

A LITERATURE REVIEW OF CORPUS LINGUISTICS RESEARCH IN REPRESENTATION OF HALLYU

DUNG QUOC DANG

Haiphong University of Management and Technology | dangquocdung2232000@gmail.com

1. INTRODUCTION

Hallyu is defined as "the growing popularity of all things Korean, from fashion and film to music and cuisine" (Nye, 2009). Since the early 2000s, Hallyu has gained popularity not only in its home country but also in neighboring East Asian countries, eventually becoming a worldwide cultural phenomenon. The reputation of Hallyu started from the premiere of Korean television dramas and Korean singers' performances in China (Hong, Park, Park & Oh, 2019; Samosir & Wee, 2024) and Hallyu gradually expanded through social media platforms (Hong, Park, Park & Oh, 2019). Furthermore, it has political, symbolic, and economic significance. As a result, its growth defines it not just as a form of entertainment but also as a negotiation of soft power and a national branding strategy.

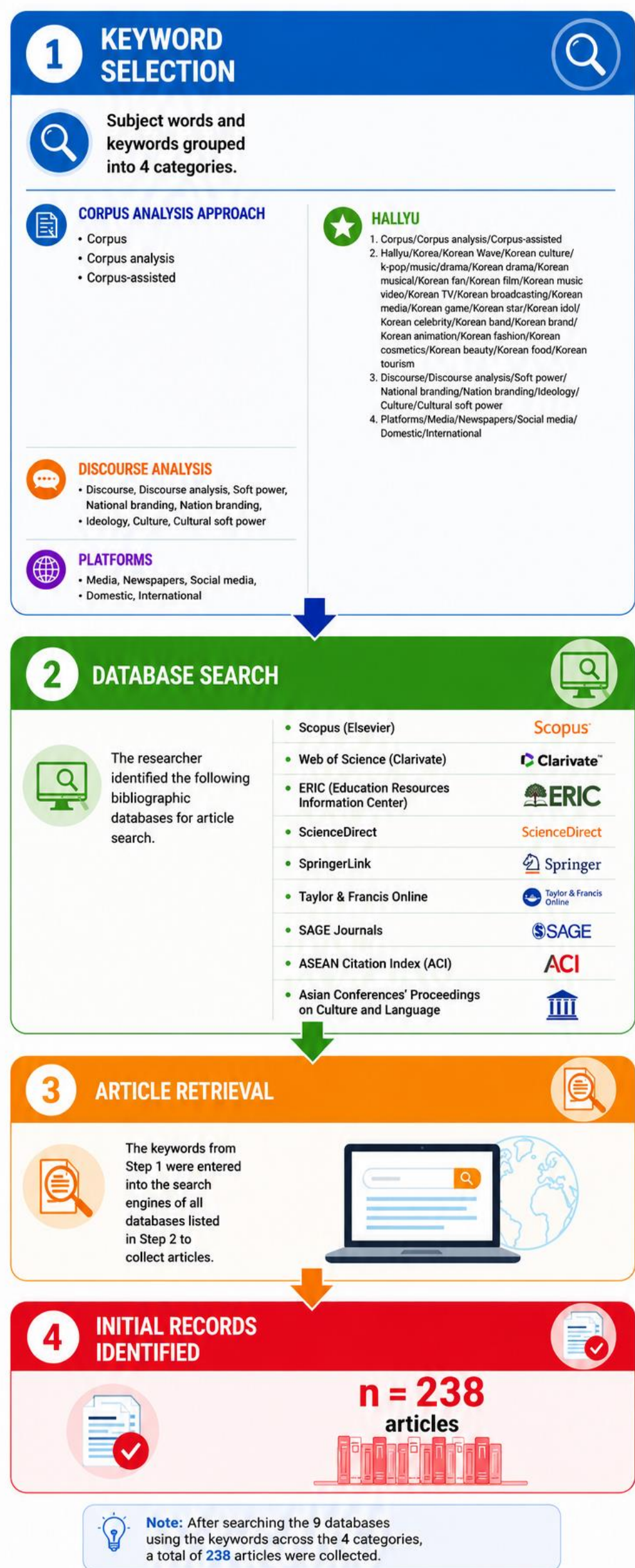
Although there are a certain number of corpus-based studies on the global representation of Hallyu in international discourse, there are limited systematic literature reviews in this research topic. Therefore, this study aims to fill in this gap by reviewing 10 corpus-based studies in the representation of general Hallyu wave in international corpus, specifically in terms of how corpus reveal international audience perceive this cultural trend and construct discourse on the country, in order to reveal understanding on the trends of corpus-based studies of Hallyu as well as future directions for this research aspect.

