography, (D) abbreviation and (E) translation. Although the (E) translation category pertains only to English and some of the subcategories in the typology pertain only to either Japanese or English,⁴ we believe that it would be more convenient to show the term variations of both languages together than separately.

(A) Syntax/morphology

1. Insertion/omission

- ‘健康保険証/*kenko-hoken-sho*’ — ‘保険証/*hoken-sho*’
(health insurance card — insurance card)
- **personal** seal registration — seal registration

2. Permutation

- ‘市町村区/*shi-cho-son-ku*’ — ‘市区町村/*shi-ku-cho-son*’
(city-town-village-ward — city-ward-town-village)
- **entry** procedure — procedure **for entering**
- total period **insured** — total **insured** period
- **two-wheeled** compact motor **vehicle** — small **two-wheeled vehicle**

3. Morpho-syntactic

- ‘必要書類/*hitsuyo-shorui*’ — ‘必要な書類/*hitsuyo-na-shorui*’
(necessity document — necessary document)
- **residence** status — **residential** status
- resident tax — resident’s tax
- copy of **the** resident certificate — copy of **your** resident certificate

4. Phrasal change

- ‘就労の在留資格/*shuro no zairyu-shikaku*’ — ‘就労が可能な在留資格/*shuro ga kano-na zairyu-shikaku*’
(working resident status — resident status which permits work)
- ‘可燃ゴミ/*kanen gomi*’ — ‘燃やせるゴミ/*moyaseru gomi*’
(combustible garbage — garbage to be burned)

110 *Controlled document authoring*

- **working** resident status — status of residence **which permits work**
 - **holders** of multiple nationality — **those who hold** multiple nationality
5. Coordination
- ‘老齡・障害・遺族基礎年金/*rorei · shogai · izoku nenkin*’ — (1) ‘老齡年金/*rorei nenkin*’; (2) ‘障害年金/*shogai nenkin*’; (3) ‘遺族基礎年金/*izoku kiso nenkin*’
(old-age pension; disabled pension; survivor’s basic pension)
 - combustible and non-combustible garbage — (1) combustible garbage; (2) non-combustible garbage
6. Combination with parenthesis
- ‘認可保育所（園）/*ninka hoiku-sho (-en)*’ — (1) ‘認可保育所/*ninka hoiku-sho*’; (2) ‘認可保育園/*ninka hoiku-en*’
(accredited nursery school)
 - mother’s (parent’s) classes — (1) mother’s classes; (2) parent’s classes
7. Composition/decomposition
- ‘外国人住民基本台帳制度/*gaikokujin-jumin-kihon-daicho-seido*’ — ‘外国人住民の基本台帳制度/*gaikokujin-jumin no kihon-daicho-seido*’
(Foreign Resident Resident Registration System — Resident Registration System for foreign resident)
8. Preposition change
- change **in** address — change **of** address
 - noise **from** everyday living — noise **created by** everyday living
9. Ellipsis
- mid-**term** to long-term resident — mid to long-term resident

(B) *Synonym*

1. Whole term
- ‘印鑑/*inkan*’ — ‘認印/*mitomein*’
(personal seal)
 - ophthalmology — eye clinic
2. Substitution
- ‘健康診査/*kenko-shinsa*’ — ‘健康診断/*kenko-shindan*’
(health checkup)
 - health insurance **certificate** — health insurance **card**
3. Explanation
- ‘日本人/*nihonjin*’ — ‘日本国籍の人/*nihon-kokuseki no hito*’
(Japanese — those who are Japanese national)
 - miscellaneous school — school in the miscellaneous category

(C) Orthography

1. Parenthesis
 - Technical Trading Intern Type 1 — Technical Intern Training (Type 1)
2. Emphasis
 - “key” money — key money
3. Declensional Kana ending (only Japanese)
 - ‘各種手当/*kakushu teate*’ — ‘各種手当て/*kakushu teate*’
(various allowance)
4. Middle dot (only Japanese)
 - ‘阪神・淡路大震災/*Hanshin · Awaji daishinsai*’ — ‘阪神淡路大震災/*Hanshin Awaji daishinsai*’
(Hanshin-Awaji Earthquake)
5. Kana/Kanji character (only Japanese)
 - ‘もえないごみ/*moenai gomi*’ — ‘燃えないゴミ/*moenai gomi*’
(non-combustible garbage)
6. And, hyphen, slash (only English)
 - **Health and Disease** Prevention Division — **Health & Disease** Prevention Division
 - **intracompany** transferee — **intra-company** transferee
 - **Respect for the Aged** Day — **Respect-for-the-Aged** Day
 - **pregnancy induced** hypertension — **pregnancy-induced** hypertension
 - **Health and Sports** Day — **Health-Sports** Day
 - health check-up for **infant and child** — regular **infant/child** health check-up
7. Space (only English)
 - childbirth — child birth
8. Plural/singular form (only English)
 - Mother-Child Health Handbooks — Mother-Child Health Handbook
9. Upper/lower case (only English)
 - Seal **R**egistration — seal **r**egistration

(D) Abbreviation

1. Initialism/acronym
 - ‘健康診査/*kenko-shinsa*’ — ‘健診/*ken-shin*’
(health checkup)

- Japan Automobile Federation — **JAF**
2. Clipping
- ‘独立行政法人/*dokuritsu gyose hojin*’ — ‘(独)/(*doku*)’
(Incorporated Administrative Agency)

(E) Translation

1. Transliteration/translation (only English)
- Shinjuku-**ku** — Shinjuku **City**
2. Literal/interpretative translation (only English)
- motor-assisted bicycle — motorcycle (50 cc or less)

Note that at the term extraction phase, we normalised the variations of the following categories: **(A)-5 coordination**, **(A)-6 combination with parenthesis**, **(C)-8 plural/singular form** and **(C)-9 upper/lower case** (see Section 6.1.2).

6.1.4 Terminology control

What we need to do next is to (1) define preferred terms and proscribed terms in Japanese, and (2) assign authorised terms in English. We take into account the three criteria in Table 6.5 to examine the variant terms.

Let us return to the example of Table 6.4. We check the four Japanese terms and the seven English terms in light of the three criteria above. Table 6.6 summarises how each term meets each of the criteria. From this, we can define, for instance,

Table 6.5 Criteria for defining preferred and proscribed terms

Dictionary evidence:	If a term is registered as an entry form in dictionaries,* we regard it as preferable.
Frequency evidence:	Higher frequency in the corpus is preferable.
Typological preference:	The following variations identified in the typology are not preferable. (A)-1 omitting necessary information/ inserting unnecessary information (A)-3 possessive case/personal pronoun (C)-2 unnecessary use of emphasis (C)-5 unnecessary use of Kana characters (C)-6 unnecessary use of hyphens (D)-1 initialisms/acronyms (D)-2 clipping (E)-1 unnecessary use of transliteration

*In this study, we consulted Grand Concise Japanese–English Dictionary.

Table 6.6 Examination of term variations

	<i>Term</i>	<i>Dic.</i>	<i>Freq.</i>	<i>Typology</i>
1	健康診査/ <i>kenko shinsa</i>	✓	30	
2	健康診断/ <i>kenko shindan</i>	✓	5	
3	検査/ <i>kensa</i>	✓	51	(A)-1: omission
4	健診/ <i>ken-shin</i>	✓	12	(D)-1: initialism
1	health medical examination		1	
2	health check-up	✓	17	(C)-6: hyphen
3	medical check-up	✓	3	(C)-6: hyphen
4	health checkup	✓	37	
5	check-up		14	(A)-1: omission (C)-6: hyphen
6	health check		1	
7	physical check-up	✓	2	(C)-6: hyphen

Table 6.7 The basic statistics of controlled terminology

	<i>(a) Uncontrolled types</i>	<i>(b) Controlled types</i>	<i>b/a</i>	<i>Tokens</i>
Japanese	3012	2802	93.0%	15313
English	3465	2740	79.1%	15708

‘健康診査/*kenko shinsa*’ as a preferred term since it is registered in the dictionary and also observed frequently (30 times) in the corpus, while the other two can be defined as proscribed terms. For the English term, we can choose ‘health checkup’ as a standard translation. Though ‘check-up’ (with a hyphen) is also frequently used in the corpus, we prefer ‘checkup’ (without a hyphen), based on the typological preference policy we adopted above.

Table 6.7 gives the basic statistics of our terminologies, showing the reduced number of term types after the variations were controlled. It can be noted that the number of English term types was reduced by about 20%, and the number of controlled term types in Japanese and in English became closer. This is not surprising because it is reasonable to assume that Japanese terminology and English terminology should contain the same number of *concepts* (or *referents*) in the parallel corpus. Controlling the variant forms of terms can be regarded as assigning one (authorised) linguistic form to one concept. We can estimate that the number of municipal concepts in our corpus is around 2700–2800.

Finally, we look at the revised list of 20 most frequent terms, highlighting the terms that obtained a higher frequency after the variations were controlled (Table 6.8). We can see, for instance, that the variant form ‘Shinjuku-ku’ (245 tokens) was merged into the preferred term ‘Shinjuku City’ (179 tokens), resulting 424 tokens of the term.

Table 6.8 The 20 most frequent controlled terms occurred in the corpus

Rank	Japanese	# token	English	# token
1	新宿区/ <i>shinjuku-ku</i>	369	Shinjuku City	424
2	区/ <i>ku</i>	249	foreign resident	203
3	問い合わせ/ <i>toiawase</i>	237	municipal administrative office	180
4	外国人/ <i>gaikokujin</i>	202	procedure	177
5	手続/ <i>tetsuzuki</i>	196	contact	169
6	市区町村の役所/ <i>shi-ku-cho-son no yakusho</i>	194	status of residence	168
7	申請/ <i>shinsei</i>	178	application	162
8	在留資格/ <i>zairyu-shikaku</i>	168	handling fee	161
9	浜松市/ <i>hamamatsu-shi</i>	128	city	160
10	届出/ <i>todokede</i>	110	health checkup	159
11	区役所/ <i>kuyakusho</i>	101	notification	148
12	保険料/ <i>hokenryo</i>	99	inquiry	147
13	健康診査/ <i>kenko shinsa</i>	98	childbirth	134
14	外国人登録証明書/ <i>gaikokujin-toroku-shomeisho</i>	89	parent	128
15	日本人/ <i>nihonjin</i>	88	foreign national	125
16	市区町村/ <i>shi-ku-cho-son</i>	87	junior high school	114
17	問合せ先/ <i>toiawase-saki</i>	87	free	109
18	無料/ <i>muryo</i>	85	income	108
19	印鑑/ <i>inkan</i>	84	various allowance	104
20	窓口/ <i>madoguchi</i>	83	personal seal	102

We are now in the position to address the question: How do we evaluate the terminology and the controlled terminology we constructed? In the following sections, we propose a way to quantitatively evaluate the coverage of terminology and the quality of variation control, and evaluate our terminologies.

6.2 Coverage estimation

As we mentioned, there are no comprehensive terminologies maintained in the Japanese municipal domain for comparison. We thus estimate the coverage of our terminology without using external terminologies.

6.2.1 Methods

6.2.1.1 Self-referring coverage estimation

To estimate the coverage of the terminologies using only constructed ones, we employ the self-referring quantitative evaluation method proposed by Kageura and Kikui (2006). The basic idea is to (1) extrapolate the size of N to infinity using

the observed data and estimate the saturation point as the ‘model’, and (2) evaluate the current status of the $V(N)$ in comparison with the ‘model’ point.

While Kageura and Kikui (2006) estimated the coverage of the lexical items of a Japanese travel expression corpus, specifically focusing on the content words (nouns, verbs and adjectives), we assume this method can be applied to our task, that is, the estimation of the coverage of the terms (mostly noun compounds) that appeared in our municipal corpus. They also emphasised that this method presupposes that the corpus qualitatively represents the whole range of relevant language phenomena in the given domain. Though the size of our municipal corpus itself is not large, it is possible to apply the method to our case, as the corpus focuses on a narrow domain (municipal-life information) and covers a wide and well-balanced range of linguistic phenomena.

6.2.1.2 Conditions for evaluation

We compare two parameters: (i) controlled and uncontrolled and (ii) Japanese and English. Thus, four conditions of terminology were prepared: (1) uncontrolled Japanese terminology, (2) uncontrolled English terminology, (3) controlled Japanese terminology and (4) controlled English terminology. While much previous work has been devoted to analysing monolingual lexical data, we attempt to compare the characteristics of terminologies with different languages.

To estimate the coverage of *terms*, we investigate the uncontrolled conditions. Our previous observation showed that uncontrolled English terminology is more varied than uncontrolled Japanese terminology, which may affect the population size of the terminologies. On the other hand, investigating the controlled conditions is important to see the coverage of *concepts* in the domain.

From the point of view of validating how well our terminologies are controlled, we explore the controlled conditions of the terminologies. Our hypothesis is that if the terminologies are well controlled, the estimated population number of Japanese and English terms become closer, as both represent the same set of concepts.

6.2.1.3 Expected number of terms

A number of methods have been proposed to estimate the population item size (Efron and Thisted, 1976; Tuldava, 1995; Baayen, 2001). Here we adopt large-number-of-rare-event (LNRE) modelling, which has been used in the field of lexical statistics (Khmaladze, 1987; Baayen, 2001; Kageura, 2012). We outline the computational steps behind the method, following Baayen (2001).

Let the population number of types be S and let each type be denoted by w_i ($i = 1, 2, \dots, S$). With each w_i , a population probability p_i ($i = 1, 2, \dots, S$) is associated. The probability that w_i occurs m times in a sample of N is calculated using the binomial theorem:

$$Pr(f(i, N) = m) = \binom{N}{m} p_i^m (1 - p_i)^{N-m}, \quad (6.2)$$

where $f(i, N) = m$ indicates the frequency of w_i in a sample of N tokens.

The expected number of types that occur m times in a sample of N is, therefore, given as follows:

$$\begin{aligned}
 E[V(m, N)] &= \sum_{i=1}^S \Pr(f(i, N) = m) \\
 &= \sum_{i=1}^S \binom{N}{m} p_i^m (1 - p_i)^{N-m} \\
 &= \sum_{i=1}^S \frac{(Np_i)^m}{m!} e^{-Np_i}.
 \end{aligned} \tag{6.3}$$

For large N and small p , a binomial distribution with parameters (N, p) is well approximated by Poisson distribution with parameter $\lambda = np$:

$$\binom{N}{m} p_i^m (1 - p_i)^{N-m} \approx \frac{(Np_i)^m}{m!} e^{-Np_i}. \tag{6.4}$$

At the final step of (6.3), the Poisson approximation is applied.

On the other hand, in order to express $E[V(N)]$, the expected number of types, we focus on the types that do not occur. The probability that type w_i does not occur in the sample N tokens is $\Pr(f(i, N) = 0)$. Taking the complement of this probability, i.e. $(1 - \Pr(f(i, N) = 0))$, we obtain the probability that w_i occurs at least once in the sample N . Hence, the $E[V(N)]$ is given as follows:

$$\begin{aligned}
 E[V(N)] &= \sum_{i=1}^S (1 - \Pr(f(i, N) = 0)) \\
 &= \sum_{i=1}^S (1 - \binom{N}{0} p_i^0 (1 - p_i)^{N-0}) \\
 &= \sum_{i=1}^S (1 - \frac{(Np_i)^0}{0!} e^{-Np_i}) \\
 &= \sum_{i=1}^S (1 - e^{-Np_i}).
 \end{aligned} \tag{6.5}$$

Note that Poisson approximation (6.4) is used again in the middle step of (6.5).

For mathematical convenience, we rewrite the Poisson models in integral forms using the structural type distribution $G(p)$, the cumulative number of types with probabilities equal to or greater than p , which is defined as follows:

$$G(p) = \sum_{i=1}^S I_{[p_i \geq p]}, \tag{6.6}$$

where $I = 1$ when $p_i \geq p$, and 0 otherwise. We can renumber the subscript of p for $p_j > 0$, such that $p_j < p_{j+1}$ ($j = 1, 2, \dots, \kappa$). As $G(p)$ is a step function, jumps at the probabilities p_j , in other words, the number of types in the population with probabilities p_j , are given by:

$$\Delta G(p_j) = G(p_j) - G(p_{j+1}) \quad (6.7)$$

We can now restate the equations (6.3) and (6.5):

$$\begin{aligned} E[V(m, N)] &= \sum_{j=1}^{\kappa} \frac{(Np_j)^m}{m!} e^{-Np_j} \Delta G(p_j) \\ &= \int_0^{\infty} \frac{(Np)^m}{m!} e^{-Np} dG(p) \end{aligned} \quad (6.8)$$

$$\begin{aligned} E[V(N)] &= \sum_{j=1}^{\kappa} (1 - e^{-Np_j}) \Delta G(p_j) \\ &= \int_0^{\infty} (1 - e^{-Np}) dG(p) \end{aligned} \quad (6.9)$$

Using some hypotheses about the form of distributions such as inverse Gauss-Poisson distribution, we can obtain models to extrapolate the $V(N)$ and $V(m, N)$ for $N \rightarrow \infty$.

6.2.1.4 Growth rate of lexical items

The constructed model also gives us insight into the growth rate, or how fast the number of types increases as we extract more terms from texts in the domain. The growth rate is obtained by taking the derivative of $E[V(N)]$ as follows:

$$\begin{aligned} \frac{d}{dN} E[V(N)] &= \frac{d}{dN} \int_0^{\infty} (1 - e^{-Np}) dG(p) \\ &= \int_0^{\infty} (-p \cdot -e^{-Np}) dG(p) \\ &= \frac{1}{N} \int_0^{\infty} Np e^{-Np} dG(p) \\ &= \frac{E[V(1, N)]}{N} \end{aligned} \quad (6.10)$$

6.2.2 Results and discussions

6.2.2.1 Population types and present status of terminologies

Table 6.9 gives the estimated population number of term types $E[S]$, together with the coverage ratio $CR (= V(N)/E[S])$.

Table 6.9 Population types $E[S]$ and coverage CR

		<i>Model</i>	$E[S]$	$V(N)$	$CR(\%)$	χ^2	p
Uncontrolled	Ja	GIGP	5505.3	2953	53.6	35.260	0.0008
		fZM	4626.2	2953	63.8	33.930	0.0012
	En	GIGP	7616.4	3255	42.7	23.857	0.0325
		fZM	6083.0	3255	53.5	28.197	0.0085
Controlled	Ja	GIGP	5111.9	2753	53.9	34.620	0.0010
		fZM	4299.0	2753	64.0	27.905	0.0093
	En	GIGP	5380.2	2611	48.5	35.354	0.0007
		fZM	4444.5	2611	58.7	36.525	0.0005

Though there are several models of LNRE, we chose the following two models, which were shown to be effective in the task of vocabulary size estimation: the generalised inverse Gauss-Poisson (GIGP) model (Sichel, 1975) and the finite Zipf-Mandelbrot (fZM) model (Evert, 2004; Evert and Baroni, 2005).⁵

The lower χ^2 -value and higher p -value indicate a better fit of the LNRE model, and Baayen (2008, p.233) remarks that a p -value above 0.05 is preferable. Though all of the p -values are below 0.05, the χ^2 -values are not bad compared to the related work by, for example, Kageura and Kikui (2006) or Baayen (2001).

According to Table 6.9, the estimated population size $E[S]$ ranges from 4299 to 7616, and the coverage ratio CR ranges from 42.7% to 64.0%. Though the values of $E[S]$ and CR depend on the models used,⁶ we can observe several important points of the result.

Firstly, focusing on the uncontrolled terminologies, we recognise very different results between Japanese and English: the population number of types of Japanese, 5505 (GIGP) and 4626 (fZM), is much smaller than that of English, 7616 (GIGP) and 6083 (fZM). Consequently, the coverage ratio of Japanese is generally higher than that of English. This may reflect the higher diversity of the uncontrolled English terminology. As we have seen in Section 6.1.3, the ratio of variations in the English terminology is much higher than that in the Japanese terminology, which suggests the potential diversity of English terminology in the population.

Secondly, controlled terminologies tend to exhibit a lower $E[S]$ and higher CR than the uncontrolled terminologies. For example, the CR of controlled terminology when fZM is adopted is 64.0% for Japanese and 58.7% for English, which means that around two thirds of the concepts in the domain are included in our terminologies. It is worth noting that the coverage of the controlled terminologies exceeds that of the uncontrolled ones. Though further development of the terminologies is necessary, this result is fairly good as a starting point and encourages practical use of the terminologies.

Finally, related to the second point, the differences of $E[S]$ and CR values between Japanese and English in controlled conditions are much smaller than

those in the uncontrolled conditions. In principle, the (population) size of the concepts in the parallel data of a given domain should be the same across the languages. The closer values of $E[S]$ between Japanese and English demonstrate that our constructed terminologies have a desirable nature.

6.2.2.2 Terminology growth

We have analysed the present status of the terminologies vis-à-vis the extrapolated population size of the terminologies when N is infinity. From the practical point of view, however, it is impossible to observe the infinite size of N within the limited textual data that is available. Our next question is to what extent we can enlarge the size of the terminologies and extend their coverage within the realistic range. To address this question, here we take a closer look at the dynamic trends of terminology growth.

We first observe how the expected number of term types $V(N)$ shifts as the number of term tokens N increases. Figure 6.3 draws for each LNRE model the growth curves of $V(N)$, the estimated number of term types, as N grows to 100000, which is approximately 6.5 times as large as the present N .⁷ The vertical dotted line indicates $N = 15000$, which is close to the present N .

Comparing the growth curves of the four conditions, we can easily recognise the general tendencies that conform to what we pointed out in the previous section. We summarise them as follows:

1. English uncontrolled terminology (thin solid line) grows more rapidly than Japanese one (thick solid line).

Figure 6.3 Growth curve of terminologies in municipal corpus

Table 6.10 Growth rate

		$V(1,N)$	<i>Growth rate</i>
Uncontrolled	Ja	1227	0.080
	En	1557	0.099
Controlled	Ja	1126	0.074
	En	1138	0.072

Table 6.11 Shift in the coverage ratio (%)

		<i>Model</i>	0.5 <i>N</i>	<i>N</i>	1.5 <i>N</i>	2 <i>N</i>	2.5 <i>N</i>	3 <i>N</i>
Uncontrolled	Ja	GIGP	39.0	53.9	63.2	69.8	74.6	78.4
		fZM	45.9	63.8	74.5	81.5	86.4	89.8
	En	GIGP	29.9	42.9	51.6	58.1	63.1	67.2
		fZM	37.2	53.5	64.2	71.8	77.4	81.8
controlled	Ja	GIGP	39.4	54.2	63.4	69.9	74.7	78.4
		fZM	46.3	64.0	74.6	81.5	86.3	89.8
	En	GIGP	35.0	48.9	57.8	64.3	69.2	73.1
		fZM	41.9	58.7	69.3	76.5	81.8	85.7

2. Controlled terminologies (dotted line) show more moderate growth than uncontrolled ones (solid lines).
3. The growth curves of controlled Japanese (thick dotted line) and English (thin dotted line) align very closely.

The growth curves also enable us to visually grasp the shift of the growth rate. We can observe that all of the curves grow rapidly in the beginning and become gentler when N reaches around 30000, about twice the size of the present N . Although within the size of 100000, all the growth curves do not seem flattened out, we can gain insight into how to effectively extend the size of the terminologies.

Table 6.10 shows the growth rates at the current size of terminologies. For uncontrolled conditions, to obtain a new term, 12 ($\approx 1/0.08$) terms should be added to the Japanese terminology, and 10 ($= 1/0.10$) terms should be added to the English terminology. For controlled conditions, to obtain a new concept, 14 ($\approx 1/0.07$) terms should be added to the Japanese or English terminologies.

These results above motivate us to further enlarge the size of our terminologies. Considering the difficulty in compiling a bilingual (or multilingual) parallel municipal corpus on a large scale, we further restrict ourselves to a realistic size of N . More specifically, we examine a coverage ratio up to three times as large as the original size N , as we can expect a rapid growth of $V(N)$ until N reaches that point.

Table 6.11 shows the shift in the estimated coverage ratio at $0.5N$ intervals up to $3N$ (about 450000 tokens). The column N stands for the current size of tokens, i.e. about 150000.

If we double the token size N , we achieve approximately 80% coverage of Japanese terms, 70% coverage of English terms and 80% coverage of concepts in the domain (when estimating by fZM), showing an increase of more than 15% compared to the original size N . If we treble N , we achieve an additional increase of at most 10% in the coverage ratio, with some of the values reaching nearly 90%. Therefore, it may be a promising idea to continue to extract more terms, to enlarge the terminologies until substantial coverage, say, 80%, is attained.

6.3 Summary

To answer the research question **RQ-F4** (Can municipal terms be comprehensively captured and well controlled?), in this chapter, we constructed bilingual municipal terminologies from scratch and evaluated the current coverage of the terminologies. The results showed that our terminologies cover (1) about 55–65% (Japanese) and 45–60% (English) of the terms, and (2) about 55–65% (Japanese) and 50–60% (English) of the concepts in the municipal domain. Though further extension of the terminologies is needed, their coverage seems good enough to be practically applied. Also, the closer values of population number of term types and the similar shapes of the terminology growth curves for Japanese and English demonstrated that our terminologies are well controlled.

A list of Japanese controlled terms and variant terms we formulated in this study plays a crucial role in controlled authoring. On the other hand, a list of controlled English terms that correspond to Japanese terms allows MT systems to produce correct translated terms consistently. The practical implementation of our managed terminologies will be presented in Chapter 7.

