

Computational Thematic Analysis of Academic Representations of the Gisaeng in the Korea Citation Index (2000–2024)*

Ribas Ramírez, César Augusto **

 0009-0007-8187-6280

Martín Casares, Aurelia ***

 0000-0002-3810-1961

Kang, Eun Kyung ****

 0000-0002-3440-4471

Table of Contents

1. Introduction
2. The Gisaeng in Scholarship and the Rise of Digital Humanities
3. Topic Modelling on the KCI Corpus
4. Analysis of the results
5. Conclusions

Abstract

This article undertakes a computational thematic analysis of academic discourse on *gisaeng* (기생, 妓生) in contemporary scientific production in the Republic of Korea. The study draws on a corpus of 613 abstracts published in Korean between 2000 and 2024 in academic journals indexed in the *Korea Citation Index* (KCI), systematically pre-processed and analysed through the Ko-SRoBERTa model and BERTopic. By combining Korean natural language processing, semantic embeddings, dimensionality reduction, and unsupervised clustering, the analysis identified ten distinct topics, subsequently organized into four macro-themes: 1) Literature, 2) Performing Arts and traditional music, 3) Colonialism, modernity and exoticization, and 4) Popular *gisaeng*. The results reveal the predominance of literary and performative representations of these women, with particular emphasis on popular figures such as *Hwang Jini* and *Chunhyang*. Simultaneously, the findings underscore lack of academic studies, particularly regarding everyday lives, maternity, and vulnerabilities of *gisaeng*, which remain largely excluded from academia. The result reveals the interpretive potential of deep learning models in historiographical, cultural and gender analysis, not as a substitute for close reading but as complementary tools capable of mapping discursive patterns, silences, and emerging lines of inquiry. By tracing a thematic cartography of recent Korean academic production, this article contributes both to digital humanities methodology and to a more nuanced understanding of how the *gisaeng* are constructed in academic discourse, contested, and reconfigured in contemporary scholarship.

Keywords: Gisaeng, Korean Studies, Digital Humanities, Topic Modelling, BERT

1. Introduction

Gisaeng women have been predominantly represented as both courtesans and professional artists in Korean cultural tradition. They have captured the attention of scholars across disciplines such as literature, history, gender studies, musicology, or performing arts and have been symbolized as active cultural agents during the Joseon dynasty, objects of sexual exploitation or as icons of resistance during the Japanese colonial period. This fragmentation within their representation in academic literature generates discourses that oscillate between

* Grant PID2023-150610NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”, by the “European Union” or by the “European Union NextGenerationEU/PRTR”.

**1st Author: Area of East Asian Studies, Department of Historical Sciences, University of Malaga, cribas@uma.es

***Coauthor: Area of East Asian Studies, Department of Historical Sciences, University of Malaga, aureliam@uma.es

****Coauthor: Area of East Asian Studies, Department of Historical Sciences, University of Malaga, eunkang@uma.es

aesthetic admiration, feminist critique and ethnographic analysis.

This study proposes a methodological approach centred on computational thematic analysis to explore how the *gisaeng* have been represented in contemporary Korean academic production. To this end, a corpus was built from academic articles indexed in the *Korea Citation Index* (KCI) between 2000 and 2024, written in Korean and including the term *gisaeng* (기생) in the title, keywords, or abstract. A deep learning model based on S-BERT was applied to this corpus, designed to generate contextualized semantic representations and enable thematic clustering of the texts through the BERTopic tool. This technique identified 13 emerging topics, which were subsequently grouped into 4 representative macro-themes through a process of qualitative interpretation: (1) literature and gender; (2) performing arts and traditional music; (3) colonialism, modernity and exotization; and (4) popular *gisaeng* figures. This article focuses on the critical analysis of these macro-themes, combining automatic semantic segmentation with close reading of the most representative texts in each group. Through this approach, the study examines the structural, conceptual, and rhetorical differences that shape the ways in which *gisaeng* are represented in current academic production in the Republic of Korea, attending both to dominant discourses and to marginal or emerging narratives. This methodological procedure makes it possible to trace a thematic cartography of recent academic imaginary on *gisaeng*, while enabling a critical reading of its continuities, ruptures, and silences. Ultimately, this study seeks to appreciate the interpretative potential of deep language models in cultural analysis and to offer a panoramic, nuanced, and relational view of how *gisaeng* are constructed and reconfigured in current Korean studies.

2. The Pertinence of Using Digital Humanities to Understand Discourses on *Gisaeng* in Recent Academic Production

Gisaeng occupy an ambiguous and highly symbolic place in Korea's cultural history and have been traditionally defined as women who performed functions of entertainment, with artistic skills, such as playing musical instruments, reciting poetry, and dancing, within the patriarchal context of Confucian Korea. These women operated according to official frameworks and there were several categories, such as the *gwangi* (官妓), attached to government offices, *gunggi* (宮妓), linked to palace contexts, or *mingi* (民妓), linked to private domains¹. Over time and across sources, *gisaeng* have been represented in dual and contradictory ways: on the one hand, idealized as aesthetic symbols of national culture and Korean femininity, and, on the other, objectified as sexual commodities and stigmatized due to their association with sexuality and prostitution. Their role during the Joseon dynasty and the Japanese colonial period has been undeniably studied and the core of these divergent interpretations, oscillating between their recognition as precursors of female emancipation and their association with sexual exploitation². From a historiographical perspective, several studies have highlighted the terminological richness surrounding the concept of *gisaengs*, showing how terms such as *ginyeo* (妓女), *changgi* (娼妓), or *sucheonggi* (守廳妓) were used interchangeably at different times to describe similar functions under shifting ideological frameworks, especially with the rise of Neo-Confucianism in mid-Joseon³. These categorizations not only reflect the changing status of these women, but also the social and political discourses that defined their visibility and legitimacy.

Despite growing scholarly interest, significant gaps persist in research on the *gisaeng*. As Jung Byungsul (2007) has noted, the everyday lives of these women have been little explored, with analyses often privileging exceptional figures such as *Hwang Jini* or *Maejang* over the study of their daily experiences. The author also points out the existence of profound childhood traumas among many *gisaeng*, with multiple references to systematic sexual abuse. Such experiences undoubtedly shaped their life trajectories in lasting ways, as well as their relationships with their environment, bodies, and art. Yet these aspects remain scarcely addressed in

