Cross-Cultural Emotion Concept Representation: A Comparison of English, Korean, and Large Language Model Representations

Mijin Kwon[1], Sean Dae Houlihan[1, 2], Jonathan Phillips[1, 2, 3]
[1] Department of Psychological and Brain Sciences, Dartmouth College
[2] Program in Cognitive Science, Dartmouth College
[3] Department of Philosophy, Dartmouth College

Abstract

Each person develops a unique emotional landscape shaped by their experiences and linguistic-cultural contexts, partly personal and partly shared with others. This enables personally unique emotional experiences while maintaining shared understanding. This work aims to advance a framework for investigating what's shared and distinct across individuals, beginning with linguistic communities as an essential level of analysis, using English and Korean speakers as our case study. We examined how emotion concept representations differ between English and Korean speakers using representational similarity analysis and network analysis. English and Korean speakers' judgments of pairwise similarity between 57 emotion concepts evidenced both substantial shared structure and language-specific patterns (Spearman's ρ = 0.72, indicating 48% unshared variance). While valence emerged as a key organizing dimension in both languages, network analyses with strength centrality showed distinct patterns for each language. First, the Korean emotion concept network demonstrated higher strength centrality across all emotion concepts than the English network, indicating higher interconnectedness between concepts. Second, high-centrality emotions were predominantly negative in both languages but formed language-specific local networks with different sets of neighboring concepts. The statistics of language usage encode a substantial part of the conceptual structure of emotion, enabling large language models to capture aspects of human emotion. Despite their advanced multilingual capabilities, GPT4-o and Claude-3.5 showed stronger alignment with English speakers' representations, regardless of prompt language. These findings demonstrate that while languages reflect common principles in emotion representation, they shape distinct patterns, with implications for cross-linguistic/cultural emotion understanding and AI system development.

Keywords: emotion concepts; cross-linguistic comparison; network analysis; large language models

Introduction

People possess rich mental models of emotions that serve as foundations for understanding and communicating emotional experiences (Breithaupt et al., 2022; Houlihan et al., 2023; Ong et al., 2015, 2019; Saxe & Houlihan, 2017). While these models contain unique elements shaped by one's life experience, they simultaneously incorporate substantial shared components derived from biological constraints, affective primitives, and common social factors such as language and culture. This integration enables humans to have distinct emotional experiences while maintaining emotional understanding with others. Given that emotions cannot be directly observed in others but must be inferred, language functions as a crucial medium for learning, conceptualizing, and communicating about emotional states (Lindquist, 2017, 2021; Nook et al., 2017; Pritzker et al., 2020; Satpute et al., 2020). Consequently, the linguistic frameworks people use likely create systematic variations in how speakers of different languages conceptualize emotions (Jackson et al., 2019).

Indeed, research examining cross-linguistic differences in emotion concepts has established that speakers of different languages conceptualize and categorize emotions in somewhat different ways (Jackson et al., 2019; Majid, 2012; Pavlenko, 2014; Wierzbicka, 2009). However, beyond documenting differences, we know less about how emotion concepts are systematically organized and interrelated across different languages, and what these organizational patterns reveal about emotional processing in different linguistic communities.

Recent methodological approaches have demonstrated that examining relationship patterns between emotion concepts provides rich insights into latent conceptual organization (Brooks et al., 2019; Brooks & Freeman, 2018; Kwon et al., 2022; Skerry & Saxe, 2015). This relational approach is particularly valuable for cross-linguistic research, as it compares the patterns of similarity between concepts rather than relying solely on finding lexically equivalent translations. By mapping these relationship patterns between languages, we can better understand both commonality and divergence in how emotions are conceptualized across linguistic communities.

Most of the work studying the similarity structure of emotion concepts has focused on English speakers, leaving open the question of how representations might vary across individuals who share different languages and cultures. Understanding such variation is crucial not only for theoretical models of emotion but also for developing culturally-sensitive artificial intelligence systems that can effectively engage in emotional communication across different linguistic contexts.

