Contrast in Cohesion between Human and Machine Translations

Ibnu Senna Sholiqin Putra[1], Donald Jupply[2]

ibnu.senna2016@student.uny.ac.id , donaldjupply@uny.ac.id

Universitas Negeri Yogyakarta
Yogyakarta, Indonesia

Abstract. While many consider that machine translation is helpful and handy, a large body of research has revealed that machine translation remains marred by inaccuracy issues. This is a concern, especially when it comes to cohesion as it determines the coherence of the text. However, this area seems to have not caught the attention of translation researchers and remains under researched. This study attempts to investigate the contrast in cohesion of the cohesive devices attempted by machine translation and human translation. Using qualitative method, the findings suggest that while the two analyzed texts contain similar number of cohesive devices, the machine translations of the cohesive devices incur significant accuracy issues which affect the overall coherence of the target text.

Keywords: human translation, machine translation, cohesion, cohesion shift

1 Introduction

In recent years, machine translation has been the language solutions for instant and convenient agent of translation. According to the blog "Ten Years of Google Translate" by Google, back in 2016, there were 500 million active users of the service with over a hundred billion words being translated a day. Being one of the most popular choices among many users, Google Translate is indeed an established player in the machine translation community. With that being said however, Google Translate is not without its flaws. Concerns regarding inaccuracy issues such as complete mistranslation often plague the end results of a machine translation. Studies including one conducted by Dewi Kartika of Universitas Negeri Semarang in 2017 found that the quality of machine translation (or in this case Google Translate) was deemed as "less accurate, less acceptable, and less readable" and is suggested for users to employ "simple instead of complex sentences" when dealing with machine translation (Kartika 2017). In other words, public confidence in machine translations is limited to the extent that machine translated texts are generally better suited to general purpose, less formal uses. To explore more of this inaccuracy issues, the research explores the flaws machine translation makes through the analysis on cohesive devices.

In regard to translation, cohesion is one of the crucial factors in determining the coherence of a given text, especially those with regard to the treatment of translation process. These considerations are deemed essential when it comes to assessing the quality of translated works, especially ones being done by machine translation. Hence, contrasting analysis of the

ELTLT 2021, August 14-15, Semarang, Indonesia

DOI 10.4108/eai.14-8-2021.2317630

result between human and machine translation-based texts is essential to determine and assess the quality and accuracy of the translated works in question. These are based on the thought that the two methods significantly differ in nature and might contribute to the end results of the translation.

To investigate the translated cohesive devices of machine translation compared to human translation, the research proposes two sets of objectives. The first is to determine how cohesion in the source text is realized and translated by human and machine translation. The second is to find out how the cohesive devices of human and machine translation differ compared to the source text. The object of the research is the novel Harry Potter and the Half Blood Prince. The analysed data are correlated in terms of cohesion between the machine based and human based translation

2 Method

The research used a qualitative method to analyse the cohesive devices in the Half Blood Prince novel. A simple counting was used in the mathematical calculations to determine the number of occurrences of the cohesive types in the Harry Potter novel, which includes data enumeration. Also, the use of descriptive and expositional means of analysing of each cohesive devices to obtain conclusive findings signifies the use of qualitative measurement of the research.

To answer the first and partially answer the second question, the research categorizes the acquired data from the two texts into Halliday & Hasan model of cohesive devices and compares the two into a unified table consisting of the number of occurrences each time a cohesive device presents on each respective text. Should any contrasting cohesive elements be discovered between the texts, the elements are analysed through descriptive and expositional analysis.

In addition to Halliday & Hasan model of cohesive devices, the research also employs Corder's model of error analysis to categorize the errors made by machine translation. Corder (1973) classifies errors into four categories, namely omission, addition, misinformation, and misordering. The classified data were then used as a definitive factor in order to determine the errors made by machine translation.

The data of the research were based on the novel "Harry Potter and The Half Blood Prince" by J. K. Rowling, the sixth novel in the series of magical fantasy narrative of the titular character Harry Potter during his (and his friends) adventures in Hogwarts School of Wizardry. The data, consisted of the first 100 lines from the book, were uploaded to the Google Translate for processing. The restriction in data sampling was due to the inadequate capabilities of the internet bandwidth to upload the whole novel to the machine translation website.

There were two main instruments of the research. One was the researchers; whose effort included data collecting, data analysis, interpreting, and report processing. The second instrument was the data sheet, which consisted of identified cohesive devices of the texts.

During data collection, the data were firstly identified of their cohesive device. Then, the data were collected and assigned into their respective rows based on the identified cohesive devices. The cohesive types were then determined according to their respective categorized data, and then be marked in their respective columns of the table. The categorized data may contain more than one set of cohesive devices.

In the data analysis, data classification was the first step taken to determine whether any shifting had occurred. Next, each of the cohesion type was analysed by the means of

contrastive analysis. To answer question number one, a sample of each cohesive devices was brought up for analysis. The data sample of the second question were taken from the data of shifting and then analysed for the causes that generate the shifting.

