A Statistical Analysis of English in Contemporary J-pop: Time Series of Lyrics and Identity, 2012-2021

David Calica Mariko Takahashi[1]
[1]Faculty of International Studies, Setsunan University
ddcalica@gmail.com mariko.takahashi@ilc.setsunan.ac.jp

Abstract

This paper analyzes the amount and variety of English within contemporary Japanese popular music (J-pop) over time and across a variety of artist identity factors. First, this study identified the quantity and variety of English tokens in the annual top 50 best-selling J-pop singles from 2012-2022. Over 70% of songs analyzed contained English lyrics, but songs on average had only 15% of total tokens in English. The average yearly amount of English increased by 6 percentage points from 2012 to 2022. Artist gender, K-pop affiliation, and group/solo status were all statistically significant with respect to amount of English in the songs.

1 Introduction and Background

Mixing English in the lyrics of songs in the local language is a phenomenon that has been widely observed in many countries (Moody, 2020). This use of English has been considered as an indication of bilingual creativity, and various functions have been identified. For example, Chan (2009) pointed out that mixing English in Cantonese popular songs achieved functions such as highlighting part of the lyrics and producing rhyming. Luk (2013) added that English in the lyrics was also used for language play through phonological and semantic crossover between English and Cantonese. Indeed, researchers have claimed that code ambiguation, or the construction of messages in two languages at the same time, is a prime example of creative use of English in popular songs of Asia including Korea (Lee, 2022) and Japan (Moody & Matsumoto, 2003; Moody, 2006). Another function of using English in popular songs has been to construct and project the intended identities of the singers (Benson, 2013; Kachru, 2006; Jin & Ryoo, 2014; Stanlaw, 2021).

Previous studies on this issue have primarily taken a qualitative approach to analysis. As such, Takahashi and Calica (2015) integrated a quantitative aspect and conducted an analysis of 100 CD singles that sold the most in Japan in 2013. The study selected gender, which had been one of the topics discussed in relation to Japanese popular music (e.g., Cogan & Cogan, 2006; Toth, 2008), as a potential factor contributing to the use of English. The analysis showed that more than 70% of the songs contained English in their lyrics and that male performers used English statistically at a higher proportion than female performers. In addition, younger male performers used more English in their lyrics compared to older male performers. The study then pointed out that English in Japanese popular songs (J-pop) fulfilled both referential and affective functions.

However, it remains to be seen if the use of English in J-pop has fluctuated over time. There is limited research tracing the change of the lyrics of popular songs from a quantitative perspective. One such example includes a study by Anglada-Tort, Krause, and North (2021), which utilized machine learning techniques to analyze the transition of gender distribution of popular songs and its relation to the content of the lyrics, but did so within a native-English UK context. As statistical research on English within musical contexts where English is not an official language is even more limited, this study aims to describe and analyze the use of English in top Japanese popular songs in the period of 2012 to 2021 by using a statistical approach.

2 Methodology

First, the 50 best-selling CD singles were identified for each year from 2012 to 2021 based on the Oricon

annual chart (https://www.oricon.co.jp/rank/). Second, for each of the 500 songs, lyrics were prepared in the text file format. Third, MeCab was used for dividing lyrics into tokens (http://taku910.github.io/mecab/). Each file was then checked manually to correct errors resulting from sentence fragments, colloquial expressions, and unique use of words. Fourth, AntConc (Anthony, 2020) was used for generating a word list and counting the number of tokens and types for each song. Finally, the number of English and Japanese tokens and types were calculated based on the word list.

Each song was tagged with its Oricon list year, what Oricon rank the song was listed at, whether the song was performed by a male or female lead singer, whether the song was performed by a group or solo artist, and whether the song was produced by a Korean company (K-pop). For purposes of this paper, Korean-based songs as published in Japan and listed on the Oricon top 50 are tagged as "K-pop".

For statistical analysis, each song's token, type, year, and artist factor data was input into Excel and R and analyzed for summary statistics. Finally, a multi-factored linear regression model was created using percent of each song's tokens as the dependent variable and artist factors and year as the independent variables.

3 Results and Analysis

We analyzed a total of 500 songs, spanning 10 years. 61% of songs were performed by a male-lead, 94% of songs were sung by a group of 2+ members, and 8% of songs were categorized as K-pop.

Fig. 1
Summary Stats – Song Attributes

Total Songs	500
Total Years (2012-2021)	10
Features	**% of Songs**
% Male Led	61.2%
% Group	94.4%
% K-pop	8.0%

Most of the songs contained some English, but overall English was low within each song when compared with native Japanese (see also Figure 10 in Appendix).

