

Machine Translation of Idiomatic Expressions and Slang Based on Korean Drama

Sungran Koh
(Kookmin University)

Koh, Sungran (2022). Machine translation of idiomatic expressions and slang based on Korean drama. *The Linguistic Association of Korea Journal*, 32(3), 45-63.

Machine translation has advanced significantly over the years. However, current machine translation models have limitations in handling various translation patterns. The most notable examples are idioms and slangs. Identifying idiomatic expressions and slang is frequently the most challenging aspect when acquiring proficiency in a new language. Within the domain of machine translation, idiomatic expressions have not received significant attention mainly due to their inherent complexity. This study seeks to improve the machine translation of idiomatic expressions and slang across languages and to discover a more effective approach for enhancing the quality of Korean-to-English translation using ChatGPT for EFL learners. This article examines the differences between human interpretation (HI) and machine translation (MT) using Papago and ChatGPT regarding Korean-English translation of idiom and slang and discuss advantages and limitations of machine translation. The data are collected from the script Korean Drama *The Glory* (2022, Kim). Any errors in ChatGPT's Korean-to-English translations were corrected by providing additional prompts to achieve the best possible translations. The results demonstrated that the overall translation quality of idiomatic expressions and slang was significantly improved by these prompts and showed increased accuracy as well. It is suggested that a systematic framework for automated interpretation of idiomatic expressions and slang be introduced into the machine translation system. Additionally, to effectively translate them, it is necessary to provide appropriate prompts along with their meanings using ChatGPT.

Key Words: idiomatic expressions, slang, ChatGPT, translation, Korean-to-English

1. Introduction

The impact of globalization and the advancement of technology has changed a variety of aspects of people's daily routines. Among them, the ability to convey information in multiple languages has become a crucial aspect of effective communication. This requires high levels of speed and accuracy in languages translation. Automated translation systems have become important for enabling communication between languages in everyday activities. Modern machine translation is quite intelligently offering fascinating insights into the nature of language and how to achieve excellent language comprehension. Machine Translation, also referred to as automated language translation, use computer programs to transform sentences from one language into another without any human participation. Evaluation of translation products typically considers factors such as speed, cost, and accuracy. Although machine translation can ensure speed and cost-effectiveness, achieving accuracy remains a challenging aspect that necessitates human intervention. The success of accurate machine translation lies in certain small-sized expressions that are already listed in its extensive database.

However, much research has claimed that machine translation has significant errors in conveying its meaning and implication (Hsu, 2014; Vilar et al., 2006). Mohamed and Shafeen (2017) state, "Machine translation poses certain challenges since exact translation of one language into another makes up a complicated problem. An attempt has been made to identify these machine translation challenges (p. 55)." Machine translation is only a time-saving tool rather than a substitute for human involvement. To achieve high-quality translations, human post-editing is necessary. Human translation aims to ensure the natural flow of the translation in the target language even though it requires deviating from the source language to achieve a more effective expression. On the contrary, machine translation focuses on analyzing the provided data rather than understanding the context, and it doesn't provide various plausible solutions based on the given situations. Machine translation provides only basic options, which prove to be simplistic when dealing with the intricate linguistic complexities in most translation.

The errors in machine translation can be attributed to certain reasons. One factor of these errors is the inability of machine translation to comprehend context. Wisniewski, Kübler and Yvon (2014) assert that since languages heavily depend on contextual understanding, machine translation is not able to provide correct meaning if users do not offer the full context of the text. In addition, machine translation has difficulties in

dealing with the translation of infrequent multi-word expressions. Especially, phrases like idioms pose a significant challenge due to their figurative meaning, which cannot be deduced from the individual meanings of their components. (Baziotis et al., 2023). Machine translation models often render these expressions literally, translating them word by word not considering the context meaning, resulting in inaccurate translations. Current Machine Translation systems face a specific challenge with idiomatic expressions since their translation is often not literal but rather logical. The lack of precision in machine translations results from depending on dictionaries and adhering to the "situational limitations" established by the programmer. The programmer assigns fixed equivalents to individual words based on their preferences, informed by their language and equivalent knowledge, and referencing the latest bilingual dictionaries (Abdulhaq, 2017).

John Hutchins (2010) asserts that machine translation, particularly in handling colloquial language found in various source texts like social media, is characterized by low quality. Slang is one of the most context-dependent aspects in translating colloquial language, representing a prevalent form of informal communication that extensively utilizes words in a flexible manner. Human translation has an intuitive ability to comprehend the meaning of slang and informal language, especially in their native tongue. However, in case of machine translation, the accurate understanding of the intended meaning of specific terms becomes even more challenging. This poses as a significant obstacle for a machine translation system's interpretation of certain language. Current machine translation methods for understanding slang typically depend on context but overlook the frequent semantic extensions inherent in slang word usage.