3 Findings and Discussions

3.1. Findings

The research results reveal that the ST has 636 number of cohesive devices, while the human translation and machine translation have 601 data and 548 respectively. The data are presented based on model of Halliday and Hasan's cohesive devices as seen in the table below

Table 1. The overall realization of cohesive devices

Cohesive Devices	Source Text	Human Translation	Machine Translation
Reference	377	338	319
Substitution	25	24	23
Ellipsis	2	2	2
Conjunction	35	35	33
Lexical Cohesion	198	203	185
Total	637	602	562

The source text has 637 number of cohesive devices. On the other hand, the human translation and machine translation texts have 602 and 562 data respectively. The amount of data in the source text consists of 377 references, 25 substitutions, 2 ellipsis, 35 conjunctions, and 198 lexical cohesions. The 602 data of cohesion in human translation text consists of 338 references, 24 substitutions, 2 ellipsis, 35 conjunctions, and 203 lexical cohesions. In the 562 data of machine translation text, 319 of them are references, with 23 substitutions, 2 ellipsis, 33 conjunctions, and 185 lexical cohesions.

In addition to the Halliday & Hasan's model of cohesive devices, the research results also reveal several variances (i.e., alterations) of the machine translation that is notably different from the two benchmarked texts. The alterations are then classified into Corder's model of error analysis as shown below.

Table 2. Corder's Error Analysis Classification

Types	Addition	Omission	Misordering	Misinformation
Occurrences	11	10	8	50

The error analysis follows the same model of Corder's error analysis as shown in chapter II. The errors were found in the machine translation compared to the human translation. The numbers are 11 errors of addition, 10 errors of omission, 8 errors of misordering, and 50 errors of misinformation

Table 3. Cohesion Shift Between ST and HT

ST			HT												TOTAL	
							Shifted									
			Not Shifted		Loss		Reference		Substitution		Ellipsis		Conjunction		Lexical Cohesion		Num of Data	% of Data
R	377	100%	327	87%	36	10%	11	3%	1	0%	0	0%	0	0%	2	1%	377	100%
S	25	100%	23	92%	0	0%	0	0%	0	0%	0	0%	0	0%	2	8%	25	100%
E	2	100%	2	100%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	2	100%
C	35	100%	35	100%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	35	100%
LC	198	100%	192	97%	0	0%	0	0%	0	0%	0	0%	0	0%	7	3%	203	100%
Total	637	100%	579	91%	36	6%	11	2%	1	0%	0	0%	0	0%	11	2%	637	100%

Note: R: Reference, S: Substitution, E: Ellipsis, C: Conjunction, LC: Lexical Cohesion, ST: Source Text, HT: Human Text

Several cases of shifting between cohesion within the ST and HT were detected. In reference, 327 (87%) data were not shifted, with 36 (or 10%) of the data were lost in the translation process. The remaining 4% of the data was shifted into substitution (with 1 data or less than 1%) and lexical cohesion with 2 data (1%). An addition of 11 references were found in the HT which account for 3% of the data.

In substitution, 23 of the 25 (92%) of the data were not shifted, with the remaining 2 (8%) shifted into lexical cohesion. No case of shifting in ellipsis and conjunction were found. 192 out of 198 data in lexical cohesion (which account for 97% of the data) were not shifted, with addition of 7 data were found.

In total, the source text contains 636 cohesive devices, 578 (or 91%) of which were not shifted in the human translation. 36 (or 6%) of the data were lost during the translation process due to multiple factors such as word choices and language typology. 11 (or about 2%) of the data were shifted into reference, and 11 (also about 2%) were shifted into lexical cohesion. The remaining 1 datum was shifted into substitution, which accounts for less than 1% of the total data.

The next page represents the number of shifting between HT and MT. The methods and classification between ST – HT and HT – MT are otherwise similar. However, additional columns to analyse Corder's error analysis are included in the data sheet

.

Table 4. Cohesion Shift and Errors between HT and MT

HT			MT																						TOTAL	
			Not Shifted		Loss		Shifted										Error									
							Reference		Substitution		Ellipsis		Conjunction		Lexical Cohesion		ADDITION		OMISSION		MISORDERING		MISINFORMATION		Cohesion	Cohesion + Errors
Reference	338	100%	275	81%	20	6%	22	7%	0	0%	0	0%	0	0%	0	0%	9	2%	6	2%	6	2%	34	9%	319	374
Substitution	24	100%	22	96%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	4%	0	0%	4	14%	23	28
Ellipsis	2	100%	2	100%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	2	2
Conjunction	35	100%	33	94%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	3%	0	0%	0	0%	1	3%	33	35
Lexical Cohesion	203	100%	176	95%	7	3%	1	1%	1	1%	0	0%	0	0%	0	0%	1	0%	3	1%	2	1%	11	5%	185	202
Total	602	100%	508	90%	27	5%	23	4%	1	0%	0	0%	0	0%	0	0%	11	2%	10	2%	8	1%	50	9%	562	641

Note: MT: Machine Translation.

