

EVOKE: Emotion Vocabulary Of Korean and English

Yoonwon Jung¹, Hageong Shin², Benjamin K. Bergen¹

¹Department of Cognitive Science, University of California San Diego, USA

²Work done at Department of Linguistics, University of California San Diego, USA
{y5jung, hashin, bkbergen}@ucsd.edu

Abstract

This paper introduces EVOKE (Emotion Vocabulary of Korean and English), a Korean-English parallel dataset of emotion words. The dataset offers comprehensive coverage of emotion words in each language, in addition to many-to-many translations between words in the two languages and identification of language-specific emotion words. The dataset contains 1,426 Korean words and 1,397 English words, and we systematically annotate 819 Korean and 924 English adjectives and verbs. We also annotate multiple meanings of each word and their relationships, identifying polysemous emotion words and emotion-related metaphors. The dataset is, to our knowledge, the most systematic and theory-agnostic dataset of emotion words in both Korean and English to date. It can serve as a practical tool for emotion science, psycholinguistics, computational linguistics, and natural language processing, allowing researchers to adopt different views on the resource reflecting their needs and theoretical perspectives. The dataset is publicly available at <https://github.com/yoowonj/EVOKE>.

Keywords: emotion words, cross-linguistic dataset, annotations, polysemy, metaphors

1. Introduction

Emotion words reveal how humans conceptualize and communicate emotional experiences through language (Wierzbicka, 1999). They also provide a foundation for lexicon-based sentiment analyses and emotion detection (Liapis et al., 2025; Raji and De Melo, 2020). Yet, emotion words are diverse and notoriously difficult to identify (Fehr and Russell, 1984; Wierzbicka, 1999; Wu and Zhang, 2025). While some words transparently denote emotional states (e.g., *sadness*, *anger*), others evoke or reflect bodily sensations, cognitive evaluations, or physiological or behavioral expressions that may or may not indicate emotional states proper (e.g., *tense*, *captivating*, *blushing*). Consequently, existing datasets of emotion words vary in how they define and annotate emotion words, reflecting a range of theoretical positions and methodological frameworks (Clore et al., 1987; Fehr and Russell, 1984; Hong and Jeong, 2009; Johnson-Laird and Oatley, 1989; Lee, 2006; Park and Min, 2005).

A maximally useful dataset of emotion words in a language would have a few features. First, it would be comprehensive—covering all words that could be, on any theory, emotion words. Second, it would explicitly identify selection criteria that establish whether words are or are not emotion terms, and would allow users to apply their preferred criteria to create bespoke theory-driven views on the data. However, existing datasets generally cover subsets of the possible emotion words in any language. Moreover, they often apply different criteria to identify emotion words based on specific theoretical frameworks, which makes it challenging to integrate or compare datasets (Fehr and Russell,

1984; Baron-Cohen et al., 2010; Clore et al., 1987; Hong and Jeong, 2009; Johnson-Laird and Oatley, 1989; Lee, 2006; Park and Min, 2005).

Cross-linguistic datasets extend the value of monolingual resources, as comparisons of emotion words can reveal emotions that may be shared or specific across cultures (Jackson et al., 2019; Russell, 1991). To date, however, the coverage of words in cross-linguistic datasets is limited. Because meanings do not always translate one-to-one across languages (Catford, 1965; Thompson et al., 2020), a word in one language may not have any direct translation in other languages (and thus create lexical gaps), or might align with multiple words in another language. For example, while English *dismay* does not have an exact translation in Korean, English *sad* can correspond to multiple, subtly distinct Korean words (i.e., *서럽다* (*söröpta*)¹ and *슬프다* (*sülpüta*)², each of which also has several different English correspondents of its own. This underscores the need for cross-linguistic datasets of emotion words with comprehensive lexical coverage in each of the languages and many-to-many translational mappings.

Lastly, emotion words are complex in that they are often polysemous. For example, Korean *부끄럽다* (*pukküröpta*) encompasses the meaning of both *shy* or *shameful*. Other polysemous senses are often metaphorical. One example is the emotional and spatial senses of *low* in English. Metaphorical relations among senses of emotion words,

¹To be sad and depressed (National Institute of Korean Language, 2025a).

²Sad and sorrowful enough to make one cry (National Institute of Korean Language, 2025b).

and polysemy more generally, reveal how emotion semantics are organized within and across languages (Kövecses, 2003; Sauciuc, 2009), yet they are rarely documented in emotion word datasets.

In light of these gaps, we present a dataset of emotion words in two languages with (1) comprehensive lexical coverage and annotations by native speakers based on several different theories to accommodate diverse views and uses, (2) many-to-many translational mappings established by bilingual experts using existing lexicographic resources, and (3) identification of polysemous words and metaphorical relations among word senses.

2. Background

2.1. Theories Defining Emotion Words

Psychological and linguistic work has produced varied methods to define emotion words. Some early studies relied on unguided intuition, with raters judging words as emotional or not without explicit criteria (Fehr and Russell, 1984; Storm and Storm, 1987). More recent work uses acceptability judgments on emotion words, placing them in frame sentences for expressing emotions like “I feel X” or “I am X” to evaluate whether they sound acceptable in those contexts (Baron-Cohen et al., 2010; Clore et al., 1987; Johnson-Laird and Oatley, 1989). Another recent approach uses exclusion criteria, asking if a word describes bodily sensation or epistemic state without emotionality, to filter out some obvious non-emotion words (Baron-Cohen et al., 2010; Johnson-Laird and Oatley, 1989; Park and Min, 2005).

Research on Korean emotion words classifies emotion-related predicates into different types: state-oriented (a state experienced by an animate subject), evaluation-oriented (properties of external entities), and expression-oriented (behavioral expressions and responses) (Hong and Jeong, 2009; Lee, 2006; Park and Min, 2005). The distinction between stative and evaluative predicates is closely tied to causality, echoing the proposal to distinguish externally caused emotions (e.g., “I am saddened by his death”) from those concerning internal goals (e.g., *desire*) (Johnson-Laird and Oatley, 1989).

2.2. Cross-linguistic Studies

The majority of existing cross-linguistic studies of emotion words prioritize the prototypicality of emotion over exhaustiveness, focusing on small sets of emotion terms presumed to be universal (typically fewer than 30) across large numbers of languages (e.g., List et al., 2018), or utilize restricted sets of words from bilingual studies (e.g., Bromberek-Dyzman et al., 2021; Tang et al., 2023). These approaches limit the exploration of finer-grained distinctions of emotion words within a language.

Another line of research highlights culture-specific emotion words, captured as words lacking equivalents in other languages (i.e., *lexical gaps*). For example, Korean words *정* (*chöng*) or *한* (*han*) lack equivalents in English, and are expressible only with longer descriptions, like “long-lasting affection and caring based on perceived closeness” (Choi and Choi, 2002) or “Unresolved resentment, grief, and anger, or a negative and long-lasting emotion encapsulating the grief of historical memory” (Kim, 2017), respectively. Cross-linguistic lexical gaps signal how different language communities represent and express emotions differently (Lomas, 2018; Lupyan, 2012; Rissman et al., 2023; Winawer et al., 2007). However, they are often discovered in small numbers through case studies (e.g., Choi and Choi, 2002; Kim, 2017; Schmidt-Atzert and Park, 1999), and are rarely identified at scale within a broader lexicon of emotion in one language in relation to other languages.

2.3. Semantic Relations and Metaphors

The meanings of polysemous emotion words are often connected through metaphorical extension from one meaning to other meanings (Bartsch, 2002). Comparing patterns of such emotion-related metaphors across languages can reveal how emotional meanings are structured similarly or differently across cultures (Kövecses, 2003). Identifying these multiple senses thus provides a basis for cross-linguistic studies on emotion semantics.