Human translation is focused on ensuring a natural tone in the target language even though it takes a lot of time and effort compared to machine translation. Human translation has several pivotal factors, including the delivery of meaning between source and target languages, the impact of cultural context, and the distinct abilities of individual translators (Li & Cai, 2014). On the contrary, machine translation relies on analyzing the limited given data rather than understanding the context, and it does not provide various and plausible solutions for the problem based on different situations. To figure out the difficulties in both human translation and machine translation, the fields of Artificial Intelligence study has rapidly developed. With the emergence of advanced Artificial Intelligence (AI) models such as ChatGPT's translation, the process of achieving accurate language translation has become more efficient and immediate. ChatGPT, the recently introduced chatbot based on the Generative Pre-training Transformer (GPT),

represents a powerful pre-trained language model developed by OpenAI. Although ChatGPT wasn't explicitly designed for translation, it is quickly gaining recognition as a reliable tool for language translation. ChatGPT translation utilizes deep learning techniques to conduct language translation across various languages. It is one of the systems that has demonstrated outstanding proficiency not only in handling the translation of multiple languages, but also in translating a single language. Its ability to comprehend and generate human-like responses across diverse contexts serves as proof of its excellent performance.

Due to ChatGPT's human-like interactive nature and intelligent conversational system, the translation it produces adapts to appropriate responses and modifications based on the given prompt and context. Enhanced with Reinforcement Learning from Human Feedback (RLHF), ChatGPT has gathered widespread user engagement owed to its remarkable proficiency in natural language understanding and generation. Originally designed as an intelligent conversational system, ChatGPT shows the ability to perform a variety of human-like tasks, such as composing poems. Jiao et al. (2023) highlights that when provided with simplistic prompts, ChatGPT shows a noticeable performance gap compared to other translation tool like Google Translate and DeepL Translate. Unlike machine translation tools such as Google Translate, which offer consistent translations irrespective of attempts, ChatGPT strives to grasp contextual subtleties with the goal of the most precise translation. Additionally, the ability of ChatGPT to provide human-like communication enhances its value as a resource to give accurate and reliable translations.

Yehoshua (1953) observes that machines struggle to offer satisfactory translations for idioms. This difficulty arises because machines adhere to specific rules and dictionaries, and he suggests resolving the issue by updating the existing rules and incorporating new ones. Basil and Jeremy (2004) state that it is more realistic to try to develop machines that collaborate with humans. To address this challenge, the goal of this study is to suggest that human translators work in conjunction with machine translators such as ChatGPT, especially when dealing with idiom and slang so complex that human intervention is necessary to fully grasp their contextual and cultural nuances. This study offers an examination of idioms and slang in machine translation, aiming to support and pave the way for future research in this area. To do this, as a first step, the human translation script was collected of the Korean streaming television series *The Glory* (2022) that was released on 2022 (part one) and 2023(part two). As a next step, Korean idioms and slang are identified and translated using machine translation tools like Papago and

ChatGPT. Subsequently, the translations are compared and analyzed. Next, incorrect machine translations are grouped to improve the precision of ChatGPT's Korean-to-English translations. These were later retranslated by providing additional prompts to ChatGPT in coordination with human input. Finally, the output translations were reanalyzed to achieve the highest level of accuracy in its translations.

2. Literature Review

2.1. The limitation of Google Translate and ChatGP from a Cultural and Contextual Perspective

Idioms and slang pose a considerable challenge for human translators in translating a source language into the target language. It is much harder for machine translation system. Human translators can often intuitively understand and interpret the meaning of idiom and slang, especially in their native language. However, when dealing with machine translation, translating the precise meaning of specific terminology becomes even more challenging. While the idiom dictionary provides a meaning, the translator still needs to select from various options or align the idiom with the context. Consequently, machine translation system fails to identify and translates them as straightforward text, resulting in errors or causing errors.

Idioms refer to multiword expressions with a fixed, often metaphorical meaning that cannot be deduced from their individual components. They represent one of the most used forms of non-literal language. Volk (1999) states that nearly every idiom has the potential to be used in its non-idiomatic, literal sense and then nearly any idiom has the possibility to introduce ambiguity. Idioms are intricate multiword expressions, and machine translation systems generally favor simpler compositional expressions that can be processed sequentially. Consequently, machine translation faces difficulties when translating infrequent multiword expressions. Among multiword expressions, idioms (e.g., "piece of cake") pose a significant challenge due to their figurative nature, and their meaning cannot be inferred from the meanings of individual words (Liu, 2017). Machine translation often translate these expressions literally, i.e., word by word, resulting in inaccurate translations. Volk (1999) argues these systems should ideally be equipped with idiom recognition, if not full idiom translation capabilities. If the system struggles to

distinguish whether a potential idiom should be interpreted literally or figuratively, this decision could be left to the user.

Idiomatic expressions originate from a nation's way of life, values, and historical events. For instance, the Korean proverb "닭 쫓던 개 지붕 쳐다본다" directly describes a specific situation in Korean society, but applying this to another culture might lead to misunderstanding the context. In Korea, the image of a dog and a chicken is a familiar everyday scene, but in other cultures, this relationship may not be as intuitive or common.

Slang refers to a prevalent form of informal language used by certain groups or subcultures to convey specific concepts, making words in a flexible and extended manner. The use of slang words and expressions can differ significantly across cultures, countries, and regions (Aguilar, 2023). The examples are in the following.

