



## A quantitative characterisation of English in K-Pop through the Korean Wave 2000–2020

Simon Barnes-Sadler, Hyejeong Ahn & Jieun Kiaer

**To cite this article:** Simon Barnes-Sadler, Hyejeong Ahn & Jieun Kiaer (21 Jul 2025): A quantitative characterisation of English in K-Pop through the Korean Wave 2000–2020, Asian Englishes, DOI: [10.1080/13488678.2025.2533537](https://doi.org/10.1080/13488678.2025.2533537)

**To link to this article:** <https://doi.org/10.1080/13488678.2025.2533537>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 21 Jul 2025.



Submit your article to this journal [↗](#)



Article views: 10



View related articles [↗](#)



View Crossmark data [↗](#)

# A quantitative characterisation of English in K-Pop through the Korean Wave 2000–2020

Simon Barnes-Sadler <sup>a</sup>, Hyejeong Ahn <sup>b</sup> and Jieun Kiaer <sup>a</sup>

<sup>a</sup>Faculty of Asian and Middle Eastern Studies, University of Oxford, Oxford, UK; <sup>b</sup>Faculty of Education, University of Melbourne, Melbourne, Australia

## ABSTRACT

This paper provides an, empirical account of changes in the prevalence of English in K-Pop lyrics from 2000 to 2020. It contrasts with earlier, smaller-scale investigations by leveraging computational techniques and a bespoke corpus of over 2000 songs. We find the prevalence of English, i.e. raw number of English tokens, proportion of English tokens relative to Korean tokens, and proportion of songs including English tokens, has risen over the surveyed period. We also present two case studies. The first suggests this rise is attributable to changes in songwriting practice, rather than artists who use less English being replaced by those who use more. The second finds English translation equivalents do not displace Korean words thematically central to K-Pop. Our results support the hypothesis that the observed increase is motivated, at least partly, by the Korean Wave. Further, we highlight the complementarity of quantitative and qualitative approaches.

## ARTICLE HISTORY

Received 20 December 2024  
Accepted 6 July 2025

## KEYWORDS

Korean English; Korean wave; popular culture; lyrics

## Introduction

This paper aims to provide a rigorous, empirical account of the change in the prevalence of English in the lyrics of South Korean pop music (hereafter K-Pop) over the early twenty-first century.

The consumption of and engagement with K-Pop has increased enormously across the English-speaking world over the past two decades. This phenomenon is part of the broader, global increase in engagement with South Korean popular culture known as *Hallyu* or The Korean Wave. As a consequence, the influence of the Korean language has been making itself felt throughout the Anglosphere in diverse forms including, the adoption of vocabulary originating in South Korea among fans of K-Pop (Ahn, Kiaer, & Barnes-Sadler, 2025) the inclusion of words of Korean origin in the Oxford English Dictionary (Ahn & Kiaer, 2024), and the explosive growth of formal and informal study of Korean as a Foreign Language (KFL).<sup>1</sup> The direction of linguistic influence has not all been one way. The role of English in South Korean society in general (e.g. Song, 2012; Park, 2021) and particularly within K-Pop has long

**CONTACT** Simon Barnes-Sadler  [simon.barnes-sadler@ames.ox.ac.uk](mailto:simon.barnes-sadler@ames.ox.ac.uk)  Faculty of Asian and Middle Eastern Studies, University of Oxford, 1 Pusey Ln, Oxford OX1 2LE, UK

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

been an object of scholarly interest. Before describing the materials and methods employed to approach the question of how we understand the prevalence of English in K-Pop lyrics and examine its change over the early twenty-first century, we review some previous work on English in K-Pop to contextualise this study and its contribution.

Qualitative analyses of the form and function of English in K-Pop lyrics have been carried out over the early twenty-first century (e.g. Lawrence, 2010; Lee, 2004). While these recognise the frequent appearance of English in K-Pop and the heterogeneity of its use, its prevalence, while remarked upon, is understandably not operationalised quantitatively as it falls outside their scope. Dedicated quantitative analyses of English in K-Pop are rarer. Among those, many focus on song titles (e.g. Ahn, 2021, p. 218) or stage names (e.g. Kang, 2023). A representative example of this tendency may be found in Jin and Ryoo (2012). In this paper, a quantitative analysis of English in K-Pop accompanies an article primarily focused on the cultural hybridisation of K-Pop under conditions of globalisation. Taking the top 50 songs of each year as recorded on the charts of *melon.com* between 1990 and 2010, the proportion of English in the names under which artists performed and song titles were calculated at five-year intervals (i.e. 1990, 1995, etc.). While this revealed an enormous increase in the proportion of artists using English names (from none in 1990 to 41 of the 50 artists appearing on the 2010 chart) and in song titles (from three in 1990 to 22 in 2010), it cannot be interpreted to mean the prevalence of English in K-pop lyrics increased to a similar extent over the surveyed period. Lyrics were discussed in this study at a much smaller scale; only three songs were examined in detail, even then without rigorous quantification. While Jin and Ryoo take the appearance of English in 28 of the top 50 K-Pop songs in December 2010 to indicate that ‘the amount of English mixing in K-Pop lyrics has soared’ (p. 125), not only is no baseline measure provided as a basis for comparison but the prevalence of English within each song (single word interjections, occasional words, whole verses) is not explored.