The important contributions of this study to the field of terminology study are summarised as follows:

- Empirical observation of the corpus endorsed the high proportion of terminology variation in actual texts. More specifically, 12.4% of Japanese term types and 36.3% of English term types are deemed variations. These results can be compared to other literature such as Daille (2005). Furthermore, the English translated terms tend to vary more than the Japanese source terms, which gives clear evidence for the claim by Warburton (2015b).
- To the best of our knowledge, this study is the first attempt to verify the coverage of bilingual terminologies using LNRE modelling. While previous research applied the method to monolingual data, focusing on words or morphemes (Baayen, 2001; Kageura and Kikui, 2006; Kageura, 2012), we analysed bilingual data, focusing on terms, and revealed different characteristics depending on the language.
- The typology of Japanese and English term variations can be used as a point of reference for future investigation of term variations. Whereas there are

several established monolingual typologies of variants, such as Jacquemin (2001) for English, Daille (2003) for French and Yoshikane et al. (2003) for Japanese, and a few bilingual typologies, Carl et al. (2004) for English and French, no typology has been proposed that deals with Japanese and English.

As future work, first of all, we will expand the size of our terminologies. Based on the estimation, to achieve about 80–90% coverage of municipal terms and concepts, we need to check 15000–30000 more term tokens. At this stage, automatic term recognition (ATE) techniques would be a viable option to efficiently collect term candidates. Another option would be crowdsourcing. The term candidate registration platform we developed is easy to use even for non-terminologists and is also accessible via web browsers, which makes it possible to recruit wider support for terminology work from the crowd.

Secondly, it will inevitably be required to examine and improve the validity of the control we made. This study took descriptive and prescriptive approaches to define controlled terms: we selected preferred terms from actual term occurrences based on three criteria—dictionary evidence, frequency evidence and typology preference. However, they are not necessarily the ‘best’ ones, as we can suppose that more reliable evidence can be attained from other sources, such as specialised legal and medical terminologies or official documentation provided by the government. Moreover, the most preferable terms may not even have been included in our limited corpus. Thus we plan to investigate other external materials to further validate our controlled terminologies.

Finally, related to the previous point, we will deal with *unseen but possible* term variations in order to enlarge the synsets of preferred terms and proscribed terms. For example, the term ‘medical checkup’ is not observed in our present corpus, but is assumed to be observed, considering the presence of the terms ‘medical check-up’ and ‘health checkup’. We intend to adopt a ‘generate and validate’ method (Sato et al., 2013), which makes use of constituents of terms to obtain new term candidates, and a ‘meta-rule’ method (Yoshikane et al., 2003; Jacquemin, 2001), which systematically produces the variant forms based on the handcrafted term transformation rules. The current terminologies and variation typology created in this study enable us to pursue both methods.

Notes

- 1 Although general (monolingual/bilingual) dictionaries register some municipal terminology, we found that the entry forms are not always consistent between the dictionaries and are sometimes different from the actual term occurrences.
- 2 Repeated use of sentences in a given domain gives us insight into the recycling of existing sentences when authoring. Though these longer expressions as such are beyond the scope of this chapter, we will discuss how to make use of existing sentences typically used in municipal procedures in Chapter 7.
- 3 In this figure, ‘personal’, the starting word of ‘personal seal registration card’, has been selected and ‘card’, the ending word of the term, is about to be clicked.

- 4 In particular, for the (C) orthography category, Japanese and English have significantly different systems of writing and characters, so the overlap between the two languages is relatively small.
- 5 Though we tried two other LNRE models, the lognormal model (Carroll, 1969) and the Yule-Simon model Simon (1960), the fit of the models to our data was not good compared to the GIGP and fZM models, so we did not adopt these models.
- 6 For all conditions, the fZM model produced higher values of *CR* than the GIGP model.
- 7 Note that the actual present *N* is 15313 for Japanese and 15708 for English as shown in Table 6.7.

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part III

MuTUAL: An authoring support system

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

7 System development

This chapter introduces an integrated controlled authoring support system, **MuTUAL**, which is designed to help non-professional municipal writers to create well-structured procedural documents that are both machine-translatable and human-readable. We first present the concept behind the system in Section 7.1. In relation to the previous chapters, we outline its overall modules in Section 7.2 and describe the implementation of each module in Section 7.3. Finally, we summarise the current status of the system development in Section 7.4.

7.1 Design principle

First of all, we emphasise that MuTUAL is an *authoring support* system, not a mail-merge-like generation system. The core mission of the system is to support human decision-making in each process of source-document authoring. At the stage of translating the source documents, we chiefly rely on MT systems to generate multilingual outputs since most municipalities cannot afford to hire professional translators.

MuTUAL supports controlled authoring process at both document-level and sentence-level. At the document-level, the system provides the DITA-based document template defined in Chapter 3 to help users formulate well-organised documents. While we are currently focusing on municipal procedural documents, this can be extended to other types of documents. At the sentence-level, the system offers a CL authoring assistant that helps writers to check the conformity to CL rules (Chapter 4) and controlled terminology (Chapter 6). Accordingly, users can create STs that are both machine-translatable and human-readable. The principal novel feature of the system is that it enables MT to generate outputs appropriate to their functional context within the target document. To do so, MuTUAL integrates context-dependent CL rules in combination with the pre-/post-translation processing described in Chapter 5.

The target users of our system are (monolingual) non-professional writers who are, in many cases, unaccustomed to the principle of controlled authoring and authoring tools. While commercial XML editors such as FrameMaker and XMetaL provide a number of functions for professional use, it is quite difficult to learn these functions and customise them to fit municipal document production.

Furthermore, an XML editor itself does not offer a CL authoring assistant, let alone MT systems. These commercial tools are also generally too expensive to adopt for smaller municipalities. We thus conjecture that existing authoring tools are not suitable for our municipal document authoring and multilingualisation scenario. It is crucial to provide an integrated authoring support system that is easy to learn and is tailored to a specific purpose.

It should be noted that MuTUAL currently does not integrate a post-editing environment to further improve MT outputs, given that municipal writers are chiefly monolingual Japanese speakers. However, this does not mean that we will exclude post-editing in future study.

Considering the practical deployment of the system in Japanese municipalities, we would consider a web-based application to be more suitable than a desktop application. The system is implemented in HTML/CSS/JavaScript (client side), PHP (server side) and MySQL (database), and runs in a normal web browser without users needing to install and update it. We continue to use two MT systems, TransGateway and TexTra, since the former is the most widely adopted RBMT system in municipalities and the latter is a state-of-the-art SMT system that is freely available.¹ Other important reasons to adopt these two MT systems are that (1) the APIs (application programming interfaces) are available, which enables MuTUAL to seamlessly obtain the translation results, and that (2) user dictionaries are accepted, which is a prerequisite for implementing controlled terminology.

7.2 Modules

MuTUAL comprises modules for document structuring, controlled writing, and multilingualisation (see Figure 7.1).

Figure 7.1 Modules of MuTUAL

DOCUMENT STRUCTURING

- **Topic template** is the main user interface for municipal staff to author a particular topic from scratch or by reusing previously written topics from the topic database, seamlessly invoking other modules.
- **Topic database** archives the topics written in multiple languages.
- **Map organiser** helps authors to assemble multiple topics from the topic database and compose documents tailored for different purposes.
- **Document database** preserves in DITA format the documents that are the final products of the system.

CONTROLLED WRITING

- **CL authoring assistant** analyses each sentence in the text box and highlights any segment that violates CL rules or controlled terminology, along with diagnostic comments and suggestions for rewriting.
- **Similar text search** helps authors reuse existing sentences, phrases and terms provided by the topic database.

MULTILINGUALISATION

- **Pre-/post-translation processing** automatically modifies the ST/TT segments in accordance with the transformation rules defined in the functional elements.
- **MT** automatically translates and back-translates the pre-processed texts from the topic template.
- **Terminology database** registers multilingual parallel terms extracted from existing municipal websites that authors can or must refer to, and also provides resources for an MT dictionary and the terminology check function.

We first elaborate on the key modules: topic template, CL authoring assistant, MT and pre-translation processing, where our framework of controlled authoring described in Chapters 3–5 is chiefly materialised. We also explain the similar text search module, which is designed to enhance consistent and efficient authoring and translation.

7.3 Implementation

7.3.1 Topic template

The topic template is the core interface for authoring self-contained topics. Authors can easily formulate well-organised topics by filling in the DITA elements specified in Chapter 3. The template consists of three panes for document structuring, controlled writing, and rendering. Figure 7.2 shows the Task topic template for municipal procedures.

Figure 7.2 Task topic template

The left-hand pane (document structuring pane) provides the basic DITA Task topic structure. Authors can populate (or delete) the additional elements such as **Step 1**, **Step 2** and **Step 3** in **Steps**, and further specify the appropriate functional element shown in Table 3.3. Tool tips are also provided for guidance on DITA elements.

If we select one of the DITA elements, a simple text box to be filled in appears in the middle pane (controlled writing pane). The key mechanism for enhancing authoring and translation is to invoke different CL authoring assistants tuned to the respective functional elements of the document. For example, the CL authoring assistant for the **Steps** element implements the context-dependent CL rule ‘Use polite speech style with declarative form *shimasu* (します)’ while others do not.

It is also useful to display the entire text while authors are writing individual elements. The topic template seamlessly connects with our MT systems (see Section 7.3.3) and, in the right-hand pane (rendering pane), it displays both the Japanese source text and the English translated text in real time.

7.3.2 CL authoring assistant

The CL authoring assistant is the core module of MuTUAL. Our system is intended for writing source texts from scratch, which led us to design a real-time interactive system to continuously check the conformity to CL rules and controlled terminology. Whenever writers enter text that violates any of the working CL rules, the system detects it and helps them to make corrections.

Writing in accordance with CL rules and controlled terminology is not an easy task for authors. Unlike common spelling and grammar checkers, our system

detects language that is grammatically correct but, according to the pre-defined writing rules, should be avoided. Writers need to change elements of their usual writing styles that are allowed in non-technical writing. Moreover, some CL rules require linguistic knowledge that may be unfamiliar to even native speakers, such as ‘*sahen-noun*’ and ‘giving and receiving verbs’. Thus it is of particular importance to provide adequate descriptions of the rules and editing instructions. In Section 4.3, we formulated two sets of CL guidelines that consist of rule descriptions and rewrite examples. We implemented an interface that enables seamless consultation of the guidelines.

7.3.2.1 Use scenario

Our CL authoring assistant consists of four components for *detection*, *suggestion*, *ranking* and *correction* based on the Table 2.2 in Chapter 2. The use scenario is as follows (see Figure 7.3):

1. Users enter Japanese text in the **input box**, guided by the instructions about implemented CL rules.
2. The system automatically splits the input text by sentence, counts and shows the number of characters in each sentence, and displays warning messages

Figure 7.3 CL authoring assistant, MT and back translation (BT)

Figure 7.4 Terminology check function

- if the sentence exceeds the threshold levels of 30 characters ('long') and 50 characters ('too long').
3. The system automatically analyses each sentence and displays any detected **CL violations** in red and **proscribed terms** in blue (*detection*).
 4. Users modify the problematic segments based on the **diagnostic comments**, referring to **detailed rule descriptions**, if needed.
 5. For particular highlighted segments, the function offers alternative expressions, which are displayed by clicking the segments (*suggestion*). If there is more than one, suggestions are presented in the order of priority (*ranking*).
 6. If the author clicks a suggestion, the offending segment in the input box above is automatically replaced (*correction*).

Figure 7.3 depicts the system detecting a violation of the CL rule R18 (Avoid using verbose wording), which is one of the general CL rules we have implemented. The system detects rule violations in the initial draft '加入することになっています/*kanyu-suru koto ni natte imasu*' (is supposed to join) and suggests two candidates, '加入する必要があります/*kanyu-suru hitsuyo ga arimasu*' (need to join) and '加入してください/*kanyu-shite kudasai*' (please join). The author can easily choose the most appropriate suggestion to replace the segment in the input box and make, if necessary, further small revisions to ensure the naturalness of the source sentence. As soon as the text is modified, MT output is automatically generated. Figure 7.4 shows the term variation check function. If authors enter a term that has an authorised form in the database, the checker highlights the term in blue with suggestions displayed on mouse-over. In this figure, the system offers a preferred form '印鑑登録証明書/*inkan-toroku-shomeisho*' (seal registration certificate) instead of '印鑑証明/*inkan-shomei*'. As in the CL check function, authors can simply click one of the suggestions to replace the term in the text box.

In step 3, false-detection alerts generated by the system could annoy and even misguide writers. Hence, we allow users to select which rules to run through a modal window (Figure 7.5). Users can simply switch off the particular rule if they feel certain about checking the rule by themselves, or keep it active while being fully aware of possible mis-detections. The possibility of mis-detections will be examined through benchmark evaluation in Chapter 8.

In step 4, we prepare the interface to seamlessly consult the particular CL guidelines (see Section 4.3) by clicking the 'i' (information) button next

Figure 7.5 CL rule selection modal

to the **diagnostic comments**. Figure 7.6 shows an example of **detailed rule descriptions** of CL rule T10 (Do not use *reru/rareru* to express the potential mood or honorifics), which appears in the modal window.

7.3.2.2 Implementation of CL rules and terminology

We have implemented the 30 context-independent rules described in Chapter 4 and the four context-dependent rules described in Chapter 5. To implement the CL violation detection component, we created simple matching rules based on part-of-speech (POS) information, using the Japanese morphological analyser McCab.

Figure 7.6 CL guideline modal

Compared to the CL check function, it is much easier to implement the term variation search mechanism as it can be materialised by simple text string matching if we have a list of synsets of preferred terms and proscribed terms. We implemented a longest matching retrieval mechanism, which finds the longest proscribed term in the database that is contained in an ST string. Currently, the term database has registrations for the 210 Japanese proscribed terms we identified in Section 6.1.4.

7.3.3 *MT and dictionary*

MuTUAL currently adopts two MT systems, TransGateway and TexTra, through APIs. We also registered in the user dictionaries the bilingual controlled terminology of the municipal domain constructed in Chapter 6. The statistical model of the TexTra system can be retrained by in-house parallel corpora. If we collect a sufficient amount of controlled source texts and their translated texts as training data, further improvement of MT quality can be expected, though such data is yet to be collected.

In the practical context, each municipality has to choose one system to use.² The topic template connects with the system and returns outputs next to the Japanese input box (Figure 7.3). Authors can refer to the outputs in real time while writing source texts. Moreover, taking into consideration the fact that writers in Japanese municipalities do not necessarily have sufficient knowledge of the target language(s), the back translation (BT) is also provided right below the MT output for reference.

BT, also called round-trip translation, is the process of translating target texts that have already been translated back to their original language. While Somers (2005) concluded that BT is not suitable for gauging the quality of MT output, it has been demonstrated that MT quality can be improved by rewriting ST with reference to BT in Japanese–English, Japanese–Chinese, and Japanese–Korean translations (Miyabe et al., 2009). Language practitioners sometimes remark that BT may give them an indication of the target text quality.

Our system shows back-translated sentences with inconsistent morphemes between ST and BT highlighted in red, together with an agreement score. The agreement score is calculated as follows:³

$$\begin{aligned}
 ST\ score &= \frac{\# \text{ common morphemes}}{\# \text{ morphemes in ST}} \\
 BT\ score &= \frac{\# \text{ common morphemes}}{\# \text{ morphemes in BT}} \\
 \text{Agreement score} &= \frac{2 \times ST\ score \times BT\ score}{ST\ score + BT\ score}
 \end{aligned}$$

This gives authors hints as to which source segments degrade the MT output quality and how much of ST is assumed to be correctly translated.

As of now, we have implemented Japanese-to-English MT and English-to-Japanese BT. When we deal with additional target languages, such as Chinese, Korean, and Portuguese, it is more likely that BT will be referred to since it is unrealistic for any one writer to refer to outputs in multiple languages.

7.3.4 Automatic pre-/post-translation processing

To translate the Japanese texts, we employ MT systems coupled with pre-/post-translation processing. As mentioned in Section 5.3, we adopt pre-translation processing (and post-translation processing only for the SMT system) to achieve a higher MT performance without degrading the quality of the source, and have now implemented the two rules defined in Section 5.3.

This processing can be fully automated by defining simple transformation rules and using a Japanese morphological analyser such as MeCab, on the condition that the linguistic patterns of the source text are sufficiently controlled in conjunction with the functional elements. For instance, the CL_{st}-compliant segment ‘持参します/*jisan-shimasu*’ is decomposed into three morphemes, a verbal-noun ‘持

Figure 7.7 Similar text search

参/jisan’ (to bring), a verb ‘し/shi’ (do) and an auxiliary verb ‘ます/masu’ (style of polite speech), and is changed into ‘持参しろ/shiro’ through the transformation of the final morpheme.

7.3.5 Similar text search

Needless to say, if we can make use of previously written segments, it would be better to do so. As we observed in Section 6.1.1, the contents of municipal websites are highly repetitive. For example, we see repeated uses of expressions like ‘詳しくはお問い合わせください/Kuwashiku wa otoiawase kudasai’ (Please inquire for more information). These kinds of typical patterns can/should be consistently used when authoring, which directly leads to a consistent translation. When we incorporate translation memories (TMs), further productivity gains can be expected.

The similar text search function is materialised using an API provided by the TexTra platform, which is based on the string search algorithm proposed by Okazaki and Tsujii (2010). As authors input text, this module searches the database for similar segments. At the point where the similarity score exceeds a certain threshold—currently 30%—the segment is presented below the diagnosis of the CL authoring assistant (Figure 7.7). Authors can adopt suggestions as appropriate by simply clicking the button with the similarity score.

This module can also be connected to the document elements; for example, the search space of **Personal condition** should be limited to the text embedded in the same element, which helps authors to efficiently search and appropriately reuse the previously written segments.

7.4 Summary

In this section, we presented the design principle of the MuTUAL system and described the details of the implementation. The system is currently operational online, implementing core modules of controlled authoring and translation. The principal innovation in the system is that it makes use of document structuring based on the DITA framework, which affords a basis for fine-grained, context-dependent CL rules coupled with pre-translation processing. It consequently enables MT to generate outputs appropriate to their functional context without degrading the quality of the source. To the best of our knowledge,

this is the first authoring environment that seamlessly combines document templates, CL authoring assistants and MT systems.

The next issue is to evaluate the system. While the system implements various functions to be evaluated, we focus on the CL authoring assistant module, since writing in accordance with a number of CL rules and controlled terms is the most difficult part of controlled authoring and should be examined in detail.⁴ The evaluation of the CL authoring assistant module will proceed in two steps: (1) precision and recall benchmarking of the CL violation detection component (Chapter 8); (2) usability evaluation of the whole module when the system is used in the human authoring process (Chapter 9).

Notes

- 1 As of 2020, TexTra only provides NMT models instead of SMT models.
- 2 Technically, we can allow users to choose the best output from different MT outputs, but MuTUAL caters to only one of the two systems as the implemented CL rules are tailored to a particular system.
- 3 Agreement scores are the harmonic mean of the ST score and the BT score.
- 4 Although evaluation of document-level authoring is also an important issue to be tackled, we assume that filling in the topic template is not as difficult as sentence-level CL authoring because the number of document elements to be followed is much smaller than that of the rules and terms defined by the CL and controlled terminology.

8 Evaluation of CL violation detection component

This chapter presents the benchmark evaluation results of the CL violation detection component of the MuTUAL system, which implemented 30 selected CL rules. To address the research question **RQ-A1** (How accurately does the system detect CL rule violations in text?), we measure the precision and recall performance of the component using a benchmark test set in which all CL violations are manually annotated.

In Section 8.1, we explain the evaluation setup. We present the overall results in Section 8.2 and further examine the results in Section 8.3. Finally, Section 8.4 summarises the results, indicating directions for future improvement and deployment of the component.

8.1 Setup

For our evaluation, we used the same dataset used in the human evaluation in Section 4.2.1.1, which was collected from the Toyohashi City website. We first selected Japanese sentences that violated each of our 30 CL rules, with four sentences per rule, i.e. a total of 120 sentences. We then checked whether each sentence violated more than one rule. A total of 307 CL rule violations were manually detected, and 95 sentences exhibited multiple violations. We also added four sentences without any violations. Our dataset thus consists of 124 sentences.

For each rule, we counted the number of cases where: violations were correctly detected (true positives: TP); non-violations were mistakenly detected as violations (false positives: FP); and violations were not detected (false negatives: FN). This enabled us to compute the metrics of precision, recall and F-measure as follows:

$$Precision = \frac{TP}{TP + FP} \quad (8.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (8.2)$$

$$F\text{-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8.3)$$

8.2 Overall results

Table 8.1 shows the results of the evaluation of the detection component for each rule. We can see that 23 rules—rules T06, T07, T09, T13, T16, T18, T21, R04, R05, R07, R10, R11, R12, R16, R17, R18, R19, R25, R26, R27, R34, R35 and R37—achieve more than 0.7 F-measure, with 13 of those rules—rules T07, T09, T16, T18, T21, R07, R11, R12, R16, R17, R18, R25 and R35—obtaining the perfect score of 1.0. These are encouraging results for practical deployment of the detection component, although larger-scale evaluation will be needed to confirm the results.

Overall, the total recall attains a high score of 0.886. In contrast, the overall precision is 0.673, meaning that about one third of the detections are false. We further observe that, focusing on the rules for which the F-measure is less than 0.7—rules T10, R02, R03, R08, R09, R20 and R33—precision is much lower than recall.

To find ways to improve the performance of the system, we further analysed the results of individual rules. This is presented below.

8.3 Detailed analysis

First, we look at the rules with good performance (both precision and recall higher than 0.7), followed by the rules with poor performance (both precision and recall lower than 0.7) and, finally, the rules with mixed performance (precision above/below 0.7 and recall below/above 0.7).

We note that for the rules with both high precision and high recall (more than 0.7)—rules T06, T07, T09, T13, T16, T18, T21, R07, R10, R11, R12, R16, R17, R18, R19, R25, R26, R27, R34 and R35—deviations from string matching rules can be captured using part-of-speech (POS) information alone, and the range of deviations is limited. Therefore, it is rather easy to formulate corresponding matching rules that are comprehensive. The following is an example of a sentence that violates rule R11, for which the violation was correctly identified (thus a true positive (TP) example); we also provide the reference human translation (HT).

Rule R11: Avoid using attributive use of *shika-nai* (しか-ない).

TP 自生地には観察会の2日間しか入れません

(Jisei-chi ni wa kansatsu-kai no futsuka-kan *shika haire-masen*)

HT One **can only** enter the wildlife area during the two days of the observation event.

This sentence consists of a variant form of ‘*shika-nai*’ construction; ‘ません/*masen*’ is an honorific mode of an auxiliary verb ‘ない/*nai*’. These kinds of variants are easily and reliably covered by a small number of matching rules using the morphological analyser.

Table 8.1 Results of the benchmark evaluation

Rule No.	Violation (#)	TP	FP	FN	TN	Precision	Recall	F-measure
T06	7	7	1	0	117	0.875	1.000	0.933
T07	4	4	0	0	120	1.000	1.000	1.000
T09	8	8	0	0	116	1.000	1.000	1.000
T10	11	11	16	0	99	0.407	1.000	0.579
T13	7	7	2	0	116	0.778	1.000	0.875
T16	4	4	0	0	120	1.000	1.000	1.000
T18	6	6	0	0	118	1.000	1.000	1.000
T21	4	4	0	0	120	1.000	1.000	1.000
R02	28	17	14	11	82	0.548	0.607	0.576
R03	16	11	33	5	75	0.250	0.688	0.367
R04	22	22	10	0	97	0.688	1.000	0.815
R05	7	4	0	3	117	1.000	0.571	0.727
R07	6	6	0	0	118	1.000	1.000	1.000
R08	9	4	11	5	104	0.267	0.444	0.333
R09	4	3	25	1	100	0.107	0.750	0.188
R10	7	6	0	1	117	1.000	0.857	0.923
R11	4	4	0	0	120	1.000	1.000	1.000
R12	5	5	0	0	119	1.000	1.000	1.000
R16	24	24	0	0	103	1.000	1.000	1.000
R17	4	4	0	0	120	1.000	1.000	1.000
R18	5	5	0	0	119	1.000	1.000	1.000
R19	47	47	5	0	83	0.904	1.000	0.949
R20	5	3	4	2	115	0.429	0.600	0.500
R25	5	5	0	0	120	1.000	1.000	1.000
R26	14	13	0	1	111	1.000	0.929	0.963
R27	9	8	0	1	115	1.000	0.889	0.941
R33	7	7	11	0	107	0.389	1.000	0.560
R34	9	8	0	1	115	1.000	0.889	0.941
R35	7	7	0	0	119	1.000	1.000	1.000
R37	12	8	0	4	115	1.000	0.667	0.800
Total	307	272	132	35	3317	0.673	0.886	0.765

In contrast, rules for which both precision and recall are below 0.7—rules R02, R03, R08 and R20—are not easily handled with morphological information alone. We illustrate this below with rule R02.

Rule R02: Do not omit subjects.

In Japanese sentences, subjects are apt to be followed by the particles ‘が/*ga*’ or ‘は/*wa*’.

TP 今後、広報等による啓発活動などで認定事業を応援していきます
(*Kongo, koho-to ni yoru keihatsu katsudo nado de nintei jigyo o oen shite iki-masu*)

HT In the future, **we** will support certified business by educational activities through advertisements, etc.