1. What corpora and tools were used in previous studies?

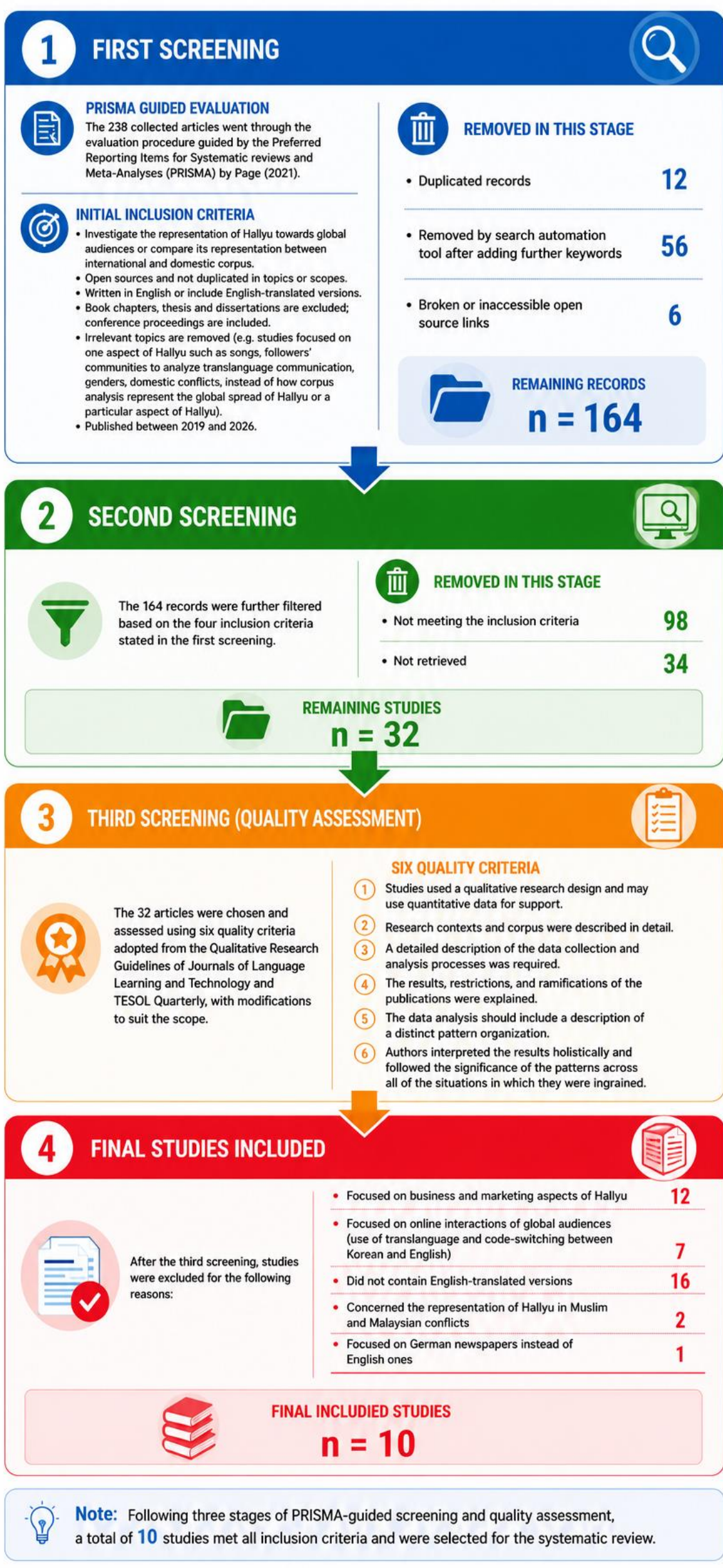
2. From what theoretical perspective have researchers conducted corpus-based studies in the global representation of Hallyu?