The table 4.A.2.2 above represents the amount of shifting between HT and MT. In total, there are 601 data in the HT and 562 data in the MT. An error analysis discovers additional 79 data, which brought the tally into 641.

In reference, 275 (81%) data were not shifted, with 22 (7%) additions were found. 1 of the data (account less than 1%) was shifted into ellipsis. 55 errors were found, with 9 (2%) being classified into addition error, 6 (2%) being omission, 6 (2%) being misordering, and 34 data (9%) being a misinformation.

In substitution, 22 (96%) of the data were not shifted, with 1 data (accounts for 4%) missing as an omission error. The remaining 4 (which accounts for 14% of the data) were classified into misinformation error.

No shifting in ellipsis were detected. By contrast, there are 1 (3%) addition and 1 (3%) misinformation in conjunction. The remaining 33 (94%) data in conjunction is not shifted.

In lexical cohesion, 176 (95%) of the data were not shifted. 7 of the data (3%) were lost. 1 (1%) of lexical cohesion data were shifted into reference, 1 (1%) shifted into substitution, 1 (less than 1%) were classified as addition error, 3 data (which account about 1%) being omission error, 2 of the data is a misordering, and 11 (5%) of the data being a misinformation error.

3.2.Discussions

3.2.1. The Realization of Cohesive Devices

The section discusses how the cohesive devices are used in the three versions of Harry Potter and the Half Blood Prince. Each cohesive device is discussed in their respective sections. The discussion of cohesive devices themselves is further split into their respective text types, which include narratives and dialogues.

3.2.2. References

An example of shifting in reference can be found in the following text, where a dialogue-type text requires different form of references to function as addressing pronouns.

*ST: At this, the Prime Minister had found his voice at last. "**You**'re — **you**'re not a hoax, then?" It had been his last, desperate hope.*

*HT: Mendengar ini, akhirnya Perdana Menteri bisa bicara lagi. "Jadi, **Anda** bukan olok-olok?" Itu harapannya yang terakhir, harapan dalam keputusasaan.*

The use of reference in this example can be found in the use of first-person pronouns in the spoken lines of the Prime Minister. The dialogues in which the Prime Minister is speaking to Cornelius Fudge, the magic minister, employs the word "You" and "Anda" as a way to refer to the magical minister. The use of third person reference reappears as the narration part of the story continues with the line of "It had been **his** last, desperate hope", which again refers to the Prime Minister of Muggles.

One important point to note is the use of "Anda" to replace the word "You" in the human translation. The use of "Anda" reflects the social setting where in this instance a conversation between two high-ranking government officials is taking place. The word "Anda" is then used as a way to address someone in a more polite fashion. This implies context of sociocultural politeness is not detected by machine translation, as evident in the example below

*HT: Mendengar ini, akhirnya Perdana Menteri bisa bicara lagi. "Jadi, **Anda** bukan olok-olok?" Itu harapannya yang terakhir, harapan dalam keputusasaan.*

*MT: Mendengar ini, Perdana Menteri akhirnya menemukan suaranya. "**Kau** - lalu kau bukan tipuan?" Itu adalah harapan terakhirnya yang putus asa*

In the machine translation datum, the word "you" to refer Cornelius Fudge has been replaced by the word "Kau." In the context of Indonesian socioculture, the word "Kau" and "Kamu" are considered to be an informal way to address someone. Using the utterance in formal setting would be considered impolite. The datum serves as an evidence that machine translation is not able to pick up sociocultural context of the text.

3.2.3. Substitution

An example of substitution in the data can be found in the following text, where "The President of a distant country" is substituted with "the wretched man" in the source text.
ST: "He was waiting for a call from the **President** of a far distant country, and between wondering when **the wretched man** would telephone, and trying to suppress unpleasant memories of what had been a very long, tiring, and difficult week, there was not much space in his head for anything else."

*HT: "Dia sedang menunggu telepon dari **presiden** Negara yang jauh, dan diantara bertanya-tanya kapan **orang sialan** itu akan menelepon dan berusaha menekan ingatan tak*

menyenangkan akan minggu yang sangat panjang, melelahkan, serta sulit, nyaris tak ada ruang tersisa di otaknya untuk hal-hal lain."

In this case, the word "wretched man" is translated into "orang sialan" in the human translation and into "orang malang" in the machine translation respectively. This is particularly important where the definitive meaning of "wretched" can be either "sad or miserable" or be used to express annoyance of said person. This may result in two different meanings, as the readers may assume that the Prime Minister is feeling somewhat sorry for the President, or that he despises the man in question.

