



The Representation of Islam in European Parliament Sessions: A Corpus Study of 'Europarl 3: German'

Tri Edliani Lestari¹, Miftahulkhairah Anwar¹, Prihantoro²

¹ Applied Linguistics, Universitas Negeri Jakarta, Jakarta, Indonesia

² Department of English, Universitas Diponegoro, Semarang, Indonesia

*Corresponding author email: tri.edliani@mhs.unj.ac.id

Article Info

Article history:

Received September 15, 2024
Approved December 12, 2024

Keywords:

Collocation Analysis, Corpus Linguistics, Semantic Categorization

ABSTRACT

This study investigates the representation of Islam in European Parliament sessions using the Europarl 3: German corpus, focusing on collocations related to Islam to understand their associated sentiments and semantic distributions. Political discourse plays a critical role in shaping public perceptions, and Islam is often framed within the context of security and conflict in contemporary European narratives. This study aims to explore two key aspects: (1) the sentiments associated with collocations related to Islam in European Parliament sessions, and (2) the semantic categorization of these collocations within the corpus. Using collocation and concordance analysis via Corpus Query Processing (CQP), the study finds that themes with absolute negative sentiments are predominantly associated with the term Terrorismus 'terrorism,' which is frequently linked to Islam. However, some other collocations reveal variations in sentiment, including the presence of limited but notable positive sentiments. The study also categorizes these collocations into semantic groups based on established corpus linguistics frameworks, identifying patterns of representation that indicate bias in political discourse toward Islam. These findings highlight the securitization of Islam in political narratives, which may reinforce negative stereotypes and contribute to social polarization. Nevertheless, the existence of references with positive sentiments suggests opportunities to promote more balanced and constructive discourse. This study underscores the importance of inclusive political language to foster social cohesion. Future research is recommended to expand the corpus to other languages, conduct longitudinal analyses, and further explore positive narratives related to Islam to achieve a more balanced political discourse in Europe.

Copyright © 2024, The Author(s).

This is an open access article under the CC-BY-SA license



How to cite: Lestari, T. E., Anwar, M., & Prihantoro, P. (2024). THE Representation of Islam in European Parliament Sessions: A Corpus Study of 'Europarl 3: German'. *Jurnal Ilmiah Global Education*, 5(4), 2100–2108. <https://doi.org/10.55681/jige.v5i4.3504>

INTRODUCTION

In Era 5.0, research involving corpus studies has become increasingly widespread, considering its ability to support various fields. Corpus studies are a linguistic approach that utilizes large collections of texts (corpora) that have been processed to empirically and systematically analyze language. This approach differs from traditional text analysis as it relies on data that can be computed and processed computationally (Baker, 2006). This is further emphasized by Tognini-Bonelli (2001), who explains that corpus studies enable the empirical and systematic analysis of language by utilizing large, computationally processed text collections.

For example, research utilizing the corpus study approach was conducted by Irschara in her study titled "Using a Corpus-Assisted Discourse Studies Approach to Analyse Gender: A Case Study of German Radiology Reports" (2023). In this study, corpus analysis and discourse analysis were used to examine gender representation in German radiology reports, focusing on linguistic patterns and potential bias in discourse surrounding male and female patients in a substantial medical corpus. In another field, Zhao (2022) also used the corpus study approach. Her research investigated gender representation in English as a Foreign Language (EFL) textbooks for Chinese secondary schools, analyzing gendered words and pronouns using corpus linguistic techniques. This study highlighted a tendency towards gender-neutral terms but also revealed persistent gender stereotypes and the underrepresentation of women in these materials.

Corpus studies offer significant advantages across various fields, particularly in vocabulary acquisition, communication, and syntactic analysis. By leveraging large datasets, researchers can more freely uncover patterns to enhance the understanding of language usage. According to Luyun & Liu (2023), corpus-based studies reveal patterns in language usage, provide empirical data on learner strategies, and enable the analysis of various reading behaviors. This approach enhances understanding of reading processes and informs targeted teaching strategies for English language learners. More broadly, corpus-based approaches can foster interdisciplinary collaboration among academics, facilitate more detailed textual analysis, and offer broader multilingual perspectives by examining translated texts, as demonstrated in a case study on Adventures of Huckleberry Finn (2022).