“Barbie”: While you may know this word as the name of the popular doll produced by toy manufacturer Mattel, it is also slang for the word “barbecue” in Australia. So don’t panic if you hear an Australian say: “Let’s fire up the barbie” - they’re likely talking about lighting a barbecue grill rather than setting a doll on fire!

“I’m dead”: A popular slang phrase among Generation Z individuals where the speaker means they find something so funny that they’re dying from laughter. (Aguilar, 2023)

Aguilar(2023) mentions that the frequent use of slang in spoken or written communication can lead to the integration of that slang into the language of the culture. In some cases, the slang might become so widely adopted that it is officially recognized with a formal definition in the dictionary.

As mentioned above, slang is often shaped by specific groups or generations and can evolve rapidly, making it difficult for machine translation to keep up. Without context, slang can be inaccurately interpreted or lead to misunderstandings. Waner (2022) mentions that machine translation systems should use a strategy based on context to enhance the precision of translations containing slang.

For example, Korean slang includes expressions like ‘꿀잼’ which means ‘very fun’ and is frequently used among younger generations; however, interpreting it literally as "honey" and "fun" does not convey the correct meaning. Therefore, understanding the cultural background and linguistic differences is crucial for accurately translating slang.

This prompt approach uses the surrounding context of a particular word to make sure its possible meaning, paraphrasing the input and translating the paraphrase. A semantically informed approach, one that attempts to establish a connection between the conventional meaning and the slang meaning, could potentially be a more effective method for translating specific slang expressions.

3. Methodology

The first stage in this study is to collect the human translation script of the Korean streaming television series *The Glory* (2022) aired on Netflix. The reason for this selection is because it centers on a South Korean seeking absolute revenge against her childhood bullies, who ruined her life, in a psychological thriller and consequently included significant Korean colloquial languages such as idiom and slang. It is also acknowledged as a distinguished piece, having received eight nominations at the 59th Baeksang Arts Awards and winning in three categories: Best Drama, Best Actress awarded to Song Hye-kyo, and Best Supporting Actress awarded to Lim Ji-yeon.

As a next step, Korean idioms and slang from *The Glory* (2022) are classified and translated using two widely-used translation systems in Korea, Papago and ChatGPT. Papago, developed by the prominent IT company Naver Corporation in Korea, is a multilingual machine translation system that utilizes AI-based neural machine learning, continually learning from its errors. Afterward, the translated outputs are examined and evaluated. Subsequently, inaccurate machine translations are categorized to enhance the accuracy of ChatGPT's Korean-to-English translations. To refine the translations, additional prompts are presented to ChatGPT in collaboration with human input, and the resulting translations are reevaluated, striving for the utmost precision in the translated content. AI models, such as ChatGPT, are helpful for translating various languages, but idioms and slang can be interpreted very differently depending on the language and culture, resulting in less accurate translations. Especially, Korean idioms and slang often have nuances or contexts that AI might not fully grasp. Therefore, having a human translator first explain the meaning in a prompt and then asking the AI for a translation can help provide a more accurate result. For example, the Korean idiom "닭 쫓던 개 지붕 쳐다본다" is directly translated or misinterpreted when ChatGPT is asked to handle it alone, such as 'A dog chasing a chicken ends up staring at the roof'. However, when a human

translator explains that this idiom refers to ‘a situation where someone has failed to achieve a goal and is left in confusion,’ ChatGPT can offer a translation that better fits the context such as ‘Like a dog who missed its chance, left confused and helpless’. By following this approach, where the human translator clarifies cultural or linguistic nuances by asking ChatGPT for a translation, it is possible to achieve a more natural and context-appropriate outcome.

4. Result and Discussion

4.1. Korean Idiomatic Expressions

Lee (2022) categorized the idiomatic expressions in Korean as a foreign language into three items: Idiom, Semantic collocation, and proverb while Kim (2014) grouped them into two items such as idiom and proverb. Along with Kim (2014)’s line, this study categorizes the idiomatic expressions into two groups: Idioms and proverbs. Idioms combine two or more words to form an entirely new unit of meaning and it can be considered the most representative type of idiomatic expression because without explicit explanation, it is impossible to guess its meaning. Proverbs have unique characteristics such as satirical or instructive qualities that are exclusive to proverbs alone. The idiomatic expressions of *The Glory* (2022) are in the following.

Table 1. The Categorization of Korean Idiomatic Expression *The Glory* (2022)

Idiomatic Expression	Korean	Human Translation
Idioms	나랑같이 칼춤춰줄 망나니가 필요하거든요	I need an executioner with me, who can help me in the hunt.
	그렇게 정성껏 엇을 맥인다고?	Why would he try to screw you over?
Proverbs	뭐 식은죽 먹기죠	Well, it was all very easy
	하나는 알고 둘은 몰라 어떡하지	You’re not the sharpest tool in the shed, are you?