In this true positive case, the lack of ‘*ga*’ or ‘*wa*’ correctly corresponds to an omission of the subject.¹ However, this is not always the case, as the false positive (FP) example below shows:

FP 実行委員会一同努力しています
(*Jikko-iinkai ichido doryoku-shite imasu*)

HT The executive committee is working hard as one

Though the subject ‘the executive committee’ (実行委員会/*jikko-iinkai*) is present in the sentence, the system mistakenly detected an omission of subject because this sentence lacks the particle.

In the false negative (FN) case below, the sentence lacks a subject in the latter clause. The human translator inferred ‘they’ as a subject, but the system failed to detect this subject omission, because the sentence includes both clue particles ‘*ga*’ and ‘*wa*’.

FN 家庭や地域は、子どもが多くの時間を日常的に過ごす場所であり、生活の中で様々なことを学んでいきます。
(*Katei ya chiki wa, kodomo ga oku no jikan o nichijo-teki ni sugosu basho de ari, seikatsu no naka de samazamana koto o manande iki-masu*)

HT Homes and communities are places where children spend a lot of time every day, and where **they** learn many things about life.

Generally speaking, the detection of missing elements—subjects (rule R02), objects (rule R03) and parts of words in enumeration (rule R20)—requires a deep analysis of sentence structure in addition to surface morphological information. To do so, we need to incorporate other tools, such as parsers and chunkers, and techniques such as machine learning. Likewise, rule R08 (Avoid inserted adverbial clauses) needs sentence structure parsing.

Finally, for the rest of the rules with mixed performance of precision and recall—rules T10, R04, R05, R09, R33 and R37²—, the surface POS constructions are simple, but correct identification requires fine-grained distinctions, as illustrated below.

Rule R05: Avoid using particle *ga* (が) for objects.

Particle ‘*ga*’ usually follows the subject, and we regulate ‘*ga*’, which follows the object. It is however difficult to distinguish subjects and objects. Based on the heuristic knowledge of municipal texts that the particle ‘*ga*’ is likely to be used for objects when an auxiliary verb, in this case, ‘*dekiru*’ (can), is used after it, we decided to capture only this pattern.

- TP** 防災航空隊は、災害発生時に直ちに防災ヘリコプターが運航できるように、24時間勤務体制とする。
(*Bosai-kokutai wa, saigai-hassei-ji ni tadachini bosai herikoputa ga unko dekiru yoni, 24-jikan-kinmu-taisei to suru.*)
- HT** The Disaster Prevention Fleet has a 24-hour duty system so that they can operate their emergency helicopters promptly if a disaster occurs.

Although this particular pattern was comprehensively captured, there are other patterns where ‘ga’ is used for objects. One of these patterns we missed is ‘ga’ with an auxiliary verb ‘*reru/rareru*’.³ Here is an example of the pattern.

- FN** 外国人市民が豊橋市のまちづくりなどに対して考えや意見が述べられるよう審議会等への登用を積極的に進める。
(*Gaikoku-jin shimin ga Toyohashi-shi no machizukuri nado ni taishite kangae ya iken ga nobe-rareru yo shingikai-to e-no toyo o sekkyokuteki ni susumeru.*)
- HT** We are proceeding proactively to appoint foreign residents to the council, etc., so that they can express their thoughts and opinions on making Toyohashi a better city.

This is a false negative case; ‘ga’ indeed follows the objects ‘考えや意見/*kangae ya iken*’ (thoughts and opinions), but it was missed. To comprehensively cover the rule-violating patterns and improve recall, we need to investigate more positive and negative examples and create additional matching rules.

Next, we look at rule T10, for which precision is low and recall is high.

Rule T10: Do not use *reru/rareru* to express the potential mood or honorifics.

‘*reru/rareru*’ has multiple grammatical usages: passive, potential and honorific. Since distinguishing these usages is quite difficult, we decided to simply detect this auxiliary verb ‘*reru/rareru*’ without considering its usage.

- TP**すでに請求された方は対象になりません
(*Sude-ni seikyu sa-re-ta kata wa taisho ni nari-masen*)
- HT** This does not apply to those who have already **claimed**

In this case, ‘*re*’ (inflected form of ‘*reru*’) happens to be honorific, and it is correctly detected. There are, however, many cases where ‘*reru/rareru*’ is used as passive:

- FP** 在留期間が3か月を超えて適法に在留する外国人の方も、住民票に記載されるようになります
(*Zairyu kikan ga 3-kagetsu o koete tekiho ni zairyu suru gaikoku-jin no kata mo, jyumin-hyo ni kisai-sa-reru yo-ni nari-masu*)

HT Foreigners who remain in the country legally for a residential period of more than three months will also begin to **be recorded** in the resident register

Considering the low precision (0.407) of rule T10, further improvement is necessary. We plan to apply machine learning methods as they have been shown to be effective in this kind of disambiguation task.

Similarly, rule R33 (Use Chinese kanji characters for verbs as much as possible instead of Japanese kana characters) seems easy to implement at first sight. However, we found that there are (auxiliary) verbs which tend to be, or should be, written in Japanese kana characters. In this experiment, we saw 11 false positive detections from 124 sentences and observed that some of the verbs are common ones, such as ‘みる/*miru*’ (see), ‘かける/*kakeru*’ (put) and ‘いう/*iu*’ (say). This implies that false positive detections could occur frequently unless we rule out these verbs when implementing the rule. Therefore, it is necessary to supply a list of verbs commonly written in kana.

8.4 Summary

In this chapter, to answer the research question **RQ-A1** (How accurately does the system detect CL rule violations in text?), we constructed a benchmark test dataset by manually annotating all CL violations in the selected sentences and evaluated the precision and recall performance of the CL violation detection component using the dataset. The results showed that 20 rules achieved high benchmark scores (both precision and recall higher than 0.7), with 13 of these rules obtaining perfect precision and recall scores. This encourages us to deploy these rules in actual scenarios and to proceed to the next phase of evaluation to assess the usability of the system.

There is, however, still much room for improvement. The detailed analysis of the results gave us insight into how we can improve the performance of the remaining rules: in summary, (1) language tools such as parsers and chunkers, (2) machine learning techniques and (3) lexical resources. It is also important to further validate the created matching rules using a larger developmental dataset that includes a wide variety of positive and negative examples (Adachi et al., 2013; Miyata et al., 2014). Finally, we should remain aware of the difficulty of achieving perfect benchmark scores for all rules because the language tools and machine learning techniques are not perfect as such. We aim to alleviate some of the problems through our interface design, which we will address in Section 9.1.1.

Notes

- 1 A human translator inferred ‘we’ as a subject in the HT.
- 2 In more detail, for rules R05 and R37, precision is high and recall is low, while for rules T10, R04, R09 and R33, precision is low and recall is high.
- 3 Note that there are several usages of ‘*reru/rareru*’, as we will describe just below. If we apply rule T10 first, we might change this pattern into ‘*ga*’ + ‘*dekiru*’, which is now captured by the current rule R05.

9 System usability evaluation

This chapter presents the experimental results of the usability evaluation of our CL authoring assistant, a core module of MuTUAL. The key feature of this module is that it supports users' decision-making at each step in validating the source. In the previous chapter, we benchmarked how accurately our system can detect CL violations. The results showed that 20 of 30 CL rules achieved reasonably high precision and recall (more than 0.7). Here, we move on to the next phase in assessing the actual usability of the system, implementing selected CL rules and controlled terminology. The research questions to be answered in this chapter are **RQ-A2** (Is our system usable for non-professional writers?) and **RQ-A3** (Does the use of the system help improve the quality of ST and TT?).

The purpose of the usability evaluation is to assess how usable our authoring assistant is for non-professional users and to get feedback from users, which enables us to further develop and improve the system's functions and interfaces.

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 chapter, we follow the ISO standard for human-computer interaction (ISO, 2010) and related studies, such as Doherty and O'Brien (2013) and Sauro and Lewis (2012).

Based on the ISO definition of usability introduced in Section 2.5.3.2, we can further elaborate our research questions as follows: (1) Does the system help reduce CL violations and improve MT quality? (**effectiveness**); (2) Does the system help reduce the burden of users as regards controlled writing? (**efficiency**); (3) Is the system easy for non-professional writers to use and is it favourably accepted? (**satisfaction**) Here research question **RQ-A1** pertains to all the questions, while research question **RQ-A2** specifically corresponds to the first question about **effectiveness**.

To assess these three aspects, we design a rewriting task in which two groups of participants are asked to amend Japanese source sentences that violate CL rules and terminology, respectively with and without the aid of the system. Although in our scenario the CL authoring assistant is mainly intended for drafting source texts from scratch, as a pilot study, we emulate the post-hoc revision settings. Even in the drafting process, 'self-revision' in accordance with CL rules and terminology is the difficult part in controlled authoring for writers, which is what this system

Table 9.1 Quantitative aspects to be measured

Effectiveness: accuracy and completeness with which users achieve specified goals	(1-a)	The number of violations
	(1-b)	The quality of the product (both ST and MT)
Efficiency: resources expended in relation to the accuracy and completeness with which users achieve goals	(2-a)	Time spent on correcting violations
	(2-b)	The number of edits needed to correct violations
Satisfaction: freedom from discomfort and positive attitudes towards the use of the product	(3-a)	Satisfaction with the task
	(3-b)	Satisfaction with the system

mainly supports. Thus it is reasonable to presume that the usefulness of the system for drafting can be measured through the revision settings. In terms of conducting an experiment, the notable merits of the revision settings are as follows:

- We can ask participants to rewrite the same source text, which enables us to systematically analyse the revision outputs generated by different participants.¹
- We can prepare source text dataset so that it contains violations of all the CL rules, which makes it possible to investigate which CL rules are particularly difficult to follow among others.

As quantitative measures, we (1-a) count the number of corrected violations and (1-b) evaluate quality of ST and MT outputs, and (2-a/b) measure the time/edit taken to correct violations, and (3-a/b) gauge subjective satisfaction with the task/system through questionnaires (see Table 9.1). In addition, we conduct follow-up interviews to get qualitative feedback from users, and inspect user products in detail.

Section 9.1 describes the experimental setup of the usability evaluation. We discuss the results of the evaluation in Section 9.2 and summarise the findings of the study in Section 9.3, showing future research venues for further implementation and evaluation.

9.1 Experimental setup

9.1.1 Task design

9.1.1.1 MT system and guideline

We continued to focus on the two MT systems, TransGateway (system B) and TexTra (system D). Therefore in this experiment we used two CL guidelines, guidelines B and D, formulated in Section 4.3. Guideline B comprises 30 rules while guideline D comprises 31. The total number of distinct rules is 36, with 25 rules belonging to both guidelines (Table 9.2).

Table 9.2 CL rules and implementation (with confidence scores)

No.	Rule	Guideline		Implement
		B	D	
T01	Try to write sentences of no more than 50 characters.	✓	✓	✓ (10)
T02	Do not interrupt a sentence with a bulleted list.	✓	✓	
T05	Ensure the relationship between the modifier and the modified is clear.	✓	✓	
T06	Use the particle <i>ga</i> (が) only to mean ‘but’.	✓	✓	✓ (9)
T07	Do not use the preposition <i>tame</i> (ため) to mean ‘because’.	✓		✓ (10)
T09	Avoid using multiple negative forms in a sentence.	✓		✓ (10)
T10	Do not use <i>reru/rareru</i> (れる/られる) to express the potential mood or honorifics.	✓	✓	✓ (4)
T11	Avoid using words that can be interpreted in multiple ways.	✓	✓	
T13	Avoid using the expression <i>to-iu</i> (という).	✓		✓ (8)
T16	Avoid using the expressions <i>omowa-reru</i> (思われる) and <i>kangae-rareru</i> (考えられる).	✓		✓ (10)
T18	Avoid the single use of the form <i>tari</i> (たり).		✓	✓ (10)
T20	Use words from a general Japanese–English dictionary.	✓	✓	
T21	Avoid using compound <i>sahen</i> -nouns.		✓	✓ (10)
T22	Ensure there are no typographical errors or missing characters.	✓	✓	
R02	Do not omit subjects.	✓	✓	✓ (5)
R03	Do not omit objects.	✓	✓	
R04	Do not use commas for connecting noun phrase enumeration.	✓	✓	✓ (7)
R05	Avoid using the particle <i>ga</i> (が) for objects.	✓	✓	✓ (8)
R07	Avoid using <i>te-kuru</i> (てくる) / <i>te-iku</i> (ていく).		✓	✓ (10)
R08	Avoid inserted adverbial clauses.		✓	
R09	Do not end clauses with nouns.	✓	✓	
R10	Avoid using <i>sahen</i> -noun + auxiliary verb <i>desu</i> (です).	✓	✓	✓ (10)
R11	Avoid using attributive use of <i>shika-nai</i> (しかない).	✓	✓	✓ (10)
R12	Avoid using verb + <i>you-ni</i> (ように).		✓	✓ (10)
R16	Avoid using the particle <i>nado</i> (など/等).		✓	✓ (10)
R17	Avoid using giving and receiving verbs.	✓	✓	✓ (10)
R18	Avoid using verbose wording.	✓	✓	✓ (10)
R19	Avoid using compound words.	✓	✓	✓ (9)
R20	Do not omit parts of words in enumeration.	✓	✓	✓ (4)
R25	Do not omit expressions that mean ‘per A’.	✓	✓	✓ (10)
R26	Avoid using the conjunctive particle <i>te</i> (て).	✓	✓	✓ (10)
R27	Avoid using the ‘if’ particle <i>to</i> (と).	✓		✓ (10)
R33	Use Chinese kanji characters for verbs as much as possible instead of Japanese kana characters.	✓	✓	✓ (4)
R34	Avoid leaving bullet marks in texts.	✓	✓	✓ (10)
R35	Avoid using machine-dependent characters.	✓	✓	✓ (10)
R37	Avoid using square brackets for emphasis.	✓	✓	✓ (10)

9.1.1.2 CL rule implementation

The benchmark evaluation results in the previous chapter enabled us to estimate how reliable the CL detection is (see Table 8.1). In order to avoid or alleviate the problems of false detection, we took the following measures in this experiment. If the precision of the rule was below 0.4, we chose not to implement the rule, as the false alerts might cause annoyance. We thus did not implement these three rules: R3 (Do not omit objects), R8 (Avoid inserted adverbial clauses) and R9 (Do not end clauses with nouns). If the precision was above 0.4, we mapped the precision scores to a 10-point scale and assigned a **confidence score** to each rule, which informs users how reliable the detection presented by the system is (shown in Figure 9.1). The rightmost column in Table 9.2 shows the 28 implemented rules with their confidence scores.

9.1.1.3 Data

To count the number of corrected violations (**1-a: effectiveness**), we prepared a manually annotated dataset. We used the same source sentences used in the previous benchmark evaluation (Section 8.1) and selected 30 sentences to ensure that the dataset contained at least one violation of each of the 36 rules. We also artificially modified two proper nouns from the municipal domain into proscribed forms. The final dataset consisted of 67 violations of guideline B and 76 violations of guideline D, including two term violations of each.

Figure 9.1 User interface

9.1.1.4 *Condition*

For each of the two CL guidelines, guidelines B and D, one group of participants rewrites sentences with the sole aid of a print copy of the guideline and a term list² without access to the system’s support functions (**control**), while the other group can use the full assistance of the system (**treatment**). Thus, four conditions were prepared: (1) Control group with guideline **B** (**CB**); (2) Control group with guideline **D** (**CD**); (3) Treatment group with guideline **B** (**TB**); (4) Treatment group with guideline **D** (**TD**).

9.1.1.5 *Procedure*

Each participant is presented with a sentence in the input box of the system (see Figure 9.1) and is asked to amend any segments that violate the CL rules or terminology, while maintaining the meaning of the source. All functions of the authoring assistants 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 an interim editing result and the elapsed time each time a participant presses the ‘Enter’ or ‘Backspace’ keys (**2-a/b: efficiency**).

9.1.1.6 *Post-task questionnaire*

To investigate **3-a/b: satisfaction** with the task and the system, we employed two standardised questionnaires that are 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 9.3).³ To evaluate the usability of the system itself, we used an SUS with ten questions on a five-point Likert scale from ‘1: strongly disagree’ to ‘5: strongly agree’ (Table 9.4).⁴ Odd-numbered questions are worded positively, while even-numbered questions are worded negatively.

9.1.1.7 *Follow-up interview*

To get detailed user feedback for the system interfaces and functionalities, we conducted a follow-up interview in a semi-structured manner. Most of the

Table 9.3 After-scenario questionnaire (ASQ)

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 (on-line help, messages, documentation) when completing the tasks.

Table 9.4 System usability scale (SUS)

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.

questions were based on the answers of the questionnaire(s); we asked participants why they chose the answers in the questionnaires. We also specifically asked the following questions, showing their resultant sentences:

- for all** How did you check the conformity to the CL guideline and the term list? And what did you find difficult?
- for all** Which rules were difficult to follow? And why?
- for all** Which sentences were difficult to amend?
- for all** What kind of support do you think would be needed to effectively and efficiently complete this task?
- for treatment** Did you make use of all functions the system provides? If not, why?
- for treatment** How did you feel about the functions and interface of the system?

9.1.1.8 MT evaluation

To evaluate the resultant MT outputs (**1-b: effectiveness**), we conducted traditional human evaluation based on LDC (2005). An evaluator judges each MT output in terms of *fluency* from ‘5: Flawless English’ to ‘1: Incomprehensible’ and *adequacy* from ‘5: All’ (of the meaning correctly correctly expressed) to ‘1: None’. We translated the resulting sentences of groups CB and TB by system B and those of CD and TD by system D into English, registering the user dictionaries of both systems (respectively, CB-dic, TB-dic, CD-dic and TD-dic). As a baseline output, we translated the original ST with and without dictionaries (Original-dic and Original). Furthermore, in order to see how much improvement can be attained when ST perfectly conforms to the CL guideline, we ourselves created ST that complies with each of the guidelines B and D, and translated them with and without dictionaries (Oracle-B-dic, Oracle-B, Oracle-D-dic and Oracle-D).

9.1.1.9 ST evaluation

To evaluate the ST quality (**1-b: effectiveness**), we followed the method presented by Hartley et al. (2012). An evaluator judges each ST in terms of *readability* on a four-point scale—‘4: Easy to read’; ‘3: Fairly easy to read’; ‘2: Fairly difficult to read’; ‘1: Difficult to read’, comparing all versions of STs (one original, three CB, three TB, three CD, three TD and two oracle STs).

9.1.2 Implementation

We recruited 12 university students, all of them native speakers of Japanese who regularly write 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 municipal writers. Three participants were randomly placed in each condition: CB, CD, TB and TD.

The flow of the whole experiment is as follows:

1. Instruction
2. Preliminary session (5 sentences for exercises)
3. Task proper
 - 3-1. Set 1 (10 sentences)
 - 3-2. Set 2 (10 sentences)
 - 3-3. Set 3 (10 sentences)
4. Post-study questionnaire
5. Interview

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 consisting of 10 sentences, and let participants take a short rest between the sets.

After the task, we asked them to answer the ASQ and SUS,⁵ 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. For the ST evaluation task, we employed three native Japanese speakers to evaluate all versions of the source sentences.

9.2 Results and discussions

9.2.1 Effectiveness

9.2.1.1 Violation correction

Table 9.5 shows the results of the violation correction. The 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 ($t = -2.878$, $df = 10$, $p = 0.016$).

The detailed results for each rule in Table 9.6 revealed that the correction rate for the 14 rules of the treatment group is higher than that of the control group, 13 rules of which are implemented in the system. In particular, six rules—T09, T10, T11, R07, R26 and R27—saw improvement of more than 30%. For example, R07 (Avoid using *te-kuru/te-iku*) and R26 (Avoid using conjunctive particles *te*) deal with the subtle linguistic elements which are not focused on in a usual writing. This suggests that the system assistance was of great help for participants in comprehensively spotting the violations, especially minor ones, in the text.

On the other hand, four rules—T20, T22, R03 and R20—of the treatment group showed results lower than those of the control group (Table 9.6). We also noted that three of these four rules—T20, T22 and R03—are not yet implemented in our system. As for R20 (Do not omit parts of words in enumeration), we realised that the system failed to capture the CL-violating segment ‘市立小・中学校/*shiritsu sho- · chu-gakko*’ and ‘月・水・金曜日/*getsu- · sui- · kin-yobi*’.⁶ This implies that users tended to rely on the system and overlook any violations the system did not detect or failed to detect. It is worthwhile to point out that (1) rules T20 and T22 can be implemented by utilising existing dictionaries and spell checkers, (2) rule R03 can be implemented by integrating deeper language tools such as parsers and chunkers and (3) rule R20 can be improved by adding more string matching rules, a task for future work.

Finally, the correction rate for term violations (see the row ‘Term’ in Table 9.6) did not achieve a high rate even for the treatment group, which was contrary to our expectations since all the term violations were correctly detected by the system and we assumed it was easy to correct them. We discuss the issue of human oversight of the system detection in Section 9.2.4.

Table 9.5 Effectiveness for each condition

	Treatment			Control		
	TB	TD	Mean	CB	CD	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	73.2	74.2

Table 9.6 Correction rate for each rule (* implemented rule)

<i>Rule</i>	<i>Treatment</i>	<i>Control</i>	<i>Improvement (%)</i>
T01*	75.9	68.5	7.4
T02	100.0	100.0	0.0
T05	66.7	66.7	0.0
T06*	100.0	100.0	0.0
T07*	100.0	100.0	0.0
T09*	66.7	0	66.7
T10*	100.0	66.7	33.3
T11	66.7	33.3	33.3
T13*	100.0	100.0	0.0
T16*	100.0	100.0	0.0
T18*	100.0	100.0	0.0
T20	50	83.3	-33.3
T21*	100.0	100.0	0.0
T22	44.4	72.2	-27.8
R02*	83.3	83.3	0.0
R03	27.8	33.3	-5.6
R04*	100.0	100.0	0.0
R05*	83.3	66.7	16.7
R07*	100.0	33.3	66.7
R08	50	50	0.0
R09	100.0	100.0	0.0
R10*	83.3	75	8.3
R11*	100.0	83.3	16.7
R12*	100.0	100.0	0.0
R16*	93.3	86.7	6.7
R17*	100.0	100.0	0.0
R18*	100.0	100.0	0.0
R19*	83.3	58.3	25.0
R20*	66.7	83.3	-16.7
R25*	100.0	100.0	0.0
R26*	95.8	50	45.8
R27*	83.3	50	33.3
R33*	100.0	91.7	8.3
R34*	100.0	100.0	0.0
R35*	100.0	100.0	0.0
R37*	93.3	83.3	10.0
Term*	58.3	50	8.3
Total	82.9	74.2	8.7

We can assume that the fewer CL violations an ST contains, the more machine-translatable and more readable it is. In the next, we look at the results of the product quality evaluation of MT outputs and ST.

9.2.1.2 *MT quality*

Tables 9.7 and 9.8 summarise the human-evaluation results of fluency and adequacy for systems B and D.⁷

Table 9.7 Result of MT quality evaluation (system B)

<i>Condition</i>	<i>Fluency</i>	<i>Adequacy</i>
TB-dic	2.74	3.20
CB-dic	2.72	3.19
Original-dic	2.36	2.80
Original	2.31	2.68
Oracle-B-dic	2.87	3.46
Oracle-B	2.83	3.38

Table 9.8 Result of MT quality evaluation (system D)

<i>Condition</i>	<i>Fluency</i>	<i>Adequacy</i>
TD-dic	2.58	2.89
CD-dic	2.57	2.80
Original-dic	2.32	2.62
Original	2.34	2.62
Oracle-D-dic	2.70	3.34
Oracle-D	2.79	3.32

We can see that the user's products (TB-dic, CB-dic, TD-dic and CD-dic) achieved higher scores of fluency and adequacy than Original(-dic), and Oracle(-dic) achieved the highest scores. We can also point out that the scores of the treatment (TB-dic and TD-dic) are slightly higher than those of the control (CB-dic and CD-dic). This demonstrates, as we expected, that the ST becomes more machine-translatable in accordance with the degree to which the ST conforms to CL.⁸

Table 9.9 shows examples of several versions of ST and corresponding MT (system D) outputs. The Original version is of the lowest quality and the Oracle-dic version is of the highest, while the quality of one of the TD-dic versions falls in between the two. We can observe that the ST of the TD-dic version conforms to rule R36 (Avoid using square brackets for emphasis), but it still needs to be improved according to rule T05 (Ensure the relationship between the modifier and the modified is clear). Although the treatment group achieved a fairly high correction rate (about 83%), yet a higher correction rate is needed to further improve the MT quality.

It should also be noted that the MT dictionary contributed to improving the MT quality in the example. Without a dictionary, the term '地域資源回収/*Chiki-shigen-kaishu*' was translated into 'The local resource recovery', while with the dictionary, it was correctly translated into 'Community resource collection'. However, we notice that for system B, the dictionary consistently has an positive effect on the MT outputs (Table 9.7), but for system D, it does not

Table 9.9 Example of ST and MT (system D) of different conditions

Original	<p>地域資源回収は、家庭から出る「ごみ」をリサイクルに回し、「資源」として再利用する、「エコ」な活動です。 (<i>Chiki-shigen-kaishu wa, katei kara deru "gomi" o risaikuru ni mawashi, "shigen" toshite sairiyo-suru, "eko" na katsudo desu.</i>)</p> <hr/> <p>The local resource recovery, “garbage” household recycled and reused as “resources”, “Eco” activities. (Fluency: 1.33, Adequacy: 3.33)</p>
TD-dic	<p>地域資源回収は、家庭から出る廃棄物をリサイクルに回し、資源として再利用するというエコな活動だ。 (<i>Chiki-shigen-kaishu wa, katei kara deru haikibutsu o risaikuru ni mawashi, shigen toshite sairiyo-suru eko na katsudo da.</i>)</p> <hr/> <p>Community resource collection, waste from households to recycling, environmentally friendly to reuse as resources. (Fluency: 2.33, Adequacy: 3.67)</p>
Oracle-dic	<p>地域資源回収はエコな活動です。家庭から出るごみをリサイクルに回し、資源として再利用します。 (<i>Chiki-shigen-kaishu wa eko na katsudo desu. Katei kara deru gomi o risaikuru ni mawashi, shigen toshite sairiyo shimasu.</i>)</p> <hr/> <p>Community resource collection is ecological activities. The waste from home recycled and reused as resources. (Fluency: 2.67, Adequacy: 4)</p>
Reference	<p>Community resource collection is the eco-friendly activity of recycling the waste produced by our households and re-using it as a resource.</p>

necessarily improve the MT quality (Table 9.8). We consider this to be because integrating the dictionary into the SMT system (system D) is rather difficult as SMT systems sometimes ‘overreact’ to small changes of, say, a single word.

