MASTERARBEIT | MASTER'S THESIS

Titel | Title

Productivity and Effort in Intralingual Translation: A Comparative Process-oriented Study of Human Text-Simplification and MT/LLM-assisted Text-Simplification into Easy Japanese in Crisis Settings

verfasst von | submitted by

Ryunosuke Sannomiya

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 070 378 331
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Translation Japanisch Deutsch
Betreut von | Supervisor:	Univ.-Prof. Dragos Ioan Ciobanu PhD

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
BGG	Behindertengleichstellungsgesetz
BLEU	Bilingual Evaluation Understudy
CAT tool	Computer-assisted (-aided) Translation tool
CEFR	Common European Framework of Reference for Languages
DRM	Disaster Risk Management
GPT	Generative Pre-trained Transformer
GTM	General Text Matcher
HTER	Human-targeted Translation Edit Rate
ICT	Information and Communication Technology
IPSS	Institute of Population and Social Security Research
JLPT	Japanese-Language Proficiency Test
LLM	Large Language Model
MT	Machine Translation
MTPE	Machine Translation Post-Editing
NASA-TLX	NASA Task Load Index
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NMT	Neural Machine Translation
PCR	Pause-to-Character Ratio
PWR	Pause-to-Word Ratio
RBMT	Rule-based Machine Translation
SMT	Statistical Machine Translation
TAPs	Think-aloud protocols
TER	Translation Edit Rate
TWB	Translators without Borders
UN CRPD	UN Convention on the Rights of People with Disabilities
WER	Word Error Rate

Abstract (English)

This Master's thesis investigates the applicability of Machine Translation (MT) and Large Language Models (LLMs) for simplifying text into Easy Japanese in crisis settings. Easy Japanese is in high demand among non-native speakers in Japan, often surpassing the demand for English. However, public servants often lack the necessary manpower, expertise and training, leading them to rely on MT tools for simplification.

Based on prior research demonstrating productivity and quality gains from MT Post-Editing (MTPE) in interlingual translation, the following research questions arise: “What is the impact of working with MT and LLMs' output in a Japanese intralingual text simplification task on the **temporal, technical, and cognitive effort (RQ1)** of citizen translators with Japanese as their native language and what is the impact on the **readabilityof the target text (RQ2)**?”

The data for the three types of effort (temporal, technical, and cognitive) will be collected using a keylogging tool (Translog-II)**,** which records duration, pauses, and keystrokes. In addition, subjective cognitive effort will also be assessed using the NASA-TLX questionnaire, and text quality will be evaluated using the Japanese readability tool (jReadability). Three comparable crisis-related texts will be used in the experiment: one processed with the MT tool called “Tsutaeru-Web (伝えるウェブ)”, one with an LLM (Gemini) trained on Easy Japanese guidelines, and one translated by humans without machine assistance.

The findings show a positive and negative impact of working with MT and LLMs in a Japanese intralingual translation task. While high-grade gains in productivity in the Post-Editing process and readability of the target text can be reached using an LLM, the productivity gain of working with an MT was moderate, and the readability of the target text decreased compared to Human Text-Simplification.

Furthermore, based on the collected data, the discussion suggests that the readability of source texts and machine-translated texts can affect the grade of each effort in the MT/LLM-assisted Text-Simplification stage. This indicates that as the readability of the source texts and the MT/LLM outputs increases, the temporal, technical, and cognitive effort required for MT/LLM-assisted Text-Simplification decreases. Moreover, it can be said that less readable source texts have a negative impact on the quality of MT/LLM outputs, requiring a greater degree of Post-Editing effort.

Abstract (Deutsch)

Diese Masterarbeit erforscht die Anwendbarkeit maschineller Übersetzung (MÜ) und großer Sprachmodelle (LLMs) für die Textvereinfachung in das Leichte Japanisch in der Krisenkommunikation. Es besteht ein hoher Bedarf an Leichtem Japanisch unter Nicht-MuttersprecherInnen in Japan und übertrifft oft die Nachfrage nach Englisch. Beamten fehlen jedoch oft die erforderlichen Arbeitskräfte, Fachkenntnisse und Schulungen, was dazu führt, dass sie sich auf die MÜ-Tools für die Textvereinfachung verlassen.

Basierend auf früheren Forschungen, die die Zunahme der Produktivität sowie Qualität von Post-Editing der MÜ (MTPE) in interlingualen Übersetzungen zeigen, stellen sich die folgenden Forschungsfragen: „Welche Auswirkungen hat das Arbeiten mit Outputs von MÜ und LLMs beim japanischen intralingualen Übersetzen auf den **temporalen, technischen und kognitiven Aufwand (RQ1)** von Citizen Translators mit Japanisch als Muttersprache und welche Auswirkungen auf die **Lesbarkeit der Zieltexte (RQ2)**?“

Die Daten der drei Arten von Aufwand (temporal, technisch und kognitiv) werden anhand von einem Keylogging-Tool (Translog-II), das Dauer, Pausen und Keystrokes aufzeichnet, gesammelt. Außerdem wird der subjektive Aufwand mithilfe des NASA-TLX-Fragenbogens bewertet und die Textqualität anhand des japanischen Lesbarkeit-Tools (jReadability) beurteilt. Drei vergleichbare krisenbezogene Texte werden für das Experiment herangezogen: einer übersetzt von einem MÜ-Tool namens „Tsutaeru-Web (伝えるウェブ)“, der zweite übersetzt von einer LLM (Gemini), die zuvor mit Richtlinien zu Leichtem Japanisch trainiert worden ist, und der dritte ohne jegliche maschinelle Übersetzung.

Das Ergebnis zeigt positive und negative Auswirkungen auf das Arbeiten mit MÜ und LLMs bei einer japanischen intralingualen Übersetzungsaufgabe. Während mithilfe der LLM eine hohe Zunahme der Produktivität im Post-Editing-Prozess und der Lesbarkeit des Zieltextes erreicht werden konnte, war die Produktivität bei der Arbeit mit der MÜ mittelmäßig und die Lesbarkeit des Zieltextes hat sich im Vergleich zur Human-Textvereinfachung verschlechtert.

Des Weiteren weist die Diskussion darauf hin, dass die Lesbarkeit des Ausgangstextes und des maschinell übersetzten Textes den Grad jedes gemessenen Aufwands in der MT/LLM-gestützten Textvereinfachung beeinflussen kann. Dies zeigt, je höher die Lesbarkeit des Ausgangstextes und der MÜ-Outputs ist, desto mehr reduziert sich der verlangte temporale, technische und kognitive Aufwand für MT/LLM-gestützte Textvereinfachung. Weiters haben

wenig lesbare Ausgangstexte eine negative Auswirkung auf die Qualität der MÜ-Outputs und fordern ein größeres Maß an Post-Editing-Aufwand.

1 Introduction

In this master's thesis, a comparative process-oriented study of the productivity and effort of Post-Editing of Machine Translation (MT) and Large Language Model (LLM) and Human Text-Simplification of Japanese texts in crisis settings will be conducted. There are three categories of translation: interlingual, intralingual and intersemiotic translation, and all of them can be a subject of translation studies. Text simplification, in the sense of translation into Easy Language, is positioned under the concept of intralingual translation (Roman 1959). The main focus of this study is to investigate and compare the effort, productivity and quality, especially readability of Human Text-Simplification and MT/LLM-assisted Text-Simplification into Easy Japanese. Especially, the translation from scratch in Easy Language will be referred to as "Human Text-Simplification", and the Post-Editing process of MT/LLM output in Easy Language will be referred to as "MT/LLM-assisted Text-Simplification" in this thesis. Furthermore, crisis translation will be a parallel theme of this research because crisis settings are used as context, and volunteer participants will be performing the tasks as citizen translators. The definition of crisis translation is described as follows:

> Citizen Translation is defined as translation practice conducted, sometimes voluntarily, sometimes by people who are volunteered, by an individual, or a community of individuals who may be trained or untrained linguists. The translational activity is conducted with the assumption of achieving a common good and may be paid or unpaid. (Federici & Cadwell 2018: 22)

The details of citizen translators will be discussed in a later section. In this respect, it can be argued that the subject of the study is highly specialised, yet it also exhibits multidisciplinary characteristics. Following this, the research questions, state of research and methodologies used in this study will be described. Then, findings will be presented and analysed, and finally, conclusions and discussion will be drawn.

2 Theoretical Background and Literature Review

Several varieties of text simplification have been documented, such as Plain Language and Easy Language Plus. However, in this paper, Easy Language, in particular Easy Japanese, will be focused on, which is not very easy to define, not only in terms but also in concepts. While there is a relatively clear difference between Easy Language and Plain Language in German, English and many other languages, research and papers which mention the differences between Easy

and Plain Language in the Japanese language can hardly be found. However, the concept and the term of "Yasashii Nihongo (やさしい日本語)", also known as "Easy Japanese", is widely known as a Japanese text simplification. Nevertheless, Yasashii Nihongo does not have an official English equivalent yet. However, the term "Easy Japanese" will be used in this paper as a matter of convenience in order to distinguish the differences between Easy and Plain Language.

On the one hand, capito (n.d.), an Austrian company from Graz offering Easy Language and accessibility services, indicates that Easy Language targets the A2 to A1 level of the Common European Framework of Reference for Languages (CEFR) and Plain Language targets approximately the B1 level. On the other hand, according to a guidebook of Yokohama City (2017), Easy Japanese targets the N4 level of the Japanese-Language Proficiency Test (JLPT) (2012), which corresponds approximately to the A2 level of the CEFR. For this reason, Easy Japanese can be categorised as Easy Language, and will be chosen as the focus of this thesis. However, as it is useful to know about both easy and plain language in order to make better decisions, literature from both disciplines will be referenced and discussed in this thesis.

2.1 Easy Language and Plain Language as Accessibility

According to Article 9 of the UN Convention on the Rights of People with Disabilities (UN CRPD), "Accessibility" is described as follows:

> To enable persons with disabilities to live independently and participate fully in all aspects of life, States Parties shall take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, to the physical environment, to transportation, to information and communications, including information and communications technologies and systems, and to other facilities and services open or provided to the public, both in urban and in rural areas. (UN CRPD 2008)

As can be understood from this definition, eliminating barriers to information and communication is a part of accessibility, and offering Plain Language is listed as a measure of that by the UN CRPD (Maaß 2020: 49). Article 9 also claims that national public institutions shall take the initiative to realise the accessibility. For example, under the Obama administration in the United States, the Plain Writing Act of 2010 came into effect, which ensures that Government documents that are for the public must be written in a way that is not difficult to

understand. At the time, PlainLanguage.gov[1] was a tool for government communicators to support the Plain Writing Act of 2010; however, since the 1st of October 2025, the website has not been updated under the Trump administration.

Plain Language Act 2022 was also enacted in New Zealand. The Act is supposed to make public service agencies and Crown agents more effective and accountable, and also make it easier for the public to get hold of certain documents. As a reason for enacting the law, the New Zealand government emphasises that a mere 16% of the adult population of New Zealand is considered to have attained a high level of literacy. Furthermore, the government mentions not only the level of literacy, but also the unfamiliarity with the subject and the limitation of computer skills. The government claims that offering information in Plain Language can contribute to enhancing transparency and fostering trust (Department of Internal Affairs Government Information Services 2023).

Moreover, besides Plain Language (Einfache Sprache), Easy Language (Leichte Sprache) is also offered as part of communication accessibility according to the German **B**ehinderten**g**leichstellun**g**sgesetz (BGG 2002) (Maaß 2020: 49). The ratification of the UN CRPD in 2009 supported the engagement of requiring "Easy Language" in Germany at that time (Kimura 2023: 61). Moreover, Easy Language and Plain Language are becoming a discipline under the Translation Studies in universities, such as the University of Hildesheim, and it is on the process of the professionalism (Kimura 2023: 63; Sugatani 2020). This phenomenon is also creating employment in the industry.

Although there are some exceptions, such as the current situation in the U.S. as shown above, the movement of introducing Plain Language and Easy Language as an accessibility service is moving forward, and this trend is also no exception in Japan. In 2019, the Japanese government introduced the guidelines of Easy Japanese, which are used across the governmental organisations. The Ministry of Internal Affairs and Communications firmly established its policy to make administrative and community information for foreigners in Japan more accessible by introducing multilingual support and Easy Japanese. However, despite the policy and guidelines, no acts have been established yet that are legally binding (Kimura 2023: 60–61). From the legal perspective, Japan and Easy Japanese are behind the other countries. However, there should be a reason why no legally binding act of Easy Language has been established. In the next section, the history of Easy Japanese and Japanese society will be discussed to find a possible answer to the gap between Japan and others regarding the

[1] https://digital.gov/guides/plain-language (accessed: 03.05.2026)

legislation of text simplification, while comparing mainly the history of Easy Japanese and Easy German.

2.2 Differences between Easy Language and Plain Language

In this section, differences between Easy Language and Plain Language will be discussed. Both Easy and Plain Language are varieties of natural languages, and they share the following characteristics: linguistic features are reduced on the word, sentence and text levels (Maaß & Schwengber 2022: 44). However, as briefly introduced above, Easy Language targets the A2 to A1 level, while Plain Language targets approximately the B1 level. Beyond the target level, the section describes the detailed differences between Easy and Plain Language.

First, on the one hand, Easy Language has a strict set of regulations, which are largely based on the rules of Netzwerk Leichte Sprache e.V. (Netzwerk Leichte Sprache e.V. 2022). On the other hand, for Plain Language, there is ISO 24495-1:2023 "Plain language Part 1: Governing principles and guidelines" (ISO 2023), which provides only recommendations for Plain Language. Easy Language "aims for maximal comprehension of texts and equality of interaction" (Lindholm & Vanhatalo 2021: 19), while Plain Language "aims for wording, structure, and design so clear that the intended audience can easily find what they need, understand what they find, and use that information" (Lindholm & Vanhatalo 2021: 19).

Second, Easy Language is a linguistic resource designed to facilitate the comprehension of written content for individuals with communication impairments, especially with cognitive disabilities and individuals affected by conditions such as aphasia, dementia, autism, and severe to moderate forms of illiteracy. In comparison, the target audience of Plain Language is everyone, and its objective is to facilitate access to the expert context for non-expert users (Maaß & Schwengber 2022: 45–46).

Third, the Easy Language guidelines regulate not only grammatical rules, such as word structure, vocabulary, and sentence structure, but also visual and medial design of the texts, such as format, paper quality, font and font style, type-size, use (or rather: ban) of quotation marks, line and page design and use of imagery (Maaß 2020: 76). Hoewever, any striking layout features is not regulated in Plain Language, but in principle, good layout is necessary as defined in the respective international norms (Maaß & Schwengber 2022: 46). Table 1 illustrates the differences of characteristics between Plain and Easy Language.

Table 1: Comparison of approximate characteristics of Plain Language and Easy Language (adapted from Lindholm & Vanhatalo 2021: 19–20)

	Plain Language	Easy Language
Purpose	Aims for wording, structure, and design so clear that the intended audience can easily find what they need, understand what they find, and use that information	Aims for maximal comprehension of texts and equality of interaction
Target audience	Everyone	People with various linguistic limitations (e.g., neuro-biological reasons, immigrant background, low literacy)
Level of complexity	Seeks simple words, structures, and constructions, avoids jargon	Deliberately and often heavily simplified at all levels: content, grammar, lexicon
Usage area	Legal or administrative texts, official documents	All purposes: informative texts, fiction and non-fiction literature, news texts, also oral communication
Legal status	In many countries, laws mandate that public agencies use Plain Language to increase access to services	No legal status in most countries, Germany an exception
Producers	Mainly authorities, companies, organizations	The more skilled party in any asymmetric communication situation

2.3 History of Easy Japanese

Japan is widely acknowledged for its high incidence of natural disasters. It is estimated that Nankai megathrust earthquakes can happen with an 80% probability, and Greater Tokyo Area earthquakes can occur with a 70% probability within the next 30 years, both of which should cause massive human casualties and economic damage (Disaster Management, Cabinet Office 2025). Indeed, Japan has experienced over 80 major natural disasters in the past 20 years; however, Japan is well equipped to manage such crises and ranks at low and stable risk in the 2021 INFORM Index for Risk Management, placing 144 out of 191 countries (Cadwell 2023: 18). Moreover, A number of significant events have exerted influence on crisis translation and interpreting policy. For instance, the Great Hanshin-Awaji Earthquake in 1995, in which language support was deficient, is a pertinent case study.

The history of Easy Japanese goes back to the Great Hanshin-Awaji Earthquake in 1995, which hit the western regions of Japan and took a heavy toll not only on Japanese but also on non-Japanese victims. At that time, the percentage of the casualties of foreigners was around 2%, while the Japanese casualties were about 1% of the total victims (Immigration Services

Agency of Japan 2020), which was caused by language barriers and a high number of foreign factory workers and students living in buildings where there was a high risk of fire (Cadwell 2023: 19). In short, the casualties of foreigners were double that of the Japanese. After the crisis, researchers analysed the background and reason for this tragic reality and concluded that non-Japanese residents could not acquire enough disaster information related to evacuation and infrastructure, among others, compared to Japanese residents because of their insufficient Japanese comprehension ability. This led to many more non-Japanese casualties compared to Japanese ones (Sato et al. 2004). From this background, Sato, a professor at Hirosaki University, and his colleagues introduced Easy Japanese, and through its practical use in the Chūetsu earthquakes in 2004 and the aftermath of the 2011 Tōhoku earthquake and tsunami, Easy Japanese is nowadays widely used also in different contexts, such as the medical scene and tourism (Takeda et al. 2021). Based on studies by Sato and Iori, several public institutions and local bodies, such as the council of Mie Prefecture, publish guidelines about Easy Japanese (Mie Prefecture 2025).

2.3.1 Comparison of Easy languages

It is apparent that Easy Japanese has been introduced in the context of crisis communication to reduce casualties, especially for foreigners in Japan, and it provides information about disasters in a more accessible and easier-to-understand way. On the other hand, the introduction of Easy/Plain Language in Germany originated from another context. The history of Easy/Plain Language was originally influenced by the "Easy Read" activism in 1974 in the U.S., which was initiated by an organisation called "People First" to provide easy-to-understand information for people with learning disabilities. This movement arrived in Germany in the late 90s, and an Easy German dictionary was published in 2000. After that, "Mensch zuerst − Netzwerk People First Deutschland", which supports people with learning disabilities and "Netzwerk Leichte Sprache", which plays a central role in the movement of Easy Language, were established in the early 2000s (Kimura 2023: 54).

These backgrounds and history of Easy language between Japan, Germany and the U.S. contrast the differences of the target groups. On the one hand, Easy German mainly focuses on accessibility for people with learning disabilities. On the other hand, Easy Japanese primarily focuses on accessibility for non-native Japanese speakers and foreign residents in Japan. Kimura (2023: 53) claims that this difference can be seen from the construction of the websites of public organisations in Japan and Germany. According to his research, the option of Easy

German is usually located under the tab of “barrier-free” together with sign language. In contrast, the option of Easy Japanese is generally located under the tab of multilingual support. From this example, it can be understood that the target groups of Easy Language between these countries are fundamentally different.

Furthermore, Kimura (2023: 57) claims that Easy Japanese puts emphasis on another point compared to Easy German. In Germany, the formation of Easy German is strictly defined; therefore, an examination of whether texts in Easy German follow the instructions is the most important point, and it should be done by an expert, in other words, a person who is directly concerned and benefits from it. However, in Japan, it is desired that not only experts or researchers, but also ordinary citizens as members of society should be involved in creating accessibility with Easy Language. For those reasons, it can be comprehended that educational and training institutions for Easy/Plain Language were well established in Germany. On the other hand, Japanese public officers are being confronted with text simplification tasks due to the lack of knowledge and training, as is described in the following section. Moreover, Sugatani (2020) argued that the initiative from the side of the person concerned was insufficient in Japanese society, and more active participation from them is required. It can therefore be surmised that the lack of such an action results in the delay of legislation in Japan.

Additionally, researchers of Easy Japanese set some goals with the introduction of Easy Japanese so that non-native Japanese speakers can achieve a high level of Japanese language proficiency as early as possible, participate in Japanese society on an equal footing with native Japanese people and take proactive action (Masden 2022: 205). According to the researchers, Easy Japanese is designed with the aim that foreigners can proactively acquire a higher level of Japanese step by step. Moreover, it is proposed that establishing Easy Japanese as a common language would facilitate more effective communication with Japanese citizens and thereby ensure equal citizenship (Masden 2022: 205–206). While another critical linguist argues that the proliferation of Easy Japanese leads to a disregard for multilingualism, the advocates of Easy Japanese claim that it is not realistic for each local authority to offer information in all languages, including minority languages (Masden 2022: 209). They highlight that, in contrast to the utilisation of minority languages that are characterised by a paucity of resources for information, Easy Japanese is a program designed to support foreign citizens in their integration into Japanese society, rather than being confined to isolated communities (Masden 2022: 210).

2.3.2 Statistics and Demand for Easy Japanese

Japan is sometimes seen as a monocultural and monolingual society. However, the number of foreign residents in Japan is growing from year to year, and the Japanese Immigration Services Agency reports that there are approximately 3.6 million foreign residents as of 2024. The variety of their nationalities includes 196 countries and regions (Immigration Services Agency 2024). Although the actual level and speed of multiculturalism in Japan is not reaching that of Europe or the United States, it is steadily advancing. On this account, 56 out of 71 organisations in Tokyo are facing a great difficulty in dealing with multilingual support (Bureau of Citizens and Cultural Affairs 2020: 5). Therefore, they are trying to utilise Easy Japanese to publish information. Nevertheless, it must be acknowledged that public servants are confronted with limitations concerning translation into Easy Japanese due to a lack of manpower, knowledge, training and education. For this reason, they are attempting to use software and websites offering MT for text simplification (Bureau of Citizens and Cultural Affairs 2020). Additionally, in crisis settings in Japan, information in Easy Japanese is more in demand than in English by non-Japanese native speakers in the ration of 76 to 68 people (Tokyo International Exchange Committee 2018: 42). It can be claimed that the difference in the demand between Easy Japanese and English is very slight, but those people expect translation by native English speakers. On the other hand, the demand for translation by non-native English speakers and MT is extremely low, which can happen very often in crisis settings, where it is not easy to summon native English speakers. The reason why Easy Japanese is in higher demand can be that the nationalities of the foreign residents who dominate the ranking, such as China, Korea and Vietnam, do not have English as their first language. For this reason, they prefer Easy Japanese to English if their first language is not available for the information in case of crisis settings. In a survey, 126 out of 415 people wished for Easy Japanese while only 36 of them asked for English. The rest of them, 253 people, desire their first language, which is, however, much more difficult to offer (Long 2012: 186).

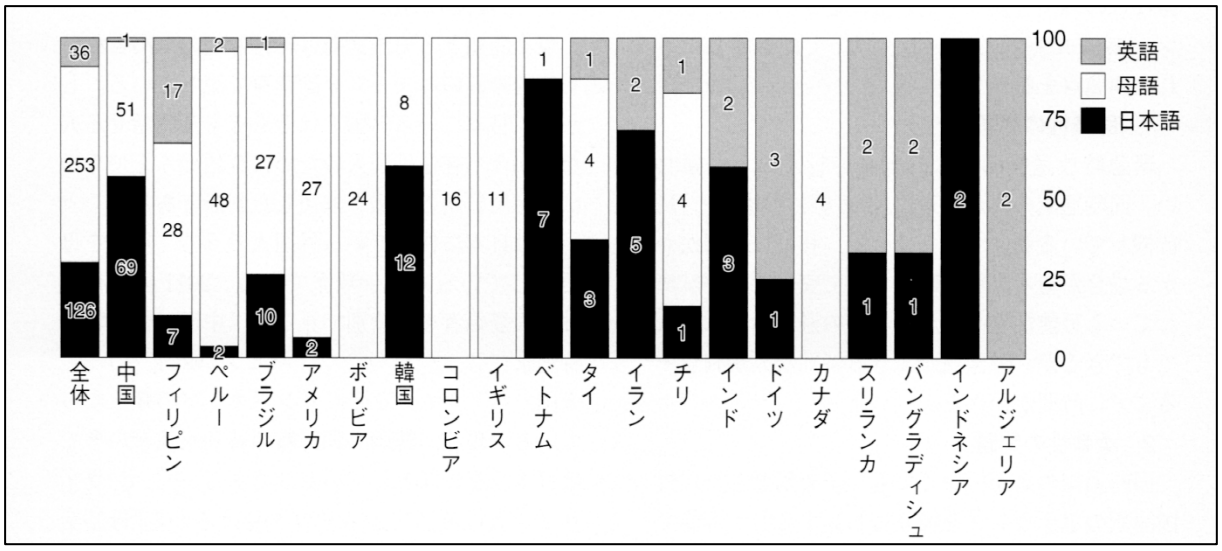


Figure 1: Languages selected by foreign residents for post-disaster consultations (adapted from Long 2012)[2]

Another survey also reports that 61.7% of the interviewed foreign residents choose Japanese as a language in which they can live without any problem, while just 36.2% of them choose English (National Institute for Japanese Language and Linguistics 2009: 4).

Furthermore, older adults also benefit from Easy Language (Gutermuth 2020; Maaß 2020; Mie Prefecture 2025), which is extremely important in Japan's super-ageing society. In Japan, the proportion of the population aged 65 and over in the total population has been rising consistently since 1950 (4.9%), reaching 10% in 1985, over 20% in 2005 and a record high of 29.3% in 2024. According to estimates by the National Institute of Population and Social Security Research (IPSS), this proportion is expected to continue to rise, reaching 34.8% in 2040 and 36.3% in 2045, when the generation born during the second baby boom (1971-1974) will be aged 65 and over (Ministry of Internal Affairs and Communications 2024: 3). A comparison of the proportion of the population aged 65 and over to the total population in 2024 reveals that Japan (29.3%) has the highest proportion of the population aged 65 and over in the world (out of 200 countries and territories with a population of 100,000 or more). Furthermore, Japan has the highest proportion of the population aged 65 and over in major countries, both in the 65-74 and 75+ age groups (Ministry of Internal Affairs and Communications 2024: 4). As can be interpreted from those statistics, Easy Japanese stands at an advantage over English,

[2] English, First Language, Japanese. The countries' names are listed from left to right: Overall, China, the Philippines, Peru, Brazil, the U.S., Bolivia, South Korea, Colombia, the U.K., Vietnam, Thailand, Iran, Chile, India, Germany, Canada, Sri Lanka, Bangladesh, Indonesia and Algeria.

because not only non-native Japanese foreign residents desire Easy Japanese, but also elderly people with Japanese as their first language can profit from Easy Japanese at the same time.

2.4 Translation Effort and Correlation with Readability

The present section will address the productivity and effort in previous process-oriented studies, which will be the central theme in this intralingual translation research. Moreover, literature on the correlation among various measures of translation effort and between objective effort and subjective rating, as well as the relationship between text readability and translation effort, will be reviewed, which will be discussed after data collection and analysis in a later chapter.

2.4.1 Effort and Productivity

As mentioned above, public officers encounter difficulty in translating into Easy Japanese due to a lack of experience, know-how and training under the existing circumstances. Therefore, they rely on software and websites as an aid; nonetheless, only 21.2% of the interviewed organisations are currently using such tools (Bureau of Citizens and Cultural Affairs 2020: 9). In such circumstances, Post-Editing of MT and LLM output has the potential to resolve issues relating to insufficient manpower and experience. This is based on the findings of numerous studies, which demonstrate that interlingual MT Post-Editing increases productivity and requires less effort compared to Human Translation (O'Brien 2007; Stasimioti & Sosoni 2020b). Jiang (2023) reports that translating with LLMs is also superior to traditional Human Translation in both speed and quality.

Regarding productivity, Krings (2001) and O'Brien (2011) claim that time effort alone is not enough to indicate productivity in Human Translation and Machine Translation Post-Editing (MTPE). Hence, Krings (2001: 178–182) introduced a model of three-dimensional effort, namely temporal effort (time spent), cognitive effort (mental processing), and technical effort (physical action). Furthermore, Lacruz claims that:

> Krings' classification of post-editing effort into these three categories applies just as well to translation itself or the editing of human translations. However, as Krings points out, these different types of effort are by no means mutually exclusive. Krings views cognitive effort and technical effort as contributors to temporal effort. (Lacruz 2017: 387)

This statement suggests that the Krings' Model can be applied not only to the investigation of MT/LLMs' Post-Editing effort but also to the effort of translation from scratch. Moreover, the

three types of effort are not independent of one another, but are interconnected. However, the relationships between cognitive, temporal, and technical effort in translation are complex. Conversely, high cognitive effort can result in high temporal effort with minimal technical effort, while translating long but simple texts may involve high temporal and technical effort but low cognitive effort. The nature of these relationships is influenced by individual differences among translators with different levels of expertise or stylistic preferences (Lacruz 2017: 387).

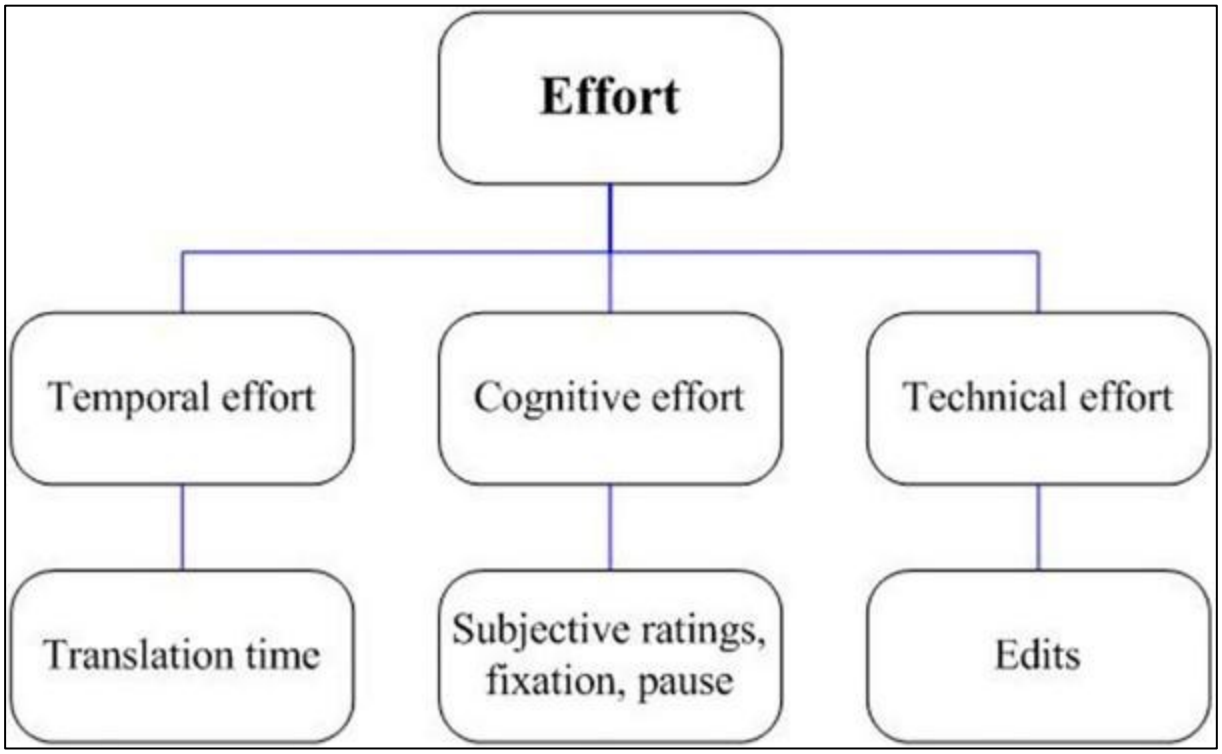


Figure 2: Three dimensions of effort (adapted from Cui et al. 2023)

First, Tatsumi's (2009) study shows that MTPE is faster than Human Translation; however, she mentions that it can depend on the source text characteristics and MT errors. An empirical study on AI-assisted translation of German healthcare texts into Plain Language shows that Post-Editing significantly reduces the time required for translations (Kröger et al. 2025). The time taken for MTPE and HT is one of the simplest indicators for the productivity of both tasks. Therefore, researchers have measured the time spent in order to investigate the temporal effort so far in their process-oriented studies (Koponen et al. 2019). They concluded that MT Post-Editing is faster than Human Translation, and therefore, it is more productive (Elming et al. 2014). As described above, the time spent is the most commonly used indicator and a simple method for investigating the effort of MTPE and HT; nevertheless, according to Krings (2001), the cognitive effort and the technical effort correlate with the temporal effort,

and therefore, it is necessary to measure the other two components of the effort beside the temporal effort (Lacruz 2017).

Second, in the matter of the technical effort for Post-Editing, Human-targeted Translation Edit Rate (HTER) is commonly used. These metrics show how many words have been changed (addition, deletion and replacement) between the machine-translated and post-edited versions; therefore, they reflect the technical effort required (Koponen 2016). In theory, the formula of HTER is mathematically expressed as follows:

$$HTER = \frac{\#\ of\ required\ edit}{\#\ of\ reference\ words}$$

However, the metrics can only measure the shortest way to the end product, and therefore, the full physical action during the Post-Editing process will be ignored (Koponen 2012). For the reason above, it is also mentioned that the utilisation of keylogging software to monitor the implementation of changes provides a more accurate assessment of the technical effort involved (Koponen 2016: 140). A study shows that fewer keystrokes are required in MTPE toward Human Translation (Elming et al. 2014).

