

Cultural Pragmatics and Causal Connectives: A Contrastive Study of Korean and English Using the AI-Hub Parallel Corpus

Sujeong Choi¹ and Sin-hye Nam^{2,*}

¹ KDI School of Public Policy and Management,
263 Namsejong-ro, Sejong-si, 30149, Republic of Korea
schoi@kdischool.ac.kr

² Kyung Hee University,
26 Kyungheedaero, Dongdaemun-gu, Seoul, 02447, Republic of Korea
namsh@khu.ac.kr

Abstract. This interdisciplinary study analyzes Korean and English causal connective expressions using the AI-Hub Korean-English parallel corpus. The primary objective is to identify the unique linguistic and cultural features of Korean causal connectives by comparing them with their English counterparts. Korean includes a wide range of causal connectives, many of which exhibit additional pragmatic features such as [+negative], [+uncertainty], and [+plurality]. From both linguistic and cultural perspectives, this study investigates whether these features are exclusive to Korean and explores the cultural factors contributing to their prevalence.

To extend the analysis into a computational framework, the study defines a formal task for evaluating the preservation of pragmatic meaning in translation. Specifically, each Korean-English sentence pair is annotated for the pragmatic features expressed in Korean, and the extent to which those features are retained in the English translation is assessed. Two task formulations are proposed: (1) a binary classification indicating full preservation vs. loss or shift, and (2) a continuous “pragmatic shift score” ranging from 0.0 to 1.0. This enables future implementation of rule-based or learning-based models to detect pragmatic mismatches in translation.

The analysis of seven Korean causal connectives reveals that the additional pragmatic features are specific to Korean and rarely appear in English. These features are culturally grounded: the [+negative] feature aligns with Korean speakers’ tendency to express disapproval indirectly to preserve politeness; [+uncertainty] reflects a cultural preference for hedging and softening assertions; and [+plurality] indicates an avoidance of definitive statements, consistent with indirect and euphemistic communication strategies common in Korean discourse.

Ethical considerations regarding data licensing and cultural bias are addressed. This research offers practical implications for computational linguistics, translation studies, and Korean language education. By uncovering culturally embedded differences in how causality is expressed, the study enhances cross-cultural understanding and contributes to improved communication in multilingual contexts.

Keywords: Parallel Corpus, Contrastive Study, Causal Connective Expressions, Interpretive Ethno-grammar

* Corresponding author

1. Introduction

1.1. Linguistic and Cultural Motivation

Computer science, particularly in the field of natural language processing (NLP), has significantly contributed to the development and analysis of linguistic corpora. Technological advances have enabled efficient collection, organization, and large-scale processing of corpora, facilitating various applications in language research and education. Among different types of corpora, the parallel corpus, which consists of aligned translations in two or more languages, offers especially rich resources for contrastive linguistic analysis. By providing a one-to-one mapping of corresponding expressions across languages, parallel corpora are instrumental in examining syntactic, semantic, and pragmatic differences.

In the context of Korean-English comparison, one particularly fruitful area of inquiry involves causal connective expressions, which are more diverse and nuanced in Korean than in many other languages. This richness is not merely grammatical but reflects culturally embedded communicative preferences. For instance, the Korean language often employs a variety of causal expressions to convey subtle shades of intention, obligation, and emotion. Such phenomena are similar to the lexical diversity observed in the Hanunoo language of the Philippines regarding rice types (Conklin, 1957), or the range of friendship-related terms in Russian that contrast with the single English word "friend" (Wierzbicka, 1997). These cases exemplify how cultural values shape language and influence not only vocabulary but also the structure and use of expressions.

Understanding this connection between language and culture is essential for effective cross-cultural communication. As Saville Troike (2003) notes, language proficiency includes not only mastery of phonology, vocabulary, and grammar but also the ability to communicate appropriately in various social and cultural contexts. In this light, the present study seeks to explore how causal meaning is expressed and translated across Korean and English, aiming to shed light on underlying cultural differences that shape linguistic expression.

1.2. Research Objectives and Scope

This study conducts a contrastive linguistic analysis of Korean and English causal connective expressions using a large-scale Korean-English parallel corpus. The primary goal is to identify how specific Korean causal connectives are translated into English and to evaluate whether the pragmatic meanings embedded in the Korean expressions are preserved, altered, or lost during translation.

By doing so, the study aims to uncover both linguistic and cultural asymmetries between the two languages. This contrastive analysis has practical implications for second language education, particularly for English-speaking learners of Korean. A better understanding of how Korean causal connectives function—both linguistically and culturally—can help learners avoid misinterpretations and improve pragmatic competence. Ultimately, the findings aim to support cross-cultural communication by enhancing awareness of how causality and intention are linguistically encoded in different language communities.

1.3. Overview of the Computational Task

To extend the linguistic analysis into a computational framework, this study proposes a formal task that quantitatively evaluates the preservation of pragmatic meaning in Korean-English translations of causal connectives. Using the AI-Hub Korean-English parallel corpus, the task involves sentence pairs in which a Korean sentence containing a specific causal connective (e.g., *Ast-(eun) tase*, *Vst-neun tase*, *Ast/Vst-(eun) nameoji*, *Vst-neurago*, *Vst-neun barame*, *Vst-neun tonge*, *Ast/Vst-a/eoseo geureonji*, *Ast/Vst-go haeseo*) is aligned with its English translation. Each Korean sentence is annotated for pragmatic features such as [+negative], [+uncertainty], or [+plurality].

The objective is to determine the extent to which these pragmatic features are retained in the English translation. Two task formulations are proposed: (1) a binary classification, where a label of 1 indicates full preservation of meaning and 0 indicates loss or change, and (2) a continuous scoring system, where a “pragmatic shift score” ranges from 0.0 (fully preserved) to 1.0 (completely lost or altered). This task definition paves the way for developing rule-based or machine learning models capable of detecting pragmatic mismatches in translation.