73% of the 500 songs contained English tokens. On average, 15% of tokens per song were English and 11% of types per song were English. Songs ranged from 0 to 100% English, with 27% of songs containing 0% English and 0.6% of songs containing 100% English.

Fig. 2
Summary Stats – English Per Song

% of Songs with English	72.6%
English as a % of Tokens per Song	
Average	15.4%
Median	4.6%
English as a % of Types per Song	
Average	10.6%
Median	3.6%

Fig. 3
% of Songs, by % of Tokens in English

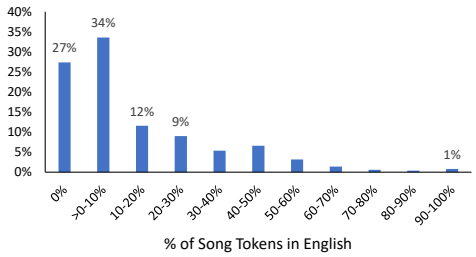


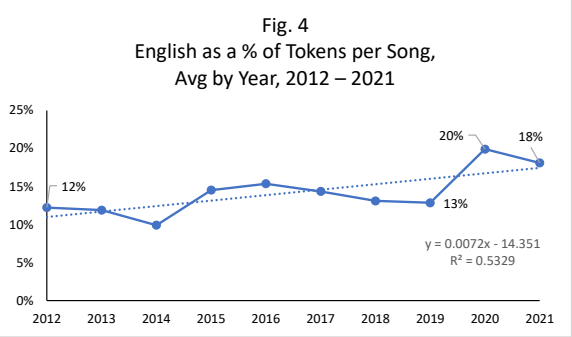
Figure 3 shows a distribution of the songs based on the percentage of song tokens in English. Nearly two-thirds (61%) of songs analyzed contain between 0% and 10% English, with the vast majority (94%) containing less than 50% English. Although English continues to play a significant role in conveying messages and musicality in contemporary J-pop, it overall remains a predominately minor language compared to native Japanese.

3.1 Trends over Time

Compared with Takahashi and Calica (2015) this paper tasked to answer how the use of English in J-pop has changed over time. To do so, we developed a time series calculating the average English usage year-over-year (see also Figure 11 in Appendix for % of types over time).

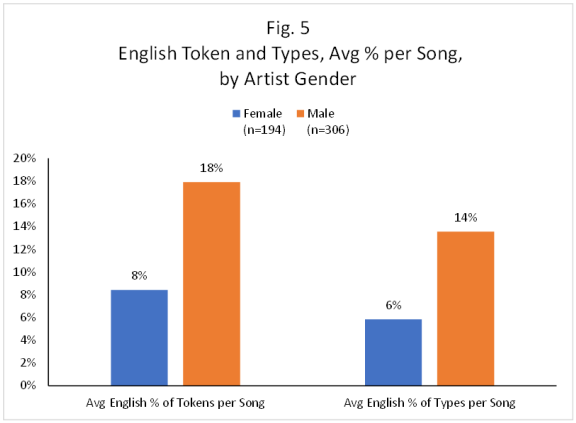
Over the past 10 years, the amount and variety of English has increased over time. Both tokens per song and types per song increased, as denoted by the significantly positive slope trendline. In 2012, songs on average had 12% of tokens as English, while in 2021, songs on average had 18% of tokens as English.

There is a +6-7 percentage points (pp) increase of English tokens and types between 2019 and 2020, which represents the largest year-over-year increase in the data. 2020 was a notable year for the music industry due to COVID-19 effects on global cultures and markets. In the next section, we note other localized effects such as gender and group factors in 2020 that also played a role in this English increase.



3.2 Gender as Related to English

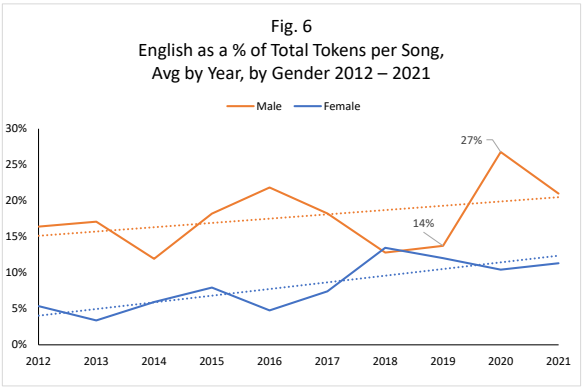
In Takahashi and Calica (2015), performer gender was found to be a significant factor in understanding the use of English in 2013 songs. This trend continues when looking across the 2012-21 timeframe.