The present work compared the representational structures of emotions in English and Korean, serving as a case study aimed at developing a framework for examining how language and culture affect emotion concept representations. Our approach aims to characterize the specific ways in which emotional understanding depends on cultural and linguistic contexts, from global organizational principles to local relationships between specific emotions, to how these

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

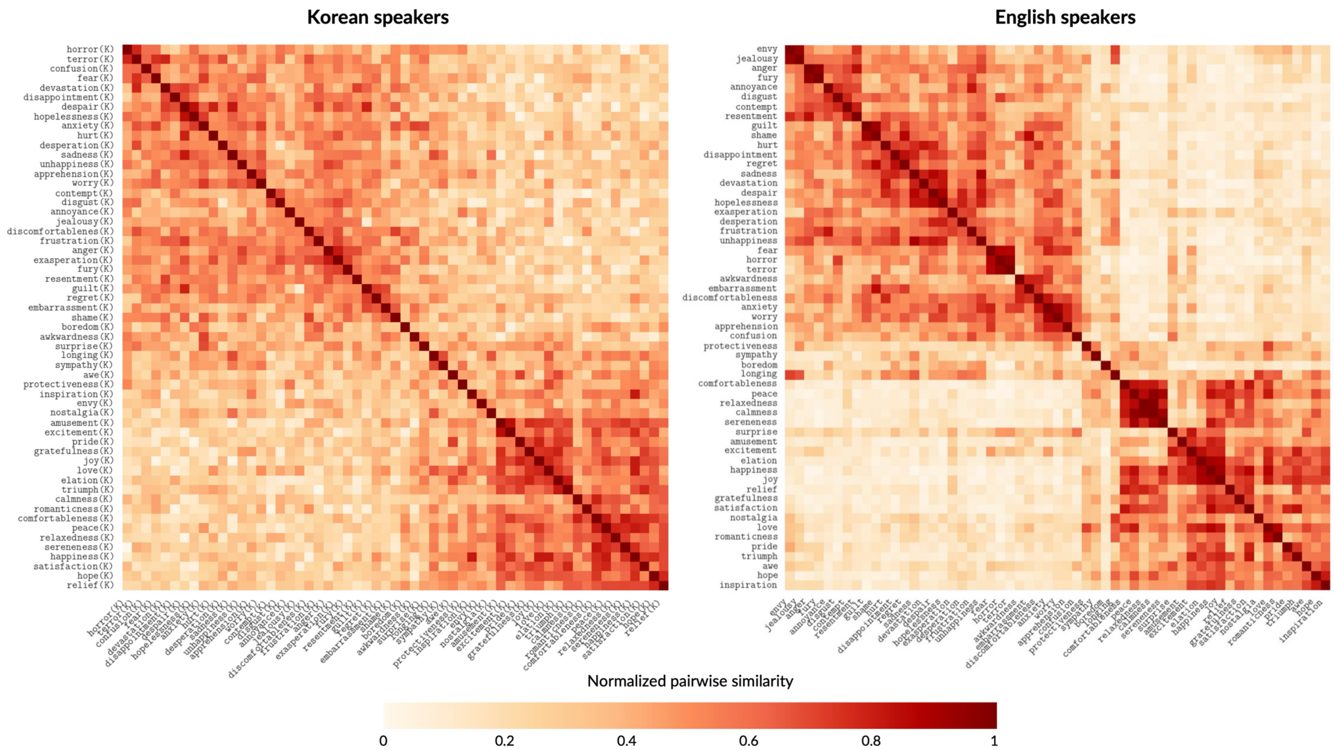

Figure 1: Pairwise similarity matrix with hierarchical clustering and min-max normalization for 57 emotions from the pairwise similarity judgment task from Korean (left) and English (right) speakers.

commonalities and differences are reflected in large language models (LLMs). By testing whether LLMs trained primarily on English data can capture these language-specific patterns, we examine the challenges and limitations artificial systems face in representing emotion concepts across different linguistic contexts. While we focus here solely on Korean and English, our aim is to develop and refine a framework that can be extended to examine emotion concept representations across a broader range of languages and cultures, and ultimately contextual and individual differences.