2. Related Works

2.1. Parallel Corpus and Linguistic Research

A parallel corpus is a corpus consisting of pairs of texts in two or more languages. Since the proposal of the first parallel corpus alignment system by Kay and Röscheisen (1988), it has been widely utilized in linguistics and NLP (Natural Language Processing) fields. The application areas of parallel corpora include machine translation, Cross-language Information Retrieval (CLIR), lexicography, language education, contrastive linguistics, and more. Over the years, various studies using parallel corpora have been conducted in Korean language and culture research. Recently, with the global spread of Korean cultural content, there has been a growing interest in Korean language education and the development of automatic translation technologies, leading to an increasing demand for language resources.

In response to this demand, institutions like the National Institute of Korean Language and AI-Hub, which are focusing on Korean language data construction, are actively involved in building Korean-foreign parallel corpora. Research on parallel corpora can be broadly categorized into studies in the field of Natural Language Processing (NLP) and studies in the fields of linguistics and applied linguistics. In the NLP field, discussions mainly revolve around the construction of parallel corpora. For instance, Vu et al. (2020) reported on the development of the Ulsan Parallel Corpora (UPC), which includes Korean-English and Korean-Vietnamese datasets. Park et al. (2022) reported on the construction of seven types of parallel corpora, including Korean, by AI Hub, and evaluated their performance. Additionally, Bang et al. (2022) constructed the English-Korean speech parallel corpus (EnKoST-C) and presented the evaluation results of its performance.

On the other hand, in the fields of linguistics and applied linguistics, the majority of research reports are based on the utilization of parallel corpora. Studies conducted using parallel corpora, with Korean as the focus, mostly involve Korean-Chinese (Sim, 2015; Li 2021; Yu 2022; Shim, 2023), Korean-English (Seo, 2008; Park, 2017; Park & Lim,

2020), and Korean-Japanese (O & Takiguchi, 2015; Kim & Jang, 2011) corpora. Particularly, parallel corpora have been utilized for contrastive linguistic studies, and a few specific examples are as follows: Seo (2008) described the English equivalents of the Korean sentence-ending suffix *'-get-'* using a Korean-English parallel corpus; Yu (2002) used a Korean-Chinese parallel corpus to describe the Chinese counterparts of the Korean expression *'-(eu)leo.'*; Xue (2021) conducted a contrastive analysis of honorific expressions based on a Korean-Chinese parallel corpus; Wilczynski (2021) conducted a contrastive study on the *'ida'* (to be) construction based on a trilingual Korean-Polish-English parallel corpus.

Although we have examined examples of such research, the parallel corpora traditionally utilized by linguists have had limitations in terms of their size and balance, which in turn might have imposed certain constraints on the research outcomes. However, as noted earlier, due to the recent demand in the fields of artificial intelligence and natural language processing, high-quality Korean-foreign language parallel corpora have been rapidly constructed on a large scale. This allows for the utilization of these new language resources to conduct contrastive linguistic studies of even higher quality than previous research. Especially, this is true for Korean-English parallel corpora. Therefore, in this study, our aim is to take advantage of the latest parallel corpora, which offer high precision, large scale, and balance, driven by the achievements in the field of NLP, to conduct applied linguistic research with improved precision.

2.2. Korean Causal Connective Expressions

According to Choi (2022), Korean has the highest number of grammatical connective expressions used to indicate causation. Choi (2022) conducted an analysis of grammatical items from ten Korean textbooks and three grammar references, organizing them according to their meanings. This study identified a total of 68 pragmatic categories, with the [cause] category exhibiting the most synonymous expressions. Specifically, 24 synonymous grammar items that express the concept of [cause] in Korean were identified. Although these expressions share a common function of indicating the cause for the subsequent clause, they differ in terms of syntactic conditions, contextual formality, and pragmatic nuances. For instance, the following causal grammar items from Korean textbooks are noted for their additional pragmatic features compared to the more neutral causal items.

- (1) *Ast-(eun) tase, Vst-neun tase*: This expression conveys a negative cause or reason.
- (2) *Ast/Vst-(eun) nameoji*: This is used when an action or situation in the first clause deteriorates, leading to negative consequences in the subsequent clause.
- (3) *Vst-neurago*: This indicates a cause, reason, or purpose that results in a negative outcome.
- (4) *Vst-neun baram*: This expression denotes a cause or reason where the preceding context negatively impacts the subsequent action.
- (5) *Vst-neun tonge*: This indicates a cause or reason that leads to a negative situation or results in the following clause.
- (6) *Ast/Vst-a/eoseo geureonji*: This is used when a preceding action or situation appears to be a cause but remains uncertain.

(7) *Ast/Vst-go haeseo*: This suggests that the preceding clause represents one of several reasons for what follows.

As underlined above, the causal connective expressions in examples (1) to (7) feature additional pragmatic attributes: the grammatical items in (1) to (5) include a [+negative] feature, the item in (6) denotes a [+uncertainty] feature, and the item in (7) implies a [+plurality] aspect. Thus, Korean causal connective expressions not only indicate the cause for the subsequent clause but also convey additional pragmatic nuances and complexities. As we have seen above, there are a variety of grammar items in Korean that express cause. Due to the diversity of connectives expressing cause, there has been a lot of research on the similarities and differences between Korean causal connectives expressions (Chin, 2005; Ahn, 2007; Park, 2008; Li, 2011; Yoo, 2015; Park, 2018; Jeon, 2021, etc.) However, there is a lack of comparative linguistic research on whether these diversities and characteristics are unique to Korean or whether they are also present in other languages. If these features are unique to Korean, it would be possible to explore their causes from a linguistic and cultural perspective, and in terms of language education, it would be a particularly important topic to focus on. To explore this, this study uses a parallel Korean-English corpus to compare seven causal connectives that contain pragmatic qualities additional to the meaning of cause with their counterparts in English. Through the analysis of these findings, this study aims to interpret the reasons for the substantial number of causal connective expressions in Korean from a cultural standpoint, and to explore the pedagogical implications of these differences in the context of language education.

3. Data and Methodology

The data for this study were sourced from the Korean-English parallel corpora created by *AI-Hub*, an AI integration platform managed by Korea's Ministry of Science and ICT and the National Information Society Agency. These corpora were produced as part of a project aimed at constructing data for artificial intelligence training.