Third, contrary to the temporal and technical effort, it is not possible to measure the cognitive effort directly. On the one hand, numerous studies have investigated the cognitive effort using eye-tracking (Elming et al. 2014; Koglin 2015), screen recording (Zhaohui Wang 2019), or think-aloud protocols (TAPs) (Lörscher 1991), or a combination of these methods. In previous research, Krings used TAPs; however, there are also critical views that this method can slow down the translation process (Koponen 2016: 141). On the other hand, the Pause-to-Word Ratio (PWR) is also widely recognised as an indicator in the translation and Post-Editing process studies to investigate the cognitive effort using a keylogging tool combined with keystrokes investigation for technical effort (Lacruz et al. 2012, 2018; O'Brien 2006). The PWR, the ratio of the number of pauses to the number of words (for Japanese texts, the ratio of the number of pauses to the number of Japanese characters, Pause-to-Character Ratio (PCR)), is one of the simplest metrics which is commonly adopted in the investigation of cognitive effort research. Each researcher uses different pause thresholds, but 300 to 5,000 milliseconds is the commonly used range. However, Lacruz and Shreve (2014) found that during the MTPE process, a number of short pauses appear in several segments in which high cognitive effort is demanded, and therefore, a pause threshold with a short range, such as 300 milliseconds, is suitable for the investigation because it can follow the cognitive effort of the Post-Editing and translation process more precisely. Wang (2023) reports in her paper that the PWR in Post-

Editing was considerably lower than that of Human Translation. From the findings of those studies, it can be deduced that MTPE is more productive and demands less effort than Human Translation under certain conditions.

Additionally, cognitive effort can be measured by subjective ratings. On the one hand, some researchers claim that subjective ratings are not reliable indicators (Moorkens et al. 2015); on the other hand, however, some studies conclude that subjective measures can provide an overall picture of workload, and the subjective ratings of translators are closely linked to objective measures (Cui et al. 2023; Liu & Cui 2024; Vieira 2017). For this reason, the combination with objective measurements is necessary to obtain reliable data.

However, the quality of AI tools' outputs, language pairs, text types, and the post-editors' familiarity with the tools and individual working processes can influence the productivity gain (Koponen 2016: 136). Therefore, Koponen (2012: 181, 2016: 132) claims that editing poor MT outputs can be unproductive compared to translation "from scratch".

2.4.2 Correlation of Translation Efforts and Readability

As discussed above, the three types of effort are not independent of one another; rather, they are interconnected. For this reason, in several process-oriented studies on translation effort and productivity, the correlation among objective measures as well as between objective and subjective measures is also discussed. In the domain of interlingual translation, while one study suggests large variations and only weak correlations between temporal effort and technical effort (Macken et al. 2020), other research shows instead that there is a strong correlation in Human Translation and Post-Editing tasks (Moorkens et al. 2015). In particular, Alvarez et al. (2020) highlight that temporal effort and keystrokes are directly related.

Lacruz mentions the relationship between temporal effort and objective cognitive effort as follows: "[...] there is a good deal of evidence that the total time taken and the time per word both correlate with cognitive effort in translation tasks, [...]" (Lacruz 2017: 395). Additionally, she also points out that "[...] cognitive effort is a significant contributor to temporal effort [...]" (Lacruz 2017: 395). Furthermore, Alvarez et al. (2020: 2) also claim that cognitive effort correlates with temporal effort in the interlingual Post-Editing process.

In addition to the correlations between temporal effort and technical effort, and between temporal effort and objective cognitive effort, Cui et al. (2023) also examined the relationship between technical and objective cognitive effort, and found a strong correlation between these objective measures.

Previous research mentioned above has highlighted that there is a positive correlation among objective measures such as duration, keystrokes, and pauses. Moreover, as Cui et al. (2023: 12) observe, while objective measures are subject to inter-rating with each other, objective and subjective measures demonstrate correlative relationships as well.

Besides the discussion on the correlation between different types of translation effort, scholars analyse the relationship between effort and text quality or text readability. Researchers point out that although the level of impact on temporal effort differed depending on sentence structures, it can take longer to translate very short or very long sentences (Koponen 2016). Tatsumi mentions that "various conditions including source text characteristics and MT errors may affect PE speed" (Tatsumi 2009: 7). Additionally, Dai & Liu (2024) investigated the correlation between the Post-Editing effort and source text readability, and they concluded that the impact of readability on cognitive effort is significant. In contrast, the consequences for the temporal and technical aspects are not as significant. Moreover, high NMT quality may have a mitigating effect on the impact of readability.

Furthermore, Tatsumi (2009) conducted a study on the correlation between Post-Editing speed and automatic evaluation metric scores, such as TER (Translation Edit Rate), BLEU (Bilingual Evaluation Understudy), GTM (General Text Matcher) and NIST (National Institute of Standards and Technology), and reported that the relationship between Post-Editing speed and textual differences measured by automatic metric scores may not always be linear. In a later study, Sanchez-Torron and Koehn (2016) researched MT quality and Post-Editing productivity using the BLEU score in the interlingual translation discipline, and they, however, found a linear relationship between the quality of MT and Post-Editing speed. This study shows that machine-translated texts with a higher BLEU score require less Post-Editing time, indicating that higher-quality MT outputs require less temporal effort. To be accurate, an increase of 1 point in BLEU score results in a decrease in Post-Editing time of around 0.16 seconds per word, indicating about a 3–4% speed increase for each BLEU point.

Tatsumi & Roturier (2010) researched the source text characteristics and the technical Post-Editing effort, and found a robust correlation between ambiguity and complexity and the technical Post-Editing effort. They measured source text characteristics using three different scoring systems: Systran, which measures the syntactic complexity and lexical ambiguity of the source sentences; After the Deadline, which checks spelling, style, and grammar; and acrolinx IQ, which facilitates controlled authoring by subjecting the document to a process of verification against a predetermined set of terms and rules to minimise ambiguity and promote consistency. It was noted that a robust relationship was identified between SYSTRAN's

complexity and ambiguity scores and technical Post-Editing effort in the interlingual translation discipline, specifically for the English-Japanese language pair.

Stasimioti & Sosoni (2020a) carried out a study on the impact of the MT outputs' quality on the technical effort of MTPE tasks. In their study, the outputs of SMT and NMT were evaluated using automatic evaluation systems such as BLEU and Word Error Rate (WER) metrics, as well as by two professional human translators in terms of adequacy and fluency, in order to obtain a more detailed and comprehensive depiction of the MT output. Regarding human evaluation, on the one hand, Escribe (2019: 37) claims that, as automatic evaluation systems, such as BLEU, don't take into account syntactic structures and grammar, human evaluation focusing on adequacy and fluency is necessary. On the other hand, such human evaluation is time-consuming and expensive. Furthermore, human judgment is highly subjective. The text translated by the NMT system was evaluated as higher on both BLEU and WER than the output of the SMT system. Furthermore, human translators evaluated the output of the NMT as superior to that of the SMT in terms of adequacy and fluency. After conducting Post-Editing tasks by professional translators and the analysis, the findings of the present study demonstrated that the BLEU and WER scores, as well as the human evaluation, exhibited higher scores in the Post-Editing of the NMT output, which led to a marginal decline in technical effort (Stasimioti & Sosoni 2020a: 170). These results suggest that high-quality MT output requires less technical effort in Post-Editing tasks.

In addition, Lacruz (2017) claims that the quality of MT output correlates with cognitive effort measured by the PWR. The study found a negative correlation between MT output quality and objective cognitive load, suggesting that low-quality MT outputs require higher pause ratios.

Regarding automatic readability evaluation of Japanese texts, "jReadability" is widely used, for example, for readability assessment of medical texts (Kohari et al. 2025). Furthermore, Lee & Hasebe (2025) used this system in their research on text simplification of Japanese news texts using ChatGPT to compare the ease of the original text and the translated texts by humans and LLMs. The tool was created for learners and teachers of the Japanese language (Lee et al. 2013; Okuhara et al. 2020). This readability evaluation system was developed based on the data from 100 Japanese language textbooks. Initially, the comprehensive dataset was segmented into 958 plain text files, with each file comprising 1,000 Japanese characters. The tool automatically calculates the readability of a text based on various linguistic factors, including the average sentence length, the word difficulty, the distribution of parts of speech, and the types of characters used per sentence. The formula of a readability score can be represented as follows:

$$
\begin{aligned}
y = {} & \{\text{average sentence length} \times -0.056\} \\
& + \{\text{kango}^{3} \text{ ratio} \times -0.126\} \\
& + \{\text{wago}^{4} \text{ ratio} \times -0.042\} \\
& + \{\text{verb ratio} \times -0.145\} \\
& + \{\text{particle}^{5} \text{ ratio} \times -0.044\} \\
& + 11.724
\end{aligned}
$$

2.5 Intralingual Translation

It should not be ignored that the most of those previous productivity studies were conducted in the scope of interlingual translation, but not intralingual translation; however, from the point of view of meaning transfer, from a source text into a target text, both types of translation activities share a certain similarity in terms of the translation effort, especially in cognitive activities (Whyatt et al. 2017: 143). Therefore, Krings' (2001) effort model and the same methodology for measuring three dimensions of the effort can be applied to a comparative process-oriented study of intralingual Machine and Human Translation, as Whyatt (2017) carried out one for paraphrasing research. Moreover, the effort model by Kring was originally suggested for a process-oriented study of MTPE; however, this model can also be used in comparative process-oriented research of Human Translation and MTPE, as Cui et al. (2023) conducted in their study. These research backgrounds should be a justifiable reason to apply this effort model and its investigative method to this comparative process-oriented study of MT Post-Editing and Human Translation into Easy Japanese for this master's thesis.

2.5.1 Easy Language and Pre-Editing

As mentioned in the introduction, text simplification, in the sense of translation into Easy Language, is located under the concept of intralingual translation (Jakobson 1959). Sakai (2021) suggests that through Easy Japanese, the accuracy of MT outputs into other languages can be improved. Similar to this approach, it has been researched that Pre-Editing can also improve the output quality of MT systems (Sánchez-Gijón & Kenny 2022). Sánchez-Gijón & Kenny define Pre-Editing as "[...] rewriting parts of source texts in a way that is supposed to

[3] A Japanese word originally derived from classical Chinese

[4] An original Japanese word

[5] A postpositional word functioning as an auxiliary to a main word

ensure better quality outputs when those texts are translated by machine" (Sánchez-Gijón & Kenny 2022: 81). It is also noticed that Pre-Editing is usually recommended in multilingual translation workflows because the advantages of Pre-Editing should be evident in each of the target language texts (Sánchez-Gijón & Kenny 2022: 82). Moreover, it can also be observed that there are some common points between Pre-Editing guidelines and Easy Language guidelines, such as avoiding unnecessarily complex sentences and using terminology consistently and coherently. From this perspective, Easy Language can also be seen as a pre-edited source text for MT.

With the Japanese-to-English language pair using Google MT, however, Sakai (2021) shows that Easy Japanese improved the MT output compared to using the original source text. It is mentioned that understanding the characteristics of the Japanese language is necessary when translating Japanese into other languages. It is also advised to replace compound and complex sentences with simple sentences, to state the subject explicitly, to supplement pronouns and so forth (Sakai 2021: 155). Furthermore, as Masden (2022: 208–209) found, there is a case in which Easy Japanese and foreign-language links appear on the same webpage. In this case, the Easy Japanese text can be regarded as the pre-edited source material. Should the English version correspond with the Easy Japanese content, it is to be considered a translation of the Easy Japanese text. This example indicates the importance of offering source information which can be understood by non-native speakers, besides translated information into other languages, namely information in Easy Language. As can be seen from this example, the pre-edited source text, in other words, text in Easy Language, is beneficial in instances where the source text is anticipated to be translated into multiple target languages. Moreover, the whole translation process from source text to target text can be accelerated when translation into Easy Language can be supported by MT tools, which is the main theme of this thesis.

2.5.2 Intralingual Translation with Artificial Intelligence

In this section, the basic functioning of language models will be described, as well as how Artificial Intelligence (AI) can help with translating into Easy/Plain Language. Language models are trained based on huge amounts of text from the internet, which are called corpora. Those trained language models, so-called LLMs, learn words and word combinations based on this big text data, and recognise the possibility of word combinations, not only between the words which are adjoining, but also over the entire sentence. While human translators write by viewing the whole text, LLMs generate texts word-for-word. Anschütz et al. (2025: 59) highlight that contexts are necessary for text generation, especially when a text handles

polysemy, which has different meanings depending on the context. Humans can distinguish polysemy in accordance with the global context and situation; however, language models need contexts to resolve the ambiguity correctly. Anschütz et al. (2025: 60) claim that when working with LLMs, it is important to consider that a word can have polysemy, and therefore, the meaning will be determined by the context.

The models are categorised into three types, namely encoder-based models, decoder-based models (Generative Pre-trained Transformer – GPT), and encoder-decoder models. The decoder-based models, such as ChatGPT, are specialised in generating texts (Yılmaz et al. 2025: 3). In the field of text simplification, an encoder-decoder model consists of an encoder which extracts the most important information and key statements from the input and a decoder which restates it with an easy structure. In the encoder, the text contents will be processed to create a numerical representation, the meaning of which will be comprehensible to the machine. In other words, the input text is initially mapped to an embedding. Anschütz et al. (Anschütz et al. 2025: 61) state that, on the one hand, encoder-decoder models are smaller and more efficient as they are trained for a specific purpose; on the other hand, they can generalise poorly and lose details and nuances easily, unlike GPT models.

Natural language processing (NLP) refers to forms of automated language processing, also including rule-based machine translation (RBMT). RBMT functions well in adding definitions and replacing difficult words with easier synonyms. This process is referred to as lexical simplification. A significant number of lexicographical databases contain word pairs that are synonymous with one another. Similarly, word definitions can be consulted in dictionaries and incorporated when required. In contrast, LLMs can deal with words having several meanings, as LLMs take the context into account. According to Anschütz et al. (2025: 63), RBMT models are less prone to hallucinations, and linguists can easily comprehend the strategy employed in text simplification. The researchers claim, however, that merely rewriting sentences is frequently inadequate; more extensive simplification may be required. In such cases, the limits of rule-based approaches become apparent. Consequently, the integration of multiple methods—for instance, the utilisation of rule-based models to segment extensive sentences before further processing—is frequently advocated for the purpose of achieving comprehensive text simplification. The quality of LLM outputs depends on the volume and quality of the training data, and they are mostly trained on standard languages. Therefore, for the purpose of text simplification tasks, LLMs need finetuning with domain-specific data and data in Easy/Plain Language as well. All in all, Anschütz et al. (2025: 63) recommend using the combination of different models, such as rule-based models and LLMs, and they also emphasise

the importance of human editors and translators performing critical reviews and taking responsibility for the publications (Anschütz et al. 2025: 70).

Kröger et al. (2025: 380) point out that, due to a lack of a scientific database of the performance of AI tools in the domain of intralingual translation, such tools for intralingual translation have not yet been established, while enough MT tools for interlingual translation are available on the market. For this reason, it is also not sufficiently empirically examined whether the intralingual translation process can be accelerated by MT systems. Kröger et al. (2025: 382) also highlight that the insufficiency of the high-quality parallel text data prevents the development of the intralingual MT systems.

Nevertheless, since LLMs have appeared on the market, several studies have been conducted in the field of automatic text simplification and intralingual translation by LLMs and MT systems. For instance, the readability, correctness and syntax complexity of simplified German texts, which were generated by ChatGPT, were measured in a study of Deilen et al. (2023). However, they concluded that while the texts became easier than the source texts, they did not satisfy the Easy Language standards. Moreover, the target texts contained some incorrectness and omitted crucial information. In their later work (Deilen, Garrido et al. 2024), an MT tool, which was developed by SUMM AI, specialised for intralingual translation (text simplification), was involved in a similar experiment as in the former study. It was also concluded that the machine-translated target texts contained some errors, and that the level of correctness and syntax complexity has not yet reached that of humans. Therefore, just like interlingual MT, intralingual MT also requires a Post-Editing process, especially for texts in safety-critical domains, such as medical, crisis, etc. Furthermore, in the case of using LLMs, prompting is needed to improve the comprehensibility of the machine-translated target texts (Deilen et al. 2023: 8).

Furthermore, besides the studies of automatic text simplification, the use of computer-assisted (-aided) translation (CAT) tools for intralingual translation has been researched for a decade. However, several studies (Deilen et al. 2023; Deilen, Garrido et al. 2024; Deilen, Lapshinova-Koltunski et al. 2024) claim that there are some issues with introducing CAT tools for the domain of intralingual translation in practice. First, in Easy Language translation, the description, explanation and definition must be clear in the text and cannot be hidden in the termbase. Second, in the context of intralingual sentence alignment, it is generally observed that there is typically no exact correspondence between the source text and the target text due to the splitting and merging of segments, the provision of additional explanations, and the

reordering of information within the source text. Therefore, alignment must be done manually by translators, thus expanding the workload.

2.5.3 Automatic Text Simplification in Japanese

Not only automatic text simplification of German context, but also that of Japanese context has been researched over the last decade. In a study (Iori et al. 2010), researchers built an automatic text simplification system with a statistical translation model which was trained by a corpus of 3,700 pairs of sentences. The results show that the accuracy was not high according to the BLEU score, a metric employed for the automatic evaluation of machine-translated text. They concluded that the low accuracy of the MT system was caused by the small volume and low quality of the corpus (Iori et al. 2010: 42). In a later study (Kumano et al. 2015), researchers built three statistical translation systems with distinct countermeasures applied. It was anticipated that these measures would result in enhancing the frequency of rewritable words and in a reduction of errors in word-simplifying, as they serve as a guide for managing unknown words and low-frequency words during the process of word alignment. As a result, each of them scored a higher BLEU score compared to the MT system without any countermeasures. As the outcome of these studies describes, it can be considered that some factors, such as the volume and quality of the corpus, architecture and countermeasures of MT systems, also affect the accuracy of the outputs in the discipline of automatic intralingual translation.

Several studies with LLMs have been conducted since they can handle the Japanese language. Similar to the other studies for German texts in the last section above, Hasebe & Lee (2025) conducted a comparative analysis of the readability between original texts and texts generated by ChatGPT (GPT-4). The study shows that the readability score of the AI-generated texts was higher than that of both the source and the human-simplified target texts. They find a potential that LLMs can collaborate with humans in text simplification tasks. In their supplementary research (Lee & Hasebe 2025), GPT-4o also joined the study, and it was found that GPT-4o could perform better than human translators as well as GPT-4. However, in terms of the readability score, GPT-4 exceeded GPT-4o. Nonetheless, both GPT-4o and GPT-4 perform as well as human translators or even better than they do. In contrast to the German context, in which not only LLMs, but also intralingual MT tools are getting focused on in the research of automatic text simplification, no study can be found where intralingual MT systems have been used.

In addition, as discussed above, public servants are confronted with limitations concerning translation into Easy Japanese due to a lack of manpower, knowledge, training and

education. Therefore, automatic text simplification tools also come into use in the field of practical applications in Japan, such as in local governments and public institutions. As a representative, “Tsutaeru-Web (伝えるウェブ)” (Morisaki 2018) has been widely introduced in more than 200 NPOs and municipal offices (Alfasado 2025b). The Tsutaeru-Web, an MT system for Easy Japanese, has been developed by a Japanese company, “Alfasado”, offering web accessibility. This MT tool is based on a program using NLP technology and is combined with its own dictionary database with around 35,000 words (Alfasado 2025a). Although the automatic text simplification tool has already been employed in the practical field, there is still no research on translation productivity and text readability using this tool.

As can be seen above, some studies on intralingual MT have been conducted in the last few years; however, almost all of the studies just focus on the output quality, more accurately, readability, correctness and syntax complexity. Although researchers are pointing out the necessity of the Post-Editing process, process-oriented studies of Post-Editing and Human Translation into Easy Language have hardly been done, in contrast to sufficient research on interlingual Post-Editing and Human Translation. For this reason, the study of this master’s thesis might be one of the earliest process-oriented studies of Post-Editing and Human Text-Simplification into Easy Japanese to date.

2.6 Crisis Translation

In crisis research, it is confirmed that communication and information play a significant role in reducing damage and saving lives. Additionally, the importance of Information and Communication Technology (ICT) is established in disaster settings. However, Cadwell and O’Brien (2016: 557) claim that the role of language, culture and translation has not been well researched yet.

Several terms, such as “crisis” and “disaster”, which generally indicate an unexpected event, are used synonymously in disaster studies. Therefore, both terms are used and handled in the same way in this thesis. According to Disaster Risk Management (DRM) or Disaster Risk Reduction, there are four stages of a disaster, namely Reduction, Readiness, Response and Recovery, which are called the “4R” derived from the initial letters of each word (O’Brien 2019). Each stage is defined as follows:

1 Reduction

> Identifying and analysing long-term risks to human life and property from hazards; taking steps to eliminate these risks if practicable, and, if not, reducing the magnitude of their impact and the likelihood of their occurring.
> **2 Readiness**
> Developing operational systems and capabilities before an emergency happens, including self-help and response programs, to ensure communities are trained and ready for when an emergency occurs.
> **3 Response**
> Actions taken immediately before, during, or directly after an emergency to save lives and protect property, and help communities recover.
> **4 Recovery**
> The coordinated efforts and processes to bring about the immediate, medium-term and long-term regeneration of a community following an emergency. (All About People 2017)

In each stage of a crisis, the accuracy of texts and information is very critical; moreover, especially in the readiness and response stages, speed will be one of the most critical factors to avoid unnecessary harm. According to the findings of the interviews, which were conducted by Cadwell (2016) as part of his qualitative study, lack of speed and consistency, especially in technical terminology, was the biggest problem in the 2011 disaster in Japan. He mentions that translation technologies, such as MT and Translation Memory, have the possibility to provide quick production and delivery of translation and boost the consistency of terminology. Moreover, developing such technologies further should also be beneficial for volunteer translators without any specific experience and training in translation, so that they can carry out tasks efficiently and collaboratively.

However, conducting disaster-related research on-site can present significant ethical challenges, given that a disaster area can be characterised by chaos and danger. There may be a lack of supplies and disorder in the infrastructure. Moreover, the individuals affected by these conditions can be traumatised, and the public servants need to work under the pressure of time and mental stress. To avoid such ethical concerns, hypothetical disaster settings can be used as an alternative method so that researchers do not need to enter a disaster zone and do not put the victims or helpers under unnecessary excessive stress. However, researchers must ensure that reliable and correct information and texts are provided to participants in a hypothetical experiment (O'Brien 2019). Otherwise, they could be unaware of the information they may need during a crisis in the future. For this reason, some authentic sample texts as well as guidelines published by local governments and public institutions are to be used in the experiment of this thesis under hypothetical crisis settings.

2.6.1 Translation Technologies and ICT in Crisis Settings

From this point of view, the use of MT in crisis settings is beneficial, but at the same time, the step of Post-Editing to check the accuracy, omission and other concerns in a target text cannot be skipped since an MT system might hallucinate (Deilen et al. 2023). Indeed, MT systems have been used before in some disaster events. One of the most well-known examples is a Haitian Creole-English MT engine, which was developed by Microsoft in four days during the 2010 Haiti earthquake, where translation of medical information was in high demand (Lewis 2010). The MT engine was available on multiple interfaces and could be integrated into a variety of apps and web pages. Another instance is the use of translation technologies in the aftermath of the 2011 Tōhoku earthquake and tsunami. While translation services were limited due to problems with power and connectivity, Yahoo! and Google Translate enabled access to crisis information in Japanese, English, Chinese and Korean languages on their websites (Cadwell 2016). Furthermore, Cadwell et al. (2019) carried out a study of the application of MT in crisis settings with citizen translators, including those who had been untrained in Post-Editing, in Kenya, and it was found that MTPE was not only faster than Human Translation but also reached the desired quality. From this result, they concluded that an MT system can be applied in disaster settings.

As can be seen above, some research on conventional MT tools was conducted, and those tools have been used in a real disaster situation for the last decade. However, cases cannot be found yet where LLMs have already been introduced in crisis settings to translate texts into Easy Language, although there is research on text simplification with an LLM in the medical domain (Makhmutova et al. 2024), which is a neighbouring subject to crisis settings.

In the context of Japan, Cadwell (2023: 26–27) reports, based on his research on Japan's annual White Paper on Disaster Management, that the Japanese government has invested enormously in ICT as well as language and translation technologies. For instance, a multilingual voice translation system and a "Safety Tips" application. ICT usually refers to the hardware, software, networks, and media employed for the collection, storage, processing, transmission, and presentation of information, in addition to associated services, which are collectively referred to as the information infrastructure, such as mobile phones, email, SMS, websites, television, social media, and the Internet. Cadwell & O'Brien (2016) also add that "In multilingual scenarios, ICT could also come to include machine translation, terminology management systems, devices for subtitling, and Translation Memory technology" (Cadwell & O'Brien 2016: 558).

Based on their research on translation technologies and ICT in the Great East Japan Earthquake 2011, they found that those technologies play a significant role in crisis settings

(O'Brien & Federici 2019: 132), and a model which captures translation as a part of the communication ecosystem was introduced (Cadwell & O'Brien 2016). The ICT ecosystem is characterised by four principles: diversity, flexibility, roundput and locality. From the point of diversity, technologies, which are used in disaster communication, differ depending on environmental conditions and limitations, such as individuals, their language and cultural backgrounds, time and zone. As ecosystems adapt to such changing environmental limitations, humans can also have an impact on the ecosystems they are a part of. They can adapt to their environment and even adjust it to suit themselves. Cadwell & O'Brien (2016: 567) introduced an example where people directly affected by the earthquake used signals received through their friend's car navigation system in order to watch a translated TV broadcast. From this perspective, humans, namely both information senders and recipients, are required to be flexible in disaster settings. In crisis communication, ICT plays a vital role in conveying information. As soon as information is introduced into the system, the information is refined, changed or distorted. In short, information circulates in the system. This information circulation is affected by environmental conditions such as power, infrastructure, location, income, language and culture. These conditions act as limiting factors within the information ecosystem. To summarise, the circulation of information through ICT during disasters depends not only on linguistic, cultural and technical experts but also on a range of technological, socio-cultural and natural environmental factors (Cadwell & O'Brien 2016: 570).

Even the enormous investment in translation technology and ICT by the Japanese government can be observed from corpus analysis of the White Paper by Cadwell (2023: 23), as translation technology-related words, such as multilingual voice translation technology and translation apps, appear frequently; however, it is also pointed out that human translators and interpreters who should be involved in such technologies and take part in the ecosystem are ignored in the White Paper. In this respect, citizen translators as well as a non-profit organisation such as Translators without Borders (TWB), which profits from translation technologies and ICT and where volunteer translators are participating actively, will be discussed in the following section.

2.6.2 Easy Language in Crisis Settings

Simplified text is also another effective way of communication in disaster settings, as people in these situations may not possess high levels of literacy, and interlingual translation may be unavailable due to a lack of linguists and technology. Therefore, intralingual translation is selected as one of the central themes in this study. However, translation into Easy Language

needs to follow some specific rules or guidelines, which are not that easy for volunteer translators without any specific experience or training to make full use of. Regarding this point, O'Brien (2019: 311) mentions that the use of technology, such as Acrolinx, which utilises AI to automatically assist with style, clarity, grammar, and more, should be advantageous.

As described in the section above, the history and background of Easy Japanese and Easy Language in other countries differ significantly. On the one hand, Easy Japanese was introduced in the context of disaster communication to reduce casualties, especially for foreign nationals in Japan; on the other hand, for instance, Plain/Easy German originates from the context of accessibility for people with learning disabilities. For this reason, the direction and perspective of the research on Easy language between Japan and other countries is slightly different. While studies in Japan focus on disaster communication in Easy Japanese, especially with emphasis on providing information for foreign citizens, research in other countries directs attention to accessibility for those who have difficulties comprehending information, which stems from differences in the level of expertise between laypeople and experts.

From this point of view, it can be considered that Japan leads research on Easy Language in the domain of crisis communication for foreigners. Sato (2024) reports that in the response after a disaster, it generally takes longer to provide multilingual support because volunteer translators are also affected by the disaster, and traffic access to the disaster-stricken area is blocked. Moreover, the translators and interpreters get overstressed as the number of language professionals available is limited, or even no replacement personnel are available for a specific language. As a result, each individual translator needs to engage in his or her tasks continuously. As the statistics in the former section show, the majority of the non-native Japanese foreign residents with various nationalities, who do not have English as their first language, prefer information in Easy Japanese to information in English. Sato (2024: 148) also mentions that most of them can manage enough Japanese in daily life, such as taking a bus or going shopping. Consequently, Easy Japanese is regarded as one of the most efficacious means of communication with foreign residents and a "lingua franca" in disaster communication.

Easy Japanese was originally so conceptualised that foreigners who have lived in Japan for about one year can understand it. For the creation of Easy Japanese, words which are related to disaster information, such as magnitude and location of the earthquake, outlook for aftershocks, prospects for restoring electricity, gas and water services, were categorised and organised along a timeline of a disaster. All those collected words were rephrased using approximately 2,000 easy words which can be understood by beginners in Japanese. The level of proficiency in these words is N4/N5 of the JLPT, which is equivalent to A2/A1 of the CEFR.

The phrases in Easy Japanese were developed through repeated verification experiments (Sato 2024). In 2004, Sato published the first guidelines for Easy Japanese, and the revised edition (Sato 2013) was published afterwards. The guidelines are intended for those who engage in supporting foreign citizens in crisis settings, such as local authorities, mass media and volunteer organisations. They include information which should be released within 72 hours after a disaster occurs. Matsuda et al. (2000: 146) justify based on the research on the Great Hanshin-Awaji Earthquake and the 1994 Northridge earthquake in Los Angeles that it takes at least 72 hours until multilingual support will be available. Sato (2013), however, highlights that various information in Easy Japanese is still required beyond the initial 72-hour period after the outbreak of a disaster.

2.6.3 Citizen Translators

Under the academic domain, a translator is defined as a person who has the necessary qualifications to act through adequate training and/or experience. According to the ISO 17100:2015 Translation services — Requirements for translation services, translator qualifications are defined as follows:

> a) a recognized graduate qualification in translation from an institution of higher education;
> b) a recognized graduate qualification in any other field from an institution of higher education plus two years of fulltime professional experience in translating;
> c) five years of full-time professional experience in translating. (ISO 2015: 6)

Additionally, the ISO 18587:2017 Translation services — Post-Editing of MT output — Requirements defines the criteria of the qualifications for the post-editor as follows:

> a) has obtained a degree in translation, linguistics or language studies or an equivalent degree that includes significant translation training, from a recognized institution of higher education;
> b) has obtained a degree in any other field from a recognized institution of higher education and has the equivalent of two years of full-time professional experience in translating or post-editing;
> c) has the equivalent of five years of full-time professional experience in translating or post-editing. (ISO 2017: 8)

However, in disaster settings, anyone who is able to serve as an intermediary between two or more linguistic and cultural systems, with or without specific training or qualifications, can be a translator (O'Brien & Federici 2019: 132). As mentioned shortly in the citation in the introduction, citizen translators who do not have experience and training as professionals contribute voluntarily to crisis settings. The translation service can sometimes be paid and

sometimes unpaid. O'Brien describes that "when people are faced with a crisis, the luxury of a trained professional is often just that – an unattainable luxury" (O'Brien & Federici 2019: 132). Moreover, in the domain of translation studies and professional practice, the term "translation" is employed to denote the process of written language transfer, while "interpreting" is used to denote spoken transfer. In crisis communication, these roles frequently overlap, with an individual often assuming responsibility for both (O'Brien 2016: 86).

As described above, citizen translators are involved and play a crucial role in crisis settings when professional translators are unavailable. In practice, Sendai City (Sendai City 2021) reports that 184 translators and interpreters engaged in translation services as a part of Sendai City Disaster Language Volunteers. This is no exception in the case of Easy Language translation. Indeed, a study of the Bureau of Citizens and Cultural Affairs (Bureau of Citizens and Cultural Affairs 2020: 9) shows that 22 out of 71 public organisations in Tokyo outsource text simplification tasks to volunteers. However, since the crisis information needs high-quality translation, and often concerns human and security contents, O'Mathúna et al. (2020) discuss an ethical justification and the importance of non-professional translators being involved in crisis communication. They argue that even though the quality of the translation of citizen translators may be inferior to that of professionals, the risk of harm that arises from a lack of vital information that is not available in languages they comprehend must be averted as a top priority.

Furthermore, O'Mathúna et al. (2020: 10–11) highlight that regarding the "trust", citizen translators can stand at an advantage over professional translators. Although the possession of specialised knowledge and training is indicative of their expertise, trust between volunteer translators and citizens from a disaster-stricken area can be more easily established, or sometimes it has already been built when the volunteers are coming from that region. Individuals from diverse backgrounds, including not only language educators from local educational institutions, university students, but also hospitality industry professionals with experience in communication, can contribute to this initiative. Citizen translators often belong to the communities they serve, giving them an advantage in building trust. Their existing social ties, networks, and understanding of community dynamics allow them to connect more effectively than professional translators. Moreover, volunteer citizen translators tend to engage not only in the response stage, but also in the reduction stage in that community.

As mentioned above, the quality of translation service by citizen translators may be inferior to that of professionals, and errors lead to fatal consequences in crisis scenarios. Therefore, O'Mathúna et al. (2020: 16) recommend the use of citizen translators as peer

reviewers, and professional translators could engage as supervisors. Moreover, when selecting citizen translators, their age, minimum language proficiency, and psychological well-being should be considered. Ideally, linguistic as well as ethical training should be held. In Japan, such a workshop is also offered in the scope of intralingual translation, namely Easy Japanese (Tokyo Metropolitan Foundation 'TSUNAGARI' 2025). O'Mathúna et al. (2020: 20) also state that providing guidelines can avoid such quality and ethical issues. Indeed, the guidelines of Easy Japanese will be provided to the participants of the experiment beforehand.