Existing annotation work on polysemous words mainly targets resolving general lexical ambiguities, not emotion words specifically (Haber and Poesio, 2024; Passonneau et al., 2010; Rumshisky and Batiukova, 2008). Similarly, cross-linguistic studies of emotion-related metaphors often focus on a small set of recurrent metaphorical mappings (e.g., “emotion as force” or “anger as a container”; Lakoff and Kövecses, 1987; Sauciuc, 2009; Türker, 2013), rather than examining metaphorical structures across a wider variety of emotion words. These gaps leave open questions about how emotion words develop multiple meanings and metaphorical extensions, and how such patterns distribute within and across languages.

3. Objective of the Dataset

We introduce EVOKE (Emotion Vocabulary Of Korean and English), a Korean-English parallel dataset of emotion words with comprehensive coverage in both languages plus cross-linguistic mappings (publicly available at <https://github.com/yoonwonj/EVOKE>). The dataset is constructed by compiling lists of words to annotate from prior studies on and datasets of emotion

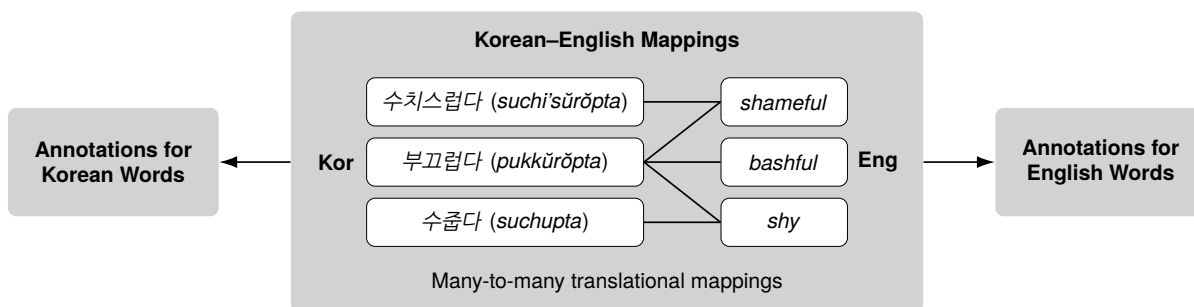


Figure 1: The structure of the Korean-English parallel emotion word dataset. Words in both languages are connected through many-to-many translational mappings, with annotations for Korean and English words.

words. Those words are then annotated according to multiple criteria used in earlier studies to investigate their characteristics as emotion words. The dataset also provides many-to-many translations of words between Korean and English, which enables discovering emotion concepts specific to each language. Lastly, the dataset identifies polysemous words and annotates the relationships between their senses, enabling the identification of emotion-related metaphors.

We focus on Korean and English for several reasons. English serves as a well-studied reference point, yet there does not yet exist a comprehensive resource with structured annotations like the one constructed here. Korean, meanwhile, is spoken in distinct cultural settings, and contains purportedly unique, culture-specific emotion terms (e.g., *정 (chöng)*, *한 (han)*) that lack direct English equivalents (Choi and Choi, 2002; Kim, 2017). Comparison of the two allows for the investigation of both shared and culture-specific aspects of emotion lexica. Additionally, the research team’s expertise in linguistics, cognitive science, and knowledge as native speakers of English and Korean makes it feasible to construct a reliable dataset.

4. Dataset Construction

The dataset consists of three separate components (see Figure 1): (1) Korean-English mappings, (2) annotations for Korean words, and (3) annotations for English words.

4.1. Word Selection and Translation

Candidate words were gathered from previous work on emotion words in English (Baron-Cohen et al., 2010; Morgan and Heise, 1988; Storm and Storm, 1987) and Korean (Jeon et al., 2022; Park and Min, 2005; Rhee and Ko, 2013).

To construct translational mappings (illustrated in Figure 1), two Korean-English bilingual speakers judged whether each word had translational equivalents in the other language. A manual translation

approach was adopted to capture nuanced and precise meanings that are difficult to obtain through automatic translation. The translators consulted two bilingual Korean-English dictionaries (one Korean-to-English and one English-to-Korean) for translation (National Institute of Korean Language, 2025c; NAVER, 2025). In addition, definitions from Korean (National Institute of Korean Language, 2025e) and English monolingual dictionaries (Oxford University Press, 2025; Cambridge University Press, 2025) were compared to ensure accurate and semantically rich matching, especially when translations provided in bilingual dictionaries varied across sources or when forward and backward translation yielded inconsistent results.

A word is considered to have a translational equivalent when it can be translated to a single word form (i.e., one or more words exist in both languages to denote the same meaning). Accordingly, lexical gaps were identified when (1) bilingual dictionary queries for words in one language only yielded multi-word expressions or idiomatic phrases in the other language,³ or when (2) back-translation produced inconsistent results. In the latter cases, the translators consulted additional native speakers and monolingual dictionaries to determine whether the dictionary equivalent reflected a true translational match that fully captured the meaning of the query word.

Moreover, lexical gap was defined as the absence of a word for a concept in one language when such a word exists in the other (e.g., Janssen, 2004; Li et al., 2024). Therefore, gaps arising at the morphosyntactic level were not included, as those do not indicate the absence of conceptual or semantic content (Ivir, 1977; Lomas, 2018; Wierzbicka, 1999).⁴

³Hyphenated forms were treated as single lexical items (e.g., *cliché-ridden*). Korean words that may appear either as single words or with internal spaces were treated as single-word entries and were not considered as lexical gaps (e.g., *사려 깊은 (saryökipün)* ‘considerate’).

⁴Some Korean verbs and nouns are systematically

ID	Label	Annotation Question
Part 1. Acceptability judgments		
ACPT1	1st person “feel”	Does “I feel X” sound acceptable?
ACPT2	3rd person plural “feel”	Does “They feel X” sound acceptable?
ACPT3	Inanimate subject “feel”	Does “It feels X” sound acceptable (inanimate “it”)?
ACPT4	1st person “am”	Does “I am X” sound acceptable?
Part 2. Semantic experienter judgments		
EXP5	Subjectivity of experience	In “It feels X”, can the sentence express what “it” experiences?
EXP6	Evaluation as experience	In “It feels X”, can the sentence express my evaluation of “it”?
EXP7	Caused experience	In “I am X”, does X denote a caused state of its associated noun/verb?
EXP8	Causing experience	In “I am X”, does X denote a causing state of its associated noun/verb?
Part 3. Exclusion criteria		
EXCL9	Pure bodily sensation	Does X describe a pure bodily/physical sensation?
EXCL10	Behavioral expression	Does X denote a behavioral expression?
EXCL11	Pure epistemic state	Does X denote a non-emotional epistemic state?
Part 4. Multiple meanings		
POLY12	Additional Meaning	Does X have another distinct meaning to annotate separately? → If yes: create a new row for that sense and complete annotations in Parts 1–3.
POLY13	Distinctiveness	Are the two senses of X in different domains?
POLY14	Relatedness	Are the two senses of X systematically related?

Table 1: Codebook for the annotation scheme. Part 1 collects acceptability judgments of the target word in four different sentences. Part 2 includes two follow-ups for ACPT3 (EXP5–6) and ACPT4 (EXP7–8) each. Part 3 applies exclusion criteria to filter relatively obvious non-emotion terms. Part 4 documents polysemy and links senses via distinctiveness and relatedness judgments. All questions used two binary labels (*acceptable/unacceptable* in Part 1, *yes/no* in Parts 2–4), with an additional option to choose *unsure*. ACPT1, ACPT2, ACPT3, EXP5, and EXP6 were only applied to adjectives, and the rest of the others were applied to both verbs and adjectives, with modifications for verb annotations. Annotation criteria were translated into Korean for Korean word annotations with appropriate modifications applied.