3.1. Corpus Overview and Genre Distribution

The *AI-Hub* Korean-English corpus includes 1.6 million sentence pairs across three styles: literary, colloquial, and conversational. From this dataset, we selected a subset of 300,000 sentence pairs, distributed as follows.

In total, 300,000 sentences (3,438,086 words) from the Korean-English parallel corpora were analyzed to investigate the English expressions that correspond to Korean reason/causal connective expressions.

3.2. Preprocessing and Data Filtering

The data were processed using a hybrid approach that combined automatic filtering and manual validation. First, automatic filtering was conducted based on keyword matching to identify sentences containing one or more of seven pre-defined Korean causal connective expressions. Following this, manual validation was performed to remove false

Table 1. Information of Korean-English Parallel Corpora

No.	Style	Context	The Number of Sentences/Words	
1	Literary	News Articles	200,000 sentences (2,658,545 words)	300,000 sentences (3,438,086 words)
2	Colloquial (Conversation)	Conversation in a meeting, shopping, school, restaurant, etc.	100,000 sentences (779,541 words)	

positives and ensure contextual appropriateness. The overall preprocessing procedure included sentence tokenization, normalization, filtering based on the presence of the target expressions, and manual verification of the extracted sentence pairs. Figure 1 provides an illustrative example of this process. A total of 479 sentences containing the seven target Korean causal connective expressions were extracted from the Korean-English parallel corpus. For each sentence pair, the Korean causal connective was identified (as shown in Column A), and its corresponding English expression was extracted (Column B). Column C displays the original Korean sentence, and Column D presents the corresponding English translation. For example, in row 310, the Korean connective expression *Vst-neun baram* corresponds to the English connective because. While some expressions showed a direct causal correspondence, others were translated with non-causal expressions. Based on this analysis, each English expression was manually categorized as either a causal or non-causal expression.

	A	B	C	D	E
1	Korean	English	Korean Sentences	English Sentences	
309	<i>Vst-neun baram</i>	and	발에 안보내지마는 '무용'기여 할 '부' 있었지만, 기세 앞날날로 가는 위장하는 바람에 이틀이 지난 2일 아침에야 현장에 도착했다.	when the help from Korean embassy in Indonesia, but due to issues with the aircraft, it had to go back and she was able to arrive at the site in the morning of 2nd after 2 days.	국제, 아시아
310	<i>Vst-neun baram</i>	because	후반 8분 수비 역파스트 킥 골키퍼가 공을 제대로 처리하지 못하는 바람에 코로바키치의 선제 역파스트(25 코로바키치(후반)의 첫골을 내주고, 후반 35분엔 후카 오모리치(33회)의 미드필드에서 중거리포를 허용했다.	At the 8th minute of the second half, they gave out the first goal to Croatia's Ante Rebić (25. Frankfur because the goalkeeper could not handle the ball properly during the back pass, and they allowed the mid-range goal to the Luka Modrić (33. Real Madrid) at the 35th minute of the second half.	스포츠, 축구, 한국프로축구
311	<i>Vst-neun baram</i>	and	몇 차례 역전 기회를 살리지 못한 양측은 결국 후반 41분 상대 수를 끌어내려한 공이 양측의 골대에 맞고 골문 안으로 떨어지는 바람에 골을 내주고 말았다.	41 minutes into the second half of the game, Korea, which was not able to make use of the several opportunities to flip the game, had the ball that I'm Seon-joo was trying to pull away from the opponent's shooting hit his head and head into the goalposts, and gave away the final goal.	스포츠, 축구, 해외축구
312	<i>Vst-neun baram</i>	so	홍익표 국회의장을 비롯한 민주당 의원들은 오전 회의가 무산되자 이날 오후 다시 회의를 재개했지만 이 밤에도 한 국당 의원들이 단체로 출석하지 않는 바람에 만장일치가 무산됐다.	At the morning meeting was canceled. Members of the Democratic Party of Korea, including Hong Ikpyo, the chairman of the subcommittee, tried to resume the meeting again in the afternoon. But, this time again, 6 members of the Liberty Korea Party came as a group and protested. So, the passing of the bill was founded.	정치, 국회, 정당
313	<i>Vst-neun baram</i>	due to	검교보고 하는 그루주 집행위원장(홍남진전환자치 전남보 부장)이 '여론사건은 (여우시민이 아니라) 좌익 군인이 일으킨 반란'이라고 발언하는 바람에 소란이 일어났다.	During the report, Executive Committee Head Go Hyo-ju/Vietnam Veterans Association Jeonnam Branch Head, stated that the uproar occurred due to the statement: "the Yeosu incident was a Revolt caused by leftist soldiers (and not the citizens of Yeosu)".	지역, 전남
314	<i>Ast-(eun) tase, Vst-neun tase</i>	due to	제품이 제대로 검수를 하지 않은 탓에 하자 제품이 출고되었는데 고객님이 원하시면 환불 처리해드릴게요.	The defected good was delivered due to our lack of inspection, so if you want, we will refund it.	여행/쇼핑
315	<i>Ast-(eun) tase, Vst-neun tase</i>	because	국내 텔레미디가 금지된 탓에 해외로 눈을 돌렸다.	Because domestic telemedicine is banned, it has turned to foreign countries.	사회, 의료, 건강
316	<i>Ast-(eun) tase, Vst-neun tase</i>	because	이 같은 지장이 끊긴 탓에 노동계의 반발은 거세졌다.	Labor resistance has intensified because such support has been cut off.	사회, 노동, 복지
317	<i>Ast-(eun) tase, Vst-neun tase</i>	because	이미 취직자를 수용한 탓에 수출도 어려움이 크다.	Because it has already accepted the quota system, exportation is also very difficult.	경제, 산업, 기업
318	<i>Ast-(eun) tase, Vst-neun tase</i>	because of	처음부터 크게 지은 탓에 골짜기 곳곳에 있었다.	The vacant lots were everywhere because of the large construction from the beginning.	경제, 산업, 기업
319	<i>Ast-(eun) tase, Vst-neun tase</i>	due to	스타트업이 끊임없이 높은 회전율을 보인다는 점도 특징이다.	Another characteristic is that it has high turnover due to dense startups.	경제, 취업, 창업
320	<i>Ast-(eun) tase, Vst-neun tase</i>	because	가동률이 절반으로 떨어진 탓에 공장 전체에 활력이 떨어졌다.	Because of the halving in operating rate, the entire factory has lost vitality.	경제, 자동차
321	<i>Ast-(eun) tase, Vst-neun tase</i>	because	기업이 투자에 가려줄 돈 탓에 순자산 조달이 위축됐다.	net fund raising has been discouraged because companies have distanced themselves from investments.	경제, 산업, 기업
322	<i>Ast-(eun) tase, Vst-neun tase</i>	because			