As mentioned in the previous section, citizen translators and NGOs such as TWB, in which volunteer translators are actively involved, are one of the biggest bodies that benefit from translation technology and ICT. MT systems, the internet and crowdsourcing solutions are the vital technologies which enable those NGOs and their associated volunteer translators to engage in translation services in a crisis situation. Rogl (2017: 237) mentions that ICTs and the Internet are very important for translators, as they facilitate access to online lexicons and glossaries, parallel texts, and web-based platforms for collaborative translation projects. These mechanisms enable the global sharing of Translation Memories and enhance the capabilities of MT engines. In addition, O'Brien (2016: 93) highlights that crowdsourced volunteer translation can therefore be an effective method of communication during crises; however, effective management is required to maintain engagement and ensure accuracy, as volunteer translators may lack the necessary translation and linguistic skills for reliable crisis communication.

Based on the literature review, the following hypothesis might be proposed for this thesis: MT and LLMs also make the process of intralingual translation into Easy Japanese more productive and less effortful than Human Text-Simplification in crisis settings, and improve the readability of the target text. The following chapter outlines the methodology for the empirical experiment, which aims to verify the hypothesis.

3 Research Questions and Methodology

Based on these statistics, backgrounds and previous research, the following research questions arise.

RQ1:

What is the impact of working with MT and LLM output in a Japanese intralingual text simplification task on the **temporal, technical, and cognitive effort** of citizen translators with Japanese as their native language?

To obtain the data on temporal effort, the total time spent on the MT/LLM-assisted Text-Simplification and Human Text-Simplification was measured. The total time was normalised by the number of Japanese characters in the source texts. For the measurement of the technical effort, the Human-targeted Translation Edit Rate (HTER) was adopted to mathematically calculate the edit distance between the outputs of an MT and the post-edited target texts. Additionally, keystroke counts, including data on insertion, deletion, navigation, copy, cut-and-paste and mouse operations, were systematically collected. As a measurement of objective cognitive effort, the Pause-to-Character Ratio (PCR) with a 300-ms pause threshold was used. Finally, the participants were asked to rate their workload level to measure subjective cognitive effort.

RQ2:

What is the impact of working with MT and LLM output in a Japanese intralingual text simplification task on the **readability of the target text**?

To investigate RQ2, the quality of the target texts was quantitatively measured. To obtain quality assessment, objective readability will be calculated using the Japanese text readability measurement system called “jReadability”.

Following, the condition of participants who joined the empirical study and performed Human Text-Simplification and MT/LLM-assisted Text-Simplification tasks is mentioned. Also, the way to select the source texts and experiment procedures is described.

3.1 Participants

Six participants with Japanese as their first language were recruited. Each of them has a different age and occupation. They have neither any particular experience and training in translation and Post-Editing, nor any knowledge of Easy Japanese, because it is supposed that untrained volunteer translators should often be involved in crisis settings. Furthermore, it is not realistic to distinguish participants with and without experience and recruit them only for this purpose. The present experiment focuses on comparing the effort and productivity of MT/LLM-assisted Text-Simplification and Human Text-Simplification. However, the study does not address variations related to translators' experience or training. Ahead of the experiment, each participant agreed to the declaration of consent and signed the document (see appendix). The list of participants is presented in Table 2 below.

Table 2: List of participants

Participants' No.	Age	Gender	Occupation	Text order of tasks
P_1	Early 30s	Male	Master's student in Translation Studies	T_1, T_2, T_3
P_2	Late 40s	Female	Pastry chef	T_1, T_3, T_2
P_3	Late 20s	Male	Master's student in East Asian Economy and Society	T_2, T_1, T_3
P_4	Early 20s	Male	Bachelor's student in Law	T_2, T_3, T_1
P_5	Early 30s	Female	Japanese teacher	T_3, T_1, T_2
P_6	Early 50s	Female	Singer / Japanese teacher	T_3, T_2, T_1

3.2 Source Text Selection

For each Human Text-Simplification and MT/LLM-assisted Text-Simplification task, informational texts with almost the same amount and volume of words, which are published by public administrations and released online as a template, or archived in a webpage of a municipal office, were selected. These texts include emergency information about evacuation indications and nearby shelters, which will be sent via Email, SMS, or social network by local governments, or published on the webpages of public companies when a crisis hits.

For the main experiment, three different texts about disaster information were selected. Text_1 was collected from the disaster information archive of Oiso Town (2025b), which was published on the 11th of September 2025 and informed of tornadoes in that area. Text_2, published on the 30th of July 2025, was collected from the homepage of Shizuoka Prefecture (2025). The text handled the damage situation due to the Kamchatka earthquake on the same day. Text_3 was derived from Weathernews Inc. (2023), a public weather forecasting company, and it reported the earthquake in the Noto Peninsula on the 5th of May 2023 and its expected damage due to heavy rain. The source texts and their English translation (as a reference) are enclosed in the appendix.

According to the readability score by jReadability (see section 4.4.4), Text_1 with 128 Japanese letters scored 3.33, categorised as the "Upper intermediate (somewhat difficult)", Text_2 with also 128 letters was rated with 2.20, falling within the "Lower advanced (difficult)", and Text_3 with 138 letters scored 2.84, corresponding to the "Upper intermediate (somewhat difficult)". As can be seen here, the collected source texts have almost the same volume of

words, and the difficulty between the texts is not very different to each other. From this respect, the comparative study can be carried out fairly.

3.3 Experimental Procedures

Before the experiment is carried out, a briefing, declaration of consent and guidelines for Easy Japanese have been sent to each participant, and they have been asked to take a look at the provided materials. The briefing describes the details of their role, tasks, situation and notification. The participants were asked to complete the task according to a set of guidelines inspired by the ones published by Mie Prefecture (Mie Prefecture 2025). These guidelines include some key points to follow, such as using less kanji and avoiding double negation, when translating into Easy Japanese. The whole guidelines are enclosed in the appendix.

The MT/LLM-assisted Text-Simplification and Human Text-Simplification tasks were conducted within the Translog-II interface (Carl 2012). The source texts have been copied and pasted into the source text area of the Translog-II.

Figure 3: Screenshot Translog-II interface

For the MT/LLM-assisted Text-Simplification tasks, the two source texts (Text_2 and Text_3) were automatically translated into Easy Japanese by an MT system called Tsutaeru-Web and a custom Gem of Google's Gemini (2.5 Flash) (Google 2025) in advance, and then the outputs of that have also been copied and pasted into the target text area of the Translog-II. The custom

Gem of Google's Gemini, which was used for machine-translating another source text, has already been trained by the experiment conductor with the guidelines of Mie Prefecture (Mie Prefecture 2025) and instructed to translate texts into Easy Japanese automatically.

Each translator took on three tasks: Human Text-Simplification and Post-Editing of the MT tool, Tsutaeru-Web and of the LLM, the custom Gem of Google's Gemini. To avoid confusion, the following point should be made clear: for the MT/LLM-assisted Text-Simplification tasks, the source texts have been machine-translated in advance; therefore, translators do not need to operate the MT tool and LLM by themselves during the tasks. Furthermore, extra prompts are not needed to get a translated text in Easy Japanese because it has been commanded to translate texts automatically, as mentioned above. In short, the time spent using the MT tool and the LLM is excluded from this study. In the experiment, six translators translated three different source texts, each in a unique order, because there are three source texts, and therefore six different permutations are possible. According to this method, it can be shown that the order of the tasks should not affect the grade of effort and productivity for each task. To standardise the control variable, the same computer was used, and the same interface of the screen was set up for each participant. The experiment conductor was available in case of questions from the participants and technical problems.

3.4 Measurements

In this subchapter, the detailed measurement systems for each effort (Temporal, Technical and Cognitive) during the Human Text-Simplification and Post-Editing are explained. Finally, the quality measurement of the target texts is also described.

3.4.1 Temporal Effort (time spent)

As mentioned in the chapter on the theoretical background, the effort of Post-Editing and Human Translation is separated into three dimensions: temporal (time spent), cognitive (mental processing), and technical (physical action) (Krings 2001), and the temporal and technical efforts can be measured directly, while the cognitive effort cannot be measured directly.

In this paper, the time spent will also be measured. To be precise, the total time spent on the MT/LLM-assisted Text-Simplification and Human Text-Simplification will be measured for the temporal effort. Strictly speaking, the total time will be normalised by the number of Japanese characters in the source texts so that it can fairly compare the spent time per character,

because the word counts of the texts can slightly differ, even though almost the same volume of texts should be used in the research (Alvarez et al. 2020; Alvarez-Vidal & Oliver 2023).

3.4.2 Technical Effort (physical action)

For the measurement of the technical effort, the Human-targeted Translation Edit Rate (HTER) (Snover et al. 2006) was adopted to mathematically calculate the edit distance between the outputs of an MT and the post-edited target texts. This HTER was modified based on the TER (Snover et al. 2006), which originally quantifies the edit operations needed toward a reference translation, and the HTER was additionally amended to be used for Japanese texts (see Appendix). However, the metrics can only measure the shortest way to the end product, and therefore, the full physical action during the MT/LLM-assisted Text-Simplification process will be ignored. Moreover, the metrics are only usable for the measurement of the technical effort of the MT/LLM-assisted Text-Simplification, but not for the Human Text-Simplification.

For this reason, keystrokes using a software-based keylogging tool, which systematically collects data on such user activity data as insertion, deletion, navigation, copy, cut-and-paste and mouse operations, can additionally be used for the measurement of the technical effort for MT/LLM-assisted Text-Simplification as well as for Human Text-Simplification (Cui et al. 2023; Martín et al. 2025). In this research, the Translog-II was used as a keystroke logger to collect the data of keyboard and mouse actions (Alvarez-Vidal & Oliver 2023). The number of total keystrokes and mouse behaviour was normalised by the number of Japanese characters in the source texts as well, for the same reason as clarified above. In short, to measure the technical effort of the MT/LLM-assisted Text-Simplification, both keystroke and HTER were used, while only the keylogging data was collected and analysed for the Human Text-Simplification.

3.4.3 Cognitive Effort (mental processing)

The cognitive effort is the most complicated element because it cannot be measured directly, as mentioned above. Therefore, the Pause-to-Character ratio (PCR) with the most commonly used 300-ms pause threshold was employed in the experiment of this thesis, as previous studies have done (Lacruz et al. 2016; Liu & Cui 2024). After each experiment, an RTF file was exported from the Translog-II, and the file was analysed using the tool (Tobisch 2025), which can automatically calculate the pauses.

[Start][•01.953][•00.390][•01.234][▼][▲][•15.360][▼][▲][•03.875][▼][▲][•01.375][A][•00.547][B][U][N][A][I][•01.000]あぶない[•03.375]◄◄◄◄[•00.750][A][•00.375][B][U][N][A][I][•00.656][Space][•00.484]危ない[•04.235]←←[•01.235][K][A][K][K][O][Space][•00.500][Space][•00.719][Space][Space][Space][Space][Space][Space][Space][Space][•07.812] （） [▼][▲][•02.875]←←←←←[•01.015][A][B][U][•00.344][N][•00.781]あぶn [•00.969]◄[•10.531][▼][▲][•32.906][▼][▲][•03.625][K][A][K][K][O][•00.672][Space][•04.641] （）

Figure 4: Sample of linear view

Until now, objective methods have been introduced for the collection of data for each effort. To collect the data on the subjective cognitive effort, subjective ratings were additionally carried out after participants had completed the Human Text-Simplification and MT/LLM-assisted Text-Simplification tasks. In this study, a retrospective survey was carried out in the form of the adapted NASA Task Load Index (NASA-TLX) (Dai & Liu 2024; Sun 2012), a scale for measuring subjective cognitive effort of translation and MT/LLM-assisted Text-Simplification. After completing a Human Text-Simplification and MT/LLM-assisted Text-Simplification task, the participants rated their cognitive load (see Figure 5).

Cognitive Effort: How hard did you have to work to accomplish your level of performance?

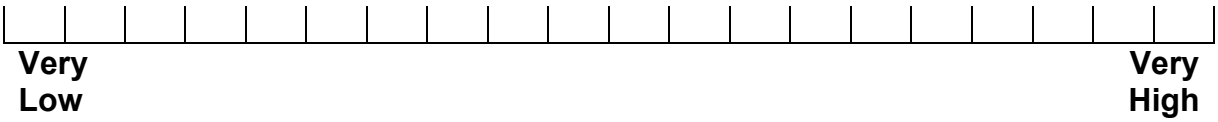


Figure 5: NASA-TLX subscale (adapted from NASA's Ames Research Center 2022)

3.4.4 Readability Score (Quality)

In this study, not only the translation efforts but also the quality of the target texts were quantitatively measured. To obtain quality assessment, objective readability was calculated using the Japanese text readability measurement system called "jReadability", which had been introduced in the literature review. The system returns a score from 0.5 to 6.4 in six-level categories (upper advanced, lower advanced, upper intermediate, lower intermediate, upper elementary, and lower elementary) (Hasebe & Lee 2015). For detailed score ranges, please refer to Table 3 below.

Table 3: jReadability Score Guidelines (adapted from Hasebe & Lee 2025)

Score Range	Level	Description
0.5–1.4	Upper advanced (very difficult)	Can understand highly specialized texts without difficulty. Does not experience challenges with any Japanese texts.

1.5–2.4	Lower advanced (difficult)	Can understand most specialized texts. Can comprehend complex structures found in literary works.
2.5–3.4	Upper intermediate (somewhat difficult)	Can grasp the general content of somewhat specialized texts and can handle most texts encountered in daily life without difficulty.
3.5–4.4	Lower intermediate (average)	Can understand relatively simple texts and grasp the content of moderately long passages.
4.5–5.4	Upper elementary (easy)	Can understand basic vocabulary and grammatical items. Can comprehend basic complex sentences using *te*-form.
5.5–6.4	Lower elementary (very easy)	Can understand basic Japanese expressions centered around simple sentences. Cannot understand complex sentence structures such as relative clauses.

As the evaluation tool, jReadability, is widely used for readability assessment and in research on text simplification of Japanese news texts using LLMs to compare the ease of the original text and the translated texts by humans and LLMs, the use of this readability evaluation system was used for this thesis's study.

3.5 Pilot Study

Before doing the main experiments, a small-scale pilot study was conducted to test the functionalities of the Translog-II and the measurement system of each effort. A woman in her early thirties participated as a volunteer translator in this small-scale pilot study, which was conducted on the 11th of August 2025.

In contrast to using three different texts in the main experiment, only two warning texts about tsunamis, both of which appeared after an outbreak of the Kamchatka earthquake on the 30th of July 2025, were used, as this small pilot study mainly focused on testing the tool and system. As materials, a tweet text from the Crisis Management Division, General Affairs Bureau, Crisis Management Department (2025) was used for the Human Text-Simplification, and the other text, which appeared on the homepage of Oiso Town (2025a), was used for the MT/LLM-assisted Text-Simplification task. The second text was machine-translated into Easy Japanese by Gemini. The first source text contains 126 Japanese letters with a readability score of 3.03, categorised as the "Upper intermediate (somewhat difficult)", while the second one has

144 letters with a readability score of 3.34, also falling within the “Upper intermediate (somewhat difficult)” according to the jReadability.

A few days before the pilot study, the participant was informed about the tasks and asked to skim the provided guidelines of Easy Japanese in advance, which could be referred to during the experiment. In this pilot study, the Human Text-Simplification was performed first, and the MT/LLM-assisted Text-Simplification followed after a few minutes’ break.

For the hardware setup, a laptop with Windows OS was connected to a large monitor so that the translator could refer to the guidelines on the laptop screen while she was translating and Post-Editing on the big screen. For the setup of the Translog-II, a source text on the left side and a target text on the right side were used. Nevertheless, difficulties were encountered in the act of moving the cursor between the two screens using a mouse. For this reason, the laptop screen will be mirrored on the big screen, and translation and referring to the guidelines will be operated exclusively on the large screen. However, participants can open a browser to see the guidelines.

As for the temporal effort of the tasks, the following data on the duration was collected. The Human Text-Simplification took 10.2 seconds per character, while Post-Editing took 5.5 seconds per character. As for the technical effort, Human Text-Simplification required 12.2 keystrokes per character, while Post-Editing required 3.6 keystrokes per character. The HTER score for the Post-Editing was 22.8. As for the cognitive effort, the value of the PWR for Human Text-Simplification was 28.0 at a 300-ms threshold, while that of Post-Editing was 17.0. Additionally, the participant scored the subjective cognitive load for Human Text-Simplification with 6 points out of 20, while that for Post-Editing was 8. Finally, as for the readability scores as a quality of the target texts, according to the jReadability, the human-translated text marked 4.66, while the post-edited text scored 5.43, corresponding to the “Upper elementary (easy)”.

4 Results

This chapter presents the results of the empirical experiment conducted in accordance with the methodology outlined above. The experiment was carried out in the period from the 20th of October to the 6th of November 2025 in a room of a Gemeindewohnung in the 4th district of Vienna under labour conditions. All translated and post-edited outputs by the participants are attached as an appendix.

Due to the use of several measurement units, abbreviations are employed. The abbreviations and their meaning are described as follows:

- Total time normalised by the number of source text Japanese characters (TSC)
- Keystrokes normalised by the number of source text Japanese characters (KSC)
- Ratio of the number of pauses to the number of Japanese characters with a 300-ms threshold (300-PCR)
- NASA Task Load Index (NASA-TLX)
- jReadability Score (jRS)

4.1 Median and Mean Scores of Each Effort

Table 4 below presents the overview of the median and mean scores of the temporal, technical and cognitive efforts and also the readability score as a quality indicator for the Human Text-Simplification (T_1) and two MT/LLM-assisted Text-Simplification tasks with Tsutaeru-Web (MT) (T_2) and Google Gemini (LLM) (T_3).

Table 4: Median and mean scores of each effort and quality

	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
Temporal Effort			
Duration (TSC)			
Median	6.1	3.9	1.6
Mean	6.3	4.4	1.7
Technical Effort			
Keystrokes per character (KSC)			
Median	8.4	3.2	0.3
Mean	8.2	3.4	0.6
HTER			
Median	N/A	34.4	4.1
Mean	N/A	34.6	5.1
Cognitive Effort			
300-PCR			
Median	15.2	11.5	4.8
Mean	14.5	13.0	5.3
NASA-TLX			

	Median	10.5	8.5	4.0
	Mean	10.8	7.8	4.5
Quality				
	jReadability Score (jRS)			
	Median	5.13	3.51	5.36
	Mean	5.06	3.61	5.41

As can be seen from the table, the result shows that the MT/LLM-assisted Text-Simplification tasks required less effort from each temporal, technical and cognitive dimensions than the Human Text-Simplification. Figure 6 below supplements that Post-Editing of LLM outputs consistently achieves the largest improvements in temporal, technical, and cognitive effort, while Post-Editing of MT outputs provides moderate efficiency gains but decreases readability compared to Human Text-Simplification.

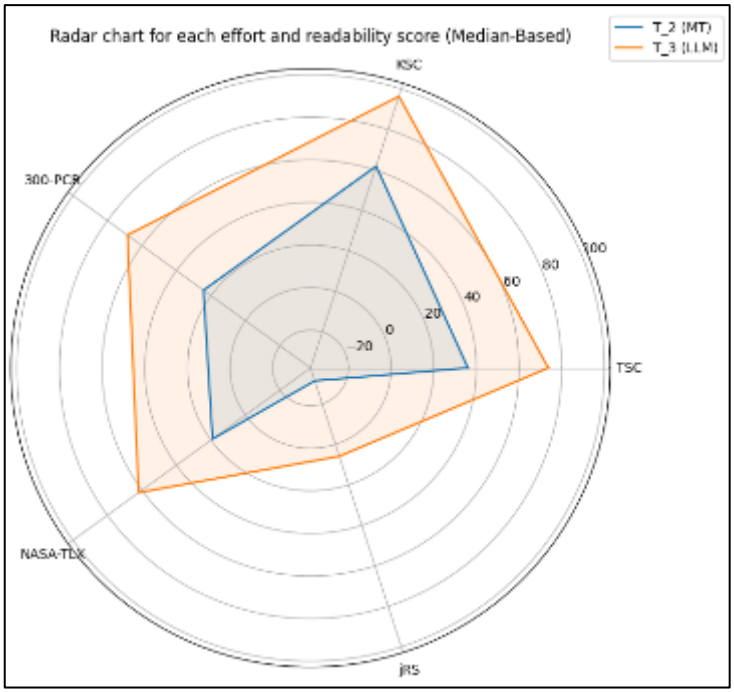


Figure 6: Reder chart for each effort and readability score[6]

First, the TSC value of both T_2 (MT) and T_3 (LLM) was smaller than that of T_1 (HTS). Compared to Human Text-Simplification (T_1), Post-Editing of T_2 (MT) was 36.1% faster, and Post-Editing of T_3 (LLM) was 73.8% faster based on the median values. That

[6] The radar chart visualises the percentage improvement of MT and LLM over HS, based on median values across five metrics: Duration, keystrokes per character, 300-PCR, NASA-TLX and readability.

means the translators require less temporal effort for Post-Editing than translation from scratch in this research.

Second, the KSC score of both T_2 (MT) and T_3 (LLM) was also smaller than that of T_1 (HTS). For Post-Editing of T_2 (MT), 61.9% fewer keystrokes per character and for Post-Editing of T_3 (LLM), 96.4% fewer keystrokes per character were required. Additionally, the HTER score shows that significantly less Post-Editing was needed for T_3 (LLM) than for T_2 (MT). The results show that Post-Editing is less technically demanding than Human Text-Simplification for translators, especially with the support of the LLM rather than the MT tool in this case.

Third, as an objective indicator for cognitive effort, the PCR value with a 300-ms threshold for both T_2 (MT) and T_3 (LLM) was smaller than that of T_1 (HTS). The Post-Editing task of T_2 (MT) was 24.3%, and the Post-Editing task (LLM) was 68.4% less cognitively demanding compared to T_1 (HTS). In addition, the subjective indicator, NASA-TLX, shows that Post-Editing requires less cognitive effort than Human Text-Simplification. Based on the retrospective feedback, the participants felt 19.0% less for the Post-Editing of T_2 (MT) and 61.9% less cognitive load for the Post-Editing of T_3 (LLM) than that for T_1 (HTS). The results of the subjective as well as objective indicators show that using the MT and LLM makes intralingual translation tasks less cognitively demanding compared to translation from scratch.

Finally, the jRS was 31.4% lower with the MT system than without AI assistance, while the jRS was 5.9% higher with the LLM. In short, the MT tool has a negative impact on the readability of the text, and on the other hand, the LLM has a positive effect on readability in this study. The reason for the negative impact of the MT system will be discussed in a later chapter.

4.2 Temporal Effort (time spent)

In this section, the results of the temporal effort will be described. Figure 7 below shows the maximum and minimum values, median, and mean for each text. As can be seen from the boxplot, it is indicated that T_1 (HTS) requires the longest translating time per character (TSC), followed by Post-Editing of T_2 (MT) and then T_3 (LLM).

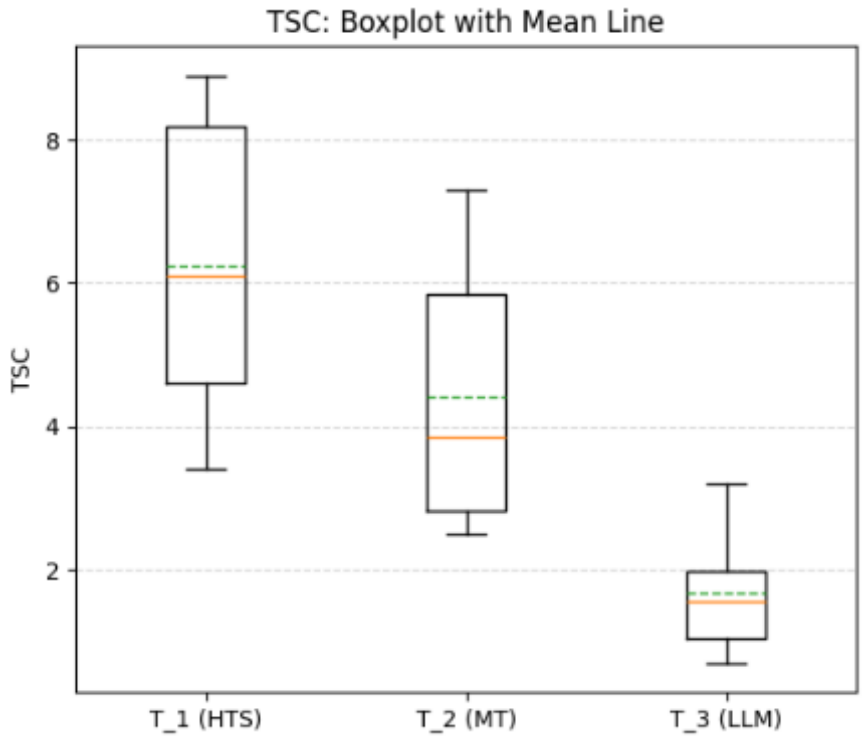


Figure 7: Boxplot of TSC[7]

The summary statistics of the boxplot of TSC are as follows. (sample size n, mean $\bar{x}$, median $\tilde{x}$[8], quartiles Q_1, Q_3, interquartile range IQR, and observed range). T_1 (HTS): n = 6, x̄ = 6.3, x̃ = 6.1, Q1 = 4.5, Q3 = 8.5, IQR = Q3 − Q1 = 4.0, range = [min, max] = [3.4, 8.9], no outliers (1.5 × IQR rule). T_2 (MT): n = 6, x̄ ≈ 4.4, x̃ = 3.9, Q1 = 2.6, Q3 = 6.4, IQR = 3.8, range = [2.5, 7.3], no outliers (1.5 × IQR rule). T_3 (LLM): n = 6, x̄ ≈ 1.7, x̃ = 1.6, Q1 = 1.0, Q3 = 2.0, IQR = 1.0, range = [0.7, 3.2], no outliers (1.5 × IQR rule).

Table 5: TSC for each participant on each task

TSC	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
P_1	4.5	2.6	2.0
P_2	4.9	3.5	1.0
P_3	7.3	7.3	1.9
P_4	8.5	6.4	3.2
P_5	8.9	4.2	1.2
P_6	3.4	2.5	0.7

[7] The Solid line indicates the median, and the dotted line indicates the mean value.

[8] All medians $\tilde{x}$are the average of the 3rd and 4th order statistics for $n = 6$. IQR was used to determine whiskers and outliers per the conventional $Q_1 - 1.5 \cdot \text{IQR}$, $Q_3 + 1.5 \cdot \text{IQR}$ rule.)

Median ($\tilde{x}$)	6.1	3.9	1.6
Mean ($\bar{x}$)	6.3	4.4	1.7

Table 5 above displays the TSC scores for each participant on each task. Overall, T_3 (LLM) demonstrates the least dispersion and the lowest central tendency, suggesting that it results in the fastest completion times of the three translation and Post-Editing conditions.

4.3 Technical Effort (physical action)

This section presents the results of the technical effort (physical action). As introduced in the methodology chapter, the technical effort was measured by using two methods: keystrokes per character and Human-targeted Translation Edit Rate (HTER).

4.3.1 Keystrokes per character

First, the results of the keystrokes per character will be described. Figure 8 below shows the maximum and minimum values, median, and mean for each text. As can be seen from the boxplot, evidence suggests that the most keystrokes per character (KSC) are required for T_1 (HTS), followed by Post-Editing of T_2 (MT) and then T_3 (LLM).

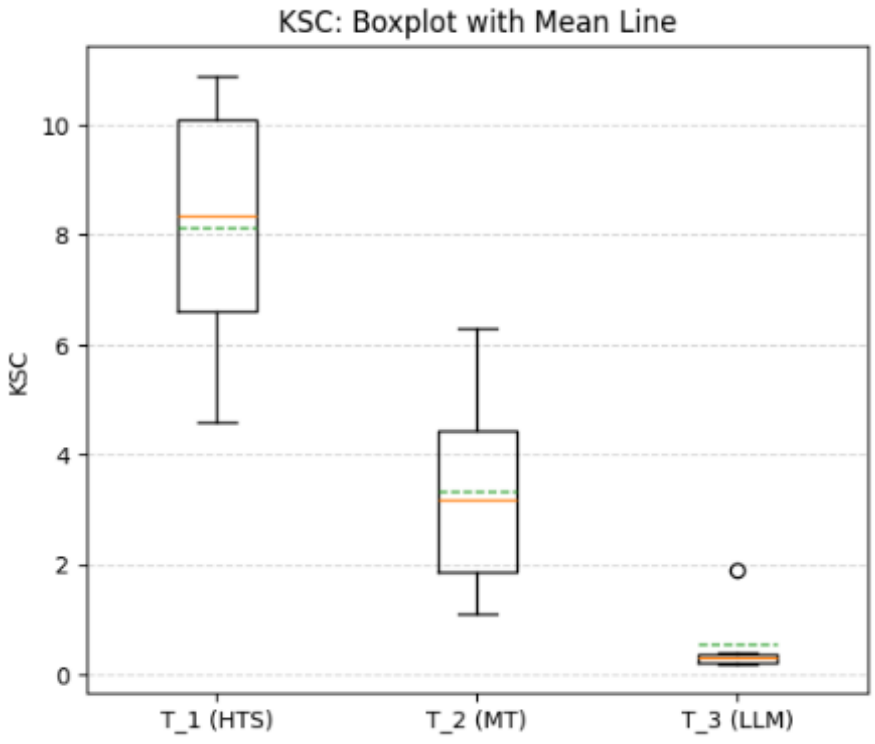


Figure 8: Boxplot of KSC

The summary statistics of the boxplot of TSC are as follows. T_1 (HTS): n = 6, $\bar{x}$ = 8.2, $\tilde{x}$ = 8.4, Q1 = 6.5, Q3 = 10.2, IQR = Q3 − Q1 = 3.7, range = [4.6, 10.9], no outliers (1.5 × IQR rule). T_2 (MT): n = 6, $\bar{x}$ = 3.4, $\tilde{x}$ = 3.2, Q1 = 1.8, Q3 = 4.5, IQR = 2.7, range = [1.1, 6.3], no outliers (1.5 × IQR rule). T_3 (LLM): n = 6, $\bar{x}$ = 0.6, $\tilde{x}$ = 0.3, Q1 = 0.2, Q3 = 0.4, IQR = 0.2, range = [0.2, 1.9], one outlier detected (1.9), based on the 1.5 × IQR rule.

Table 6: KSC for each participant on each task

KSC	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
P_1	6.9	1.1	0.2
P_2	4.6	2.1	0.2
P_3	10.2	6.3	1.9
P_4	9.8	4.3	0.3
P_5	10.9	4.5	0.4
P_6	6.5	1.8	0.3
Median ($\tilde{x}$)	8.4	3.2	0.3
Mean ($\bar{x}$)	8.2	3.4	0.6

Table 6 above displays the KSC scores for each participant on each task. To sum up, the results of the KSC analysis demonstrate a marked decline from T_1 (HTS) to T_2 (MT) and subsequently to T_3 (LLM), suggesting a reduction in the technical effort required. The highest number of keystrokes per character was required by Human Text-Simplification, whereas MT reduced the effort by more than half, and LLM output required almost no manual editing. The distribution for T_3 (LLM) exhibited a single high-value outlier; however, this was not reflected in the overall variability, which remained minimal in comparison to the other conditions.

4.3.2 Human-targeted Translation Edit Rate (HTER)

Secondly, the results of the HTER will be described. Figure 9 below shows the maximum and minimum values, median, and mean for T_2 (MT) and T_3 (LLM). As demonstrated by the boxplot, the data indicate that the output of T_2 (MT) required more Post-Editing, such as word addition, deletion and replacement, than that of T_3 (LLM).

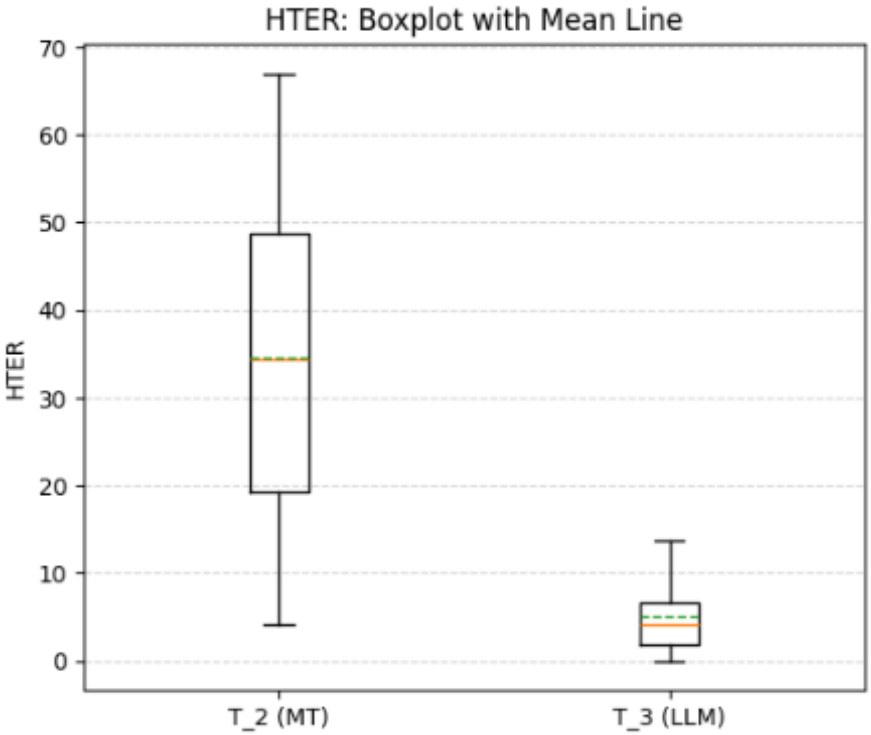


Figure 9: Boxplot of HTER

The summary statistics of the boxplot of HTER are as follows. T2 (MT): n = 6, $\bar{x} \approx 34.6$, $\tilde{x} = 34.4$, Q1 = 15.9, Q3 = 51.9, IQR = Q3 − Q1 = 36.0, range = [4.2, 67.0], no outliers (1.5 × IQR rule). T3 (LLM): n = 6, $\bar{x} \approx 5.1$, $\tilde{x} = 4.1$, Q1 = 1.3, Q3 = 7.5, IQR = 6.2, range = [0.0, 13.8], no outliers (1.5 × IQR rule).