4.1.1. Korean Idioms

The subcategorization idioms of Korean idiomatic expression in *The Glory* (2022) are translated by machine translation, Papago and Chatgpt. In Korean ‘칼춤을 추다’ means

'To skillfully wield a sword' and '나랑같이 칼춤춰줄 망나니가 필요하거든요' is translated as 'I need an executioner with me, who can help me in the hunt' by the human translator. However, Papago translates it as 'I need a guy who can dance with me' whereas Chatgpt translates it as 'I need a rogue who will dance the knife dance with me'. Both machine translation does not recognize the meaning of the context meaning. Chatgpt just translates '칼춤' as the meaning of the each word 'dance the knife' as shown in Table 2.

Table 2. The Machine Translation of Korean Idioms *The Glory* (2022)

Korean	Papago	ChatGPT
Korean Idioms 나랑같이 칼춤춰줄 망나니가 필요하거든요	I need a guy who can dance with me	I need a rogue who will dance the knife dance with me.
그렇게 정성껏 엿을 멕인다고?	You're going to do that with all your heart?	Do you think I'm going to swallow your nonsense so easily?

Korean idiomatic expression '엿먹이다' means 'deviously manipulate or deceive (someone) sneakily'. In Korean sentence '그렇게 정성껏 엿을 멕인다고?' is translated into 'Why would he try to screw you over?' by human translator. However, it is translated into 'You're going to do that with all your heart?' by Papago whereas 'Do you think I'm going to swallow your nonsense so easily?' by ChatGPT in the wrong way. In this case, the collaboration between human translators and ChatGPT, AI-based systems is needed. The Korean idiom meaning as additional prompt is added in ChatGPT such as '엿을 멕이다 means deviously manipulate or deceive(someone) sneakily'. Then human translator asks the same sentence to retranslate. The result is close to the appropriate meaning such as 'Do you think I'm deviously manipulating you like that?'. By using additional prompts as the form of the conversation, ChatGPT shows better translation result.

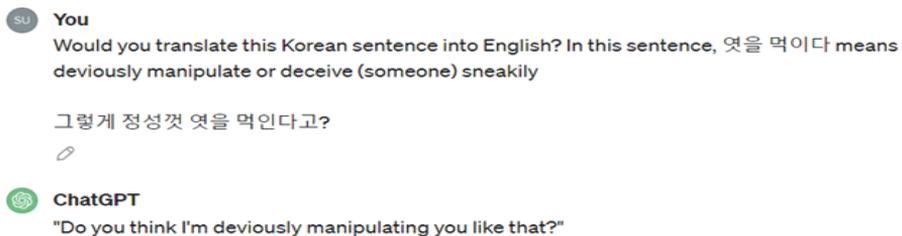


Figure 1. Idiom Information Prompt

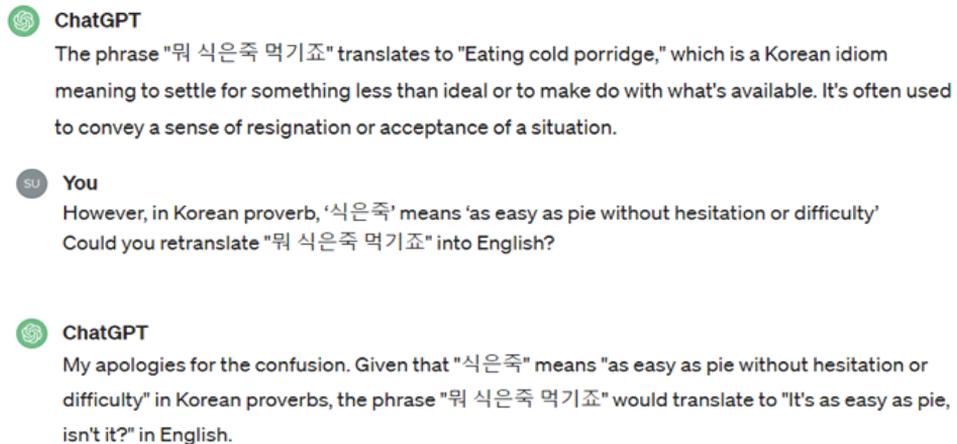
4.1.2. Korean Proverbs

The proverbs, the subset of Korean idiomatic expression are translated by machine translation, Papago and ChatGPT. In Korean proverb, '식은죽' means 'as easy as pie without hesitation or difficulty' 'like taking candy from a baby'. '뭐 식은죽 먹기죠' is translated into 'Well, it was all very easy' by human translator. It is translated into 'It's a piece of cake'(Papago) by Papago whereas 'Eating cold porridge' by ChatGPT in an inappropriate way (See Table 3).

Table 3. The Machine Translation of Korean Proverbs *The Glory (2022)*

Korean Proverb	Papago	ChatGPT
뭐 식은죽 먹기죠	It's a piece of cake	Eating cold porridge
하나는 알고 둘은 몰라 어떡하지	I know one, but I don't know what to do	Knowing one and not knowing the other, what should I do?

Then, the Korean proverb meaning prompt is input into ChatGPT before asking it to retranslate. As shown in Figure 2, ChatGPT shows better result by using prompt like a human conversation.