*HT: "Dia sedang menunggu telepon dari **presiden** Negara yang jauh, dan diantara bertanya-tanya kapan **orang sialan** itu akan menelepon dan berusaha menekan ingatan tak menyenangkan akan minggu yang sangat panjang, melelahkan, serta sulit, nyaris tak ada ruang tersisa di otaknya untuk hal-hal lain."*

*MT: "Dia sedang menunggu telepon dari **Presiden** negara yang jauh, dan antara bertanya-tanya kapan **orang malang** akan menelepon, dan mencoba untuk menekan kenangan tidak menyenangkan tentang apa yang telah terjadi Minggu yang sangat panjang, melelahkan, dan sulit, tidak ada banyak ruang di kepalanya untuk hal lain."*

In the case of machine translation, the word "the wretched man" was translated into "orang malang" instead. The difference may result in two different impressions of the Prime Minister's stance to the President (as stated earlier), with the Prime Minister in human translation being annoyed by the President or the President's call, and the machine translation version is more benevolent. Although the impressions were of little importance to the plot as a whole, the case still presents a unique challenge faced by the translators as mere word choices may affect the reader's perception.

One possible (albeit rather odd) explanation as to why the human translator chooses to employ "orang sialan" instead of the more sympathetic "orang malang" is that "orang malang" can also be used to refer to people of Malang, a notable city located in East Java, thus "orang sialan" is used, arguably to avoid confusion. The effect of the word choice is that the human translation and machine translation carry different connotations. In this context, the word "orang sialan" carries a negative connotation, as it implies that the narrator or the Prime Minister hates the President. Conversely, the word "orang malang" carries a positive connotation where it implies that the narrator or the Prime Minister is feeling sympathetic to the President.

3.2.4. Ellipsis

Example of ellipsis in the data source can be observed in the following dialogues, which also happens to be a dialogue type text.

*ST: "I — what — dragons?" spluttered the Prime Minister. "**Yes, three**," said Fudge. "And a sphinx. Well, good day to you."*

*HT: ""Saya—apa—naga?" gagap Perdana Menteri. "**Ya, tiga**," kata Fudge. "Dan satu sphinx. Nah, selamat siang."*

*MT: ""Aku - apa - naga?" kata Perdana Menteri."**Ya, tiga**," kata Fudge. "Dan sphinx. Baiklah, hari baik untukmu"*

As mentioned, ellipsis omits the repeating items in the text that the readers may be able to comprehend completely without their existence. As evident in the data above, there were supposed to be a further reference to the three dragons that Fudge had mentioned. As the readers are more likely able to comprehend the meaning based on the previous mentions during the conversation, the word "**dragons**" and "**naga"** in Fudge's part of the conversation would be an unnecessary repetition and thus omitted in the form of ellipsis.

3.2.5. Conjunction

One example of conjunction in a narrative setting can be found in the following text.

*ST: He had gazed hopelessly at the Prime Minister for a moment, **then** said, "Well, sit down, sit down, I'd better fill you in…Have a whiskey…"*

*HT: Dia menatap Perdana Menteri tanpa harapan selama beberapa saat, **kemudian** berkata, "Nah, duduklah, duduklah, sebaiknya saya beritahu Anda... silakan minum wiski..."*

*MT: Dia telah menatap dengan putus asa pada Perdana Menteri sejenak, **lalu** berkata, "Baiklah, duduk, duduk, saya lebih baik mengisi Anda di… Nikmati wiski…"*

In this case, the temporal conjunction was used to further advance the narrative, as it signifies the subsequent action that Cornelius Fudge does. Given the vast amount of variety in signifying temporal conjunction, it is too little wonder that the three texts, despite having different items of conjunction, all carries the same function and meaning.

3.2.6. Lexical Cohesion

An example of lexical cohesion in the data can be found in the hypernymy text below.

*ST: Even the **weather** was dismal; all this **chilly mist** in the middle of July…It wasn't right, it wasn't normal…*

*HT: Bahkan **cuaca** pun suram; banyak **kabut dingin** di tengah bulan Juli... ini tidak benar, ini tidak normal...*

*MT: Bahkan **cuacanya** suram; semua **kabut dingin** di tengah Juli… Itu tidak benar, itu tidak normal…*

The use of hypernym can be identified by the presence of the word "weather" and "chilly mist," where weather is a superordinate of "chilly mist" (i.e., chilly mist is a type of weather.) Those superordinate and subordinates carry the same general idea, in which they both refer to the climate that was told by the author.

3.2.7. Differences in Cohesive Devices Between Human and Machine translation Compared to the Source Text

In this section, the differences between the three texts are presented. The contrastive analysis itself consists of two parts, which includes the form of transitional/shift analysis and machine translation error analysis. In cases where no transition/shift between cohesion were observed, the machine translation error is discussed based on their respective types in the transitional/shift analysis section instead.