4.2. Annotation Objective and Process

We annotated words whose part-of-speech is identified as adjectives and verbs in each language. Adjectives linked to identifiable lexical roots⁵, either nouns or verbs, were annotated as noun–adjective or verb–adjective pairs. Adjectives without such identifiable lexical roots, along with all verbs, were annotated independently. Nouns were not annotated because the sentence contexts most commonly used for acceptability judgments in the literature are designed around predicative uses of words (Baron-Cohen et al., 2010; Clore et al., 1987; Johnson-Laird and Oatley, 1989), and several existing annotation criteria target semantic properties specific to predicates (see Section 4.3.2; Hong and Jeong, 2009; Lee, 2006; Park and Min, 2005).

translated into English using *be-* constructions (verb) and *being-* constructions (noun). This creates gaps in verbs and nouns for English relative to Korean in a systematic way. See Bentivogli et al. 2000 for additional examples and discussion of morphosyntactic gaps.

⁵A form commonly perceived as a base form without its morphological derivation.

We recruited three native English speakers and three native Korean speakers as annotators⁶, who had backgrounds in cognitive science or linguistics. Annotators completed one week of training followed by ten weeks of annotation. During weekly meetings, researchers and annotators discussed questions and edge cases. Those discussions resulted in modifications to the annotation guidelines meant to improve the robustness of subsequent annotations.

Annotators made binary judgments (*acceptable/unacceptable*; *yes/no*) for each annotation criterion, with the option to choose *unsure*. We chose binary judgments over graded ratings because some of the questions are intrinsically categorical. Binary ratings also avoid subjectivity introduced by continuous scales (e.g., differential use of Likert scales per annotator) and allow for easier combination of annotation results from the different criteria to select words that qualify as emotion words (see Section 6.1 for an example of a use case).

⁶Annotators received course credits as compensation.

Items marked `unsure` were reviewed in weekly meetings. Cases that could be resolved through group discussion were annotated again with binary labels. However, given the subjective nature of emotion word judgments, we did not force consensus. When annotators continued to find it difficult to make a binary decision, we kept the `unsure` label.

The dataset was split so that each annotator received a unique 30% of the words, while an additional 10% of the words were assigned to all three annotators for inter-rater reliability assessment.

4.3. Annotation Criteria

The annotation criteria used features of emotion words identified from the existing literature. Those criteria were selected to include or exclude candidate words from the category of emotion words. The full list is presented in Table 1.

Adjectives were annotated on all 14 criteria, while verbs were annotated on 9, excluding ACPT1, ACPT2, ACPT3, EXP5, and EXP6. In addition, ACPT4 was modified to suit part-of-speech differences. Because the criteria differ by language and by part-of-speech, Table 1 reports the annotation scheme for English adjectives, which encompasses all 14 of the criteria we devised. For Korean annotations, the criteria were translated and adapted to Korean syntax and usage patterns. Details of the modifications to Korean word annotations are in Appendix A.

4.3.1. Part 1: Acceptability Judgments

The first part of the annotation collected acceptability judgments to elicit native speakers' usage patterns of the words.⁷ Annotators judged whether target words sound `acceptable` or `unacceptable` in given sentences, with an option to choose `unsure`. The sentences served as standardized contexts for evaluating each emotion word's appropriateness for expressing emotion. When the word alone sounded unnatural, annotators could suggest an alternate phrasal structure, which was then adopted for the annotation (e.g., "I feel cared" could be replaced by "I feel cared for").⁸

The sentences for acceptability judgments were selected based on their previous adoption as inclusion criteria for emotion words (Baron-Cohen et al., 2010; Clore et al., 1987; Johnson-Laird and Oatley, 1989), from psychological and linguistic studies of emotion (Hong and Jeong, 2009; Liu, 2016; Lee, 2006; Niedenthal, 2008; Paul et al., 2020).

ACPT1: 1st person "feel", ACPT2: 3rd person plural "feel", ACPT3: Inanimate subject "feel"

⁷We ground the acceptability judgments in pragmatic acceptability instead of grammatical acceptability.

⁸This was in response to differences between English, which commonly uses phrasal verbs, and Korean, which instead tends to use morphological derivation.

Three sentence contexts based on "I feel X" but with varying subjects were used to assess whether words activate subjective feelings of emotion (Lee, 2006; Niedenthal, 2008; Paul et al., 2020): "I feel X" (ACPT1: 1st person "feel"), "They feel X" (ACPT2: 3rd person plural "feel"), and "It feels X" (ACPT3: Inanimate subject "feel"), where X is a target word. Participants were instructed to interpret "I" and "they" as animate experiencers (1st person and 3rd person plural), and "it" as an inanimate object. If a word denotes a subjective feeling, it is expected to be judged as acceptable with animate subjects, but not with inanimate ones (Baron-Cohen et al., 2010; Clore et al., 1987; Dowty, 1991; Lee, 2006; Liu, 2016).

ACPT4: 1st person "am" "I am X" (ACPT4: 1st person "am") was used to contrast with "I feel X" (1st person "feel") and capture words denoting states versus evaluations. For example, *caring* fits "I am caring" but not "I feel caring". This contrast helps reveal whether words refer to internal states or external evaluations. However, some words (e.g., *pretty*) can still be acceptable in both "I feel X" and "I am X". Follow-up questions in Section 4.3.2 further clarify these cases (Clore et al., 1987; Hong and Jeong, 2009; Lee, 2006; Park and Min, 2005).

4.3.2. Part 2: Semantic Experiencer Judgments

The second part introduced semantic experiencer judgments as follow-up questions to ACPT3 and ACPT4. These questions aimed to handle patterns in acceptability judgments like those for *pretty*. Specifically, these questions helped determine whether words describe subjective experiences or external evaluations, reflecting the contrast between stative and evaluative words (Choi, 2008; Hong and Jeong, 2009; Lee, 2006).

EXP5: Subjective experience, EXP6: Evaluation Annotators judged whether the word in "It feels X" (ACPT3: Inanimate subject "feel") could express what the inanimate subject "it" experiences (EXP5), and whether it could express the speaker's evaluation of "it" (EXP6). These follow-up judgments clarify whether ACPT3 acceptability stems from evaluative readings. For instance, "it feels merciless" may sound acceptable if "it" refers to a fighter jet, but such usage reflects evaluation rather than the inanimate subject's feelings. These follow-up questions ensure the frame captured the intended semantics and help sort out such exception cases.

EXP7: Caused, EXP8: Causing Follow-up questions for ACPT4 (1st person "am") were designed to complement the distinction between experiential and evaluative words, following the "caused emotion" distinction of Johnson-Laird and Oatley (1989). Annotators were asked to judge whether the target word describes a "caused" (EXP7) or "caus-

ing” (EXP8) state within the “I am X” sentence context. Korean annotators were allowed to adjust the subject to consider additional naturalistic contexts, incorporating the annotators’ suggestions for more comprehensive evaluation.

To guide these judgments, adjectives were paired with their identifiable root nouns or verbs (e.g., *sadness*–*sad*), and were asked to judge whether each adjective is a caused state or a cause of the paired root nouns or verbs. For example, annotators judged whether *sad* refers to the caused state of *sadness* or its causing state. For the remaining adjectives, annotators were asked to imagine hypothetical noun forms.

These causality judgments clarify whether a word denotes an internal state or an external evaluation: internal states typically correspond to caused forms (e.g., *depressed*), while evaluations often align with causing forms (e.g., *depressing*). A word may encompass both caused and causing meanings, or words may represent each meaning separately. Moreover, words with similar meanings in different languages may capture these caused and causing meanings differently. Such differences could help reveal varying semantic properties of emotion words across languages.