Methods

Study Materials and Participants

We selected 57 emotion concepts frequently used across multiple emotion research studies (Cowen & Keltner, 2017; Gross & Levenson, 1995; O'Reilly et al., 2016; Scherer, 2005; Skerry & Saxe, 2014; Tottenham et al., 2009; Watson et al., 1988). Concepts are selected in English first for a previous study (Kwon et al., 2022) and translated into Korean through a multi-stage process. First, 25 translators fluent in both languages independently provided translations (Average agreement across all emotions: *66.5%*; Krippendorff's alpha = *0.53*). Four additional bilingual experts then reviewed translations using accuracy verification and back-translation approaches, followed by targeted resolution of cases where concepts in either language had multiple potential matches in the other. Final translations were determined based on majority agreement among these experts.

We recruited 839 English speakers through Amazon Mechanical Turk and Prolific (M_{age} = 40.01, SD_{age} = 13.69, N_{female} = 427) and 369 Korean speakers through a survey panel in Korea (M_{age} = 40.89, SD_{age} = 12.70, N_{female} = 168). All participants were native speakers of their respective languages.

Similarity Judgment Task

Participants rated pairs of emotion concepts (Supplementary methods) on a scale from 0 (completely different) to 100 (very similar). English and Korean speakers rated 30 and 60 randomly selected pairs, respectively. The overall similarity value for each concept pair was calculated by taking the mean of all participants' ratings for that pair.

LLM Data Collection

To examine how well LLMs capture language-specific patterns in emotion concept representation, we collected similarity ratings from GPT4-o and Claude-3.5, two widely used "foundation" LLMs, using equivalent prompts for the same similarity judgment in both languages. For each concept pair, we collected multiple ratings (*n=10*) using a temperature setting of 0.95, with the final rating calculated as the mean.

Representational Similarity Analysis

To analyze relationships between emotion concept representations across languages and LLMs, we employed representational similarity analysis (RSA; Kriegeskorte et al., 2008). Our approach builds on principles that position

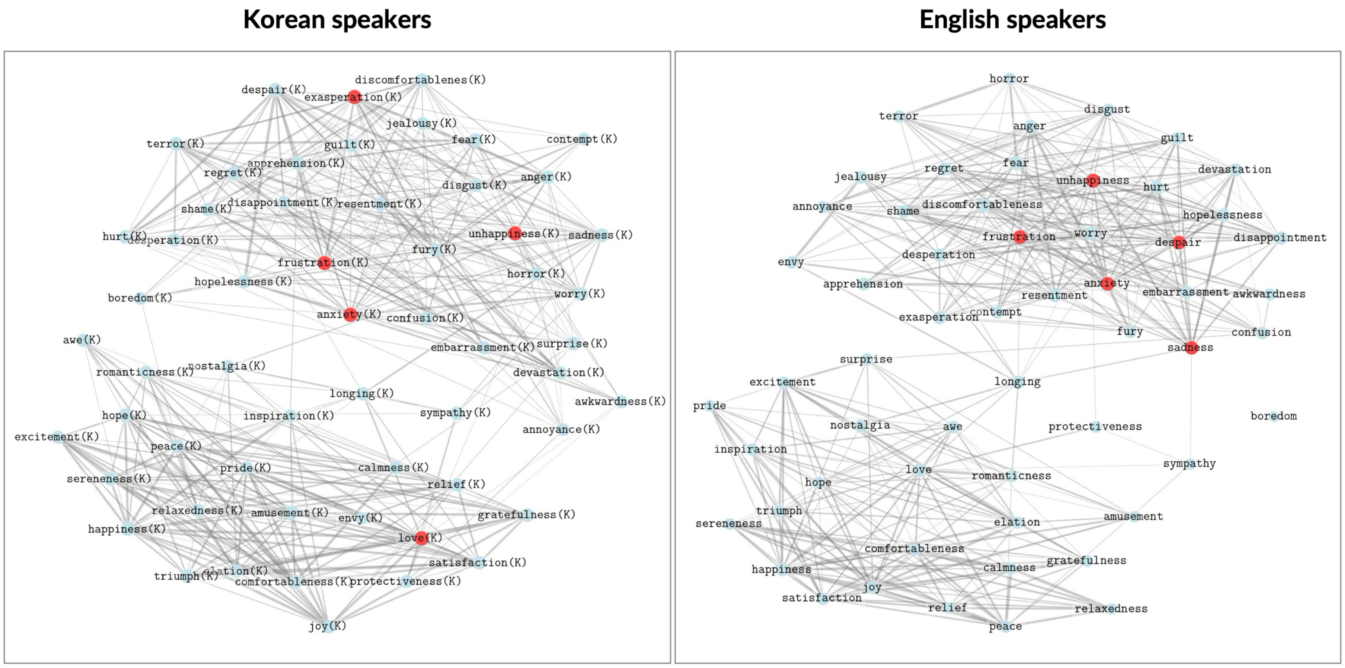

Figure 2: Network graph of emotion concepts in English and Korean based on pairwise similarity ratings. Red circles indicate emotion concepts with the highest strength centrality in each network.