9.2.1.3 ST quality

Table 9.10 shows the results of source readability.⁹ Readability scores indicate the average scores of the four-point-scale readability judgement. We also calculated the ratio of acceptable-level sentences that were judged as ‘4: Easy to read’ or ‘3: Fairly easy to read’.

We can see that for all versions of STs the readability score is generally high, and that about 90% of source sentences achieved acceptable-level readability. Notably, the rewritten versions (TB, TD, CB, CD, Oracle-B and Oracle-D) all had higher readability than that of the Original version. Although it is difficult to draw a clear conclusion from these results because of the low consistency of the

Table 9.10 Result of ST quality evaluation

<i>Condition</i>	<i>Readability score</i>	<i>Acceptable level (%)</i>
TB	3.35	85.2
TD	3.31	85.2
CB	3.52	91.5
CD	3.36	87.0
Original	3.23	84.4
Oracle-B	3.39	86.7
Oracle-D	3.40	90.0

Table 9.11 Time efficiency

	<i>Treatment</i>			<i>Control</i>		
	<i>TB</i>	<i>TD</i>	<i>Mean</i>	<i>CB</i>	<i>CD</i>	<i>Mean</i>
Total time (sec.)	2405	2206	2306	3744	2844	3294
— per sentence (sec.)	80.2	73.5	76.9	124.8	94.8	109.8
— per correction (sec.)	43.7	35.2	39.5	75.3	51.1	63.2

human judgements, we can say that in general our CL rules and the term list would contribute to improving readability.

9.2.2 Efficiency

9.2.2.1 Time

Table 9.11 shows the results of the time-efficiency measures. The 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 ($t = 2.826$, $df = 10$, $p = 0.018$). These results demonstrate that our system greatly enhanced the efficiency of checking for and correcting violations.

9.2.2.2 Editing effort

Editing effort or burden can be measured by the edit distance, or how different given two text strings are. In this study, we used the Levenshtein distance, the minimum number of single-character edit operations—insertions, deletions, and substitutions—to transform one given string into another (Levenshtein, 1966). We present the scores of edit distance (E), accumulative edit distance (AE) and edit efficiency.

Table 9.12 Detailed edit log

Version	Edit log	E	AE
Original	入場券は4月22日（金）午前10時販売開始です。	0	0
1	入場券は4月22日（金）午前10時に販売開始です。	1	1
2	入場券は4月22日（金）午前10時に販売を開始です。	1	2
3	入場券は4月22日（金）午前10時に販売を開始で。	1	3
4	入場券は4月22日（金）午前10時に販売を開始。	1	4
5	入場券は4月22日（金）午前10時に販売を開始します。	3	7
6	入場券は4月22日（金）午前10時に販売開始します。	1	8
7	入場券は4月22日（金）午前10時に販売が開始します。	1	9
8	入場券は4月22日（金）午前10時に販売が開始ます。	1	10
9	入場券は4月22日（金）午前10時に販売が開始されます。	2	12
10	入場券は4月22日（金）午前10時に販売が開始さます。	1	13
11	入場券は4月22日（金）午前10時に販売が開始ます。	1	14
12	入場券は4月22日（金）午前10時に販売が開始します。	1	15
13	入場券は4月22日（金）午前10時に販売開始します。	1	16
Final	入場券は4月22日（金）午前10時に販売を開始します。	1	17
Reference	Ticket sales will start at on Friday, April 22 at 10:00 AM.		

As an indication of the editing effort, previous literature usually computed the edit distance (E) between an original sentence and its final rewritten version. However, this does not necessarily reflect the actual ‘trial-and-error’ operations during the task. We thus introduce accumulative edit distance (AE), which is calculated by accumulating the edit distances between the sequential pairs of interim sentences. Also, edit efficiency can be calculated by dividing the score of AE by E, which indicates how efficiently the editing was performed. A smaller value indicates higher efficiency.

Let us take a look at a detailed edit log in Table 9.12. The original sentence ‘入場券は4月22日（金）午前10時販売開始です。/Nyuujoken wa 4-gatsu 22-nichi (kin) gozen 10-ji hanbai kaishi desu.’ was finally rewritten into ‘入場券は4月22日（金）午前10時に販売を開始します。/Nyuujoken wa 4-gatsu 22-nichi (kin) gozen 10-ji **ni** hanbai **o** kaishi **shimasu**.’ by adding two characters, ‘に/ni’ and ‘を/o’, and substituting two characters ‘です/desu’ into ‘しま/shima’, i.e. the edit distance between the original sentence and the final version is four. On the other hand, the accumulative edit distance is 17 and thus the edit efficiency is 4.25 (=17/4). This indicates that the participant tried several ways of rewriting before producing the final version. Indeed, in this example the participant changed the active voice ‘販売を開始します/hanbai o kaishi simasu’ (Version 5) into the passive voice ‘販売が開始されます/hanbai ga kaishi saremasu’ (Version 9), but reverted this edit afterwards and eventually adopted the active voice in the final version.

Table 9.13 shows the overall results of the three indicators. The scores of edit efficiency are about 1.8–1.9, and there is no significant difference between the

Table 9.13 Edit distance

	<i>Treatment</i>			<i>Control</i>		
	<i>TB</i>	<i>TD</i>	<i>Mean</i>	<i>CB</i>	<i>CD</i>	<i>Mean</i>
Edit distance (E)	276.0	288.7	282.4	263.7	252.7	258.2
— per sentence	6.2	5.9	6.1	7.9	7.3	7.6
— per correction	4.9	4.6	4.8	5.3	4.6	5.0
Accumulative E (AE)	498.7	532.3	515.5	504.7	457.3	481.0
— per sentence	11.0	10.9	11.0	15.1	13.5	14.3
— per correction	8.9	8.5	8.7	10.0	8.3	9.2
Edit efficiency (AE/E)	1.807	1.844	1.826	1.914	1.810	1.862

treatment group and the control group ($t = -0.111$, $df = 10$, $p = 0.914$). Whereas the time efficiency was improved by more than 30% with system assistance (Table 9.11), the edit efficiency showed no significant improvement. This suggests that the system greatly enhanced the speed of the operations, but did not contribute to reducing the trial-and-error edit operations. There is thus ample scope for improvement in terms of edit efficiency. To reduce unnecessary edit operations, we think that the system should provide concrete directions for rewriting in a more prescriptive manner.

9.2.3 Satisfaction

9.2.3.1 Satisfaction with the task

ASQ (satisfaction with the task) revealed no statistically significant difference between the control group and the treatment group,¹⁰ nonetheless we can see that for all three questions the mean scores of the treatment group are higher than those of the control group (Table 9.14). It is also evident that, while there are only two negative answers (i.e. Likert scale of 1–3) from the treatment group, there are seven from the control group. This suggests that participants assisted by the system were generally satisfied with the task completion.

In more detail, even for the treatment group, two participants negatively answered Question 1 ('Overall, I am satisfied with the ease of completing the tasks in this scenario'). There needs to be further investigation into the difficulty of the task and the system, which we will address in Section 9.2.4.

The mean score of Question 2 ('Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario') for the treatment group is one point higher than that for the control group. All participants in the treatment group also positively answered this question, while half of the participants in the control group negatively answered. This result is endorsed by the previous finding that the system greatly improved users' time efficiency.

Table 9.14 Result of questionnaire ASQ (satisfaction with the task)

Q.	Treatment						Mean
	TB1	TB2	TB3	TD1	TD2	TD3	
1	5	3	6	4	3	6	4.5
2	5	4	6	5	6	4	5.0
3	6	5	7	4	5	5	5.3

Q.	Control						Mean
	CB1	CB2	CB3	CD1	CD2	CD3	
1	4	2	3	5	6	3	3.8
2	3	2	3	5	6	5	4.0
3	6	6	4	5	5	3	4.8

Finally, 11 participants positively answered Question 3 ('Overall, I am satisfied with the support information (on-line help, messages, documentation) when completing the tasks'), which shows that our CL guidelines and system instructions were favourably received.

9.2.3.2 Satisfaction with the system

The SUS results (satisfaction with the system) pertain only to the treatment group (Table 9.15). To calculate overall SUS scores 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. The higher the score, the more usable the system was judged to be. The mean score is 75.0, which seems reasonably high, even if there is room for improvement.

In particular, 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,¹¹ which we further discuss in the next section.

9.2.4 User feedback

This section summarises the user feedback collected from follow-up interviews from the viewpoints of the **task**, **guideline** and **system**.

Firstly, we list the feedback about the rewriting task below.

- The longer the sentence is, the more difficult it is to comprehensively check its conformity to CL.
- They became accustomed to the task as they progressed.
- Some felt unsure about oversight of the violations (both for the treatment and the control group).

Table 9.15 Result of questionnaire SUS (satisfaction with the system)

<i>Q.</i>	<i>TB1</i>	<i>TB2</i>	<i>TB3</i>	<i>TD1</i>	<i>TD2</i>	<i>TD3</i>	<i>Mean</i>
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
SUS Score	70	68	100	54	80	78	75.0

Secondly, the interviews revealed the following difficulties of our CL guidelines and the term list.

- It was difficult to infer omitted subjects (rule R02) and objects (rule R03).
- Some felt uneasy about applying rules R16 and R20 as they are sometimes contrary to daily language use.
- Some wanted to know which rules were particularly important and should be given priority.
- Some felt unsure whether some of the words would be listed in a general Japanese–English dictionary (rule T20).
- It was difficult to detect typographical errors or missing characters (rule T22).
- Some participants in the control group found it bothersome to check the term list frequently.

The first point can be tackled by providing more detailed descriptions and model examples of the rules. It is also effective to train human writers to properly handle these omitted elements.

The second point is related to the degradation of the source readability. Humans are likely to be reluctant to write (and read) sentences that run counter to their language use. It is necessary to provide a means to maintain the source readability as much as possible, whereas the human writers themselves should make efforts to become accustomed to CL.

The third point is practically important as sometimes we have difficulties in properly dealing with all the violations that are identified. Users sometimes cannot think of alternative expressions that are appropriate for the context. A priority ranking of the rules would allow us to give up less important rules.

For the rest of the points, we would consider further mechanical support to be helpful for human writers, especially when the number of CL rules and terms that should be consulted increases.

Finally, with regard to the feedback for the system, all participants in the treatment group said that if they had not used the system, the task would have been more difficult. A noteworthy finding is that most of the participants tolerated mis-detections by the system; they said that to prevent oversight of CL violations, the detection support is helpful even if its precision was low. They would prefer higher recall at the expense of lower precision, which is contrary to what was claimed in the previous literature (Section 2.5.3.1). We assume that this is because (1) they were fully aware of the possible mis-detections of the system in advance, and (2) our system shows diagnostic messages in a less invasive manner and does not force users to adopt the system suggestions (decision-making is largely entrusted to users). Meanwhile, the following issues were also revealed in the interviews.

- Some participants failed to use the various support functions, such as suggestions of alternative expressions.
- Confidence scores were sometimes not reliable and could be confusing.

Moreover, contrary to our expectations, we found that some participants failed to correct the proscribed terms highlighted in blue, even though they recognised them, simply because they forgot what the blue highlights indicated. To make the system more effective, we need to provide more detailed user instructions and further simplify the interface.

9.2.5 Detailed observation

To ascertain the issues we faced above and find ways to improve the CL deployment, we further investigate the product and process of the participants, focusing on two aspects : (1) variance of the products and (2) transition of the task time.

9.2.5.1 Variance across participants

We noticed that even though participants referred to the same rules in the guidelines, the resultant sentences varied. To quantify the variance of text, we check the **text similarity** of the products made by different participants. Text similarity is calculated by the following formula:

$$\text{Text similarity} = \frac{\sum_{i,j}^{n_p} L_{\text{normalised}}(x_i, x_j)}{n_p C_2} \quad (9.1)$$

$$L_{\text{normalised}}(x_i, x_j) = 1 - \frac{L(x_i, x_j)}{\text{MaxLen}(x_i, x_j)} \quad (9.2)$$

Where $L(x_i, x_j)$ is the Levenshtein distance of given two sentences, and $\text{MaxLen}(x_i, x_j)$ is the length of the longest of the two sentences. The score of $L_{\text{normalised}}(x_i, x_j)$ varies between 0 and 1, and the higher the score is, the more

Table 9.16 Text similarity between participants

No.	Guideline B			Guideline D		
	Treatment	Control	Both	Treatment	Control	Both
1	0.919	0.963	0.945	0.698	0.752	0.726
2	0.549	0.869	0.707	0.882	0.881	0.878
3	0.692	0.830	0.774	0.705	0.773	0.766
4	0.940	0.915	0.922	0.780	0.869	0.825
5	0.882	0.926	0.908	0.922	0.872	0.894
6	0.891	0.911	0.905	0.954	0.738	0.842
7	0.898	0.822	0.866	0.756	0.930	0.838
8	0.865	0.827	0.847	0.965	0.966	0.962
9	0.946	0.838	0.874	0.662	0.803	0.756
10	0.814	0.768	0.798	0.739	0.753	0.761
11	0.810	0.816	0.832	0.627	0.946	0.800
12	0.713	0.769	0.748	0.921	0.398	0.671
13	0.900	0.803	0.860	0.932	0.722	0.808
14	0.947	1.000	0.966	1.000	0.944	0.943
15	0.901	0.958	0.927	0.765	0.765	0.812
16	0.842	0.691	0.729	0.579	0.707	0.671
17	0.932	0.830	0.841	0.868	0.794	0.834
18	0.915	0.855	0.893	0.790	0.782	0.780
19	0.921	0.808	0.832	0.935	0.982	0.949
20	0.510	0.526	0.559	0.543	0.454	0.482
21	0.864	0.769	0.810	0.831	0.952	0.879
22	0.903	0.910	0.898	0.895	0.938	0.921
23	0.459	0.813	0.641	0.943	0.920	0.911
24	0.707	0.564	0.641	0.499	0.911	0.683
25	0.893	0.603	0.724	0.798	0.829	0.832
26	0.905	0.905	0.921	0.808	0.954	0.860
27	0.540	0.893	0.683	0.786	0.650	0.745
28	0.795	0.817	0.766	0.889	0.767	0.819
29	0.668	0.657	0.700	0.808	0.836	0.813
30	0.917	0.844	0.881	0.839	0.849	0.832
Mean	0.815	0.817	0.813	0.804	0.815	0.810

similar the two sentences are. n_p is the number of sentences to be compared. In our experiment, three versions were created for each condition ($n_p = 3$).¹²

Table 9.16 summarises the text similarity scores for each of the 30 sentences. The scores above 0.8 are highlighted in bold and the scores below 0.6 are shaded in grey. The mean scores are all slightly above 0.8, which seems fairly high, but some sentences exhibit low scores, such as No. 20.

Let us first look at the case of high text similarity. Table 9.17 shows, for example, an original sentence and six rewritten versions. This original sentence only violates rule R09 (Do not end clauses with nouns). While participants TB1, CB1 and CB2 amended the sentence in exactly the same way—splitting the sentence and inserting the omitted element ‘*desu*’, the others made additional

Table 9.17 Example of high-text-similarity sentence

Original	天井まで火が燃え広がったら消火は困難 [R09] 素早く避難しましょう。 (<i>Tenjo made hi ga moe-hirogattara shoka wa konnan, subayaku hinan shimasho.</i>)
HT	If the fire spreads to the ceiling, it will become difficult to extinguish, so promptly evacuate.
TB1	天井まで火が燃え広がったら消火は困難です。素早く避難しましょう。 (<i>Tenjo made hi ga moe-hirogattara shoka wa konnan desu. Subayaku hinan shimasho.</i>)
TB2	天井まで火が燃え広がったら、消火は困難です。素早く避難しましょう。 (<i>Tenjo made hi ga moe-hirogattara, shoka wa konnan desu. Subayaku hinan shimasho.</i>)
TB3	天井まで火が燃え広がったら、自分で消火するのは困難です。素早く避難しましょう。 (<i>Tenjo made hi ga moe-hirogattara, jibun de shoka suru no wa konnan desu. Subayaku hinan shimasho.</i>)
CB1, CB2	天井まで火が燃え広がったら消火は困難です。素早く避難しましょう。 (<i>Tenjo made hi ga moe-hirogattara shoka wa konnan desu. Subayaku hinan shimasho.</i>)
CB3	天井まで火が燃え広がった場合は、消火は困難です。素早く避難しましょう。 (<i>Tenjo made hi ga moe-hirogatta baai wa, shoka wa konnan desu. Subayaku hinan shimasho.</i>)

small changes, such as replacing the particle ‘*tara*’ with ‘*baai wa*’, though these were not necessary operations.

We noticed that low text similarity often arose from the variant ways to apply structure-level CL rules. Table 9.18 illustrates rewritten versions of sentence number 20, the text similarity score of which is below 0.6. The original sentence violates rules T05, T22, R16 and R20. Though rules T22, R16 and R20 were correctly amended by almost all participants in a similar way, rule T05 (Ensure the relationship between the modifier and the modified is clear) was applied differently or not applied at all.

To improve the consistency between different people, we need to define how to modify the violated segments in detail, providing more rule descriptions and model examples for rewriting.

9.2.5.2 Transition of task time

Next, we observe the transition of the time spent on the task. In the follow-up interviews, participants remarked that they were becoming accustomed to the rewriting tasks as they advanced. To roughly estimate how much their time

Table 9.18 Example of low-text-similarity sentence

Original	経済的支援を必要とする市立小・中学校[R20]に通う児童・生徒のいる家庭の給食費や学用品など [R16] のの [T22]補助を行います。 (<i>Keizai-teki shien o hitsuyo to suru shiritsu-sho · chugakko ni kayou jido · seito no iru katei no kyushoku-hi ya gakuyo-hin nado no no hojo o okonai-masu.</i>)
HT	We will give assistance such as food stipends and school materials, etc., for families with children and students in the city elementary and middle schools who require economic support.
TD1	経済的な支援を必要とする市立小・中学校に通う児童・生徒のいる家庭の給食費や学用品の補助を行います。 (<i>Keizai-teki shien o hitsuyo to suru shiritsu-sho · chugakko ni kayou jido · seito no iru katei no kyushoku-hi ya gakuyo-hin no hojo o okonai-masu.</i>)
TD2	市は、市立の小学校・中学校に通う児童・生徒のいる、経済的な支援を必要とする家庭の給食費や学用品の補助を行います。 (<i>Shi wa, shiritsu no shogakko · chugakko ni kayou jido · seito no iru, Keizai-teki na shien o hitsuyo to suru katei no kyushoku-hi ya gakuyo-hin no hojo o okonai-masu.</i>)
TD3	市立小学校・中学校に通う児童・生徒を持つ、経済的支援の必要な家庭に、給食費や学用品の補助をします。 (<i>Shiritsu shogakko · chugakko ni kayou jido · seito o motsu, keizai-teki shien no hitsuyo na katei ni, kyushoku-hi ya gakuyo-hin no hojo o shimasu.</i>)
CD1	給食費や学用品の補助は経済的支援を必要とする市立小学校・市立中学校に通う児童・生徒がいる家庭に行います。 (<i>Kyushoku-hi ya gakuyo-hin no hojo wa keizai-teki shien o hitsuyo to suru shiritsu shogakko · shiritsu chugakko ni kayou jido · seito ga iru katei ni okonai-masu.</i>)
CD2	経済的支援を必要とし、市立小学校・市立中学校に通う児童・生徒のいる家庭の給食費や学用品の補助を行います。 (<i>Keizai-teki shien o hitsuyo to shi, shiritsu shogakko · shiritsu chugakko ni kayou jido · seito no iru katei no kyushoku-hi ya gakuyo-hin no hojo o okonai-masu.</i>)
CD3	市立小学校・市立中学校に通う、経済的支援を必要とする児童・生徒のいる家庭の給食費や学用品の補助を行います。 (<i>Shiritsu shogakko · shiritsu chugakko ni kayou, keizai-teki shien o hitsuyo to shuru jido · seito no iru katei no kyushoku-hi ya gakuyo-hin no hojo o okonai-masu.</i>)

efficiency increased in accordance with the progression of the task, we compare the average elapsed time of the three sets.

Tables 9.19 and 9.20 show the transitions of the average task time from Set 1 to Set 3 for each condition. The former table presents the average time per sentence, while the latter presents the average time per character, which normalises the length of the sentences as the sentence length may influence the task time. The percentage value in the tables indicates the ratio of the task time of the set against that of Set 1.

Table 9.19 Task time transition (time per sentence, in seconds)

	<i>B</i>			<i>D</i>		
	<i>Treatment</i>	<i>Control</i>	<i>Mean</i>	<i>Treatment</i>	<i>Control</i>	<i>Mean</i>
Set 1	100.9	148.3	124.6	81.0	107.7	94.3
Set 2	78.0 (77.3%)	123.5 (83.3%)	100.7 (80.8%)	81.4 (100.5%)	100.5 (93.3%)	91.0 (96.4%)
Set 3	61.6 (61.1%)	102.7 (69.3%)	82.2 (66.0%)	58.2 (71.9%)	76.2 (70.7%)	67.2 (71.2%)

Table 9.20 Task time transition (time per character, in seconds)

	<i>B</i>			<i>D</i>		
	<i>Treatment</i>	<i>Control</i>	<i>Mean</i>	<i>Treatment</i>	<i>Control</i>	<i>Mean</i>
Set 1	2.6	3.6	3.1	1.8	2.5	2.1
Set 2	1.7 (64.0%)	2.9 (80.2%)	2.3 (73.3%)	1.8 (98.2%)	2.3 (92.2%)	2.0 (94.7%)
Set 3	1.5 (58.3%)	2.7 (76.2%)	2.1 (68.6%)	1.6 (85.9%)	2.1 (84.0%)	1.8 (84.8%)

Comparing the results of Sets 1–3 shows that the task time decreased in accordance with the progression of the task. Particularly, when guideline B is used, the average time (both per sentence and per character) of Set 3 is lower than that of Set 1 by about 30%. This clearly indicates that participants became accustomed to the task at a later stage, implying further increased time efficiency of the CL deployment. It is therefore necessary to investigate to what extent users can learn to efficiently use CL and what kind of training is necessary to reach that point.

9.3 Summary

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 the term list. 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**).

These results above positively answered the research question **RQ-A2** (Is our system usable for non-professional writers?), demonstrating that the system can particularly enhance the efficiency of CL authoring by non-professional writers. As for the research question **RQ-A3** (Does the use of the system help improve the quality of ST and TT?), though the contributions of the system itself to the MT and ST quality were not significant, a comparison of the original ST and controlled versions of ST revealed that CL authoring (even without the help of the system) was shown to be effective in improving MT quality without degrading ST quality. The MT evaluation results of the oracle ST suggested that, however, there is still much 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 thereby 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. Finally, we plan to investigate usability in the ‘drafting-from-scratch’ setting, and to establish how far total translation productivity can be increased when we integrate MT dictionaries with controlled writing.

Notes

- 1 On the other hand, in the ‘drafting-from-scratch’ setting, outputs would vary greatly depending on the participants, which makes it difficult to analyse them consistently.
- 2 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 evaluation.
- 3 Since the questionnaire is originally in English, we translated it into Japanese. We also changed the phrase ‘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.
- 4 We also translated this into Japanese.
- 5 Participants assigned to the control group answered only the ASQ.
- 6 CL-compliant versions are ‘市立小学校・市立中学校/shiritsu shogakko・shiritsu chugakko’ (municipal elementary school and municipal junior high school) and ‘月曜日・水曜日・金曜日/getsu-yobi・sui-yobi・kin-yobi’ (Monday, Wednesday and Friday).
- 7 The inter-rater agreement scores are 0.521 for fluency and 0.487 for adequacy based on the Krippendorff’s α (Krippendorff, 2004). Though these scores are not high enough to let us draw conclusion from the results, we assume that some general tendencies can still be observed.
- 8 Note that the treatment group amended more violations than the control group.
- 9 The inter-rater agreement score is 0.175 based on the Krippendorff’s α . Though this is quite low, we can still observe some general tendencies.

- 10 The results of an independent samples t-test are as follows: Question 1 ($t = -0.810$, $df = 10$, $p = 0.437$); Question 2 ($t = -1.369$, $df = 10$, $p = 0.201$); Question 3 ($t = -0.785$, $df = 10$, $p = 0.451$).
- 11 ISO (2011) defines learnability as the ‘degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use.’
- 12 Therefore, for both the treatment and control conditions, a total of six versions were created ($n_p = 6$).