Table 7: HTER for each participant on each task

HTER	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
P_1		4.2	1.3
P_2		29.4	7.5
P_3		67.0	13.8
P_4		15.9	4.4
P_5		51.9	3.8
P_6		39.4	0
Median ($\tilde{x}$)		34.4	4.1
Mean ($\bar{x}$)		34.6	5.1

Table 7 above displays the HTER scores for each participant on each task. Statistical analysis revealed that HTER scores were substantially lower in T_3 (LLM) than in T_2 (MT). Moreover, T_2 (MT) exhibited considerable variability. In contrast, T_3 (LLM) remained

consistently low with minimal dispersion. Additionally, it can be observed that P_6 did not undertake any editing whatsoever and delivered the LLM output as the final product. Overall, the findings suggest that the LLM-generated output necessitates considerably fewer manual corrections compared to the output of the conventional MT.

4.4 Cognitive Effort (mental processing)

This section presents the results of the cognitive effort (physical action). As introduced in the methodology chapter, the cognitive effort was measured by using two methods: the Pause-to-Character Ratio (PCR) as an objective indicator and the NASA Task Load Index (NASA-TLX) as a subjective indicator.

4.4.1 Pause-to-Character Ratio (PCR)

First, the results of the ratio of pauses to Japanese characters with a 300-ms threshold (300-PCR) will be described. Figure 10 below shows the maximum and minimum values, median, and mean for each text. As illustrated by the boxplot, evidence suggests that the Pause-to-Character Ratio (PCR) for T_1 (HTS) was slightly higher than that of Post-Editing of T_2 (MT); however, T_3 (LLM) required significantly fewer pauses than T_1 (HTS) and T_2 (MT).

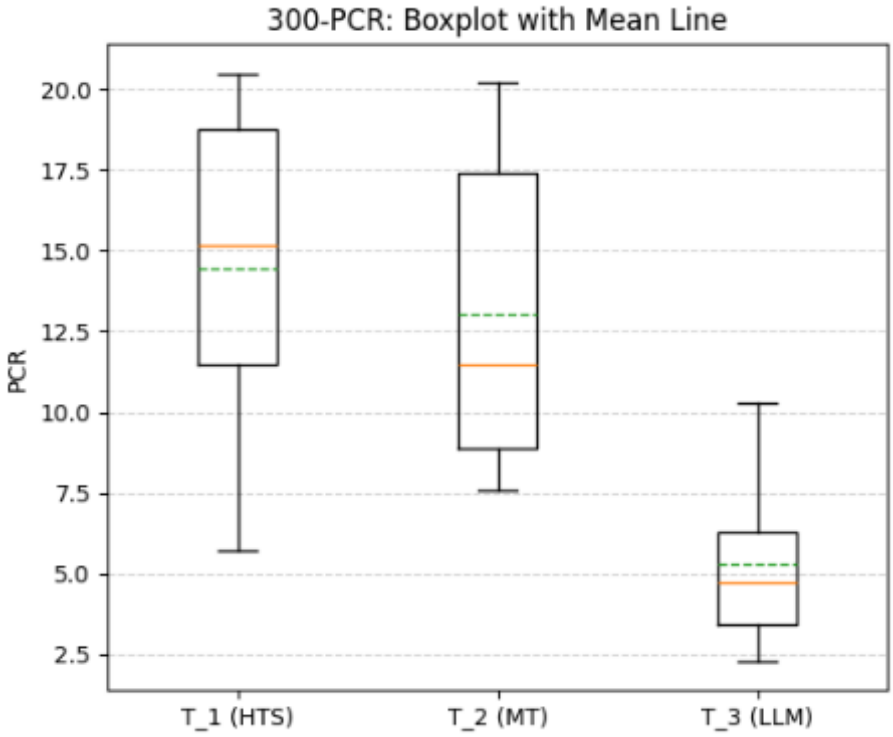


Figure 10: Boxplot of 300-PCR

The summary statistics of the boxplot of 300-PCR are as follows. T1 (HTS): n = 6, $\bar{x} \approx$ 14.5, $\tilde{x}$ = 15.2, Q1 = 10.6, Q3 = 19.6, IQR = Q3 − Q1 = 9.0, range = [5.7, 20.5], no outliers (1.5 × IQR rule). T2 (MT): n = 6, $\bar{x} \approx$ 13.0, $\tilde{x}$ = 11.5, Q1 = 8.2, Q3 = 19.2, IQR = 11.0, range = [7.6, 20.2], no outliers (1.5 × IQR rule). T3 (LLM): n = 6, $\bar{x} \approx$ 5.3, $\tilde{x}$ = 4.8, Q1 = 3.3, Q3 = 6.5, IQR = 3.2, range = [2.3, 10.3], no outliers (1.5 × IQR rule).

Table 8: 300-PCR for each participant on each task

300-PCR	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
P_1	10.6	8.2	6.5
P_2	14.2	11.0	3.3
P_3	16.2	20.2	5.6
P_4	19.6	19.2	10.3
P_5	20.5	12.0	3.9
P_6	5.7	7.6	2.3
Median ($\tilde{x}$)	15.2	11.5	4.8
Mean ($\bar{x}$)	14.5	13.0	5.3

Table 8 above displays the 300-PCR scores for each participant on each task. The PCR scores were found to be substantially higher in T_1 (HTS) than in T_2 (MT) and T_3 (LLM), indicating markedly greater cognitive effort during the process of Human Text-Simplification. T_2 (MT) exhibited intermediate PCR values, with a wider distribution than T_3 (LLM), indicating greater variability in post-editors' cognitive workload for MT outputs than LLM outputs. It can be additionally noticed that P_3 and P_6 required more pauses for Post-Editing of T_2 (MT) than T_1 (HTS). T_3 (LLM) demonstrated minimal objective cognitive load and the most efficient editing process among the three translation conditions, as indicated by consistently low PCR scores.

4.4.2 NASA Task Load Index (NASA-TLX)

Secondly, the results of the NASA Task Load Index (NASA-TLX) will be described. Figure 11 below shows the maximum and minimum values, median, and mean for each text. As demonstrated by the boxplot, the available evidence suggests that the subjective cognitive workload was perceived to be highest in T_1 (HTS), lower in T_2 (MT), and lowest in T_3 (LLM).

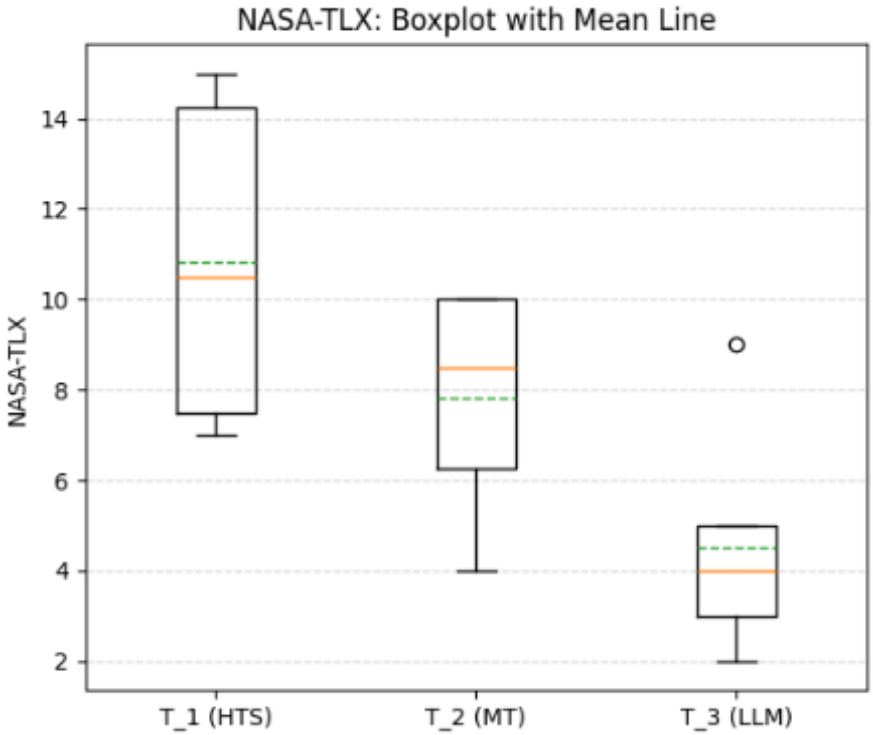


Figure 11: Boxplot of NASA-TLX

The summary statistics of the boxplot of NASA-TLX are as follows. T1 (HTS): n = 6, $\bar{x} \approx 10.8$, $\tilde{x} = 10.5$, Q1 = 7.0, Q3 = 15.0, IQR = Q3 − Q1 = 8.0, range = [7, 15], no outliers (1.5 × IQR rule). T2 (MT): n = 6, $\bar{x} \approx 7.8$, $\tilde{x} = 8.5$, Q1 = 6.0, Q3 = 10.0, IQR = 4.0, range = [4, 10], no outliers (1.5 × IQR rule). T3 (LLM): n = 6, $\bar{x} = 4.5$, $\tilde{x} = 4.0$, Q1 = 3.0, Q3 = 5.0, IQR = 2.0, range = [2, 9], one outlier detected (9), based on the 1.5 × IQR rule.

Table 9: NASA-TLX for each participant on each task

NASA-TLX	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
P_1	7	4	9
P_2	7	6	3
P_3	12	10	5
P_4	15	10	3
P_5	9	7	2
P_6	15	10	5
Median ($\tilde{x}$)	10.5	8.5	4.0
Mean ($\bar{x}$)	10.8	7.8	4.5

Table 9 above displays the NASA-TLX scores for each participant on each task. The analysis demonstrated that cognitive effort was highest in T_1 (HTS), moderate in T_2 (MT),

and lowest in T_3 (LLM). T_1 was found to require the greatest subjective workload, with median and upper-quartile scores that were notably higher than those of the other time points. In contrast, T_3 (LLM) produced consistently low ratings of workload, with only one high-score outlier of P_1, indicating more cognitively effortless processing.

4.5 Readability Score (Quality)

Finally, the results of the jReadability Score (jRS) will be described. Figure 12 below shows the maximum and minimum values, median, and mean for each text. As illustrated by the boxplot, the available evidence indicates that T_3 (LLM) achieved the highest and most consistent readability across participants. T_1 (HTS) demonstrated moderate readability, while T_2 (MT) exhibited the lowest scores.

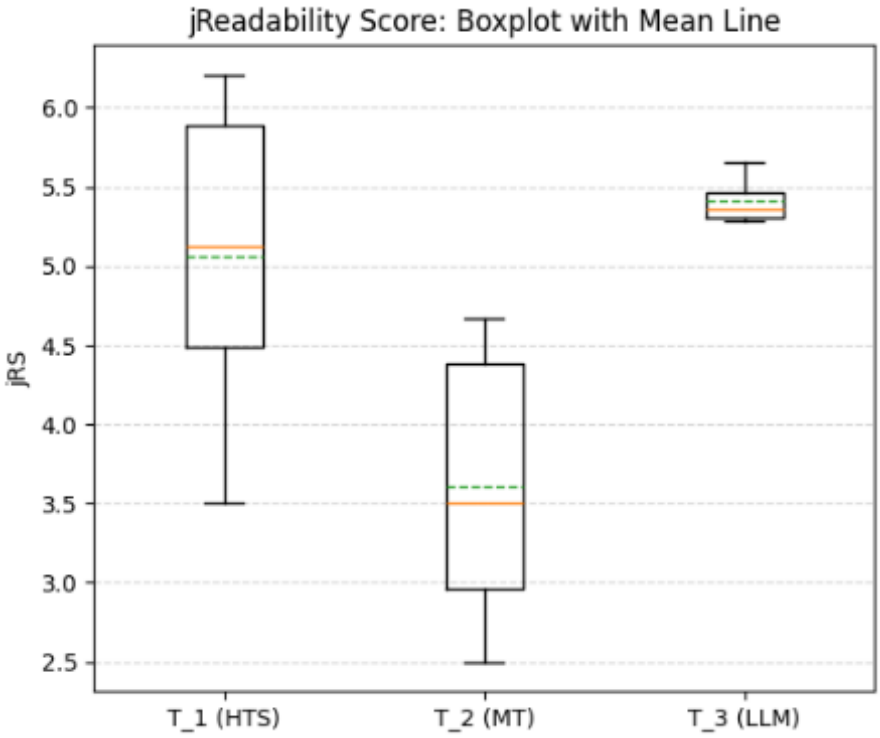


Figure 12: Boxplot of jRS

The summary statistics of the boxplot of jRS are as follows. T1 (HTS): n = 6, $\bar{x} \approx 5.06$, $\tilde{x} = 5.13$, Q1 = 4.32, Q3 = 6.10, IQR = Q3 − Q1 = 1.78, range = [3.50, 6.21], no outliers (1.5 × IQR rule). T2 (MT): n = 6, $\bar{x} \approx 3.61$, $\tilde{x} = 3.51$, Q1 = 2.82, Q3 = 4.64, IQR = 1.82, range = [2.50, 4.67], no outliers (1.5 × IQR rule). T3 (LLM): n = 6, $\bar{x} \approx 5.41$, $\tilde{x} = 5.36$, Q1 = 5.29, Q3 = 5.49, IQR = 0.20, range = [5.28, 5.66], no outliers (1.5 × IQR rule).

Table 10: jRS for each participant on each task

jRS	T_1 (HTS)	T_2 (MT)	T_3 (LLM)
P_1	4.32	2.82	5.39
P_2	3.50	3.38	5.49
P_3	6.21	4.67	5.66
P_4	5.26	2.50	5.28
P_5	6.10	4.64	5.29
P_6	4.99	3.63	5.33
Median ($\tilde{x}$)	5.13	3.51	5.36
Mean ($\bar{x}$)	5.06	3.61	5.41

Table 10 above displays the jReadability Score for each participant on each task. T_1 (HTS) and T_3 (LLM) exhibited analogous high readability values, with T_3 (LLM) demonstrating the greatest consistency across participants. Moreover, the narrow interquartile range (IQR) in T_3 indicates consistent readability across all samples, while T_1 (HTS) and T_2 (MT) exhibited greater variability. T_2 (MT) exhibited notably lower readability compared to HT and the post-edited text of LLM outputs. Overall, T_3 (LLM) was rated as being substantially more legible than both T_1 (HTS) and T_2 (MT).

Additionally, the following diagram illustrates the comparison of the MT/LLM outputs and the target text's readability score. As can be observed from the chart, all the participants increased the readability of the target text (T_2) by the MT-assisted Text-Simplification task over the raw MT output to some extent, with the exception of P_4. In contrast, the readability of the target text (T_3) by the LLM-assisted Text-Simplification task remained consistent with the raw LLM output, exhibiting no significant increase or decrease. The discussion on the readability follows in the later section of the next chapter.

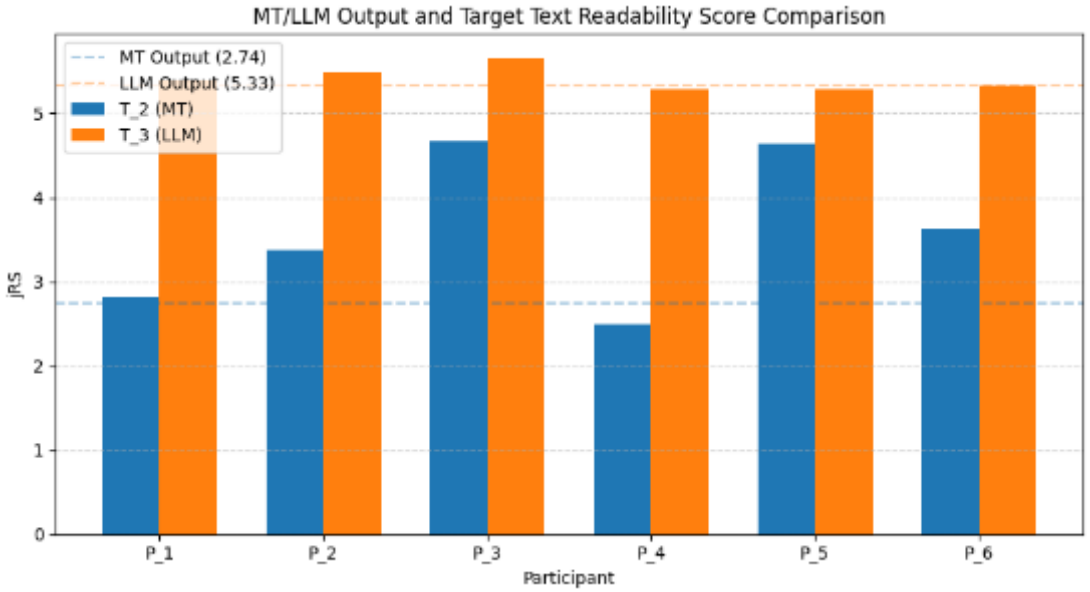


Figure 13: Bar chart on MT/LLM outputs and the target text's readability score

5 Discussion

In this chapter, the results of the intralingual process-oriented translation study and the correlation between each effort and the text readability will be discussed, with reference to previous studies.

5.1 Correlation between Translation Effort

The following six subchapters are especially related to the **RQ1**. Here, the impact of MT tools and LLMs in intralingual translation on the temporal, technical, and cognitive effort will be discussed. Moreover, the discussion is supported by understanding the correlation of translation effort.

RQ1:
What is the impact of working with MT and LLM output in a Japanese intralingual text simplification task on the **temporal, technical, and cognitive effort** of citizen translators with Japanese as their native language?

In the following sections, the correlation between temporal, cognitive, and technical effort in Human Text-Simplification tasks and MT/LLM-assisted Text-Simplification tasks will be analysed and discussed. In the first three sections, the correlation among objective measures (duration, keystrokes per character, and Pause-to-Character Ratio) will be addressed. In the

latter three sections, the correlation between objective measures and a subjective measure (NASA Task Load Index) will be examined.

5.1.1 Temporal Effort and Technical Effort

In the domain of interlingual translation, research shows that there is a strong correlation in Human Translation and Post-Editing tasks (Moorkens et al. 2015). In particular, Alvarez et al. (2020) highlight that temporal effort and keystrokes are directly related.

In this section, it will be analysed how duration (TSC) and keystrokes per character (KSC) correlate with each other under the condition of intralingual translation, based on previous studies. HTER will be excluded from this analysis, since Alvarez et al. (2020: 7–8) mention that there is no direct relationship between duration and HTER, and Macken et al. (2020: 16) also claim that HTER is not a suitable indicator to investigate a relationship between temporal and technical effort. The following diagram illustrates the relationship between TSC and KSC.

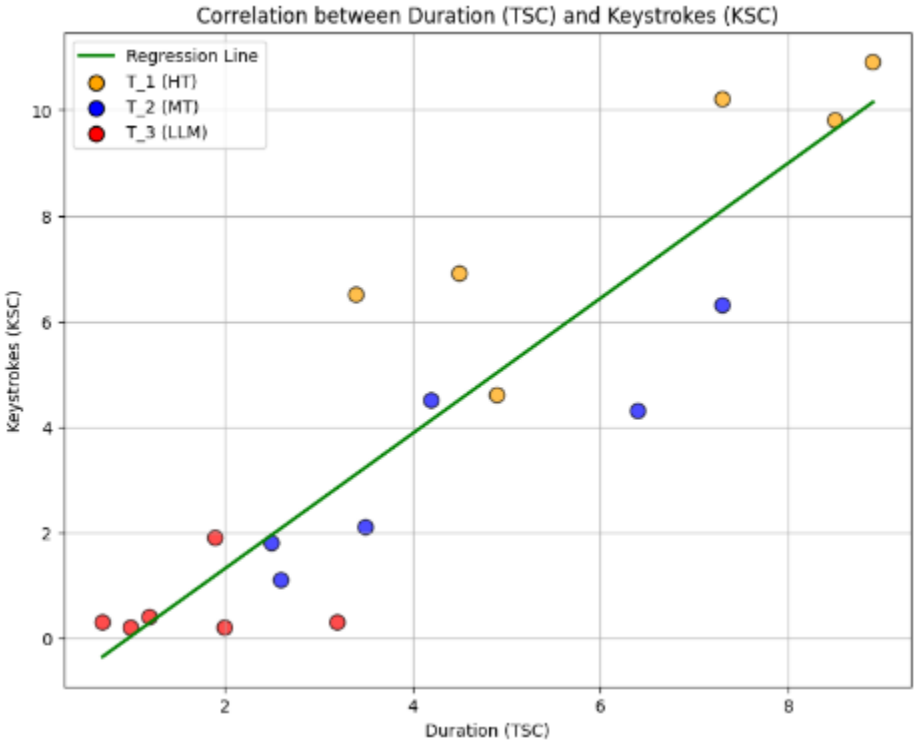


Figure 14: Scatter plot of the correlation between TSC and KSC

As Figure 14 visualises, a robust positive correlation between TSC and KSC can be observed ($r \approx 0.90$). It has been demonstrated that an increase in the duration of translation is indicative of a greater number of keystrokes per character. This suggests that tasks which require a greater degree of editing effort tend to require a greater duration of time. Across

conditions, both duration and keystrokes per character decrease systematically from T_1 (HTS) to T_2 (MT) to T_3 (LLM).

It can be said that the analysed correlation of this intralingual translation study corresponds to the previous research, which claims that there is a robust relationship between temporal and technical effort in Human Translation tasks and Post-Editing tasks (Moorkens et al. 2015), and these two types of effort are directly related to each other (Alvarez et al. 2020).

Moreover, temporal effort can serve as a reliable proxy for technical effort, and vice versa. This is a useful methodological approach, especially if one of the measures is missing or unreliable, not only for the study of interlingual translation, but also for intralingual text simplification.

Furthermore, one important implication of the findings is that reducing temporal and technical effort, particularly in the LLM condition, is highly relevant in crisis communication contexts, where a rapid response is essential. In such situations, the ability to produce clear, accessible text in a short amount of time is crucial, as delays can directly affect the dissemination of information and public safety. Therefore, LLM output, which facilitates faster and more efficient text simplification, can play a pivotal role in facilitating timely and effective communication in urgent situations.

5.1.2 Temporal Effort and Objective Cognitive Effort

As discussed in the previous section, a strong positive relationship between temporal and technical effort can be found in this study. In this section, the correlation between duration and pauses will be discussed. Lacruz mentions that it has been demonstrated that cognitive effort has a substantial impact on temporal effort (Lacruz 2017: 395). In addition, Alvarez et al. (2020: 2) also claim that cognitive effort correlates with temporal effort in the interlingual Post-Editing process.

In this section, the analysis will examine the correlation between duration (TSC) and cognitive load under the condition of intralingual translation, as evidenced by previous studies. As introduced in the methodology chapter, the Pause-to-Character Ratio (300-PCR) is adopted as an indicator of objective cognitive effort in this research. The following diagram illustrates the relationship between TSC and 300-PCR.

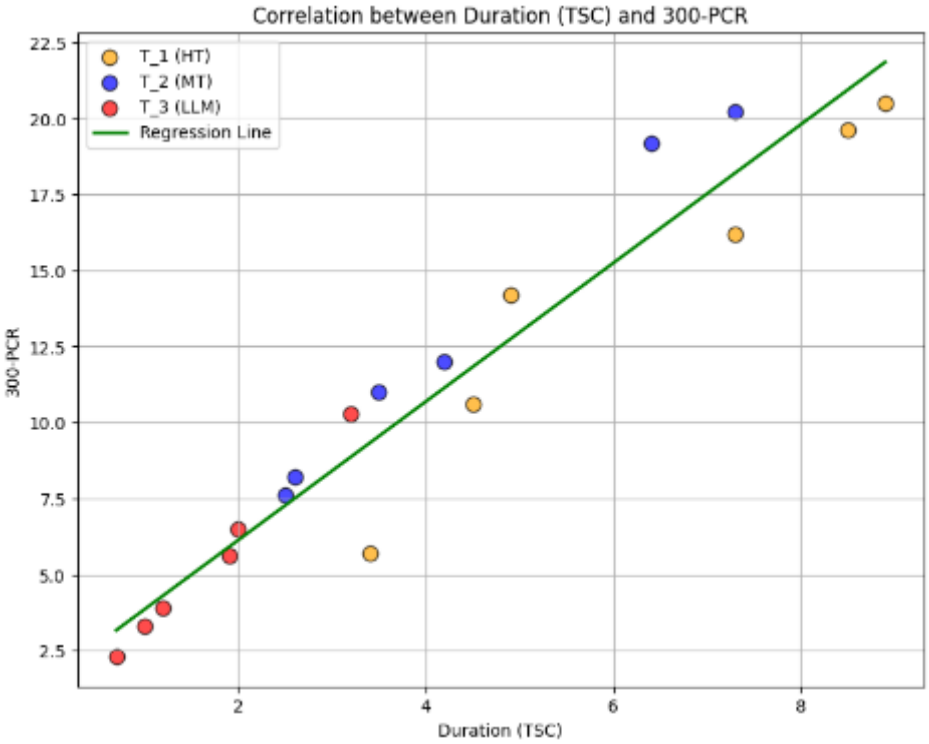


Figure 15: Scatter plot of the correlation between TSC and 300-PCR

As demonstrated in Figure 15, a significant positive correlation is evident between TSC and 300-PCR ($r \approx 0.96$), indicating that there is a strong and reliable relationship between these variables. In the scatter plot, points located towards the upper right indicate longer translation duration and higher cognitive processing load, as reflected by increased 300-PCR values. Across conditions, a systematic shift is observed from T_1 (HTS) to T_2 (MT) and further to T_3 (LLM), characterised by shorter duration and lower 300-PCR. Overall, the analysis demonstrates a robust positive correlation between the duration of translation and 300-PCR, thus indicating that an increase in processing time is associated with greater cognitive load.

In short, as Alvarez et al. (2020) and Lacruz (2017) report, a strong positive correlation between temporal effort and cognitive effort can also be recognised in this intralingual translation study.

Another important consideration in crisis settings is that translators can experience cognitive strain due to the stressful nature of the events. In emergency situations, they frequently work under pressure, processing urgent and potentially distressing information, which increases cognitive load and makes the task more demanding. Tools that reduce the effort required are valuable in these contexts. Working with MT and especially LLM output may help ease this burden by lowering the amount of cognitive effort needed during the simplification process.

5.1.3 Technical Effort and Objective Cognitive Effort

In the final section of the discussion on the correlation among the objective measures, the relationship between keystrokes per character and pauses will be addressed. Cui et al. (2023) also examined the relationship between technical and objective cognitive effort, finding a strong correlation between these objective measures.

In this section, it will be analysed how keystrokes per character (KSC) and the Pause-to-Character Ratio (300-PCR) correlate with each other using the data collected from this intralingual translation study. The following diagram illustrates the relationship between KSC and 300-PCR.

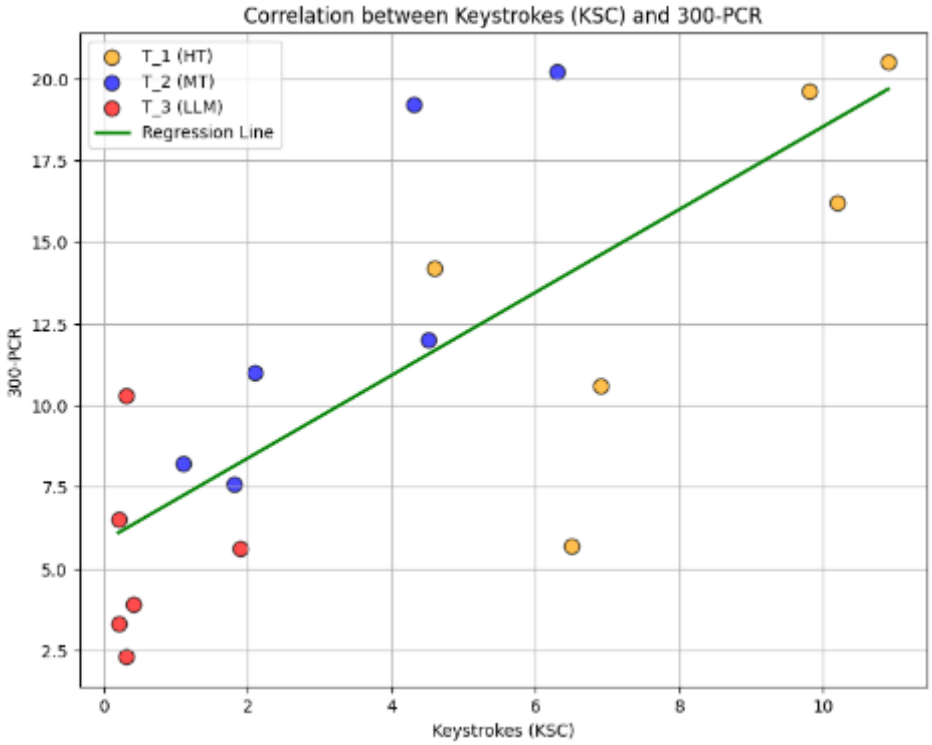


Figure 16: Scatter plot of the correlation between KSC and 300-PCR

As Figure 16 describes, although there is a positive correlation between KSC and 300-PCR ($r \approx 0.76$), the scattering of data is noticeable, and the correlation between them is not stronger compared to that of TSC and KSC or TSC and 300-PCR. While both data of 300-PCR and KSC for T_3 (LLM) are relatively concentrated in the lower left, indicating that both cognitive and technical load were small, some data points of T_2 (MT) and T_1 (HTS) deviate from the regression line, indicating that some translators required significantly more pauses for Post-Editing of T_2 (MT), and others required notably fewer pauses for Post-Editing of T_1 (HTS).

From this data analysis, the discussion suggests that there is a positive relationship between technical and objective cognitive effort, as Cui et al. (2023) also claim; however, it is not as robust as that of temporal and technical effort or temporal and objective cognitive effort.

Moreover, it can also be said that under certain circumstances, translators require extremely higher or lower objective cognitive effort. In addition, the number of pauses required by translators depends on how familiar they are with the keyboard and how fast they can type, since some participants, who do not use a computer frequently in their job or life, were also recruited as citizen translators without any specific experience. For instance, it can be interpreted that the data points for T_1 (HTS), which are located far from the regression line, suggest that the participants are used to typing fast. Furthermore, the data shows that Human Text-Simplification necessitates longer pauses than MT/LLM-assisted Text-Simplification in most cases, although for MT/LLM-assisted Text-Simplification, more reading time is needed, where translators do not actively operate the keyboard or mouse, since they need to read source texts as well as machine-translated texts. This can indicate that cognitive load might be a contributing factor not only during the process of reading but also during the production of written texts.

In addition, there is a weak point in the Translog-II in collecting the data related to technical effort and the number of pauses. The tool can only track actions that occur within the interface. Consequently, the act of moving and scrolling the mouse to refer to the guidelines is disregarded as a technical action, and the time allocated for reading the guidelines is recognised as a pause. This suggests that participants with a low KSC but a high 300-PCR may have dedicated time to reviewing the guidelines. This weak point can be overcome by utilising additional tools and methods, such as eye-tracking or screen recording, in conjunction with a logging tool.

5.1.4 Temporal Effort and Subjective Cognitive Effort

While the previous three sections discussed the correlation among objective measures, the following three sections will focus on the correlation between objective measures and subjective cognitive ratings. Cui et al. (2023: 12) highlight that objective measures are inter-rated with each other, and that objective and subjective measures are also related to each other.

Firstly, the present section will analyse the correlation between duration (TSC) and the NASA-TLX using data collected from this intralingual translation study. The following diagram demonstrates the relationship between TSC and NASA-TLX.