academic publications, despite their relevance for a fuller understanding of the agency of these women. This omission is especially significant considering the persistence of the *gisaeng* as a cultural symbol in present-day Korea, where they continue to evoke tensions between memory, gender, and nation⁴. A more nuanced understanding of these women therefore requires a methodological approach that combines social history, discourse analysis, and cultural studies, attending both to the material conditions of their existence and to the ways in which they have been represented and reinscribed in the collective imagination. In this respect, the tools and approaches of the digital humanities open new possibilities for systematically tracing the representations, discourses, and silences that have shaped academic knowledge about the *gisaengs*. To date, no large-scale automated thematic analysis has been carried out that would allow a structured observation of the predominant and emerging discursive lines in the field. This study aims precisely to provide an initial cartography, generated through *deep language models*, as a starting point for future qualitative research to explore the identified thematic clusters in greater detail and thereby move toward a deeper, more critical, and contextualized understanding of the figure of the *gisaengs*. In this context, it becomes pertinent to ask how *gisaeng* have been treated in recent academic production. Beyond studies focused on emblematic cases or on literary and visual representations, there remains a need for a panoramic and systematic perspective that makes it possible to observe the discursive dynamics that have structured scholarly knowledge about these women. In this regard, the quantitative analysis of scientific production offers a suitable methodological avenue for detecting patterns, silences, and shifts in the academic treatment of the subject. In recent years, the quantitative analysis of scholarly output has gained increasing relevance in the Humanities, driven both by the development of specialized databases and by the incorporation of tools from computational linguistics and artificial intelligence. This methodological turn has not only enabled the systematization and visualization of large volumes of scientific literature but has also opened new questions about the ways in which knowledge is constructed, circulates, and is transformed within specific academic communities.

Over the past two decades, digital humanities have become a key interdisciplinary field that integrates traditional humanistic approaches with computational tools, visualization, and quantitative analysis, opening new ways of asking questions and representing knowledge⁵. In Korea, this expansion has been particularly significant thanks to the early digitization of archives, open-data policies, and institutional support. As Cha and Wall note⁶, Koreanists now have privileged access to large digital repositories, though Cha⁷⁸ warns against reducing this research to purely computational analyses. Scholars such as Kim Baro⁹ have also emphasized the importance of combining technical training with humanistic perspectives.

Within this context, computational approaches such as Latent Dirichlet Allocation¹⁰ and deep learning models like BERT¹¹ allow large-scale analyses of academic discourse. These techniques go beyond word counts, revealing thematic clusters, semantic shifts, and emerging lines of meaning, often supported by visualizations such as UMAP. For the case of this study, these tools enabled the mapping of discursive patterns, silences, and transformations about *gisaeng* in the Korean scholarly discourse between 2000 and 2024. However, it is important to underline that automated analysis cannot replace critical reading. As Moretti¹², Jockers¹³ and Underwood¹⁴ stress, distant reading must complement close reading. Similarly, Ricoeur¹⁵ and later Braun and Clarke¹⁶¹⁷ highlight the need to situate interpretation within its social and historical conditions. In this study, BERTopic serves not as a substitute for qualitative work but as a starting point to identify thematic clusters that future research can explore in greater depth.

3. Topic Modelling on the KCI Corpus

This study focuses on the application of BERT, specifically the *jhgan/ko-sroberta-multitask*¹⁸, to analyse South Korean academic discourse on the *gisaeng*, with the aim of identifying the main thematic frameworks

through which they were constructed and problematized. To this end, a methodological process was developed that combines Korean natural language processing with unsupervised machine learning algorithms, integrating tools for morphological tokenization, semantic vector representation, dimensionality reduction, clustering, and topic modelling.

The corpus analysed consists of 632 abstracts of academic articles published in Korean, extracted from the *Korea Citation Index* (KCI). To build the corpus, a search was conducted in March 2025 through the Journal Search interface (논문 검색) of the Korea Citation Index (KCI) portal¹⁹. The term “기생” (*gisaeng*) was entered in the general keyword field, with the search scope set to all available fields (title, abstract, and keywords). The search was restricted to publications classified under the Humanities (인문학) subject category. This initial search yielded a total of 969 articles. A subsequent manual screening was carried out to refine the results, since a significant number of texts used the term *gisaeng* not in reference to the historical women so designated, but as a synonym for the biological or metaphorical concept of “parasite” (a homograph in Korean language). This first conceptual filtering added to those whose metadata did not include an abstract, or whose abstract was in a language other than Korean, reduced the corpus to a total of 632 abstracts written in modern Korean.

To prepare the texts for computational analysis, morphological tokenization was applied using the Okt (Open Korean Text) analyser, focusing exclusively on nouns with a minimum length of two characters. This decision was based on the consideration that nouns constitute the lexical units with the highest semantic density in academic discourse and are therefore the most suitable candidates for topic modelling tasks. However, this grammatical filtering was complemented by a second level of refinement through an extensive stopword list specifically developed for this study. This list was not limited to standard dictionaries but was instead constructed and refined manually through an iterative inspection of the corpus, evaluating the frequency and informational value of each term. This dual filtering, by word class and by lexical content, is not redundant but complementary: while restricting to nouns excludes particles, generic adjectives, and grammatical verb forms, the stopword filtering removes highly frequent but thematically uninformative nouns in the corpus context. It should be noted that although textual preprocessing was based on grammatical filtering focused exclusively on nouns, the final stopword list also included some words that do not formally belong to this category. This is because certain terms, such as particles, adverbs, or complex grammatical expressions, are occasionally tagged as nouns by the Okt analyser due to contextual ambiguities or imperfect segmentation. The manual inclusion of these cases in the stopword list was intended to remove lexical units of low semantic density which, although not true nouns, recurrently appeared as such in the analyser’s output. This methodological decision optimized the quality of the final corpus without contradicting the principle of focusing on informative nouns, thereby enhancing the accuracy of topic modelling and the interpretability of the results, in line with previous studies on Korean lexical analysis for NLP tasks in the digital humanities²⁰. Once the texts were pre-processed and their embeddings generated using the Ko-SRoBERTa model, the next step was to determine the optimal parameter configuration for the BERTopic model by means of a systematic grid search across possible hyperparameter combinations for the UMAP (for dimensionality reduction) and HDBSCAN (for cluster detection) algorithms. A total of 5,460 different BERTopic configurations were evaluated, systematically combining values of the parameters `min_cluster_size`, `min_samples`, `n_neighbors` and `min_dist`. For each configuration, key metrics were calculated, including the total number of topics detected (`n_topics`), the number of documents unassigned to any topic (`n_unassigned`), the size of the largest topic (`max_topic_size`), and the standard deviation of topic sizes (`std_topic_size`). Based on these results, the following values were selected as optimal: `min_cluster_size`=13, `min_samples`=2, `n_neighbors`=10, and `min_dist`=0.0, which offered a robust balance between thematic granularity, semantic coherence, and a low rate of unassigned documents. Figure 1 visually compares this optimal configuration with two extreme reference cases: the one that produces the largest number of topics and

the one that achieves the smallest number of unassigned documents. The comparative analysis shows that the selected configuration not only maintains a controlled distribution of topic sizes but also maximizes effective document assignment, thereby ensuring meaningful thematic segmentation that can be utilized for subsequent qualitative analysis.