Only in recent scholarship has the simple empirical question of the changing prevalence of English in K-Pop been rigorously addressed. I. Schneider (2024) carried out a mixed-methods analysis constructing linear regression and Poisson regression models. These quantitative techniques model the strength and significance of the relationship between two variables. Drawing upon the real-world observations of the release dates of 58 K-Pop songs by four internationally popular K-Pop artists and the proportion of unique English headwords (‘types’ in the terminology of corpus linguistics) appearing in them, the model attempts to ‘predict the percentage of English types as a function of release date’ (I. Schneider, 2024, p. 107). Taken in aggregate, a statistically significant increase in the proportion of English types was found over the surveyed period (2015–2022). This is a clear indication that English has recently increased in prevalence in K-Pop lyrics. The current study investigates the possibility of expanding the generality of that assertion by looking over a longer period and including more artists. Furthermore, our analysis focuses on the lyrics of songs released as constituent parts of albums which attained high positions in South Korean annual charts. This contrasts with I. Schneider’s focus on the lyrics of songs released as singles produced by selected international breakout artists. As such, our results provide a greater indication of the position of English in the South Korean domestic pop-music landscape. We include the important caveat, though, that our

results represent a high-level overview, rather than nuanced reflections of artists' or listeners' lived experience.

## Materials and methods

The main question this paper tackles is whether the prevalence of English in K-Pop lyrics has changed over the early twenty-first century. That is, has the quantity of English used in K-Pop lyrics increased over this time and how does this relate to the quantity of Korean used in the lyrics? To answer this question, we constructed a collection of song lyrics for quantitative analysis, drawing on the frameworks of corpus linguistics (for overviews, see McEnery & Wilson, 2001 or Hardie, 2016; for concrete methods, see Desagulier, 2017 or Brezina, 2018) and text analysis (for overviews see Fiordistella Iezzi & Celardi, 2020 or Schweinberger, 2024). For concrete methods, see Schneider, 2024), as detailed in this section.

Our collection of texts consists almost in its entirety<sup>2</sup> of user-generated transcriptions hosted on *music.bugs.co.kr* as of November 2023 for the lyrics to the songs collected in the top 10 best-selling albums of each year from 2000 to 2020 according to the Circle/Gaon Chart and the Recording Industry Association of Korea (RIAK).<sup>3</sup> This period notably stretches from the inception of the Korean Wave, through its growth, and into its establishment as a global phenomenon. The outcomes of increased international engagement with South Korean popular culture are necessarily complex and reflexive. This international engagement may provide a commercial incentive to incorporate additional English into K-Pop to access a larger, global market for songs more easily understood by English-speaking people. Simultaneously, as noted by Yeo (2018, p. 96), being founded upon engagement with specifically Korean popular culture, the global consumption and distribution of K-Pop that constitutes, and is perpetuated by, the Korean Wave leads to a tension between the global and the local, that is, between English and Korean. By embracing the accessibility enabled by an international language of wider communication, such as English, too enthusiastically, K-Pop risks shedding one of the qualities upon which its global popularity rests: its indexing of contemporary South Korea. Acknowledging that correlation does not prove causation, the relationship between the Korean Wave and the changing prevalence of English in K-Pop suggested in prior work is a consideration motivating both the selection of this timeframe and, indeed, this study as a whole. We return to this theme in the discussion section.

Turning to data collection, while the number of songs sampled constitutes only a small fraction of the amount of popular music released in South Korea, it represents a number of songs orders of magnitude larger than have been considered to date. Furthermore, as it consists of albums purchased, played on-air, and streamed enough across multiple services to chart highly in year-end charts, it may be considered representative of the musical and lyrical landscape of K-Pop for the surveyed period.

There were, however, some exclusion criteria. Albums released explicitly in languages other than Korean were excluded from the corpus. Also, songs re-released with no modification on so-called 'repackaged' albums which also appeared on the first issue release of those albums (if the first issue also appeared in the same year's top 10 best-selling albums) were removed to avoid counting the same data twice. We further excluded instrumental tracks that did not feature lyrics.

As these transcriptions were based upon user-generated content, each one was manually inspected to ensure accurate, consistent, rather than abbreviated transcription, and amended against recordings as appropriate. A quantitative summary of the corpus can be found in [Table 1](#).

The corpus consists of 2199 transcriptions of lyrics from 196 albums by 215 artists. There are some additional, minor caveats to bear in mind about the quality of the data. The number of songs and artists included each year varies enormously, affecting the balance of this corpus. Particularly striking is the large number of artists represented in the data for 2001 due to the unusual success of a large compilation album that year. The factors underlying the rest of this variation include the following: the change in marketing practices meaning more albums and songs met our exclusion criteria (outlined above) over time, the increasing prevalence of shorter albums (EPs) in the charts over time, and the increasing tendency for single artists to enjoy chart success with multiple album releases in one year (i.e. artists with multiple albums among the top 10 best-selling albums of the year). While it would have been possible to address any one of these factors, we chose not to do so as the impact on data consistency and ecological validity was deemed unacceptable. For example, standardising the number of albums considered would, in combination with our exclusion criteria, have expanded the range of chart positions from which the collection of texts was constructed to those outside the predetermined sampling frame, problematising year-on-year comparisons. Other approaches, for example attempting to include a uniform number of artists, have similar issues for consistency. Following a strategy which drew a sample from the sampling frame (each surveyed year's top 10 albums) to ensure balance would have reduced the ecological validity of our findings; our data reflects the reality of the situation in which the reception individual artists enjoy varies and albums contain variable numbers of songs. Despite these caveats, we

**Table 1.** Summary of number of songs, albums, and artists per year 2000–2020.

Year	Number of Songs	Number of Albums	Number of Artists
2000	147	10	8
2001	183	10	65
2002	142	10	10
2003	142	10	9
2004	125	10	8
2005	133	10	9
2006	136	10	10
2007	160	10	8
2008	98	8	8
2009	98	9	8
2010	77	9	6
2011	89	10	8
2012	72	10	9
2013	62	7	5
2014	80	9	8
2015	67	8	5
2016	81	8	5
2017	70	10	5
2018	76	9	6
2019	67	9	8
2020	96	10	6

assert that the analysis of this sizeable, diverse collection should provide a rigorous account of change in prevalence of English in K-Pop lyrics.