Fig. 1. An Example of Data Analysis

This classification was conducted using a hybrid approach combining automatic keyword-based filtering and manual validation. The annotation was performed by two trained annotators, and the inter-annotator agreement, measured using Cohen's Kappa, was 0.87, indicating substantial consistency. Due to licensing restrictions, the preprocessing scripts

and annotation guidelines are not publicly available; however, they can be provided upon reasonable request for academic and research purposes. (see Section 3.7).

3.3. Train/Dev/Test Split

As this study is not aimed at model training but rather focuses on the qualitative and quantitative analysis of linguistic patterns, the dataset was not divided into separate training, development, or test subsets. Instead, a total of 479 sentence pairs containing Korean causal connectives were extracted and analyzed. This annotated subset can be made available for evaluation or further research upon reasonable request.

3.4. Annotation Guidelines and Procedure

The annotation process was carried out by two trained annotators with backgrounds in linguistics. Their task was to categorize the corresponding English expressions as either causal or non-causal. To ensure consistency, detailed annotation guidelines were developed and followed. These guidelines included definitions of causal and non-causal relations, illustrative examples, descriptions of edge cases, and clearly defined decision criteria for handling ambiguous expressions. Before annotating the main dataset, the annotators participated in a training session using a pilot dataset of 50 sentence pairs. They then independently annotated the full dataset, after which discrepancies were reviewed and resolved collaboratively, leading to refinements in the annotation protocol.

3.5. Inter-Annotator Agreement

To evaluate the reliability of the annotation, inter-annotator agreement was measured using Cohen's Kappa (κ). The resulting κ value was 0.87, which indicates a high level of agreement and suggests substantial consistency between the two annotators.

3.6. Data Analysis Procedure

The overall procedure for analyzing the Korean-English sentence pairs is illustrated in Figure 2. First, Korean sentences containing one or more of the seven target causal connective expressions were identified within the parallel corpus. The corresponding English sentences were then extracted, and the English expressions aligned with the Korean causal markers were manually examined. Based on pragmatic criteria, each English expression was categorized into either a causal expression group or a non-causal expression group. Expressions that conveyed explicit causal meaning were included in the causal group, while others that did not indicate causality were classified as non-causal.

Additionally, the study investigated the specific English equivalents of each Korean causal connective. For instance, in the case of the Korean expression ‘-(eun) tase’, 145 sentence pairs were retrieved from the corpus. Among the corresponding English sentences, some included causal expressions while others did not. These were divided into the two aforementioned categories and further analyzed to determine the frequency and proportion of each English counterpart. The results of this frequency analysis, including representative expressions and their distribution, are presented in Tables 2 through 4 in Section 4.

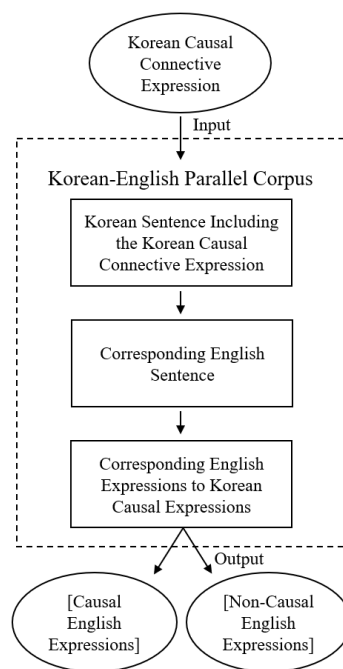


Fig. 2. Procedure of Data Analysis

3.7. Data Licensing and Access

The original corpus used in this study is distributed by *AI-Hub* under specific licensing terms that restrict public redistribution. Researchers who wish to access the original dataset may do so by applying through the *AI-Hub* website and agreeing to the platform’s usage policies. To support reproducibility and facilitate further research, supplementary materials—including preprocessing scripts, annotation guidelines, and a sample of annotated data with licensing-safe examples—can be made available upon request to qualified researchers for non-commercial academic purposes.

4. Contrastive Analysis on Causal Expressions in Korean and English

As described in Chapter 2, the Korean causal connective expressions *Ast-(eun) tase*, *Vst-neun tase*, *Ast/Vst-(eun) nameoji*, *Vst-neurago*, *Vst-neun barame*, *Vst-neun tonge* connote [+negative] meaning, the expression *Ast/Vst-a/eoseo geureonji* has [+uncertainty] meaning, and the expression *Ast/Vst-go haeseo* contains [+plurality] meaning additionally. In this chapter, the study investigates whether the additional pragmatic features of Korean causal connective expressions are reflected in their English counterparts, through a contrastive analysis of the results from a Korean-English parallel corpus. This section analyzes the degree of pragmatic preservation between Korean causal connectives and their

English translations. While the computational task is defined in Section 1.3 as a classification or scoring problem, this study implements a manual and qualitative evaluation, rather than applying an automated classifier. Each connective is examined to assess whether the pragmatic features are retained or lost in translation, providing a foundational dataset for future computational modeling.