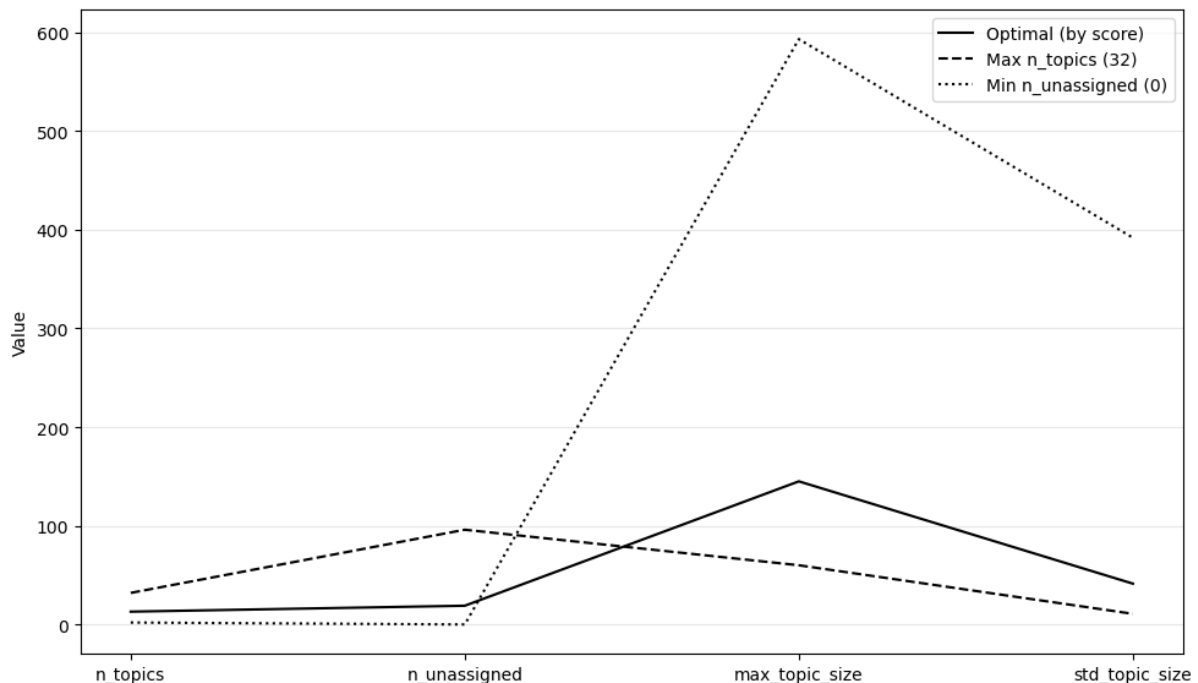


Figure 1. Parallel Coordinates of Key Metrics by S-BERT Configuration

Source: figure produced by the authors

To ensure the full reproducibility of the topic model, a comprehensive system was implemented for saving all intermediate and final products of the analysis. Instead of recalculating the model in each run, which could introduce slight variations due to the non-deterministic nature of the HDBSCAN algorithm and differences across hardware environments, the generated embedding vectors for each text, the fully trained BERTopic model (including UMAP, HDBSCAN, and the vectorizer), and the results with topic assignments for each document were persistently stored. This decision follows current methodological recommendations in reproducible data science, which are particularly relevant in contexts involving morphologically rich or non-segmented writing systems, such as Korean, where tokenization and preprocessing choices have a significant impact on topic modelling outcomes²¹. By preserving these objects in reusable formats (.npy for embeddings, .csv for results, and a structured folder for the model), the need for recalibration is avoided and external replication of the results by other researchers is facilitated, thereby ensuring methodological transparency and consistency.

Based on this analysis, a final configuration was selected that produced 13 well-differentiated topics, with 613 documents assigned and only 19 unassigned, representing a robust balance between analytical depth and thematic coverage. This strategy of exhaustive parameter evaluation ensured that the results of topic modelling were not only computationally efficient but also analytically interpretable and methodologically sound. In the same vein, special attention was paid to the semantic quality of the pre-processed corpus, particularly with respect to the definition of the stopwords list (Table 1), whose impact on topic segmentation is especially significant in morphologically rich corpora such as Korean. The criterion for inclusion in the stopwords list was based both on frequency of occurrence and on the low informational value of certain terms for the detection of thematic content. Grammatical particles and discourse connectors that, although occasionally tagged as nouns by the Okt analyser, lacked significant semantic density were removed. Common expressions in academic writing were also excluded,

as well as impersonal and collective pronouns or deictics whose presence hindered precise semantic segmentation.

In addition, vague determiners and quantifiers, temporal expressions, and contextual adverbs were removed due to their tendency to appear redundantly without directly contributing to topic detection. Furthermore, the term ‘gisaeng’, the central theme of the study, was deliberately excluded to prevent its high frequency from artificially dominating the topic modelling results. This exclusion allowed the topics to emerge around the contextual, functional, and symbolic aspects associated with the gisaeng, rather than from their literal mention. The stopword list was progressively refined throughout the modelling process, following iterative evaluations of the results generated through embeddings. This refinement resulted in a more coherent and analysable thematic representation.

Table 1. Table Title

Category	Korean stopwords
Grammatical and functional particles	우리, 자기, 자신, 여러분
Common expressions in research	문제, 내용, 자료, 의미, 과정, 기준으로, 관련, 기타, 생각한대로, 연구, 분석
Personal and demonstrative pronouns	우리, 자기, 자신, 여러분
Vague determiners and quantifiers	무엇, 무슨, 아무도, 약간, 정도
Temporal and contextual expressions	때문, 때문에, 다시, 시기, 언제, 여기, 이전, 이후, 지금, 우선, 최대한
Other adverbs and discursive expressions	가까스로, 오히려, 만약, 만일, 가령, 단지, 거나, 거의, 경우, 다른, 여덟
Term filtered by thematic content	기생

Subsequently, the texts were represented in a high-dimensional vector space using semantic embeddings generated by the jhgan/ko-sroberta-multitask model, a variant of SBERT trained for Korean through multiple semantic tasks such as textual inference, classification, and information retrieval. This model, available via the sentence-transformers library, produces dense representations that capture global semantics and enable the detection of latent similarities between documents. These embeddings, like other BERT-base models, have 768 dimensions, a standard size that ensures rich contextual information. To facilitate clustering and visualization, dimensionality reduction was performed with UMAP (Uniform Manifold Approximation and Projection), configured with ten neighbours, five components, and a minimum distance of zero, using cosine similarity. These parameters, selected through grid search optimization, preserved both local coherence and the global structure of the corpus in the reduced space. The resulting distribution is shown in Figure 2, where the documents can be observed in two dimensions prior to the qualitative colouring of clusters.



Figure 2. UMAP Distribution of the Abstract Corpus

Source: authors' visualization produced with BERTopic on SBERT embeddings trained on the Korean corpus.