similarity judgments as fundamental to accessing mental representations (Roads & Love, 2024), revealing broader - organizational principles than single-task paradigms by capturing the underlying conceptual structure across contexts rather than task-specific responses.

RSA enables the comparison of different representational spaces through second-order isomorphism without requiring perfect one-to-one mapping. This is crucial when comparing emotion concepts across languages, where we assume minimal equivalence between translation pairs (e.g., "sadness" and "슬픔") as necessary anchor points, while acknowledging these concepts occupy different positions within their respective networks.

For RSA, we constructed similarity matrices from pairwise ratings for both language groups and LLM responses, after rescaling the raw similarity ratings to 0-1. We, then, calculated Spearman's rank correlations with permutation tests (10,000 permutations) to assess shared structure between languages and LLM alignment.

Results

Representational Structure Comparisons

Comparison of similarity matrices between English and Korean speakers revealed substantial shared structure in emotion concept representations (*Spearman's* $\rho = 0.72$, $p < 0.001$, 10,000 permutations for significance testing), with 48% unshared variance pointing to language-specific patterns. Hierarchical clustering (Ward method) revealed valence as a key organizing dimension in both languages, with positive and negative emotions forming distinct clusters (Fig. 1).

We further examined emotion pairs showed the most consistent and most divergent similarity ratings between languages by calculating absolute differences between normalized similarity values across two groups and identified the top 10 most similar and different pairs (Table 1). The most consistent similarity ratings across languages emerged for pairs rated either distinctly different (e.g., fear-nostalgia, similarity of 0.106 in Korean and 0.107 in English) or moderately similar (e.g., peace-triumph, similarity of 0.502 in both), rather than pairs rated as highly similar. For emotion pairs where similarity ratings differed most between languages, different patterns emerged: Korean speakers indicated higher similarity between emotions with different valences or arousal levels (e.g., envy-relaxedness, boredom-horror, and happiness-regret), assigning moderate similarity while English speakers rated them as rather dissimilar. In contrast, English speakers indicated higher similarity ratings between emotions sharing similar valence (e.g., guilt-shame, awe-joy), rating these as highly or moderately similar while Korean speakers rated them as moderately similar or dissimilar.

Network Analysis

To further examine the global and local structures of emotion concept representations of Korean and English speakers, we transformed the similarity matrices into weighted networks, with emotion concepts as nodes and normalized similarity ratings as edge weights. This network representation allowed us to analyze both the overall organization of emotion concepts and their specific local relationships (e.g., which emotions are most conceptually similar to "joy" in each language). To focus on the most robust relationships, we thresholded the normalized edge weights at the 75th percentile, retaining only the top quartile of weighted connections in our analysis.

Network analysis reinforced the role of valence as a key organizing dimension in both languages (Fig. 2). However, several emotions bridged these clusters by connecting to both positive and negative emotions: “longing” and “sympathy” in both networks, “boredom”, “nostalgia”, and “annoyance” in the Korean network, and “surprise” and “protectiveness” in the English network.

Table 1: Conceptually more similar emotion concept pairs for each language group.