ChatGPT
The phrase "뭐 식은죽 먹기죠" translates to "Eating cold porridge," which is a Korean idiom meaning to settle for something less than ideal or to make do with what's available. It's often used to convey a sense of resignation or acceptance of a situation.

You
However, in Korean proverb, '식은죽' means 'as easy as pie without hesitation or difficulty'. Could you retranslate "뭐 식은죽 먹기죠" into English?

ChatGPT
My apologies for the confusion. Given that "식은죽" means "as easy as pie without hesitation or difficulty" in Korean proverbs, the phrase "뭐 식은죽 먹기죠" would translate to "It's as easy as pie, isn't it?" in English.

Figure 2. Korean Proverb Information Prompt 1

Korean proverb '하나는 알고 둘은 모른다' means 'not being able to see all sides of things and being narrow-minded, lacking flexibility and being stubborn'. '하나는 알고 둘

은 몰라 어떡하지' is translated into 'You're not the sharpest tool in the shed, are you?' by human translator. However, it is translated into 'I know one, but I don't know what to do?' by Papago whereas 'I know one thing, but I don't know the other. What should I do?' in English' by ChatGPT. Both machine translations do not recognize the context meaning and it is translated in the literal meaning of the sentence. After adding the Korean proverb as an additional prompt in ChatGPT and asking for the sentence to be retranslated, the result closely matches the intended meaning, 'Knowing one aspect but ignorant of the other. What should I do?' Using additional prompts in the form of conversation seems to yield better translation results.

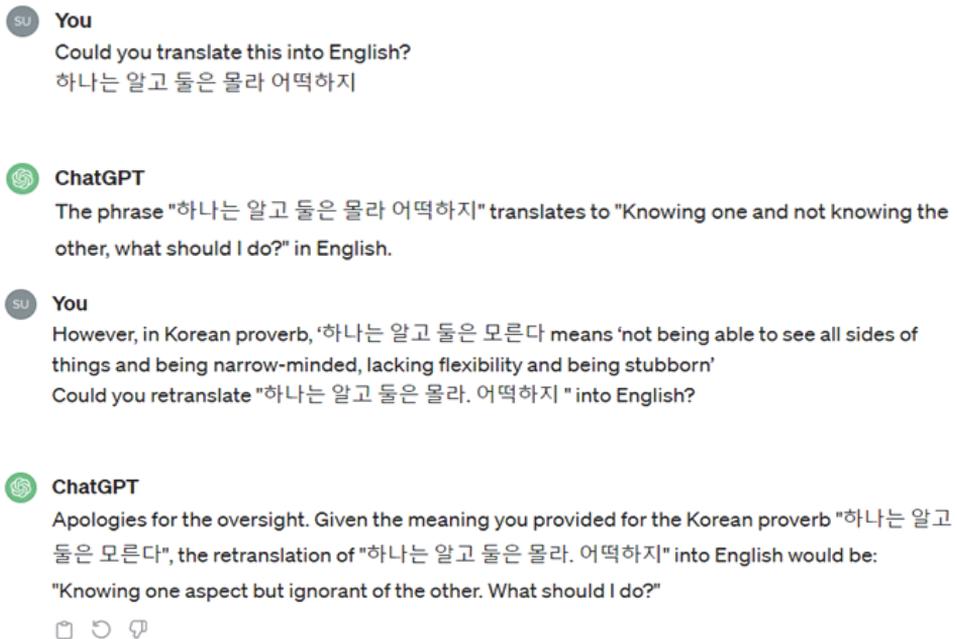


Figure 3. Korean Proverb Information Prompt 2

4.2. Korean Slangs

In this section, some items with higher error rates are analyzed through a closer examination: incorrect disambiguation (subject), wrong lexical choice, and missing words. Some detail points are described in the following. Translating slang plays a crucial role in content localization, which involves tailoring content to align with the culture and

context of the intended audience. This process can include incorporating cultural idioms and slang specific to the target audience to ensure the content feels genuinely local (Aguilar, 2023). KakaoTalk is a Korean messenger tool that uses the function "1" next to a text to signify that the receiver has not read the message. Only when the receiver reads the message, the "1" disappears from next to the text. The Korean '얼른 읽어라, 1 지워라' means 'Read my text' but both machine translation translated it to 'Erase 1' and 'delete page 1'. Since machine translation do not recognize Korean messenger tool, it is not translated in an appropriate way.

Slang represents a prevalent type of informal language characterized by its flexible and extensive utilization of words, posing a notable challenge for natural language processing (NLP) systems in terms of interpretation. Humans frequently possess an innate ability to understand and interpret the significance of slang and informal language, particularly in their native language. Interestingly, many individuals employ slang expressions without fully comprehending their meanings. However, in machine translation, accurately determining the intended meaning of slang becomes even more challenging. This study categorized the Korean slang from *The Glory* (2022) and the result shows in the followings (See Table 2).