3.3. Transitional Shift Analysis

3.3.1. Reference

An example of shifting in reference can be observed in the following lines, where a case of reference is omitted due to the word selection of human translation.
*ST: "Er," said the Prime Minister, "listen…It's not a very good time for me…I'm waiting for a telephone call, **you see**…from the President of —"*
*HT: "Er," kata Perdana Menteri, "ini bukan saat yang cocok untuk saya... saya sedang menunggu telepon, **soalnya**... dari presiden ne—"*

The word "you see" is replaced by "soalnya" which carries similar function yet differs in terms of cohesive devices. In both cases, it shows the hesitation of the Prime Minister to contact Fudge, as he himself was waiting for a call from the President of a distant country. A literal translation of the same line can be seen in the machine translation.

*HT: "Er," kata Perdana Menteri, "ini bukan saat yang cocok untuk saya... saya sedang menunggu telepon, **soalnya**... dari presiden ne—"*

*MT: "Er," kata Perdana Menteri, "dengar… Ini bukan waktu yang tepat untuk saya… Saya menunggu panggilan telepon, **Anda tahu**… dari Presiden -"*

The datum shows that the machine translation follows the source text word-by-word and thus retains the cohesive element of reference in the form of "Anda tahu." Anda in this case refers to the man in the painting whose appearance was narrated in the very previous line in the novel.

*ST: "To the Prime Minister of Muggles. Urgent we meet. Kindly respond immediately. Sincerely, Fudge." **The man in the painting** looked inquiringly at the Prime Minister*

*HT: "Kepada Perdana Menteri Muggle. Perlu sekali kita bertemu. Mohon segera ditanggapi. Salam, Fudge." **Pria dalam lukisan memandang** Perdana Menteri dengan ingin tahu.*

*MT: "Untuk Perdana Menteri Muggle. Mendesak kita bertemu. Mohon tanggapi segera. Hormat kami, Berbuat curang."**Pria dalam lukisan itu** menatap Perdana Menteri dengan penuh rasa ingin tahu*

As it happens, the datum also contains a slight difference in reference between the two translated texts, as the machine translation contains the reference signifier "itu" to refer to the

man in the painting. No visible effects in meaning were observed during the change though, as the two text carries the same identical meaning, with or without the reference signifier.

3.3.2. Substitution

In the data acquired during the research, the shift in substitution largely consists of the decision of human translation to substitutes the substitutive cohesive devices of the source text. There are several considerations behind the changes, namely the fact that several substitutional words in the source text may be unfamiliar to the target language readers, as shown in the following dialogues.

*ST: Before the Prime Minister could ask why he was dripping all over the **Axminster**, Fudge had started ranting about a prison the Prime Minister had never heard of, a man named "Serious" Black, something that sounded like "Hogwarts," and a boy called Harry Potter, none of which made the remotest sense to the Prime Minister.*

*HT: Sebelum Perdana Menteri sempat bertanya kenapa air menetes netes dari pakaiannya membasahi **karpetnya**, Fudge sudah mulai berteriak-teriak tentang penjara yang belum pernah didengar Perdana Menteri, seorang laki laki bernama "Serius" Black, sesuatu yang kedengarannya seperti Hogwarts dan seorang anak laki-laki bernama Harry Potter, tak satu pun masuk akal bagi Perdana Menteri.*

In this case, the word "Axminster," a form of substitution to the Prime Minister's carpet is translated into "karpetnya," which functions both as a possessive reference (as the word "nya" is to imply Prime Minister's rug) and repetition, as the word "carpet" has been used prior in the story. This is particularly interesting where "Axminster" itself, in real life, is a manufacturer and the branding of a certain luxurious carpet in the United Kingdom. According to Encyclopaedia Britannica, Axminster carpets are made from hand knotted wool, with the very first batches made in Axminster, Devon, England in 1755. Given the particularity of the carpets, it is relatively unlikely that an average Indonesians would have known or even heard of such fine mats. Hence, the translator opted to replace the peculiar substitute word into a more generalized, yet not much different, form of cohesion in the means of repetition and reference.

Given the ineptitude of machine translation to consider cultural, social, and historical aspects of a translation elements, it is not surprising that the form "Axminster" is retained in the machine translation.

*HT: "Sebelum Perdana Menteri sempat bertanya kenapa air menetes netes dari pakaiannya membasahi **karpetnya**, Fudge sudah mulai berteriak-teriak tentang penjara yang belum pernah didengar Perdana Menteri, seorang laki laki bernama "Serius" Black, sesuatu yang kedengarannya seperti Hogwarts dan seorang anak laki-laki bernama Harry Potter, tak satu pun masuk akal bagi Perdana Menteri."*

*MT: "Sebelum Perdana Menteri bisa bertanya mengapa dia menetes di seluruh **Axminster**, Fudge mulai mengomel tentang penjara Perdana Menteri belum pernah mendengar tentang, seorang pria bernama "Serius" Black, sesuatu yang terdengar seperti itu"Hogwarts," dan*

seorang anak laki-laki bernama Harry Potter, tidak ada satupun yang masuk akal bagi Prime Menteri."