2. PRISMA SELECTION PROCEDURE OF CORPUS-BASED STUDIES

2.1. Initial article collection (Identification stage)

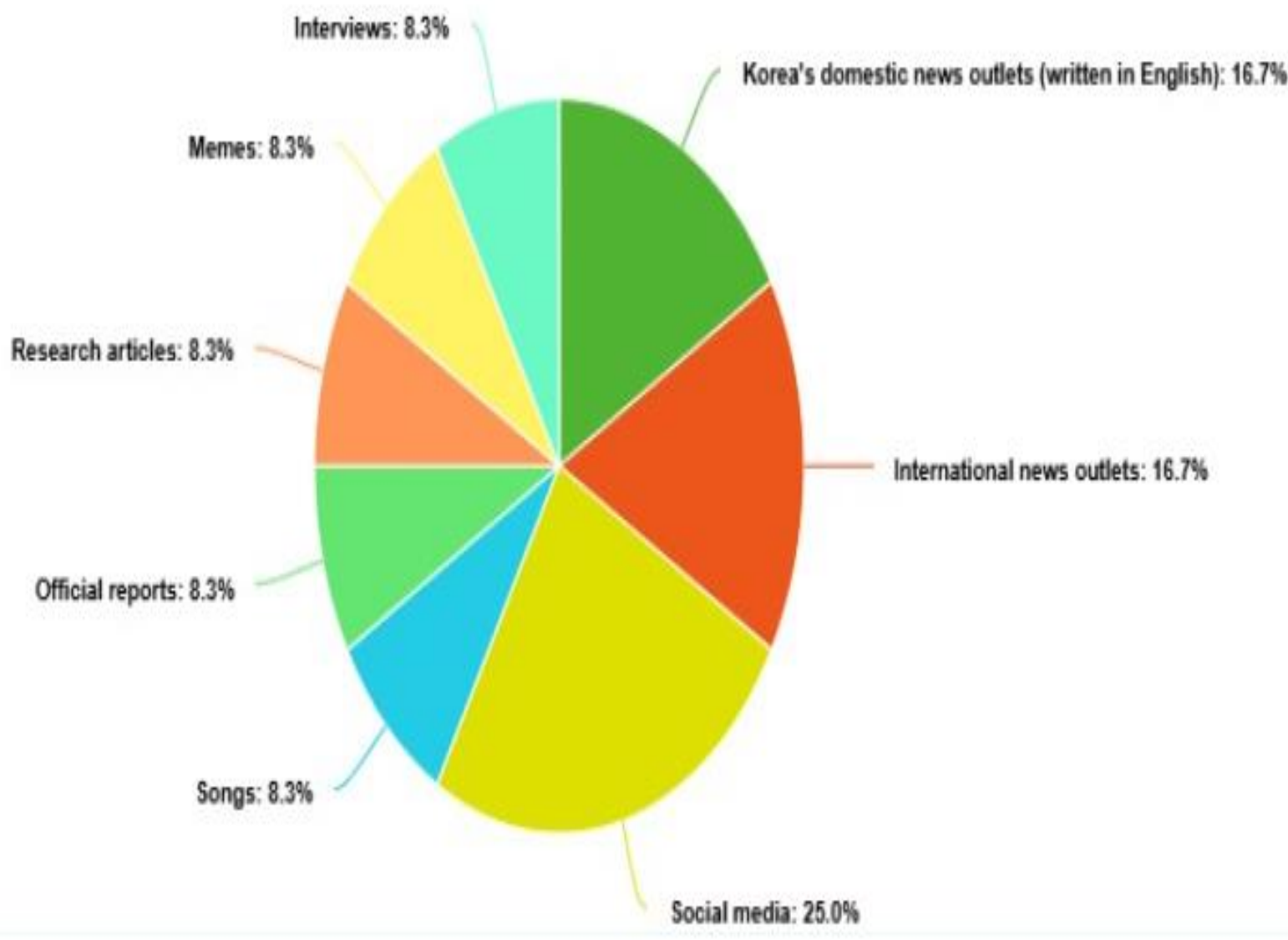


2.2. Data evaluation for collected articles (Screening stage)

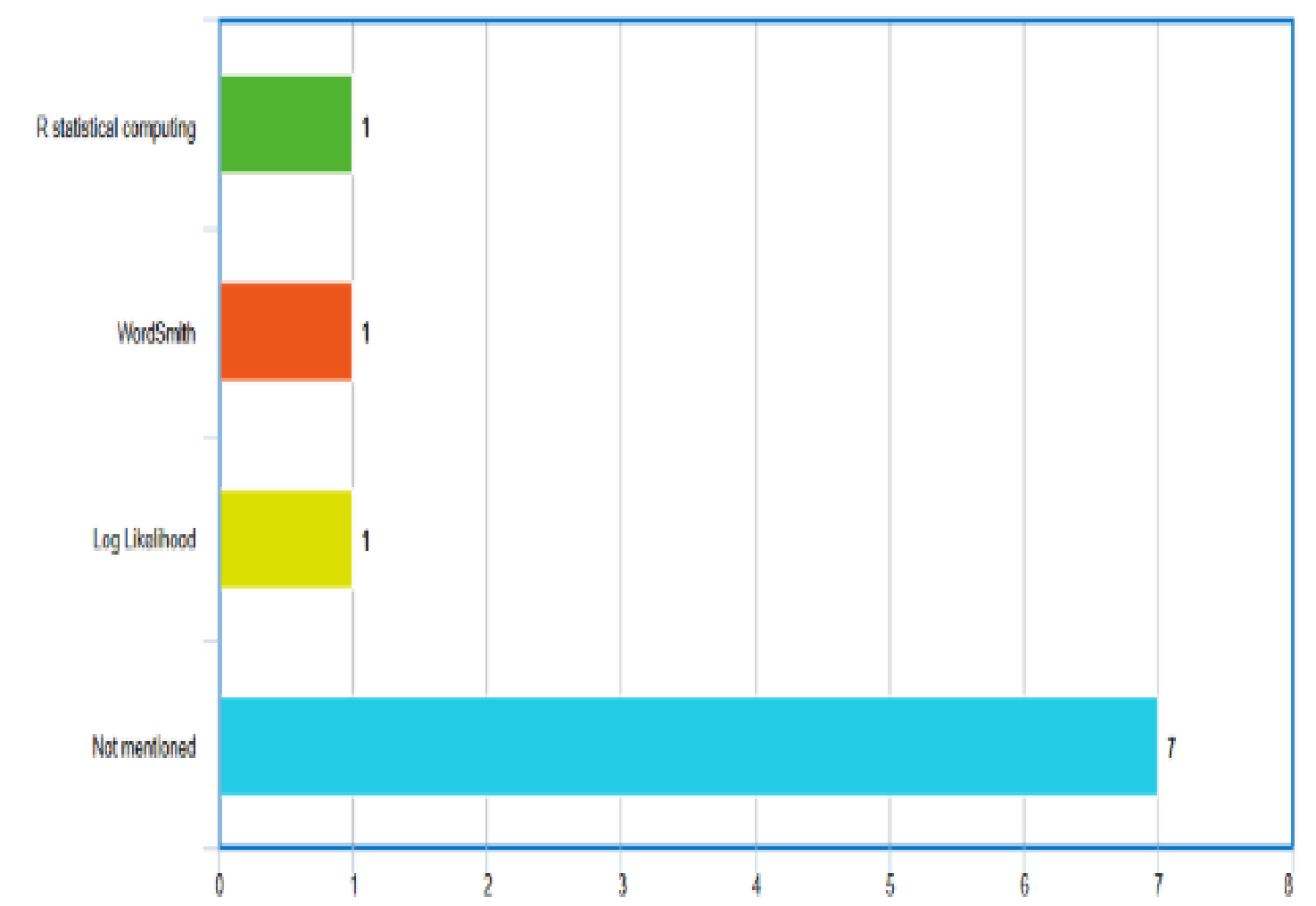


3. ANALYSIS

3.1. Corpus sources



3.2. Tools employed



3.3. Theoretical and data analysis frameworks

SUMMARY OF THEORETICAL AND ANALYTICAL FRAMEWORKS OF STUDIES			
AUTHORS & YEARS	THEORETICAL FRAMEWORK	ANALYTICAL FRAMEWORK	
1 Rafilah, Triarisanti & Widyana (2025)	Framing Theory (Entman, 1993)	➔	Pan & Kosicki's (1993) framing model, analyzing syntax, script, thematic, and rhetorical to show how a reality is structured in news
2 Moon (2023)	Critical Discourse Analysis by Fairclough & Wodak (1997) and Van Dijk (1985, 1988, 2000, 2006, 2009) and Framing Theory by Goffman (1974)	➔	Corpus-assisted Critical Discourse Analysis to analyze how news frame reality, with corpus linguistic tools for statistical analysis, along with qualitative description
3 Jin (2024)	Critical Discourse Analysis to describe cultural soft power	➔	Not mentioned
4 Hidayati (2024)	Critical Discourse Analysis, Discourse Historical Analysis to analyze the construction of a social reality, in connection to audiences' response	➔	Not mentioned
5 Tempora (2024)	Sociolinguistic Theory by Wardhaugh & Fuller (2021) and Functional Theory of Language (Halliday, 2009)	➔	Content Analysis for sociolinguistic approach to analyze sociolinguistic phenomena, registers