4.3.3. Part 3: Exclusion Criteria

The third part applied exclusion criteria (Park and Min, 2005; Baron-Cohen et al., 2010; Lee, 2006). While emotions involve coordinated mind–body processes, most theories distinguish them from pure bodily sensations (e.g., *thirsty*) or cognition (e.g., *enlightened*) (Anderson and Adolphs, 2014; Johnson-Laird and Oatley, 1989; Lee, 2006; Paul et al., 2020). Many also separate emotional states from physical expressions (e.g., *cry*), which are byproducts rather than core states (Anderson and Adolphs, 2014; Johnson-Laird and Oatley, 1989; Hong and Jeong, 2009; Lee, 2006; Paul et al., 2020). These judgments are used to guide the exclusion of non-emotion words.

Annotators judged whether each target word describes pure bodily or physical sensations (EXCL9), behavioral expressions (EXCL10), or pure epistemic states (EXCL11). These criteria were designed to flag words that primarily fell into these categories for potential exclusion from the set of emotion words. This procedure provides a direct way to filter out non-emotion terms according to widely shared theoretical assumptions, still allowing researchers to retain those as emotion words if desired.

4.3.4. Part 4: Multiple Meanings

POLY12: Additional meaning The fourth part of the annotation was introduced to capture whether each word had other meanings distinct from its

primary emotional meaning (POLY12). When annotators identified an additional meaning, they were instructed to complete annotations in Parts 1–3 for the additional meaning identified. To further specify semantic relationships among multiple senses, two follow-up questions assessed the relationship between the identified meanings.

POLY13: Distinctiveness, POLY14: Relatedness Annotators judged whether the two senses concern different domains or concepts (Distinctiveness; POLY13), and whether they are connected in a systematic way (Relatedness; POLY14). For example, the word *low* exemplifies a prototypical metaphorical relationship. Its physical and emotional meanings belong to distinct domains yet remain systematically linked through the conceptual mapping between spatial position and emotional valence. In contrast, *blue* demonstrates a looser connection in that its color and emotional senses are distinct, but their association lacks systematic grounding. Although the metaphorical status of *blue* is debatable, it does not exhibit the systematicity typical of metaphors.

- (1)
 - a. “The ceiling is low.” → HEIGHT
 - b. “I am feeling low today.” → FEELING
 - c. → POLY13: yes, POLY14: yes
- (2)
 - a. “The sky is blue.” → COLOR
 - b. “I feel blue.” → FEELING
 - c. → POLY13: yes, POLY14: no

These dimensions derive from conceptual metaphor theory, and specifically the Metaphor Identification Protocol (Declercq and van Poppel, 2023; Group, 2007; Kovecses, 2010), to annotate words that could have emotional meaning through metaphorical extension (e.g., *low*). The annotation results could help reveal how emotional meanings emerge and relate to other meanings.

5. Dataset Analysis and Evaluation

5.1. Word characteristics

5.1.1. Part-of-Speech Statistics

Summary word statistics are in Table 2. A total of 1,426 Korean words and 1,397 English words were included. Translation equivalents were identified for all of these, forming many-to-many mappings across Korean and English (Figure 1).

5.1.2. Translational Equivalence

The majority of words had corresponding translations in the other language. For Korean words, 574 nouns (97.12%), 587 adjectives (96.86%), and 213 verbs (93.01%) had one or more translational equivalents in English. For English words, 495 nouns (97.44%), 630 adjectives (93.89%), and 213 verbs

	Nouns	Adjectives	Verbs	Total
Korean words	591	606	229	1,426
With translation(s) in English	574 (97.12%)	587 (96.86%)	213 (93.01%)	1,374 (96.35%)
Without translation(s) in English	17 (2.88%)	19 (3.14%)	16 (6.99%)	52 (3.65%)
English words	508	671	218	1,397
With translation(s) in Korean	495 (97.44%)	630 (93.89%)	213 (97.71%)	1,338 (95.80%)
Without translation(s) in Korean	13 (2.56%)	41 (6.11%)	5 (2.29%)	59 (4.22%)

Table 2: The statistics of Korean and English lexical entries with and without translational equivalents. Words without translational equivalents indicate lexical gaps. Polysemous words with either (1) multiple parts of speech, or (2) one of the meanings having translations in the other language but not in the other, were counted as separate lexical entries. See Appendix B for policies regarding counting polysemous word entries.

(97.71%) had one or more corresponding words in Korean.

The analysis of translational equivalents revealed a relatively similar degree of mapping complexity between the two languages. Korean words had an average of 1.61 translational equivalents in English (SD = 0.93), and English words had a nearly identical average of 1.65 translational equivalents in Korean (SD = 1.15). A total of 472 words in English (33.79%) and 542 words in Korean (38.01%) had two or more translational equivalents in the other languages. Some words demonstrated particularly rich translational mappings. The Korean adjective *우울하다* (*uulhata*) exemplified this complexity with six English equivalents: *depressed*, *dismal*, *gloomy*, *low*, *down*, and *blue*, capturing a range of physical and emotional states that Koreans express through a single word.

5.1.3. Lexical Gaps

Among Korean words, 17 nouns (2.88%), 19 adjectives (3.14%), and 16 verbs (6.99%) lacked translational equivalents in English. In English, 13 nouns (2.56%), 41 adjectives (6.11%), and 5 verbs (2.29%) had no corresponding translations in Korean.

These lexical gaps may indicate conceptual gaps, though this is not always the case. For example, some of those words indicate areas where each language expresses emotion concepts uniquely, like *답답하다* (*taptaphata*)⁹ (Schmidt-Atzert and Park, 1999). No single English word is equivalent, nor is there a conventionalized phrasal way to express the concept it denotes. However, other lexical gaps do have phrasal equivalents, like *moody*, which has a multi-word equivalent to “having ups and downs”

⁹Feeling dissatisfied and uncomfortable when someone’s behavior or the situation makes it hard to meet one’s expectation (National Institute of Korean Language, 2025d; Schmidt-Atzert and Park, 1999).

(*기분 변화가 심한* (*kipun byŏnhwaka simhan*)) in Korean. Systematically distinguishing these instances from each other is beyond the scope of this paper. However, because the identified lexical gaps include both cases, this dataset can be utilized to further explore empirical questions regarding the relationship between lexical gaps and conceptual gaps and the effects—if any—of the conventionality of multi-word descriptions for emotion concepts.

5.2. Annotation Results

A total of 819 Korean words (602 adjectives, 217 verbs) and 923 English words (678 adjectives, 245 verbs) were annotated.¹⁰

Annotation criteria differed by part of speech, which made averaging results across adjectives and verbs infeasible. We thus report results for adjectives, which constitute the largest portion of the dataset, and use all criteria (see Section 4.3 for details). Results of verb annotations are provided in Appendix C, and Figs. 5 to 7.

5.2.1. Agreement Scores

Agreement scores were calculated for the annotation criteria in Parts 1–3, as not all words had multiple meanings to be annotated. See Figures 3 and 4 in Appendix for the full pairwise annotation scores.

Overall, Korean annotations showed higher reliability (mean Cohen’s $\kappa = 0.74$) than English annotations (mean $\kappa = 0.60$) in average. This pattern may reflect greater consistency in the features of Korean emotion words or, alternatively, additional training adjustments introduced for the Korean word annotation after the annotation of English words.

¹⁰Discrepancies between the number of words annotated and those included in the translational mappings arose from post-hoc corrections to part-of-speech coding and translation mappings.