Given its advantages, such as utilizing large data sources, this approach is increasingly applied to analyze online media. Yizhang (2022) discussed the construction of a sarcasm corpus on the Chinese Internet, consisting of 2,000 labeled texts. This corpus aids in the study and detection of sarcasm, addressing the lack of open resources for linguistic research and sentiment analysis. Another example is The Research of English Teaching Corpus Construction Based on Internet and Information Technology (Depeng, 2021). This study discusses the development of an English teaching corpus leveraging the Internet and information technology, highlighting its advantages over traditional methods in terms of professionalism, timeliness, and flexibility, ultimately enhancing student engagement with traditional English and Chinese cultures.

Currently, research utilizing the CQPweb platform for corpus analysis is still relatively scarce, although many studies reference it. A comprehensive study on CQPweb was conducted by Hardie in 2012. CQPweb is a sophisticated web-based corpus analysis tool designed to balance usability with analytical depth. This tool employs a flexible data model compatible with various corpora, enabling users to perform diverse analyses such as concordance, collocation,

and frequency lists. Built on IMS Open Corpus Workbench and MySQL, it allows advanced queries while remaining accessible to untrained users.

Hardie (2012) highlighted CQPweb's flexibility, noting that its general data model supports all types of corpora, enhancing its usability across research domains. He also emphasized the platform's user-friendly interface, adapted from BNCweb, making navigation and feature utilization straightforward. Other strengths include comprehensive analysis options, allowing users to perform distribution tables, generate charts, and identify keywords, facilitating in-depth linguistic research. Additionally, CQPweb supports dynamic bilingual corpora integration for translation studies, enhancing bilingual translation memory systems (Tianmei, 2020). Furthermore, the platform can support the creation of knowledge graphs from large document corpora, aiding in data exploration and fact extraction (Staar et al., 2020).

According to the Pew Research Center (2017) report *The Future of World Religions: Population Growth Projections, 2010-2050*, Islam is the second-largest religion globally, following Christianity. The study estimates that the number of Muslims was around 1.8 billion in 2015 and is projected to grow to approximately 2.76 billion by 2050. These results align with findings from the World Religion Database (2020) and the United Nations' World Population Prospects (2017).

The representation of Islam and Muslims is a multifaceted issue shaped by various media, artistic expressions, and political discourses. This representation often oscillates between negative stereotypes and more nuanced portrayals, influenced by cultural contexts and historical events. American print media frequently uses negative descriptors for Muslims, associating them with terrorism and extremism, contributing to distorted public perceptions (Jamil et al., 2023). Western media often portrays Muslim women through an essentialist lens, while Indian media tends to sensationalize their issues, focusing on personal laws and education, but failing to capture emerging narratives of Muslim women's agency (Jaiswal & Kapoor, 2024).

While dominant narratives often lean towards negative representations, a growing movement among Muslim artists and activists seeks to reclaim their identity and present more complex and positive images of Islam and Muslims. This shift underscores the potential of art and media to foster understanding and challenge stereotypes. Currently, issues related to the representation of Islam are more frequently discussed in countries like the United States, India, or France. Based on these findings, this research aims to explore a corpus study approach concerning the narratives or issues surrounding Islam's representation in Europe, specifically in the European Parliament corpus *Europarl 3: German*. This corpus contains transcripts of debates in the European Parliament, representing the language used in formal settings. This study is expected to provide new insights and perspectives on the representation of Islam globally using a corpus study approach through the CQPweb platform.

METHODS

Using the *Europarl 3: German* corpus through Corpus Query Processing (CQP), the research process for the study on the representation of Islam in European Parliament sessions consists of a series of structured procedures. Each stage is designed to ensure a thorough and methodical examination of linguistic material. A detailed description of the research process is provided below:

1. Corpus Selection and Preparation:
 - a. Select Relevant Corpus: Choose the Europarl 3: German corpus because it contains transcripts of European Parliament sessions relevant to the research topic and has not been previously utilized.
 - b. Retrieve Relevant Texts: Use queries focused on the keyword "Islam" to explore European perspectives on Islam as a major religion.
 - c. Organize Data: Filter texts based on session dates, themes, and speakers to facilitate further analysis.
2. Using CQPweb for Linguistic Analysis:
 - a. Access CQPweb: Utilize the CQPweb interface for linguistic analysis, including collocation and concordance analysis.
 - b. Keyword and Frequency Analysis: Perform keyword searches and frequency analyses to identify the most frequently occurring terms related to Islam. In this study, the four highest frequencies are used, with each word appearing more than ten times.
 - c. Collocation and Concordance Analysis: Focus on the top four collocations as they are essential to understanding the overall interpretation of Islam in the European Parliament. Use a ± 3 word range from the keyword for collocation analysis to reinforce the contextual meaning of collocations. In the case of German, a ± 1 word range is not feasible due to the presence of articles or grammatical gender, which typically occupy those positions.
3. Mapping and Coding Categories:
 - a. Define Categories Based on the Muslim Sketch Framework (Baker, 2012): Use categories such as Conflict, Religion, Culture, and Ethnic/National Entities.
 - b. Map Collocations to Categories: Assign terms to relevant categories, ensuring that collocations such as "Islamic terrorism" fall under the Conflict category, while "Islamic community" is categorized as Group/Organization.
 - c. Refine Coding: Adjust or recategorize terms based on the mapping process.
 - d. Recheck Data: Verify data consistency to ensure more reliable analysis.

RESULTS AND DISCUSSION

- A. The sentiments associated with collocations related to Islam that appear in European Parliament sessions within the "Europarl 3: German" corpus

The focus of this analysis is to examine the sentiments associated with key collocations related to Islam that appear in European Parliament sessions. The analysis utilizes the "Europarl 3: German" corpus, identifying and analyzing specific terms frequently associated with Islam to determine their sentiment.

The research involved selecting relevant texts from the Europarl 3: German corpus. This selection focused specifically on discussions where Islam was mentioned, ensuring the data was pertinent to the study's objectives. After identifying these texts, collocation analysis was conducted to examine the terms most frequently associated with Islam. Sentiment analysis was then performed on these collocations to determine whether they carried positive (POS), negative (NEG), or neutral (NET) sentiments. The analysis prioritized four key collocations: "Terrorismus" (terrorism), "Religion" (religion), "Christentum" (Christianity), and "Europa"

(Europe). These terms were chosen based on their high frequency and relevance in discussions surrounding Islam, as they provided significant insights into the framing of Islam within the European Parliament discourse.

The collocations identified during the analysis were systematically examined to determine their frequencies and contextual associations. Each collocation was assigned a sentiment label—positive, negative, or neutral—based on the specific context in which it appeared within the corpus. The findings revealed that "Terrorismus" was predominantly linked to negative sentiment, highlighting the frequent association of Islam with themes of conflict and security. In contrast, "Religion" and "Christentum" exhibited more varied sentiments, ranging from positive to neutral, depending on the context of the discussions. Similarly, "Europa" reflected neutral to slightly positive sentiments, often in contexts related to cultural integration or interfaith dialogue. The frequency distribution of sentiments for each collocation is summarized below, providing a clear overview of how these terms were framed within the corpus.

Table 1. Collocation Sentiment

COLLOCATE	POSITIVE	NEGATIVE	NETRAL	TOTAL OCCURANCE
TERRORISMUS	0	4	0	4
RELIGION	10	4	2	16
CHRISTENTUM	1	10	8	19
EUROPA	4	6	4	14

The analysis revealed that "Terrorismus" (terrorism) is strongly associated with negative sentiment, particularly in discussions where Islam is explicitly mentioned. This frequent collocation underscores a tendency to frame Islam within the context of conflict and security. In contrast, "Religion" (religion) exhibits a more balanced sentiment distribution. Positive references are notably significant, especially in discussions focusing on the peaceful aspects of Islam. This suggests that, while negative associations exist, there are also attempts to highlight Islam's role in fostering peace and coexistence. The terms "Christentum" (Christianity) and "Europa" (Europe), when analyzed together, tend to convey neutral to slightly positive sentiment. These terms frequently appear in contexts that discuss cultural integration and interfaith dialogue, reflecting a more inclusive narrative compared to the framing of "Terrorismus". The analysis also uncovered some unexpected patterns. Despite the frequent negative collocation of "Islam" (Islam) with "Terrorismus" (terrorism), there are instances where Islam is framed positively. These positive associations are particularly evident in discussions about religious tolerance, suggesting that Islam is not exclusively viewed through a conflict-driven lens.