4.1. [+Negative] feature in causal connective expressions

First, the English expressions corresponding to the Korean causal connective expressions *Ast-(eun) tase*, *Vst-neun tase*, *Ast/Vst-(eun) nameoji*, *Vst-neurago*, *Vst-neun barame*, *Vst-neun tonge*, which imply the [+negative] feature, were analyzed. Table 2 presents the frequency and proportion of these Korean causal connective expressions with the [+negative] feature in the Korean-English parallel corpus, along with their corresponding English expressions.

As shown in Table 2, in English no causal expressions exhibiting the [+negative] feature observed in Korean expressions. Instead, only basic or neutral causal expressions such as *because*, *due to*, *as*, *since* were used in English as their corresponding counterparts. The following (a) (e) in Figure 3 are the examples of Korean sentences containing the causal connective expressions with [+negative] feature and the corresponding English sentences.

In (a) (e) in Figure 3, the causal connective expressions used in the Korean sentences all indicate that the cause of the negative outcome in the second clause is in the first clause, while implying a negative attitude towards the cause described in the first clause. On the other hand, the corresponding English causal expressions are neutral, indicating that the speaker is conveying a neutral view of the causal event without a negative attitude. Rather, the analysis revealed a strong tendency in English to use direct negative vocabulary when the speaker intends to convey a negative meaning. In English, words with negative connotations are employed directly to express negative intentions. In Korean, however, grammatical expressions that imply negativity are often used in a more indirect manner. This can also be seen in the example that some of the sentences that use causal connective expressions that connote the [+negative] feature in Korean are not expressed by grammatical expressions of cause in English, but rather verbs with negative connotations. For example, in some cases, the speaker's negative attitude expressed through '*Vst-neun tonge*' in Korean was corresponded to by the negative verb '*confuse*' in English, and in other cases, the negative attitude expressed through '*Vst-neun barame*' in Korean was corresponded to by the negative verb '*disturb*' in English. This means that in Korean, even when speakers do not use direct vocabulary to convey negative intentions, negative nuances can still be communicated indirectly through grammatical expressions. Searle (1969, 1979) argued that the primary motivation for indirect speech is politeness. The discussion of politeness in linguistics was developed by Lakoff (1972) and culminated in Brown & Levinson (1987) with the concept of personal politeness strategy. Since individuals have a sense of face, and any act that threatens their desire for face tends to be viewed as a face-threatening act, speakers and listeners will adopt politeness communication strategies to avoid face-threatening acts as much as possible and minimize conflict. In Korean, the speaker's negative attitude toward the situation can be interpreted as an intention to increase politeness and minimize status threats by using causal connectives that have negative connotations rather than using direct negative vocabulary.

Table 2. Korean Causal Connective Expressions Containing the [+negative] feature and Corresponding English Expressions

No.	Korean	N	Category	English	N	Rate
1	Ast-(eun) tase, Vst-neun tase	145	Causal Expressions (N=138, R=95.1)	because (of)	57	39.31
				due to	32	22.07
				as	25	17.24
				since	14	9.66
				cause	4	2.76
				so	2	1.38
				by	1	0.69
				for	1	0.69
				therefore	1	0.69
				so ... that	1	0.69
				after	2	1.38
				and	1	0.69
				as a result	1	0.69
				result in	1	0.69
				lead	1	0.69
				while	1	0.69
			Non-causal Expressions (N=7, R=4.82)	because	5	20.83
				as	3	12.50
2	Ast/Vst-(eun) nameoji	24	Causal expressions (N=13, R=54.16)	because	5	20.83
				as	3	12.50
				so ... that	3	12.50
				for	2	8.33
			Non-causal expressions (N=11, R=45.84)	and	5	20.83
				∅	5	20.83
				by ing	1	4.17
3	Vst-neurago	13	Causal expressions (N=6, R=46.16)	because	4	30.77
				due to	1	7.69
				as	1	7.69
				ing	2	15.38
			Non-causal expressions (N=7, R=53.84)	with	1	7.69
				from	1	7.69
				and bring, on	1 2	7.69 15.38
4	Vst-neun baram	153	Causal expressions (N=122, R=79.74)	because	61	39.87
				so	24	15.69
				as	14	9.15
				since	11	7.19
				due to	6	3.92
				cause	6	3.92
				and	12	7.84
				∅	9	5.88
			Non-causal expressions (N=31, R=20.26)	, which	2	1.31
				after	2	1.31
				when from, drive, disrupt, get	2 4	1.31 2.60
5	Vst-neun tonge	4	Causal expressions (N=2, R=50)	since	1	25
				due to	1	25
			Non-causal expressions (N=2, R=50)	and	1	25
				confuse	1	25

- a. Korean: 기업이 투자에 거리를 둔 탓에 순자금 조달이 위축됐다.
 kiʌbi tʰudzæ ɡʌritʰur dun tʰasɐ sundzɑɡʌm dzodari witeʰuktʰwettʰɑ.
 English: Net fund raising has been discouraged *because* companies have distanced themselves from investments.
- b. Korean: 한때 우리나라 농업은 생산성을 너무나도 중시 한 나머지 농약과 비료를 너무 많이 써서 생명력을 잃은 논이 많다.
 hantʰɐ urinara noŋʌbʌm seŋsʌnsʌŋʰur ʌmunado dzʌŋsian namʌdzi
 noŋjakkʰwa birjorʰur ʌmu mani sʰʌsʌ seŋmjʌŋŋʌɡʰur iɾʌm noni mantʰɑ.
 English: There are many rice paddies that have lost their vitality due to excessive use of pesticides and fertilizer *because* productivity was much too important for Korean agriculture.
- c. Korean: 진짜? 나도 출장 다녀오느라고 전혀 몰랐지.
 teinteʰɑ? nado teʰuldzʌŋ danjʌomurago dzʌŋjʌ mollatteʰi.
 English: Really? I didn't know *because* I was travelling for work.
- d. Korean: 지하철이 고장으로 멈추는 바람에 하마터면 지각할 뻔했어.
 teiateʰʌɾi godzʌŋtʰʌro mʌmtɐʰʌmun baramɐ amatʰʌmjʌn dzigakʰar
 pʰʌnɐsʰʌ.
 English: I was almost late *because* the subway stopped due to a failure.
- e. Korean: 대형마트나 홈쇼핑, 백화점에서 이벤트를 벌이며 가격 할인행사까지 하는 통에 소상공인들은 더욱 찬바람만 느끼고 있다.
 tejʌŋmatʰʌma omsjopʰiŋ bekʰwadzʌmɐsʌ ibentʰʌrʰur bʌrimjʌ ɡʌɡjʌɡ
 ʌɾineŋsakʰadzi ʌmun tʰoŋɐ sosʌŋɡoŋindʰʌrʌm dʌug teʰʌmbaramman nʌtʰikʰigo
 ittʰɑ.
 English: Small business owners are feeling the chill of the wind more *since* the large supermarkets, home shopping and department stores are holding events and even price discount events.