6 Yihuan, Chong, Yap & Juan (2025)	Critical Discourse Analysis by Fairclough & Wodak (1997) and Van Dijk (1985, 1988, 2000, 2006, 2009) to analyze legitimization, based on Van Leeuwen's (2008) legitimization categorisation	➔	Longitudinal Critical Discourse Analysis to study changes of discourse over different periods
7 Hong, Park, Oh & Park (2019)	Not mentioned	➔	Literature meta-analysis, Knowledge network analysis to analyze patterns of coding
8 Yoon (2023)	Critical Discourse Analysis to reveal underlying cultural soft power	➔	Not mentioned
9 Barnes-Sadler, Ahn & Kiaer (2025)	Not mentioned	➔	Corpus Linguistics to analyze frequencies of coding
10 Valljärvi & Khan (2025)	Fauconnier's (1997) Conceptual Blending Theory	➔	Qualitative corpus-based analysis for analyzing semiotics of delivering messages and their psychological, social functions

4. DISCUSSION

4.1. Diversity and Representativeness of Corpora

Corpus diversity reveals varied representations, ideologies, and perspectives of Hallyu. Social media dominates as a corpus source, shaping audience and governmental discourses, while newspapers highlight cultural soft power. However, gaps remain in governmental studies, phase-based analysis, and unconventional corpus sources such as academic research.

4.2. Efficiency of Tools employed

The selected studies showed limited use of corpus analysis tools, with few reporting quantitative and qualitative methods. Future research should employ corpus software and mixed-methods approaches to improve analytical transparency, efficiency, reproducibility, and the comparison of Hallyu representation across different developmental phases.

4.3. Theoretical and analytical frameworks