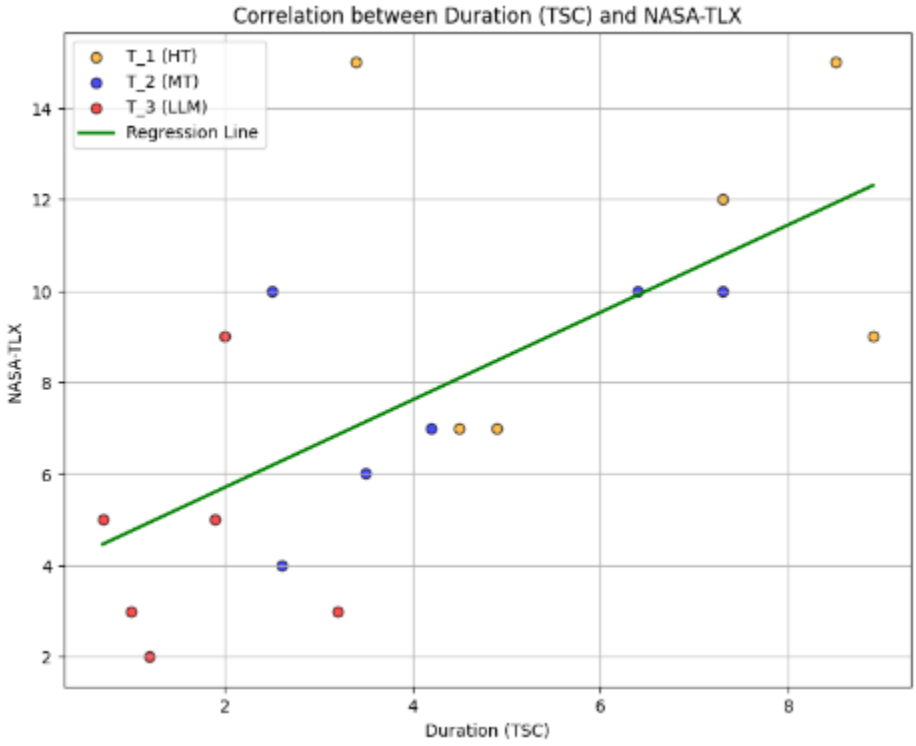


Figure 17: Scatter plot of the correlation between NASA-TLX and TSC

As Figure 17 shows, the correlation between NASA-TLX and TSC is not very strong but moderate ($r \approx 0.64$). It can be inferred that longer processing times would be associated with a higher subjective workload for some participants; however, duration should not be the primary factor for all translators to determine their subjective cognitive rating. Nevertheless, across the different conditions, data points systematically shifted towards the lower-left quadrant from T_1 (HTS) to T_3 (LLM), suggesting some grade reduction in both temporal and subjective cognitive effort.

Overall, the relationship between temporal effort and subjective cognitive workload in this intralingual translation study is not as robust as that asserted by Cui et al. (2023) but is moderate.

5.1.5 Technical Effort and Subjective Cognitive Effort

As the second discussion of the correlation between objective measures and subjective cognitive ratings, the relationship between keystrokes per character and subjective cognitive effort will be addressed in this section. The study conducted by Cui et al. (2023: 13) reports that keystrokes per character and subjective cognitive load also correlate positively.

Based on the previous research data, the analysis on the correlation between KSC and NASA-TLX will be carried out in this section. The following diagram illustrates the relationship between KSC and NASA-TLX.

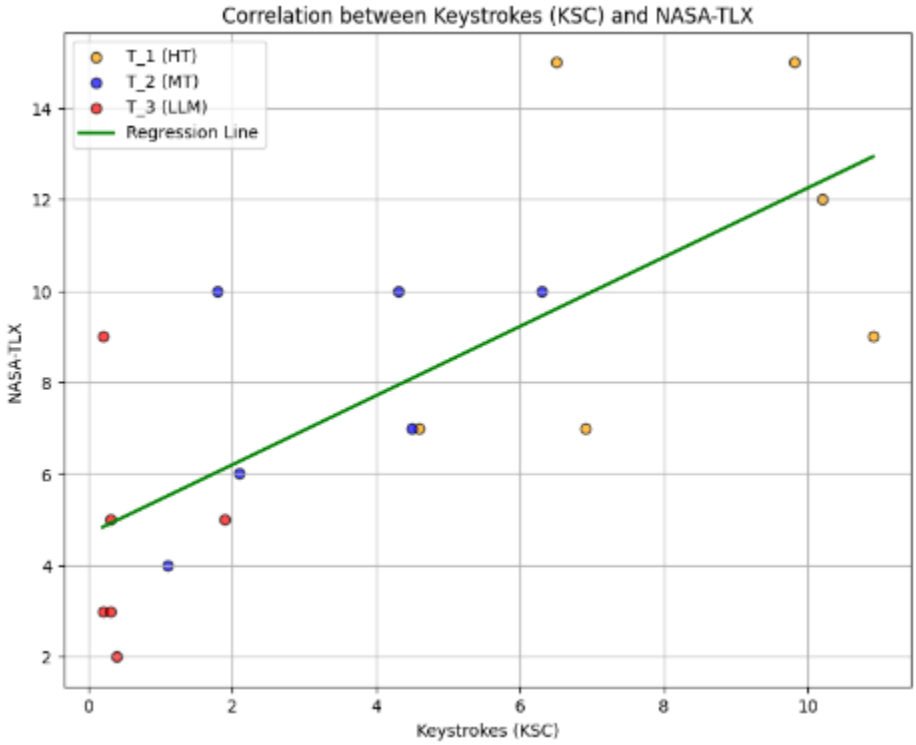


Figure 18: Scatter plot of the correlation between NASA-TLX and KSC

As Figure 18 demonstrates, the correlation between NASA-TLX and KSC is robust ($r \approx 0.72$) and stronger than that between NASA-TLX and TSC. It can be inferred that more technical effort would be associated with a higher subjective workload. Moreover, the systematic shift from T_1 (HTS) to T_2 (MT) and further to T_3 (LLM) towards lower keystroke per character counts and reduced NASA-TLX scores suggests a consistent reduction in perceived cognitive load. This indicates that participants experience a greater cognitive workload when more typing and editing are required. In short, it can also be said that subjective cognitive rating may have a more direct correlation with technical effort rather than temporal effort.

5.1.6 Objective Cognitive Effort and Subjective Cognitive Effort

As the final discussion of the correlation between objective measures and subjective cognitive ratings, this section will address the relationship between 300-PCR and subjective cognitive effort. The research conducted by Cui et al. (2023: 13) claims that pauses and subjective cognitive load also correlate positively.

Based on the previous research data, this section will present the analysis of the correlation between 300-PCR and NASA-TLX. The following diagram illustrates the relationship between 300-PCR and NASA-TLX.

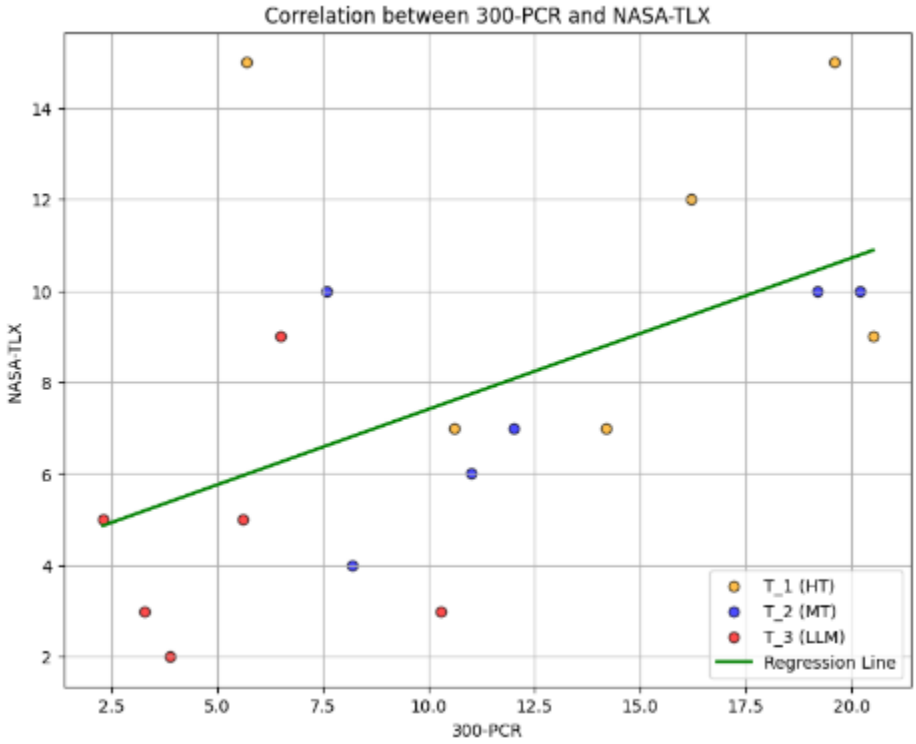


Figure 19: Scatter plot of the correlation between NASA-TLX and 300-PCR

As Figure 19 illustrates, although there is a positive correlation between NASA-TLX and 300-PCR ($r \approx 0.52$), it is moderate and weaker than the relationship between NASA-TLX and either duration or keystrokes per character. This can indicate that the length of pauses may not always have an impact on the subjective cognitive rating. This can be associated with the fact that translators tend to feel cognitive load if the translation and Post-Editing process takes a long time and requires more keyboard and mouse operations. It can be inferred that the length of reading and thinking time, which is recorded as pauses, might be difficult to perceive directly for translators in comparison to technical or temporal effort. For this reason, subjective cognitive effort cannot always be linked to objective cognitive effort.

Furthermore, there are various possible causes of the variability in the subjective cognitive effort data. This might be, for instance, due to the order in which the texts were translated by the participants, or the difference in which text was used as the standard for comparing the degree of effort. In this experiment, the participants evaluated their subjective cognitive effort using only numerical values. However, it would be interesting to conduct retrospective interviews in a further qualitative study to identify extreme examples and understand how and why translators evaluate their effort.

With regard to cognitive effort in general, it is important to note that the experiment was conducted under controlled laboratory conditions. As such, the participants were not exposed to the stressors typically present in a real crisis. This means that both the objective and

subjective measures of cognitive effort observed in this study may not be representative of those that would be observed under actual emergency conditions, where heightened stress and pressure are likely to influence cognitive processing.

In the last three sections, the correlation between subjective cognitive effort and three types of objective effort was discussed. It has become evident that each relationship between those has a robust or moderate positive correlation. However, the analysis's findings indicate a strong positive correlation between NASA-TLX and KSC ($r \approx 0.72$) compared to TSC ($r \approx 0.64$) or 300-PCR ($r \approx 0.52$), suggesting that technical effort may be the primary factor in determining subjective cognitive rating, at least in this experiment.

5.2 Text Readability and Translation Effort

The following six subchapters are especially related to **RQ2**. Here, the impact of MT tools and LLMs in intralingual translation on the readability of the target text will be discussed. Moreover, the discussion is supported by understanding the correlation between translation effort and the source text and MT/LLM output readability.

RQ2:
What is the impact of working with MT and LLM output in a Japanese intralingual text simplification task on the **readability of the target text**?

In the following sections, it will be analysed and discussed how the readability of source texts, as well as MT/LLM outputs, should influence temporal, cognitive, and technical effort.

5.2.1 Source Text Readability and Temporal Effort

Several interlingual process-oriented studies (Cui et al. 2023; Tatsumi 2009) show that participants spent more time on Human Translation compared to Post-Editing, indicating that more temporal effort was required in Human Translation tasks than in Post-Editing tasks. While MT systems have a positive effect on the temporal effort of interlingual translation and Post-Editing, it was found in this intralingual process-oriented study that MT tools and LLMs have the potential to shorten the time spent on translating Japanese warning texts about disaster information into Easy Japanese by citizen translators who have Japanese as their first language but do not have any experience in text simplification. In other words, MT tools and LLMs also have a positive effect on the temporal effort of intralingual translation and Post-Editing.

However, as the results show, compared to the consistently and significantly reduced temporal effort (73.8%) using an LLM, Post-Editing of T_2 (MT) was just 36.1% faster than Human Text-Simplification. Studies show that Post-Editing effort varies with sentence length, structure and source text characteristics, with very short or long sentences often taking longer to translate (Koponen 2016). Research also indicates that readability affects cognitive effort, though its impact on time and technical effort is weaker (Dai & Liu 2024). High-quality NMT systems can reduce the negative effects of low readability.

Based on the studies mentioned above, the correlation between duration (TSC) and source text readability score by jReadability will be discussed. The following diagram illustrates the relationship between TSC and the readability scores of the source texts.

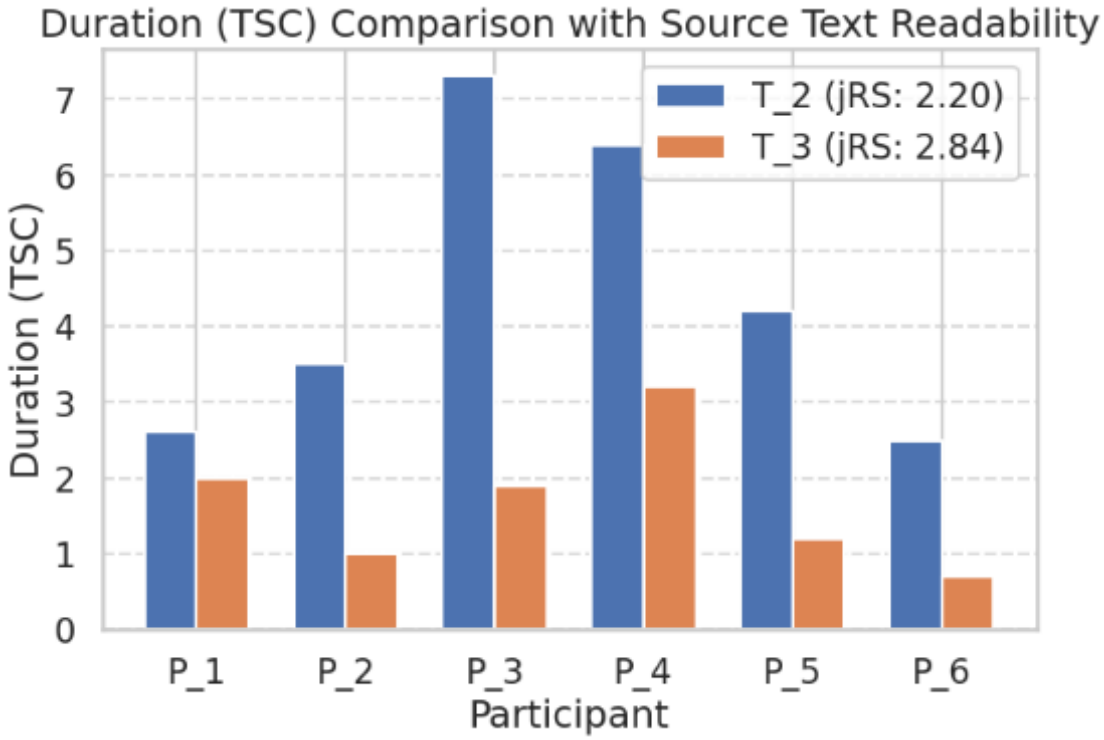


Figure 20: Chart of correlation between TSC and source text readability

As visualised in Figure 20 above, there is a clear correlation between higher source text readability scores and faster Post-Editing speeds. The readability score of T_3 (2.84) is higher than that of T_2 (2.20), and it can be observed that each translator required less time for each MT/LLM-assisted Text-Simplification task. The higher the readability of the source text, the less time per character translators require, indicating reduced time processing effort. The prevailing tendency indicates that text readability is a robust predictor of Post-Editing efficiency.

Regarding T_1 (HTS), however, as can be observed from Figure 7 and Table 4, even though the source text readability score of T_1 (3.33) is higher than the other texts, the longest time was required for each translator to accomplish the Human Text-Simplification task. From these statistics, it can be inferred that, irrespective of the source text readability, Human Text-Simplification is the most time-consuming task in comparison to MT/LLM-assisted Text-Simplification.

Overall, in contrast to the study of Dai & Liu (2024), the findings of this study suggest a correlation between higher source text readability scores and faster Post-Editing speeds.

5.2.2 MT/LLM Output Readability and Temporal Effort

In the previous section, it was discussed how the readability of the source texts can affect Post-Editing speed. Next, the correlation between temporal effort and MT/LLM output quality will be addressed in this section. Sanchez-Torron & Koehn (2016) researched MT quality and Post-Editing productivity in the interlingual translation discipline. Their study shows a linear relationship between the quality of MT and Post-Editing speed, indicating that higher-quality MT outputs require less effort.

As this study, however, handles intralingual translation, and BLEU is therefore not suitable to measure MT/LLM output quality, the readability score by jReadability will be adopted to measure MT and LLM output quality, analogous to the measurement of source text readability. It is essential to clarify from the onset that the correlation between technical effort and other efforts will be discussed using HTER as a measurement of technical effort in the later sections. The following diagram shows the relationship between duration (TSC) and the machine-translated texts' readability scores.

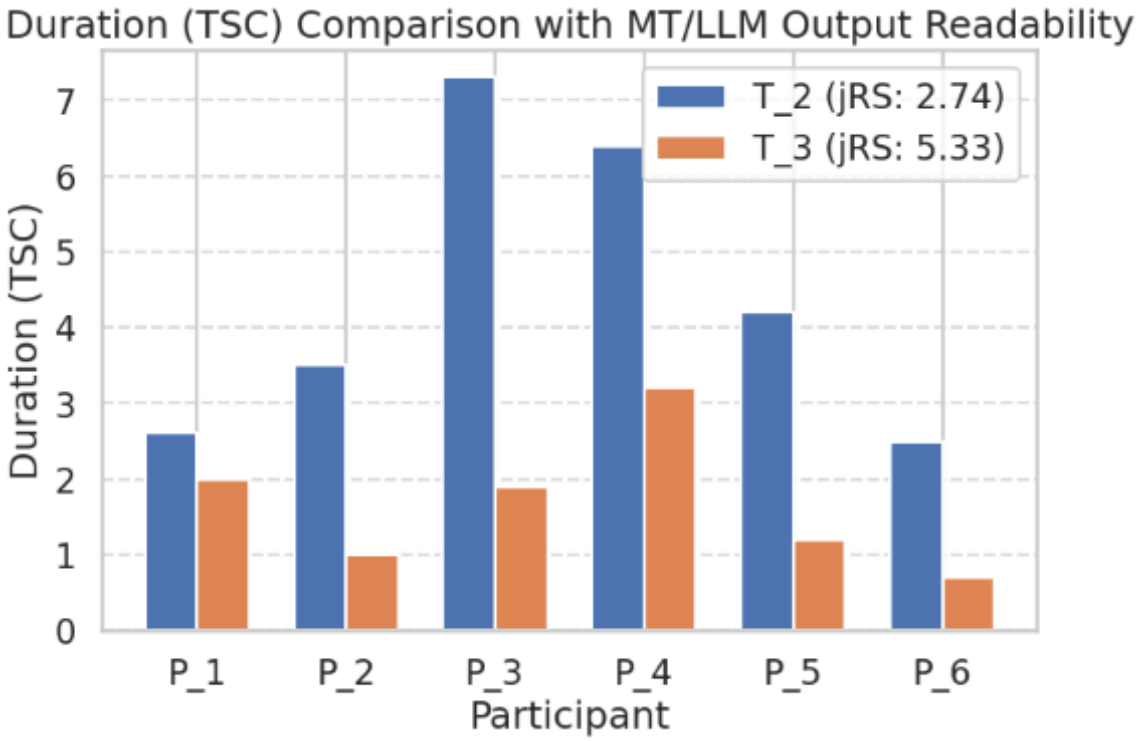


Figure 21: Chart of correlation between TSC and machine-translated text readability

As visualised in Figure 21 above, there is also a clear correlation between higher machine-translated texts' readability scores and faster Post-Editing speeds. The output of T_2, which was machine-translated by Tsutaeru-Web, scored 2.74, indicating a slight increase in readability against its original text (2.20). The output of T_3, which was machine-translated by Gemini, scored 5.33, indicating a significant increase in readability against its original text (2.84). A robust and statistically significant negative correlation was identified between the readability score of MT/LLM outputs and Post-Editing duration. This finding suggests that texts which are more readily comprehensible tend to be post-edited more expeditiously. The observed relationship appears to be unlikely to be due to chance, suggesting that there may be a meaningful link between the readability of machine-translated texts and Post-Editing speed.

Dai & Liu (2024) mention the impact of the source text quality on the MT output as follows:

> [...] it should be noted that the impact of the source text is not only confined to the allocation of cognitive resources, but could be extended to the resources allocated to the MT output, since source text features are mirrored in the MT to a certain extent. (Dai & Liu 2024: 208)

This phenomenon can also be observed between the readability scores of the source texts and the MT/LLM outputs in this study, as the source text with a low readability score marked a lower readability score for a machine-translated text and vice versa for the text with a high

readability score. From this perspective, not only the quality of source texts but also that of machine-translated texts can influence the Post-Editing speed. In summary, as Sanchez-Torron and Koehn (2016) suggest, low-quality MT outputs require more temporal effort.

5.2.3 Source Text Readability and Technical Effort

In this section, the correlation between source text readability and technical effort, especially keystrokes per character (KSC) and HTER, will be analysed and discussed. Tatsumi & Roturier's (2010) research examined the source text characteristics and the technical Post-Editing effort, identifying a significant correlation between ambiguity and complexity and the technical Post-Editing effort.

Based on the study above, the correlation of the readability of the source text and the technical effort required in intralingual translation will be discussed. Similar to the analysis of the relationship between source text readability and temporal effort in the previous sections, the jReadability Score will be used as a measure of source text readability. As measurements of technical effort, keystrokes per character (KSC) as a main indicator and HTER as a supportive indicator will be used. The following diagram shows the relationship between keystrokes per character (KSC) and the source texts' readability scores.

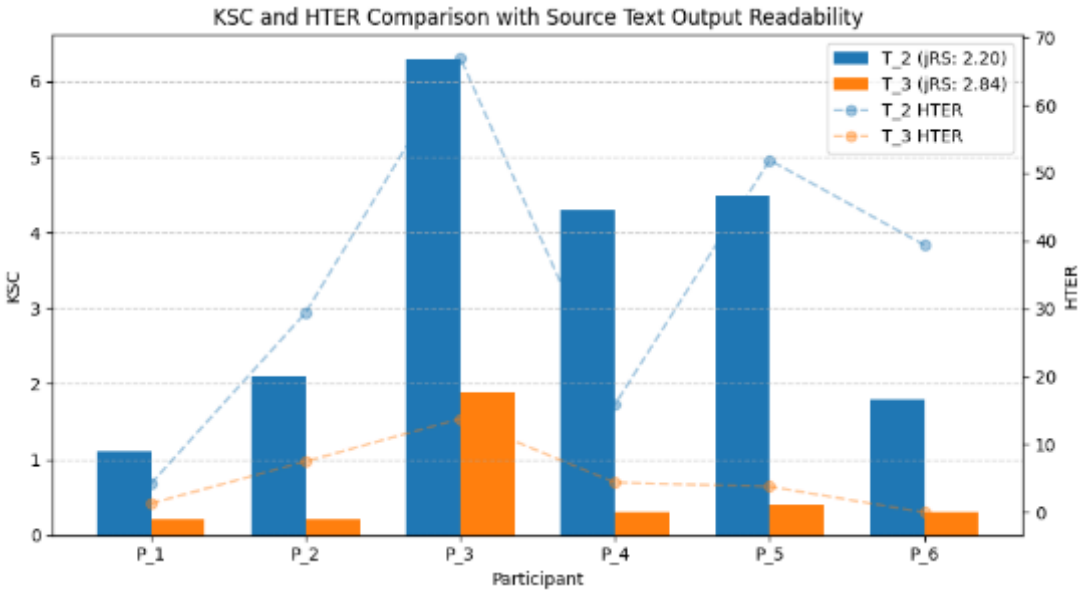


Figure 22: Chart of correlation between KSC and source text readability

Figure 22 illustrates that the source text with higher readability is associated with fewer keystrokes per character, while larger HTER values cluster of T_2 (2.20), indicating increased technical effort for Post-Editing compared to T_3 (2.84). Translators necessitate a greater degree of technical effort when confronted with source texts characterised by reduced

readability. This observation signifies that a more substantial investment of technical effort is required, encompassing deletion, insertions, text re-ordering and other pertinent activities, in order to ensure the delivery of final texts that meet the requisite standards of quality. This finding largely corresponds to the results of the study by Tatsumi & Roturier (2010), which demonstrated that the complexity of source texts and the technical effort of Post-Editing should correlate.

Additionally, even though the source text readability score of T_1 (3.33) is the highest among the three texts, as can be observed from Figure 7 and Table 4, the most keystrokes per character were required to accomplish the Human Text-Simplification task. These statistics indicate that, regardless of source text readability, Human Text-Simplification is the most technically demanding task compared with MT/LLM-assisted Text-Simplification.

Furthermore, as demonstrated in Figure 22, the KSC and HTER scores of the participants largely correspond to one another in both T_2 and T_3. However, it is noteworthy that P_4 exhibited a comparatively high proportion of KSC in T_2, suggesting that the participant engaged in a substantial number of actions, including typing and deleting. However, the HTER score is found to be very low, suggesting that the editing distance between the raw MT output and the target text was minimal. It is evident that the working methods employed by T_4 were not as efficient as those utilised by the other participants when observed from the perspective of technical effort. This may be attributable to insufficient experience and training in text simplification tasks and Post-Editing.

5.2.4 MT/LLM Output Readability and Technical Effort

Next, the correlation between the readability of MT/LLM outputs and the technical effort will be analysed and discussed. Stasimioti & Sosoni (2020a) researched the impact of the SMT and NMT outputs' quality on the technical effort of MTPE tasks. The automatic metrics and the human evaluation exhibited higher scores in the Post-Editing of the NMT output, which led to a marginal decline in technical effort (Stasimioti & Sosoni 2020a: 170). These results suggest that high-quality MT output requires less technical effort in MTPE tasks.

In this discussion, the correlation between the readability of the MT and LLM outputs and the technical effort required in intralingual translation will be examined, as outlined in the preceding study. As was the case in the preceding sections, which analysed the correlation between the readability of the source text and the temporal effort involved, the jReadability Score will be employed as a measurement of the readability of the source text. The measurement of technical effort will be achieved by the use of keystrokes per character (KSC) as a primary

indicator and HTER as a secondary indicator. The following diagram illustrates the relationship between keystrokes per character (KSC) and the readability scores of the MT/LLM outputs.

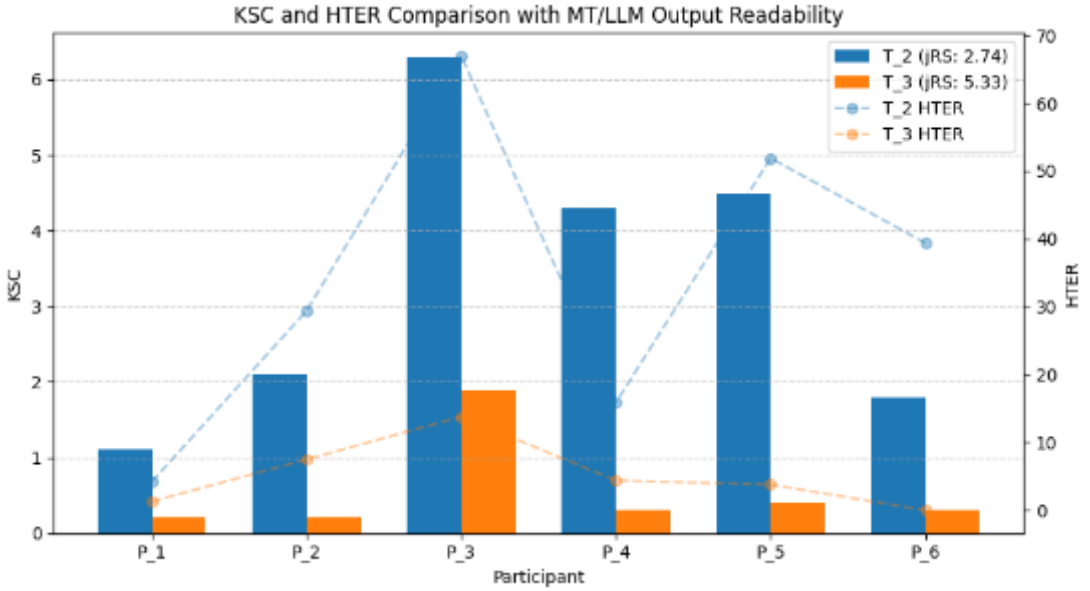


Figure 23: Chart of correlation between KSC and source text readability

As Figure 23 describes, higher readability scores of MT/LLM outputs are associated with a substantially reduced number of keystrokes per character and lower HTER values, indicating a reduction in the technical effort required during the Post-Editing stage. It can be observed from the diagram that each participant required more KSC for the MT/LLM-assisted Text-Simplification task of T_2 (2.74) compared to that of T_3 (5.33). This finding suggests that more readable machine-translated outputs lower the technical load for translators. The observed strong and statistically significant negative correlation further supports the claim that MT/LLM output readability is a key predictor of Post-Editing efficiency.

As mentioned above, Stasimioti & Sosoni (2020a: 170) highlight in their study that more fluent and grammatical MT outputs, which achieve higher scores on automatic metrics, require less technical effort in Post-Editing tasks. Corresponding to the previous study, the analyses of this study revealed that machine-translated texts, which are more readable, require fewer technical activities such as deletion, insertions, and text re-ordering in MT/LLM-assisted Text-Simplification.

Additionally, as mentioned in the findings chapter, the HTER score for LLM-assisted Text-Simplification (T_3) by P_6 was “0”, indicating that P_6 did not undertake any editing whatsoever and delivered the LLM output as the final product. Therefore, the target text readability score stays at 5.33, which is the same as that of the raw LLM output. It can be argued that the participant might have conducted some editing, but decided to use the raw LLM output

at the end, because the HTER score reflects only the editing distance between the raw MT/LLM output and the post-edited text.

[Start]••[•01:34.484][Stop]

Figure 24: Linear view of T_3 by P_6

However, as shown in Figures 23 and 24, the KSC was minimal, and there were only pause plots in the linear view, indicating that no editing was performed; instead, the mouse was moved just to stop logging. Nevertheless, the target text readability score was not the lowest among all the participants, suggesting that, in cases where MT tools or LLMs can generate texts of a sufficiently high standard, the need for further Post-Editing may be reduced or eliminated. This can be very important in making sure that messages are provided quickly and are clear and easy to understand, especially in crises.

5.2.5 Source Text Readability and Cognitive Effort

In this section, the correlation between source text readability and cognitive effort will be analysed and discussed. Dai & Liu (2024: 222) claim that source text readability has a significant effect on cognitive effort.

Based on previous research, this section will present an analysis of the correlation between the jReadability Score and objective cognitive measures. The following diagram illustrates the relationship between the jRS of the source texts and 300-PCR.

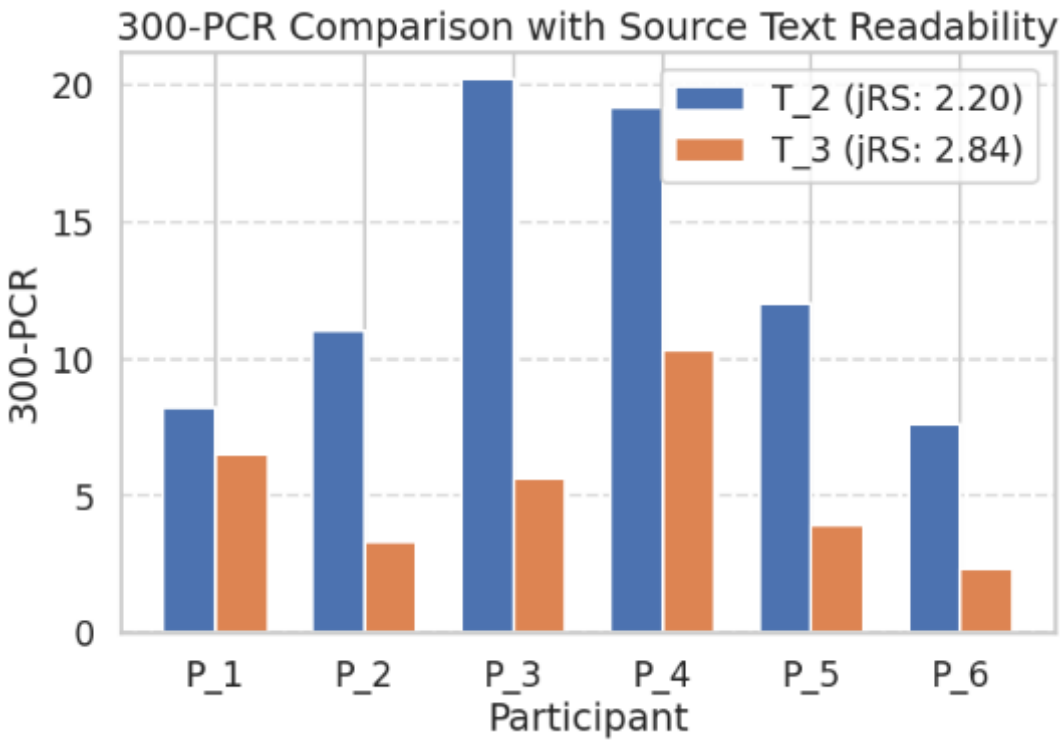


Figure 25: Chart of correlation between 300-PCR and source text readability

Figure 25 illustrates that the higher readability scores of source texts are consistently associated with fewer pauses across all participants during MT/LLM-assisted Text-Simplification. It can be understood that the translators required more pauses for Post-Editing of T_2 (2.20) compared to T_3 (2.84). In other words, enhanced source text readability leads to a reduction in cognitive processing load during the Post-Editing stage. This finding largely corresponds to the results of the study by Dai & Liu (2024: 222), which demonstrated that the complexity of source texts and the cognitive effort required for Post-Editing should be correlated.