Similar between two groups			
Emotion Concept Pairs	*Kor*	*Eng*	*Diff*
peace - triumph	0.502	0.503	0.000
fear - nostalgia	0.107	0.107	0.000
disappointment - relief	0.146	0.145	0.000
relaxedness - sympathy	0.310	0.310	0.001
fury - hopelessness	0.428	0.428	0.001
Korean speakers			
Emotion Concept Pairs	*Kor*	*Eng*	*Diff*
envy - relaxedness	0.576	0.063	0.512
excitement - relaxedness	0.632	0.138	0.494
envy - relief	0.506	0.050	0.455
boredom - horror	0.448	0.025	0.423
anxiety - peace	0.476	0.054	0.422
English speakers			
Emotion Concept Pairs	*Kor*	*Eng*	*Diff*
guilt - shame	0.390	0.917	0.526
awe - joy	0.189	0.704	0.514
envy - jealousy	0.471	0.936	0.464
disappointment - regret	0.400	0.836	0.436
envy - resentment	0.335	0.763	0.427

Strength Centrality

Next, we calculated strength centrality, which represents the sum of weighted edges for each emotion concept. Strength centrality is suitable for our similarity-based networks as it captures immediate relationships without assumptions about indirect pathways or information flow, while incorporating both the number and strength of similarity relationships between emotion concepts (Bringmann et al., 2019). Here, high strength centrality indicates that an emotion has higher similarity with many other emotions in the network, while low strength centrality indicates that an emotion is more distinct, sharing lower similarity with other emotions.

Strength centrality analysis revealed differences in emotion network at both global and local levels, while also identifying common patterns across languages. At the global level, the Korean network showed consistently higher strength centrality values compared to the English network, indicating greater overall interconnectedness between emotion concepts (Fig. 3). In both networks, negative emotions dominated the high-centrality concepts, with “love” as the only positive exception (Tables 2 and 3). Many of these high-centrality emotions overlapped between languages—“unhappiness”, “frustration”, and “anxiety” showed high centrality in both networks. However, despite this overlap in

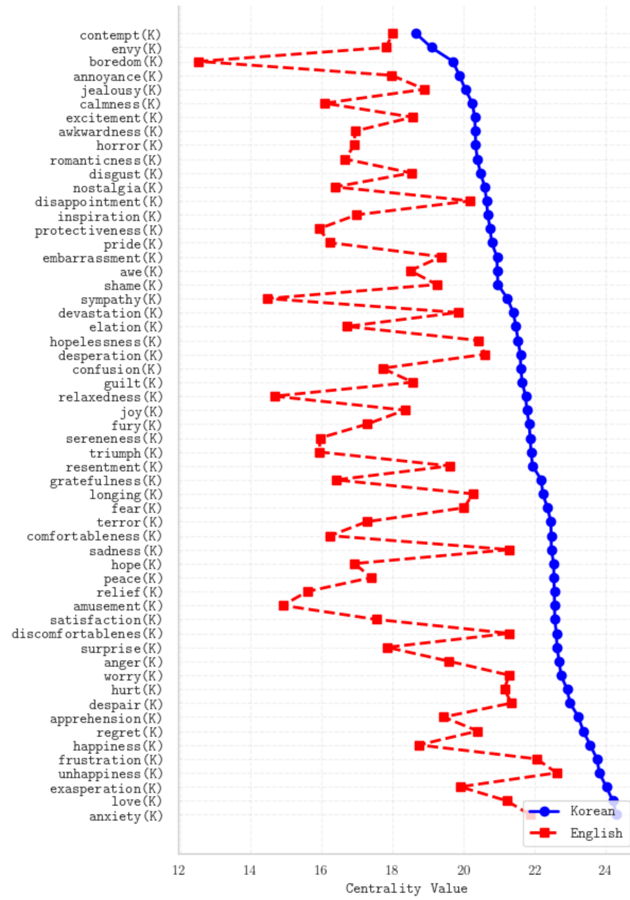

Figure 3; Strength centrality values across all emotion concepts in the Korean (blue) and English (red) emotion concept networks. Ordered by centrality values of Korean emotion concepts for visualization.

central emotions, their local network structures differed markedly: for example, “unhappiness” clustered with active negative emotions (e.g., anger, exasperation) in Korean, while in English, it clustered with depressive states (e.g., despair, hopelessness, devastation).

At the other end of the centrality spectrum, we examined emotions with the lowest strength centrality—those that are most dissimilar to others in the network, being more distinct and differentiated. The lowest five emotions in English were “boredom”, “sympathy”, “relaxedness”, “amusement”, and “relief”, while in Korean they were “contempt”, “envy”, “boredom”, “annoyance”, and “jealousy”. Except for “boredom”, which showed low centrality in both networks, the remaining low-centrality emotions were primarily positive states in English but social or interpersonal emotions in Korean.