On this reduced semantic space, clustering was performed with HDBSCAN, configured with a minimum cluster size of ten and two samples per cluster. This produced dense groups of semantically related abstracts, while documents without sufficient affinity were classified under the label -1 and erased from the visualization. From this segmentation, the BERTopic model generated thematic groups defined by their most distinctive lexical features. To refine topic assignments, a count-based vectorizer was used with the option `calculate_probabilities=True`, enabling not only the identification of each document's dominant topic but also its probabilistic distribution across others.

In the following section, these topics are qualitatively categorized, and the resulting thematic map is presented. Beyond the static visualization included in this article, an interactive version of the UMAP is accessible online through Hugging Face Spaces, allowing readers to explore the corpus in greater detail²². All datasets, preprocessing scripts, and the full codebase of the application have been deposited in Zenodo²³.

4. Analysis of the Results

Based on the topic model generated with BERTopic, it was possible to identify the keywords that define each of the thematic clusters (Table 2). The title assigned to each topic was the result of an interpretive analysis grounded in the systematic reading of the titles and abstracts of all the articles it comprised, complemented by the keywords generated through the c-TF-IDF (class-based Term Frequency–Inverse Document Frequency) model, which calculates the relative weight of each word within a topic compared to the rest of the corpus. Subsequently, the topics were manually organized into four analytical macro-themes, according to their semantic affinity, object of study, and methodological approach: (1) Literature and gender; (2) performing arts and traditional music; (3) colonialism, modernity and exoticization; and (4) popular gisaeng figures. The following table synthesizes this categorization, showing for each topic its identifier number, the interpretive title assigned, the corresponding macro-theme, the most relevant keywords in Korean and in Spanish, the total number of documents grouped in

each and the respective percentage. This classification constitutes the basis for the in-depth thematic analysis developed in the following sections.

Table 2. Thematic Breakdown of BERTopic
Source: Author's elaboration. Percentages calculated based on n=613 assigned documents.

Topic Number	Topic Name	Macro theme	Keywords in Korean	Keywords in English	N° of Doc.	%
Topic 1	Literary Constructions of <i>Gisaeng</i> and Female Subjectivity in Korea	Literature and gender	여성, 남성, 사랑, 사회, 욕망, 담론, 주체, 그녀, 근대, 소설	woman, man, love, society, desire, discourse, subject, she, modernity, novel	145	23,66%
Topic 2	Associations, Training Houses, and the Performance Practices of <i>Gisaeng</i>	Performing arts and traditional music	문화, 활동, 공연, 음악, 권번, 조합, 극장, 전통, 연행, 정재	culture, activity, performance, music, <i>gwonbeon</i> (gisaeng guild), association, theatre, tradition, stage performance, court dance (jeongjae)	104	16,97%
Topic 3	From Classical Fiction to Colonial Novels: Female Figures, Desire, Gender and Transgression Discourses	Literature and gender	소설, 작품, 이야기, 인물, 문학, 작가, 창작, 서사, 역사, 한국	novel, work, story, character, literature, author, creation, narrative, history, Korea	79	12,88%
Topic 4	Premodern Narratives, Historiography, and Self-Representations of <i>Gisaeng</i>	Literature and gender	기녀, 시인, 한시, 문학, 사대부, 퇴계, 작품, 여성, 남성, 문인	<i>gisaeng</i> , poet, classical literature, literature, sadaebu (Confucian literati), Toegye, work, femininity, masculinity, literary person	63	10,28%
Topic 5	Anthologies, Performance, and the Musical Legacy of <i>Gisaeng</i>	Performing arts and traditional music	시조, 가곡, 가집, 작품, 향유, 가사, 편찬, 세기, 노래, 여창	<i>sijo</i> , <i>gagok</i> (classical Korean song genre), song anthology, work, appreciation/enjoyment, <i>gasa</i> (narrative verse form), compilation/editing, century, song, female singer	48	7,83%
Topic 6	Visualizing and Controlling <i>Gisaeng</i> : Tourism, Print Media, and Colonial Discourse	Colonialism, modernity and exotization	조선, 관광, 평양, 일본, 경성, 잡지, 학교, 일본인, 발행, 사진	<i>Joseon</i> , tourism, Pyeongyang, Japan, Gyeongseong (Seoul under Japanese rule), magazine, school, Japanese person, publication, photograph	47	7,66%
Topic 7	Chunhyangjeon and Its Adaptations: From Classical Fiction to Modern Film	Popular gisaeng figures	춘향전, 도령, 영화, 인물, 사랑, 남원, 서사, 신분, 이몽룡, 옥중	Chunhyangjeon, Nobleman (<i>Doryeong</i>), film, character, love, Namwon, narrative, social status, Yi Mongryong, in prison	40	6,53%
Topic 8	Revisiting the Archival Files on <i>Gisaeng</i> and Prostitutes: Documentation, Errors, and New	Colonialism, modernity and exotization	창기, 조합, 서류철, 공창제, 통감부, 일제, 대한제국, 성매매, 속령, 오류	prostitute, association, file, system of public prostitution, Office of the Resident-General, Imperial Japan, Korean Empire, prostitution, ordinance, error	16	2,61%

	Perspectives					
Topic 9	Pansori, Female Singers, and the Cultural Transmission of Vocal Traditions	Performing arts and traditional music	판소리, 지역, 소리, 전승, 명창, 대목, 바디, 전주, 통속, 사설	<i>pansori</i> , region, sound/voice, transmission (heritage), master singer (<i>Myeongchang</i>), passage/section (<i>Daemok</i>), variant/version (<i>Badi</i>), <i>Jeonju</i> , popular/common (<i>Tongsok</i>), lyrics/text (<i>Saseol</i>)	16	2,61%
Topic 10	<i>Hwang Jini</i> and the Cultural Reimagining of Korea's Most Iconic <i>Gisaengs</i>	Popular gisaeng figures	황진이, 그녀, 인간, 황진, 중력, 자아, 송도, 사랑, 이미지, 작품	<i>Hwang Jini</i> , she, human being, <i>Hwangjin</i> , gravity, self, Songdo, love, image, work	15	2,45%
Topic 11	Gender, Class, and Narrative Conflicts in Korean Literature from Joseon to Modernity	Literature and gender	무정, 인물, 사건, 안평대군, 이데올로기, 이광수, 가부장제, 특옥, 진술, 계급	Mujeong (The Heartless, novel by <i>Yi Gwang-su</i>), character, event/incident, Prince Anpyeong, ideology, Yi Gwang-su, patriarchy, Deuk-ok (character name), statement/testimony, social class	14	2,28%
Topic 12	<i>Gisaeng</i> , Activism, and National Movements in Modern Korean History	Colonialism, modernity and exoticization	운동, 태극기, 활동, 구제, 참여, 고아원, 단체, 수재민, 자원, 김구	movement, South Korean flag, activity, humanitarian aid, participation, orphanage, organization, flood victims, volunteering, Kim Gu	13	2,12%
Topic 13	Regional Identity and Cultural Imagination in <i>Pyeongyang's</i> Literary and Artistic Representations	Performing arts and traditional music	평양, 논개, 관서, 지역, 주요한, 조선, 작품, 대시, 대동강, 이야기	<i>Pyeongyang</i> , <i>Nongae</i> , <i>Gwanseo</i> (northwestern region of Korea), region, important, <i>Joseon</i> , work (literary/artistic piece), courtship/declaration of love (<i>Daesi</i>), <i>Daedong</i> River, story/narrative	13	2,12%