This duality may reflect the complexities of European Parliament discourse, where Islam is simultaneously addressed in the context of security issues and cultural integration. Another unexpected finding is the positive framing of "Religion" (religion) when referring to Islam. This occurs primarily in discussions emphasizing religious tolerance and diversity. Such framing may indicate an underlying effort to acknowledge the variety of religious beliefs and the importance of fostering dialogue within European communities.

The findings align with Baker's (2012) observations regarding the complex representation of Islam in political discourse, where the religion is often linked to both

conflict and the promotion of peace. Additionally, the literature highlights that political discourse in European contexts, especially within legislative bodies like the European Parliament, mirrors broader societal tensions between security concerns and multiculturalism (Wodak, 2015). The nuanced results of this study reveal dual narratives within European Parliament discussions. On one hand, the association of Islam with "Terrorismus" underscores prevailing concerns about security. On the other, positive references to "Religion" acknowledge Islam's peaceful dimensions and contributions to societal harmony. This duality reflects the ongoing balancing act in European political discourse between addressing security challenges and promoting religious tolerance.

B. Semantic distribution of collocations related to Islam in the corpus

This study involved extracting collocations associated with the term "Islam" from the Europarl 3: German corpus using the Corpus Query Processor (CQP) web interface. After extracting these collocations, the next step was to categorize them semantically into predefined categories that reflect societal attitudes and frameworks associated with Islam. The semantic categories were defined based on existing frameworks, such as Baker's (2012) categorization method for analyzing representations, which include categories like Conflict, Religion, Culture, and Ethnic/National Identity.

Results of the Collocation Distribution Analysis:

The analysis results are divided into several categories, including:

1. Conflict: Words like "terrorismus" (terrorism) frequently co-occur with "Islam," indicating discussions that often associate Islam with conflict.
2. Religion: Terms like "Christentum" (Christianity) and "Religion" reflect comparative discussions between Islam and other religions.
3. Ethnic/National Identity: Collocations such as "Europa" (Europe) highlight a focus on the relationship between Islam and European identity or policies.
4. Culture and Social Practices: Terms related to social and cultural aspects, such as "Frauen" (women), emphasize discussions on cultural practices often linked to Islam.

Visualization & Tables:

A distribution chart illustrates the frequency representation of each category based on the occurrence of collocations.

Semantic Category	Frequency
Conflict	25
Religion	18
Ethnic/Identity	15
Culture	10

Table 2. Semantic Category

The table above demonstrates that discussions linking Islam with conflict occur more frequently than those in other categories.

C. Unexpected Results and Possible Explanations

One unexpected result is the strong association between Islam and terrorism (i.e., the high frequency of "terrorism"). This indicates a bias or focus on the negative representation of

Islam, potentially reflecting the sociopolitical context in which these sessions were held. Additionally, the relatively lower occurrence of positive or neutral terms related to Islam, such as those discussing religious coexistence or multiculturalism, suggests a skewed focus on securitization and conflict-related discourse.

This result may be influenced by specific historical contexts during the corpus timeframe, such as increased attention to security issues in Europe post-9/11 or during debates about Muslim community integration.

D. Connecting Results to Broader Literature

These findings align with previous research by Baker (2012), which observed that media and political discourse often frame Islam in the context of conflict and security, particularly in Western contexts. The emphasis on the collocation "terrorism" is consistent with the literature on Islamophobia, where Islam is frequently framed as a security threat in European public discourse (Richardson, 2004).

The prevalence of collocations associated with "Christentum" (Christianity) and "Europa" (Europe) reflects Huntington's (1993) theory of the "clash of civilizations," wherein Islam is culturally and ideologically positioned as distinct from the West.

CONCLUSION

The analysis highlights that European Parliament sessions, as reflected in the Europarl 3: German corpus, often associate Islam with negative collocations. Terms such as "Terrorismus" (terrorism) and "Extremismus" (extremism) are commonly linked to Islam, indicating a framing of the religion within a securitized and conflict-driven narrative. Positive or neutral representations, such as references to cultural contributions, integration, or community, are notably less frequent. This framing underscores a tendency to associate Islam primarily with negative themes, reinforcing stereotypes and potentially shaping public perceptions.