Turning to our concrete methods, text files containing the lyrics to individual songs were coded for year of appearance on the chart, album, and artist, then checked for accuracy. English tokens were extracted using the *quanteda* package (Benoit et al., 2018) in the R statistical computing environment (R Core Team, 2021). For this, we accepted transcribers' orthographic choices as accurate. That is, we understood English to be transcribed using the Roman Alphabet rather than Hangul. We then computed the number of tokens of English and Korean used in each song, calculated the normalised frequency of English per thousand tokens (ptt) per year, and identified the number of songs per year which did not contain any tokens of English. We note the point of difference with I. Schneider (2024) in considering tokens, that is, the concrete word forms which appear in texts with no stipulation that they be unique, rather than types for our analysis. Acknowledging the 'repetitive nature of song lyrics' and the fact that '[c]ounting the number of unique words in a text is more indicative of the complexity of language use than counting all words individually' (I. Schneider, 2024, p. 107), we consider the inclusion of even repetitious lyrics or short interjections of English in K-Pop lyrics significant for indicating the overall prevalence of the language in our collection. We further argue that relying on tokens rather than types better reflects a lay listener's understanding of 'how much' English is in a song.

Summary, aggregate measures of the prevalence of English in K-Pop, such as normalised frequency ptt, erase the variability in the collection by song, album, and artist. Thus, they are supplemented by visualisations to provide more comprehensive, intuitive overviews. Alongside this communicative function, visualisations are important tools for data inspection and interpretation, as they may reveal qualities of the data that are otherwise obscured by quantitative summaries (Anscombe, 1973).

We further include two short case studies to provide complementary perspectives on our observation of a general, aggregated increase in the prevalence of English. The first looks at the change in the average number of tokens of English used in the songs of two artists that appear in the chart over multiple years. This case study provides speculative insights as to whether any changes observed in the prevalence of English are due to the increase in popularity of artists who use different amounts of English or whether individual artists have changed the amount of English they include in their lyrics over time. The second case study examines the change in the relative frequency of the keywords 'love' and *sarang* (love). It provides an initial clue about the extent to which English may be displacing Korean to express concepts thematically central to pop music. As well as supplementing our general findings, this case study provides an illustration of the limitations of quantitative and so-called distant reading (Moretti, 2000) approaches.

## Results

This section is divided into three parts. In the first, we present our global findings based on aggregate analysis of the number of English tokens recorded for each surveyed year. We then turn to our two case studies to further explore the contours of the change in prevalence of English in K-Pop lyrics.

## Aggregate analysis

The raw number of tokens of English, the raw total of tokens combining both English and Korean,<sup>4</sup> and the normalised frequency of English tokens (ptt) for each surveyed year is presented in Table 2.

The information in Table 2 demonstrates the prevalence of English in K-Pop has increased in absolute and relative terms over the first 21 years of the twenty-first century. Our discussion here focuses on the normalised measure as this allows for year-on-year comparison that accounts for the variation of the number of songs considered in each surveyed year. Only twice, in 2001 and 2007, did the normalised frequency of English tokens ptt fall below that observed for 2000. Furthermore, the final two years of the surveyed period contained a proportion of English tokens over three times as great as the first surveyed year. The absolute increase in English tokens is even more extreme. Just over four times as many English tokens appeared in the data for 2020 (14,881) compared with the data for 2000 (3,465). This increase was neither entirely linear nor monotonic, but is starkly discernible in a visualisation which plots the mean number of tokens of English per song per album over the surveyed period, as in Figure 1.

From Figure 1, we can see that none of the albums included in our collection for the second half of the surveyed period featured lyrics exclusively in Korean throughout. The lowest mean number of English tokens per song per album from the year 2011 onwards was 10.6; the highest was 304.5. A further indication of the increase in prevalence of English in K-Pop lyrics is the fact that lowest mean number of tokens of English in this collection for any album which charted in 2020 (74) is higher than that of the album with the highest mean number of tokens of English in this collection for 2000 (45). Finally, it is notable that the range of mean tokens of English per song per album varies considerably between years from just 45 in 2000 to 368.7 in 2010, although the

**Table 2.** Absolute number and normalised frequency per year of English tokens in corpus.

Year	Number of Tokens of English in Corpus	Total number of English and Korean Tokens in Corpus	Normalised Frequency of English Tokens (ptt) in Corpus
2000	3,465	30,592	113.26
2001	2,318	27,438	84.48
2002	4,302	23,124	186.04
2003	4,609	23,960	192.36
2004	5,090	24,886	204.53
2005	2,774	21,073	131.64
2006	4,989	27,106	184.06
2007	2,692	29,470	91.35
2008	4,832	20,437	236.43
2009	10,247	28,478	359.82
2010	8,412	23,074	364.57
2011	7,000	23,049	303.70
2012	4,541	18,213	249.33
2013	3,892	15,565	250.05
2014	4,558	19,573	232.87
2015	4,792	20,127	238.09
2016	7,107	24,089	295.03
2017	6,314	18,483	341.61
2018	6,629	22,173	298.97
2019	6,999	18,944	369.46
2020	14,881	35,771	416.01



**Figure 1.** Mean number of tokens of English per song per album for the top 10 albums in South Korean charts 2000–2020.

small number of albums sampled for each year means that this fact is amenable only to highly speculative interpretation.

We now consider the prevalence of songs in which no English tokens appear to demonstrate that the increase in prevalence of English is occurring at the level of the song as the increase in prevalence of English ptt could be accounted for by a relatively small number of songs with lyrics exclusively or largely in English. This supplementary analysis demonstrates that the prevalence of English is increasing not only in terms of the raw number of tokens or as a proportion of the tokens in the combined English and Korean ‘bags of words’ that each year’s top albums contain, but that English also appears in an ever-greater proportion of individual song’s lyrics. Both the raw number of songs and the proportion of songs in which no English tokens appear as a percentage per year can be seen in [Table 3](#). [Figure 2](#) visualises only the number of songs.