Figure 3 presents the distribution of documents in the UMAP space, where each colour represents one of the themes identified by BERTopic. The shades within the red, yellow, blue or purple scales indicate the correspondence of each theme to a common macro theme. Colour intensity corresponds to the number of documents per theme, enabling visualization of their relative density. This visualization makes it possible to more clearly perceive the global organization of the corpus around the major discursive lines that structure academic production on the gisaeng.



Figure 3. UMAP Distribution of the Abstract Corpus

Source: authors' visualization produced with BERTopic on SBERT embeddings trained on the Korean corpus.

The thematic content analysis shows that, among the 613 documents assigned to a specific theme, the largest macro-theme is Literature and Gender with 301 documents (49.1%). It is followed by Performing arts and traditional music with 181 documents (29.5%), Colonialism, modernity and exotization with 76 (12.4%), and finally Popular *gisaeng* figures with 55 (9.0%). This distribution indicates that KCI scholarship on *gisaeng* is primarily anchored in literary and performative perspectives, while studies on colonial discourse, tourism, and visibility, as well as on emblematic individuals such as Hwang Jin-i and Chunhyang, appear as more specialized but nonetheless significant areas of inquiry. More than half of the corpus is concentrated in three themes: Topic 1 on literary constructions of *gisaeng* and female subjectivity (145 documents), Topic 2 on associations, training houses, and performance practices (104), and Topic 3 on female figures, desire, and gender in fiction (79). By contrast, the least represented are Topics 11, 12, and 13, with 14 and 13 documents each, focused respectively on the genderized role of *gisaeng*, class, and narrative conflicts in the context of Korean Literature, on modern activism and national movements, and on regional cultural representations in *Pyeongyang*. The UMAP visualization of the corpus (Figure 3) reflects this distribution: at the centre lies Literature and gender (red), the largest and most structuring nucleus, from which other clusters radiate; to its left appears Performing arts and traditional music (yellow), forming a relatively autonomous block; Colonialism, modernity and exotization (blue) occupies a transversal position in the lower-left, linked to visual and political discourses; while Popular *gisaeng* figures (purple) remains more peripheral on the upper and upper-right area, bridging literature and performance through emblematic figures. This cartographic overview provides the structural outline of *gisaeng* scholarship. To understand how these clusters are constituted, it is necessary to turn to the thematic cores themselves.

The corpus on Literary Constructions of *Gisaeng* and Female Subjectivity in Korea (Topic 1) demonstrates how these women have been mobilized as a narrative device to interrogate chastity, virtue, desire, and modernity. Premodern studies revisit the “virtuous *gisaeng*” and the rhetoric of chastity not as fixed categories but as contested fields of symbolic negotiation. For example, Kim Boram’s work on the ideology and practice of chastity among *Joseon gisaeng* reveals virtue as a performative construct rather than a stable moral code, while Cho Hyun Woo reframes the “righteous *gisaeng*” as a narrative effect oscillating between moral celebration and didactic exploitation. Comparative studies linking Korean legends with sinicized sources, such as Du Shiniang, situate the *gisaeng* within a broader trans-Asian constellation of loyalty and tragic desire. Modern and contemporary analyses shift the focus toward celebrity, media, and gender politics. Jiyoung Suh’s studies on “public intimacy” explain how *gisaeng* became media celebrities whose erotic and artistic capital was commodified through print and visual culture. Similarly, scholarship on Kim Mal-bong, highlights how female characters were reconfigured as sex

workers or precarious youth in postwar fiction, intertwining romantic redemption, stigma, and gendered economies. Even in popular Cold War genres such as thrillers, works like *gisaeng* spy Kim Sosan reveal the *gisaeng* figure as a threshold between desire, surveillance, and nationalism. Methodologically, these studies combine discourse analysis, feminist hermeneutics, and media sociology, emphasizing the *gisaeng* not as a fixed occupation but as a semantic node for social imaginaries. From Classical Fiction to Colonial Novels (Topic 3) traces the transition from *Joseon* narratives to colonial-era novels, focusing on how female desire and transgression became metaphors for modern subjectivity. Studies of Kim Dong-in's "When the Eyes Barely Open" and works such as *The Song of Bu-yong's Longing and Danbalryeong* illustrate how inherited conventions were reshaped under colonial conditions. *Yi Gwang-su's Mujeong* (The Heartless) emerges as a key paradigm, embodying tensions between patriarchy, nationalism, and female subjectivity. Analyses highlight the ambivalence of women as desiring subjects and disciplined bodies, making literature a space of negotiation between tradition and modernity. Keywords such as "novel," "narrative," "woman," "desire," and "modernity" underscore the interdisciplinary orientation of this topic, where female figures serve as metaphors for the dilemmas of modernization.

The theme Premodern Narratives, Historiography, and Self-Representations of *Gisaeng* (Topic 4) turns to *sijo*, *gasa*, and *hanmun* literary records to show how *gisaeng* were both objects of literati projection and active poetic voices. Studies of poem exchanges (*sujak sijo*) and circles like the *Samhojeong* poetry society reveal *gisaeng* not only as muses but also as creators who deployed irony and critique within Confucian constraints. Figures such as *Hwang Jini* and *Non Gae* appear as "righteous *gisaeng*," invested with patriotic virtue in contexts of war and crisis. While most sources are male-mediated, attributed *sijo* and *gasa* demonstrate moments of female self-representation that articulate desire, loneliness, and aspirations for recognition. These contributions highlight *gisaeng* as cultural agents whose voices shaped literary memory and challenge a purely masculinized historiography. Finally, Gender, Class, and Narrative Conflicts in Korean Literature from Joseon to Modernity (Topic 11) foregrounds *Yi Gwang-su's Mujeong* as a crucible for debates on patriarchy, ideology, and class conflict. Feminist readings interpret the female characters as embodying both promises of emancipation and enduring patriarchal constraints, while Marxist critiques underscore class hierarchies and nationalist aspirations embedded in the novel's love plots. References to Prince Anpyeong connect premodern concerns with colonial literature, showing continuity in ideological preoccupations. These studies consistently emphasize that gender and class were constitutive of literary modernity, with female figures functioning as symbolic sites where collective anxieties about nation and modernization were inscribed.