Furthermore, as can be observed from Figure 7 and Table 4, it required the most pauses to accomplish the Human Text-Simplification task, although the source text readability score of T_1 (3.33) is the highest among the three texts. As demonstrated by the statistics, regardless of the readability of the source text, Human Text-Simplification is the most cognitively demanding task compared to MT/LLM-assisted Text-Simplification.

5.2.6 MT/LLM Output Readability and Cognitive Effort

Finally, the relationship between the readability of MT/LLM outputs and cognitive effort will be addressed in this section. Lacruz (2017: 392) highlights that lower MT quality can be associated with more cognitive load.

In accordance with extant research, the correlation of MT/LLM output and cognitive effort will be analysed below. As an indicator of the quality of machine-translated texts, the jReadability Score will also be adopted in this discussion. The following diagram illustrates the relationship between the jRS of MT outputs and 300-PCR.

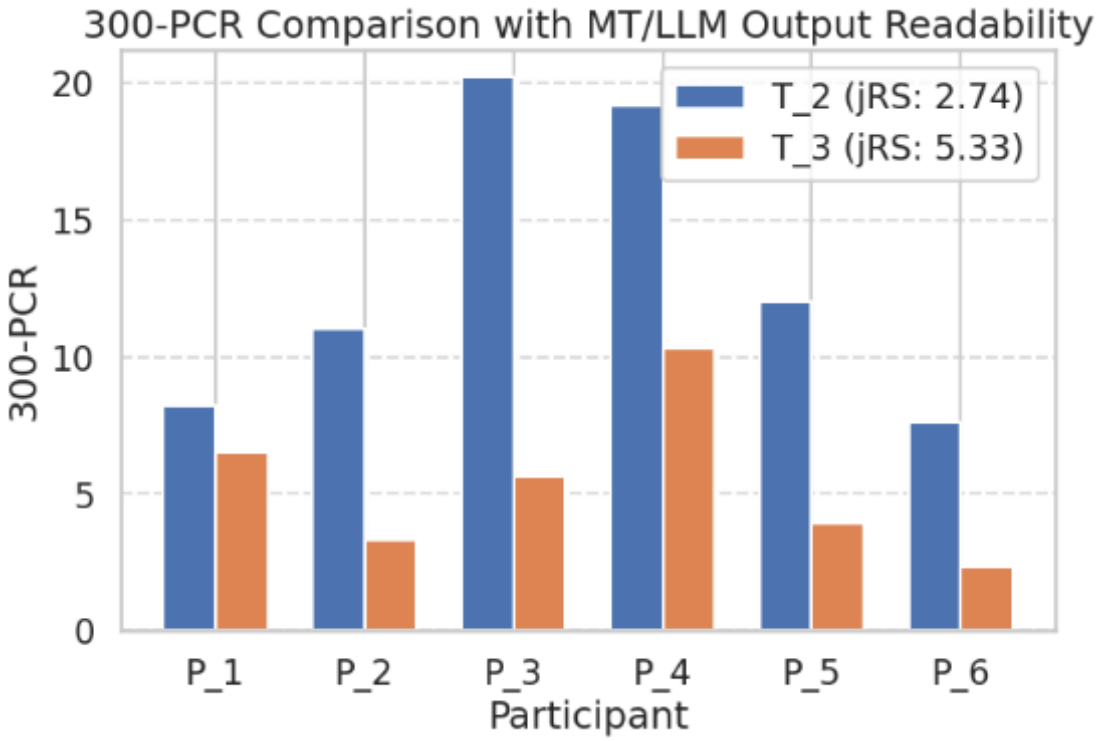


Figure 26: Chart of correlation between 300-PCR and source text readability

As demonstrated in Figure 26, it can be observed that translators required more pauses for the Post-Editing of T_2 (2.74) than for T_3 (5.33), indicating that a lower readability score is associated with higher cognitive load in MT/LLM-assisted Text-Simplification. Across all participants, higher readability scores have been consistently demonstrated to be associated with fewer pauses, indicating reduced cognitive load. In other words, enhanced readability of MT/LLM output results in more efficient Post-Editing. This analysis illustrates that working with Gemini (LLM) is less cognitively demanding compared to Tsutaeru-web (MT) in this intralingual translation study.

In summary, the findings of this study correspond to previous research by Lacruz (2017: 392) that the higher the readability of the MT/LLM output, the less cognitive effort is required for MT/LLM-assisted Text-Simplification.

In the final six sections, the correlation between the readability of source texts or MT/LLM outputs and different types of the Post-Editing effort has been discussed. From the analysis above, it can be concluded that both the readability of source and machine-translated texts influences the grade of the efforts in the MT/LLM-assisted Text-Simplification stage. It has been demonstrated that as the readability of source texts and MT/LLM outputs increases, the temporal, technical, and cognitive effort required for MT/LLM-assisted Text-Simplification decreases. Figure 27 describes this relationship between the readability of the source text and the MT/LLM output, and the readability of the target text and translation effort.

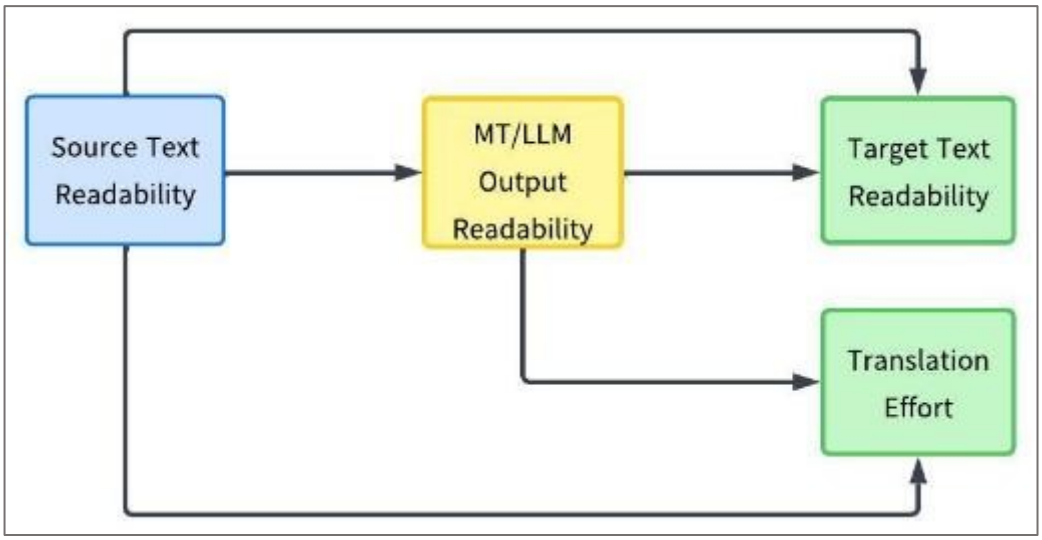


Figure 27: Flowchart illustrating the relationship between readability and translation effort

Moreover, as Dai & Liu (2024: 209) suggest, the complexity of source texts has a negative impact on the quality of MT outputs, necessitating a greater degree of Post-Editing effort. Furthermore, the findings demonstrate that when source and machine-translated texts are characterised by complexity, there is often a corresponding decline in the target texts' readability scores, even though translators invest substantial temporal, technical, and cognitive effort.

As mentioned in the finding section, Figure 13 illustrates that the majority of participants exhibited enhanced readability in T_2 through the implementation of MT-assisted simplification, with the exception of P_4. Conversely, T_3, derived from LLM-assisted simplification, demonstrated no significant deviation from the original output. However, this fact implies that if the quality of raw MT/LLM output is quite high, for instance, especially T_3 in this case, it is difficult to make the text simpler, even though translators put in a great deal of effort. Moreover, as shown in the diagram, making unnecessary changes may actually result in a lower standard of quality. Another interesting point is that all the participants got a better score in Human Text-Simplification (T_1) rather than MT-assisted Text-Simplification (T_2). This might be since the source text readability of T_1 (3.33) was higher than that of T_2 (2.20) and even the raw MT output readability of T_2 (2.74). However, it is still unclear whether the quality of source text or MT/LLM output has a greater impact on the target text. This would be clarified in further experiments involving different combinations of source texts and MT tools or LLMs.

Furthermore, it should be noted that the experiment was carried out by the participants without any experience in Post-Editing and text simplification, although guidelines for Easy

Japanese were provided beforehand. The involvement of experienced or trained translators in the experiment could have resulted in enhanced stability in the quality of the target texts, and they might have been less influenced by the degree of the MT/LLM output quality.

A further point for discussion is the quality assessment of the experiment, which was conducted with the help of the jReadability tool. This tool is designed to calculate the objective readability score of the target texts. This system is beneficial for translators, as it enables them to quickly assess the ease and accessibility of the simplified target text. Additionally, the tool is open-source and cost-free, and it is efficient in terms of both time and cost, which is crucial in crisis settings. However, it is imperative to acknowledge the significance of the involvement of the primary target group in the domain of accessibility (Maaß 2020: 136–137). In consideration of the preceding points, the quality assessment in future research may be conducted by the target users of Easy Japanese, such as foreign residents in Japan, as this was not possible in this thesis. Consequently, the objective readability score may not be the same as the subjective judgment. Moreover, the involvement of the target group in text assessment may contribute to the development of MT/LLM systems for text simplification.

6 Limitations and Future Work

Even though the scale of this research was small, the desired data was collected, and the objectives of the research were achieved. However, it should be noted that the study had several limitations.

At the beginning, it is highly possible to offer a critique of this master's thesis in its entirety. This is due to the fact that the methodology for the experimentation was borrowed and repurposed from the one employed for the process-oriented study of interlingual translation. Furthermore, the majority of the discussion was also conducted based on the comparison with prior research on interlingual translation, which might not be suitable for a direct comparison with intralingual text simplification. However, as mentioned in the literature review, there is hardly any research on the process-oriented intralingual translation, which can be cited in this thesis at the moment. Although such criticism was therefore to be expected, the same methodology was employed, and the findings were discussed based on the previous interlingual translation studies. Nonetheless, a considerable number of results were consistent with previous studies. Whether this is merely a coincidence or not will likely become clear through further

research. Therefore, it is hoped that this master's thesis will serve as a catalyst for further research on process-oriented intralingual translation study.

Next, participants were limited to only six volunteer translators who, at the time of the experiment, possessed neither any professional translation experience nor any knowledge of Easy Japanese. In future work, some Japanese language teachers or public servants working in a cultural affairs division may be recruited to find whether translators with professional experience can steadily perform more efficiently and provide final products of higher quality. In addition, ideally, more participants would be available so that more reliable quantitative data could be collected.

Furthermore, only one MT system (Tsutaeru-Web) and one LLM (Gemini) were used and compared in this study. Although there are only a few MT tools for Easy Japanese available on the market at the moment, the company is continually training the system with more training data, which should improve the quality of MT outputs. Regarding LLMs, several LLMs, such as ChatGPT or Copilot with different versions, can be introduced in future research. In this study, the LLM was just instructed to translate the source texts into Easy Japanese in accordance with the Easy Japanese guidelines, which were uploaded into the system. However, it is possible to use other updated guidelines or to do further prompting.

Finally, it is also possible to use other methodologies and tools to collect data. In this research, a keylogging tool (Translog-II), which can also record pauses, was adopted to collect the data on cognitive effort. In future work, the use of an eye-tracking system, Think-aloud protocols or the combination of those tools and methodics could be considerable.

7 Summary and Conclusion

This master's thesis comprises a comparative process-oriented study of the applicability of MT and LLMs for the simplification of Japanese texts in crisis settings. Easy Japanese is in high demand among non-native speakers in Japan, often surpassing the demand for English. However, public servants often lack the necessary manpower, expertise and training, leading them to rely on MT tools for simplification.

In view of the extant research demonstrating productivity and quality gains from MTPE in interlingual translation, the following research questions are posited, and the findings, as derived from an analysis of the data, provide the following conclusions.

RQ1:

What is the impact of working with MT and LLM output in a Japanese intralingual text simplification task on the **temporal, technical, and cognitive effort** of citizen translators with Japanese as their native language?

The findings show both positive and negative impacts of working with MT and LLMs in a Japanese intralingual translation task. Compared to Human Text-Simplification, the data from working with the LLM show strong gains in readability and productivity, in terms of temporal, technical, and cognitive effort. The duration of Post-Editing (T_3) was 73.8% shorter than translation from scratch (T_1). The Post-Editing task (T_3) required 96.4% fewer keystrokes per character, indicating less technical demand. The participants required 68.4% fewer pauses, and the mean score of NASA-TLX was 61.9% smaller, which indicates that the LLM-assisted Text-Simplification task (T_3) was both objectively and subjectively less cognitively demanding. Furthermore, the readability of the post-edited text (T_3) was 5.9% higher than the human-translated text.

On the other hand, compared to the LLM, the productivity gain using the MT system was more moderate, and the readability of the post-edited final text does not constitute an improvement compared to Human Text-Simplification. In addition, the duration of Post-Editing (T_2) was 36.1% shorter than translation from scratch (T_1), and the MT-assisted Text-Simplification task (T_2) required 61.9% fewer keystrokes per character, indicating a moderate reduction in technical demand. The participants demonstrated a 24.3% reduction in pauses, and the mean score on the NASA-TLX scale decreased by 19.0%. These findings suggest that the reduction of the objective and subjective cognitive demands of the MT-assisted Text-Simplification task (T_2) was also moderate. In addition, the readability of the post-edited text (T_2) was even 31.4% lower than the human-translated text.

Furthermore, it has been discussed how objective measures, such as duration, keystrokes per character, and pauses, correlate with one another and with a subjective measure. The analysis shows that there are positive correlations both among objective measures and between objective and subjective measures. However, the robustness of the correlation varies. For instance, a strong relationship between TSC and NASA-TLX could be found, which indicates that the technical effort may be the primary factor in determining the subjective cognitive rating.

RQ2:

What is the impact of working with MT and LLM output in a Japanese intralingual text simplification task on the **readability of the target text**?

The findings suggest that, moreover, if the source and machine-translated texts are characterised by intricacy, a corresponding decline in the target texts' readability scores can be observed. As Figures 12 and 13 illustrate, the target text readability of T_3 was higher than that of T_2 for all participants and its mean. This is presumed to be due to the higher source text readability (2.84) and the higher LLM output readability (5.33) of T_3 compared to the source text (2.20) and MT output (2.74) of T_2. In other words, it can be concluded that both translation effort and target text readability are influenced by source text readability as well as MT/LLM output readability. Furthermore, the findings demonstrate that should the MT output prove to be inaccurate, the resultant post-edited text will likely exhibit a lower quality than that of a translation from scratch. In addition, the discussion on the correlation between the readability of source texts or MT/LLM outputs and different types of effort suggests that the more complicated the source and machine-translated texts, the more temporal, technical, and cognitive effort is required.

In summary, the discussion and the results of this experiment demonstrate that, as discussed in the literature review, less temporal, technical, and cognitive effort is required when working with MT and LLMs' output in a Japanese intralingual text simplification task compared to Human Text-Simplification. However, the degree of the effort reduction depends on the source text's readability and the MT/LLM output's readability. Consequently, the hypothesis that “MT and LLMs also make the process of intralingual translation into Easy Japanese more productive and less effortful than Human Text-Simplification in crisis settings, and improve the target text readability” has only been partially verified.

Finally, as mentioned in the theoretical sections, translation technologies and ICT play an important role in crisis settings, and the Japanese government has invested enormously in such technologies. However, as the results show, the grade of the productivity gain and the quality of the target texts depend on the MT/LLM systems and the capability of translators. For those reasons, it should be reasonable not only to invest in the technologies but also to secure and develop personnel who are available to engage in a crisis setting as citizen translators by offering workshops or training for Easy Japanese communication, which is indeed a critical point overlooked by the Japanese government.

Acknowledgements

This research was supported by “Alfasado” and its personnel in charge, who allowed me full access to the Tsutaeru-Web system. I would like to thank the participants who generously gave up their time, and Sarah Tobisch, for proofreading and developing the TranslationAnalyzer for automatic counting of pauses. I would also like to express my heartfelt gratitude to Univ.-Prof. Dragos Ioan Ciobanu, PhD and to all the professors who accompanied me during the semesters for their constructive comments and insightful feedback. Lastly, I am thankful for my parents who supported me throughout my studies.

References

Alfasado (2025a). 伝えるウェブ. https://tsutaeru.cloud/assets/tsutaeruweb_intr.pdf (accessed: 30.06.2025).

Alfasado (2025b). 「伝えるウェブ」が、導入200件を突破しました！. https://tsutaeru.cloud/news/achievement200.html (accessed: 03.10.2026).

All About People (2017). The 4R's of Emergency Management. https://www.allaboutpeople.co.nz/wp-content/uploads/2017/05/The-4Rs-of-Emergency-Management-1.pdf (accessed: 08.10.2025).

Alvarez, Sergi; Oliver, Antoni & Badia, Toni (2020). Quantitative Analysis of Post-Editing Effort Indicators for NMT. In: Martins, André, Moniz, Helena, Fumega, Sara, Martins, Bruno, Batista, Fernando, Coheur, Luisa, Parra, Carla, Trancoso, Isabel, Turchi, Marco, Bisazza, Arianna, Moorkens, Joss, Guerberof, Ana, Nurminen, Mary, Forcada, Mikel L. & Marg, Lena (Eds) *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. Lisboa: European Association for Machine Translation, 411–420. https://aclanthology.org/2020.eamt-1.44/.

Alvarez-Vidal, Sergi & Oliver, Antoni (2023). Assessing MT with measures of PE effort. *Ampersand* 11, 100125. DOI: https://doi.org/10.1016/j.amper.2023.100125.

Anschütz, Miriam; Deilen, Silvana & Lapshinova-Koltunski, Ekaterina (2025). Künstliche Intelligenz: Wie funktioniert KI für Einfache Sprache? In: Maaß, Christiane, Lapshinova-Koltunski, Ekaterina & Hörner, Julian (Eds) *Einfache Sprache mit KI-Tools*. Wiesbaden: Springer Fachmedien Wiesbaden, 57–71. DOI: https://doi.org/10.1007/978-3-658-47867-4_5.

Bureau of Citizens and Cultural Affairs (2020). 「やさしい日本語」に関する調査結果（概要）. https://www.seikatubunka.metro.tokyo.lg.jp/documents/d/seikatubunka/26_hearing_kekka (accessed: 17.06.2025).

Cadwell, Patrick (2016). A Place for Translation Technologies in Disaster Settings: The Case of the 2011 Great East Japan Earthquake. In: O'Hagan, Minako & Zhang, Qi (Eds) *Conflict and Communication: A Changing Asia in a Globalizing World - Language and Cultural Perspectives*. Hauppauge: Nova Science Publishers, Inc, 169–194.

Cadwell, Patrick (2023). Translation, interpreting, language, and foreignness in crisis communication policy. In: Declercq, Christophe, & Kerremans, Koen*The Routledge Handbook of Translation, Interpreting and Crisis*. 1st ed. London: Routledge, 17–30. DOI: https://doi.org/10.4324/9781003207580-3.

Cadwell, Patrick & O'Brien, Sharon (2016). Language, culture, and translation in disaster ICT: an ecosystemic model of understanding. *Perspectives* 24 (4), 557–575. DOI: https://doi.org/10.1080/0907676X.2016.1142588.

Cadwell, Patrick; O'Brien, Sharon & DeLuca, Eric (2019). More than tweets: A critical reflection on developing and testing crisis machine translation technology. *Translation Spaces* 8 (2), 300–333. DOI: https://doi.org/10.1075/ts.19018.cad.

capito (n.d.). Easy Language: What is it and why is it important? The most frequently asked questions about easy-to-understand language and barrier-free information. https://www.capito.eu/en/easy-language/ (accessed: 25.06.2025).

Carl, Michael (2012). Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research. In: Calzolari, Nicoletta, Choukri, Khalid, Declerck, Thierry, Doğan, Mehmet Uğur, Maegaard, Bente, Mariani, Joseph, Moreno, Asuncion, Odijk, Jan & Piperidis, Stelios (Eds) *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA), 4108–4112.

Center for Japanese-Language Testing (2012). Indication of the CEFR Level for Referance. https://www.jlpt.jp/e/about/cefr_reference.html (accessed: 25.06.2025).

Crisis Management Division, General Affairs Bureau Crisis Management Department (2025). 横浜市総務局危機管理室. https://x.com/yokohama_saigai/status/1950455242342224111 (accessed: 30.07.2025).

Cui, Ying; Liu, Xiao & Cheng, Yuqin (2023). A Comparative Study on the Effort of Human Translation and Post-Editing in Relation to Text Types: An Eye-Tracking and Key-Logging Experiment. *Sage Open* 13 (1), 1–15. DOI: https://doi.org/10.1177/21582440231155849.

Dai, Guangrong & Liu, Siqi (2024). Towards Predicting Post-editing Effort with Source Text Readability: An Investigation for English-Chinese Machine Translation. *The Journal of Specialised Translation* (41), 206–229. DOI: https://doi.org/10.26034/cm.jostrans.2024.4723.

Deilen, Silvana; Garrido, Sergio Hernández; Lapshinova-Koltunski, Ekaterina & Maaß, Christiane (2023). Using ChatGPT as a CAT tool in Easy Language translation. In: Štajner, Sanja, Saggio, Horacio, Shardlow, Matthew & Alva-Manchego, Fernando (Eds) *Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability*. Varna: INCOMA Ltd., Shoumen, Bulgaria, 1–10. DOI: https://doi.org/10.48550/arXiv.2308.11563.

Deilen, Silvana; Garrido, Sergio Hernández; Lapshinova-Koltunski, Ekaterina; Maaß, Christiane; Hörner, Julian; Theel, Vanessa & Ziemer, Sophie (2024). Towards AI-supported Health Communication in Plain Language: Evaluating Intralingual Machine Translation of Medical Texts. In: Demner-Fushman, Dina, Ananiadou, Sophia, Thompson, Paul & Ondov, Brian (Eds) *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*. Torino: ELRA and ICCL, 44–53.

Deilen, Silvana; Lapshinova-Koltunski, Ekaterina; Garrido, Sergio Hernandez; Maaß, Christiane; Horner, Julian; Theel, Vanessa & Ziemer, Sophie (2024). Evaluation of intralingual machine translation for health communication. In: *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*. Sheffield: European Association for Machine Translation (EAMT), 469–479.

Department of Internal Affairs Government Information Services (2023). Plain language. https://www.digital.govt.nz/standards-and-guidance/design-and-ux/content-design-guidance/writing-style/plain-language (accessed: 03.06.2026).

Disaster Management, Cabinet Office (2025). 地震災害. https://www.bousai.go.jp/kyoiku/hokenkyousai/jishin.html (accessed: 15.11.2025).

Elming, Jakob; Balling, Laura Winther & Carl, Michael (2014). Investigating User Behaviour in Post-editing and Translation using the CASMACAT Workbench. In: O'Brien, Sharon, Balling, Laura Winther, Carl, Michael, Simard, Michel & Specia, Lucia (Eds) *Post-editing of machine translation: processes and applications*. Newcastle upon Tyne: Cambridge Scholars Press, 147–169.

Escribe, Marie (2019). Human Evaluation of Neural Machine Translation: The Case of Deep Learning. In: *Proceedings of the Second Workshop Human-Informed Translation and Interpreting Technology associated with RANLP 2019*. Varna: Incoma Ltd., Shoumen, Bulgaria, 36–46. DOI: https://doi.org/10.26615/issn.2683-0078.2019_005.

Federici, Federico M. & Cadwell, Patrick (2018). Training citizen translators: Design and delivery of bespoke training on the fundamentals of translation for New Zealand Red Cross. *Translation Spaces* 7 (1), 20–43. DOI: https://doi.org/10.1075/ts.00002.fed.

Google (2025). Gemini. https://gemini.google.com/gem/1BoAYLrqqmGkCJ8RMQCZD0WJpi6YjRgvq?usp=sharing (accessed: 16.10.2025).

Gutermuth, Silke (2020). Leichte Sprache für alle? Eine zielgruppenorientierte Rezeptionsstudie zu Leichter und Einfacher Sprache. Berlin: Frank & Timme.

Hasebe, Yoichiro & Lee, Jae-Ho (2015). INTRODUCING A READABILITY EVALUATION SYSTEM FOR JAPANESE LANGUAGE EDUCATION. https://jreadability.net/file/hasebe-lee-2015-castelj.pdf (accessed: 13.10.2025).

Hasebe, Yoichiro & Lee, Jae-Ho (2025). Divergent strategies in text simplification: A comparative analysis of AI and human approaches in language processing. *Intercultural Pragmatics* 22 (2), 203–230. DOI: https://doi.org/10.1515/ip-2025-2002.

Immigration Services Agency (2024). 令和６年６月末現在における在留外国人数について. https://www.moj.go.jp/isa/publications/press/13_00047.html (accessed: 26.06.2025).

Immigration Services Agency of Japan (2020). 在留支援のためのやさしい日本語ガイドライン. https://www.bunka.go.jp/seisaku/kokugo_nihongo/kyoiku/pdf/92484001_01.pdf (accessed: 25.06.2025).

Iori Isao; Iwata Kazunari; Tsutsui Chie; Mori Atsushi & Matsuda Makiko (2010). 「やさしい日本語」を用いたユニバーサルコミュニケーション実現のための予備的考察. 一橋大学留学生センター. DOI: https://doi.org/10.15057/18797.

ISO (2015). ISO 17100:2015 Translation services — Requirements for translation services. Geneva: ISO. https://www.iso.org/standard/59149.html.

ISO (2017). ISO 18587:2017 Translation services — Post-editing of machine translation output — Requirements. Geneva: ISO. https://www.iso.org/standard/62970.html.

ISO (2023). ISO 24495-1:2023 Plain language Part 1: Governing principles and guidelines. Geneva. https://www.iso.org/standard/78907.html.

Jakobson, Roman (1959). On Linguistic Aspects of Translation. In: *On Translation*. Harvard University Press, 232–239. DOI: https://doi.org/10.4159/harvard.9780674731615.c18.

Jiang, Chunlan (2023). Investigation on the Application of Artificial Intelligence Large Language Model in Translation Tasks. In: Yacob, Shakila, Cicek, Berat, Rak, Joanna & Ali, Ghaffar (Eds) *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*. Paris: Atlantis Press SARL, Vol. 779. 1341–1351. DOI: https://doi.org/10.2991/978-2-38476-126-5_147.

Kimura, Goro Christoph (2023). 「やさしいことば」は誰のためか―「やさしい日本語」とドイツの「やさしいことば」（Leichte Sprache）の比較から („Leichte Sprache“ für wen? – Ein deutsch-japanischer Vergleich). *Bulletin of the Faculty of Foreign Studies, Sophia University* 58, 49–73.

Koglin, Arlene (2015). An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *Translation & Interpreting* 7 (1), 126–141. DOI: https://doi.org/ti.106201.2015.a06.

Kohari, Junya; Okuhara, Tsuyoshi & Shirabe, Ritsuko (2025). Comprehensiveness, Quality, and Readability of Non-Invasive Prenatal Testing Information on Japanese Medical Institution Websites. *SSRN* . DOI: https://doi.org/10.2139/ssrn.5389009.

Koponen, Maarit (2012). Comparing human perceptions of post-editing effort with post-editing operations. In: Callison-Burch, Chris, Koehn, Philipp, Monz, Christof, Post, Matt, Soricut, Radu & Specia, Lucia (Eds) *Proceedings of the Seventh Workshop on Statistical Machine Translation*. Montréal: Association for Computational Linguistics, 181–190.

Koponen, Maarit (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.

Koponen, Maarit; Salmi, Leena & Nikulin, Markku (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation* 33 (1–2), 61–90. DOI: https://doi.org/10.1007/s10590-019-09228-7.

Krings, Hans P. (2001). Repairing texts: empirical investigations of machine translation post-editing processes. Kent, Ohio: Kent State University.

Kröger, Janina; Deilen, Silvana; Hörner, Julian; Lapshinova-Koltunski, Ekaterina & Maaß, Christiane (2025). Entwicklung eines redaktionellen Workflows für die KI-gestützte Übersetzung von Gesundheitsinformationen in Einfache Sprache. *Trans-Kom* 18 (1), 380–404.

Kumano Tadashi; Goto Isao & Tanaka Hideki (2015). Automatic News Rewriting into Easy Japanese by Statistical Machine Translation. *ITE Annual Convention*. DOI: https://doi.org/10.11485/iteac.2015.0_32D-2.

Lacruz, Isabel (2017). Cognitive Effort in Translation, Editing, and Post-editing. In: Schwieter, John W. & Ferreira, Aline (Eds) *The Handbook of Translation and Cognition*. 1st ed. Wiley, 386–401. DOI: https://doi.org/10.1002/9781119241485.ch21.

Lacruz, Isabel; Carl, Michael & Yamada, Masaru (2018). Literality and Cognitive Effort: Japanese and Spanish. In: Calzolari, Nicoletta, Choukri, Khalid, Cieri, Christopher, Declerck, Thierry, Goggi, Sara, Hasida, Koiti, Isahara, Hitoshi, Maegaard, Bente, Mariani, Joseph, Mazo, Hélène, Moreno, Asuncion, Odijk, Jan, Piperidis, Stelios & Tokunaga, Takenobu (Eds) *The LREC 2018 Proceedings : Eleventh International Conference on Language Resources and Evaluation*. Miyazaki: European Language Resources Association, 3818–3821. http://www.lrec-conf.org/proceedings/lrec2018/pdf/811.pdf.

Lacruz, Isabel; Carl, Michael; Yamada, Masaru & Aizawa, Akiko (2016). Pause Metrics and Machine Translation Utility. Paper presented at ‘The 22nd Annual Meeting of the Association for Natural Language Processing. NLP 2016’, Sendai: The Association for Natural Language Processing, 1213–1216.

Lacruz, Isabel & Shreve, Gregory M. (2014). Pauses and cognitive effort in post-editing. In: O’Brien, Sharon, Carl, Michael, Simard, Michel & Specia, Lucia (Eds) *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, 246–272.

Lacruz, Isabel; Shreve, Gregory M. & Angelone, Erik (2012). Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study. In: O'Brien, Sharon, Simard, Michel & Specia, Lucia (Eds) *Workshop on Post-Editing Technology and Practice*. San Diego, California: Association for Machine Translation in the Americas.

Lee, Jaeho & Hasebe, Yoichiro (2025). Simplifying Japanese News Using ChatGPT. *Mathematical Linguistics* 34 (8), 563–573. DOI: https://doi.org/10.24701/mathling.34.8_563.

Lee, Jaeho; Sunakawa, Yuriko; Hori, Keiko & Hasebe, Yoichiro (2013). jReadability. https://jreadability.net/ (accessed: 15.07.2025).

Lewis, William (2010). Haitian Creole: How to Build and Ship an MT Engine from Scratch in 4 days, 17 hours, & 30 minutes. In: Yvon, François & Hansen, Viggo (Eds) *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*. Saint Raphaël: European Association for Machine Translation.

Lindholm, Camilla & Vanhatalo, Ulla (ed.) (2021). Handbook of Easy Languages in Europe. Berlin: Frank & Timme. DOI: https://doi.org/10.26530/20.500.12657/52628.

Liu, Xiao & Cui, Ying (2024). The impact of text type on Chinese-English translation effort: an investigation with reference to translation entropy. *Humanities and Social Sciences Communications* 11 (1), 1–9. DOI: https://doi.org/10.1057/s41599-024-03187-y.

Long, Daniel (2012). 緊急時における外国人住民のコミュニケーション問題 : 東日本大震災と阪神大震災から学べること. *The Journal of Japan Academy of Health Sciences* 14 (4), 183–190. DOI: https://doi.org/10.24531/jhsaiih.14.4_183.

Lörscher, Wolfgang (1991). Translation performance, translation process, and translation strategies: a psycholinguistic investigation. Tübingen: Narr, 307.

Maaß, Christiane (2020). Easy Language – Plain Language – Easy Language Plus : Balancing Comprehensibility and Acceptability. Berlin: Frank & Timme. DOI: https://doi.org/10.26530/20.500.12657/42089.

Maaß, Christiane & Schwengber, Laura Marie (2022). Easy Language and Plain Language in Germany. *Rivista internazionale di tecnica della traduzione n.24 - 2022* 43–61. DOI: https://doi.org/10.13137/2421-6763/34290.

Macken, Lieve; Prou, Daniel & Tezcan, Arda (2020). Quantifying the Effect of Machine Translation in a High-Quality Human Translation Production Process. *Informatics* 7 (2), 12. DOI: https://doi.org/10.3390/informatics7020012.

Makhmutova, Liliya; Salton, Giancarlo; Perez-Tellez, Fernando & Ross, Robert (2024). Automated Medical Text Simplification for Enhanced Patient Access. In: *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies*. Rome, Italy: SCITEPRESS - Science and Technology Publications, 208–218. DOI: https://doi.org/10.5220/0012466100003657.

Martín, Ricardo Muñoz; Sun, Sanjun; Du, Zhiqiang & Puerini, Sara (2025). Keylogging. In: Rojo López, Ana María & Muñoz Martín, Ricardo (Eds) *Research Methods in Cognitive Translation and Interpreting Studies*. Amsterdam: John Benjamins, Vol. 10. 157–182. DOI: https://doi.org/10.1075/rmal.10.07mar.

Masden, Kirk (2022). The need for links to original texts on Japanese government websites : toward the inclusion of Japanese in linguisitc services for foreigners. *KGU Journal of Language and Literature* 29 (2), 187–219.