Table 4. The List of Korean Slang *The Glory* (2022)

	Korean	English	Human Translation	Frequency
1	씨발	ssibal	Fuck this!	41
2	년	nyeon	bitch	22
3	새끼	saekki	bastard	18
4	개새끼	gaesaekki	Son of a Bitch	15
5	쌍년	ssangnyeon	bitch	11
6	존나	jonna	so freaking	11
7	좆	jot	fuck	5
8	지랄	jiral	bullshit	5
9	돼지다	dwejida	dead	2
10	빡치다	ppakchida	super pissed	2
11	대박	daebak	awesome	1
12	양아치	yang-achi agari	real bad guy	1
13	아가리	agari	mouth	1
14	고딩	goding	kids	1
15	씹다	ssipda	ignore	1
16	꼴값	kkolgap	play around	1
17	꼰대	kkondae	ruder	1
18	눈깔	nunkkal	eyes	1

The most common slang is '씨발(ssibal)' and the next is '년(nyeon)' and then '새끼(saekki)'. 씨발(ssibal) is commonly used to express one's dissatisfaction or anger. 년(nyeon) is term used to derogatorily refer to a woman whereas 개새끼(gaesaeikki) is a derogatory term used to refer to a man who engages in despicable or dirty actions and is considered unpleasant or undesirable. '새끼(saekki) is a term used affectionately to refer to a young person. It is commonly use when Koreans have conversation informally, but it is regarded as inappropriate to use in public or in front of the elderly.

'대박(daebak)' is a term used to express something as 'awesome' or 'amazing'. It is often used not only when you're dumbfounded by a situation, but when you say '대박 나세요' meaning 'I wish you great success' to wish someone good luck. It's commonly used when you're astonished by a situation, and it's frequently exchanged between friends, as well as appearing in advertisements. It is a noun and can be paired with various verbs. When combined with '이다 (to be),' 대박 + 이다' means 'to be awesome,' while 대박 + 나다 (to come out)' signifies 'to become successful' or 'to be a hit (product). Sometimes '대박' can be used as an adverb, as in '대박 멋있다(really cool)'. The usages are as follows.

Table 5. Conjugations and Common Usages of 대박

Noun/Adjective Form	Exclamation
대박	대박
Adjective + "to be"	Noun + Verb
대박이다	대박(이) 나다
대박이야 (Present tense)	대박이 터지다
대박이었어 (Past tense)	

'빡치다 (bachida)' refers to 'feeling irritated or angry'. There are primarily two explanations for how this term came to hold these meanings. One theory suggests that '빡치다' originated from the expression 'to have sexual intercourse' and evolved into another profanity related to sex. Another theory proposes that it stemmed from '빡,' which denotes someone's forehead or head, and the action of hitting one's head on a desk, brick wall, or another object in frustration. Nevertheless, '빡치다(bachida)' is used to mean 'very angry' among young Koreans.

In '꼴값 (kkolgap)', '꼴' means 'shape' or 'face', and it's usually used in conjunction

with another word rather than independently. '값' means 'price', so '꼴값' literally translates to 'the price of one's face'. Typically paired with the verbs '하다 (to do)' or '떨다 (to shake)', '꼴값' is used to express that someone is an 'eyesore' or that a person is acting or being 'ridiculous'. As this is an insulting expression, it is used in informal language, such as '꼴값떠네 (making a fool of oneself)' or '꼴값 떨고 있네(acting ridiculous)' (Lee, 2023). The slangs are translated by two machine translation, Papago and ChatGPT.

Table 6. Korean Slang by Machine Translation (*The Glory, 2002*)

	Korean	English	Papago	ChatGPT
1	씨발	ssibal	fuck	fuck
2	년	nyeon	year	year
3	새끼	saekki	young	“child,” “kid,” or “young one
4	개새끼	gaesaekki	Son of a Bitch	son of a bitch. It's a derogatory term in Korean.
5	쌍년	ssangnyeon	a double year	“bitch” or “double bitch”
6	존나	jonna	Damn	“fucking” or “really” in English. It is vulgar and often used to emphasize something in a strong or offensive way. For example, “존나 좋아” would translate to “fucking great” or “really great.”
7	좆	jot	jot	“dick” or “penis” in English. It is highly offensive and considered very rude.
8	지랄	jiral	Flip out	The Korean term “지랄” (jiral) is a slang term used to describe someone acting crazy, throwing a tantrum, or causing a scene. In English, it can be translated to “bullshit,” “crazy,” or “freaking out.” It is a derogatory term often used to express frustration or disdain towards someone's behavior.
9	뉘지다	dwejida	croak	It appears to be a misspelling or an uncommon word.
10	빡치다	ppakchida	get mad	to be annoyed” or “to be pissed off”
11	대박	daebak	jackpot	“awesome,” “amazing,” “fantastic,” or “incredible”
12	양아치	yang-achi	Gangster	“hoodlum,” “thug,” or “delinquent”
13	아가리	agari	trap, hole, gob	mouth or maw,
14	고딩	goding	High School Student	high school student
15	씹다	SSIPDA	Chew	to chew or to bite
16	꼴값	kkolgap	last value	being made a fool of or being taken for a ride.
17	꼰대	kkondae	an old-fashioned person	old-fashioned, stick in the mud, or fuddy-duddy.
18	눈깔	nunkkal	peeper	eyes or gaze