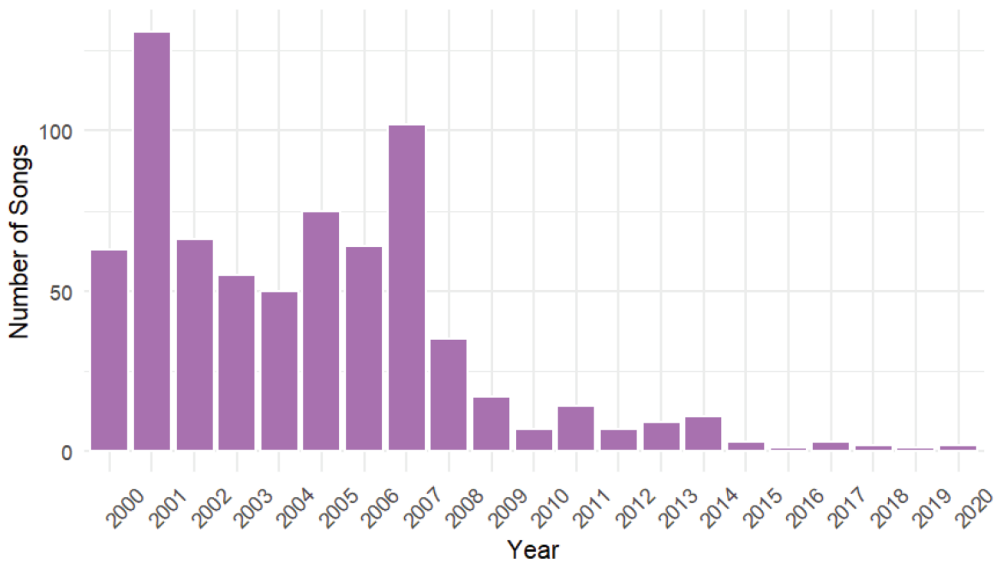
Once more, we see evidence of the increasing prevalence of English. While over one third of all songs included in the corpus for each of the first nine surveyed years did not include a single English token, the number and proportion of songs not to feature English declined sharply from 2009. From 2015 until the end of the surveyed period fewer than one in twenty songs included in the collection featured exclusively Korean lyrics.

This decrease is perceivable in the context of the entire corpus when we visualise the number of tokens of English per song over the surveyed period, as in [Figure 3](#).

In this visualisation, the points represent individual songs. The graphical alpha parameter of the points has been reduced. Thus, areas which appear more opaque represent higher concentrations of points. The reduction in opacity of the points along the bottom of the graph indicates a reduction in the number of songs which do not feature tokens of English over the surveyed period. Prior to 2008, it was common for over 50 songs per year among those included in the collection to contain no tokens of English,

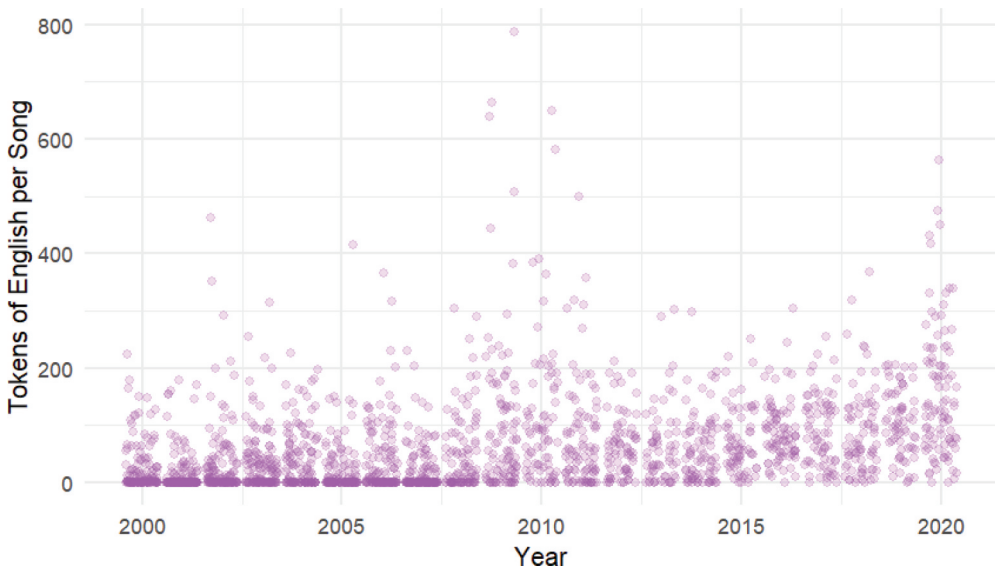
**Table 3.** Number of songs per year with No English tokens.

Year	Number of Songs with No English Tokens	Proportion of Songs (%) with No English Tokens
2000	63	42.9
2001	131	71.6
2002	66	46.5
2003	55	38.7
2004	50	40.0
2005	75	56.4
2006	64	47.1
2007	102	63.8
2008	35	35.7
2009	17	17.3
2010	7	9.1
2011	14	15.7
2012	7	9.7
2013	9	14.5
2014	11	13.8
2015	3	4.5
2016	1	1.2
2017	3	4.3
2018	2	2.6
2019	1	1.5
2020	2	2.1

**Figure 2.** Number of songs with No English tokens on the top 10 albums in South Korean charts 2000–2020.

with a high point of 131 songs in 2001. Since 2008, far fewer songs have contained no tokens of English. The high point was in the first year of that period, 2008, with 35 songs. In nine of the 13 years between then and 2020 (inclusive), fewer than 10 songs without any English lyrics appeared among those included in the collection.