Fig. 3. Example 1: Korean sentences and corresponding English sentences

4.2. [+Uncertainty] feature in causal connective expression

The English expressions corresponding to the Korean causal connective expression *Ast/Vst-a/eoseo geureonji*, which indicates the [+uncertainty] feature, were examined. Table 3 presents the frequency and proportion of Korean causal connective expressions with the [+uncertainty] feature in the Korean-English parallel corpus, along with their corresponding English expressions.

Table 3. Korean Causal Connective Expression Containing the [+uncertainty] feature and Corresponding English Expressions

No.	Korean	N	Category	English	N	Rate
1	Ast/Vst-a/eoseo geureonji	124	Causal expressions (N=113, R=90.4)	because	31	24.80
				maybe (it's) because	28	22.40
				so	13	10.40
				probably because	11	8.80
				perhaps because	7	5.60
				since	7	5.60
				due to	3	2.40
				maybe that's why	2	1.60
				maybe the reason	2	1.60
				not sure if it is because	2	1.60
				with	2	1.60
				guess it's because	1	0.80
				perhaps ... so	1	0.80
				perhaps due to	1	0.80
				whether it's from	1	0.80
				so ... that	1	0.80
			Non-causal expressions (N=12, R=9.6)	∅	8	6.40
				and	2	1.60
				as to whether	1	0.80
				no wonder	1	0.80

Conjectural terms such as "*maybe*," "*probably*," and "*perhaps*" were absent in over 50% of the English expressions corresponding to *Ast/Vst-a/eoseo geureonji* within the causal expression category. This observation suggests that instances of speculation in causal expressions are significantly more frequent in Korean.

The Korean sentences in (a) (c) in Figure 4, all use causal expressions that imply the [+uncertainty] feature. However, in the corresponding English sentences, only (a) contains '*probably*', which expresses [+uncertainty]. The English sentence in (b) uses only the causal connective expression '*since*' without the [+uncertainty] feature, and (c) has no corresponding causal connective expression at all. As noted, this aligns with the phenomenon of frequent use of hedging expressions in Korean. When expressing causation, there is a tendency to convey the message in a mild rather than strong or explicit manner, which is closely related to demonstrating politeness toward the listener. This distinction is particularly apparent in examples (b) and (c). In (b), if the hearer is struggling with an event while the speaker is fine, even though the speaker does not have to use the [+uncertainty] feature because "*I'm used to it because I go there every day*" is a recurring event for the speaker and the speaker's personal belief, to show consideration for the hearer, the speaker softens her stance by adding [+uncertainty] to the explanation of her ease with the

- a. Korean: 시험 기간 *이라서* 그런지 도서관에 사람이 엄청 많네.
 siam giganirasA guurAndzi dosAgwane sarami amte^hΛη manne.
 English: There are a lot of people in the library *probably because* it's the exam season.
- b. Korean: 나는 매일 다 *녀서* 그런지 익숙한데, 어떤 게 가장 힘들니?
 nanun meir danjAsA guurAndzi iks*uk^handε At*Λη ge gadzan imdtuni?
 English: I'm used to it *since* I walk there every day, but what is most difficult?
- c. Korean: 제가 일한 지 얼마 안 *돼서* 그런지 업무량이 조금 버거운데, 혹시 비법 같은 것 없어요?
 tεεgα iran dzi Λma an dwesA guurAndzi ammurjanj dzogum bagΛunde
 oks*i bibAb gat^hun gAs aps*Λjo?
 English: It hasn't been long since I've worked here. I feel like the workload is a little heavy. Is there a secret you can tell me?

Fig. 4. Example 2: Korean sentences and corresponding English sentences

situation. In the case of (c), the speaker is reducing the burden on the listener by adding [+uncertainty] to the reason that she is feeling overwhelmed, instead of saying that her workload is heavy.

4.3. [+Plurality] feature in causal connective expression

The English expressions corresponding to the Korean causal connective expression *Ast/Vst-go haeseo*, which denotes the [+plurality] of the reason, were examined. Table 4 presents the frequency and proportion of Korean causal connective expressions featuring the [+plurality] aspect in the Korean-English parallel corpus, along with their corresponding English expressions.

Table 4. Korean Causal Connective Expression Containing the [+plurality] feature and Corresponding English Expressions

No.	Korean	N	Category	English	N	Rate
1	<i>Ast/Vst-go haeseo</i>	9	Causal expressions (N=7, R=77.78)	because	2	22.22
				so	2	22.22
				since	1	11.11
				due to	1	11.11
				as	1	11.11
			Non-causal expressions (N=2, R=22.22)	∅	2	22.22

As shown in Table 4, [+plurality] feature of the causal expression is included only in Korean. In English, there is no expression which indicate the [+plurality] feature at all. In that, the [+plurality] feature of cause is the unique characteristic of Korean causal expressions. The following (a) and (b) in Figure 5 are the examples of Korean sentences containing the causal connective expressions with [+plurality] feature and the corresponding English sentences.