'년(nyeon)' in Korean has two meanings such as 'year' and 'bitch' in slang, which is translated into 'year' in both Papago and ChatGPT. The cause of these errors is that machine translation systems are not connected to dictionaries specifically designed to provide ambiguous multiple meanings for such expressions. One possible way to solve machine translation with these expressions is to feed appropriate prompts along with their meanings. It is retranslated in ChatGPT by adding a prompt and the result shows correct translation. It is explained as a very offensive word and the same word as 'bitch' in English.(See Figure 4)

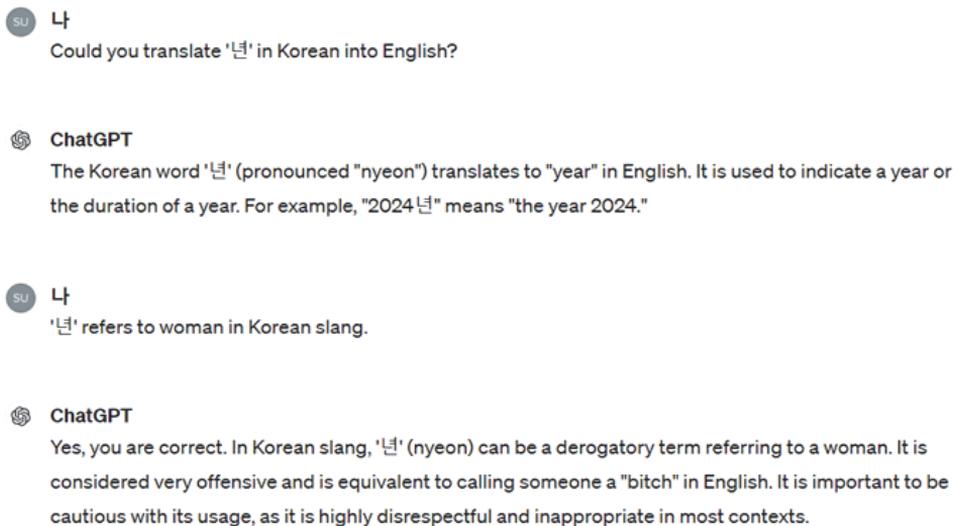


Figure 4. *Yeon* Information Prompt

'년(nyeon)' can be used with some kinds of words such as '미친년(crazy bitch)', '나쁜년(that bitch)', '씨발년(you fucking bitch)', '거머리 같은년(fucking leech)', '쓰레기 같은년(You piece of garbage)', '그년(the bitch)', '이년(The bitch)', '독오른년(that bitch)', '어리기만 한년(young bitch)'. The sentences included '년(nyeon)' are translated both Papago and ChatGPT. The finding shows that ChatGPT has better translation capability than Papago. Papago continues to translate '년(nyeon)' as 'year' in the following. In addition, ChatGPT includes some additional explanations such as "Please note that "씨발년" is a highly offensive term in Korean, equivalent to "bitch" in English, and should be used with caution.'

Table 7. The Sentence List of Korean Slang *Yeon* (*The Glory*, 2022)

Korean	Human Translation	Papago	ChatGPT
1 씨발 그래야 쌍년 안되고 난년이 되는 거야	That's how you become a successful bitch.	That's how it's going to be two years and two	That way, you won't become a two-faced [explicit] and instead become an old [derogatory term for a woman].
2 아 시랄년이 갈데가 있어서	This arrogant bitch has somewhere to be.	Oh, there's a place to go in the year of birth	"The bitch has somewhere to go." Please note that "시랄년" is a highly offensive term in Korean, equivalent to "bitch" in English, and should be used with caution.
3 이런 거머리 같은 년	You goddam leech of a woman!	You're a leech	Such a foolish woman" or "This idiotic woman.
4 너같이 나쁜 년도 이렇게 잘만 사는데	Even an evil bitch like you is living so well	I'm living in a bad year like you	Even a bad woman like you can live so well.
5 쌍년이	Fucking bitch	a double-headed person	"two-faced woman" or "deceitful woman
6 얼다대고 질이야 빈년이	Don't try to teach me. You fucking feather brain	You're so mean. You're so mean	"Who do you think you are to give advice, you fucking empty-headed bitch." The text in Korean is quite offensive.

In Korean, '존나' '쫏' do not translate in ChatGPT, but '존잼' is translated into "super fun" or "really fun." The Korean expression "존잼" is an abbreviation of "존나 재미있어" (jonna jaemitseo), which translates to "super fun" or "really enjoyable" in English. It is a casual and colloquial way of expressing that something is very entertaining or amusing.

5. Conclusion

Machine translation lacks fundamental cultural and contextual understanding of certain idiomatic expressions. Machine translation gives literal meanings to certain idiomatic expressions, but it does not mean that they are all wrong since a reader and a translator can tell that the literal translation hints at something beyond the literal meaning. Shao et al. (2018) proposed a technique for calculating the occurrence of machine translation errors, which requires the development of custom-made lists specific to each language (known as blocklists) containing terms that represent literal translation

mistakes. When idioms are translated, machine translator could be a mean as a sort of dictionary of idioms. The dictionary of idioms would give a meaning but still the translator would have to choose from among different options or to match the idiom with the context.