Part IV

Conclusion

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

10 Research findings and outlook

In this book, we established frameworks and a system environment to help non-professional writers produce well-organised municipal documents that are both machine-translatable and human-readable. The key concept, controlled document authoring, encompasses a wide range of textual control mechanisms, including document formalisation, CL and terminology management. Primarily focusing on the task of Japanese-to-English translation of municipal procedural documents, we have addressed the above aspects of controlled document authoring. The purpose of this chapter is to synthesise what we have carried out and revealed.

Closing this book, this chapter begins with a summary of the research findings of Chapters 3–9 in relation to the research questions (Section 10.1). It then discusses the main contributions of this study to the MT and CL research fields, and examines the methodological perspective underlying the book (Section 10.2). Finally, it both details remaining tasks and outlines future opportunities for work surrounding this topic (Section 10.3).

10.1 Findings and achievements of the study

This research was prompted by the observation that municipal documents are often ill-structured and that low-quality MT outputs are used to produce multilingual versions of these documents. Given that many municipalities in Japan cannot afford to hire translators and post-editors, we have tackled this issue by implementing upstream controlled authoring of document multilingualisation. The principal research question to be answered was as follows:

RQ: Can controlled document authoring help non-professional writers to create well-structured source texts that are machine-translatable and human-readable?

Answering this question demands not just proposing effective frameworks of controlled authoring, but also developing an authoring support system to help writers implement the frameworks in actual work scenarios.

In Part II, Chapters 3–5, we established the frameworks of controlled authoring at the different textual levels, namely, document-level, sentence-level and terminology-level.

At the document-level, Chapter 3 answered the research question **RQ-F1** (Can municipal documents be well modelled?) by comprehensively collecting and categorising functional document elements and mapping them onto the well-established document structure DITA. The results demonstrated that municipal procedures were well modelled using DITA Task topic.

At the sentence-level, Chapter 4 addressed the research question **RQ-F2** (To what extent can a Japanese CL improve the quality of ST and TT?) by formulating Japanese CL rules intended for municipal documents and evaluating the effectiveness of individual rules on both MT quality (using four different MT systems) and source readability. The results of human-evaluation experiment led to two major findings:

- If we select optimal rules tailored to particular MT systems, we can obtain a great improvement in MT quality.
- Most of our CL rules improved, or at least retained, the source readability.

Based on these results, we selected effective rules for intended MT systems and compiled two sets of CL guidelines, supplying the description and rewriting examples of each rule to facilitate users' understanding of the rule.

In Chapter 5, we undertook the research question **RQ-F3** (Can the combination of controlled language and document structure further improve the TT quality without degrading ST quality?), which requires unification of the document-level DITA structure and sentence-level CL. We defined four document-element-dependent CL rules for both Japanese and English sides. To resolve the incompatibilities between ST and TT quality when MT is used, we further incorporated an internal pre-translation processing mechanism in the MT work flow. The diagnostic observation of outputs generated by four MT systems showed that the pre-translation processing of ST is particularly effective in the case of RBMT, while further mechanisms, such as post-translation processing are needed for SMT to produce desirable outputs.

At the terminology-level, Chapter 6 answered the research question **RQ-F4** (Can municipal terms be comprehensively captured and well controlled?) by constructing and evaluating Japanese–English controlled terminologies. We first manually collected bilingual term pairs from a parallel corpus of municipal domain and defined preferred and proscribed terms for both Japanese and English. To evaluate the constructed terminologies in terms of coverage and quality of control, we statistically estimated the population size of terms and concepts in the domain, and assessed the current status of terminologies compared with the population status. The results indicated that our terminologies covered about 45–65% of the terms and 50–65% of the concepts in the municipal domain, and were well controlled.

Based on the frameworks established, in Part III, Chapters 7–9, we detailed the development and evaluation of a controlled authoring support system named MuTUAL.

Chapter 7 described the design principles and the overall module organisation of MuTUAL. It materialised the frameworks proposed in the previous chapters: the topic template is based on the DITA structure; the CL authoring assistant implements global/local CL rules and controlled Japanese terminology; the MT systems (RBMT and SMT) are customised with Japanese–English terminologies; and the automatic pre-/post-translation processing is integrated. The most innovative feature of the system is that it invokes different CL authoring assistants and pre-/post-translation processing tailored to each functional element of the topic template. This consequently enables MT to produce outputs appropriate to the functional element within the document.

Assuming that (re)writing in accordance with the CL rules and controlled terminology is especially difficult for non-professional writers, we evaluated the CL authoring module of MuTUAL. Evaluation proceeded in two phases: performance evaluation and usability evaluation. In Chapter 8, we first gauged the precision and recall performance of a CL detection component using a benchmarking test set in which CL violations are manually annotated. The answer for the research question **RQ-A1** (How accurately does the system detect CL rule violations in text?) is that 20 of the 30 implemented rules achieved high benchmark scores (both precision and recall higher than 0.7), with 13 rules obtaining perfect scores. While this is an encouraging result for future implementation of CL, the benchmarking evaluation also disclosed the issue of low precision which we need to resolve or alleviate from the point of view of usability.

In Chapter 9, we conducted a usability evaluation to address the two research questions: **RQ-A2** (Is our system usable for non-professional writers?) and **RQ-A3** (Does the use of the system help improve the quality of ST and TT?). At this stage, we decided not to implement low-precision (precision below 0.4) rules and also incorporated ‘confidence scores’ to inform users of how accurately the system is supposed to detect CL violations. Adhering to the usability definition of ISO (2010), we designed and implemented a usability test to assess the effectiveness, efficiency and user satisfaction of the CL authoring assistant, comparing the two conditions with and without the help of the system. The results led us to the following three major points:

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

The MT evaluation results also revealed that user-generated STs were generally more suitable for MT than the original STs regardless of the support from the system, but were less suited to MT compared to oracle STs that perfectly complied with our CL rules. This demonstrates that controlled authoring is indeed effective

in improving MT results if we can correctly and comprehensively follow the rules, but in terms of that, there is still much room for improving methods for assisting non-professional writers.

Looking back at the principal research question, we can now conclude that controlled authoring helps create well-organised municipal procedural documents that are more machine-translatable than uncontrolled ones, without degrading ST quality. Although further improvement is needed, especially in the phase of deployment of controlled authoring by non-professional writers, the results were generally positive. This opens up the promising prospect of practical deployment of controlled document authoring for contextual MT in real-world scenarios.

10.2 Contributions and implications

10.2.1 Document-level framework for extending MT and CL research

The most noteworthy contribution of this study to the fields of MT and CL is that we examined the feasibility of connecting functional document elements with sentence-internal features. Introducing document structure contributes not only to creating well-organised documents, but also to guiding MT to produce appropriate results within the document by contextualising the CL rules and pre-translation transformation rules to particular document elements.

It is not that MT and CL research communities did not recognise the necessity of exploiting document-level features in dealing with sentence-level phenomena prior to this work. For example, much effort has been devoted to research on anaphora resolution for MT (Mitkov, 1999). More recent work includes the *Workshop on Discourse in Machine Translation*, which has been held since 2013 (DiscoMT, 2013, 2015) and focuses on MT research that makes use of textual properties that go beyond individual sentences. In previous CL studies, as summarised in Section 2.3.3, the need for document-level CL was duly advocated (e.g. Hartley and Paris, 2001; Bernth, 2006; Hartley, 2010).

Nevertheless, relatively few studies have explicitly tackled document-level features. There are several challenges which we assume could hamper document-related work.¹ Firstly, the concept of ‘document’ itself is vaguely defined. There appears to be no clear consensus about the terms (or words) ‘document’ and ‘discourse’; they tend simply to be used to mean ‘beyond-sentence’. As DiscoMT (2015, p.iii) summarised, textual properties beyond individual sentences involve:

- ‘document-wide properties, such as style, register, reading level and genre;
- patterns of topical or functional sub-structure;
- patterns of discourse coherence, as realized through explicit and/or implicit relations between sentences, clauses or referring forms;
- anaphoric and elliptic expressions, in which speakers exploit the previous discourse context to convey subsequent information very succinctly’

Although this list is useful in order to grasp the broad picture of document/discourse-level properties, it is difficult to see the relationship between them. What is first needed is to clarify what properties should be captured and from what viewpoint. In this study, we took a genre perspective, in which we view a document as a complete text that has a certain communicative goal, and textual elements in the document have a particular functional role in relation to the larger parts of text. This is different from perspectives that focus on inter-sentence relationships such as anaphora and cause–effect relationships.²

Focusing on the CL studies, we notice that document-level features are not well examined. The document elements are identified in rather coarse-grained terms, such as broad distinctions of ‘procedures’, ‘descriptions’ and ‘warnings/cautions’ (see also Section 2.3.3). As we have illustrated in this book, however, fine-grained functional roles of document elements are closely tied to particular linguistic patterns. For instance, elements that specify procedural steps require imperative forms in English, while elements that specify the eligibility conditions when certain events occur require conditional clauses like ‘when you’ and ‘if you’ (see Section 5.1 for detail). To make the most of document-level properties, in-depth analysis and explication of the properties is indispensable. The genre analysis (functional structure analysis) conducted in Section 3.1 is one example of such analysis.

More practically, the lack of resources and document-aware evaluation methods encumbers document-related studies. Researchers might be reluctant to undertake work that is difficult to benchmark using conventional evaluation metrics such as BLEU. As Guzmán et al. (2014, p.688) noted, ‘current automatic evaluation metrics such as BLEU are inadequate to capture discourse-related aspects of translation quality’ and ‘there is consensus that discourse-informed MT evaluation metrics are needed in order to advance research in this direction’. On the one hand, nowadays document-level automatic MT evaluation metrics are being proposed more frequently (e.g. Giménez et al., 2010; Guzmán et al., 2014; Gong et al., 2015). On the other hand, task-based human evaluations have been drawing more attention (Voss and Tate, 2006; Berka et al., 2011; Matsuzaki et al., 2015, 2016), as they can be useful to capture document-level MT quality. Looking further into related fields such as technical writing, document design, NLG and translation studies, we can gain an insight into document quality evaluation (see also Section 2.1.3). Establishing document-level evaluation methodologies is a crucial agenda for the MT and CL research communities.

10.2.2 *From descriptive to prescriptive*

The notion of *controlled* authoring inevitably requires a prescriptive perspective; it defines preferred usage of document structure, language and terminology. The methodological steps underlying our studies in Chapters 3–6 are as follows: *describe* the range of (document, linguistic or terminological) items actually appearing in the given collection of documents, and then *prescribe* preferred usage based on the descriptions. This ‘from descriptive to prescriptive’ approach itself

is not new and was/will be applied in many research fields. What we would emphasise are two fundamental questions that naturally arise, but were rarely addressed explicitly:

- Do the items identified in the given data properly reflect and substantially cover the whole range of items that we might encounter in the given domain?
- How to decide the preferred and proscribed usage among several (sometimes numerous) other possibilities?

The first question pertains to descriptive methods in general. In the case of our study, we investigated 123 municipal procedural documents for collecting functional document elements; we used 100 sentences and three MT systems to identify textual and linguistic features that are putatively effective for machine-translatability; we exploited 15391 aligned sentence pairs to extract municipal terms. We are, however, still unsure as to what extent we should continue to observe and investigate other data/phenomena. As we pointed out in Section 2.4.4, the sufficiency of items have not yet been examined in much detail due to the difficulty in obtaining a population range of items. Yet, we can estimate it as Kageura and Kikui (2006) did. Since this book only examined the coverage of terminologies, future work should further scrutinise the status of descriptively obtained results.

Regarding the second question, we took different measures in each study: for the task of modelling *well-organised* documents, we employed DITA, an external standard that is widely accepted; for the task of specifying *effective* CL rules, we conducted human evaluation to assess the impact of the rules on MT quality and ST readability; for the task of defining *preferred* terms, we referred to dictionary evidence, frequency evidence and typological preference. We should keep it in mind that DITA is not the only standard, that machine-translatability and source readability might not cover all requirements of CL rules, and that other external materials can be referenced for controlling terminologies. It is essential to create a clear standard for prescribing observed phenomena, taking into account the foreseen applications.

10.3 Remaining issues and future directions

Outstanding questions and tasks were mentioned in the last section (summary section) of each chapter. We describe below the major work still to be addressed regarding our research, and sketch out future plans.

Firstly, this book focused on the Japanese-to-English translation of municipal-life information, more specifically, procedural documents. We will extend our framework to other text domains (types) and target languages. Although what we have created in the study is dependent on the particular domain and languages at hand, the methodologies proposed and adopted are domain- and language-independent. For document formalisation, while municipal procedures have proved to be well-structured when using the DITA Task topic,

other types of municipal documents remain unexamined. We intend to specialise the Concept topic and the Reference topic of DITA by further investigating document elements of other documents. For CL formulation, we will extend target languages to, for example, Chinese and Korean. We will first examine the effectiveness of our 60 Japanese CL rules on MT quality and additionally investigate other linguistic patterns by using the rewriting-trial-based protocol which we proposed in Section 4.1.2. For controlled terminology construction, we will collect and validate terms from multiple languages. The list of Japanese and English terms created in Section 6.1 can be utilised to extract terms from other language corpora.

Another important extension of this work is conducting a field study to assess the feasibility and effectiveness of implementing controlled authoring in actual work scenarios. The results of the usability study demonstrated that the core module of MuTUAL—CL authoring assistant—greatly facilitated the rewriting task of non-professional writers and was favourably accepted by them. This encourages us to further develop and evaluate the full range of functions of MuTUAL, including document structuring modules. The evaluation focus must be more on document-level quality. At this stage, task-based, reader-focused methods can be beneficial to investigate, for example, whether readers can successfully perform tasks indicated by the document. Furthermore, in the course of product evaluation of the MT raw outputs, we have unsurprisingly been confronted with the limitation that controlled authoring is still not sufficient for constantly achieving publishable MT quality. Even if STs perfectly conform to our CL and controlled terminology, the task of Japanese-to-English MT is difficult. Therefore, in practice, post-editing (PE) remains to be an essential step to correctly communicate the necessary information to intended readers. Considering the limited resources of Japanese municipalities, we need to seek a solution to minimise PE cost by providing simple tools to amend raw MT outputs. MuTUAL can seamlessly incorporate a PE tool in the main interface and we suppose that functions and interfaces developed for the CL authoring assistant can be ported to the PE tool.

Finally, it should be noted that the functional document elements introduced in this study are not necessarily instantiated in linear textual forms. Municipal documents do contain tables, charts, illustrations and photos together with (or instead of) text. For example, documents provided by CLAIR frequently use tables to compactly present information. Importantly, the segments in these tables are often embodied in fragments of text—simple words and phrases. Technical illustrations are also included to facilitate readers' understanding of the documents, which are, in nature, language independent. Though the combination of textual and non-textual forms is beyond the scope of the present study, as MT deals only with texts, we underscore that the notion of functional document elements is applicable to all forms of information instantiation. Our next question is: What form is most appropriate to instantiate each element? The effective integration of various forms of information (including visual materials) has been tackled in many fields such as technical communication, document design and

information architecture. At this stage, we should reconsider the role of MT in streamlining the document production work flow in multilingual settings. In future work, we will strive to address this issue with a view towards bettering human communication through documents.

Notes

- 1 An extensive review and discussion of discourse-level SMT is provided by Hardmeier (2014).
- 2 A notable framework to analyse inter-sentence relationships is Rhetorical Structure Theory (RST) (Mann and Thompson, 1988), which enables us to determine a functional role of each sentence/clause in relation to other sentences/clauses.

Appendices

A. CL guidelines (Japanese original version)

T01: 一文はできる限り50文字以内におさめてください。

補足説明:

一文に多くの内容を詰め込みすぎない。一文が複数の述語を含む場合は、なるべく短文に分割する。

書き換え例:

(1) 文を分割する

住民登録とは～～～であり、入国後に～～～しなければなりません。

⇒ 住民登録とは～～～です。入国後に～～～しなければなりません。

(2) 箇条書きを使う

日本で働くには、～～～、～～～、～～～の条件を満たす必要があります。

⇒ 日本で働くには、以下の条件を満たす必要があります。

- ～～～
- ～～～
- ～～～

T02: 箇条書きで書くときは、列挙項目の前後の文を完結させてください。

補足説明:

文の途中に、箇条書きをはさまない。箇条書きの導入文を完結させてから、項目を列挙する。

書き換え例:

日本で働くには、

- ~~~
- ~~~
- ~~~

の条件を満たす必要があります。

⇒ 日本で働くには、以下の条件を満たす必要があります。

- ~~~
- ~~~
- ~~~

T05: 修飾語と被修飾語の関係を明確にしてください。

補足説明:

修飾語がどの要素に係るか、明確に示す。

書き換え例:

(1) 語を補う・語順を変える

大きな庭の木の下

⇒ 庭に生えている大きな木の下

(2) 読点を追加する

大きな庭の木の下

⇒ 大きな庭の、木の下

T06: 「が」を使って文をつなげるのは、「しかし」の意味を持つ場合（逆接用法）だけにしてください。

補足説明:

「が」は「しかし」（but）の意味を持つ逆接用法のみで使用する。順接（and）の場合は、「が」を使用しない。

書き換え例:

(1) 文を分割する

市役所は平日に開いていますが、第二土曜日にも開いています。

⇒ 市役所は平日に開いています。また第二土曜日にも開いています。

(2) 順接に変更する

市役所は平日に開いていますが、第二土曜日にも開いています。

⇒ 市役所は平日に開いており、第二土曜日にも開いています。

T07: 「ので」の意味で「ため」を使わないでください。「ので」を使ってください。

補足説明:

理由を示す「ため」(because)は「ので」に変更する。目的を示す「ため」(in order to)は「ので」に変更しない。

書き換え例:

生活用品が足りないため、物資の支援を要請する。

⇒ 生活用品が足りないので、物資の支援を要請する。

T09: 1文に複数の否定表現を使わないでください。

補足説明:

二重否定(複数の否定表現)を避け、なるべく肯定文を使用する。

書き換え例:

(1) 対偶をとる

登録しないと使用できません。

⇒ 使用するためには登録する必要があります。

(2) 肯定文に変更する

休日も使用できないわけではありません。

⇒ 休日も使用できます。

T10: 可能や尊敬の意味で「～れる」「～られる」を使わないでください。

補足説明:

「～れる」「～られる」は、受け身の意味で使用する。曖昧性を避けるため、尊敬や可能の意味では使わない。

書き換え例:

(1) (尊敬用法) 通常表現に変更する

登録された方のみ使用できます。

⇒ 登録した方のみ使用できます。

(2) (可能用法) 「～できる」の形を使用する

当施設は50人まで受け入れられます。

⇒ 当施設は50人まで受け入れることができます。

T11: 複数の意味に解釈できる言葉ではなく、なるべく明確な意味を持つ言葉を使ってください。

補足説明:

「とる」「出す」「動く」「もつ」などの動詞は、解釈の幅が広いため、文脈に合わせて「採用する」「提出する」「発車する」「担当する」など、なるべく明確な意味を持つ動詞を使う。ただし、不必要に難解な語彙を使用しない。

書き換え例:

(1) 語を変更する

列車が動くまで待機してください。

⇒ 列車が発車するまで待機してください。

(2) 語を変更する

今年は愛知の本社で新人をとる予定です。

⇒ 今年は愛知の本社で新人を採用する予定です。

T13: 「～という」表現はなるべく省いてください。

補足説明:

特に意味を持たない「～という」表現を避ける。ただし、「のんほいパークという名の公園」のように、「という」を削除できない場合もある。

書き換え例:

まだ更新するという段階ではありません。

⇒ まだ更新する段階ではありません。

T16: 「思われる」「考えられる」は必要なとき以外は省いてください。

補足説明:

「思われる」「考えられる」は事実を伝達すべき行政文書ではなるべく避ける。削除する、もしくは他の明確な表現に置き替える。

書き換え例:

(1) 削除する

避難する方がよいと思われる。

⇒ 避難する方がよい。

(2) 明確な表現を使う

消防活動の目標としては次のように考えられる。

⇒ 消防活動の目標としては次の通りです。

T18: 「たり」は単独で使わず、「～たり～たり」と並列で使ってください。

補足説明:

助詞「たり」を使うときは、列挙項目の全てに「たり」をつける。

書き換え例:

(1) 「たり」を追加する

地震の時に窓ガラスが割れたり、食器棚が倒れると大変危険です。

⇒ 地震の時に窓ガラスが割れたり、食器棚が倒れたりすると大変危険です。

(2) 他の並列表現を使う

地域で話し合ったり、学習会を開催することをお勧めします。

⇒ 地域での話し合いや学習会の開催をお勧めします。

T20: なるべく標準的な和英辞典に載っている語を使ってください。

補足説明:

行政文書は、幅広い層の地域住民が理解できる語、もしくは日本語非母語話者が一般的な辞書を使いながら理解できる語で執筆する。ただし、地域名や制度名などの固有名詞は許容する。

書き換え例:

各機関は、定期的又は随時に通信訓練を実施し、発災時に備えるよう努力する。

⇒ 各機関は、定期的又は随時に通信訓練を実施し、災害発生時に備えるよう努力する。

T21: 複合サ変名詞＋「する」の形を使わないでください。

補足説明:

サ変名詞とは、「する」を後ろにつけて動詞化する名詞を指す。例えば、サ変名詞「作業」は「する」をつけると動詞「作業する」になる。「受信転送する」のように、サ変名詞を複数組み合わせる動詞化する表現は避ける。

書き換え例:

(1) 分解する

文房具を購入配分する。

⇒ 文房具を購入し、配分する。

(2) 中黒を使用する

メールを受信転送する。

⇒ メールを受信・転送する。

T22: 誤字、脱字がないように注意してください。また、同音異義語や助詞の抜けにも注意してください。

補足説明:

日本語入力の誤変換や、コピー&ペーストに伴う脱字に特に注意する。

書き換え例:

施設の会館時間は、ホームページご覧ください。

⇒ 施設の開館時間は、ホームページをご覧ください。

R02: 主語はなるべく省略しないでください。

補足説明:

日本語は主語が省略されやすい。文脈や常識から無理なく判断できる範囲で、なるべく主語を明示する。

書き換え例:

会場では、市からの説明にメモを取るなど、真剣に耳を傾けているとともに、議題について活発に議論しました。

⇒ 会場では、市民は、市からの説明にメモを取るなど、真剣に耳を傾けているとともに、議題について活発に議論しました。

R03: 目的語はなるべく省略しないでください。

補足説明:

文脈や常識から無理なく判断できる範囲で、なるべく目的語を明示する。多少の冗長さは許容する。

書き換え例:

外国語での情報提供を行っている目的は、外国の方が読めるようにするためです。

⇒ 外国語での文書提供を行っている目的は、外国の方が文書を読めるようにするためです。

R04: 並列要素を読点でつなげないでください。

補足説明:

読点「、」は、節の区切りでも使用されるので、並列要素をつなぐ際は、他の接続詞「と」「や」「か」「また」や中黒「・」を使用する。

書き換え例:

こうした被害の多くは、平面計画、施工になんらかの弱点がある建物に顕著にみられます。

⇒ こうした被害の多くは、平面計画や施工になんらかの弱点がある建物に顕著にみられます。

R05: 目的語につける助詞は、「～が」ではなく「～を」を使ってください。

補足説明:

目的語に対する助詞は「を」を使用する。

書き換え例:

適切なサービスが選択できるようにコーディネート体制の確立に努めます。

⇒ 適切なサービスを選択できるようにコーディネート体制の確立に努めます。

R07: 「～てくる」「～ていく」を使わないでください。

補足説明:

「～てくる」「～ていく」は多くの場合、削除しても文意が大きく変わらない。不要な「～てくる」「～ていく」は削除する。

書き換え例:

今後、県など関係機関と連携しながらパトロールを実施していく予定です。

⇒ 今後、県など関係機関と連携しながらパトロールを実施する予定です。

R08: 副詞節を主語と述語の途中になるべく挿入しないでください。

補足説明:

長い副詞節（動詞を修飾する要素）を主語と述語の間に挟むと、文の構造が複雑になる。なるべく副詞節は、文頭に出すとよい。

書き換え例:

イベントは、雨や強風の日を除き、市民公園で開催されます。

⇒ 雨や強風の日を除き、イベントは市民公園で開催されます。

R09: 体言止めは使わないでください。

補足説明:

体言止め（節や文の末尾が名詞で終わる形）は避け、節や文を完結させる。

書き換え例:

(1) 節を完結させる

すばやく避難所に移動、情報収集に努めましょう。

⇒ すばやく避難所に移動し、情報収集に努めましょう。

(2) 文を完結させる

準備するものは、パスポート・印鑑・在留カード。

⇒ 準備するものは、パスポート・印鑑・在留カードです。

R10: サ変名詞＋「です」の形をなるべく使わないでください。

補足説明:

サ変名詞とは、「する」を後ろにつけて動詞化する名詞を指す。例えば、サ変名詞「作業」は「する」をつけると動詞「作業する」になる。「誕生です」「締め切りです」のように、サ変名詞＋「です」の表現はなるべく避ける。

書き換え例:

(1) 動詞化する

市民の憩いの場の誕生です。

⇒ 市民の憩いの場が誕生しました。

(2) 動詞化する

応募の締め切りです。

⇒ 応募を締め切りました。

R11: 「～しか～ない」の形を使わないでください。

補足説明:

「～しか～ない」の形は、「～のみ」「～だけ」に変更する。

書き換え例:

(1) 「～のみ」を使う

指定のゴミ袋しか使えません。

⇒ 指定のゴミ袋のみ使えます。

(2) 「～だけ」を使う

大会には6年生しか参加できません。

⇒ 大会には6年生だけ参加できます。

R12: 動詞＋「ように」の形を使わないでください。

補足説明:

不要な「ように」は削除する。

書き換え例:

こまめな休憩をとるようにしましょう。

⇒ こまめな休憩をとりましょう。

R16: 「など」「等」をなるべく使わないでください。

補足説明:

「など」「等」は、なるべく削除する。

書き換え例:

身分証明書等を持参してください。

⇒ 身分証明書を持参してください。

R17: 授受動詞「～あげる」「～くれる」は使わないでください。

補足説明:

「～あげる」「～くれる」といった授受動詞は、冗長なので削除する。

書き換え例:

本人の代わりに書類を提出してあげてください。

⇒ 本人の代わりに書類を提出してください。

R18: 冗長な表現をなるべく使わないでください。

補足説明:

冗長な表現はなるべく削除する。特に、「ものとする」「ものとなる」「ものにする」「ものになる」「こととする」「こととなる」「ことにする」「ことになる」を削除対象とする。

書き換え例:

(1) 「こととする」を削除する

次の各号に掲げる列車の運転取扱いを実施することとする。

⇒ 次の各号に掲げる列車の運転取扱いを実施する。

(2) 「ものとなる」を削除する

この公園は災害時に市が避難所として指定するものとなります。

⇒ この公園は災害時に市が避難所として指定します。

R19: 長い複合名詞（連続した3語以上の名詞の連なり）をなるべく使わないでください。

補足説明:

連続した3語以上の名詞が連なる複合名詞をなるべく避ける。ただし、「印鑑登録証明書」などの固有表現（ひとまとまりの定型的な表現）は書き換えない。

書き換え例:

(1) 複合名詞を展開する

図書館資料や市民提供資料を閲覧できます。

⇒ 図書館資料や市民が提供した資料を閲覧できます。

(2) 複合名詞を展開する

主要拠点間情報伝達を強化する。

⇒ 主要な拠点間の情報伝達を強化する。

R20: 項目を列挙する際、項目の一部を省略しないでください。

補足説明:

中黒で複数の項目を並列する際、各項目の表現の一部を省略しない。

書き換え例:

新宿・渋谷・豊島区にお住まいの方

⇒ 新宿区・渋谷区・豊島区にお住まいの方

R25: 単位表現には「～につき」「～あたり」を使ってください。

補足説明:

単位表現は、「1回1000円」のように簡略化せずに、「～につき」「～あたり」を明示する。

書き換え例:

- (1) 「～につき」を使う
子ども1人1000円です。
⇒ 子ども1人につき1000円です。
 - (2) 「～あたり」を使う
1時間1000円かかります。
⇒ 1時間あたり1000円かかります。
-

R26: 節の区切りに「～て」をなるべく使わないでください。

補足説明:

節の区切りに「～て」の形はなるべく使わず、「～し」の形を使う。

書き換え例:

- 母子健康手帳及び個人通知を持参して、健康診断を受診してください。
- ⇒ 母子健康手帳及び個人通知を持参し、健康診断を受診してください。
-

R27: 条件ifの用法で「～すると」を使わないでください。

補足説明:

「と」は等位接続詞として使われるので、条件節では「～と」は避け、「～たら」「～れば」「～場合」を使う。

書き換え例:

- (1) 「～れば」を使う
届出を出すと手当を受け取れます。
⇒ 届出を出せば手当を受け取れます。
- (2) 「～場合」を使う
期限内に手続きをしないと、手当を受けることができなくなることがあります。
⇒ 期限内に手続きをしない場合、手当を受けることができなくなることがあります。

R33: 動詞はなるべく平仮名ではなく漢字で表記してください。

補足説明:

曖昧性を避けるため、動詞はなるべく平仮名ではなく漢字で表記する。ただし、常用漢字の範囲外の漢字や普段目にしない漢字は使わない。また固有名詞の表記は変更しない。

書き換え例:

(1) 漢字を使用する

仕事をやめずに日本に滞在する。

⇒ 仕事を辞めずに日本に滞在する。

(2) 漢字を使用する

ぜひ、みなさんでいきましょう。

⇒ ぜひ、みなさんで行きましょう。

R34: 文頭の記号は削除してください。

補足説明:

「○」「●」「▲」「☆」「＊」などの記号を文頭につけない。なお簡条書きの先頭の記号は中黒のみ許容する。

書き換え例:

(1) 記号を削除する

◎自宅の電話より、公衆電話の方が比較的つながります。

⇒ 自宅の電話より、公衆電話の方が比較的つながります。

(2) 中黒を使用する（簡条書きの場合）

持ち物：

○ 運転免許証

○ 申請書

⇒ 持ち物：

・ 運転免許証

・ 申請書

R35: 機種依存文字を使わないでください。

補足説明:

特定の環境でしか正常に表示されない機種依存文字を使わない。

書き換え例:

敷地面積は14 m^2 です。

⇒ 敷地面積は14平方メートルです。

R37: カギ括弧を使って単語を強調しないでください。

補足説明:

カギ括弧は、引用やタイトル（名称）を示す場合に使用し、単なる強調のためには使用しない。

書き換え例:

(1) カギ括弧を削除する

環境に負荷をかけず、自分のペースで移動できる「自転車」を取り入れたライフスタイルを提案するイベントを開催します。

⇒ 環境に負荷をかけず、自分のペースで移動できる自転車を取り入れたライフスタイルを提案するイベントを開催します。

(2) カギ括弧を削除する

市のカレンダーは「無料で」お持ち帰りください。

⇒ 市のカレンダーは無料でお持ち帰りください。

B. CL guidelines (English translated version)

T01: Try to write sentences of no more than 50 characters.

Description:

Do not include too much content in a single sentence. If it contains multiple predicates, separate them into multiple short sentences.

Example:

(1) Split the sentence.

Residence registration is [...], and you need to submit [...].

⇒ Residence registration is [...]. You need to submit [...].

(2) Use a bulleted list.

To work in Japan, [Requirement A], [Requirement B], and [Requirement C] should be satisfied before entering Japan.

⇒ To work in Japan, the following requirements should be satisfied before entering Japan:

- [Requirement A]
- [Requirement B]
- [Requirement C]

T02: Do not interrupt a sentence with a bulleted list.

Description:

Do not insert a bulleted list in the middle of the sentence. Complete the sentence first and then use a bulleted list.

Example:

To work in Japan,

- [Requirement A]
- [Requirement B]
- [Requirement C]

should be satisfied before entering Japan.

⇒ To work in Japan, the following requirements should be satisfied before entering Japan:

- [Requirement A]
- [Requirement B]
- [Requirement C]

T05: Ensure the relationship between the modifier and the modified is clear.**Description:**

Avoid ambiguous attachment of the modifier to the modified.

Example:

- (1) Complement phrases/Change word order.
 a large-scale waste disposal project
 ⇒ a large-scale project of waste disposal
- (2) Add a punctuation mark (specific to the Japanese language).
 okina niwa no ki
 ⇒ *okina niwa no, ki*
 (large garden tree)
-

T06: Use the particle *ga* only to mean ‘but’.**Description:**

The Japanese coordinating conjunction *ga* can be used to mean ‘but’ (adversative) and ‘and’ (conjunctive). Use it only to mean ‘but’.

Example:

- (1) Split the sentence.
 Shiyakusho wa heijitsu ni hiraite-imasu ga, daini doyobi mo hiraite-imasu.
 ⇒ *Shiyakusho wa heijitsu ni hiraite-imasu. Mata daini doyobi mo hiraite-imasu.*
 (The city hall is open on weekdays. It is also open on the second Saturday.)
- (2) Use a conjunctive particle.
 Shiyakusho wa heijitsu ni hiraite-imasu ga, daini doyobi mo hiraite-imasu.
 ⇒ *Shiyakusho wa heijitsu ni hiraite-ori, daini doyobi mo hiraite-imasu.*
 (The city hall is open on weekdays and on the second Saturday.)
-

T07: Do not use the preposition *tame* to mean ‘because’. Use the preposition *node*.**Description:**

The Japanese preposition *tame* has multiple meanings such as ‘in order to’ and ‘because’. Use *tame* to mean ‘in order to’ and *node* to mean ‘because’.

Example:

- Seikatsu yohin ga tarinai tame, bussshi no shien o yosei-shita.*
 ⇒ *Seikatsu yohin ga tarinai node, bussshi no shien o yosei-shita.*
 (We requested relief supplies because there is a lack of daily necessities.)

T09: Avoid using multiple negative forms in a sentence.**Description:**

Multiple negation in a sentence can complicate its meaning. Use positive expressions as much as possible.

Example:

- (1) Rephrase the sentence using contraposition.
If you do not register, you cannot use it.
⇒ To use it, you need to register.
 - (2) Convert a double negation to an affirmative.
It is not impossible to use it on weekends.
⇒ It is possible to use it on weekends.
-

T10: Do not use *reru/rareru* to express the potential mood or honorifics.**Description:**

The Japanese auxiliary verbs *reru/rareru* can be used to express several meanings, such as potential mood, passive voice, and honorifics. Use *reru/rareru* only to express passive voice.

Example:

- (1) Use a direct style instead of honorifics.
Toroku sareta kata nomi shiyo deki-masu.
⇒ *Toroku shita kata nomi shiyo deki-masu.*
(Only registered users can use it.)
 - (2) Use a *dekiru* form instead of a potential mood form.
To-shisetsu wa 50-nin made ukeire-rare-masu.
⇒ *To-shisetsu wa 50-nin made ukeireru koto ga deki-masu.*
(The facility can accommodate up to 50 people.)
-

T11: Avoid using words that can be interpreted in multiple ways.**Description:**

Some simple verbs, such as ‘do’, ‘take’, ‘have’ and ‘get’, can be interpreted in multiple ways. Use verbs that have concrete meanings as much as possible. However, do not use unnecessarily difficult words.

Example:

- Please take your ID card.
⇒ Please bring your ID card.
-

T13: Avoid using the expression *to-iu*.**Description:**

The Japanese functional expression *to-iu* often does not have a particular meaning. In such cases, delete *to-iu* from the sentence.

Example:

- Mada shikaku o koshin-suru to-iu dankai dewa arimasen.*
⇒ *Mada shikaku o koshin-suru dankai dewa arimasen.*
(It is not the time to renew your qualification.)

T16: Avoid using the expressions *omowa-reru* and *kangae-rareru*.**Description:**

The Japanese expressions *omowa-reru* (feel) and *kangae-rareru* (think) do not have important meanings in municipal documents.

Example:

- (1) Delete the expression.

Hinan-suru ho-ga yoi to omowareru.

⇒ *Hinan-suru ho-ga yoi.*

(It is better to evacuate.)

- (2) Use a more concrete expression.

Tsugi no yoni kangae-rareru.

⇒ *Tsugi no tori desu.*

(It is as follows:)

T18: Avoid the single use of the form *tari*.**Description:**

When you use the particle *tari*, attach it to all of the enumerated items.

Example:

- (1) Add *tari*.

Mado-garasu ga ware-tari, shokkidana ga taoreru to taihen kiken desu.

⇒ *Mado-garasu ga ware-tari, shokkidana ga taore-tari-suru to taihen kiken desu.*

(It is very dangerous if window glass is broken or a cupboard collapses.)

- (2) Use another expression.

Chiki de hanashia-ttari, gakushu-kai no kaisai o osusume shimasu.

⇒ *Chiki de no hanashiai ya gakushu-kai no kaisai o osusume shimasu.*

(We recommend discussions and study sessions in local communities.)

T20: Use words from a general Japanese-English dictionary.**Description:**

Municipal documents should be written in simple language that even non-native speakers of Japanese can manage to understand using general bilingual dictionaries. Proper nouns and technical terms are allowed.

Example:

発災/*hassai*

⇒ 災害発生/*saigai hassei*

(occurrence of disasters)

T21: Avoid using compound *sahen*-nouns.**Description:**

A *sahen*-noun is a noun which can be connected to the verb *suru* to form a new verb. For example, the *sahen*-noun *sousa* (operation) can be converted to *sousa-suru* (operate). Complex combinations of multiple *sahen*-nouns, such as *henshu-sousa-suru*, should be avoided.

Example:

- (1) Decompose compound *sahen*-nouns.

Bunbogu o konyu-haibun-suru.

⇒ *Bunbogu o konyu-shi, haibun-suru.*

(We will purchase and distribute stationery.)

- (2) Use a centred dot.

メールを受信転送する。/*Meru o jushin-tensou-suru.*

⇒ メールを受信・転送する。/*Meru o jushin・tensou-suru.*

(Receive and forward e-mail.)

T22: Ensure there are no typographical errors or missing characters.**Description:**

In particular, carefully check for typographical errors when you use kana/kanji conversion and missing characters when you copy and paste text.

Example:

施設の会館時間は、ホームページご覧ください。/*Shisetsu no kaikan jikan wa, homu-peji goran kudasai.*

⇒ 施設の開館時間は、ホームページをご覧ください。/*Shisetsu no kaikan jikan wa, homu-peji o goran kudasai.*

(Please see the website for the operating hours of the facility.)

R02: Do not omit subjects.**Description:**

In Japanese texts, subjects tend to be omitted. Insert subjects as much as possible using contextual knowledge and common sense.

Example:

Gidai ni-tsuite kappatsu ni giron-shimasita.

(The agendas were actively discussed.)

⇒ *Shimin wa gidai ni-tsuite kappatsu ni giron-shimasita.*

(Citizens actively discussed the agendas.)

R03: Do not omit objects.**Description:**

Insert objects as much as possible using contextual knowledge and common sense. Some redundancy is allowed.

Example:

Honyaku no mokuteki wa, gaikoku no kata ga yomeru youni suru tame desu.

(The purpose of the translation is to help foreign residents to read.)

⇒ *Honyaku no mokuteki wa, gaikoku no kata ga bunsyo o yomeru youni suru tame desu.*

(The purpose of the translation is to help foreign residents to read the documents.)

R04: Do not use commas for connecting noun phrase enumeration.**Description:**

Japanese commas can be used roughly to separate chunks of information. When you enumerate noun phrases, use coordinating conjunctions such as *to*, *ya*, *ka*, and *mata*.

Example:

建物の設計、施工に問題がある。/*Tatemono no sekkei, sekou ni mondai ga aru.*

⇒ 建物の設計や施工に問題がある。/*Tatemono no sekkei ya sekou ni mondai ga aru.*

(There are problems in the design and construction of the building.)

R05: Avoid using particle *ga* for objects.**Description:**

The particle *ga* should be used only for subjects. Use the particle *o* for objects.

Example:

Tekisetsu na sabisu ga ukeraremasu.

⇒ *Tekisetsu na sabisu o ukeraremasu.*

(You can receive the appropriate service.)

R07: Avoid using *te-kuru/te-iku*.**Description:**

The Japanese functional expressions *te-kuru* and *te-iku* often do not have particular meanings. Unnecessary *te-kuru* and *te-iku* should be omitted.

Example:

Kenko-shinsa o jisshi shi-te-iku yotei desu.

⇒ *Kenko-shinsa o jisshi suru yotei desu.*

(Health checkups are planned to be conducted.)

R08: Avoid inserted adverbial clauses.

Description:

Long adverbial clauses inserted between subjects and predicates may complicate sentence structures. Bring adverbial clauses to the head of the sentence.

Example:

The event will, except on rainy days, be held in the city park.

⇒ Except on rainy days, the event will be held in the city park.

R09: Do not end clauses with nouns.

Description:

Complete clauses and sentences.

Example:

(1) Complete the end of the clause.

Hinanjo ni ido, joho-shushu ni tsutome-masho.

⇒ *Hinanjo ni ido-shi, joho-shushu ni tsutome-masho.*

(Move to the shelter and try to gather information.)

(2) Complete the end of the sentence.

Junbi-suru mono wa, pasupoto · inkan, zairyu-kado.

⇒ *Junbi-suru mono wa, pasupoto · inkan, zairyu-kado desu.*

(What you need to prepare are your passport, personal seal, and residence card.)

R10: Avoid using *sahen-noun* + auxiliary verb *desu*.

Description:

A *sahen-noun* is a noun which can be connected to the verb *suru* to form a new verb. For example, the *sahen-noun* *sousa* (operation) can be converted to *sousa-suru* (operate). Avoid using the construction *sahen-noun* + auxiliary verb *desu* as much as possible.

Example:

Obo no shimekiri desu.

⇒ *Obo o shimekiri mashita.*

(The application has been closed.)

R11: Avoid using attributive use of *shika-nai*.

Description:

Convert the expression *shika-nai* to *nomi* or *dake*.

Example:

Shitei no gomi-bukuro shika tsukae-masen.

⇒ *Shitei no gomi-bukuro nomi tsukae-masu.*

(You can use only the specified garbage bags.)

R12: Avoid using verb + *you-ni*.**Description:**Delete unnecessary *you-ni*.**Example:***Komame ni kyukei o toru youni shimasho.*⇒ *Komame ni kyukei o torimasho.*(Take a break frequently.)

R16: Avoid using particle *nado*.**Description:**The particle *nado* means ‘etc.’ Avoid the use of *nado* as much as possible.**Example:***Mibun-shomeisho nado o jisanshite kudasai.*

(Bring your IDs, etc.)

⇒ *Mibun-shomeisho o jisanshite kudasai.*(Bring your IDs.)

R17: Avoid using giving and receiving verbs.**Description:**The Japanese giving verb *ageru* and receiving verb *kureru* do not have concrete meanings. Delete them as much as possible.**Example:***Honnin no kawari ni shorui o teishutu-shite agete kudasai.*⇒ *Honnin no kawari ni shorui o teishutu-shite kudasai.*(Submit the document on behalf of the applicant.)

R18: Avoid using verbose wording.**Description:**In particular, the following Japanese expressions should be avoided: *mono-to-suru*, *mono-to-naru*, *mono-ni-suru*, *mono-ni-naru*, *koto-to-suru*, *koto-to-naru*, *koto-ni-suru*, and *koto-ni-naru*.**Example:***Kono koen wa saigai-ji ni shi ga hinanjo to-shite shitei suru mono-to-narimasu.*⇒ *Kono koen wa saigai-ji ni shi ga hinanjo to-shite shitei shimasu.*(The city designates this park as a shelter during a disaster.)

R19: Avoid using compound words.**Description:**

Avoided using compound nouns that combine more than two nouns, except for proper nouns and fixed terms.

Example:

Shinjuku City waste disposal rules

⇒ rules for waste disposal in Shinjuku City

R20: Do not omit parts of words in enumeration.

Description:

When you enumerate items, keep the form of each item complete.

Example:

- Shinjuku, Shibuya, and Toshima City
 ⇒ Shinjuku City, Shibuya City, and Toshima City
-

R25: Do not omit expressions that mean ‘per A’.

Description:

To indicate ‘per A’, explicitly use *ni-tsuki* or *atari*.

Example:

- Kodomo hitori 1000-en desu.*
 ⇒ *Kodomo hitori ni-tsuki 1000-en desu.*
 (The cost is 1000 yen per child.)
-

R26: Avoid using conjunctive particle *te*.

Description:

At the end of clauses, use the continuative form of the verb *shi* instead of the conjunctive particle *te*.

Example:

- Boshi-techo o jisan shite, kenko-shindan o jushin shite kudasai.*
 ⇒ *Boshi-techo o jisan shi, kenko-shindan o jushin shite kudasai.*
 (Bring your mother-and-child handbook, and have your child undergo a health checkup.)
-

R27: Avoid using ‘if’ particle *to*.

Description:

The Japanese particle *to* is commonly used as a coordinate conjunction. For a conditional clause, use the conjunctive expressions *tara*, *reba*, or *baai* instead of *to*.

Example:

- Todokede o dasu to teate o uketoremasu.*
 ⇒ *Todokede o dase ba teate o uketoremasu.*
 (If you submit the notification, you will receive the allowance.)

R33: Use Chinese kanji characters for verbs as much as possible instead of Japanese kana characters.**Description:**

Japanese kana characters may be converted to Chinese kanji characters in multiple ways. To avoid ambiguity, use as much as possible the kanji characters included in the Joyo Kanji list (kanji for ordinary use). Do not change the characters of proper nouns and technical terms.

Example:

- 仕事をやめずに日本に滞在できます。/Shigoto o yamezu ni nihon ni taizai-dekimasu.
 ⇒ 仕事を辞めずに日本に滞在できます。/Shigoto o yamezu ni nihon ni taizai-dekimasu.
 (You can stay in Japan without quitting your job.)
-

R34: Avoid leaving bullet marks in texts.**Description:**

Delete decorative symbols such as ○, ●, ▲, ★ and * at the heads of sentences. Use centred dots for the bulleted items.

Example:

- Bring your ID card.
 ⇒ Bring your ID card.
-

R35: Avoid using machine-dependent characters.**Description:**

Machine-dependent characters are special characters that can be displayed only in specific computer environments and language settings.

Example:

- The area of this room is 14 m².
 ⇒ The area of this room is 14 square meters.
-

R37: Avoid using square brackets for emphasis.**Description:**

Japanese square brackets are commonly used to indicate quotations and titles (names). Avoid using square brackets to emphasise key words and phrases.

Example:

- 市のカレンダーは「無料で」お持ち帰りください。/Shi no karenda wa “muryo de” omochikaeri kudasai.
 (Please take home the city calendar “for free”.)
 ⇒ 市のカレンダーは無料でお持ち帰りください。/Shi no karenda wa muryo de omochikaeri kudasai.
 (Please take home the city calendar for free.)

Bibliography

- ACCEPT (2013a). D 2.2 Definition of pre-editing rules for English and French (final version). http://www.accept.unige.ch/Products/D2_2_Definition_of_Pre-Editing_Rules_for_English_and_French_with_appendixes.pdf.
- ACCEPT (2013b). D 9.2.2 Survey of evaluation results (version 1). http://www.accept.unige.ch/Products/D_9_2_Survey_of_evaluation_results.pdf.
- ACCEPT (2013c). D4.2 Report on robust machine translation: Domain adaptation and linguistic back-off. http://www.accept.unige.ch/Products/D_4_2_Report_on_robust_machine_translation_domain_adaptation_and_linguistic_back-off.pdf.
- Adachi, R., Takeuchi, K., Murayama, R., Fanderl, W., Miyata, R., Vogel, I., Apel, U., and Kageura, K. (2013). Development and use of a platform for defining idiom variation rules. In *Proceedings of the 5th International Language Learning Conference (ILLC)*, pages 1–19, Penang, Malaysia.
- Adamic, L. A. (2002). Zipf’s law and the Internet. *Glottometrics*, (3):143–150.
- Adriaens, G. and Macken, L. (1995). Technological evaluation of a controlled language application: Precision, recall, and convergence tests for SECC. In *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI)*, pages 123–141, Leuven, Belgium.
- Adriaens, G. and Schreurs, D. (1992). From Cogram to Alcogram: Toward a controlled English grammar checker. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING)*, pages 595–601, Nantes, France.
- AECMA (1995). A guide for the preparation of aircraft maintenance documents in the aerospace maintenance language AECMA Simplified English. AECMA Document, PSC-85-16598.
- Ahmad, K. and Rogers, M. (2001). Corpus linguistics and terminology extraction. In Wright, S. E. and Budin, G., editors, *Handbook of Terminology Management, Vol.2: Application-Oriented Terminology Management*, pages 725–760. John Benjamins, Amsterdam.
- 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.
- Alabau, V., Leiva, L. A., Ortiz-Martínez, D., and Casacuberta, F. (2012). User evaluation of interactive machine translation systems. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT)*, pages 20–23, Trento, Italy.

- Allen, J. (2003). Post-editing. In Somers, H., editor, *Computers and Translation: A Translator's Guide*, pages 297–317. John Benjamins, Amsterdam.
- Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.
- ASD (2017). ASD Simplified Technical English. Specification ASD-STE100, Issue 7. <http://www.asd-ste100.org>.
- Baayen, H. (2001). *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht.
- Baayen, H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press, Cambridge.
- Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, pages 1–15, Toulon, France.
- Baisa, V., Ulipová, B., and Cukr, M. (2015). Bilingual terminology extraction in Sketch Engine. In *Proceedings of the 9th Workshop on Recent Advances in Slavonic Natural Language Processing (RASLAN)*, pages 61–67, Karlova Studánka, Czech Republic.
- Banerjee, S. and Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan.
- Baroni, M. (2009). Distributions in text. In Lüdeling, A. and Kytö, M., editors, *Corpus Linguistics: An International Handbook*, pages 803–822. Mouton de Gruyter, Berlin.
- Bellamy, L., Carey, M., and Schlotfeldt, J. (2012). *DITA Best Practices: A Roadmap for Writing, Editing, and Architecting in DITA*. IBM Press, Upper Saddle River, New Jersey.
- Berka, J., Černý, M., and Bojar, O. (2011). Quiz-based evaluation of machine translation. *Prague Bulletin of Mathematical Linguistics*, 95:77–86.
- Bernth, A. (1997). EasyEnglish: A tool for improving document quality. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP)*, pages 159–165, Washington, DC.
- Bernth, A. (1999). Controlling input and output of MT for greater user acceptance. In *Proceedings of the 21th Conference of Translating and the Computer*, London.
- Bernth, A. (2006). EasyEnglishAnalyzer: Taking controlled language from sentence to discourse level. In *Proceedings of the 5th International Workshop on Controlled Language Applications (CLAW)*, pages 1–7, Cambridge, Massachusetts.
- Bernth, A. and Gdaniec, C. (2001). MTranslatability. *Machine Translation*, 16(3):175–218.
- Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. Continuum International, London.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4):243–257.
- Biber, D. and Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press, New York.
- Billr, O., Elhadad, M., and Netzer, Y. (2005). Interactive authoring of logical forms for multilingual generation. In *Proceedings of the 10th European Workshop on Natural Language Generation (ENLG)*, pages 24–31, Aberdeen, Scotland.
- Bojar, O., Chatterjee, R., Federmann, C., Haddow, B., Huck, M., Hokamp, C., Koehn, P., Logacheva, V., Monz, C., Negri, M., Post, M., Scarton, C., Specia, L., and Turchi, M. (2015). Findings of the 2015 Workshop on Statistical Machine Translation. In *Proceedings of the 10th Workshop on Statistical Machine Translation (WMT)*, pages 1–46, Lisbon, Portugal.