Grammar mistakes aside, the use of Axminster in this sentence can be determined to be correct in terms of cohesion standpoint. However, the fact that many Indonesians are unfamiliar to the luxurious rug in question leads to the translation being potentially problematic, namely being unable to digest the information provided by the substitutional word due to lack of background knowledge.

3.3.3. Ellipsis

Due to the fact that there are only two instances of ellipsis in the three texts, compounded by the lack of shifting in the sample itself, an analysis of shifting in ellipsis from the sample data cannot be conducted.

3.3.4. Conjunction

There is no shift in conjunction between the source text and the human translation. However, a total of two mistakes attributed to machine translation ineptitude were found, with one sample being an omission-type error and the other being mistranslation.

Example of the omission-type error can be observed in the following narration.

*ST: And was it his fault that one of his Junior Ministers, Herbert Chorley, had chosen this week to **act so peculiarly <u>that</u> he was now** going to be spending a lot more time with his family?*

*HT: Dan salahnyakah jika salah satu menteri mudanya, Herbert Chorley, telah memilih minggu itu untuk **bersikap begitu ganjil <u>sehingga</u> sekarang dia** akan melewatkan jauh lebih banyak waktu bersama keluarganya?*

*MT: Dan apakah itu miliknya kesalahan bahwa salah satu Menteri Muda, Herbert Chorley, telah memilih minggu ini untuk **bertindak begitu aneh dia sekarang** akan menghabiskan lebih banyak waktu dengan keluarganya?*

The reason why the text can be classified as an omission-error is due to the missing, underlined **<u>sehingga</u>** between the word "begitu aneh" and "dia sekarang" of the machine translation compared to the human translation. The word **<u>that</u>** and **<u>sehingga</u>** itself functions as what is called an adjectival clause, which in itself functions as a subordinate clause. Henceforth, the incorrect omission (or the lack) of conjunction in this case can be classified as an omission error.

3.3.5. Lexical Cohesion

A shift from lexical cohesion of the source text into different types of cohesive devices in the target text were not found during the course of the research. However, a particularly

noteworthy case of the opposite can be found in the following dialogues, where a baffling translation mistake occurred.

ST: "To the Prime Minister of Muggles. Urgent we meet. Kindly respond immediately. Sincerely, ***Fudge****." The man in the painting looked inquiringly at the Prime Minister*

HT: "Kepada Perdana Menteri Muggle. Perlu sekali kita bertemu. Mohon segera ditanggapi. Salam, ***Fudge****." Pria dalam lukisan memandang Perdana Menteri dengan ingin tahu.*

MT: "Untuk Perdana Menteri Muggle. Mendesak kita bertemu. Mohon tanggapi segera. Hormat kami, ***Berbuat curang****."Pria dalam lukisan itu menatap Perdana Menteri dengan penuh rasa ingin tahu*

While machine translation (or in this case Google Translate) is well known for its reliability issues, a mistake of this kind is dumbfounding. The datum, a reiteration in the form of a repetition of a character's name, has been accurately translated in other occasions, as can be seen below.

ST: "I…oh…very well," said the Prime Minister weakly. "Yes, I'll see ***Fudge****."*

HT: "Saya... oh... baiklah," kata Perdana Menteri lemah. "Ya, saya akan menemui ***Fudge****."*

MT: "Saya… oh… sangat baik," kata Perdana Menteri dengan lemah. Ya, saya akan melihat ***Fudge****."*

Minor translation errors aside, it can be seen that Cornelius Fudge's name is translated properly in the previous datum. A quick search of "Fudge" in the Merriam-Webster website indeed lists the definition as "a foolish nonsense – often used interjectionally to express annoyance, disappointment, or disbelief" and "a soft creamy candy made typically of sugar," which may have been the case behind the mistranslation.

While it can be guessed where the machine translation may have obtained its definition of "Fudge," (a feat that is barely equivalent anyway) the fact that other instances of "Fudge" is translated properly still leaves room for disbelief. One possible explanation is that there has been an error or a glitch occurred during this very specific part of the translation and thus, an error was made.

3.4. Machine Translation Error Anaysis

3.4.1. Addition

As the name implies, addition refers to the case where additional cohesive items were found in the machine translation, in which no part of it were found or exist in both original text and human translation. The item in question is classified as addition when the text within the text itself has no cohesive ties and such cannot be made coherent by standing on its own.

An interesting case of this concept is found in the first line of the book, in which machine translation algorithm made a blatantly obvious mistake where it produces a reference signifier without a reference point.