Selected studies employed diverse theoretical and analytical frameworks, with Critical Discourse Analysis emerging as the dominant approach. However, analytical frameworks remain inconsistent and often underreported. Future corpus studies should adopt transparent mixed-methods designs and clearly defined analytical frameworks to improve consistency and comparability.

4.4. Further gaps and future direction

- Knowledge gap: The lack of diachronical corpus studies on Hallyu of different developmental phases.
- Population gap: The lack of studies on Hallyu discourse representation from unconventional corpus sources, such as songs, or academic articles.
- Methodological gap: The lack of corpus studies on Hallyu discourse representation that employ mixed-methods of qualitative thick description of corpora and quantitative description of linguistic frequency.
- Analytical/Theoretical gap: Several studies lack reference to the type of analytical or theoretical gap adopted.

5. CONCLUSION

This meta-literature analysis shows that corpus linguistics is an effective approach for examining Hallyu representation, particularly as a form of cultural soft power and globalization. However, existing studies remain methodologically fragmented, rely mainly on newspaper and social media corpora, and rarely adopt longitudinal perspectives. Future research should employ longitudinal mixed-methods designs and apply consistent theoretical frameworks, such as Critical Discourse Analysis, Corpus-Assisted Discourse Studies, framing analysis, and social semiotics, to strengthen the field.

Screening process, based on the PRISMA procedure (Page, et al. 2021)