- Bouayad-Agha, N., Power, R., Scott, D., and Belz, A. (2002). PILLS: Multilingual generation of medical information documents with overlapping content. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, pages 2111–2114, Las Palmas, Spain.
- Bredenkamp, A., Crysmann, B., and Petrea, M. (2000). Looking for errors: A declarative formalism for resource-adaptive language checking. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC)*, Athens, Greece.
- Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes*, 13(1):47–59.
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In Jordan, P. W., Thomas, B., McClelland, I. L., and Weerdmeester, B., editors, *Usability Evaluation in Industry*, pages 189–194. Taylor and Francis, London.
- Brown, P. F., Cocke, J., Della, S. A. P., Della, V. J. P., Jelinek, F., Lafferty, J. D., Mercer, R. L., and Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2):79–85.
- Brown, P. F., Cocke, J., Della, S. A. P., Della, V. J. P., Jelinek, F., Mercer, R. L., and Roossin, P. S. (1988). A statistical approach to language translation. In *Proceedings of the 12th Conference on Computational Linguistics (COLING)*, pages 71–76, Budapest, Hungary.
- Brown, P. F., Della, V. J. P., Della, S. A. P., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.
- Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes*, 4(3):207–224.
- Cabré, M. T., Montané, M. A., and Nazar, R. (2012). Corpus-based terminology processing. In *Tutorial of the 10th International Congress of Terminology and Knowledge Engineering (TKE)*, Madrid, Spain.
- Cadwell, P. (2008). *Readability and Controlled Language*. MA Dissertation, Dublin City University.
- Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., and Schroeder, J. (2007). (Meta-) evaluation of machine translation. In *Proceedings of the 2nd Workshop on Statistical Machine Translation (WMT)*, pages 136–158, Prague, Czech Republic.
- Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., and Schroeder, J. (2008). Further meta-evaluation of machine translation. In *Proceedings of the 3rd Workshop on Statistical Machine Translation (WMT)*, pages 70–106, Columbus, Ohio.
- Carey, M., Lanyi, M. M., Longo, D., Radzinski, E., Rouiller, S., and Wilde, E. (2014). *Developing Quality Technical Information: A Handbook for Writers and Editors*. IBM Press, Upper Saddle River, New Jersey.
- Carl, M., Rascu, E., Haller, J., and Langlais, P. (2004). Abducing term variant translations in aligned texts. *Terminology*, 10(1):101–130.
- Carl, M. and Way, A. (2003). *Recent Advances in Example-Based Machine Translation*. Springer, Netherlands.
- Carroll, J. B. (1966). An experiment in evaluating the quality of translations. *Mechanical Translation and Computational Linguistics*, 9(3-4):67–75.
- Carroll, J. B. (1969). A rationale for an asymptotic lognormal form of word-frequency distributions. In *Research Bulletin*. Educational Testing Service, Princeton, New Jersey.
- Carroll, T. (2010). Local government websites in Japan: International, multicultural, multilingual? *Japanese Studies*, 30(3):373–392.

- Castilho, S., O'Brien, S., Alves, F., and O'Brien, M. (2014). Does post-editing increase usability? A study with Brazilian Portuguese as target language. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT)*, pages 183–190, Dubrovnik, Croatia.
- Cerrella Bauer, S. (2015). Automatic term extraction. In Kockaert, H. J. and Steurs, F., editors, *Handbook of Terminology*, volume 1, pages 203–221. John Benjamins, Amsterdam.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Colineau, N., Paris, C., and Linden, K. V. (2002). An evaluation of procedural instructional text. In *Proceedings of the International Natural Language Generation Conference (INLG)*, pages 128–135, New York.
- Colineau, N., Paris, C., and Linden, K. V. (2012). Government to citizen communications: From generic to tailored documents in public administration. *Information Polity*, 17(2):177–193.
- Colineau, N., Paris, C., and Linden, K. V. (2013). Automatically producing tailored web materials for public administration. *New Review of HyperMedia and MultiMedia*, 9(2):158–181.
- Collins, M., Koehn, P., and Kučerová, I. (2005). Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 531–540, Ann Arbor, Michigan.
- Costa, Â., Ling, W., Luís, T., Correia, R., and Coheur, L. (2015). A linguistically motivated taxonomy for machine translation error analysis. *Machine Translation*, 29(2):127–161.
- Cross, C. and Oppenheim, C. (2006). A genre analysis of scientific abstracts. *Journal of Documentation*, 62(4):428–446.
- Daille, B. (1996). Study and implementation of combined techniques for automatic extraction of terminology. In Resnik, P. and Klavans, J. L., editors, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, pages 49–66. MIT Press, Cambridge, Massachusetts.
- Daille, B. (2003). Conceptual structuring through term variations. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment (MWE)*, pages 9–16, Sapporo, Japan.
- Daille, B. (2005). Variations and application-oriented terminology engineering. *Terminology*, 11(1):181–197.
- Daille, B., Gaussier, É., and Langé, J.-M. (1994). Towards automatic extraction of monolingual and bilingual terminology. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING)*, pages 515–521, Kyoto, Japan.
- Daille, B., Habert, B., Jacquemin, C., and Royauté, J. (1996). Empirical observation of term variations and principles for their description. *Terminology*, 3(2):197–257.
- Damerau, F. J. (1990). Evaluating computer-generated domain-oriented vocabularies. *Information Processing & Management*, 26(6):791–801.
- Day, D., Priestley, M., and Schell, D. (2005). Introduction to the Darwin Information Typing Architecture: Toward portable technical information. <http://www.ibm.com/developerworks/xml/library/x-dita1/x-dita1-pdf.pdf>.
- De Jong, M. and Schellens, P. J. (2000). Toward a document evaluation methodology: What does research tell us about the validity and reliability of evaluation methods? *IEEE Transactions on Professional Communication*, 43(3):242–260.

- Désilets, A., Huberdeau, L.-P., Laporte, M., and Quirion, J. (2009). Building a collaborative multilingual terminology system. In *Proceedings of the 31st Conference of Translating and the Computer*, London.
- Dillinger, M. (2001). Dictionary development workflow for MT: Design and management. In *Proceedings of the Machine Translation Summit VIII*, pages 83–88, Galicia, Spain.
- DiscoMT (2013). *Proceedings of the Workshop on Discourse in Machine Translation (DiscoMT)*. Sofia, Bulgaria.
- DiscoMT (2015). *Proceedings of the 2nd Workshop on Discourse in Machine Translation (DiscoMT)*. Lisbon, Portugal.
- Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the 2nd International Conference on Human Language Technology Research (HLT)*, pages 138–145, San Diego, California.
- Doherty, S. (2012). *Investigating the Effects of Controlled Language on the Reading and Comprehension of Machine Translated Texts: A Mixed-Methods Approach*. PhD thesis, Dublin City University.
- Doherty, S. and O'Brien, S. (2012). A user-based usability assessment of raw machine translated technical instructions. In *Proceedings of the 10th Conference of the Association for Machine Translation in the Americas (AMTA)*, San Diego, CA.
- Doherty, S. and O'Brien, S. (2013). 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.
- Efron, B. and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika*, 63(3):435–447.
- Evert, S. (2004). A simple LNRE model for random character sequences. In *Proceedings of the 7es Journées internationales d'Analyse statistique des Données Textuelles (JADT)*, pages 411–422, Louvain-la-Neuve, France.
- Evert, S. and Baroni, M. (2005). Testing the extrapolation quality of word frequency models. In *Proceedings of the Corpus Linguistics 2005*, Birmingham, UK.
- Evert, S. and Baroni, M. (2007). zipfR: Word frequency distributions in R. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Posters and Demonstrations Session*, pages 29–32, Prague, Czech Republic.
- Feng, L. (2008). *Text Simplification: A Survey*. Technical Report, The City University of New York.
- Fischer, M. (2010). Language (policy), translation and terminology in the European Union. In Thelen, M. and Steurs, F., editors, *Terminology and Lexicography Research and Practice: Terminology in Everyday Life*, volume 13, pages 21–34. John Benjamins, Amsterdam.
- Flowerdew, J. and Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes*, 29(2):78–93.
- Foo, J. (2012). *Computational Terminology: Exploring Bilingual and Monolingual Term Extraction*. Licentiate thesis, Linköping University.
- Foo, J. and Merkel, M. (2010). Computer aided term bank creation and standardization: Building standardized term banks through automated term extraction and advanced editing tools. In Thelen, M. and Steurs, F., editors, *Terminology and Lexicography Research and Practice: Terminology in Everyday Life*, volume 13, pages 21–34. John Benjamins, Amsterdam.

- Frantzi, K., Ananiadou, S., and Mima, H. (2000). Automatic recognition of multi-word terms: The C-value/NC-value method. *International Journal on Digital Libraries*, 3(2):115–130.
- Frantzi, K., Ananiadou, S., and Tsujii, J. (1998). The C-value/NC-value method of automatic recognition for multi-word terms. In Nikolaou, C. and Stephanidis, C., editors, *Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference (ECDL)*, pages 585–604. Springer, Berlin, Heidelberg.
- Fulford, H. (2001). Exploring terms and their linguistic environment in text: A domain-independent approach to automated term extraction. *Terminology*, 7(2):259–279.
- Gaussier, É. (1998). Flow network models for word alignment and terminology extraction from bilingual corpora. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (ACL-COLING)*, pages 444–450, Montreal, Quebec, Canada.
- Gerlach, J., Porro, V., Bouillon, P., and Lehmann, S. (2013). Combining pre-editing and post-editing to improve SMT of user-generated content. In *Proceedings of the MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP)*, pages 45–53, Nice, France.
- Giménez, J., Márquez, L., Comelles, E., Castelloñ, I., and Arranz, V. (2010). Document-level automatic MT evaluation based on discourse representations. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, pages 333–338, Uppsala, Sweden.
- Gong, Z., Zhang, M., and Zhou, G. (2015). Document-level machine translation evaluation with gist consistency and text cohesion. In *Proceedings of the 2nd Workshop on Discourse in Machine Translation (DiscoMT)*, pages 33–40, Lisbon, Portugal.
- Gulati, A., Bouillon, P., Gerlach, J., Porro, V., and Seretan, V. (2015). The ACCEPT Academic Portal: A user-centred online platform for pre-editing and post-editing. In *Proceedings of the 7th International Conference of the Iberian Association of Translation and Interpreting Studies (AIETI)*, Malaga, Spain.
- Guzmán, F., Márquez, S. J. L., and Nakov, P. (2014). Using discourse structure improves machine translation evaluation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 687–698, Baltimore, Maryland.
- Haque, R., Penkale, S., and Way, A. (2014). Bilingual termbank creation via log-likelihood comparison and phrase-based statistical machine translation. In *Proceedings of the 4th International Workshop on Computational Terminology (CompuTerm)*, pages 42–51, Dublin, Ireland.
- Hardmeier, C. (2014). *Discourse in Statistical Machine Translation*. PhD thesis, Uppsala University.
- Hartley, A. (2010). Enabling multilingual applications of ‘Controlled Language’: The DITA framework. *AAMT Journal*, (48):15–18.
- Hartley, A. and Paris, C. (1997). Multilingual document production from support for translating to support for authoring. *Machine Translation*, 12(1):109–129.
- Hartley, A. and Paris, C. (2001). Translation, controlled languages, generation. In Steiner, E. and Yallop, C., editors, *Exploring Translation and Multilingual Text Production*, pages 307–325. Mouton, Berlin.
- 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 (EAMT)*, pages 237–244, Trento, Italy.

- Hoard, J. E., Wojcik, R., and Holzhauser, K. (1992). An automated grammar and style checker for writers of Simplified English. In Holt, P. O. and William, N., editors, *Computers and Writing: State of the Art*, pages 278–296. Intellect, Oxford.
- Horn, R. E. (1989). *Mapping Hypertext: The Analysis, Organization, and Display of Knowledge for the Next Generation of On-Line Text and Graphics*. Lexington Institute, Arlington.
- Horn, R. E. (1998). Structured writing as a paradigm. In Romiszowski, A. and Dills, C., editors, *Instructional Development: State of the Art*. Educational Technology Publications, Englewood Cliffs, New Jersey.
- Hoshino, S., Miyao, Y., Sudoh, K., Hayashi, K., and Nagata, M. (2015). Discriminative preordering meets Kendall's τ maximization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 139–144, Beijing, China.
- Hutchins, J. (2005a). Current commercial machine translation systems and computer-based translation tools: System types and their uses. *International Journal of Translation*, 17(1-2):5–38.
- Hutchins, J. (2005b). Example-based machine translation: A review and commentary. *Machine Translation*, 19(3):197–211.
- Hutchins, J. (2015). Machine translation: History of research and applications. In Chan, S.-W., editor, *Routledge Encyclopedia of Translation Technology*, pages 120–136. Routledge, New York.
- Hutchins, J. and Somers, H. (1992). *An Introduction to Machine Translation*. Academic Press, London.
- Inui, K. and Fujita, A. (2004). A survey on paraphrase generation and recognition. *Journal of Natural Language Processing*, 11(5):151–198. (乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理).
- Isahara, H. (2015). Translation technology in Japan. In Chan, S.-W., editor, *Routledge Encyclopedia of Translation Technology*, pages 315–326. Routledge, New York.
- ISO (2010). ISO 9241-210:2010 Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en>.
- ISO (2011). ISO/IEC 25010:2011 Systems and software engineering—Systems and software quality requirements and evaluation (SQuaRE)—System and software quality models. <https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en>.
- ISO (2012). ISO 26162:2012 Systems to manage terminology, knowledge and content—Design, implementation and maintenance of terminology management systems. <https://www.iso.org/obp/ui/es/#iso:std:iso:26162:ed-1:v1:en:term:3.2.7>.
- Isozaki, H., Hirao, T., Duh, K., Sudoh, K., and Tsukada, H. (2010a). Automatic evaluation of translation quality for distant language pairs. In *Proceedings of the 2010 Conference on Empirical Methods on Natural Language Processing (EMNLP)*, pages 944–952, Cambridge, Massachusetts.
- Isozaki, H., Sudoh, K., Tsukada, H., and Duh, K. (2010b). Head finalization: A simple reordering rule for SOV languages. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, pages 244–251, Uppsala, Sweden.
- Itagaki, M., Aikawa, T., and He, X. (2007). Automatic validation of terminology translation consistency with statistical method. In *Proceedings of the Machine Translation Summit XI*, pages 269–274, Copenhagen, Denmark.

- Jacquemin, C. (2001). *Spotting and Discovering Terms through Natural Language Processing*. MIT Press, Cambridge, Massachusetts.
- Japio (2013). Patent Documents Writing Manual. Ver.1. <https://www.tech-jpn.jp/tokkyo-writing-manual/>. (一般財団法人日本特許情報機構. 特許ライティングマニュアル).
- Jean, S., Cho, K., Memisevic, R., and Bengio, Y. (2015). On using very large target vocabulary for neural machine translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP)*, pages 1–10, Beijing, China.
- Jiang, B., Yin, J., and Liu, Q. (2015). Zipf's law for all the natural cities around the world. *International Journal of Geographical Information Science*, 29(3):498–522.
- JTCA (2011). *Style Guide for Japanese Documents*. Japan Technical Communication Association, Tokyo, Japan. (テクニカルコミュニケーション協会. 日本語スタイルガイド. テクニカルコミュニケーション協会出版事業部会).
- Kageura, K. (2012). *The Quantitative Analysis of the Dynamics and Structure of Terminologies*. John Benjamins, Amsterdam.
- Kageura, K. and Kikui, G. (2006). A self-referring quantitative evaluation of the ATR Basic Travel Expression Corpus (BTEC). In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 1945–1950, Genoa, Italy.
- Kageura, K. and Umino, B. (1996). Methods of automatic term recognition: A review. *Terminology*, 3(2):259–289.
- Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1700–1709, Seattle, Washington.
- 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 (CLAW)*, pages 51–61, Pittsburgh, Pennsylvania.
- Kando, N. (1997). Text-level structure of research articles: Implications for text-based information processing systems. In *Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research*, pages 1–14, Aberdeen, Scotland.
- Kando, N. (1999). Text structure analysis as a tool to make retrieved document usable. In *Proceedings of the 4th International Workshop on Information Retrieval with Asian Languages (IRAL)*, pages 126–135, Taipei, Taiwan.
- Karkaletsis, V., Samaritakis, G., Petasis, G., Farmakiotou, D., Androutsopoulos, I., and Spyropoulos, C. D. (2001). A controlled language checker based on the Ellogon text engineering platform. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL), Software Demonstrations*, pages 90–103, Pennsylvania.
- Karsch, B. I. (2015). Terminology work and crowdsourcing: Coming to terms with the crowd. In Kockaert, H. J. and Steurs, F., editors, *Handbook of Terminology*, volume 1, pages 291–303. John Benjamins, Amsterdam.
- Khmaladze, E. V. (1987). *The Statistical Analysis of Large Numbers of Rare Events*. Technical Report MS-R8804, Department of Mathematical Sciences, CWI, Amsterdam.
- Kilgarrieff, A., Rychlý, P., Smrž, P., and Tugwell, D. (2004). The Sketch Engine. In *Proceedings of the 11th EURALEX International Congress*, pages 105–116, Lorient, France.
- Kim, Y., Hong, M., and Park, S.-K. (2007). CL-guided Korean-English MT system for scientific papers. In *Proceedings of the 8th International Conference on Intelligent Text*

- Processing and Computational Linguistics (CICLing)*, pages 409–419, Mexico City, Mexico.
- Kittredge, R. (2003). Sublanguages and controlled languages. In Mitkov, R., editor, *Oxford Handbook of Computational Linguistics*, pages 430–437. Oxford University Press, Oxford.
- Knight, K. and Chander, I. (1994). Automated postediting of documents. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI)*, pages 779–784, Seattle, Washington.
- Koehn, P. (2009). *Statistical Machine Translation*. Cambridge University Press, New York.
- Koehn, P. and Hermann, U. (2014). The impact of machine translation quality on human post-editing. In *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation (HaCat)*, pages 38–46, Gothenburg, Sweden.
- Kohl, J. R. (2008). *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*. SAS Institute, Cary, North Carolina.
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25:131–148.
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3):411–433.
- Kruijff, G.-J. M., Teich, E., Bateman, J. A., Kruijff-Korbayová, I., Skoumalová, H., Sharoff, S., Sokolova, E. G., Hartley, T., Staykova, K., and Hana, J. (2000). Multilinguality in a text generation system for three Slavic languages. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, pages 474–480, Saarbrücken, Germany.
- Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1):121–170.
- Kupiec, J. (1993). An algorithm for finding noun phrase correspondences in bilingual corpora. In *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics (ACL)*, pages 17–22, Columbus, Ohio.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Langlais, P. and Carl, M. (2004). General-purpose statistical translation engine and domain specific texts: Would it work? *Terminology*, 10(1):131–153.
- LDC (2005). Linguistic data annotation specification: Assessment of fluency and adequacy in translations, Revision 1.5. Technical Report, Linguistic Data Consortium.
- Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In Hundt, M., Nesselhauf, N., and Biewer, C., editors, *Corpus Linguistics and the Web*, pages 133–149. Rodopi, Amsterdam.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8):707–710.
- L’Homme, M.-C. (1994). Management of terminology in a machine-translation environment. *Terminology*, 1(1):121–135.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12):317–324.
- Mann, W. and Thompson, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- Marco, C. D., Bray, P., Covvey, D., Cowan, D., Ciccio, V. D., Hovy, E., Lipa, J., and Yang, C. (2008). Authoring and generation of individualised patient education materials. *Journal on Information Technology in Healthcare*, 6(1):63–71.

- Maswana, S., Kanamaru, T., and Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, 2:1–11.
- 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. (松田成正. 産業日本語の取り組み: 特許ライティングマニュアルを中心に. 情報管理).
- Matsuzaki, T., Fujita, A., Todo, N., and Arai, N. H. (2015). Evaluating machine translation systems with second language proficiency tests. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 145–149, Beijing, China.
- Matsuzaki, T., Fujita, A., Todo, N., and Arai, N. H. (2016). Translation errors and incomprehensibility: A case study using machine-translated second language proficiency tests. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, pages 2771–2776, Portorož, Slovenia.
- McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, Cambridge.
- Mirkin, S., Venkatapathy, S., Dymetman, M., and Calapodescu, I. (2013). SORT: An interactive source-rewriting tool for improved translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL), System Demonstrations*, pages 85–90, Sofia, Bulgaria.
- Mitamura, T., Baker, K., Nyberg, E., and Svoboda, D. (2003). Diagnostics for interactive controlled language checking. In *Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (EAMT/CLAW)*, pages 237–244, Dublin, Ireland.
- Mitamura, T. and Nyberg, E. (2001). Automatic rewriting for controlled language translation. In *Proceedings of the NLPRS 2001 Workshop on Automatic Paraphrasing: Theories and Applications*, pages 1–12, Tokyo, Japan.
- Mitkov, R. (1999). Introduction: Special issue on anaphora resolution in machine translation and multilingual NLP. *Machine Translation*, 14(3-4):159–161.
- Miyabe, M., Yoshino, T., and Shigenobu, T. (2009). Effects of undertaking translation repair using back translation. In *Proceedings of the ACM International Workshop on Intercultural Collaboration (IWIC)*, pages 33–40, Palo Alto, California.
- Miyata, R., Adachi, R., Apel, U., Vogel, I., Fanderl, W., Murayama, R., Takeuchi, K., and Kageura, K. (2014). The use of corpus evidence and human introspection to create idiom variations. In *Proceedings of the 2nd Asia Pacific Corpus Linguistics Conference (APCLC)*, pages 201–202, Hong Kong.
- Møller, M. H. and Christoffersen, E. (2006). Building a controlled language lexicon for Danish. *LSP & Professional Communication*, 6(1):26–37.
- Moreno-Sánchez, I., Font-Clos, F., and Corral, Á. (2016). Large-scale analysis of Zipf’s law in English texts. *PLoS ONE*, 11(1):1–19.
- Mossop, B. (2014). *Revising and Editing for Translators*. Routledge, New York.
- Nagao, M. (1984). A framework of a mechanical translation between Japanese and English by analogy principle. In Elithorn, A. and Banerji, R., editors, *Artificial and Human Intelligence*, pages 173–180. Elsevier/North-Holland, New York.

- Nagao, M., Tanaka, N., and Tsujii, J. (1984). Support system for writing texts based on controlled grammar. *IPSJ SIG Technical Reports*, NL(44):33–40. (長尾真, 田中伸佳, 辻井潤一. 制限文法にもとづく文章作成援助システム. 情報処理学会研究報告).
- Nanjo, H., Yamamoto, Y., and Yoshimi, T. (2012). Automatic construction of statistical pre-editing system from parallel corpus for improvement of machine translation quality. *Information Processing Society of Japan*, 53(6):1644–1653. (南條浩輝, 山本祐司, 吉見毅彦. 機械翻訳の品質向上のための対訳コーパスからの統計的前編集システムの自動構築. 情報処理学会論文誌).
- Nguyen, D. and Rosé, C. P. (2011). Language use as a reflection of socialization in online communities. In *Proceedings of the Workshop on Languages in Social Media (LSM)*, pages 76–85, Portland, Oregon.
- Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann, San Francisco.
- Nielsen, J. (2012). Usability 101: Introduction to usability. <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
- Nyberg, E. and Mitamura, T. (2000). The KANTOO machine translation environment. In *Proceedings of the 4th Conference of the Association for Machine Translation in the Americas (AMTA)*, pages 192–195, Cuernavaca, Mexico.
- 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. John Benjamins, Amsterdam.
- OASIS (2010). Darwin Information Typing Architecture (DITA) Version 1.2. <http://docs.oasis-open.org/dita/v1.2/os/spec/DITA1.2-spec.html>.
- O’Brien, S. (2003). Controlling controlled English: An analysis of several controlled language rule sets. In *Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (EAMT/CLAW)*, pages 105–114, Dublin, Ireland.
- O’Brien, S. (2006a). Controlled language and post-editing. *Multilingual*, 17(7):17–19.
- O’Brien, S. (2006b). *Machine-Translatability and Post-Editing Effort: An Empirical Study Using Translog and Choice Network Analysis*. PhD thesis, Dublin City University.
- O’Brien, S. (2010). Controlled language and readability. In Shreve, G. M. and Angelone, E., editors, *Translation and Cognition*, pages 143–165. John Benjamins, Amsterdam.
- Ó Broin, U. (2009). Controlled authoring to improve localization. *Multilingual*, October/November:12–14.
- Oda, J. (2010). *Ways to improve websites for municipalities: Technique and attitude needed for PR managers*. JIJI Press, Tokyo. (小田順子. 自治体のためのウェブサイト改善術: 広報担当に求められるテクニックとマインド. 時事通信社).
- 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. (小倉英里, 工藤真代, 柳英夫. シンプルファイド・テクニカル・ジャパニーズ: 英訳を視野に入れて日本語を作る. 情報処理学会研究報告).
- Okazaki, N. and Tsujii, J. (2010). Simple and efficient algorithm for approximate dictionary matching. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 851–859, Beijing, China.
- OpenUM Project (2011). Report on the first meeting of OpenUM project working group. <http://www.slideshare.net/OpenUM/open-um-projectphase01report>. (OpenUMプロジェクト. OpenUMプロジェクト第一次検討部会報告書).
- Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 311–318, Philadelphia, Pennsylvania.