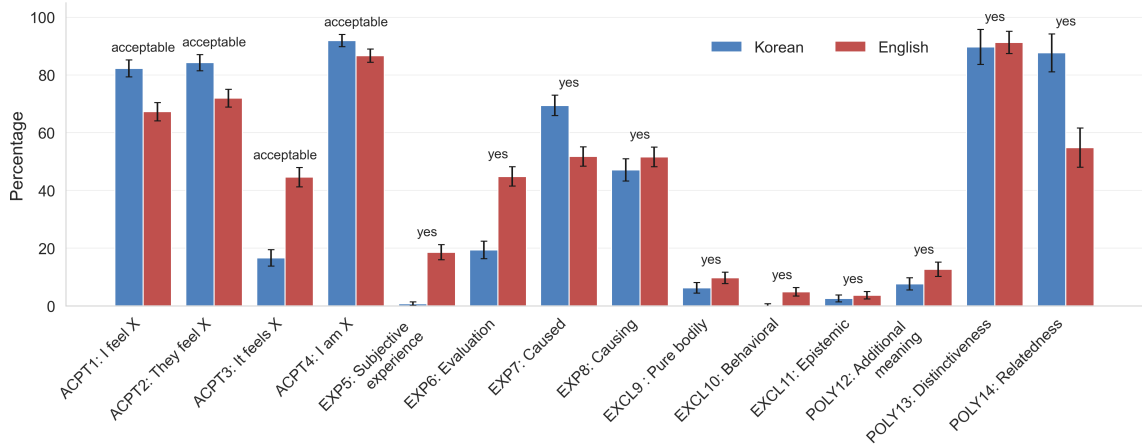


Figure 2: The percentages of annotation values for each annotation criterion for adjectives in Korean and English. The percentages for POLY12 indicate the ratio of words annotated as having more than one meaning in each language. The percentages for POLY13 and POLY14 were calculated among items annotated as having more than one meaning in POLY12. Error bars indicate 95% confidence intervals.

Agreement also varied by criterion and cross-linguistically. Exclusion criteria produced the highest agreement in both languages (mean EXCL9–11: Korean $\kappa = 0.89$, English $\kappa = 0.82$). Causality criteria showed the lowest agreement across both languages (EXP7–8: mean Korean $\kappa = 0.52$, mean English $\kappa = 0.45$), with the “causing” criterion (EXP8) being particularly challenging (Korean $\kappa = 0.46$, English $\kappa = 0.45$). For the acceptability judgments (ACPT1–4), Korean annotators reached substantial agreement (mean $\kappa = 0.73$), while English annotators reached moderate to substantial levels (mean $\kappa = 0.58$). The sub-criteria for subjective experience and evaluation (EXP5–6) achieved higher agreement in Korean (mean $\kappa = 0.76$) than in English (mean $\kappa = 0.47$).

These inter-annotator agreement patterns highlight both the subjective nature of emotion concepts and the theoretical disagreements that surround them. In particular, the relatively low agreement on causality judgments likely reflects the abstract and complex reasoning required. Annotators likely relied on their subjective judgments that relied on personal experiences and their language use. As such, criteria like EXP7 (“Caused”) and EXP8 (“Causing”) should be applied cautiously in making inclusion or exclusion decisions when defining emotion words.

5.2.2. Annotation Trend

A summary of annotation trends for all annotated adjectives is presented in Figure 2. Annotations in the agreement set were adjudicated using majority vote to produce a single combined dataset for analysis.

Words in both languages exhibited higher acceptance in “I am” (ACPT4; Korean: 91.83%, English: 86.85%) than in “I feel” (ACPT1; Korean: 82.13%,

English: 67.23%) or in “They feel” (ACPT2; Korean: 84.28%, English: 71.91%) sentences. Most words were annotated as unacceptable in “It feels” (ACPT3) in Korean (83.36%), but not in English (55.14%).

Responses for causality criteria were more evenly distributed in both languages (EXP7 Korean: 69.18% *yes*, English: 51.56% *yes*; EXP8 Korean: 46.69% *yes*, English: 51.82% *yes*). Exclusion criteria (EXCL9–11) showed a highly consistent rejection rate (around 90%) in both languages.

More English adjectives had multiple meanings annotated (12.68%) than in Korean (7.48%). Cross-linguistic differences also appeared in relatedness judgments (POLY14; Korean 87.37%, English 55.00%), whereas distinctiveness judgments converged (POLY13; Korean 89.47%, English 92.22%). Annotation of additional meaning was correlated positively with bodily meanings in both languages, though more strongly in Korean ($r = 0.46$) than English ($r = 0.26$), suggesting a pattern of metaphorical extension from physical to emotional domains.

6. Future Applications

6.1. Theory-Driven Selection of Words

This dataset is theory-agnostic, as it does not commit to a single theoretical stance on how emotion words should be defined. Researchers can apply their preferred annotation criteria to select emotion words that align with their theoretical perspectives and research needs, enabling the curation of experimental stimuli for both emotion research and computational studies of emotion word semantics.

A conservative approach to selecting words from the dataset applies decision criteria from prior studies. Specifically, (1) a word should be acceptable in

“I feel”, “They feel”, and “I am” sentences, but not in “It feels” context (Part 1; *yes* for ACPT1, ACPT2, ACPT4; *no* for ACPT3); (2) if acceptable in “It feels”, it should express human evaluation rather than an inanimate subject’s experience (Part 2; *no* for EXP5 and *yes* for EXP6); and (3) it should not denote pure bodily sensations, behavioral expressions, or non-emotional epistemic states (Part 3; *no* for EXCL9–11).

Applying these criteria yields 425 single-word emotion adjectives in Korean and 317 in English. Examples of excluded words include *mean* (못되다 (*mottwaeta*) in Korean), which feels unnatural in “I feel X”, and *clueless* (어리둥절하다 (*ǒritungchǒlhata*) in Korean), which denotes a non-emotional epistemic state.

6.2. Emotion Concepts and Metaphors

Moreover, this dataset supports both computational and behavioral research on emotion concepts. Emotion word embeddings could be utilized to conduct computational analyses of lexical gaps, revealing how multilingual language models represent language-specific words. Emotion word embeddings also enable the investigation of foundational questions about emotion structure, such as whether shared dimensions (e.g., valence and arousal) organize the emotion lexicon across languages (Niedenthal, 2008; Yik et al., 2023). High-quality translation mappings provided by the dataset, validated by bilingual speakers of both languages, offer a reliable gold standard for machine learning approaches.

Behavioral studies can further examine how conceptual gaps influence the communication or interpretation of those emotions (Rissman et al., 2023) using the emotion terms in this dataset. Finally, the annotations of multiple meanings in this dataset (POLY12–14) supports cross-cultural investigations of emotion-related metaphors. Comparing the identified metaphorical extensions, researchers can assess the universality versus specificity of emotion-related metaphors (Kövecses, 2003; Sauciuc, 2009).

7. Conclusion

The present dataset includes 1,426 Korean words and 1,397 English words, with many-to-many translational mappings documented between the two languages. Using a feature-based annotation approach, the dataset provides a theory-agnostic set of adjectives and verbs that can be selected as emotion words based on criteria drawn from prior studies. The dataset captures both within-language and cross-linguistic features of emotion words. Beyond documenting linguistic variation, the

dataset provides several principled bases for selecting emotion words and for comparing emotion-related metaphors across languages. It can serve as a resource for linguistic research, computational analysis of emotion concepts, and empirical studies of emotion.

8. Ethical Considerations and Limitations

The dataset does not involve serious ethical considerations, as it contains no personal or identifiable information. All annotators were assigned anonymized coder IDs.

Despite the dataset’s contributions, this study has certain limitations. Although the dataset was constructed to cover more than one language, there are many other languages to investigate for their emotion words. Regarding the annotation, the majority of the data was annotated by a single trained annotator, and the agreement set (about 10% of adjectives and verbs) has three annotators’ judgments adjudicated. This percentage was chosen to prioritize the annotators’ effort put into the large number of emotion words in the dataset with complex annotation criteria that required specific training with adequate domain knowledge. Future research could extend the approach to a broader range of languages and include a larger proportion of items annotated by multiple annotators.

9. Acknowledgements

We would like to thank the annotators for their dedication and hard work. We also thank anonymous reviewers for their valuable feedback and comments.

