Supporting Information for

Feminism without Morality, Neoliberalism as Feminist Praxis?:

A Computational Textual Analysis of Womad, a South Korean Online “Feminist” Community

Table of Contents

[Online Appendix A: Text-Preprocessing 1](#_Toc50416115)

[Online Appendix B: Model Selection 2](#_Toc50416116)

[Online Appendix C: Glossary 4](#_Toc50416117)

# Online Appendix A: Text-Preprocessing

The current study only used nouns to analyze the topic structure not only because, for Korean, nouns are considered most relevant in capturing the thematic structure of a certain document (Kang, Song, & Jho, 2013), but also because Korean part-of-speech tagging software in general--as of yet--are often unable to accurately lemmatize verbs and adjectives (Nam, 2016).

The following text preprocessing procedures were implemented to extract nouns from each post. First, a set of non-Korean characters was removed from the corpus, including special characters, emojis, and numbers. Second, we removed any Womad vernacular such as adding “-노(-noh)” or “-이기야 (-igiya)” at the end of sentences because such suffixes lack substantive meaning. Third, any variations and obvious typos were corrected in order to ensure consistency within the corpus in terms of word use. For example, the 20th Korean president 문재인 (Moon Jae-in) is often dubbed as 문죄인 (Moon, the Sinner), 문재앙 (Moon, the Catastrophe), and 문제인 (Moon, the Problem): we unified this as 문재인 within the corpus.

Next, we used a Korean part-of-speech tagging software (KoNLP; Jeon 2016) to extract nouns from the corpus. A set of terms including slang, names, and technical terms were added to the software dictionary in order to correctly identify nouns. KoSpacing package (Jeon 2018) was also used for any words longer than 7 Korean characters in order to correct grammatical errors due to the incorrect use of spacing. Finally, we removed less frequent terms (appearing in less than .1% of documents), too-frequent terms such as 한남 (Korean male), and non-noun terms that KoNLP program incorrectly identified as nouns. This procedure left us with a total of 20,419 documents with 4,239 unique nouns.

# Online Appendix B: Model Selection

# 

A procedure proposed by Maier et al. was implemented to determine the number of topics (Maier et al., 2018). This procedure combines a systematic approach and a qualitative investigation to construct candidate models and further narrow down topic solutions for a textual dataset.

First, we estimated a series of models by systematically varying different combinations for topic numbers (K; 20, 40, 60) and alpha parameters (α, .01, .05, .1, .2, .5, 1). The beta parameter was fixed at 1/K. Then, one candidate model from the model with the same K was selected based on the likelihood and topic coherence measures (See Table A1). Third, two researchers examined each of these candidate models to select an appropriate topic solution for the purpose of the current research. The candidate model with K = 40 was selected because automatically generated topics tend to be too specific in the model with K = 60 and were too broad and thus contained multiple issues in one topic with the model in K = 20.

Second, following the procedure outlined in Maier et al. (2018), the two researchers independently investigated each of automatically generated topics to decide whether a certain topic should be included in the analysis. Top frequent words, topic proportion, and the topic coherence score of each topic were considered. Specifically, those topics whose top frequent word list were hard to interpret while exhibiting low proportion and coherence scores were excluded. We further excluded any topics that both researchers independently examined and marked off as incoherent. For any topics that the two researchers were not able to reach a consensus, a close examination of documents (posts) randomly selected from the topic was conducted to decide whether to keep or discard a topic. A total of 9 topics were excluded after the procedure.

Third, the two researchers also collaborated to label each of the 31 automatically generated topics. Specifically, the researchers independently read a random subset of documents from each topic and independently proposed a label for each topic. The researchers then worked together to create a label that summarizes the overarching theme of each topic. This procedure resulted in a total of 31 topics for the analysis. Table 1 in the manuscript presents a total of 31 topics, along with the five most frequently appearing nouns for each topic.

***Table A1. Choice of Candidate Models from Topic Models with Varying Parameter Sets***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | K | α | β (1/*K*) | Likelihood | Mean Coherence |
| 1 | 20 | 0.01 | 0.050 | -4065204.72 | -845.11 |
| 2 | 20 | 0.05 | 0.050 | -3892700.22 | -835.33 |
| 3 | 20 | 0.10 | 0.050 | -3801797.23 | -827.51 |
| 4 | 20 | 0.20 | 0.050 | -3702486.57 | -831.42 |
| 5\* | 20 | 0.50 | 0.050 | -3556314.74 | -809.01 |
| 6 | 20 | 1.00 | 0.050 | -3448740.75 | -830.39 |
| 7 | 40 | 0.01 | 0.025 | -3798993.04 | -853.87 |
| 8\* | 40 | 0.05 | 0.025 | -3571812.40 | -847.07 |
| 9 | 40 | 0.10 | 0.025 | -3453200.14 | -852.16 |
| 10 | 40 | 0.20 | 0.025 | -3321760.83 | -860.95 |
| 11 | 40 | 0.50 | 0.025 | -3155590.43 | -849.10 |
| 12 | 40 | 1.00 | 0.025 | -3009114.62 | -869.00 |
| 13 | 60 | 0.01 | 0.017 | -3635043.48 | -859.00 |
| 14\* | 60 | 0.05 | 0.017 | -3357452.47 | -855.85 |
| 15 | 60 | 0.10 | 0.017 | -3230337.35 | -860.80 |
| 16 | 60 | 0.20 | 0.017 | -3074969.26 | -865.76 |
| 17 | 60 | 0.50 | 0.017 | -2899860.08 | -878.62 |
| 18 | 60 | 1.00 | 0.017 | -2753845.68 | -901.84 |

Note: *K* = number of topics. \* indicates the candidate model for the given K.

# Online Appendix C: Glossary

**Bonjwa:** In Womad, Bonjwa often refers to an exceptional woman excelling in her area of expertise. In the time frame of our study, Bonjwa almost exclusively referred to the woman who took a picture of a male nude model leading to the Hongik University hidden camera incident.

**Café:** Café refers to online communities serviced through two popular portal sites in South Korea: Naver.com and Daum.net

**Corset**: Corset symbolizes any social oppression and expectations towards women that reinforces notions of femininity.

**Dentured-senior**: A derogative term referring to seniors.

**Hosik-(Han):** The name of a male firefighter who made misogynic comments on social media.

**It-Man:** A pronoun (third person singular) referring to a man. It supposedly mirrors “it-woman (i.e., she),” as no pronoun referring to a man in the third person singular exists in Korean.

**Jae-gi:** A slang term for suicide, referring to the accidental suicide of a male rights activist Jae-gi Sung, who jumped off from Seoul’s Mapo bridge during a publicity stunt.

**Mutt-and-swine**: Term referring to a conformist majority, i.e., those who submit thoughtlessly to the patriarchy.

**Mother-Father**: This word reverses the Korean term for parent (i.e., “bu-mo;” ) which literally translates into “father-mother”

**Peeping-penis**: Refers to male lurkers in Womad.

**S-Man**: Our translation of “Sho-nam,” which literally means “a small man,” i.e., a male child.

**Sho-kids:** Our translation of “Shyo-rin-ee.” It is a compound noun made of Japanese slang term “shyo-ta,” which mean a young boy, and a Korean word for kid (i.e., “eo-rin-ee;”). The term is used to sexualize young boys and legitimize pedophilia.

**The sun**: Our translation of “Hat-nim (i.e., the sun),” which refers to former President Geun-Hye Park.