- Paris, C., Colineau, N., Lampert, A., and Linden, K. V. (2010). Discourse planning for information composition and delivery: A reusable platform. *Journal of Natural Language Engineering*, 16(1):61–98.
- Paris, C., Colineau, N., Lu, S., and Linden, K. V. (2005). Automatically generating effective on-line help. *International Journal on E-Learning*, 4(1):83–103.
- PLAIN (2011). Federal plain language guidelines, Revision 1. <https://plainlanguage.gov/media/FederalPLGuidelines.pdf>.
- Plitt, M. and Masselot, F. (2010). A productivity test of statistical machine translation post-editing in a typical localisation context. *Prague Bulletin of Mathematical Linguistics*, 93:7–16.
- Power, R., Scott, D., and Hartley, A. (2003). Multilingual generation of controlled languages. In *Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (EAMT/CLAW)*, pages 15–17, Dublin, Ireland.
- Pym, P. (1990). Pre-editing and the use of simplified writing for MT. In Mayorcas, P., editor, *Translating and the Computer 10: The Translation Environment 10 Years on*, pages 80–95. Aslib, London.
- Rascu, E. (2006). A controlled language approach to text optimization in technical documentation. In *Proceedings of Konferenz zur Verarbeitung Natürlicher Sprache (KONVENS)*, pages 107–114, Konstanz, Germany.
- Reiter, E. and Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge.
- Reiter, E., Robertson, R., Lennox, A. S., and Osman, L. (2001). Using a randomised controlled clinical trial to evaluate an NLG system. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 442–449, Toulouse, France.
- Resnik, P., Buzek, O., Hu, C., Kronrod, Y., Quinn, A., and Bederson, B. B. (2010). Improving translation via targeted paraphrasing. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 127–137, Massachusetts.
- Reuther, U. (2003). Two in one – Can it work?: Readability and translatability by means of controlled language. In *Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (EAMT/CLAW)*, pages 124–132, Dublin, Ireland.
- Reynolds, P. (2015). Machine translation, translation memory and terminology management. In Kockaert, H. J. and Steurs, F., editors, *Handbook of Terminology*, volume 1, pages 276–287. John Benjamins, Amsterdam.
- Roturier, J. (2004). Assessing a set of controlled language rules: Can they improve the performance of commercial machine translation systems? In *Proceedings of the 26th Conference of Translating and the Computer*, pages 1–14, London.
- Roturier, J. (2006). *An Investigation into the Impact of Controlled English Rules on the Comprehensibility, Usefulness and Acceptability of Machine-Translated Technical Documentation for French and German Users*. PhD thesis, Dublin City University.
- Roturier, J., Mitchell, L., Grabowski, R., and Siegel, M. (2012). Using automatic machine translation metrics to analyze the impact of source reformulations. In *Proceedings of the 10th Conference of the Association for Machine Translation in the Americas (AMTA)*, San Diego, California.

- Roturier, J., Mitchell, L., and Silva, D. (2013). The ACCEPT post-editing environment: A flexible and customisable online tool to perform and analyse machine translation post-editing. In *Proceedings of the MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP)*, pages 119–128, Nice, France.
- Rubens, P., editor (2001). *Science and Technical Writing: A Manual of Style*. Routledge, New York.
- Sager, J. C. (1990). *A Practical Course in Terminology Processing*. John Benjamins, Amsterdam.
- Sager, J. C. (2001). Terminology compilation: Consequences and aspects of automation. In Wright, S. E. and Budin, G., editors, *Handbook of Terminology Management, Vol.2: Application-Oriented Terminology Management*, pages 761–771. John Benjamins, Amsterdam.
- Sato, K., Takeuchi, K., and Kageura, K. (2013). Terminology-driven augmentation of bilingual terminologies. In *Proceedings of the Machine Translation Summit XIV*, pages 3–10, Nice, France.
- Sato, S. and Nagao, M. (1990). Toward memory-based translation. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING)*, pages 247–252, Helsinki, Finland.
- Sato, S., Tsuchiya, M., Murayama, M., Asaoka, M., and Wang, Q. (2003). Standardization of Japanese sentences. *IPSJ SIG Technical Reports*, NL(4):133–140. (佐藤理史, 土屋雅稔, 村山賢洋, 麻岡正洋, 玉晴晴. 日本語文の規格化. 情報処理学会研究報告).
- Sauro, J. and Lewis, J. R. (2012). *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann, Burlington, Massachusetts.
- Schmidt-Wigger, A. (1999). Term checking through term variation. In *Proceedings of the 5th International Congress of Terminology and Knowledge Engineering (TKE)*, pages 570–581, Vienna, Austria.
- Schriver, K. A. (1997). *Dynamics in Document Design: Creating Text for Readers*. John Wiley & Sons, New York.
- Schwartz, L. (2014). Monolingual post-editing by a domain expert is highly effective for translation triage. In *Proceedings of the 3rd Workshop on Post-editing Technology and Practice (WPTP)*, pages 34–44, Vancouver, Canada.
- Seretan, V., Bouillon, P., and Gerlach, J. (2014a). A large-scale evaluation of pre-editing strategies for improving user-generated content translation. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*, pages 1793–1799, Reykjavik, Iceland.
- Seretan, V., Roturier, J., Silva, D., and Bouillon, P. (2014b). The ACCEPT portal: An online framework for the pre-editing and post-editing of user-generated content. In *Proceedings of the Workshop on Humans and Computer-Assisted Translation (HaCaT)*, pages 66–71, Gothenburg, Sweden.
- 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 (CLAW)*, pages 62–75, Pennsylvania.
- Shubert, S. K., Spyridakis, J. H., Holmback, H. K., and Coney, M. B. (1995). The comprehensibility of Simplified English in procedures. *Journal of Technical Writing and Communication*, 25(4):347–369.
- Sichel, H. S. (1975). On a distribution law for word frequencies. *Journal of the American Statistical Association*, 70(351a):542–547.

- Simard, M., Goutte, C., and Isabelle, P. (2007). Statistical phrase-based post-editing. In *Proceedings of the Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 508–515, Rochester, New York.
- Simon, H. (1960). Some further notes on a class of skew distribution functions. *Information and Control*, 3(1):80–88.
- Skalicky, S. (2013). Was this analysis helpful? A genre analysis of the Amazon.com discourse community and its “Most Helpful” product reviews. *Discourse, Context & Media*, 2(2):84–93.
- Smart, J. (2006). SMART Controlled English. In *Proceedings of the 5th International Workshop on Controlled Language Applications (CLAW)*, Cambridge, Massachusetts.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of Association for Machine Translation in the Americas (AMTA)*, pages 223–231, Cambridge, Massachusetts.
- Somers, H. (2005). Round-trip translation: What is it good for? In *Proceedings of the 3rd Australasian Language Technology Workshop (ALTA)*, pages 127–133, Sydney, Australia.
- Spaggiari, L., Beaujard, F., and Cannesson, E. (2003). A controlled language at Airbus. In *Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (EAMT/CLAW)*, pages 151–159, Dublin, Ireland.
- Specia, L. (2010). Translating from complex to simplified sentences. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language (PROPOR)*, pages 30–39, Porto Alegre, Rio Grande do Sul, Brazil.
- Specia, L., Cancedda, N., and Dymetman, M. (2010). A dataset for assessing machine translation evaluation metrics. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, pages 3375–3378, Valletta, Malta.
- Spyridakis, J., Holmback, H., and Shubert, S. K. (1997). Measuring the translatability of Simplified English in procedural documents. *IEEE Transactions on Professional Communication*, 40(1):4–12.
- Sun, Y., O’Brien, S., O’Hagan, M., and Hollowood, F. (2010). A novel statistical pre-processing model for rule-based machine translation system. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT)*, Saint-Raphaël, France.
- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27 (NIPS)*, pages 3104–3112.
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, Cambridge.
- Swales, J. M. (2004). *Research Genres: Explorations and Applications*. Cambridge University Press, Cambridge.
- Tatsumi, M. (2010). *Post-editing Machine Translated Text in a Commercial Setting: Observation and Statistical Analysis*. PhD thesis, Dublin City University.
- Tessuto, G. (2015). Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over. *English for Specific Purposes*, 37:13–26.
- Thicke, L. (2011). Improving MT results: A study. *Multilingual*, January/February:37–40.

- Tsuji, K. and Kageura, K. (2004). Extracting low-frequency translation pairs from Japanese-English bilingual corpora. In *Proceedings of the 3rd International Workshop on Computational Terminology (CompuTerm)*, pages 23–30, Geneva, Switzerland.
- Tuldava, J. (1995). *Methods in Quantitative Linguistics*. Wissenschaftlicher Verlag Trier, Trier.
- Uchimoto, K., Hayashida, N., Ishida, T., and Isahara, H. (2006). Automatic detection and semi-automatic revision of non-machine-translatable parts of a sentence. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 703–708, Genoa, Italy.
- Vasconcellos, M. (2001). Terminology and machine translation. In Wright, S. E. and Budin, G., editors, *Handbook of Terminology Management, Vol.2: Application-Oriented Terminology Management*, pages 697–723. John Benjamins, Amsterdam.
- Vilar, D., Xu, J., D’Haro, L. F., and Ney, H. (2006). Error analysis of statistical machine translation output. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 697–702, Genoa, Italy.
- Vintar, Š. (2010). Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation. *Terminology*, 16(2):141–158.
- Voss, C. R. and Tate, C. R. (2006). Task-based evaluation of machine translation (MT) engines: Measuring how well people extract who, when, where-type elements in MT output. In *Proceedings of the 11th Conference of the European Association for Machine Translation (EAMT)*, pages 203–212, Oslo, Norway.
- Warburton, K. (2014). Developing lexical resources for controlled authoring purposes. In *Proceedings of LREC 2014 Workshop: Controlled Natural Language Simplifying Language Use*, pages 90–103, Reykjavik, Iceland.
- Warburton, K. (2015a). Managing terminology in commercial environments. In Kockaert, H. J. and Steurs, F., editors, *Handbook of Terminology*, volume 1, pages 360–392. John Benjamins, Amsterdam.
- Warburton, K. (2015b). Terminology management. In Chan, S.-W., editor, *Routledge Encyclopedia of Translation Technology*, pages 644–661. Routledge, New York.
- 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. (渡邊豊英. 産業日本語プロジェクトの概要: 特許・技術情報の利用性向上のために. 情報管理).
- White, J. S. and O’Connell, T. A. (1994). Evaluation in the ARPA machine translation program: 1993 methodology. In *Proceedings of the Workshop on Human Language Technology (HLT)*, pages 135–140, Plainsboro, New Jersey.
- Wright, S. E. and Budin, G., editors (1997). *Handbook of Terminology Management, Vol.1: Basic Aspects of Terminology Management*. John Benjamins, Amsterdam.
- Wright, S. E. and Budin, G., editors (2001). *Handbook of Terminology Management, Vol.2: Application-Oriented Terminology Management*. John Benjamins, Amsterdam.
- Yasui, H. (2009). *Why Are Municipal Websites Difficult To Use—A Novel Information Dissemination by E-Municipality and E-Government Through “Universal Menu”*. JIJI Press Publication Service, Tokyo. (安井秀行. 自治体Webサイトはなぜ使いにくいのか?—“ユニバーサルメニュー”による電子自治体・電子政府の新しい情報発信. 時事通信出版局).
- Yoshida, S. and Matsuyama, A. (1985). Standardizing Japanese: Standardizing dependency relations and transformation rules. *IPSJ SIG Technical Reports*, NL(31):1–6. (吉田将, 松山晶子. 日本語の規格化: 係り受け関係の規格化とそれへの変換ルール. 情報処理学会研究報告).

- Yoshikane, F., Tsuji, K., Kageura, K., and Jacquemin, C. (2003). Morpho-syntactic rules for detecting Japanese term variation: Establishment and evaluation. *Journal of Natural Language Processing*, 10(4):3–32.
- Yoshimi, T. (2001). Improvement of translation quality of English newspaper headlines by automatic pre-editing. *Machine Translation*, 16(4):233–250.
- Zhang, Z., Iria, J., Brewster, C., and Ciravegna, F. (2008). A comparative evaluation of term recognition algorithms. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, pages 2108–2113, Marrakech, Morocco.
- Zipf, G. K. (1935). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Houghton Mifflin, Boston, Massachusetts.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Hafner, New York.

Index

Note: Page references in *italic* refer to Figures; those in **bold** refer to Tables.

- adequacy 26, 27, 149, 152–153
Adriaens, G. 52
after-scenario questionnaire (ASQ) **148**, 157–158
Ahmad, K. 41
Aikawa, T. 37–38
Alabau, V. 53
Allen, J. 25
ambiguity 22, 29, 30
architecture, machine translation (MT) 20–22; *see also* machine translation (MT)
authoring support system 49–51; MuTUAL 127–128
automatic evaluation metric (AEM) 27–28, 37, 173
automatic rewriting 49
automatic term extraction (ATE) 43, 102, 122; SDL MultiTerm 43; Sketch Engine 43

Baayen, H. 105–106, 115, 121
back translation (BT) 24, 129, *131*, 135
Baroni, M. 118
Bellamy, L. 13
Bernth, A. 32, 34, 37
Bhatia, V. K. 16
Biber, D. 16, 41, 60
BLEU (Bilingual Evaluation Understudy) 27, 28, 53, 173
Brooke, J. 53, 148

Cabr, M. 41–42
Callison-Burch, C. 27
Carl, M. 24, 45, 122
Carroll, J. B. 26
Carroll, T. 3, 9
Castilho, S. 53
Cerrella Bauer, S. 46, 47
checker, controlled language (CL) *see* controlled language (CL) authoring assistant
Christoffersen, E. 44, 45
Cohen, J. 27
Colineau, N. 19–20, 48
communicative goal 12, 13, 16, 173
compatibility of requirements 33, 38, 68, 85, 94
complexity 29, 30
concept 39–40, 44, 113; coverage of 115, 121
confidence score **146**, 147, 160
Conrad, S. 16, 60
controlled authoring 7, 127, 173; controlled document authoring 7, 169; definition 7
controlled language (CL) 23, 28; commonality 34; conformity to 50, 54, 130, 158; definition 28; deployment 36–37; document-level 31–32, 172–173; effectiveness 38, 68; formulation 34–36, 69–71; guideline 87–89, 132–133, *134*;

- human-oriented 33, 38; language properties 29; lexical component 30, 101; machine translation (MT)-oriented 29, 33, 37, 71; parameters 33–34, 71; sentence-level 68; syntactic component 30; violation 23, 52, 54, 132, 138, 147; violation correction 151, **152**; violation detection 50, 51–52, 132, 133, 138, 147, 160
- controlled language (CL) authoring assistant 49–51; ACCEPT portal 50; Acrolinx 49, 50; EasyEnglishAnalyzer 32, 37; KANTOO Controlled Language Checker 50; MAXit Checker 49, 50; MuTUAL 130–134
- controlled language (CL) rule 30, 31, 68–69; context-dependent 91–92; generally effective 77–79; machine translation (MT) dependent 79–81; optimal 81; rewriting trial based (CL-R) 69–71; technical writing based (CL-T) 69
- controlled language (CL) rule set 33–34, 35, 38; AECMA Simplified English 29, 30, 32, 33, 34, 54, 101; Caterpillar Technical English 29; Easy Japanese (*Yasashii Nihongo*) 35, 102; Global English 36; Patent Documents Writing Manual 31, 35; Perkins Approved Clear English 30; PLAIN 29; Simplified Technical Japanese (STJ) 31, 35; SMART Controlled English 29
- controlled terminology 44, 101, 115, 127
- controlled vocabulary 101
- controlled writing 23, 36, 129, 130, 144
- corpus 41–42; aligned 42; bilingual 42; comparable 42; multilingual 42; municipal 102–103; parallel 42; as resource for machine translation (MT) 21–22
- corpus compilation 41–42
- Costa, Â. 28
- coverage 47, 102
- customisation, machine translation (MT) 23, 24–25; dictionary 24, 40–41, 128, 129, 134, 149, 153–154; retraining 24; terminology integration 24–25
- Daille, B. 44, 45, 108, 109, 121, 122
- Dale, R. 12
- Day, D. 13
- decision-making 32, 50, 104, 127, 160
- descriptive document analysis 16–19; functional structure analysis 17–19; genre analysis 16–17; move analysis 17, 19
- Désilets, A. 44
- desired linguistic form 91–92, 95–97
- Dillinger, M. 40
- DITA (Darwin Information Typing Architecture) 13; element 13–14; map 13; specialisation 63–65; Task topic 14, **64**, 129–130; topic 13
- document 7, 12, 172–173
- Document Automation (DA) 48
- document formalisation 12–15, 59
- document framework 13–15
- document structure 12–13; conceptual structure 12–13; IMRD (Introduction, Methods, Results and Discussion) 16; physical structure 12–13
- document template 49, 59, 129–130
- Doherty, S. 35, 53, 144
- domain 34, 39; industry 31, 34; municipal 9; user-generated content (UGC) 34
- domain expert 42–43, 44, 105
- edit distance 155; accumulative edit distance 155–156, **157**; Levenshtein distance 155, 160
- effectiveness: controlled language (CL) 37, 38, 68; system usability 53, 144, **145**, 151–155

- efficiency 52, 53, 144, **145**, 155–157
 English 3, 9, 10
 English controlled language (CL) 35
 English for Specific Purposes (ESP) 17
 errors analysis, machine translation (MT) 28
 evaluation: automatic evaluation 27–28, 37, 173; controlled language (CL) 37–39, 71; document quality 19–20; extrinsic 47; human evaluation 26–27; intrinsic 47; linguistic quality 25; machine translation (MT) output 25–28, 73–75, 149; precision and recall 51–52, 138; system usability 52–54, 144–145; task-based 19–20; terminology 47, 114–115
 Evert, S. 118

 Fischer, M. 40
 fluency 26, 27, 149, 152–153
 Foo, J. 43
 Frantzi, K. 43
 functional document element 13–14, 59–63

 Gdaniec, C. 34
 genre 16
 Gerlach, J. 35, 50
 Guzmán, F. 173

 Hardie, A. 42
 Hartley, A. 31, 32, 33, 34, 37, 38, 82, 150
 Hoard, J. E. 54
 Horn, R. E. 14, **15**
 human translation 3, 25; as gold-standard 26, 27, 74
 Hutchins, J. 20, 25

 information mapping 14, **15**
 inter-rater agreement 27
 ISO (International Organization for Standardization) standard: definition of controlled authoring 7; definition of term 39–40; definition of usability 52–53
 Itagaki, M. 47

 Jacquemin, C. 108–109, 122
 Japanese 3, 9
 Japanese controlled language (CL) 35

 Kageura, K. 39, 42, 47, 114, 115, 121
 Kamprath, C. 29, 34
 Kando, N. 17–19, 60
 Karsch, B. I. 44
 Khmaladze, E. V. 115
 Kikui, G. 47, 114, 115, 121
 Kim, Y. 35
 Kittredge, R. 23, 36
 Koch, G. G. 27
 Kohl, J. R. 36
 Kuhn, T. 28, 29

 Landis, J. R. 27
 Langlais, P. 24
 language pair 22
 learnability 52, 158
 Levenshtein, V. I. 155
 Lewis, J. R. 144, 148
 linguistic pattern 31, 69, 91–92
 linguistic specification 91–92

 machine-translatability 23–24, 33, 37, 38, 68, 86, 87, 174
 machine translation (MT) 20; contextual 8; example-based (EBMT) 6, 20–21; general-purpose 24; Google Translate 3, 10, 70, 73, 95; interactive 53; neural (NMT) 6, 20, 22; off-the-shelf 9, 24, 93; raw output 23, 25, 53, 175; rule-based (RBMT) 6, 20–21; statistical (SMT) 6, 20–21; TexTra 24, 70, 73, 87, 95, 128, 134, 145; The Hon'yaku 73, 95; TransGateway 70, 73, 87, 95, 128, 134, 145
 Macken, L. 52
 Masselot, F. 25, 28

- McEnery, T. 42
 Merkel, M. 43
 Mirkin, S. 24
 Mitamura, T. 36, 50, 51
 Møller, M. H. 44, 45
 morphological analyser 51, 133, 135, 139
 Mossop, B. 12
 multilingualisation 3, 7, 9, 128
 municipal-life information 9, 59, 60, **61**, 102, 174; CLAIR (Council of Local Authorities for International Relations) 59–60, **61**; Hamamatsu City 60, **61**; Shinjuku City 60, **61**
- Nagao, M. 30
 natural language 29
 natural language generation (NLG) 12
 natural language processing (NLP) 6, 43, 51
 Nielsen, J. 52, 53
 non-professional writer 3, 8, 49, 87, 127, 144
 Nyberg, E. 36, 38, 50, 51
- OASIS (Organization for the Advancement of Structured Information Standards) 13
 O'Brien, S. 30, 31, 32, 33–34, 37, 38, 53, 68, 144
 Ó Broin, U. 7
 O'Connell, T. A. 26
 Ogura, E. 31, 35
 out-of-vocabulary (OOV) 22, 24
- Papineni, K. 27
 Paris, C. 31
 parser 51, 141, 151
 Plitt, M. 25, 28
 post-editability 37
 post-editing (PE) 25; effort 28, 37–38
 post-translation processing 94, 98, 135
 precision 51–52, 138
 pre-editing 23–24, 50; *see also* controlled language (CL)
- pre-translation processing 93, 94–98, 135
 procedural document 15; municipal 9, 60, **61**; seal registration procedure 4, 5, 62, 63, 65, 66
 productivity 24, 25, 28, 37, 47, 50
 Pym, P. 30, 34, 37
- quality: document 19–20; machine translation (MT) output 25
 questionnaire: machine translation (MT) quality 73–75; satisfaction with system 53, 148; satisfaction with task 53, 148; source readability 82, 83
- readability *see* source readability
 recall 51–52, 138
 Reiter, E. 12, 19
 representativeness 41
 Resnik, P. 23
 Reuther, U. 33, 38
 Rogers, M. 41
 Roturier, J. 35, 37, 38–39
 Rubens, P. 15
- Sager, J. C. 39, 41, 42, 47
 satisfaction 52, 53, 144, **145**, 157–158
 Sauro, J. 144, 148
 Schriver, K. A. 19
 Seretan, V. 34, 50
 Sichel, H. S. 118
 similar text search 129, 136
 Smart, J. 29
 Somers, H. 20, 25, 135
 source input control 23–24
 source language (SL) 6
 source readability 38, 82–87, 154–155
 source text (ST) 6, 21, 22; controlled language (CL)-compliant 37, 88, 92
 Specia, L. 28
 Spyridakis, J. 38
 sufficiency 42, 47, 102; *see also* coverage
 Swales, J. M. 17, 19, 60
 synset 44, 101

- system usability scale (SUS) 53, **149**, 158
 target language (TL) 6, 33
 target output control 23, 25
 target text (TT) 6, 21, 22
 technical document 31, 34, 59
 technical writing 15, 35, 69
 term: authorised translation 44, 101; bilingual 105, 107, **108**; coverage of 115, 121; municipal 104; preferred 44, 112–113, 132; proscribed 44, 112–113, 132
 term candidate 43, 105
 term extraction 41, 42–44; automatic 43; manual 42–43, 104, 105
 term variation: definition 44; extent 44, 109; typology 44–45, 108–112
 term variation management 41, 44–47, 112–114; criteria **112**
 terminologist 42, 44
 terminology: consistency 40, 44, 47; database 43, 129; definition 39; evaluation 47; growth 119–121; population size 117–119
 terminology management 39; collaborative 43, 44, 104; descriptive 40, 102; normative 40; prescriptive 40, 102
 text similarity 160–162
 Thicke, L. 24–25, 37
 translatability 37–38; human-translatability 38; machine-translatability 23–24, 33, 37, 38, 68, 86, 87, 174
 translation 3, 4
 translation memory (TM) 43, 136
 Uchimoto, K. 23–24
 Umino, B. 42
 uncontrolled terminology 115
 understandability 73
 univocity 44
 unpredictability 93, 94
 usability 52–53; evaluation 53–54, 144–145
 use scenario, controlled language (CL) authoring assistant 51, 131–133, 144
 Vasconcellos, M. 24
 Vilar, D. 28
 Warburton, K. 40, 44, 46, 47, 108, 121
 White, J. S. 26
 Yoshikane, F. 45, 108, 122
 Zhang, Z. 102
 Zipf's law 106–107

Taylor & Francis Group
an informa business

Taylor & Francis eBooks

www.taylorfrancis.com

A single destination for eBooks from Taylor & Francis with increased functionality and an improved user experience to meet the needs of our customers.

90,000+ eBooks of award-winning academic content in Humanities, Social Science, Science, Technology, Engineering, and Medical written by a global network of editors and authors.

TAYLOR & FRANCIS EBOOKS OFFERS:

A streamlined
experience for
our library
customers

A single point
of discovery
for all of our
eBook content

Improved
search and
discovery of
content at both
book and
chapter level

REQUEST A FREE TRIAL
support@taylorfrancis.com

Routledge
Taylor & Francis Group

CRC Press
Taylor & Francis Group