10. Bibliographical References

- David J Anderson and Ralph Adolphs. 2014. A framework for studying emotions across species. *Cell*, 157(1):187–200.
- Simon Baron-Cohen, Ofer Golan, Sally Wheelwright, Yael Granader, and Jacqueline Hill. 2010. Emotion word comprehension from 4 to 16 years old: A developmental survey. *Frontiers in evolutionary neuroscience*, 2:109.
- Renate Bartsch. 2002. Generating polysemy: Metaphor and metonymy. *Metaphor and metonymy in comparison and contrast*, 20:49–74.
- Luisa Bentivogli, Emanuele Pianta, and Fabio Pianta. 2000. Coping with lexical gaps when building aligned multilingual wordnets. In *Proceed-*

- ings of the Second International Conference on Language Resources and Evaluation, Athens, Greece. European Language Resources Association.
- Katarzyna Bromberek-Dyzman, Rafał Jończyk, Monica Vasileanu, Anabella-Gloria Niculescu-Gorpin, and Halszka Bąk. 2021. Cross-linguistic differences affect emotion and emotion-laden word processing: Evidence from polish-english and romanian-english bilinguals. *International Journal of Bilingualism*, 25(5):1161–1182.
- Cambridge University Press. 2025. [Cambridge english dictionary](#).
- John Cunnison Catford. 1965. *A linguistic theory of translation*, volume 31. Oxford university press London.
- Injae Choi and Sang-Chin Choi. 2002. The effects of korean cultural psychological characteristics on coping styles, stress, and life satisfaction: Centering around cheong and weness. *Korean Journal Of Counseling And Psychotherapy*, 14(1):55–71.
- SJ Choi. 2008. The type and character of feeling verb. *Journal of the Society of Korean Language and Literature*, 58:127–159.
- Gerald L Clore, Andrew Ortony, and Mark A Foss. 1987. The psychological foundations of the affective lexicon. *Journal of personality and social psychology*, 53(4):751.
- Jana Declercq and Lotte van Poppel. 2023. Coding metaphors in interaction: A study protocol and reflection on validity and reliability challenges. *International Journal of Qualitative Methods*, 22:16094069231164608.
- David Dowty. 1991. Thematic proto-roles and argument selection. *language*, 67(3):547–619.
- Beverley Fehr and James A Russell. 1984. Concept of emotion viewed from a prototype perspective. *Journal of experimental psychology: General*, 113(3):464.
- Pragglejaz Group. 2007. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, 22(1):1–39.
- Janosch Haber and Massimo Poesio. 2024. Polysemy—evidence from linguistics, behavioral science, and contextualized language models. *Computational Linguistics*, 50(1):351–417.
- Zellig S Harris. 1954. Distributional structure. *Word*, 10(2-3):146–162.
- Jong-sun Hong and Yeon-ju Jeong. 2009. Establishing the category of emotion verb and classifying emotion verbs. *Korean Linguistics*, 45:387–420.
- Vladimir Ivir. 1977. Lexical gaps: A contrastive view. *Studia Romanica et Anglica Zagradiensia: Revue publiée par les Sections romane, italienne et anglaise de la Faculté des Lettres de l'Université de Zagreb*, (43):167–176.
- Joshua Conrad Jackson, Joseph Watts, Teague R Henry, Johann-Mattis List, Robert Forkel, Peter J Mucha, Simon J Greenhill, Russell D Gray, and Kristen A Lindquist. 2019. Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472):1517–1522.
- Maarten Janssen. 2004. Multilingual lexical databases, lexical gaps, and sim u llda. *International Journal of Lexicography*, 17(2):137–154.
- Duyoung Jeon, Junho Lee, and Cheongtag Kim. 2022. User guide for kote: Korean online comments emotions dataset. *arXiv preprint arXiv:2205.05300*.
- Philip Nicholas Johnson-Laird and Keith Oatley. 1989. The language of emotions: An analysis of a semantic field. *Cognition and emotion*, 3(2):81–123.
- Philip Nicholas Johnson-Laird and Peter Cathcart Wason. 1977. *Thinking: Readings in cognitive science*. CUP Archive.
- Sandra So Hee Chi Kim. 2017. Korean" han" and the postcolonial afterlives of" the beauty of sorrow". *Korean Studies*, pages 253–279.
- Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.
- Zoltan Kovecses. 2010. *Metaphor: A practical introduction*. Oxford university press.
- Zoltán Kövecses. 2012. *Emotion concepts*. Springer Science & Business Media.
- Peter Kuppens, Francis Tuerlinckx, James A Russell, and Lisa Feldman Barrett. 2013. The relation between valence and arousal in subjective experience. *Psychological bulletin*, 139(4):917.
- George Lakoff and Zoltán Kövecses. 1987. The cognitive model of anger inherent in american english. In Dorothy Holland and Naomi Quinn, editors, *Cultural Models in Language and Thought*, pages 195–221. Cambridge University Press.
- Won-Kyung Lee. 2006. The classification and properties of emotive verbs in korean. *Discourse and Cognition*, 13(1):133–161.

- Senyu Li, Bradley Hauer, Ning Shi, and Grzegorz Kondrak. 2024. [Translation-based lexicalization generation and lexical gap detection: Application to kinship terms](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6891–6900, Bangkok, Thailand. Association for Computational Linguistics.
- Charalampos M Liapis, Aikaterini Karanikola, and Sotiris Kotsiantis. 2025. Enhancing sentiment analysis with distributional emotion embeddings. *Neurocomputing*, 634:129822.
- Johann-Mattis List, Simon J Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi, and Robert Forkel. 2018. Clics2: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology*, 22(2):277–306.
- Meichun Liu. 2016. Emotion in lexicon and grammar: lexical-constructural interface of mandarin emotional predicates. *Lingua Sinica*, 2(1):4.
- Tim Lomas. 2018. Experiential cartography and the significance of “untranslatable” words. *Theory & Psychology*, 28(4):476–495.
- Gary Lupyan. 2012. Chapter Seven - What Do Words Do? Toward a Theory of Language-Augmented Thought. *Psychology of Learning and Motivation*, 57:255–297.
- Rick L Morgan and David Heise. 1988. Structure of emotions. *Social psychology quarterly*, pages 19–31.
- National Institute of Korean Language. 2025a. [Korean-english learners' dictionary](#).
- National Institute of Korean Language. 2025b. [Korean-english learners' dictionary](#).
- National Institute of Korean Language. 2025c. [Korean-english learners' dictionary](#).
- National Institute of Korean Language. 2025d. [Korean-english learners' dictionary](#).
- National Institute of Korean Language. 2025e. [Standard korean language dictionary](#).
- NAVER. 2025. [Naver korean-english dictionary](#).
- Paula M Niedenthal. 2008. Emotion concepts. *Handbook of emotions*, 3:587–600.
- Oxford University Press. 2025. [Oxford english dictionary](#).
- In-Jo Park and Kyung-Hwan Min. 2005. Making a list of korean emotion terms and exploring dimensions underlying them. *Korean Journal of Social and Personality Psychology*, 19(1):109–129.
- Rebecca J Passonneau, Ansaif Salieb-Aouissi, Vikas Bhardwaj, and Nancy Ide. 2010. Word sense annotation of polysemous words by multiple annotators. In *LREC*.
- Elizabeth S Paul, Shlomi Sher, Marco Tamietto, Piotr Winkielman, and Michael T Mendl. 2020. Towards a comparative science of emotion: Affect and consciousness in humans and animals. *Neuroscience & Biobehavioral Reviews*, 108:749–770.
- Shahab Raji and Gerard De Melo. 2020. What sparks joy: The affectvec emotion database. In *Proceedings of the web conference 2020*, pages 2991–2997.
- Shin-Young Rhee and Il-Ju Ko. 2013. Measuring a valence and activation dimension of korean emotion terms using in social media. *Science of Emotion and Sensibility*, 16(2):167–176.
- Lilia Rissman, Qiawen Liu, and Gary Lupyan. 2023. Gaps in the lexicon restrict communication. *Open Mind*, 7:412–434.
- Anna Rumshisky and Olga Batiukova. 2008. Polysemy in verbs: systematic relations between senses and their effect on annotation. In *Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics*, pages 33–41.
- James A Russell. 1991. Culture and the categorization of emotions. *Psychological bulletin*, 110(3):426.
- Gabriela-Alina Sauciuc. 2009. The role of metaphor in the structuring of emotion concepts. *Cognitive Semiotics*, 5(1-2):244–267.
- Lothar Schmidt-Atzert and Hyun-Sook Park. 1999. The korean concepts daphada and uulhada: A cross-cultural study of the meaning of emotions. *Journal of Cross-Cultural Psychology*, 30(5):646–654.
- Christine Storm and Tom Storm. 1987. A taxonomic study of the vocabulary of emotions. *Journal of personality and social psychology*, 53(4):805.
- Dong Tang, Yang Fu, Huili Wang, Bo Liu, Anqi Zang, and Tommi Kärrkäinen. 2023. The embodiment of emotion-label words and emotion-laden words: Evidence from late chinese–english bilinguals. *Frontiers in Psychology*, 14:1143064.