- a. Korean: 아직 못 했습니다. 새로운 바이어 분이라 취향도 모르고 해서 아직 고민 중입니다.
 adzig mot^hets^{*}umnida seroun bai^h bunira te^hwijando morugo esa
 adzig gomindzugimnida.
 English: Not yet. I don't know really know what they like since it's our first time meeting them.
- b. Korean: 어린 아이도 있고 해서 집에서 먹으려고요.
 arinaido itk^{*}o esa dzibes^h magw^hajagojo.
 English: I'm going to eat at home since I have children.

Fig. 5. Example 3: Korean sentences and corresponding English sentences

In both (a) and (b) in Figure 5, *Ast/Vst-go haeseo* is employed to indicate that the preceding clause represents one of several reasons for the information presented in the subsequent clause in Korean sentences. However, in the corresponding English sentences, this connotative meaning was not expressed. Unlike in English, the [+plurality] feature, which indicates that the presented reason is just one among several possible reasons, frequently occurs in Korean. This allows speakers to avoid definitive expressions by emphasizing that the reason given is one of multiple factors, rather than asserting it as the sole reason. For example, the speaker in (a) is in a situation where they must admit that they have not yet completed a task and must explain why. In this case, they adopt a defensive attitude by employing *Ast/Vst-go haeseo*, which implies that there are various reasons, to avoid explicitly stating the reason that is causing concern. The use of *Ast/Vst-go haeseo* also hints that the reason they have not yet finished is due to the many considerations they must consider. Should they provide a single, clear reason for their indecision, and that reason is countered, they would lose any further opportunity to defend their position. However, by establishing a situation of indecision and implying that there are various reasons, not just one, the speaker ensures they will have additional opportunities to defend themselves, even if the listener objects to the reason they have given.

This can also be interpreted as a strategy to avoid expressing one's intentions too assertively. It is often utilized in contexts where the speaker seeks to refrain from stating a definitive reason. For instance, when declining someone's request or suggestion, a speaker may opt to provide nuanced explanations to prevent the other person from losing face or may choose not to reveal their true reasons for rejecting the request. In (b), for example, the speaker faces the challenge of declining an invitation to dine outside with a large group. Instead of directly expressing a desire to refuse the listener's proposal, the speaker aims to minimize any potential damage to the proposer's face. They achieve this by im-

plicitly conveying, using *Ast/Vst-go haeseo*, that there are multiple reasons prompting their decline, rather than simply a personal intention to reject the offer.

While the current analysis does not directly implement the computational task defined in Section 1.3, it demonstrates its theoretical feasibility through manual, feature-based evaluation. The findings serve as preliminary evidence and a foundational dataset for developing future automated classifiers or scoring systems that assess pragmatic preservation in translation.

5. Conclusion

5.1. Summary of Findings

This study is an interdisciplinary investigation that utilizes the AI-Hub Korean-English parallel corpus to contrastively analyze causal connective expressions in Korean and English. The primary objective was to identify the distinctive characteristics of Korean causal connectives, particularly those that exhibit additional pragmatic features such as [+negative], [+uncertainty], and [+plurality]. These features were analyzed both linguistically and culturally to determine their uniqueness to the Korean language and their prevalence in Korean discourse.

Through the analysis of seven specific Korean causal connectives that carry these features, the study revealed that such expressions are largely absent in English. The additional pragmatic layers found in Korean are closely tied to cultural tendencies toward indirectness, politeness, and the mitigation of speaker stance. For instance, Korean speakers often employ causal connectives with [+negative] connotations to express disapproval or refusal indirectly, a strategy aligned with cultural norms that emphasize maintaining social harmony and avoiding face-threatening acts. Similarly, connectives with [+uncertainty] and [+plurality] allow speakers to hedge their intentions, introduce speculative reasoning, or diffuse responsibility, all of which align with Korean cultural values surrounding modesty and relational sensitivity.

In contrast, English tends to favor direct and explicit expressions of causality, reflecting cultural norms that value clarity, individual agency, and unambiguous communication. English causal connectives thus typically lack the additional semantic and pragmatic layers found in their Korean counterparts.

By comparing how causality is linguistically and culturally encoded in Korean and English, this study offers meaningful insights into the intersection of language, culture, and communication. The findings not only contribute to contrastive linguistics but also have practical implications for language education, particularly in the field of Korean as a foreign language. Recognizing these culturally embedded linguistic patterns can support more effective cross-cultural communication and foster greater intercultural understanding, which are essential for promoting social cohesion in a globalized society.

5.2. Ethics and Limitations

This study also acknowledges several ethical and methodological considerations. The data used for analysis come from the publicly available Korean-English parallel corpus developed by *AI-Hub*. Due to licensing constraints set by the National Information Society

Agency (NIA), the corpus itself cannot be openly redistributed. However, qualified researchers may apply for access through the *AI-Hub* platform, and, upon request, we can provide preprocessing scripts, annotation guidelines, and representative samples that comply with licensing terms for academic use.

From an ethical standpoint, we recognize the potential for cultural bias embedded in the corpus. Causal connectives often reflect nuanced pragmatic intentions that are deeply influenced by cultural norms. Therefore, any computational or rule-based interpretation of pragmatic equivalence across languages must be approached with caution. The indirectness and politeness strategies observed in Korean causal expressions may not have direct equivalents in English, and modeling such shifts automatically can oversimplify or misrepresent these meanings.

Additionally, there is a risk of overgeneralization when applying pragmatic classification tasks to other language pairs or datasets. Cultural and linguistic variability means that tools trained on one corpus may not perform reliably in other contexts. The pragmatic shift classification task proposed in this study is meant as a theoretical and methodological foundation, not as a comprehensive solution.

In terms of methodological limitations, although we developed detailed annotation guidelines and achieved high inter-annotator agreement (Cohen's $\kappa = 0.87$), pragmatic interpretation is inherently subjective. Ambiguities remain, particularly in context-dependent cases that are difficult to resolve without deeper discourse information. Moreover, while this study defines a computational task and proposes potential formulations for pragmatic shift evaluation, we have not implemented or tested these models empirically within this paper. Future work should include experimental evaluation using classification metrics such as Accuracy or F1-score, as well as error analysis to better understand the impact of cultural and linguistic nuances on model performance.