As well as the declining number of songs without tokens of English, individual songs featuring particularly high numbers of tokens of English, the lyrics of which may even be entirely in English or feature English-speaking artists, can also be seen. Representative



**Figure 3.** Number of tokens of English per song for the top 10 albums in South Korean charts 2000–2020.

examples include JYJ’s 2010 song *Ayyy Girl*, featuring Kanye West and Malik Yusef, which contains 649 tokens of English, and a version of NCT’s *Make a Wish (Birthday Song)* which contains exclusively English lyrics (563 tokens). Alongside these extreme outliers in the context of the collection and the strong tendency for the prevalence of English in K-Pop lyrics to increase on aggregate over the surveyed period, it is notable that even at the inception of the Korean Wave there were less extreme outliers which included a number of tokens of English consistent with songs released towards the end of the surveyed period, just as there are K-Pop songs to be found towards the end of the surveyed period, the lyrics of which are exclusively Korean.

Taken as a whole, the foregoing provides legitimate grounds for unambiguously concluding that the prevalence of English in K-Pop lyrics has increased over the early twenty-first century.

### **Case study 1: change in prevalence of English for individual artists**

For this case study, we examine the change in the prevalence of English tokens in the works of two artists who appear in the collection over multiple years: *dongbangsingi* (‘Rising Gods of the East’, known internationally and hereafter as TVXQ!), and Super Junior. These artists were selected for the case study as they not only contributed a large number of albums to the collection (eleven for Super Junior, nine for TVXQ!), but these releases were spread over relatively long, comparable periods of time. For TVXQ! this covered the period from 2004 to 2015, while Super Junior’s contribution covered the slightly later period from 2007 to 2019. This contrasts with other artists who may have contributed more releases, for example EXO or BTS, both of whom contributed 14 releases, but over much shorter periods, i.e. 2013 to 2019 for EXO and 2015 to 2020

for BTS. While Super Junior and TVXQ!’s releases do not perfectly overlap or cover the entirety of the surveyed period, we nevertheless consider the number of releases and timeframe sufficient to be suggestive of changes in the strategies and methods involved in the production of lyrics for these artists, that is, their songwriting practice. Furthermore, by focusing on these two bands we isolate the timing of each release’s production as a potentially causal variable rather than artists’ personal characteristics such as gender and age. After all, the artists upon whom we focus are comparable in that they are boy bands formed in the early 2010s. A final point in common is the level and quality of these artists’ international success. While both have toured internationally and, significantly, oriented much of their international promotion towards East Asia (Yeo 2018: 10), they have not enjoyed the chart success of K-Pop artists such as BTS or BLACKPINK in majority English-speaking countries. For example, neither have appeared on the US Hot 100 Billboard Chart to date. This means that songwriting for these groups is unlikely to be informed by an attempt to capitalise on their pre-existing popularity in the English-speaking world.

The change in prevalence of English in TVXQ! and Super Junior’s albums is visualised in terms of mean tokens per song per year’s album releases included in the corpus over time, as shown in Figure 4.

Again, although we do not observe a linear, monotonic increase in the prevalence of English in either artist’s output, an upward trajectory is discernible. Another commonality is the temporal proximity of two dips in the mean number of English tokens for both groups’ 2011 output. TVXQ!’s 2011 album *Wae (Keep Your Head Down)*, for example, is notable for being the only release to have a lower mean number of English tokens per song than the group’s earliest release in this collection. While Super Junior’s 2011 and

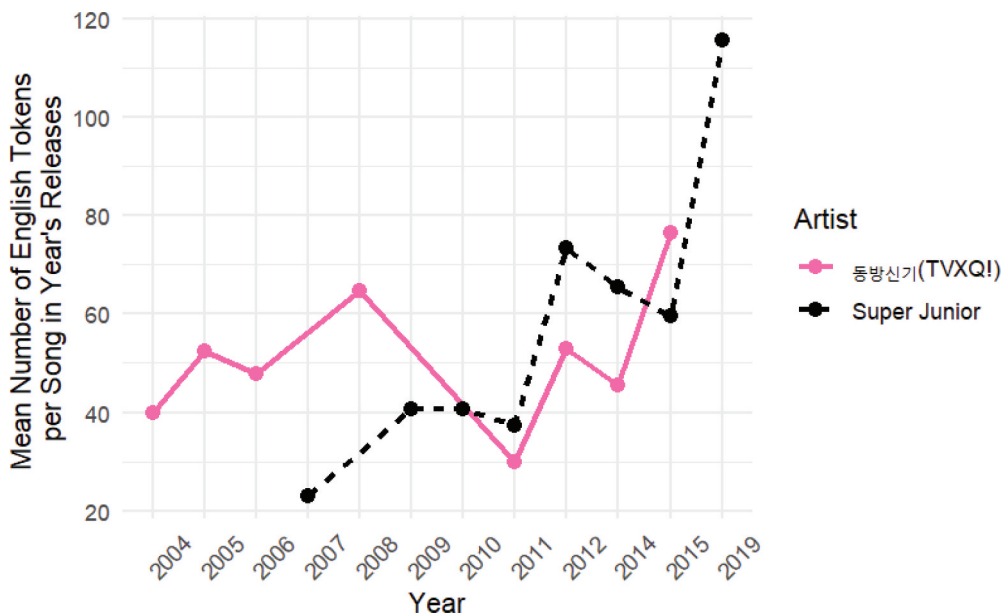


Figure 4. Mean number of tokens of English per song in year’s releases by two K-Pop artists.

2014 output shows comparable dips in the amount of English used, they are far from the low level of that group's earliest included release.