*ST: It was nearing midnight and the Prime Minister was sitting alone in **his** office, reading a long memo that was slipping through **his** brain without leaving the slightest trace of meaning behind.*
*HT: Saat itu menjelang tengah malam dan perdana Menteri sedang duduk sendirian di **kantornya**, membaca laporan panjang yang lewat begitu saja melalui **otaknya** tanpa meninggalkan makna sedikit pun*
*MT: Saat itu hampir tengah malam dan Perdana Menteri sedang duduk sendirian di **kantornya**, membaca panjang lebar memo yang menyelinap di **otaknya** tanpa meninggalkan sedikit pun makna di balik**nya***

The bolded and underlined "nya" in the aftmost part of the sentence has no direct reference in regards to the text itself, as the other "nya" functions as a possessive pronoun of the "Perdana Menteri." Hence, the word "nya" in "dibaliknya" found in the machine translation can be considered as addition by the machine translation algorithm, as the cohesive element itself cannot be found in both the original English text or the human translation. The addition itself is deemed as a moot point.

3.4.2. Omission

A case of omission error can be found in the following narration line, where machine translation was not able to distinguish the different language style of English and Bahasa Indonesia, resulted in incorrect omission of reference point.

*ST: **It was coming** — as the Prime Minister had known at the first cough — from the froglike little man wearing a long silver wig who was depicted in a small, dirty oil painting in the far corner of the room.*

*HT: **Suara itu datangnya**— seperti yang telah diketahui Perdana Menteri waktu mendengar batuk yang pertama kali— dari pria kecil bertampang kodok memakai wig perak panjang yang tergambar dalam lukisan cat minyak kecil kotor di sudut ruangan yang jauh.*

*MT: **Itu datang** - seperti yang diketahui Perdana Menteri pada batuk pertama – dari laki-laki kecil mirip katak yang mengenakan wig perak panjang yang digambarkan dalam lukisan cat minyak kecil dan kotor disudut jauh ruangan.*

It can be observed that the addition of the word "Suara" in the human-translated text. In the source and machine translation, the word "Suara" is nowhere to be found. This is caused by different language typology of Bahasa Indonesia and English, where it is acceptable (and is coherent) to refer a specific item in English without a head word. However, as (evidently in this case), machine translation could not pick up the difference in language typology, it kept the original English typology into Bahasa Indonesia transliteration, resulting in what is called by omission.

3.4.3. Misordering

Misordering is pretty straight forward; a case where the influence of first language structure is being brought upon the second language, resulting in incorrect order of phrases within the sentence. There are 7 recorded cases of misordering collected during the research.

An example of misordering which involves fundamental structural mistake, where a literal translation made by machine translation produces grammatical error can be found in the datum below.

*ST: He remembered it as though it were yesterday and knew it would haunt him until **his dying day.***

*HT: Dia ingat jelas peristiwa itu, seakan baru terjadi kemarin, dan tahu itu akan menghantuinya **sampai hari kematiannya***

*MT: Dia mengingatnya seolah-olah itu baru kemarin dan tahu itu akan menghantuinya sampai **miliknya hari kematian**.*

The grammatical mistake can be attributed to the difference in sentence structure between English and Bahasa Indonesia, where possessive pronouns are constructed in an arranged fashion of (Pron + Noun) in English and (Noun + Nya) in Bahasa Indonesia. What machine translation had made is an example where machine translation has not been able to consider the different sentence structure between languages and thus, an error of such form occurred. Although the word "hari kematian miliknya" is also incorrect (i.e., unnatural) in terms of meaning, the sentence would at least be correct in terms of structure.

3.4.4. Misinformation

An error of misinformation can be observed in the following datum, where machine translation mistakenly kept the form and structure of source text into the target language.

*ST: The Prime Minister rather resented being told to sit down in his own office, let alone offered **his own whiskey**, but he sat nevertheless.*

*HT: Perdana Menteri agak sebal disuruh duduk di kantornya sendiri, apalagi ditawari **wiskinya sendiri**, namun dia toh duduk juga*

*MT: Perdana Menteri agak kesal disuruh duduk di kantornya sendiri, apalagi menawarkan **miliknya memiliki wiski**, tapi dia tetap duduk.*

The possible explanation behind this error lies on the incompetence of machine translation to identify collocation in phrases. Due to the fact that the word "his" and "own" fulfills the purpose of a possessive marker, the machine translation algorithm may have picked the sentences as having double the amount of the marker. Added to the fact that the same kind of error in the omission datum is also present in this datum, the machine translation produces the phrase "memilikinya memiliki wiski" which made no grammatical sense whatsoever yet can be traced by the word "his own whiskey" that was literally translated.

A case of severe misinformation that directly affects a cohesive device can be found in the following lines where a conjunction is translated incorrectly.