Bill Thompson, Seán G Roberts, and Gary Lupyan. 2020. Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, 4(10):1029–1038.

Ebru Türker. 2013. A corpus-based approach to emotion metaphors in Korean: A case study of anger, happiness, and sadness. *Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics Association*, 11(1):73–144.

Anna Wierzbicka. 1999. Emotions across languages and cultures: Diversity and universals. *Cambridge UP*.

Jonathan Winawer, Nathan Witthoft, Michael C. Frank, Lisa Wu, Alex R. Wade, and Lera Boroditsky. 2007. [Russian blues reveal effects of language on color discrimination](#). *Proceedings of the National Academy of Sciences*, 104(19):7780–7785.

Chenggang Wu and Juan Zhang. 2025. Emotion word is an ambiguous concept: dividing, defining, and connecting emotion-label words and emotion-laden words. *Cognition and Emotion*, pages 1–15.

Michelle Yik, Chiel Mues, Irene NL Sze, Peter Kuppens, Francis Tuerlinckx, Kim De Roover, Felicity HC Kwok, Shalom H Schwartz, Maher Abu-Hilal, Damilola Fisayo Adebayo, et al. 2023. On the relationship between valence and arousal in samples across the globe. *Emotion*, 23(2):332.

A. Modifications to Korean word annotations

For the Korean word annotations, the core instructions were provided in English, consistent with those used for the English word annotations. The sentence frames for ACPT1–4 and EXP5–8 were translated into Korean as follows (For EXP5–8, the sentences and subjects were translated identically to those used in ACPT1–4).

For acceptability judgments, the sentence contexts were translated as: “내가/나는 X고 느낀다.” (ACPT1), “그들이/은 X고/(이)라고 느낀다.” (ACPT2), “이것이(은)/저것이(은)/그것이(은) X고/(이)라고 느낀다.” (ACPT3), and “나는 X다.” (ACPT4). As Korean doesn’t have a single word equivalent to the English *it*, three different types of subjects were allowed (*이것* ‘this thing’, *저것* ‘that thing over there’, *그것* ‘that thing’) in ACPT3. If any one of these subject forms yielded an `acceptable` judgment, the item was coded as `acceptable`.

Additionally, annotators were allowed to use whichever word ending felt most natural in each sentence context, as Korean lexical items can appear

with different inflectional endings depending on the context. Additional columns were provided for annotators to indicate the most appropriate word form for each frame. For example, 슬프다 (*sŭlpŭta*) feels natural as 슬프다고 in “내가/나는 X고 느낀다.” (ACPT1) context, and therefore the annotator recorded 슬프다고 in a separate column as a natural word form for ACPT1.

B. Policies on Counting Polysemous Word Entries

Polysemous words are counted as two separate entries under two conditions: (1) has meanings spanning multiple parts of speech (e.g., *worry* as a noun and verb were counted as two separate entries.), or (2) one of the meanings having translations in one language but not in the other (e.g., *needy* was counted as two entries, as one sense had a Korean equivalent and another did not). This approach was adopted to ensure that the total number remained interpretable; otherwise, the sum of the number of words across categories would not correspond to the total number of words in each language. Tables 3 to 5 summarizes the number of words overlapping across two conditions.

Table 3: Overlaps Across Parts of Speech in Korean

Translation		Noun	Verb	Adjective
With	Noun	–	0	0
	Verb	–	–	1
Without	Noun	–	0	–
	Verb	–	–	0

Table 4: Overlaps Across Parts of Speech in English

Translation		Noun	Verb	Adjective
With	Noun	–	35	14
	Verb	–	–	2
Without	Noun	–	1	–
	Verb	–	–	0

Table 5: Overlaps Across Translational Status Within Parts of Speech

Part of Speech	Korean	English
Noun	2	4
Verb	2	3
Adjective	3	11

C. Verb Annotation Results

C.1. Agreement Scores

The pairwise kappa scores, along with the aggregated scores across all pairs, of the verb agreement set are presented in Figs. 5 and 6. Overall, annotations of Korean verbs (mean Cohen's $k = 0.82$; aggregated across all annotation criteria and coders) showed higher reliability than annotations of English verbs (mean Cohen's $k = 0.69$; aggregated across all annotation criteria and coders), which mirrors the agreement pattern of adjective annotations.

Across individual criteria, questions in the exclusion criteria produced higher agreement (EXCL9–11; Korean $k = 0.88$, English $k = 0.83$). In contrast, annotations on causality criteria exhibited lower and more variable agreement, particularly for Korean verbs across the questions (EXP7–8; Korean $k = 0.72$, English $k = 0.48$). While the agreement on the “causing” question was at an “almost perfect” level (mean $k = 0.92$), the “caused” question (EXP7) produced lower agreement (mean $k = 0.52$) for Korean verbs. This pattern contrasts with that observed for adjectives, where “causing” criteria were challenging for both Korean and English annotators (Korean $k = 0.46$, English $k = 0.45$). Overall, agreement on exclusion criteria was similarly high for both verbs and adjectives, whereas agreement on causality annotations was higher for verbs than for adjectives.

The agreement patterns of verbs also highlight the subjective nature of emotion concepts, particularly for causality judgments. At the same time, higher agreement on causality annotations for verbs indicates relatively less variation for verbs. The very high level of agreement for “causing” criterion suggests that the causative properties of Korean emotion verbs may be more consistently interpretable than those of English emotion verbs.

C.2. Annotation Trend

A summary of annotation trends is in Figure 7. Annotations in the agreement set were adjudicated using majority vote to produce a single combined dataset for analysis.

In the causality criteria, the general trend of Korean adjectives showing higher acceptance for EXP7 but lower acceptance for EXP8 than English adjectives remains the same for verbs as well, but the difference between the two criteria was larger (EXP7 acceptance: Korean 68.51%, English 51.15%; EXP8 acceptance: Korean 7.23%, English 43.08%). Exclusion criteria (EXCL9–11) also showed highly consistent rejection in both languages, achieving around 90% rate of `no` response.

Unlike the discrepancy across languages for

adjectives, verbs had a similar rate of additional emotion-related meanings being annotated (POLY12; 7.76% in Korean, 7.83% in English). This rate is similar to that found for Korean adjectives, but remains lower than the rate for English adjectives. These findings may suggest a relatively higher degree of polysemy among English adjectives.



Figure 3: Agreement scores across annotation criteria for the English adjective agreement set.