Male performers used a higher percentage and higher variety of English in their songs, with the average male song having +10pp more English and +8pp more types. It was theorized in Takahashi and Calica (2015) that a prototypical J-pop male image is constructed through a variety of factors, potentially including heavier English usage. Given a higher statistical usage, it does appear that the modern J-pop male image remains more strongly tied with English usage.

The relationship between male gender and English is relatively consistent over time, as in Figure 6. Both genders have increased their English usage over time.

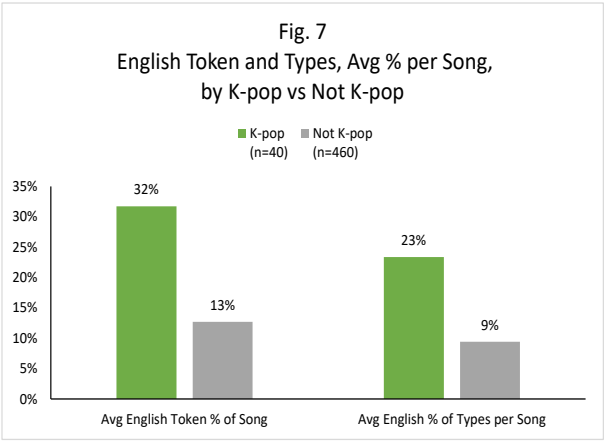
There is variation for the gender-English distribution, particularly in 2018-2020. 2018 is the first year observed where the female average usage was equal to or higher than the male average usage. This is primarily due to the increase in female K-pop groups, such as TWICE, appearing in the Oricon top 50 for the first time. As will be noted in the next section, Korean-based groups provide a significant increase of English usage in Japanese music. For example, if TWICE were excluded from the data, the average female Japanese English token percentage in 2018 would drop from 13.5% to 10.3%, which would be lower than the male average of 12.8%.



As noted in section 3.1, 2020 provided major changes to the J-pop landscape. In addition to global COVID-19 effects, many new male groups entered the top lists with increased English usage. For example, new male groups such as SixTONES, Snow Man, and JO1 debuted and reached the top 50 with 6 songs, with a combined average of 51% English tokens. These new songs replaced staple female groups such as the “48-groups” (AKB48, NMB48, etc.), which had a drop from the top 50 lists, going from an average of 11 songs in the top 50 between 2012-2019 to 8 in 2020.

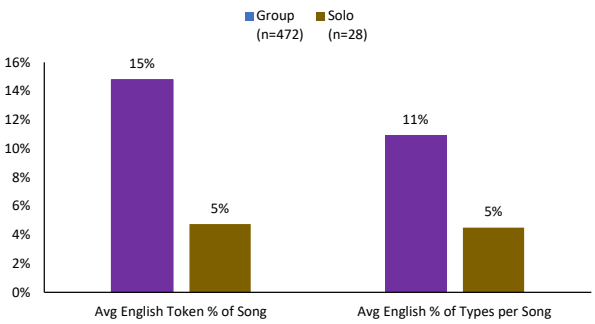
3.3 Korean and Groups Factors

K-pop has had a significant effect on English usage in Japanese music, with many Korean-based groups hitting the top 50 charts. As noted in section 3.2, TWICE for example helped increase the amount of English in Japanese music in 2018. K-pop groups, while they hold a minority of top 50 song slots, do provide a significant increase in English usage and variety.



The average Oricon listed Korean-based song has +19pp more English tokens and +14pp more types than a Japanese-based song. Similarly, being part of a group versus being an individual performer is linked to English usage. The average group has +10pp more English usage and +6pp more English types.

Fig. 8
English Token and Types, Avg % per Song, by Group vs Solo



3.4 Factor Significance

To statistically test each factor's significance with respect to English usage, we utilized a multi-factored linear regression model and tested whether each factor has statistically significant predictive effects. For this model, we regressed each song's English token percentage with respect to the song's year, performer gender, group/solo category, K-pop category, and 1-50 rank.

Fig. 9
Linear Model
English Token % = Year + Gender + Group + K-pop + Rank

Factor	Linear Estimate	P-val	Significantly Correlated with E-Tokens
Year	0.77%	2.8	Yes
Gender	10.2%	5.8	Yes
Group	11.0%	3.1	Yes
K-pop	17.8%	5.9	Yes
Rank	-0.08%	-1.3	No

"Yes" includes significance up to the 1% level

Year, gender (male), group, and K-pop flags all had a positive and statistically significant effect with respect to English usage by song. Rank was a factor which had no statistical effect, meaning that English is not correlated with how highly the song was ranked in the top 1-50 (See Appendix for details).