Taken together, the Literature and Gender macro-theme illustrates that *gisaeng* and other female figures are not static literary types but dynamic discursive constructs. Across premodern, colonial, and modern contexts, they operate as prisms for interrogating morality, desire, class, and national identity. The scholarship demonstrates that female subjectivity emerges less as a historical given than as a contested site shaped by shifting regimes of discourse and cultural production.

The research grouped under the following macro theme on Performing arts and traditional music examines the institutional, repertorial, and performative dimensions of *gisaeng* as professional artists. Far from treating their work as "remnants of tradition," these studies emphasize the processes of guild organization, professionalization, and repertoire standardization that shaped their role as cultural mediators. One strand highlights the long historical roots of *gisaeng* performance. Archival sources such as the Mukjae ilgi document the cooperative configurations between *gisaeng* and musicians as early as the sixteenth century, revealing defined obligations, wages, and hierarchies that positioned performance as regulated labour rather than occasional entertainment. A second axis focuses on the urban performance culture of the early twentieth century, particularly the role of *gwŏnbeon* guilds as institutions that centralized training in singing, dance, and etiquette. Analyses of newspapers, hand programs, and memoirs illustrate how *gisaeng* ensembles acquired public visibility while becoming dependent on intermediaries and theatre infrastructures. Studies of works like *Yet'an ilbaegin* portray *gisaeng* groups of the

1910s as professional collectives embedded in circuits of contracts, auditions, and touring. A further set of contributions addresses the musical repertoire itself, with particular attention to *sijo*, *gagok*, and *gasa*. Anthologies such as *Gyobang gayo* and early twentieth-century compilations are examined as vehicles for canon formation and cultural preservation. At the same time, research on female vocalists (여창) demonstrates how *gagok* and *sijo* were adapted to concert settings, with transformations in vocal projection, staging, and audience reception. These works underscore the reflexive reinvention of “tradition” as part of modern performance culture. Another focus is the role of *gisaeng* in pansori transmission. Studies highlight their contribution as performers and teachers in regional variants, such as those in Chungcheong, as well as their technical innovations in voice, ornamentation, and declamation. This revaluation challenges male-centred historiography by showing how *gisaeng* expanded the expressive range of pansori and sustained its pedagogical transmission. Finally, Topic 13 situates *gisaeng* practices in the cultural geography of Pyeongyang. Here, literature, lyric traditions, and performance converge in articulating regional and national identities. Figures such as Nongae and motifs tied to the Daedong River underscore the symbolic dimension of *gisaeng* artistry as markers of resistance, loyalty, and local belonging.

Overall, this macro theme demonstrates that *gisaeng* were not incidental entertainers but key agents in the institutionalization, preservation, and transformation of Korean vocal and performative traditions. By bridging classical repertoires and modern infrastructures, they ensured the continuity of *sijo*, *gagok*, *gasa*, and pansori while situating their work within broader circuits of identity, memory, and cultural modernity.

The third macro theme focusses on colonialism, modernity and exotization of the *gisaeng*. It examines how these women were redefined under colonial modernity as both cultural emblems and objects of state regulation. The 47 studies grouped in Topic 6 highlight the role of photography, tourism, and print culture in transforming *gisaeng* into icons of a picturesque “Korean tradition.” Postcards and visual media circulated in Gyeongseong and abroad displaced performers from artistic contexts to a colonial showcase, while magazines such as *Tourist Korea* instrumentalized *gisaeng* as tourist attractions and symbols of national otherness. At the same time, colonial authorities deployed legal and bureaucratic categories to discipline their bodies, conflating performers with prostitutes within a system of surveillance and economic control. All these processes emphasized Korea as a destination for sexual tourism directed to the Japanese traveller. Topic 8 focuses on the colonial administrative archives that documented *gisaeng* and registered prostitutes (창기) during the transition from the Korean Empire to the Japanese Residency-General. Studies reveal the artificiality of these records: duplications, misclassifications, and inconsistencies expose the bureaucratic production of categories rather than a reflection of social realities. The research underscores that these files must be read critically, not as neutral sources but as instruments of colonial governance aimed at redefining entertainment and sexuality under Japanese rule. Finally, Topic 12 highlights *gisaeng* as participants in modern activism and humanitarian work. Articles document their involvement in independence movements, their symbolic appropriation of the Taegeukgi flag, and their participation in relief campaigns such as the 1925 flood response. These accounts challenge reductive views of *gisaeng* as purely artistic figures or objects of sexual control, instead presenting them as civic agents engaged in community and national projects.

Taken together, this macro theme shows that colonial modernity generated a dual image of the *gisaeng*: as an exoticized icon marketed through tourism and media, and as a subject of state discipline through regulation and bureaucratic categorization. Yet, simultaneously, *gisaeng* also emerged as active participants in nationalist and humanitarian movements. This tension between colonial exotization, institutional control, and civic agency illustrates the complexity of their position within Korea’s colonial modernity.

Lastly, the macro theme Popular *Gisaeng* Figures explores how individual *gisaeng* and related female figures became crystallized in cultural memory as exemplary icons, while also situating them within broader repertoires of narrative, performance, and collective imagination. Although *Chunhyang* and *Hwang Jini* remain the most recurrent figures, the corpus reveals that their prominence is neither exclusive nor exhaustive: other characters,

voices, and historical personalities emerge as complementary embodiments of female agency, morality, and cultural symbolism.

The 40 studies in Topic 7 focus primarily on *Chunhyangjeon*, yet their scope extends beyond the heroine herself. Analyses of pansori versions and manuscript traditions emphasize not only the fluidity of *Chunhyang*'s image but also the wider narrative system that sustained the story. Scholars highlight the presence of subsidiary characters, corrupt magistrates, loyal servants, maternal figures, whose roles are crucial to understanding how morality, hierarchy, and social critique are staged in these works. Moreover, comparative studies draw connections between *Chunhyangjeon* and other premodern narratives, showing that the cultural resonance of *gisaeng* figures cannot be reduced to a single heroine but is embedded in a larger storytelling tradition where questions of virtue, loyalty, and desire were continually reconfigured. The filmic and theatrical adaptations of the twentieth century reinforce this dual dynamic: while *Chunhyang* herself is nationalized as an allegory of fidelity and resistance, the broader cast of characters anchors the narrative in social commentary and dramatizes collective tensions around status, law, and justice.