ST: It was precisely this sort of behavior that made him dislike Fudge's visits so much. He was, ***after all****, the Prime Minister and did not appreciate being made to feel like an ignorant schoolboy*

HT: Persis sikap seperti inilah yang membuatnya sangat membenci kunjungan Fudge. ***Bagaimanapun juga*** *dia Perdana Menteri dan tak suka disudutkan sampai merasa seperti murid yang tak tahu apa-apa.*

MT: Tingkah laku semacam inilah yang membuatnya begitu tidak menyukai kunjungan Fudge. Dia, ***setelah semuanya****, Perdana Menteri dan tidak menghargai dibuat merasa seperti anak sekolah yang bodoh.*

It can be observed that the word "setelah semuanya" was supposed to replace the word "after all." The problem lies in the concept of conjunction where a very specific phrases or term is required to indicate two clauses being connected by a conjunctive phrase. The word "setelah semuanya" is not a proper conjunctive phrase and thus is deemed as a misinformation in the form of mistranslation. It is suspected that direct, literal translation is the probable cause behind the word "setelah semuanya" as "setelah" and "semuanya" is a one-to-one equivalent to "after" and "all" respectively

4 Conclusion

This research finds that the realization of cohesive devices in Harry Potter and the Half Blood Prince can be attributed to references being the most frequently used device. This is due to the fact that the source text, a hybrid text consisting of narration and dialogues, uses a considerable number of references, which was used to tie the entities in the story. The realization of cohesive devices in HT is largely similar to the ST. However, the realization between HT and MT is noticeably different. This is exacerbated by the text types of Half Blood Prince which consists of narratives and dialogues. The machine translation often fails to consider the different types of the text, resulting in incorrect use of cohesive devices.

The research also discovers significant difference in cohesive devices between the translated texts and the source text. The differences can be attributed to shifting of cohesive devices between the ST and HT and errors of machine translation in the case of ST and MT translations. A comparison between HT and MT solidifies the fact that there are errors in the process of translation by the machine translation. Most of the errors can be attributed to the incompetence of machine translation to detect certain aspect of cohesive devices, such as sociocultural aspect of the language and language typology.

References

Printed Books

[1.] Halliday, M. A., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

[2.] Kuncahya, A. (2015). *Cohesion in Narrative Texts Presented in the Electronic Textbook of Senior High School Grade X Entitled "Developing English Competence".* Yogyakarta: Yogyakarta State University.

[3.] Kunz, K. (2015). *English-German Contrasts in Cohesion and Implications for Translation.* GECCo.

[4.] Mastropierro, L., & Mahlberg, M. (2017). *Key Words and Translated Cohesion in Lovecraft's At the Mountains of Madness and One of its Italian Translations.* University of Birmingham.
[5.] Mellinger, C. D., & Hanson, T. A. (2017). *Quantitative Research Methods in Translation and Interpreting Studies.* Oxon: Routledge.
[6.] Munday, J. (2001). *Introducing Translation Studies: Theories and Applications.* New York: Routledge.
[7.] Newmark, P. (1988). *A textbook of translation.* New York: Prentice Hall International.
[8.] Ninsiana, W. (2014). *Grammatical Cohesion Devices on the Indonesian Translation of English Bidding Document.* Semarang: Universitas Sebelas Maret.
[9.] Nord, C. (2018). *Translating as a Purposeful Activity. Functionalist Approaches Explained. Second Edition.* London: Routledge.
[10.] Snell-Hornby, M. (1995). *Translation Studies: An Integrated Approach.* Amsterdam/Philadelphia: John Benjamins.
[11.] Steiner, E. (2004). *Translated Texts: Properties, Variants, Evaluations.* Peter Lang.

Electronic Sources

[12.] Halliday, M., & Matthiessen, C. M. (2013). Halliday's Introduction to Fuctional Grammar. Oxon: Routledge. doi:https://doi.org/10.4324/9780203431269
[13.] Korstjens, I., & Moser, A. (2018). Part 4: Trustworthiness and publishing. 120-124. doi:https://doi.org/10.1080/13814788.2017.1375092
[14.] Maharani, W. (2018). Kohesi dan Koherensi Teks Bacaan dalam Buku Bahasa Indonesia Kelas XII Semester 2. Jakarta: Badan Pengembangan dan Pembinaan Bahasa. Retrieved from http://kbi.kemdikbud.go.id/kbi_back/file/dokumen_makalah/dokumen_makalah_1540524188.pdf
[15.] Toury, G. (2012). Descriptive Translation Studies and beyond: Revised Edition. doi:https://doi.org/10.1075/btl.100

Journal & Reports

[16.] Arvianti, Gilang. (2018). Human Translation versus Machine Translation of Instagram's Captions. Magelang: Universitas Tidar.
[17.] Rudi, H. (2010). Cohesion in the Headlines of The Jakarta Post Online Issued in November – December 2010. Yogyakarta: Universitas Negeri Yogyakarta.
[18.] Tversky, A. (1977). Features of similarity. Psychological Review 84, 327–352.
[19.] Vasquez, L., & Alberto, D. (2008). Error Analysis in a Written Composition. Envigado: Institution Universitaria de Envigado.
[20.] Yuliasri, Issy., & Pamela Allen. (2019). Humour Loss in the Indonesian Translation of Harry Potter and the Sorcerer's *Stone*. Bandung: Indonesian Journal of Applied Linguistics.