Network Structures of Similar and Different Emotion Pairs across Languages

We next examined whether differences in network relationships could help explain cross-linguistic variation in similarity ratings (identified in the representational similarity analysis; Table 1). For each pair that is more similar in

Korean or English, we analyzed their network relationships in two ways: by counting overlapping concepts among their top 10 nearest neighbors and by checking whether they were each other's nearest neighbors. When a pair showed higher similarity in one language, we expected to find stronger network relationships (more overlapping neighbors and/or direct neighbor status) in that language's network.

Indeed, pairs rated more similar by Korean speakers showed more overlapping neighbors and direct neighbor status in the Korean network (with "anxiety-peace" as a notable exception showing no direct or indirect connections). Similarly, pairs rated more similar by English speakers showed stronger connections exclusively in the English network, either through overlapping neighbors or direct connections. These findings show that cross-linguistic differences in similarity ratings are reflected in how these emotions relate to other emotions in each language's network.

Table 2: Nearest neighbors of high centrality emotion concepts in the English network.

Emotion	Centrality	5 Nearest Neighbors
unhappiness	22.624	sadness, despair, hopelessness, devastation, disappointment
frustration	22.055	annoyance, disappointment, exasperation, fury, confusion
anxiety	21.870	worry, discomfortableness, apprehension, fear, unhappiness
despair	21.344	hopelessness, sadness, unhappiness, desperation, devastation
sadness	21.290	unhappiness, despair, hopelessness, devastation, disappointment

Table 3: Nearest neighbors of high centrality emotion concepts in the Korean network.

Emotion	Centrality	5 Nearest Neighbors
anxiety(K)	24.306	worry, fear, hopelessness, awkwardness, surprise
love(K)	24.202	happiness, joy, gratefulness, elation, amusement
exasperation(K)	24.019	anger, fury, frustration, disgust, discomfortablenes
unhappiness(K)	23.825	anger, exasperation, anxiety, worry, sadness
frustration(K)	23.757	despair, exasperation, discomfortablenes, boredom, devastation

Comparison with Foundation Language Model Representation

Lastly, we tested how GPT4-o and Claude-3.5, two of the most widely used language models, represent emotion concept similarities and whether they could capture the language-specific patterns in emotion concept representation identified in our human data.

Results showed both GPT4-o and Claude-3.5's English-prompted responses showed a higher correlation with English speakers' ratings ($r = 0.91$ and 0.93, $ps < 0.001$) than with Korean speakers' ratings ($r = 0.71$ and 0.72, $ps < 0.001$). More notably, when prompted in Korean, both models' responses showed a stronger correlation with English speakers' ratings ($r = 0.88$ and 0.88, $ps < 0.001$) than with Korean speakers' ratings ($r = 0.75$ and 0.77, $ps < 0.001$), and were also more similar to its English-prompted responses ($r = 0.90$ and 0.93, $ps < 0.001$; all correlations calculated with 10,000 permutations).

To formally test these relationships, we compared three regression models predicting LLMs' similarity ratings with English and Korean human similarity ratings: a full model with both similarity ratings as predictors, and two reduced models with either English-only or Korean-only ratings (Table 4). In the full model, for both English- and Korean-prompted responses, English speakers' ratings were consistently stronger predictors (GPT4-o: $\beta = 0.908$ and $\beta = 0.697$; Claude-3.5: $\beta = 0.960$ and $\beta = 0.680$) than Korean speakers' ratings (GPT4-o: $\beta = 0.135$ and $\beta = 0.366$; Claude-3.5: $\beta = 0.116$ and $\beta = 0.438$). Adding English ratings to Korean-only models showed larger improvements in model fit (GPT4-o: $\chi^2 = 1845.792$ and 1039.001; Claude-3.5: $\chi^2 = 2052.098$ and 1011.202; all $ps < 0.001$) compared to adding Korean ratings to English-only models (GPT4-o: $\chi^2 = 39.213$ and 194.762; Claude-3.5: $\chi^2 = 31.158$ and 276.044; all $ps < 0.001$). These patterns indicate that both models' representations of emotion concepts mainly reflect that of English speakers, even when prompted in Korean, suggesting the limited ability to capture language-specific patterns in emotion concept representation.