The semantic categorization of collocations further illustrates a strong emphasis on conflict and security-related themes. Categories such as Culture, Social Practices, and Religion are underrepresented, which reinforces the dominance of securitization over inclusive narratives about diversity and integration. These findings reveal how the language used in political discussions can influence societal attitudes, particularly toward minority groups like Muslims. By framing Islam predominantly through the lenses of conflict and extremism, parliamentary discourse risks perpetuating Islamophobic sentiments and contributing to social polarization.

The study emphasizes the importance of language choice in political discourse, particularly in influential forums such as the European Parliament. Recognizing the cultural and social contributions of Muslim communities and fostering balanced discussions about Islam could contribute to greater social cohesion. The observed securitization of Islam reflects broader sociopolitical anxieties in Europe, suggesting a need for deliberate efforts to promote multicultural understanding and shared values.

Future research could expand beyond the German-language Europarl 3 corpus to explore discussions about Islam across other European languages. A cross-linguistic approach would provide comparative insights into how Islam is framed in different linguistic and cultural contexts within Europe. Additionally, longitudinal studies examining how parliamentary discourse on Islam has evolved over time could offer valuable insights. These studies could assess changes in framing, particularly in response to significant sociopolitical events such as terrorist attacks or shifts in immigration policy.

Research could also focus on identifying positive or neutral narratives about Islam, such as references to cultural contributions, social cohesion, and integration efforts. Amplifying these narratives could provide a counterbalance to the predominant negative discourse. Furthermore, studies exploring how parliamentary language influences public perceptions of Islam could provide empirical evidence of the broader impact of political discourse. Such studies could incorporate surveys, interviews, and sentiment analyses on social media platforms to assess the relationship between political language and public attitudes.

The study underscores the importance of balanced discourse in legislative and political contexts. Discussions about Islam in the European Parliament should move beyond a narrow focus on security concerns to acknowledge the cultural and social contributions of Muslim communities to European society. Policymakers should consider the influence of their language choices and strive to promote inclusive and balanced narratives.

REFERENCES

- Andrew, Hardie. (2012). CQPweb — combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380-409. doi: 10.1075/IJCL.17.3.04HAR
- Baker, P. (2012). *In other words: A coursebook on translation* (2nd ed.). Routledge.
- Bao, Tianmei. (2020). Research on English Vocabulary Translation Technology Based on Dynamic Bilingual Corpus. 288-291. doi: 10.1145/3419635.3419729
- Bleich, E. (2011). *The legacies of integration: The politics of race and nation in Europe*. Oxford University Press.
- Dapeng, Wang. (2021). The Research of English Teaching Corpus Construction Based on Internet and Information Technology. doi: 10.1145/3482632.3483069
- Huntington, S. P. (1993). *The clash of civilizations and the remaking of world order*. Simon & Schuster.
- Karoline, Irschara. (2023). Using a Corpus-Assisted Discourse Studies Approach to Analyse Gender: A Case Study of German Radiology Reports. 23(2):114-139. doi: 10.13060/gav.2022.015
- Luyun, Liu., Xuan, Li. (2023). A Corpus-Based Study of English Learners' Reading Strategies. doi: 10.54691/gwj76g68
- Peter, Staar., Michele, Dolfi., Christoph, Auer. (2020). *Corpus Processing Service: A Knowledge Graph Platform to perform deep data exploration on corpora*. 1(2) doi: 10.1002/AIL2.20
- Pew Research Center, *The Future of World Religions: Population Growth Projections, 2010-2050*.
- Richardson, J. E. (2004). The representation of Islam in the British press. *Journal of Media Studies*, 12(3), 145-158.
- United Nations, *World Population Prospects: The 2017 Revision*.
- Wodak, R. (2015). *The politics of fear: What right-wing populist discourse means*. SAGE Publications. World Religion Database, 2020 Edition.
- Xiaowen, Zhao. (2022). A Corpus Study of Gender Representation in Chinese Secondary School EFL Textbooks. *International journal of studies in English language and literature*, 10(11):8-18. doi: 10.20431/2347-3134.1011002

- Yizhang, Zhu. (2022). Open Chinese Internet Sarcasm Corpus Construction: An Approach. *Frontiers in Computing and Intelligent Systems*, 2(1):7-9. doi: 10.54097/fcis.v2i1.2484 (
- _____(2022). Benefits of a Corpus-based Approach to Translations: The Example of Huckleberry Finn. 176-190. doi: 10.4324/9781003298328-11