Matsuda Yoko; Maeda Rikako & Sato Kazuyuki (2000). A study of how to feed disaster information to non-native Japanese speakers : strategies and effrctiveness. *Japanese Linguistics* 145–159. DOI: https://doi.org/10.15084/00002034.

Mie Prefecture (2025). やさしい日本語 ガイドライン. https://www.pref.mie.lg.jp/common/content/001173183.pdf (accessed: 04.07.2025).

Ministry of Internal Affairs and Communications (2024). 統計からみた我が国の高齢者 －「敬老の日」にちなんで－. https://www.stat.go.jp/data/topics/pdf/topics142.pdf (accessed: 16.07.2025).

Moorkens, Joss; O'Brien, Sharon; Da Silva, Igor A. L.; De Lima Fonseca, Norma B. & Alves, Fabio (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation* 29 (3–4), 267–284. DOI: https://doi.org/10.1007/s10590-015-9175-2.

Morisaki, Kentaro (2018). 伝えるウェブ. Osaka: Alfasado Inc. https://tsutaeru.cloud/translation/try.html.

NASA's Ames Research Center (2022). NASA Task Load Index (TLX). https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf (accessed: 16.07.2025).

National Institute for Japanese Language and Linguistics (2009). 「生活のための日本語：全国調査」結果報告＜速報版＞. https://www2.ninjal.ac.jp/past-projects/nihongo-syllabus/research/pdf/seika_sokuhou.pdf (accessed: 26.06.2025).

Netzwerk Leichte Sprache e.V. (2022). Regeln für Leichte Sprache. https://www.netzwerk-leichte-sprache.de/fileadmin/content/documents/regeln/Regelwerk_NLS_web.pdf (accessed: 03.12.2026).

O'Brien, Sharon (2006). Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output. *Across Languages and Cultures* 7 (1), 1–21. DOI: https://doi.org/10.1556/Acr.7.2006.1.1.

O'Brien, Sharon (2007). An empirical investigation of temporal and technical post-editing effort. *Translation and Interpreting Studies* 2 (1), 83–136. DOI: https://doi.org/10.1075/tis.2.1.03ob.

O'Brien, Sharon (2011). Towards predicting post-editing productivity. *Machine Translation* 25 (3), 197–215. DOI: https://doi.org/10.1007/s10590-011-9096-7.

O'Brien, Sharon (2016). Training Translators for Crisis Communication: Translators Without Borders as an Example. In: Federici, Federico M. (Ed.) *Mediating Emergencies and Conflicts*. London: Palgrave Macmillan UK, 85–111. DOI: https://doi.org/10.1057/978-1-137-55351-5_4.

O'Brien, Sharon (2019). Translation technology and disaster management. In: O'Hagan, Minako (Ed.) *The Routledge Handbook of Translation and Technology*. 1st ed. London: Routledge, 304–318. DOI: https://doi.org/10.4324/9781315311258-18.

O'Brien, Sharon & Federici, Federico Marco (2019). Crisis translation: considering language needs in multilingual disaster settings. *Disaster Prevention and Management: An International Journal* 29 (2), 129–143. DOI: https://doi.org/10.1108/DPM-11-2018-0373.

Oiso Town (2025a). 津波警報が発令されています. https://www.town.oiso.kanagawa.jp/kinkyu/23325.html (accessed: 30.07.2025).

Oiso Town (2025b). 竜巻注意情報. https://oiso-city.site.ktaiwork.jp/?p=3161 (accessed: 16.10.2025).

Okuhara, Tsuyoshi; Ishikawa, Hirono; Ueno, Haruka; Okada, Hiroko; Kato, Mio & Kiuchi, Takahiro (2020). Influence of high versus low readability level of written health information on self-efficacy: A randomized controlled study of the processing fluency effect. *Health Psychology Open* 7 (1), 1–9. DOI: https://doi.org/10.1177/2055102920905627.

O'Mathúna, Dónal P.; Escartín, Carla Parra; Roche, Proinsias & Marlowe, Jay (2020). Engaging citizen translators in disasters: Virtue ethics in response to ethical challenges. *Translation and Interpreting Studies* 15 (1), 57–79. DOI: https://doi.org/10.1075/tis.20003.oma.

Rogl, Regina (2017). Language-related disaster relief in Haiti: Volunteer translator networks and language technologies in disaster aid. In: Antonini, Rachele, Cirillo, Letizia, Rossato, Linda & Torresi, Ira (Eds) *Benjamins Translation Library*. Amsterdam: John Benjamins Publishing Company, Vol. 129. 231–255. DOI: https://doi.org/10.1075/btl.129.12rog.

Sakai Kazumi (2021). Über leichtes Japanisch und maschinelle Übersetzung – Zur Förderung der Sprachbewusstheit –. *Neue Beiträge zur Germanistik* 162, 147–160. DOI: https://doi.org/10.11282/jgg.162.0_147.

Sánchez-Gijón, Pilar & Kenny, Dorothy (2022). Selecting and preparing texts for machine translation: Pre-editing and writing for a global audience. In: Kenny, Dorothy (Ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, 81–103. DOI: https://doi.org/10.5281/ZENODO.6759980.

Sanchez-Torron, Marina & Koehn, Philipp (2016). Machine Translation Quality and Post-Editor Productivity. In: *Conferences of the Association for Machine Translation in the Americas: MT Researchers' Track*. Austin: The Association for Machine Translation in the Americas, 16–26. https://aclanthology.org/2016.amta-researchers.2/.

Sato, Kazuyuki (2013). 〈増補版〉「やさしい日本語」作成のためのガイドライン. https://www.fdma.go.jp/singi_kento/kento/items/kento207_20_sankou5-6.pdf (accessed: 04.11.2025).

Sato Kazuyuki (2024). 大規模災害時の「やさしい日本語」表現. *Acoustical Science and Technology* 80 (3), 147–152. DOI: https://doi.org/10.20697/jasj.80.3_147.

Sato, Kumi; Okamoto, Kohei; Takahashi, Kimiaki; Tanaka, Shozo; Yamaoka, Koshun & Miyao, Masaru (2004). Hardships suffered by foreign people from the Hanshin-Awaji Earthquake and multilingual disaster information. *Bulletin of Social Medicine* 22, 21–28.

Sendai City (2021). 情報発信が活発に行われる反面 避難所では思わぬ事態も。. https://sendai-resilience.jp/shinsai10/interview/20201012_entry03_02.html (accessed: 22.10.2025).

Shizuoka Prefecture (2025). 令和 7 年 7 月 30 日　カムチャツカ半島付近の地震による被害状況. https://www.pref.shizuoka.jp/bosaikinkyu/saigai/1076694/index.html (accessed: 16.10.2025).

Snover, Matthew; Dorr, Bonnie; Schwartz, Rich; Micciulla, Linnea & Makhoul, John (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. Massachusetts: Association for Machine Translation in the Americas, 223–231.

Stasimioti, Maria & Sosoni, Vilelmini (2020a). MT output and post-editing effort: Insights from a comparative analysis of SMT and NMT output and implications for training. In: Besznyák, Rita, Fischer, Márta, &

Szabó, Csilla (Eds) *Fit-For-Market Translator and Interpreter Training in a Digital Age*. Wilmington: Vernon Press, 151–175.

Stasimioti, Maria & Sosoni, Vilelmini (2020b). Translation vs Post-editing of NMT Output: Insights from the English-Greek language pair. In: Ortega, John E., Federico, Marcello, Orasan, Constantin & Popovic, Maja (Eds) *Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation*. Virtual: Association for Machine Translation in the Americas, 109–124.

Sugatani, Yasuyuki (2020). 「ドイツ語圏のやさしいことば」. https://www.arsvi.com/2020/20200919sy.htm (accessed: 20.10.2025).

Sun, Sanjun (2012). Measuring difficulty in English-Chinese translation: Towards a general model of translation difficulty. PhD thesis, Kent State University.

Takeda, Yuko; Iwata, Kazunari & Nii, Midori (2021). 医療現場の外国人対応英語だけじゃない「やさしい日本語」. Tokyo: 南山堂.

Tatsumi, Midori (2009). Correlation between Automatic Evaluation Metric Scores, Post-Editing Speed, and Some Other Factors. In: *Proceedings of Machine Translation Summit XII: Posters*. Ottawa, 1–8.

Tatsumi, Midori & Roturier, Johann (2010). Source Text Characteristics and Technical and Temporal. In: *Proceedings of the Second Joint EM+/CNGL Workshop: Bringing MT to the User: Research on Integrating MT in the Translation Industry*. Denver: Association for Machine Translation in the Americas, 43–52. https://aclanthology.org/2010.jec-1.6/.

Tobisch, Sarah (2025). TranslationAnalyzer. https://github.com/krankatze/TranslationAnalyzer (accessed: 19.10.2025).

Tokyo International Exchange Committee (2018). 東京都在住外国人向け 情報伝達に関するヒアリング調査報告書 (Hearing Survey on Information Communication for Foreign Residents in Tokyo). https://tabunka.tokyo-tsunagari.or.jp/info/files/a70d5ac7db12bd5c538a3b38f2a01613c262657e.pdf (accessed: 17.06.2025).

Tokyo Metropolitan Foundation 'TSUNAGARI' (2025). 【世田谷区】やさしい日本語ワークショップ実践編へ（11 月 22 日開催）. https://tabunka.tokyo-tsunagari.or.jp/event/2025/10/03_10.html (accessed: 23.10.2025).

UN CRPD (2008). Article 9 – Accessibility. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-9-%20accessibility.html (accessed: 19.10.2025).

Vieira, Lucas Nunes (2017). Cognitive effort and different task foci in post-editing of machine translation: A think-aloud study. *Across Languages and Cultures* 18 (1), 79–105. DOI: https://doi.org/10.1556/084.2017.18.1.4.

Wang, Yu & Jalalian Daghigh, Ali (2023). Cognitive effort in human translation and machine translation post-editing processes: A holistic and phased view. *FORUM* 21 (1), 139–162. DOI: https://doi.org/10.1075/forum.22009.wan.

Wang, Zhaohui (2019). Application of Screen Recording in Multimodal Study of Translation Process. In: *Proceedings of the 2019 3rd International Conference on Economic Development and Education*

Management (ICEDEM 2019). Dalian: Atlantis, 395–398. DOI: https://doi.org/10.2991/icedem-19.2019.93.

Weathernews Inc. (2023). 能登地震　揺れの強かった周辺で明日から雨　寒暖差も注意. https://weathernews.jp/s/topics/202305/050215/?fbclid=IwY2xjawNbiz9leHRuA2FlbQIxMABicmlkETFKYVVubnZEeGRCbDluY2syAR65yWr1th9MTtDd-joH9N5dmy7-7TeMsOZngcRSxHqTHEXtpYaDnGwQb7VQ5w_aem_iF34w7h1kn4WLo7KW6aG6A (accessed: 16.10.2025).

Whyatt, Bogusława; Kajzer-Wietrzny, Marta & Stachowiak, Katarzyna (2017). Intralingual and interlingual translation: Designing a comparative study of decision-making processes in paraphrase and translation. In: Jakobsen, Arnt Lykke & Mesa-Lao, Bartolomé (Eds) *Benjamins Translation Library*. Amsterdam: John Benjamins Publishing Company, Vol. 133. 135–158. DOI: https://doi.org/10.1075/btl.133.05why.

Yılmaz, Yücel; Öztürk, Muhammed Maruf & Nejkovic, Valentina (2025). A Comprehensive Review of LLMs: Architecture, Performance, and Applications. In: *2025 12th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)*. Cacak, Serbia: IEEE, 1–7. DOI: https://doi.org/10.1109/IcETRAN66854.2025.11114289.

Yokahama City (2017). 「やさしい日本語」で伝える 分かりやすく 伝わりやすい日本語を目指して. https://www.moj.go.jp/isa/content/930005564.pdf (accessed: 25.06.2025).

List of AI-supported tools

AI-based permitted material	Use	Affected sections of the thesis	Notes
DeepL Write	Linguistic revision	Parts of the texts in some chapters	The suggested corrections have been accepted.
Grammarly	Linguistic revision	Parts of the texts in some chapters	The suggested corrections have been accepted.
OpenAI ChatGPT	Creation of illustrations	Figures in Chapters 4 and 5	The AI's suggestions of Python code and figures were significantly revised.

Appendix

Briefing Letter

修士論文実験へのご協力のお願い（ブリーフィングレター）

2025 年 10 月 16 日

Belvederegasse 36-38/2/2
1040 Wien, Austria

研究参加のお願い

このたび、ウィーン大学修士課程の研究の一環として、下記の要領で実験を実施いたします。本実験の目的およびシナリオは以下の通りです。

【あなたは、ある市役所の多文化共生課の担当職員として勤務しています。緊急速報として SNS やホームページに掲載される災害に関する文章を「やさしい日本語」に翻訳することになりました。添付されたガイドラインは、翻訳作業時の参考資料として活用してください。以下は緊急速報の例文です。
現在、津波注意報が発表されているため、海の公園の海水浴場は遊泳禁止となっています。ご利用の皆様にはご不便をおかけしますが、ご理解のほどよろしくお願いいたします。海の中や海岸は危険です。注意報の解除まで、海に入ったり、海岸に近づかないようにしてください。】

本実験へのご参加は任意です。ご協力いただける方は、以下の内容をご確認ください。

実験の概要

・実施期間：2025 年 10 月 16 日～2025 年 12 月 31 日（予定）

・所要時間：90 分程度

・実施場所：Belvederegasse 36-38/2/2

・参加対象：日本語を第一言語とされる方

参加に関する注意事項

・実験参加は完全に任意です。いつでも中断または辞退することができます。

・収集したデータは匿名化され、研究目的以外では一切使用されません。

・本研究はウィーン大学の研究倫理指針に基づいて実施されます。

・参加にあたり、事前に同意書をご確認いただきます。

連絡先

ご質問やご参加希望の方は、以下の連絡先までお気軽にご連絡ください。

三宮 龍之介（Ryunosuke Sannomiya）

E-mail：a12204442@unet.univie.ac.at

住所：Belvederegasse 36-38/2/2, 1040 Wien, Austria

ご検討のほど、よろしくお願いいたします。

敬具

三宮 龍之介
ウィーン大学 翻訳学科 修士課程

Briefing Letter (English MT post-edited version of the Japanese text translated with DeepL.com) (accessed: 06.03.2026)

Request for Cooperation in Master's Thesis Experiment (Briefing Letter)

16 October 2025

Belvederegasse 36-38/2/2

1040 Vienna, Austria

Request for Research Participation

As part of a master's degree research project at the University of Vienna, we shall be conducting an experiment as outlined below. The purpose and scenario of this experiment are as follows.

[You are employed as a staff member in the Multicultural Coexistence Department of a city council. You have been tasked with translating disaster-related texts published on social media and websites as emergency alerts into 'Easy Japanese'. Please use the attached guidelines as reference material during the translation process. Below is an example of an emergency alert.
Due to a tsunami advisory currently in effect, swimming is prohibited at the seaside park's bathing area. We apologise for any inconvenience caused and appreciate your understanding. The sea and coastline are dangerous. Please refrain from entering the sea or approaching the coast until the advisory is lifted.]

Participation in this experiment is entirely voluntary. If you are willing to assist, please review the following details.

Experiment Overview

・Duration: 16 October 2025 to 31 December 2025 (planned)

・Estimated time: Approximately 90 minutes

・Location: Belvederegasse 36-38/2/2

・Eligible Participants: Individuals whose primary language is Japanese

Important Notes Regarding Participation

・Participation in this experiment is entirely voluntary. You may discontinue or withdraw at any time.

・Collected data will be anonymised and used solely for research purposes.

・This study is conducted in accordance with the research ethics guidelines of the University of Vienna.

・Prior to participation, you will be required to review and sign a consent form.

Contact Information

Should you have any questions or wish to participate, please feel free to contact us at the following address.

Ryunosuke Sannomiya
E-mail: a12204442@unet.univie.ac.at
Address: Belvederegasse 36-38/2/2, 1040 Vienna, Austria

We would be grateful for your consideration.

Yours sincerely,

Ryunosuke Sannomiya
Master's Programme, Centre of Translation Studies, University of Vienna

Template of the Declaration of Consent

Declaration of Consent to participate in an experiment

Declaration of Consent

Name: _____

I hereby agree to participate in the experiment as part of the University of Vienna's master's thesis research project "Productivity and Effort in Intralingual Translation: A Comparative Process-oriented Study of Post-Editing and Human Translation into Easy Japanese in Crisis Settings". I have been adequately informed about the purpose and procedure. I agree to the collection of data as part of the experiment. I understand that this data will be recorded anonymously and used exclusively for scientific purposes. I may revoke my consent at any time.

_____ _____

(Place, date) (Signature)

Source Texts

Text_1

神奈川県東部は、竜巻などの激しい突風が発生しやすい気象状況になっています。 空の様子に注意してください。 雷や急な風の変化など積乱雲が近づく兆しがある場合には、頑丈な建物内に移動するなど、安全確保に努めてください。 落雷、ひょう、急な強い雨にも注意してください。 Source: https://oiso-city.site.ktaiwork.jp/?p=3161 (accessed: 16.10.2025)	**English machine-translated version of the Japanese text (Translated with DeepL.com)** (accessed: 06.03.2026) Eastern Kanagawa Prefecture is experiencing weather conditions conducive to severe gusts such as tornadoes. Please pay close attention to the sky. Should signs of approaching cumulonimbus clouds appear, such as thunder or sudden wind changes, take steps to ensure your safety, such as moving to a sturdy building. Also, beware of lightning strikes, hail, and sudden heavy rain.

Text_2

令和 7 年 7 月 30 日 8 時 25 分頃カムチャッカ半島付近で発生した地震により、9 時 40 分、本県沿岸に津波警報が発表されました。 海岸及び河口付近の方は、直ちに沿岸や河口付近から離れ、高台などの安全な場所に避難してください。	**English machine-translated version of the Japanese text (Translated with DeepL.com)** (accessed: 06.03.2026) At approximately 8:25 am on 30 July 2025, an earthquake occurred near the Kamchatka Peninsula. Consequently, a tsunami warning was issued for the coast of this prefecture at 9:40 am. Those near the coast or river mouths should immediately move away from the shoreline

海岸及び河口付近には絶対に近づかないでください。 Source: https://www.pref.shizuoka.jp/bosaikinkyu/saigai/1076694/index.html (accessed: 16.10.2025)	and river mouth areas and evacuate to higher ground or other safe locations. Do not approach the coast or river mouth areas under any circumstances.

Text_3

石川県能登地方・珠洲市で震度 6 強を観測する地震が発生しました。 連休最後の土日となる明日 6 日(土)以降は本降りの雨が予想されています。強い揺れで地盤が緩んでいる可能性があるため、土砂災害等に注意してください。 また、寒暖差も大きくなる可能性があるため体調管理にも注意してください。 Source: https://weathernews.jp/s/topics/202305/050215/?fbclid=IwY2xjawNbiz9leHRuA2FlbQIxMABicmlkETFKYVVubnZEeGRCbDluY2syAR65yWr1th9MTtDd-joH9N5dmy7-7TeMsOZngcRSxHqTHEXtpYaDnGwQb7VQ5w_aem_iF34w7h1kn4WLo7KW6aG6A (accessed: 16.10.2025)	**English machine-translated version of the Japanese text (Translated with DeepL.com)** (accessed: 06.03.2026) An earthquake registering a seismic intensity of 6 upper has been observed in Suzu City, Noto District, Ishikawa Prefecture. Heavy rain is forecast from tomorrow, Saturday the 6th, which marks the final weekend of the consecutive holidays. As the strong shaking may have loosened the ground, please be alert for landslides and other geological hazards. Additionally, be mindful of your health as significant temperature fluctuations are possible.

MT Outputs

Text_2 (MT)	Text_3 (LLM)
2025 年（ねん）7 月（がつ）30 日（にち）8 時（じ）25 分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の 近（ち）かくで 出（で）た 地震（じしん）に より、9 時（じ）40 分（ふん）、本県（ほんけん）の 海（うみ）の 近（ちかく）に 津波（つなみ） 警報（けいほう）（「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」という 強（つよい） 注意（ちゅうい）） が発表（はっぴょう）されました。 海岸（かいがん）と 河口（かこう）（川（がわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ））の 近（ちか）くの ひとは、急（いそ）いで 海（うみ）の 近（ちか）くや 河口（かこう）の 近（ちか）くから はなれます。高（たか）い所（ところ）などの 安全（あんぜん）な 場所（ばしょ）に 逃（に）げて ください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きい 地震（じしん）が ありました。 あした 6 日（にち）（土曜日（どようび））から、 雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）で 地面（じめん）が 弱（よわ）く なっていますから、土砂災害（どしゃさいがい）＜山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくること＞に 気（き）を つけてください。 また、 暖（あたた）かく なったり 寒（さむ）く なったり するかもしれません。 病気（びょうき）に ならないように 気（き）を つけてください。

海岸（かいがん）と　河口（かこう）（川（がわ）が、海（うみ）や　湖（みずうみ）に　つながる　場所（ばしょ））の　近（ちか）くには　絶対（ぜったい）に　近（ちか）づかないで　ください。	

Target Texts

	T_1 (HT)	T_2 (MT)	T_3 (LLM)
P_1	神奈川県の東側は、竜巻による激しい風が起こりやすい状況です。 空をよく見てください。 雷が鳴ったり、風が急に強くなったりしたら、積乱雲が近づいているサインです。 安全そうな建物の中に入って、危険な場所から離れてください。 雷、ひょう、急な強い雨にも気をつけてください。	2025年（ねん）7月（がつ）30日（にち）8時（じ）25分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の 近（ち）かくで 起（お）きた 地震（じしん）に より、9時（じ）40分（ぷん）、本県（ほんけん）の 海（うみ）の 近（ちかく）に 津波（つなみ） 警報（けいほう）（「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」という 強（つよい） 注意（ちゅうい）） が発表（はっぴょう）されました。 海岸（かいがん）と 河口（かこう）（川（かわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ））の 近（ちか）くの ひとは、すぐに海（うみ）の 近（ちか）くや 河口（かこう）の 近（ちか）くから はなれてください。高（たか）い所（ところ）などの 安全（あんぜん）な 場所（ばしょ）に 逃（に）げて ください。 海岸（かいがん）と 河口（かこう）（川（かわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ））の 近（ちか）くには 絶対（ぜったい）に 近（ちか）づかないで ください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きな 地震（じしん）が ありました。 あした 6日（か）（土曜日（どようび））から、 雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）で 地面（じめん）が 弱（よわ）く なっていますから、 土砂災害（どしゃさいがい）＜山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくること＞に 気（き）を つけてください。 また、 暖（あたた）かく なったり 寒（さむ）く なったり するかもしれません。 病気（びょうき）に ならないように 気（き）を つけてください。
P_2	神奈川県東の地域は、トルネードなどの強い風が起きやすい状態です。 空の感じに気をつけてください。 雷や急な風の変化など巨大に発達した雲が近づきそうな感じがする場合には、丈夫なビルの中に移動するなど、安全な場所にいられるように気をつけてください。 雷が落ちてきたり、氷が降ってきたり、突然の強い風にも気をつけてください。	2025年（ねん）7月（がつ）30日（にち）8時（じ）25分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の 近（ち）かくで 起きた 地震（じしん）に より、9時（じ）40分（ふん）、本県（ほんけん）の 海（うみ）の 近（ちかく）で「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」という 注意（ちゅうい） が発表（はっぴょう）されました。 海の近くと、 川（がわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ）の 近（ちか）くの ひとは、急（いそ）いでそこから はなれてください。高（たか）い所（ところ）などの 安全（あんぜん）な 場所（ばしょ）に 逃（に）げて ください。 海の近くと 川（がわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ）の 近（ちか）くには 絶対（ぜったい）に 近（ちか）づかないで ください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きい 地震（じしん）が ありました。 あした 6日（にち）（土曜日（どようび））から、 雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）で 地面（じめん）が 弱（よわ）く なっていますから、 山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくることに 気（き）を つけてください。 また、急に 暖（あたた）かく なったり 寒（さむ）く なったり するかもしれません。 病気（びょうき）に ならないように 気（き）を つけてください。

P_3	神奈川県（かながわけん） の 東（ひがし）は、つよい風（かぜ）が ふきやすくなっています。 天気（てんき）に きをつけてください。 かみなり や 風（かぜ）が 急（きゅう）に つよくなると、つよい雨（あめ） が ふるかもしれません。 そのときは、おおきな建物（たてもの） の なかにはいってください。 じぶん の いのち は じぶん で まもりましょう。 かみなり、 ひょう（そら から こおり が おちてくる こと）、つよい雨（あめ） にも きをつけてください。	2025年（ねん）7月（がつ）30日（にち）8時（じ）25分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の ちかくで地震がおきました。 9時（じ）40分（ふん）、海（うみ）の 近（ちかく）に 津波（つなみ） 警報（けいほう）（「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」というおしらせ） がだされました。 海（うみ）と 河口（かこう）（川（かわ）が、海（うみ）や 湖（みずうみ）に つながるところ）の 近（ちか）くの ひとは、急（いそ）いで そこから はなれてください。高（たか）いところなどの 安全（あんぜん）な ところに 逃（に）げて ください。 海（うみ）と 河口（かこう）には 近（ちか）づかないで ください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きい地震（じしん）が ありました。 あした 6日（にち）（土曜日（どようび））から、雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）＜地面（じめん）が おおきく ゆれること＞で 地面（じめん）が 弱（よわ）く なっています。 土砂災害（どしゃさいがい）＜山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくること＞に 気（き）を つけてください。 また、 あたたかく なったり さむく なったり するかもしれません。 病気（びょうき）に ならないように 気（き）を つけてください。
P_4	神奈川県（かながわけん）の東（ひがし）では、竜巻（たつまき）や とても強（つよ）い風（かぜ）が起（お）きるかもしれません。 天気（てんき）が変（か）わること に 気（き）をつけてください。 雷（かみなり） や 強（つよ）い風（かぜ）が 起（お）きそうなときは、強（つよ）い建物（たてもの）の中（なか）に 入（はい）ってください。 雷（かみなり）、ひょう「氷（こおり）が 空（そら）から降（ふ）ってくること」、いきなりの 強（つよ）い雨（あめ）にも 気（き）をつけてください。	2025年（ねん）7月（がつ）30日（にち）8時（じ）25分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の 近（ちか）くで おきた 地震（じしん）に より、9時（じ）40分（ふん）、ここの県（けん）の 海（うみ）の 近（ちか）くに 津波（つなみ） 警報（けいほう）（「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」という 強（つよ）い 注意（ちゅうい）） が発表（はっぴょう）されました。 海（うみ）の近（ちか）くと 川（かわ）の近（ちか）く、川（かわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ）の 近（ちか）くの ひとは、急（いそ）いで 海（うみ）の 近（ちか）くや 河口（かこう）の 近（ちか）くから はなれてください。高（たか）い所（ところ）などの 安全（あんぜん）な 場所（ばしょ）に 逃（に）げて ください。 海（うみ）の近（ちか）くと 川（かわ）の近（ちか）く、川（かわ）が、海（うみ）や 湖（みずうみ）に つながる 場所（ばしょ）の 近（ちか）くには 絶対（ぜったい）に 近（ちか）づかないで ください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きい地震（じしん）が ありました。 あした 6日（にち）（土曜日（どようび））から、雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）で 地面（じめん）が 弱（よわ）く なっていますから、 土砂災害（どしゃさいがい）＜山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくること＞に 気（き）を つけてください。 また、気温（きおん）は、 暖（あたた）かく なったり 寒（さむ）く なったり するかもしれません。 病気（びょうき）に ならないように 気（き）を つけてください。

P_5	神奈川県（かながわけん）の東（ひがし）で、とても強（つよ）い風（かぜ）や雨（あめ）があるかもしれません。 天気（てんき）に気（き）をつけてください。 ときどき空（そら）を見ましょう。 大きい雲（くも）があるときや、風（かぜ）が強（つよ）くなりそうなときは、大きくて、安全（あんぜん）な建物（たてもの）の中に入（はい）ってください。じぶんを守（まも）りましょう。 雷（かみなり）や、雪（ゆき）のかたまり、強（つよ）い雨（あめ）にも、注意（ちゅうい）してください。	2025年（ねん）7月（がつ）30日（にち）8時（じ）25分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の 近（ち）かくで 地震（じしん）がありました。そして、9時（じ）40分（ふん）、この県（けん）の 海（うみ）の 近（ちかく）に 津波（つなみ） 警報（けいほう）（「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」という お知（せ）らせ）がありました。 川（がわ）や、海（うみ）や 湖（みずうみ）の 近（ちか）くの ひとは、急（いそ）いで はなれてください。 高（たか）い所（ところ）などの 安全（あんぜん）な 場所（ばしょ）に にげてください。 川（がわ）や、海（うみ）や 湖（みずうみ）の 近（ちか）くには 、絶対（ぜったい）に 行かないでください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きい 地震（じしん）が ありました。 あした 6日（にち）（土曜日（どようび））から、雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）で 地面（じめん）が 弱（よわ）くなっていますから、 土砂災害（どしゃさいがい）＜山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくること＞に 気（き）を つけてください。 また、 暖（あたた）かく なったり 寒（さむ）くなったり するかもしれません。 病気（びょうき）や風邪（かぜ）に ならないように 気（き）を つけてください。
P_6	神奈川県（かながわけん）の東側（ひがしがわ）では たつまき（つよいかぜが うずのようになる）などの つよい風（かぜ）がおこりやすいです。 空（そら）の ようすに きをつけて ください。 雷（かみなり）や きゅうに 風（かぜ）が つよくなったり しそうな ときには しっかりした 建物（たてもの）の なかに にげて、安全（あんぜん）になるまで そこに いてください。 落雷（らくらい＝かみなりがおちること）、ひょう（こおりのかたまりがふること）、きゅうな つよいあめにも きをつけて ください。	2025年（ねん）7月（がつ）30日（にち）8時（じ）25分（ふん）頃（ごろ） カムチャッカ 半島（はんとう）の 近（ち）かくで おきた 地震（じしん）に よって、9時（じ）40分（ふん）、この県（けん）の 海（うみ）の 近（ちかく）に 津波（つなみ） 警報（けいほう）（「高（たか）い波（なみ）が 来（く）るので 逃（に）げて ください」という 強（つよい） 注意（ちゅうい）） が発表（はっぴょう）されました。 海（うみ）と 河口（かこう）（川（がわ）が、海（うみ）に つながる 場所（ばしょ））の 近（ちか）くの ひとは、急（いそ）いで はなれてください。高（たか）い所（ところ）などの 安全（あんぜん）な 場所（ばしょ）に 逃（に）げて ください。 海（うみ）と 河口（かこう）の 近（ちか）くには 絶対（ぜったい）に いかないで ください。	石川県（いしかわけん） 能登（のと）地方（ちほう）・珠洲市（すずし）で とても 大（おお）きい 地震（じしん）が ありました。 あした 6日（にち）（土曜日（どようび））から、雨（あめ）が たくさん 降（ふ）るかもしれません。 地震（じしん）で 地面（じめん）が 弱（よわ）くなっていますから、 土砂災害（どしゃさいがい）＜山（やま）が 崩（くず）れることや、 石（いし）が 落（お）ちてくること＞に 気（き）を つけてください。 また、 暖（あたた）かく なったり 寒（さむ）くなったり するかもしれません。 病気（びょうき）に ならないように 気（き）を つけてください。

TER

```
import evaluate

# TER 評価器のロード
ter = evaluate.load("ter")

#  MT 出力（predictions）とポストエディット済みの日本語訳（HTER 用の references）
predictions = ["直ちに高台に避難して"]
references = [["今すぐ高いところに避げて"]]  # 内部はリストのリスト形式にする（複数参照訳対応）

# TER（=HTER）計算
results = ter.compute(
    predictions=predictions,
    references=references,
    normalized=True,
    case_sensitive=False,
    support_zh_ja_chars=True  # 日本語・中国語対応を有効にする
)

# 結果表示
print("スコア:", results["score"])
print("編集回数:", results["num_edits"])
print("参照長:", results["ref_length"])
```

```
スコア: 100.0
編集回数: 6
参照長: 6.0
```

Easy Japanese Guideline

誰とでも、もっと簡単に
コミュニケーションをとれるように

やさしい日本語ガイドライン

三　重　県

目次

はじめに

三重県には、2023年12月末現在で62，561人の外国人が暮らしており、県民（1，758，918人）の約28人に1人が外国人です。新型コロナウイルス感染症の影響で一時的に減少しましたが、2022年以降は再び増加傾向にあり、今後さらに増えていくことが見込まれています。

三重県では、考えや習慣などの違いを文化的な豊かさの源泉と考え、お互いが尊重し協力し合い、すべての人が生き生きと暮らせる多文化共生社会の実現を目指しています。誰もが安心して暮らすためには、情報を迅速かつ正確にわかりやすく伝えることが必要です。

「やさしい日本語」は阪神淡路大震災をきっかけに生まれた、外国人にも理解しやすい日本語です。このガイドラインでは、外国人に対して有効なコミュニケーション手段である「やさしい日本語」を多くの人に知っていただき、災害時だけでなく、職場や地域で接する外国人とのコミュニケーションに役立てていただけるよう作成しました。

「やさしい日本語」とは

「やさしい日本語」とは、むずかしい言葉を言い換えるなど、相手に配慮したわかりやすい日本語のことです。外国人だけでなく、子どもや高齢者、障がいのある人とのコミュニケーションにも役立つツールとして、国や自治体、地域社会、民間企業など様々な分野で取り組みが広がっています。