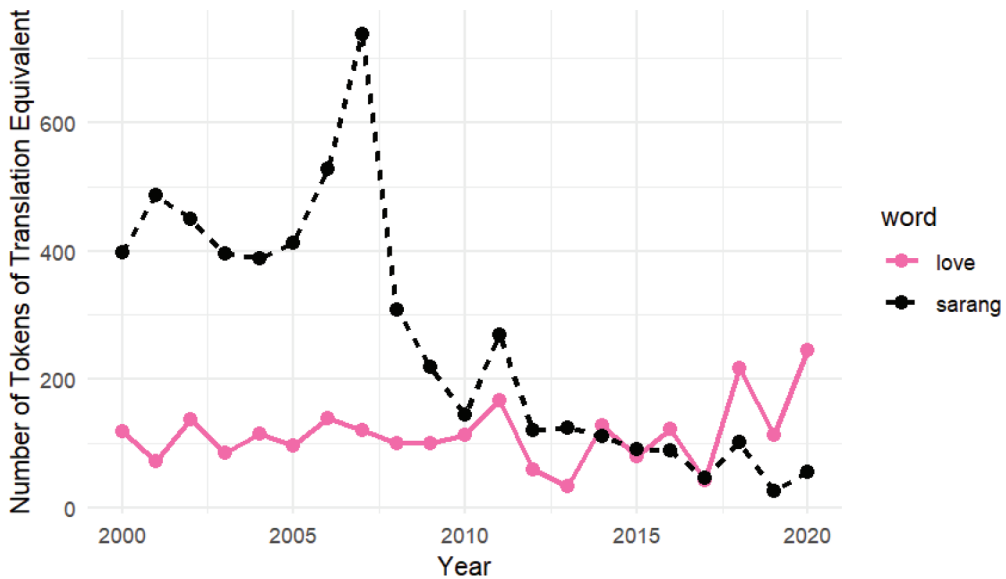
Here, we observe an increase in the mean number of tokens of English per song per album for these two groups. Taken as a whole, this suggests that the increase in the global, aggregated prevalence of English observed over the surveyed period is not attributable solely to artists who use more English becoming more popular than those which use less, thus replacing them in our collection. Rather we may speculate that even artists whose careers span the surveyed period are agentively or unconsciously increasing the amount and proportion of English they incorporate into their lyrics. The specific changes in songwriting practice which enable this may include working with more or different collaborators, or responding to a changed cultural context, but establishing the nature of and reasons behind this is, in each case, a matter for situated, granular, qualitative research. Nevertheless, the observed increase in the prevalence of English in this small subset of the collection may be interpreted as a sign of the impact that the influence of the expansion of K-Pop into markets and a listenership more familiar with English than Korean is having even on artists that enjoy relatively constrained success in such markets in comparison to international breakout artists.

### **Case study 2: change in prevalence of English for translation equivalents**

For this case study, we examine the relative frequencies of tokens of the keywords 'love' and 'sarang' over the surveyed period. In so doing, we aim to get some sense of whether the increase in the prevalence of English in K-Pop lyrics leads to the displacement of Korean for expressing concepts that are thematically central to popular music.<sup>5</sup> We note that this is in itself a speculative proposition. Nevertheless, the absolute frequencies of tokens for each surveyed year of 'love' and 'sarang' can be seen in [Figure 5](#).

There are a number of clear conclusions to be drawn from this visualisation. The number of tokens of 'sarang' has decreased substantially, but far from linearly, over the surveyed period. The number of tokens of 'love' has increased over the surveyed period. The proportionate decrease in tokens of 'sarang' is far greater than the proportionate increase in tokens of 'love'. There is no clear relationship between the number of tokens of each attested each year. For example, in 2007 when a particularly large number of tokens of 'sarang' appeared in our collection, there was neither a marked decrease nor small absolute number of tokens of 'love'.

The relationship between the increase in the number of tokens of 'love' and the decrease in the number of tokens of 'sarang' over the surveyed period can be illustrated and tested for significance using Pearson's correlation coefficient. This statistical test allows us to determine the strength and direction of the relationship between changes in two variables as well as the statistical significance of that relationship. The correlation coefficient takes values between  $-1$  and  $+1$ . Values of  $-1$  indicate that, as the value of one variable changes, its counterpart changes by the same amount in the opposite direction (e.g. as one value increases the other decreases). Values of  $+1$  indicate the opposite relationship between variables. Values close to  $0$  indicate no relationship. When a correlation coefficient reaches a threshold for significance, by convention  $p$ -values of less than  $0.05$ , the covariance observed is considered unlikely to have arisen by chance. As we start from the hypothesis that increases in the number of tokens of 'love' lead to



**Figure 5.** Number of tokens of two translation equivalents for top 10 albums in South Korean charts in 2000–2020.

a decrease in the number of tokens of ‘*sarang*’, we conducted a one-tailed test for negative correlation between the number of tokens of ‘love’ and ‘*sarang*’. This test, however, revealed no correlation,  $r(15) = 0.36$ ,  $p = 0.438$ . That is, the very weak positive correlation found ( $r(15) = 0.36$ ) could plausibly have occurred simply by chance ( $p = 0.438$ ).

One clearly discernible point, though, is the decrease in the combined number of tokens of these words. At this point, reasons underlying this are unknown. While fully addressing this falls outside the scope of this paper, it serves as an object demonstration of how such work can set the agenda for the qualitative, interpretative work needed to explain such observations as well as highlighting the limitations of our current approach. Our aggregated findings represent just a surface manifestation of complex, dynamic processes involving the adoption of new concepts, linguistic borrowing, changes in media consumption habits, and both global and local sociolinguistic and sociocultural shifts. We conjecture that promising lines of enquiry might include investigating whether consumer preferences changed to increase the representation of genres with thematic preoccupations other than romantic love over the surveyed period, e.g. more hip-hop less ballad, and qualitative content analysis of the lyrics to determine whether there was any industry-wide change in lyrical themes. In the following section, we contextualise the wider significance of our findings and further explore the potential avenues for exploration they suggest.

## Discussion

We have shown that the prevalence of English tokens in the collection of K-Pop song lyrics has increased over the course of the early twenty-first century. Our first case study suggests artists who appear in the charts increase the amount of English used in their

lyrics, rather than being displaced by new artists for whom English is more prevalent. Our second case study further suggests that the decreasing frequency of tokens of Korean keywords such as '*sarang*', that are thematically central to pop music, is not related to the increasing frequency of English translation equivalents. We further note that the timing of these developments has coincided with the immense growth in popularity of Korean-language popular culture including K-Pop throughout the world, including the wider Anglosphere.