Discussion

Common Principles and Language-Specific Patterns

Our findings reveal both shared and distinct patterns in how English and Korean speakers represent and organize emotion concepts. The substantial correlation between similarity matrices points to common principles in emotion concept organization across languages, while the considerable unshared variance indicates the presence of language-specific patterns.

Network analyses revealed that while valence emerged as a fundamental organizing dimension in both languages, forming distinct clusters of positive and negative emotions, strength centrality analysis uncovered important cross-linguistic differences in network structure.

First, the Korean network showed consistently higher strength centrality values across emotion concepts (Fig. 3), indicating greater overall interconnectedness between emotions. In both networks, negative emotions dominated the high-centrality concepts (with "love" in the Korean network as the only exception). This pattern indicates that negative emotions might be more granularly represented in both languages, where similar contexts and situations can evoke multiple, related negative emotions. Counter-intuitively, this granularity leads to higher similarity between negative emotions, while positive emotions, being less granular in their representation, remain more distinct from one another, resulting in lower centrality values. Furthermore, while several of these high-centrality emotions overlapped between networks (e.g., anxiety, unhappiness, frustration), these concepts showed different sets of nearest neighbors,

Table 4. Comparison between full and reduced regression models predicting LLM similarity ratings.

Model Results	GPT4-o similarity ratings (KOR)	GPT4-o similarity ratings (ENG)	Claude-3.5 similarity ratings (KOR)	Claude-3.5 similarity ratings (ENG)
Model R^2				
Full model	0.771	0.846	0.784	0.866
Korean similarity ratings-only model	0.561	0.512	0.592	0.513
English similarity ratings-only model	0.741	0.843	0.743	0.863
Likelihood ratio (adding predictor)				
Korean similarity ratings	194.762***	39.213***	276.044***	31.158***
English similarity ratings	1039.001***	1845.792***	1011.202***	2052.098***
Standardized coefficients				
Korean similarity ratings	0.366	0.135	0.438	0.116
English similarity ratings	0.697	0.908	0.680	0.960
Intercept	-0.098	0.013	-0.125	-0.067

*** $p < .001$

revealing language-specific patterns in how even common emotions are related to others.

The networks also differed in their low-centrality emotions, representing the most distinct and dissimilar concepts in each network. The English network's low-centrality emotions primarily included positive states while the Korean network's included social or interpersonal emotions.

Also notably, when emotion pairs are rated differently across languages, these differences are systematically reflected in their network relationships with other emotions. These network patterns offer a new way to understand cross-linguistic differences in emotion concept representations. For example, an emotion concept in one language might be better understood through multiple emotion concepts in another language that together share similar relationship patterns.

Limitations in Cross-Linguistic Representation of Current AI Models

Our analysis of GPT4-o and Claude-3.5 showed remarkably high correspondence with human emotion concept representations, particularly with English speakers. However, despite their advanced multilingual capabilities, both models' representations showed stronger alignment with English even when prompted in Korean, revealing important limitations in capturing language- and culture-specific emotion representations. These results reveal an important gap in current AI systems: the ability to process multiple languages does not necessarily translate to capturing language-specific patterns in how concepts like emotions are organized and understood.

These limitations raise important implications for cross-cultural AI development. While English dominance in AI training is well-documented, our findings reveal its deeper impact on conceptual representation: the underlying semantic structures remain influenced by the dominant training language despite surface-level multilingual capabilities.

Our comparison between human and LLM emotion representations provides a novel method for evaluating cultural sensitivity in AI systems beyond traditional accuracy metrics. Unlike other AI capabilities that don't necessarily need to mirror human performance, emotion understanding requires alignment with human conceptual structures, particularly for systems designed to interact with humans in emotionally nuanced ways across cultures.

Future Directions

Building on our current findings, we plan to extend this research in several important directions. First, we will expand beyond our English-first concept selection approach to incorporate emotion concepts uniquely present in Korean culture to investigate a more complete picture of each language's emotion conceptual space.

Second, we will examine how these representations function in specific contexts to understand whether the language-specific patterns we observed are stable across contexts or whether context sensitivity itself differs across languages.