「やさしい」には２つの意味があります

なぜ英語ではなくて日本語なの？

三重県に住む外国人の国籍は多様化しており、その内訳は多い順に、ブラジル、ベトナム、フィリピン、中国、韓国、インドネシアで、英語が母語ではない国がほとんどです。その中には、「学んだ日本語を使ってみたい」「英語は話せないから日本語のほうが得意」という人もいます。

出入国在留管理庁の調査で、自身の日本語能力を『日常生活で必要な会話ができる』以上と回答した外国人が多いことから、日本語は外国人とのコミュニケーション手段として有効と言えます。

出入国在留管理庁「令和5年度　在留外国人に対する基礎調査報告書」より

「やさしい日本語」はどんなときに使うの？

災害などの緊急時に

地震や台風などの災害が起きたときは、多言語で翻訳したり、通訳できる人を探したりする時間がありません。災害の他にも、急病やケガ、事件など急いで情報を伝えないといけないときに「やさしい日本語」を使えば、多くの人に知らせることができます。外国語がわからなくても、ポイントを知れば誰にでも使えます。

日常生活で

日本語に不慣れな外国人は、日常生活の中で『役所から手紙が届いたけど読めない』、『子どもの学校の先生が話す敬語がむずかしい』と不安を感じることがあります。情報を多言語に翻訳して伝える方法もありますが、「やさしい日本語」を使って伝えることもできます。情報が伝わると不安が減り、自分で解決するきっかけになります。

外国人と日本語でやりとりするときに気をつけること

「外国人は日本語ができない」と思い込み、褒めたつもりで言った言葉や無意識の行動で相手を傷つけていることがあります。ここでは、外国人と接するときの心構えとして、彼らと日本語でやりとりするときに気をつけるポイントを3つ紹介します。

2021年の国際交流基金の調査によると、世界には約379万人の日本語学習者がいます。
この数は今後も増えつづけ、外国人と日本語でやりとりする機会も増えていくと予想されます。

外国人が母語以外の言語として日本語を学んでいることに敬意を示し、私たちも日本語を使うときは相手にわかるよう調節することで、円滑なコミュニケーションの実現が期待できるのではないでしょうか。

「これは漬物だよ、つ・け・も・の」「箸を使うのが上手だね～すごいね～！」

子ども扱いするような話し方をしない

日本語がまだ不十分な私は、何もできない子どもと同じ？
大げさに褒めなくてもいいので大人として扱ってください

「外国人に日本語で話しかけられたけど、英語で答えたほうがいいよね」

外国人には英語で話さなければ、と決めつけない

日本語で話しかけたのだから、日本語で答えてほしい
英語圏出身じゃないので英語はわからないよ～！

「小さな日本語のミスが気になる。日本で仕事をするなら正確に使ったほうがいいのでは？」

外国人に完璧な日本語力を求めない

日本語がまだ十分ではないけど、一生懸命仕事をしています
片言の日本語を笑われると傷ついて、話すのが怖くなります

参考文献:「とりあえず日本語で　もしも･･･あなたが外国人と『日本語で話す』としたら」(荒川洋平)

「やさしい日本語」の作り方

簡単な日本語であれば理解できる外国人は多いという調査結果がありましたが、その日本語能力は人それぞれ違います。私たちが普段何気なく話している日本語は、敬語やカタカナ語、あいまいな表現などがあり、日本人と話すときと同じ感覚で外国人と話すと、うまく意思疎通がとれないこともあります。

「やさしい日本語」はコツさえつかめば、誰でも使うことができます。少しの工夫で相手に伝わりやすくなりますので、以下のポイントを参考に実践してみてください。

ポイント	
1	むずかしい言葉を避けて、簡単な言葉を使う
2	１つの文に１つの情報にして、文は短くする
3	重要な言葉はそのまま使い、説明を加える
4	敬語（尊敬語・謙譲語）は丁寧語にする
5	文末表現はなるべく統一する
6	二重否定や、あいまいな表現は使わない
7	カタカナ語を使うときは気をつける
8	「分かち書き」をする
9	漢字は少なくして、使う場合はふりがなをつける
10	時間や年月日の表記はわかりやすくする
11	理解度を上げるために工夫する

各ポイントの詳細は次のページから説明していきます。

＼ポイント／

1 むずかしい言葉を避けて、簡単な言葉を使う

「やさしい日本語」は、日本語能力試験のN4・N5レベル（初級）程度の語彙を使います。
これは、小学校の2、3年生で習うくらいの漢字とひらがな及びカタカナによる表現です。
むずかしい言葉や専門用語は使わないようにして、わかりやすい言葉にしましょう。

N4 レベルの目安
趣味や食べ物、週末の予定などの身近で日常的な会話が理解できます。
友達から来たはがきやメールを読んで、理解できます。

N5 レベルの目安
『座ってください』などの簡単な指示を聞いて、何をするべきか理解できます。
駅の時刻表や案内板を見て、自分が乗る電車の時間がわかります。

参考資料：「日本語能力試験　合格者と専門家の評価による日本語能力試験レベル別 Can-do リスト」

- **漢語** ➡ 和語にする
  危険　→ 危ない
  体調が悪い → 元気がない
  お会計する → お金を払う
  ※ 漢語は中国から入ってきた言葉で、音読みで表す（雷雨、見物）
  ※ 和語はもともと日本にあった言葉で、訓読みで表す（雨、見る）
- **専門用語** ➡ 日常語にする
  捺印　→ はんこを押す
  授業参観　→ 授業を見る
  通報する　→ 連絡する ／ 電話する
- **略語** ➡ 元の言葉にする
  健診　→ 健康診断
  取説　→ 取扱説明書
  学割　→ 学生割引
- **抽象的ではっきりしない言葉** ➡ 具体例をあげる
  ペット　→ 犬や猫などの動物
  お年寄り　→ おじいさん・おばあさん
  粗大ごみ　→ イスや机など大きいごみ

＼ポイント／

2 １つの文に１つの情報にして、文は短くする

接続詞（～が、～ので）でつないで長い文にせず、1つの文に1つの情報になるよう文を短くしてください。伝える情報が多いときは、箇条書きにするのも有効です。また、解説や手順、ルールなどを説明するときは番号を振るとわかりやすくなります。

【例】　明日は清掃活動があるので、午前8時に事務所に集合してください。

→　明日は掃除をします。午前8時に事務所に来てください。

【例】　清掃活動をする際は、汗拭きタオルや飲み物、軍手を各自でご準備ください。

→　掃除をするときに持ってくるもの

- タオル　（汗を拭くために使います）
- 飲み物　（喉が渇いたときに飲みます）
- 軍手　（手を守るための手袋）

＼ポイント／

3 重要な言葉はそのまま使い、説明を加える

災害用語や日常生活でよく使われる言葉など、外国人にとってむずかしくても日本で生活する上で知っておくとよい言葉はそのまま使い、意味を説明するとわかりやすいでしょう。文章で伝える場合、言葉の後にかっこ書きで説明をつけます。

【例】	救急車	→	救急車（病気やけがをした人を助ける車）
	避難訓練	→	避難訓練（地震や台風のときに逃げる練習）
	余震	→	余震（あとからくる地震）
	停電	→	停電（電気を使うことができない）
	断水	→	断水（水道から水が出ない）
	行方不明	→	行方不明（どこにいるかわからない、見つからない）
	保護者	→	保護者（子どもの家族、お父さん、お母さんなど）
	熱中症	→	熱中症（暑くて気持ち悪くなる、頭が痛くなる病気）
	暗証番号	→	暗証番号（あなただけが知っている番号）

＼ポイント／

4 敬語（尊敬語・謙譲語）は丁寧語にする

敬語の中でも尊敬語と謙譲語は相手との関係性や状況に応じて使い分けるため、外国人には複雑で、簡単に覚えられません。日本語教育のほとんどの教科書では、「です」「ます」「～ください」を最初に習うので丁寧語を使ったほうが伝わりやすくなります。

【例】	お掛けください	→ 座ってください
	お召し上がりください	→ 食べてください
	拝見します	→ 見ます
	差し上げます	→ あげます ／ 渡します
	おいくつですか？	→ 何歳ですか？
	どちらにお勤めですか？	→ どこの会社で働いていますか？
	よろしいでしょうか？	→ いいですか？
	お大事になさってください	→ 体に気をつけてください

豆知識

「お水」「ご案内」･･･美化語はほどほどに

敬語の中に「美化語」があることを知っていますか？美化語とは、ご住所・お身体のように言葉の前に「お」や「ご」をつけることで、ものごとを上品に美しく表す言い方です。ところが、この美化語をつけることでアクセント（強く発音する部分）が変わることがあり、日本語に不慣れな外国人には聞き取りにくいことがあります。可能な範囲でいいので、「お」「ご」をつけるのをやめましょう。

【例】 「お薬はお食事の後です。お水はたくさんお取りください。」
→ 「薬は食事の後です。水をたくさん飲んでください。」

ただし、「お」「ご」を取ることで意味がわからなくなったり、乱暴になったりする言葉もあるので、その場合は言い換えを考えるなどの工夫が必要です。

【例】 お子様 → あなたの子ども
お手洗い → トイレ

＼ポイント／

5 文末表現はなるべく統一する

文法的にむずかしい表現は避けて、簡単でわかりやすい言葉にしてください。

- 可能・不可能の意味「～れる」、「～られる」 ➡ 「～することができる」
  駅まで来られますか？ → 駅まで来ることができますか？
  ここで桜が見られますよ → ここで桜を見ることができます
  その野草は食べられません → その草は食べることができません
- 不確定な表現「たぶん」、「おそらく」、「～ようです」 ➡ 「～かもしれません」
  おそらく行きます → 行くかもしれません
  台風が来る可能性があります → 台風が来るかもしれません
- 依頼表現「～しましょう」「～してもらっていい？」 ➡ 「～してください」
  時間を守りましょう → 時間を守ってください
  私の家に来てもらっていい？ → 私の家に来てください

＼ポイント／

6 二重否定や、あいまいな表現は使わない

例えば、「行きたくないわけじゃない」は否定語が2つあるので二重否定です。ストレートに伝えるよりも言葉の響きを柔らかくする効果がありますが、まわりくどい表現なのでわかりにくく、誤解を与えることがあります。相手に明確な意図を伝えたいときや、緊急性の高い情報を伝える場合は避けてください。

- 二重否定
  言わずにはいられない → 言いたい
  手続きができないとは言い切れない → 手続きできます
  （プレゼントをもらって）嬉しくないわけない → とても嬉しいです
- あいまいな表現
  ご遠慮ください → しないでください／だめです
  荷物はその辺に適当に置いてください → 荷物は机の上に置いてください
  （ペンを借りたいと聞かれて）大丈夫です → はい、いいですよ
- 複数の意味を持つ表現
  あなたは明るいですね → あなたは元気な人ですね
  （笑い話を聞いて）あ～、おかしい → あ～、面白い
  （もっと食べるか聞かれて）もう結構です → もういらないです

＼ポイント／

7 カタカナ語を使うときは気をつける

カタカナ語は、外国語や専門用語を表現するためにカタカナで表記された言葉です。古くから定着しているものから新しく生まれたものまで、さまざまなバリエーションがあります。

おもな分類	（言い換え例）
外来語 海外の言葉が日本語として定着したもの	デマ → うその話 ルール → 決まっていること サービス → してもらえること
和製英語 日本で作られた英語に似ている言葉	ＮＧ → やってはいけないこと ワイシャツ → シャツ、服 マイバッグ → 買ったものを入れる袋
日本語特有の略語 元の言葉を省略して短くしたもの 他の言葉同士を組み合わせたもの	スマホ → スマートフォン ドタキャン → 急にやめること パトカー → 警察の車

カタカナ語には元の言葉と意味や発音が違うものや、日本でしか通じない言葉もあるので、なるべく使わないようにしましょう。カタカナ語の中でも、日本語能力試験N4・N5程度で日常的によく使っているものや他の言葉での表現がむずかしいものは、そのまま使ってください。

日本語初級者が学ぶカタカナ語の例

スーパー、レストラン、デパート、コンビニ、アパート、エアコン、トイレ、
Ｔシャツ、ズボン、サンダル、カレーライス、ドーナッツ、ジュース、ビール、
タクシー、バス、バイク、ボールペン、アルバイト、ゲーム、アニメ　など

豆知識

オノマトペ（擬態語・擬音語）は伝わらない

「ばったり会う」「すらすら答える」「ちくっとする」などの言葉を、オノマトペ（擬態語・擬音語）といいます。物音や声、物事の様子・気持ちなどを具体的で感覚的に表した言葉で、日本人は日常で多くのオノマトペを使っています。しかし、オノマトペを使わない国は多く、外国人にはむずかしいです。
オノマトペはできるだけ使わないようにして、簡単な言葉に言い換えましょう。

【例】　わくわくします　→　とても楽しみです
　　すくすく育つ　→　元気に育つ
　　彼は英語がペラペラです　→　○○さんは英語が上手です

＼ポイント／

8 「分かち書き」をする

単語と単語の間や、文節にスペースを入れることを「分かち書き」といいます。分かち書きをすると、文の切れ目が明確になり文章が読みやすくなります。基本的には「ね」や「さ」を入れても不自然にならないところで区切ります。

【例】 ここではきものを脱いでください
→ ここで 靴を 脱いでください ※「履物」よりもわかりやすい表現にしました
→ ここで 着物を 脱いでください ※「着物」に限定しないなら「服」でもいいです

明日再度お越しください
→ 明日 もう一度 来て ください ※むずかしい言葉をわかりやすい表現にしました

津波は何度も繰り返し襲ってきて、後から来る津波の方が高くなることがあります。
→ 津波は 何回も きます。
大きい 波が 後から くることが あります。

昨夜、23時45分に三重県で強い地震がありました。気象庁は今後も最大で震度5までの地震が起きる可能性があるとして注意を呼び掛けています。
→ 昨日の夜 11時45分に 三重県で 大きい 地震が ありました。
この後も 大きい 地震が くるかも しれません。
注意して ください。

「分かち書き」をしない日本語、実は世界で少数派

英語で文章を書くときは、単語と単語の間にスペースを入れるルールがあります。
単語を全部つないでしまうと読みづらくなり、ひとつの単語の区切りがわかりにくいためです。

Mie Prefecture is located in the Tokai region between the cities of Nagoya to the east and Osaka to the west.
三重県は名古屋と大阪の間にあり、東海地域に位置する県です。

世界の言語のほとんどが「分かち書き」で表記されており、単語が全てつながっているのは日本語・中国語・タイ語など、一部の言語に限られています。
日本に来る外国人は年々増加しており、病院や役所、観光地で多言語表記による案内も増えています。もし見かけたときは、「分かち書き」になっているか注目してみてください。

＼ポイント／

9 漢字は少なくして、使う場合はふりがなをつける

漢字を使わない国の出身者にとって、漢字を読んだり、意味を理解したりすることは簡単ではありません。ふりがながあっても読むのに時間もかかり、発音はわかっても意味がわからない人もいます。漢字の量は多すぎないようにして、使う場合はすべての漢字にふりがなをつけてください。

※漢字の使用量は1文に3、4字程度が目安です

【例】 公共交通機関でお越しください → 電車(でんしゃ)やバスで来(き)てください

至急警察に通報してください → 今(いま)すぐ警察(けいさつ)に電話(でんわ)してください

医療機関を受診してください → 病院(びょういん)に行(い)ってください

家庭訪問を行います → 先生(せんせい)が家(いえ)に来(き)て話(はな)しをします

＼ポイント／

10 時間や年月日の表記はわかりやすくする

- 元号は使わず、西暦にする
  令和6年7月2日 → 2024年7月2日
- 年月日の表記にスラッシュ「/」やドット「.」を使わない
  2010/11/12 → 2010年11月12日
  2015.4.1 → 2015年 4月 1日
- 時間は12時間表示にして、午前・午後をはっきり示す
  15時30分 → 午後3時30分
- 「○○～△△」ではなく、「○○から△△まで」にする
  【例】 博物館の開館時間は9:00～17:00です
  → 博物館は、午前9時から午後5時まで開いています

＼ポイント／

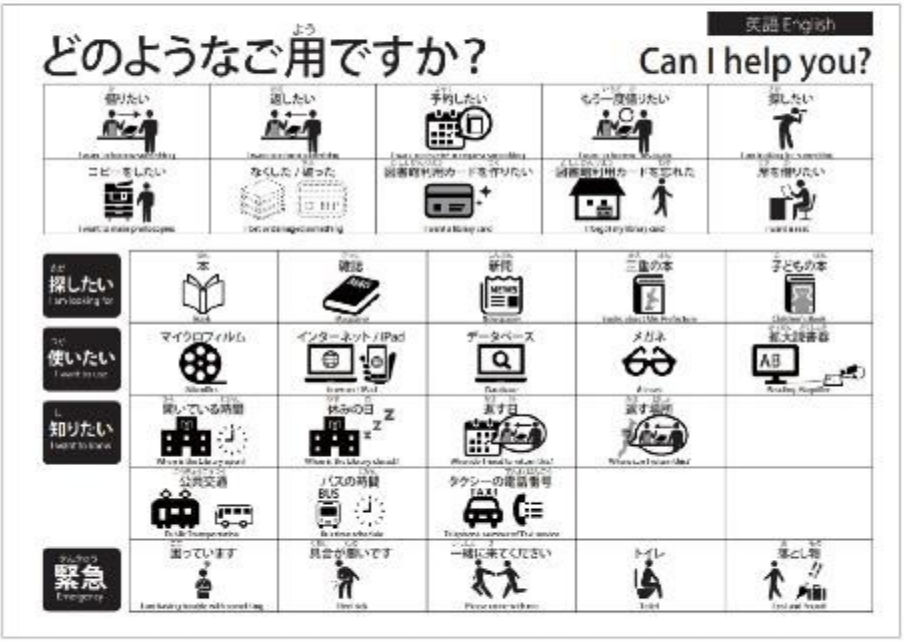
11 理解度を上げるために工夫する

- 文字だけで伝えずに、イラストや写真、図などを活用して視覚的にわかりやすくする
- 身ぶり手ぶり（ジェスチャー）を交えたり、表情で伝えたりすることも有効
  ※ただし、ジェスチャーの意味は世界共通ではないので注意が必要
- 覚えてほしいことや重要な情報はメモに書いて渡しておくと、あとで確認することができ、家族や日本語がわかる知人に尋ねることもできる
- 翻訳テキストやイラスト・写真を指し示すことで意思表示ができるコミュニケーションボードを準備する

豆知識

三重県立図書館に「コミュニケーション支援ボード」があります！

「本を借りたい」「利用カードを作りたい」など、絵や文字を指し示すことでコミュニケーションができます。図書館を利用するときに、コミュニケーションのバリアをなくしていくことの手助けになるよう作られたもので、障がいのある方、高齢者や幼児、外国人の方にも幅広く活用されています。ボードは絵とやさしい日本語に加えて、英語、ポルトガル語、スペイン語、中国語で記されたものが用意されています。

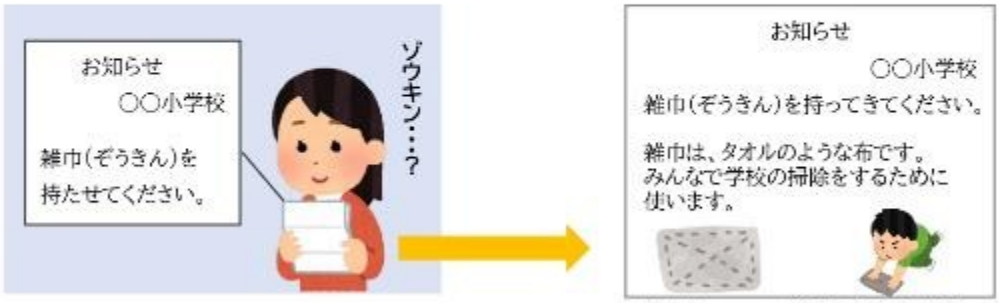


「やさしい日本語」に大切なこと

「やさしい日本語」のポイントを一通り見ていただきましたが、何が「やさしい」のかは情報を伝えたい相手の出身や経験、文化によって違います。「やさしい日本語」では、言葉を簡単な表現にするテクニックだけではなく、コミュニケーションにおける態度について考え、相手の事情を考える思いやりの心を持つことが大切です。

「あたりまえ」という思い込みを捨てる

国籍、文化・習慣、世代などの違いにより、自分が「あたりまえ」と思っていることが通用しないこともあります。例えば、日本の学校では雑巾を使って教室を掃除するのが一般的ですが、これは世界共通の文化ではありません。ふりがなをつけたり、ひらがなで表記したりしても、そもそも雑巾を知らなければ意味が伝わらないのです。自分の常識や習慣を相手に押し付けるのではなく、相手の文化や価値観を理解し、どのような配慮や工夫が効果的なのかを考えることが大切です。



※雑巾について補足の情報や説明、
実物の写真などがあるとよい。

「第三者返答」に気をつける

"第三者返答とは、話しかけてきた人の見かけの印象などから、
その人との意思疎通が問題ないにもかかわらず無視して、その人と一緒にいる人に返答すること"

オストハイダ・テーヤ(関西学院大学教授)

外国人に話しかけられたのに、「日本語がわからないだろう」「日本人に話した方が伝わる」と思い込み、付き添いの日本人に向かって話してしまうケースがあります。外国人に限らず、障がい者、高齢者も同じ経験をすることがあり、当事者は自分の存在が無視されていると感じ、傷つきます。相手の見た目や印象でコミュニケーションができないと判断せずに、話しかけられた相手の目を見て、直接本人と話すことを心がけましょう。

相手の背景に合わせた「やさしさ」を

漢字を使わない国の出身者には、ひらがなを多く使った文章の方が伝わるかもしれませんが、中国や台湾などの出身者には漢字が多い方が伝わりやすいこともあります。

日本語を話す能力についても人それぞれ違うので、来日してから数か月で流暢に日常会話ができる人もいれば、日本で長く暮らしていても、あいさつ程度しか話せない人もいます。
また、外国人でも日本語が堪能で日本社会についてよく知っている人の場合は、無理に単純な日本語を使うと不愉快に感じさせることがあります。

「やさしい日本語」に決まった正解はありません。どうしたら相手にわかりやすく伝えることができるかを常にイメージし、工夫することが大切です。相手を思いやり、『わかりやすい言葉』と『やさしい気持ち』でコミュニケーションをとりましょう。

ひらがなだけの文章は、本当に「やさしい」？

日本語を話すのは得意だけど、漢字の入った長い文章を読むのが苦手な外国人は多いです。
漢字が少しでも混じっていると、それだけで読むことを諦めてしまう人にとって、ひらがなの文は受け入れやすいです。外国人向けに情報を伝えるときに、すべてひらがなで書くという選択肢もありますが、注意が必要です。

例えば、「りすとひょうをおくります。かみをまとめたらこうかいします。」という文章があります。「リスト表を送ります。紙をまとめたら公開します」という意味で解釈する人が多いと思いますが、「リスとヒョウを送ります。髪をまとめたら後悔します」と解釈する人もいるかもしれません。

このように、日本語には同音異義語（発音が同じだけど意味が違う言葉）が多く、ひらがなだけの文章では解釈に困ることがあります。上の例文のように単語の区切りがないと、さらに混乱するでしょう。ひらがなだけの文章にする場合は、単語の区切りがわかるようにスペースを空けて、「分かち書き」にするとわかりやすくなります。

「やさしい日本語」へ変換してみよう①

1. 単語編

漢語 → 和語		
1	今朝	
2	更新	
3	不要です	
4	小児科	
5	終日運休	
カタカナ語 → 和語		
1	キャンセル	
2	(タイヤの)パンク	
3	ヘルシーな朝食	
4	ストレス	
5	ハザードマップ	
敬語 → 普通語		
1	ご両親	
2	どなた	
3	いかがですか？	
4	どちらのご出身ですか？	
5	ご連絡いたします	

2. 文章編

① 事前の申し込みが必要です

➡

② こちらはセール対象外です

➡

③ お子さんは何かアレルギーがありますか？

➡

④ お大事にしてください

➡

回答例は次のページです

回答例①

1. 単語編

漢語 → 和語		
1	今朝	今日の朝
2	更新	新しくする
3	不要です	いりません／いらないです
4	小児科	子どもの病気やけがを治す病院
5	終日運休	今日は電車(バス)が動いていません／止まっています
カタカナ語 → 和語		
1	キャンセル	やめる
2	(タイヤの)パンク	タイヤに穴が開く／タイヤが壊れる
3	ヘルシーな朝食	体にいい朝ごはん
4	ストレス	疲れること／心配なこと
5	ハザードマップ	地震や台風のときに危ないところがわかる地図
敬語 → 普通語		
1	ご両親	○○さんのお父さんとお母さん
2	どなた	だれ
3	いかがですか？	どうですか？
4	どちらのご出身ですか？	どこの国から来ましたか？／あなたの国はどこですか？
5	ご連絡いたします	電話します／メールします

2. 文章編

① 事前の申し込みが必要です
➡ ○月○日までに申し込んでください

② こちらはセール対象外です
➡ これは安くなりません

③ お子さんは何かアレルギーがありますか？
➡ ○○さんの子どもは、体がかゆくなったり、赤くなったりする食べ物がありますか？
➡ あなたの子どもは、体の具合が悪くなる薬がありますか？

④ お大事にしてください
➡ 体に気をつけてください

「やさしい日本語」へ変換してみよう②

3. 身近な場面で使われる日本語

同じ地域に暮らす外国人や一緒に働く外国人とのコミュニケーションを想定して、日本ではおなじみの制度や専門用語をわかりやすい表現にしましょう。

地域で・・・		
1	ポイ捨て禁止	
2	不在連絡票	
3	町内会	
職場で・・・		
1	給与明細	
2	有給休暇	
3	源泉徴収票	
4	雇用保険	
5	労災保険	
役所で・・・		
1	所得税	
2	住民税	
3	年金	
4	母子手帳	

回答例は次のページです

回答例②

3. 身近な場面で使われる日本語

地域のこと		
1	ポイ捨て禁止	ごみを捨てないでください
2	不在連絡票	あなたに荷物が来ました。あなたは家にいませんでした。だから、宅配便の人は荷物を持って帰りました。そのお知らせです。
3	町内会	同じ地域に住んでいる人たちが集まって、いろいろな活動をします。地域の掃除、お祭りがあります。
職場で・・・		
1	給与明細	給料のことが書いてある紙
2	有給休暇	仕事を休んでも給料がもらえる休みの日
3	源泉徴収票	去年働いた分の給料と、税金をいくら払ったか書いてある紙
4	雇用保険	仕事がなくなったとき、家族の世話で働けないときにお金をもらうことができます。
5	労災保険	仕事でケガをしたとき、病気になったときにお金をもらうことができます。
役所で・・・		
1	所得税	給料をもらうときに国に払う税金
2	住民税	あなたが住んでいる市と県に払う税金
3	年金	年をとって、働けなくなったときに国からお金をもらうことができます。日本に住む20歳から59歳の人は必ず年金を払います。（※国民年金の場合）
4	母子手帳	母子はお母さんと赤ちゃんのことです。手帳は小さいノートです。お母さんと赤ちゃんの体のことを書きます。赤ちゃんを育てるために必要なことが書いてあります。

参考:文化庁「つながるひろがる にほんごでのくらし」https://tsunagarujp.mext.go.jp/

「やさしい日本語」言い換えリスト

※言い換え例は、文節ごとにスペースを入れた「分かち書き」にしています。

緊急・災害

	用語	言い換え例
か	緊急	すぐに
	警報	つよい 注意の お知らせ 大きな 災害が おこるかもしれません
	洪水	川から 町に 水が くる
さ	地震	地面が ゆれる
	情報	お知らせ
	浸水	家の 中に 水が 入る
た	台風	とても 強い 風と 雨
	津波	地震の あとに 海の 水が たくさんくる
は	負傷者	けがをした人
	暴風雨	とても 強い 風と 雨
や	行方不明	どこに いるか わからない
	予報	これから 起きることの お知らせ

健康・医療

	用語	言い換え例
か	健康	元気なこと、病気でないこと
	健康診断	体が 元気か しらべる
さ	自己負担	自分で はらう お金
	症状	体の 悪いところ(痛い、血が出るなど)
	処方せん	くすりを 書いた 紙 (医者が 書く)
	診察	医者が からだを しらべる
な	入院	病院に 泊まって 病気や けがを なおす
は	保険証	保険に 入っている人の 紙や カード
ま	問診票	体の 悪いところを 書く 紙
や	予防接種	病気に ならないための 注射

生活・仕事

	用語	言い換え例
か	給料	はたらいて 会社から もらう お金
	契約	法律で 守らないと いけない 約束
さ	サービス	してもらえること
	残業	きまった 時間より 長く はたらく
	失業する	仕事が なくなる
	請求書	お金を 払ってくださいという 紙
た	問い合わせる	わからないことを 聞く
は	変更	変わる
	本人確認書類	運転免許証、パスポート、在留カードなど
ま	無料	お金が いらない
ら	領収書	お金を はらったときに もらう紙（レシート）

子育て・教育

	用語	言い換え例
あ	1学期、2学期、3学期	4月から7月、9月から12月、1月から3月
	上靴、上履き	学校の 中で 履く 靴
か	学童保育	小学校が 終わったあとに 子どもが 行く ところ。 そこで 宿題を します。友達と 遊びます。
	学級閉鎖	病気の 子どもが 多いので 学校が 休みになる
	給食	学校が 用意する 昼ごはん
	教科書	学校で 使う 勉強の 本
さ	出生届	子どもが うまれたことを 役所に 知らせる 紙
	奨学金	勉強のために 借りることが できる お金 勉強のために もらうことが できる お金
は	PTA	保護者(子どもの 家族)と 先生が 集まって、 子どもたちの 学校生活を 助ける 活動を します。
	保育園	働く人が 子どもを 預ける ところ
や	幼稚園	小学校へ 入る 前の 子どもが 通う ところ

参考:大阪府「やさしい日本語をつかいましょう！ ～外国人とのコミュニケーション術～」
https://www.pref.osaka.lg.jp/documents/11368/1-40.pdf

参考資料・ウェブサイト

「やさしい日本語」ガイドラインを作成する際に参考にした資料やウェブサイトを紹介します。
二次元コードからアクセスするとリンク先に移動します。(2025年1月現在の公開情報に基づく)

【出入国在留管理庁】在留支援のためのやさしい日本語ガイドライン
国や自治体が情報発信をするときに、やさしい日本語を使って在留外国人にもしっかりと情報が届くようになることを目指して作られた。特に、書き言葉に焦点を当てている。

【静岡県】やさしい日本語 文書作成のポイント
やさしい日本語の作り方の基本や、役立つツールを紹介。これらをまとめた「やさしい日本語の手引き」もダウンロードできる。

【静岡県富士市】地域活動のためのやさしい日本語のしおり
地域で暮らす外国人とのコミュニケーションに役立ててもらうために作られた。
やさしい日本語の基本ルールの他に、生活のルールや町内活動の伝え方のコツを掲載。

【愛知県】「やさしい日本語」の手引き
やさしい日本語を広く普及するため、初心者でも手軽に学べるように作られた。
やさしい日本語を作るときのポイントの他に、方言を交えた練習問題や用語集を収録。

【大阪府】「やさしい日本語」を使いましょう!~外国人とのコミュニケーション術~
やさしい日本語の書き換え練習では、災害/健康・医療/生活・仕事/年金・税金/子育て・教育と、幅広いカテゴリーに分けて掲載。回答例は分かち書きを採用している。

【群馬県】外国人に伝わりやすい「やさしい日本語」講座(基礎編)
群馬県のやさしい日本語講座で実際に使われた教材を公開したもの。やさしい日本語が求められた背景や作り方の基本、練習問題を掲載。

【栃木県宇都宮市】外国人市民への情報提供ガイドライン
外国人向けの情報提供について、4つの章で構成されたガイドライン。実用的なやさしい日本語の使い方のほか、多言語翻訳との比較やコラムなど豊富な情報を掲載。

【文化庁】日本語学習サイト つながるひろがる にほんごでのくらし
全49シーン、全135場面の学習を通じて、外国人が生活に必要な日本語を学べる。
掲載されている日本語はシンプルでわかりやすい表現になっている。

三重県で働く国際交流員に聞いてみました！

外国人とのコミュニケーションで大切なこと

オーストラリア出身　Kさん

日本では、「おいで」と手のひらを下に向けて手招きをするジェスチャーがありますが、海外だと「あっちへ行け」という意味になります。日本でよく使われるジェスチャーの中には、外国人に通じないものや誤解を招くものがあるので気をつけてください。

ブラジル出身　Mさん

日本人に話しかけると、相手がとても焦っていることがあります。外国人と話すのって緊張するのでしょうか・・・？それを見ると、こちらまで緊張してしまい、迷惑をかけているような気持ちになって日本語が出てこなくなります。だから、落ち着いてゆっくり話してくれると助かります。

中国出身　Sさん

無意識だと思いますが、日本人の話すスピードは速いです。店員さんがマニュアル通りに話す言葉は早口で、聞き取れないこともあります。丁寧に対応してくれるのは嬉しいけど、敬語が多いとわかりにくいです。ゆっくり話すことを意識して、シンプルな表現を使ってください。

理解と協力

ハートでつくる多文化共生

やさしい日本語　ガイドライン
令和7年1月

発行：三重県　環境生活部　ダイバーシティ社会推進課
〒514-0009　三重県津市羽所町700 アスト津３階
TEL:059-222-5974　FAX:059-222-5984
E-mail:tabunka@pref.mie.lg.jp