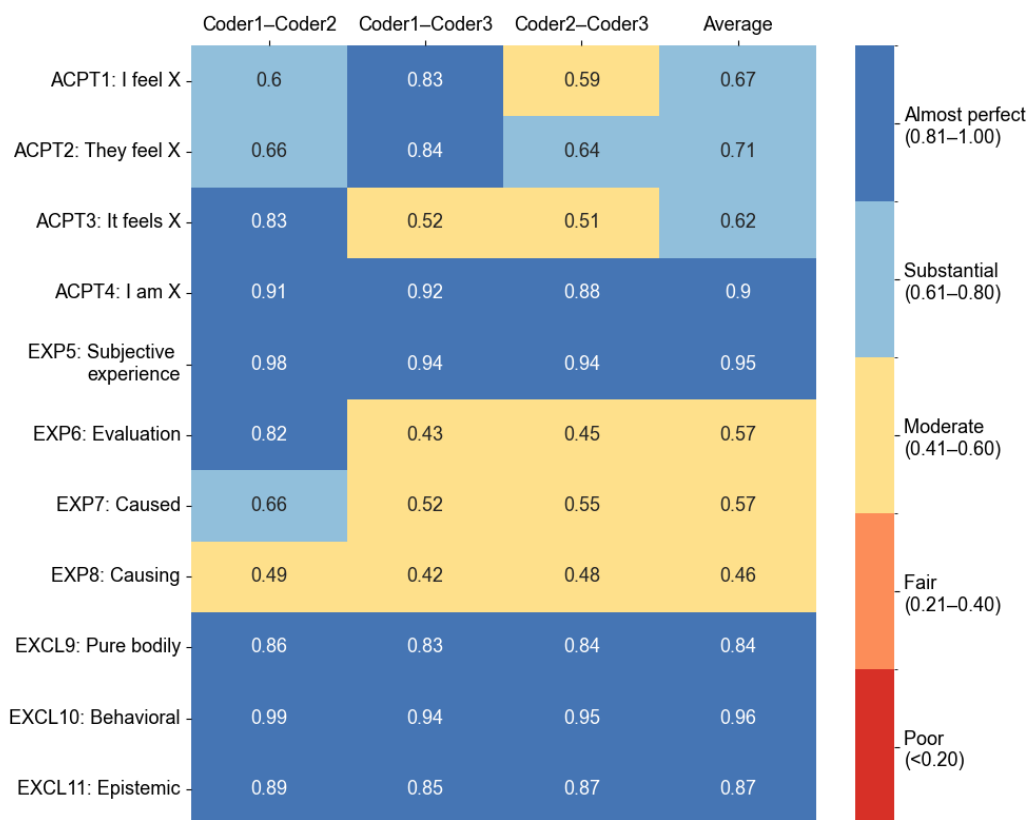


Figure 4: Agreement scores across annotation criteria for the Korean adjective agreement set.

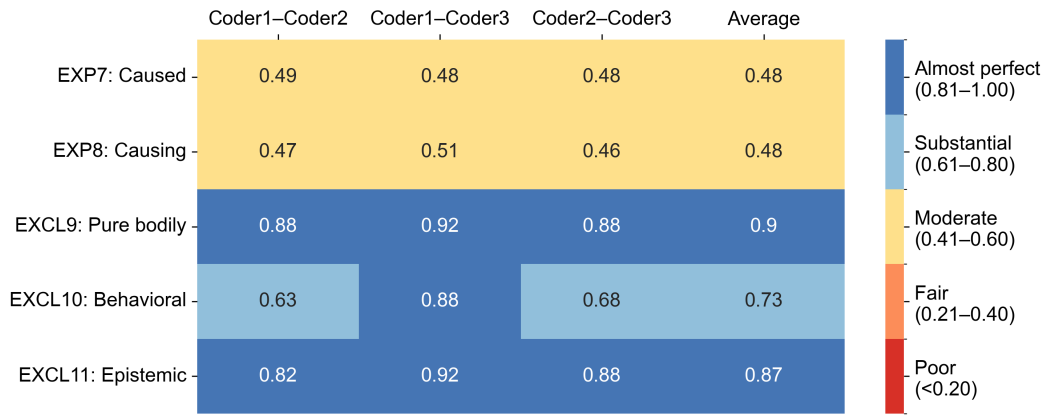


Figure 5: Agreement scores across annotation criteria for the English verb agreement set.

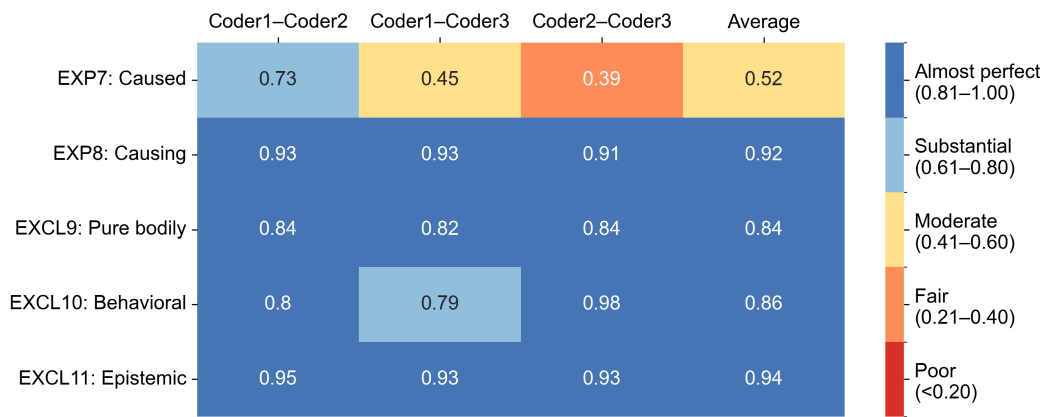


Figure 6: Agreement scores across annotation criteria for the Korean verb agreement set.

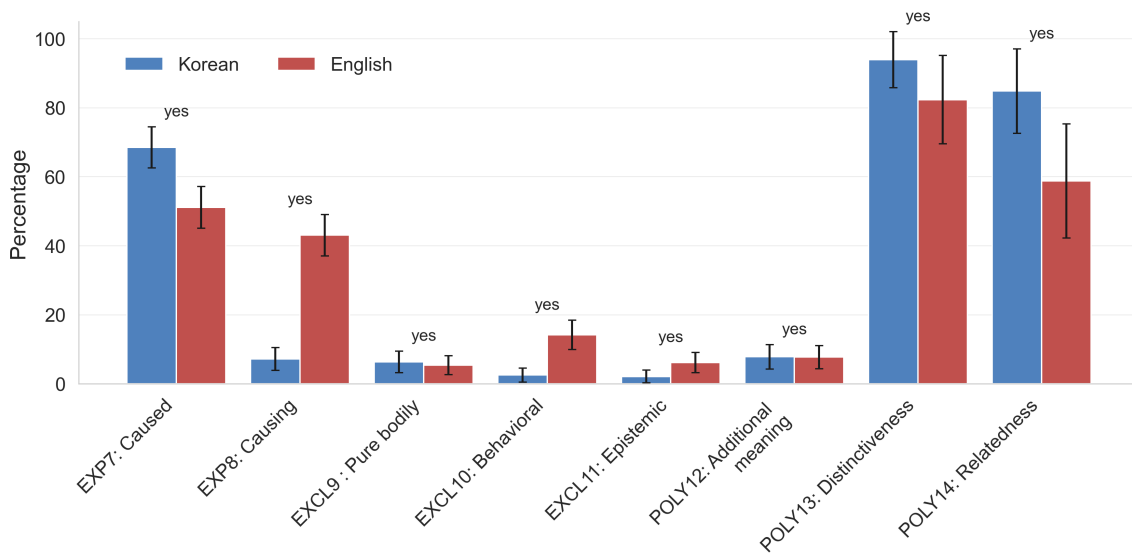


Figure 7: The percentages of annotation values for each criterion for verbs in Korean and English. The percentages for POLY12 indicate the ratio of words annotated as having more than one meaning in each language. The percentages for POLY13 and POLY14 were calculated among items annotated as having more than one meaning in POLY12. Error bars indicate 95% confidence intervals.