That the increase in prevalence of English in K-Pop lyrics and increased global engagement with Korean pop culture have coincided raises the obvious question of cause and effect. In other words, is the increasing prevalence of English in K-Pop lyrics brought about by some attempt on the part of the songwriters, conscious or not, to appeal to audiences outside Korea? The results presented here are certainly suggestive of this engagement being an underlying factor in the increasing prevalence of English in K-Pop. Appealing to the mechanism approach to causation (Brady, 2011, pp. 1078–1083), we can see that the commercial incentive of making music that is more appealing and accessible to a large, global market would be a very parsimonious explanation for our findings. The retention of thematically central Korean keywords in lyrics is also understandable through this framework as these may be taken to index Korea, thereby serving as an attempt to appeal to the globalised market for authentic Korean cultural products.

As plausible as such an interpretation intuitively seems, we emphasise two caveats. First, across the field of causal inference (e.g. Imbens & Rubin, 2015, p. 3) correlation such as we observe here is not in and of itself considered sufficient to definitively draw conclusions about causation. Thus, the Korean Wave, understood as increased global engagement with contemporary South Korean popular culture, may not be the sole or even main causal factor in the observed increase in the prevalence of English. Separate, dedicated work is required to determine the relative importance of local and global factors in the increase in the prevalence of English in K-Pop lyrics across the industry and for individual artists.

Second, consider the counterfactual case. Our study does not allow us to determine whether other social or linguistic factors would have led to the observed increase in prevalence of English in the absence of the enormously increased global engagement with K-Pop associated with the Korean Wave. For example, Kiaer and Ahn (2023) observe that English constituted an ever-increasing proportion of the words and morphemes used to coin new words in South Korea in the early twenty-first century. The diverse range of fields from which these words were drawn beyond popular culture, particularly technology and business, attest that other domains of encounter contribute to the increasing prevalence of English in Korea. Furthermore, South Korea's 'English boom' and 'English fever', both dating from the 1990s (Park, 2009, p. 52), reveal education as a significant, early site of encounter with English for the vast majority of people growing up in South Korea.

In summary, the results presented here demonstrate the increase in the prevalence of English in K-Pop lyrics over the early twenty-first century. While by no means definitive, by robustly establishing temporal proximity they contribute a degree of support to the theory that this change in prevalence is, at least partly, a response to the Korean Wave. This relationship, and other possible connections, must remain speculative without insight into the actual behaviour and motivations of individuals. Uncovering this

requires situated, granular, qualitative research. This could, for example, take the form of exploring the sociohistorical context of South Korea's English education, along with societal trends, to enhance understanding of the role K-pop artists (e.g. songwriter interviews) and audiences (e.g. market analysis) play in shaping language interactions with song lyrics.

As well as these contributions, our results' characterisation of the distribution and changes in the prevalence of English in K-Pop raise numerous questions about the position of English in this context. For example, recall the large variation between years in the range of the mean number of tokens of English per song per album reported above. Why such large year-on-year variation should be observed alongside the overall increase in prevalence of English is puzzling. Given the steep slope reported for I. Schneider's regression models (I. Schneider, 2024), it seems likely that whether or not the individual artists that enjoy chart success in South Korea orient their music towards international audiences alongside a domestic one is a key factor for within year variation. Testing this hypothesis, or determining other underlying, causal factors for this observation, requires granular, qualitative analysis of, for example, songwriting practice and marketing strategies. A further example relates to the stylistic implications of English in temporal context. Some songs dating to the early 2000s have proportions of English lyrics more in line with songs from the late 2010s and some songs from later in the surveyed period have entirely Korean lyrics. Given the changing frequency of songs with larger proportions of English lyrics over the early twenty-first century, it would be interesting to investigate whether any social, generic or functional (especially poetic) characteristics are associated with songs that 'buck the trend' of their times. For example, Lee (2004) characterised the then contemporary use of English in K-Pop lyrics as 'prevalent', arguing that, among its functions, was acting as 'a simple attention-getter for stylistic purposes' (2004, p. 446). How or even whether the same functional, stylistic effect could be achieved using linguistic resources that have gone from merely prevalent to near ubiquitous remains an open question. Alongside the stylistic, our findings raise social questions at multiple scales. For example, on the global scale we may ask how transcultural flows through diverse scapes influence or respond to this shift, while at individual level we may ask how or whether the observed change has any impact on (para)social, interpersonal dynamics among fans and artists alike.

Our case studies, too, provide observations which may motivate further research. Examples include the above-mentioned question of the global decrease in tokens of both 'love' and '*sarang*' and whether the temporal coincidence of the decrease in the prevalence of English observed for TVXQ! and Super Junior in 2011 can be attributed to a chance similarity between the releases of these two groups or some industry-wide, external factor. The questions raised by both our overall findings and case studies demonstrate the additional insights provided by the quantitative, empirical approach taken in this paper as well as its limitations. It is notable that these questions cannot be investigated taking the methodological approach adopted here. They could not, however, have been legitimately generated without recourse to such an approach. Thus, as well as fulfilling its primary objective of providing a rigorous, empirical account of the change in the prevalence of English in K-Pop lyrics and providing further grounds for considering the Korean Wave a factor

underlying this change, this paper provides an object demonstration of the complementary nature of quantitative and qualitative approaches and the benefits of mixed methods research.