Despite these limitations, we believe this study contributes to a deeper understanding of pragmatic shifts in translation and provides a meaningful step toward computational modeling of culturally embedded language use.

References

1. Ahn, Heeyoung. *A Study of the Hierarchy of Modern Korean Connective Endings*. Master's dissertation, University of Seoul (2007).
2. Bang, J.-U., Maeng, J.-G., Park, J., Yun, S., and Kim, S.-H. English–Korean speech translation corpus (EnKoST-C): Construction procedure and evaluation results. *ETRI Journal*, 45, 18–27 (2023).
3. Chin, Chongnan. *A Study on the Discourse Grammar of Korean Causative Expressions*. Doctoral dissertation, Hankuk University of Foreign Studies (2005).
4. Choi, Sujeong. *A Study of the Meaning-Based Categorization of Grammar Items for Synonymous Grammar Education of Korean Language*. Doctoral dissertation, Yonsei University (2022).
5. Conklin, Harold. *Hanunoo Agriculture*. Rome: Food and Agriculture Organization of the United Nations (1957).
6. Jeon, Hye-Young. *A Study on the Teaching of Connective Endings for Reason-Cause*. Master's dissertation, Hansung University (2021).
7. Kay, M., and Röscheisen, M. *Text-translation alignment*. Technical Report. Xerox Palo Alto Research Center (1988).

8. Kim, Sun Hye, and Jang, Seung Hye. The Method of Describing Micro-structure in Bilingual Dictionaries Using Parallel Corpus - Focusing on Description of Grammatical Elements. *Language Facts and Perspectives*, 28, 263–292 (2011).
9. Li, Yunhui. *A Study on the Spiral Education of Korean's Connective Ending '-aseo' and '-nikka' for Foreign Students*. Master's dissertation, Sejong University (2011).
10. Li, Wenhua. A Study on the Corresponding Form of the Chinese “Zai” in the Parallel Corpus. *The Journal of Learner-Centered Curriculum and Instruction*, 21(19), 59–73 (2021).
11. O, Seon Yeong, and Keiko, Takiguchi. Study on Korean “eopsi” adverbs and their Japanese equivalent patterns based on examples in the Korean-Japanese parallel corpus. *Korean Language and Literature*, 65, 37–69 (2015).
12. Park, Chan-Jun, and Lim, Heui Seok. A Study on the Performance Improvement of Machine Translation Using Public Korean-English Parallel Corpus. *Journal of Digital Convergence*, 18(6), 271–277 (2020).
13. Park, Chanjun, Shim, Midan, Eo, Sugyeong, Lee, Seolhwa, Seo, Jaehyung, Moon, Hyeonseok, and Lim, Heuiseok. Empirical Analysis of Parallel Corpora and In-Depth Analysis Using LIWC. *Applied Sciences*, 12 (2022).
14. Park, Dae Beom. *Connection Expression Education Research of the Reason Cause for the Korean Language Studying Person*. Master's dissertation, Sangmyung University (2008).
15. Park, Minsin. *Study of Semantic Category-based Korean Grammar Education, Focusing on Causative Expressions*. Doctoral dissertation, Seoul National University (2018).
16. Park, Myongsu. Investigation into English-Korean Parallel Corpus with ParaConc. *The Journal of Translation Studies*, 18(5), 29–57 (2017).
17. Saville-Troike, Muriel. *The Ethnography of Communication: An Introduction*. 3rd edn. Oxford: Blackwell (2003).
18. Seo, Se Jung. A Contrastive Study of the Korean Precedent Ending “-ket-” and its Corresponding Forms in English. *Language Facts and Perspectives*, 22, 193–215 (2008).
19. Shen, Lanji. A Study on the Korean Expressions Corresponding to the Chinese Adverbs ‘(hai)’ - focused on Chinese-Korean parallel corpus. *Language Facts and Perspectives*, 36, 247–277 (2015).
20. Sim, Ji-Young. A Study on the Corresponding Expressions of “ Causative Using Parallel Corpus - Focused on the Teaching of A→B Translation between Chinese and Korean. *The Journal of the Korea Contents Association*, 23(5), 475–483 (2023).
21. Vu, Van-Hai, Nguyen, Quang-Phuoc, Shin, Joon-Choul, and Ock, Cheol-Young. UPC: An Open Word-Sense Annotated Parallel Corpora for Machine Translation Study. *Applied Sciences*, 10 (2020).
22. Wierzbicka, Anna. English Causative Constructions. In N. J. Enfield (Ed.), *Ethnosyntax*. New York: Oxford University Press (2002).
23. Wierzbicka, Anna. *Understanding Cultures through Their Key Words: English, Russian, Polish, German, Japanese*. New York: Oxford University Press (1997).
24. Wilczynski, Tomasz. *A Contrastive Study on the Korean ‘ida’ Constructions and Their Polish and English Counterparts - Focused on Bible Parallel Corpus*. Doctoral dissertation, Keimyung University (2021).
25. Xue, Wenying. *Study on the Contrast Between Korean-Chinese Honorific Expression Based on Parallel Corpus*. Doctoral dissertation, Keimyung University (2021).
26. Yoo, Hae-jun. Contents Construction Method for Teaching Korean Grammar Based on Language Functions. *The Journal of Language & Literature*, 61, 515–538 (2015).
27. Yu, Xiangxin. A Study on Chinese Corresponding Expressions of ‘-(eu)leo’ Based on Parallel Corpus. *Teaching Korean as a Foreign Language*, 66, 161–192 (2022).

Sujeong Choi is an Assistant Professor of Practice in the Korean Language Program at KDI School. She received her Ph.D. in Korean Linguistics with a specialization in

Korean language education from Yonsei University. Her research interests include Korean pragmatics and AI-integrated Korean language education.

Sin-hye Nam Sin-hye Nam is an Assistant Professor in the Department of Korean Language and Literature at Kyung Hee University. She received her Ph.D. in Korean Linguistics from Yonsei University. Her research interests include corpus linguistics and applied linguistics.

Received: Month DD, 20YY; Accepted: Month DD, 20YY.