As seen before, '년(nyeon)' in Korean which has two meanings such as 'year' and 'bitch' in slang was translated into 'year' in both Papago and ChatGPT, not 'bitch' as was meant in the context. These expressions do not initially have ambiguous multiple meanings, which is why the exclusion of text-bound expressions is seen as another limitation of machine translation. Two suggestions can be taken from this research. First, this issue can be addressed by inputting these fixed expressions depending on the context along with their meanings and translations into the machine translation system. It might benefit machine translation to include more than one dictionary to cover all the expressions, the basic expressions and the complicated ones in the machine translation systems. Furthermore, the cause of the errors is that machine translation systems are not connected to dictionaries specifically designed to provide ambiguous multiple meanings for such expressions. Secondly, thus, it is suggested that one possible way to solve machine translation with idiomatic expressions and slangs is to feed appropriate prompts along with their meanings by using ChatGPT. Certain prompts are designed to enhance the accuracy of translations produced by ChatGPT. These prompts include sentences with common idiomatic expressions and slang. Users can assist ChatGPT in translating these phrases, ensuring the use of more natural English.

Integrating multiple dictionaries is a promising approach to overcoming the limitations of machine translation. However, in practice, implementing this approach faces technical challenges such as data consistency, context recognition, and update management. One practical alternative to integrating multiple dictionaries is the hybrid approach. This method improves translation quality through collaboration between AI-based translation systems and human translators. After the machine translation system references multiple dictionaries and provides a translation, human translators review and edit it. This process helps to overcome the limitations of machine translation.

To address these challenges, various complementary solutions, such as hybrid approaches, utilizing user feedback, and developing domain-specific systems, need to be applied. To enhance the accuracy and efficiency of machine translation systems, it is essential to combine technological advancements with strategies that deeply reflect cultural and linguistic understanding.

References

- Abdulhaq, S. (2017). Machine translation: The cultural and idiomatic challenge, *Journal of Al-Azhar University*, 19(2), 1-28. <https://digitalcommons.aaru.edu.jo/al-azhar/vol19/iss2/9>
- Aguilar, R. (2023). A guide to translating slang into other languages. Weglot. <https://www.weglot.com/blog/translate-slang>
- Basil, H., & Jeremy, M. (2004), *Translation an advanced resource book*, London and New York: Routledge
- Baziotis, C., & Mathur, P., & Hasle, E. (2023). Automatic evaluation and analysis of idioms, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 3682-3700, <https://aclanthology.org/2023.eacl-main.267>
- Hsu, J. A. (2014). Error classification of machine translation a corpus-based study on Chinese-English patent translation. *Translation Studies*, 18(4), 121-136.
- Jiao, W., Wang, W., Huang, J., Wang, X., & Tu. Z. (2023). Is Chatgpt a good translator? a preliminary study. https://www.researchgate.net/publication/367359399_Is_ChatGPT_A_Good_Translator_A_Preliminary_Study
- Kim S.-Y. (2014). Study on pragmatic use of Korean idiomatic expressions with a focus on variation tendencies, *Journal of the International Network for Korean Language and Culture*, 11(3), 23-46
- Lee, C. (2022). Suggestion for the phraseological teaching and learning of Korean idiomatic expressions. *Bilingual Research*, 89, 169-192.
- Lee, S.-Y. (2023). *Korean Slang Expressions*, Longtail Books.
- Li, H., Graesser, A. C., & Cai, Z. (2014). Comparison of Google translation with human translation. In W. Eberle & C. Boonthum-Denecke (Eds.), *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference* (pp. 21-23). Florida, USA.
- Liu, D. (2017). *Idioms: Description, comprehension, acquisition, and pedagogy*. Routledge.
- Mohamed Z. H., & Shafeen, M. N. (2017). A brief study of challenges in machine translation. *IJCSI International Journal of Computer Science Issues*, 14(2), 54-57.
- Vilar, D., Xu, J., d'Haro, L. F., & Ney, H. (2006). Error analysis of statistical machine translation output. *Proceedings of the 5th international Conference of Language*

Resources and Evaluation (LREC). pp. 697-702. Genoa, Italy.

- Volk, M. (1999). The automatic translation of idioms: Machine translation vs. translation memory systems. *Computer Science Linguistics*. DOI: 10.5167/UZH-19070
- Wisniewski, G., & Kbler, N., Yvon, F. (2014). A corpus of machine translation errors extracted from translation students' exercises. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 3585-3588. http://www.lrec-conf.org/proceedings/lrec2014/pdf/1115_Paper.pdf
- Yehoshua, B. (1953). Some linguistic problems connected with machine translation. Retrieved from <http://www.mt-archive.info/Bar-Hillel-1953.pdf> on 18 November 2015 (p.221-223)

Koh Sungran

Assistant Professor

Kookmin University

77, Jeongneung-ro, Seongbuk-gu, Seoul, Republic of Korea

Email: sungran@kookmin.ac.kr

Received on August 7, 2024

Revised version received on September 22, 2024

Accepted on September 30, 2024