The final 15 studies that are gathered in Topic 10 are largely devoted to *Hwang Jini*, whose poetry, biography, and modern reinterpretations cement her as a paradigmatic figure of beauty, intellect, and desire. Yet the corpus also includes a significant exception: the figure of *Kim Man-deok* (김만덕), the philanthropist and entrepreneur from Jeju who, although not canonized as a *gisaeng* in the same sense as Hwang Jini, has been integrated into the historiography of exemplary women for her economic acumen and charitable acts. Her inclusion signals that the field of "popular *gisaeng* figures" is porous, accommodating women whose fame derived from distinct forms of public engagement and who came to embody alternative models of female virtue and social recognition. Hwang Jini herself, moreover, is never represented monolithically: studies range from premodern *sijo* and literati encounters to modern cinematic portrayals, where she appears alternately as tragic lover, intellectual partner, nationalist heroine, or emblem of trauma and desire.

Together, these two topics show that emblematic female figures associated with the *gisaeng* tradition do not operate solely as isolated icons but as nodes in a broader cultural repertoire. *Chunhyang* and *Hwang Jini*, while central, coexist with other figures such as Kim Man-deok and with the numerous secondary characters that populate narrative and performative traditions. Their significance lies precisely in this plasticity: they are mutable cultural symbols through which Korean society has articulated ideals of loyalty, virtue, resistance, and creativity, while negotiating the contradictions of gender, class, and morality. In this way, *Popular Gisaeng Figures* captures both the enduring power of individual heroines and the wider imaginative field that elevated *gisaeng* and related women into cultural infrastructures of Korean modernity.

5. Conclusions

The present thematic analysis of 613 articles published between 2000 and 2024 in Korean academic journals indexed in the Korea Citation Index shows both the achievements and the limitations on the field of scientific publications on *gisaeng* women. Over the past two decades, scholarship has significantly advanced our understanding by recovering emblematic figures such as *Hwang Jini* and *Kim Man-deok*, analysing classical narratives like *Chunhyangjeon*, and documenting *gisaeng* performance practices in *pansori*, *sijo*, and *gagok*. These contributions have established *gisaeng* studies as a vibrant area of inquiry at the intersection of literary criticism, performance studies, and gender history.

Yet this breadth contrasts with persistent silences that continue to shape our understanding of the phenomenon. Among the most notable gaps is the scarce attention paid to the everyday lives of *gisaeng*, maternity and vulnerability. Despite the abundance of studies on their representations, artistic trajectories, and ideological projections, very little has been investigated about their concrete life experiences: the conditions of their training,

the social ties they forged, their spaces of intimacy, health, aging, or affectivity. Their maternity, for instance, remains almost completely absent from the discussion, even though this is a crucial issue given the matrilineal inheritance system that defined them. The lack of attention to this dimension prevents a comprehensive understanding of *gisaeng* as historical subjects with their own agency. Equally absent is any systematic effort to recover the names and surnames of *gisaeng* to bring them out of oblivion and anonymity to provide then an identity as the existing individuals they were. The fact that a handful of emblematic *gisaeng* enjoy such centrality also means that many others remain in anonymity, erased from cultural memory. Moreover, the trauma of child abuse and the pervasive sexual exploitation to which many *gisaeng* were subjected from very early ages are scarcely addressed. This omission is especially troubling, as these were often institutionalized systems of bodily control and artistic exploitation. To neglect these aspects is to obscure one of the most fundamental realities of their lives.

Another constant in academic production is the reproduction, whether implicit or explicit, of the dichotomy between artist and prostitute, a structuring tension in the representation of *gisaengs*. This duality, which on the one hand legitimizes their cultural value and on the other obscures their exploitation, was also observed in our earlier work analysing the character *gi* (妓) in the *Annals of the Joseon Dynasty*²⁴. This study on the character *gi* demonstrated how the associated female figure embodies both institutional recognition and moral subordination, revealing a semantic ambivalence that explicitly exists at least from the beginnings of the *Joseon* dynasty and remains active in contemporary narratives.

Considering this dispersion of approaches and the persistence of dichotomous frameworks, there is a need to develop holistic analyses that integrate the multiple dimensions already studied while also foregrounding those that remain silenced. Such an approach would generate transversal readings that recognize the complexity and historicity of the *gisaeng* women, enabling us to move beyond disciplinary compartments toward more nuanced, empathetic, and critical representations. In this context, the computational methodologies employed in this study prove valuable tools for mapping academic knowledge, identifying blind spots, patterns of concentration, recurring thematic threads, and analytical gaps. Far from replacing qualitative reading, this approach complements textual criticism by providing a panoramic view of discourses, allowing us to see the forest without losing sight of the trees.

At the same time, it is worth considering that the thematic distribution identified in this study may also be influenced by patterns of access to different types of primary sources. While the prominence of literary analyses reflects both the richness of textual materials and their centrality in the Korean Humanities, it may also indicate a tendency within scholarship to privilege certain kinds of sources over others that remain less frequently mobilized in the study of *gisaeng*. These tendencies do not necessarily arise from limitations in data infrastructure, but rather from long-standing academic habits and interpretive traditions that shape how researchers choose and use their sources.

Looking ahead, it would be desirable to expand this exercise by incorporating academic literature in other languages, monographs, dissertations, and non-indexed materials, as well as by periodically updating the analysis as new articles are published. Advances in natural language processing techniques will also enable more fine-grained topic models, with greater capacity to detect subjectivities, silenced voices, and ideological shifts, thus opening new possibilities for digital humanities applied to Korean and gender studies. We also hope that this study will encourage further research using large-scale databases such as the Korea Citation Index, demonstrating the potential of computational methods to build systematic, reproducible, and data-driven state-of-the-art reviews.

This work therefore constitutes a point of departure for reconsidering what has been said, and what remains unsaid, about the *gisaeng*, as well as an invitation to read them not only as cultural icons but as historical figures shaped by patriarchal ideals of beauty, as well as by coercion, exploitation, resistance, and memory.