Lastly, we will examine non-English language models trained primarily in Korean and other languages to determine whether the English-alignment patterns stem primarily from training data distribution or reflect fundamental challenges in capturing language-specific conceptual structures.

Supplementary Materials

Supplementary methods are available at: https://osf.io/jspq4/

References

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial Breithaupt, F., Li ,Binyan, & and Kruschke, J. K. (2022). Serial reproduction of narratives preserves emotional appraisals. *Cognition and Emotion*, *36*(4), 581–601. https://doi.org/10.1080/02699931.2022.2031906

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, *128*(8), 892–903. https://doi.org/10.1037/abn0000446

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences*, *116*(32), 15861–15870. https://doi.org/10.1073/pnas.1816408116

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour*, *2*(8), 581–591. https://doi.org/10.1038/s41562-018-0376-6

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, *9*(1), 87–108. https://doi.org/10.1080/02699939508408966

Houlihan, S. D., Kleiman-Weiner, M., Hewitt, L. B., Tenenbaum, J. B., & Saxe, R. (2023). Emotion prediction as computation over a generative theory of mind. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *381*(2251), 20220047. https://doi.org/10.1098/rsta.2022.0047

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://www.frontiersin.org/article/10.3389/neuro.06.004.2008

Kwon, M., Wager, T., & Phillips, J. (2022). Representations of emotion concepts: Comparison across pairwise, appraisal feature-based, and word embedding-based similarity spaces. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/8vj3d366

Lindquist, K. A. (2017). The role of language in emotion: Existing evidence and future directions. *Current Opinion in Psychology*, *17*, 135–139. https://doi.org/10.1016/j.copsyc.2017.07.006

Lindquist, K. A. (2021). Language and Emotion: Introduction to the Special Issue. *Affective Science*, *2*(2), 91–98. https://doi.org/10.1007/s42761-021-00049-7

Majid, A. (2012). Current Emotion Research in the Language Sciences. *Emotion Review*, *4*(4), 432–443. https://doi.org/10.1177/1754073912445827

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*, 881–889. https://doi.org/10.1038/s41562-017-0238-7

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141–162. https://doi.org/10.1016/j.cognition.2015.06.010

Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational Models of Emotion Inference in Theory of Mind: A Review and Roadmap. *Topics in Cognitive Science*, *11*(2), 338–357. https://doi.org/10.1111/tops.12371

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., Bölte, S., Baron-Cohen, S., & Lundqvist, D. (2016). The EU-Emotion Stimulus Set: A validation study. *Behavior Research Methods*, *48*(2), 567–576. https://doi.org/10.3758/s13428-015-0601-4

Pavlenko, A. (2014). *The Bilingual Mind: And What it Tells Us about Language and Thought*. Cambridge University Press. https://doi.org/10.1017/CBO9781139021456

Pritzker, S. E., Fenigsen, J., & Wilce, J. M. (2020). *The Routledge handbook of language and emotion*. Routledge, Taylor and Francis Group.

Roads, B. D., & Love, B. C. (2024). Modeling Similarity and Psychological Space. *Annual Review of Psychology*, *75*(1), 215–240. https://doi.org/10.1146/annurev-psych-040323-115131

Satpute, A. B., Nook, E. C., & Cakar, M. E. (2020). The Role of Language in the Construction of Emotion and Memory: A Predictive Coding View. In R. D. Lane & L. Nadel (Eds.), *Neuroscience of Enduring Change* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780190881511.003.0004

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a Bayesian model of theory of mind. *Current Opinion in Psychology*, *17*, 15–21. https://doi.org/10.1016/j.copsyc.2017.04.019

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Skerry, A. E., & Saxe, R. (2014). A Common Neural Code for Perceived and Inferred Emotion. *Journal of Neuroscience*, *34*(48), 15997–16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014

Skerry, A. E., & Saxe, R. (2015). Neural Representations of Emotion Are Organized around Abstract Event Features. *Current Biology*, *25*(15), 1945–1954. https://doi.org/10.1016/j.cub.2015.06.009

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wierzbicka, A. (2009). Language and Metalanguage: Key Issues in Emotion Research. *Emotion Review*, *1*(1), 3–14. https://doi.org/10.1177/1754073908097175