## Notes

1. For example, the 72% increase in higher education institutions offering some form of KFL between 2007 and 2024, from 82 to 141 (AATK, 2024; Silva, 2007, p. 11).
2. The transcriptions of 21 songs were age-restricted on the Korean website. These were not accessible to the authors due to the specifics of the age verification system in place. Instead, alternative user generated transcriptions were sourced from the site *genius.com*.
3. The Circle Chart, known as the Gaon chart from 2010 to 2022, represents South Korea's *de facto* official chart. As it does not cover the entire surveyed period, we rely on sales data from the RIAK as a proxy for chart data prior to 2010.
4. Neither English nor Korean underwent further pre-processing, such as stemming or lemmatisation. Thus, here tokens should be understood as potentially morphologically complex orthographic words, or *eojeol*. This Korean term covers, for example, combinations of nouns and particles as single tokens. While not necessarily part of stemming or lemmatisation, the form of pre-processing known as tokenisation, in the Korean context, often splits *eojeol* into their constituent morphemes. This would greatly inflate our count of Korean tokens. This would not impact our general findings, as a proportionate increase in English would still be reflected. Splitting Korean tokens in this way, though, would reduce the face-validity of our results as the count of Korean tokens could then include such morphemes as enclitic particles and verb endings, depending on the tokeniser used.
5. We consider the question of whether these words should be considered functionally distinct rather than translation equivalents in contemporary K-Pop an open one. Even if the concept of romantic love was not traditionally denoted by the word *sarang* as put forward in Lawrence (2010, p. 47), it has certainly come to be in contemporary South Korea.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Academy of Korean Studies [AKS-2021-OLU-2250004].

## ORCID

Simon Barnes-Sadler  <http://orcid.org/0000-0003-0591-316X>

Hyejeong Ahn  <http://orcid.org/0000-0003-2121-7658>

Jieun Kiaer  <http://orcid.org/0000-0003-0195-8213>

## References

- AATK. (2024, April). *Schools: Colleges/Universities*. Retrieved from <https://aatk.org/schools/>
- Ahn, H. (2021). English and K-Pop: The case of BTS. In E. L. Low & A. Pakir (Eds.), *English in east and South Asia: Policy, features and language in use* (pp. 214–225). Routledge.
- Ahn, H., & Kiaer, J. (2024). Translingual English words of Korean origin and beyond: Skinship, fighting, chimaek. *Asian Englishes*, 26(1), 69–83. doi:10.1080/13488678.2023.2216866

- Ahn, H., Kiaer, J., & Barnes-Sadler, S. (2025). BTS's and ARMYs' dynamic Translanguaging on social media. In J. Suh & E. S. Park (Eds.), *BTS and Languages: K-pop transcending language and communication* (pp. 101–126). Routledge.
- Ancombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17–21. doi:10.1080/00031305.1973.10478966
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. doi:10.21105/joss.00774
- Brady, H. E. (2011). Causation and explanation in social science. In R. Goodin (Ed.), *The Oxford handbook of political science* (pp. 1054–1107). Oxford University Press.
- Brezina, V. (2018). *Statistics in corpus Linguistics: A practical Guide*. Cambridge University Press.
- Desagulier, G. (2017). *Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics*. Springer.
- Fiordistella Iezzi, D., & Celardi, L. (2020). Text analytics: Past, present, and future. In D. F. Iezzi, D. Mayaffre, & M. Misuraca (Eds.), *Text analytics, studies in classification, data analysis, and knowledge organization* (pp. 3–15). Springer.
- Hardie, A. (2016). Corpus linguistics. In K. Allan (Ed.), *The Routledge Handbook of linguistics* (pp. 502–515). Routledge.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.
- Jin, D. Y., & Ryoo, W. (2012). Critical interpretation of hybrid KPop: The global-local paradigm of English mixing in lyrics. *Popular Music and Society*, 37(2), 113–131. doi:10.1080/03007766.2012.731721
- Kang, H. S. (2023). 'BTS, PSY, BLACKPINK' the English naming of South Korean music entertainers – its prevalence and motives. *English Today*, 155(39:4), 264–274. doi:10.1017/S0266078422000293
- Kiaer, J., & Ahn, H. (2023). *Emergence of Korean English: How Korea's dynamic English is Born*. Routledge.
- Lawrence, C. B. (2010). The verbal art of borrowing: Analysis of English borrowing in Korean Pop songs. *Asian Englishes*, 13(2), 42–63. doi:10.1080/13488678.2010.10801282
- Lee, J. S. (2004). Linguistic hybridization in K-Pop: Discourse of self-assertion and resistance. *World Englishes*, 23(3), 429–450. doi:10.1111/j.0883-2919.2004.00367.x
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh University Press.
- Moretti, F. (2000). Conjectures on world literature. *New Left Review*, 1(1), 54–68. <https://newleftreview.org/issues/ii1/articles/franco-moretti-conjectures-on-world-literature>
- Park, J.-K. (2009). 'English fever' in South Korea: Its history and symptoms. *English Today*, 97(25), 1. doi:10.1017/S026607840900008X
- Park, J. S.-Y. (2021). *In pursuit of English: Language and subjectivity in neoliberal South Korea*. Oxford University Press.
- R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schneider, G. (2024). *Text analytics for corpus linguistics and digital humanities: Simple R scripts and tools*. Bloomsbury Academic.
- Schneider, I. (2024). English's expanding linguistic foothold in K-pop lyrics. *English Today*, 40(2), 105–112. doi:10.1017/S0266078423000275
- Schweinberger, M. (2024, May 26). *Introduction to text analysis*. Brisbane: The Language Technology and Data Analysis Laboratory (LADAL). Retrieved from <https://ladal.edu.au/introta.html>
- Silva, D. J. (2007). Issues in Korean language teaching in the United States: Recent facts and figures. *The Korean Language in America*, 12, 106–125.
- Song, J. J. (2012). South Korea: Language policy and planning in the making. *Current Issues in Language Planning*, 13(1), 1–68. doi:10.1080/14664208.2012.650322
- Yeo, R. (2018). *The commodification of English in K-pop: Globalisation and multiple markets* (Master's thesis). Singapore: National University of Singapore.