- ¹ Jung, B. (n.d.). *나는 기생이다*. Retrieved 4 September 2025, from <http://library.litkorea.or.kr/originalworks/410302>
- ² Pilzer, J. D. (2006). The Twentieth-Century “Disappearance” of the Gisaeng during the Rise of Korea’s Modern Sex-and-Entertainment Industry. In M. Feldman & B. Gordon (Eds), *The Courtesan’s Arts* (pp. 295–311). Oxford University Press New York, NY. <https://doi.org/10.1093/oso/9780195170283.003.0018>
- ³ Shin, H. (2022). Nason Kim Dong-wook’s History of Gisaeng: Definition of ‘Giyeou/Gisaeng/changgi’. *The Modern Bibliography Review Society*, 26, 617–650. <https://doi.org/10.56640/mbr.2022.26.617>
- ⁴ Wang, Y. (2023). When Camera Encountered ‘Chosŏn Beauties’ Kisaeng Photographs, Tourism, and Postcards from the 1880s to Colonial-Period Korea: Kisaeng Photographs, Tourism, and Postcards from the 1880s to Colonial-Period Korea. In J.-C. An & A. Perrin, *Cultural Exchanges Between Korea and the West Artifacts and Intangible Heritage* (p. Chapter_15801). Fondazione Università Ca’ Foscari. <https://doi.org/10.30687/978-88-6969-717-3/008>
- ⁵ Burdick, A. (2012). *Digital humanities*. MIT Press.
- ⁶ Cha, J., & Wall, B. (2023). Introduction to Special Section Digital Korean Studies. *Korean Studies*, 47(1), 1–7. <https://doi.org/10.1353/ks.2023.a908615>
- ⁷ Cha, J. (2018). Digital Korean studies: Recent advances and new frontiers. *Digital Library Perspectives*, 34(3), 227–244. <https://doi.org/10.1108/DLP-04-2018-0013>
- ⁸ Cha, J. (2023). Big Data Studies: The Humanities in Uncharted Waters. *Korean Studies*, 47(1), 274–299. <https://doi.org/10.1353/ks.2023.a908625>
- ⁹ 金把路. (2016). 韓國數位人文教育現狀與課題. *漢學研究通訊*, 35(1), 27–35.
- ¹⁰ Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null), 993–1022.
- ¹¹ Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1810.04805>
- ¹² Moretti, F. (2007). *Graphs, maps, trees: Abstract models for literary history* (Paperback edition). Verso.
- ¹³ Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press. <https://www.jstor.org/stable/10.5406/j.ctt2jcc3m>
- ¹⁴ Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. the University of Chicago press.
- ¹⁵ Ricœur, P. (1976). *Interpretation theory: Discourse and the surplus of meaning*. Fort Worth : Texas Christian University Press. <http://archive.org/details/interpretationth0000ricu>
- ¹⁶ Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- ¹⁷ Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.
- ¹⁸ Park, Y., & Shin, Y. (2023). Using Multiple Monolingual Models for Efficiently Embedding Korean and English Conversational Sentences. *Applied Sciences*, 13(9), 5771. <https://doi.org/10.3390/app13095771>
- ¹⁹ <https://www.kci.go.kr/kciportal/po/search/poArtiSear.kci>
- ²⁰ Li, X. (2021). Analysis of Topics of Interdisciplinary Research in the Field of Library Science and Information Science from the Perspective of Altmetrics. *OALib*, 08(11), 1–9. <https://doi.org/10.4236/oalib.1108136>
- ²¹ Bostrom, K., & Durrett, G. (2020). Byte Pair Encoding is Suboptimal for Language Model Pretraining. Findings of the Association for Computational Linguistics: EMNLP 2020, 4617–4624. <https://doi.org/10.18653/v1/2020.findings-emnlp.414>
- ²² Access the interactive UMAP through the following URL: <https://huggingface.co/spaces/Cesarhanmun/UMAPGisaengKCI>
- ²³ Datasets, preprocessing scripts and codebase of the app available through the DOI: <https://doi.org/10.5281/zenodo.17237749>
- ²⁴ Ribas Ramírez, C. A., & Martín Casares, A. (2025). Las mujeres de Corea representadas por el carácter Gi en los Anales de la Dinastía Joseon (1392—1897). In E. K. Kang (Ed.), *Estudios sobre mujeres de Corea y Asia Oriental* (1ª ed.). Dykinson. <https://doi.org/10.14679/4082>

References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null), 993–1022.
- Bostrom, K., & Durrett, G. (2020). Byte Pair Encoding is Suboptimal for Language Model Pretraining. Findings of the Association for Computational Linguistics: EMNLP 2020, 4617–4624. <https://doi.org/10.18653/v1/2020.findings-emnlp.414>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

-
- Burdick, A. (2012). *Digital humanities*. MIT Press.
- Cha, J. (2018). Digital Korean studies: Recent advances and new frontiers. *Digital Library Perspectives*, 34(3), 227–244. <https://doi.org/10.1108/DLP-04-2018-0013>
- Cha, J. (2023). Big Data Studies: The Humanities in Uncharted Waters. *Korean Studies*, 47(1), 274–299. <https://doi.org/10.1353/ks.2023.a908625>
- Cha, J., & Wall, B. (2023). Introduction to Special Section Digital Korean Studies. *Korean Studies*, 47(1), 1–7. <https://doi.org/10.1353/ks.2023.a908615>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1810.04805>
- Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press. <https://www.jstor.org/stable/10.5406/j.ctt2jcc3m>
- Jung, B. (n.d.). 나는 기생이다. Retrieved 4 September 2025, from <http://library.ltikorea.or.kr/originalworks/410302>
- Li, X. (2021). Analysis of Topics of Interdisciplinary Research in the Field of Library Science and Information Science from the Perspective of Altmetrics. *OALib*, 08(11), 1–9. <https://doi.org/10.4236/oalib.1108136>
- Moretti, F. (2007). *Graphs, maps, trees: Abstract models for literary history* (Paperback edition). Verso.
- Park, Y., & Shin, Y. (2023). Using Multiple Monolingual Models for Efficiently Embedding Korean and English Conversational Sentences. *Applied Sciences*, 13(9), 5771. <https://doi.org/10.3390/app13095771>
- Pilzer, J. D. (2006). The Twentieth-Century “Disappearance” of the Gisaeng during the Rise of Korea’s Modern Sex-and-Entertainment Industry. In M. Feldman & B. Gordon (Eds), *The Courtesan’s Arts* (pp. 295–311). Oxford University Press New York, NY. <https://doi.org/10.1093/oso/9780195170283.003.0018>
- Ribas Ramírez, C. A., & Martín Casares, A. (2025). Las mujeres de Corea representadas por el carácter Gi en los Anales de la Dinastía Joseon (1392—1897). In E. K. Kang (Ed.), *Estudios sobre mujeres de Corea y Asia Oriental* (1^a ed.). Dykinson. <https://doi.org/10.14679/4082>
- Ricœur, P. (1976). *Interpretation theory: Discourse and the surplus of meaning*. Fort Worth : Texas Christian University Press. <http://archive.org/details/interpretationth0000ricu>
- Shin, H. (2022). Nason Kim Dong-wook’s History of Gisaeng: Definition of ‘Giyeou/Gisaeng/changgi’. *The Modern Bibliography Review Society*, 26, 617–650. <https://doi.org/10.56640/mbr.2022.26.617>
- Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. the University of Chicago press.
- Wang, Y. (2023). When Camera Encountered ‘Chosŏn Beauties’ Kisaeng Photographs, Tourism, and Postcards from the 1880s to Colonial-Period Korea: Kisaeng Photographs, Tourism, and Postcards from the 1880s to Colonial-Period Korea. In J.-C. An & A. Perrin, *Cultural Exchanges Between Korea and the West Artifacts and Intangible Heritage* (p. Chapter_15801). Fondazione Università Ca’ Foscari. <https://doi.org/10.30687/978-88-6969-717-3/008>
- 金把路. (2016). 韓國數位人文教育現狀與課題. *漢學研究通訊*, 35(1), 27–35.