Based on the independent linear regression model, the significant linear estimates are greater than 0 and can be additive. For example, all else equal, a male K-pop group song will on average have +39pp more English than a female, non-K-pop solo artist song.

4. Discussion

English continues to carry significant purpose for lyrical construction within Japanese popular music, as it is statistically significant and remains the highest non-Japanese language to appear in top 50 Oricon for 2012-2021. Over time, English slowly has begun to increase in usage and variety, making it an even stronger portion of the Japanese music scene. Like many other fields, 2020 in particular had a pronounced effect on J-pop and English in J-pop, including the debut of high English-using groups.

English usage is also significantly linked to the musical performer identity, as it is associated with gender, group/solo distinction, and being Korean-based vs. non-Korean-based. However, this statistical identity based on lyrics is a constructed or presented image from the artist. Further research, such as audience testing, could help understand how such a language-based image is perceived from the audience's perspective.

5. Conclusion

Within this study, we statistically analyzed the use and variety of English in Japanese popular music lyrics. While English had broad usage across songs and was present in 73% of songs, it remained limited in terms of average amount (15% of tokens per song) and average variety (11% of types per song).

English has increased in usage from 2012 to 2021, with the average tokens per song increasing by 6 percentage points. English was also on average higher in songs from male artists, group performers (vs solo), and K-pop artists.

We look forward to further monitoring English in J-pop, especially as we encounter new structural changes. For example, new groups like Snow Man, which uses an average of over 50% English per song, have helped increase usage overall. We could potentially see English grow from a strong minority language presence to an even more pronounced and integral unit of Japanese musical communication.

References

Anglada-Tort, M., Krause, A. E., & North, A. C. (2021). Popular music lyrics and musicians' gender over time: A computational approach. *Psychology of Music*, *49*(3), 426-444. https://doi.org/10.1177/0305735619871602

Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Benson, P. (2013). English and identity in East Asian popular music. *Popular music*, *32*(1), 23-33. https://doi.org/10.1017/S0261143012000529

Chan, B. H. S. (2009). English in Hong Kong Cantopop: language choice, code-switching and genre. *World Englishes*, *28*(1), 107-129. https://doi.org/10.1111/j.1467-971X.2008.01572.x

Cogan, B., & Cogan, G. (2006). Gender and authenticity in Japanese popular music: 1980–2000. *Popular Music and Society*, *29*(1), 69-90. https://doi.org/10.1080/03007760500167313

Jin, D. Y., & Ryoo, W. (2014). Critical interpretation of hybrid K-pop: The global-local paradigm of English mixing in lyrics. *Popular Music and Society*, *37*(2), 113-131. https://doi.org/10.1080/03007766.2012.731721

Kachru, Y. (2006). Mixers lyricing in Hinglish: Blending and fusion in Indian pop culture. *World Englishes*, *25*(2), 223-233. https://doi.org/10.1111/j.0083-2919.2006.00461.x

Lee, S. (2022). Typology of English–Korean code ambiguation: Bilingual creativity in the Korean context. *English Today*. Advance online publication. https://doi.org/10.1017/S0266078422000116

Luk, J. (2013). Bilingual language play and local creativity in Hong Kong. *International Journal of Multilingualism*, *10*(3), 236-250. https://doi.org/10.1080/14790718.2013.808200

Moody, A. J. (2006). English in Japanese popular culture and J-Pop music. *World Englishes*, *25*(2), 209-222. https://doi.org/10.1111/j.0083-2919.2006.00460.x

Moody, A. (2020). *English in Asian popular culture*. In K. Bolgon, W. Botha, & A. Kirkpatrick (Eds.), *The handbook of Asian Englishes* (pp. 763-786). Wiley-Blackwell. https://doi.org/10.1002/9781118791882

Moody, A., & Matsumoto, Y. (2003). "Don't touch my moustache": Language blending and code ambiguation by two J-pop artists. *Asian Englishes*, *6*(1), 4-33. https://doi.org/10.1080/13488678.2003.10801106

Stanlaw, J. (2021). Using English as identity markers in Japanese popular music: J-pop and beyond. In E. L. Low & A. Pakir (Eds.), *English in East and South Asia* (pp. 198-211). Routledge. https://doi.org/10.4324/9780429433467

Takahashi, M., & Calica, D. (2015). The significance of English in Japanese popular music: English as a means of message, play, and character. *Proceedings of the 21st Annual Meeting of the Association for Natural Language Processing*, 868-871. https://www.anlp.jp/proceedings/annual_meeting/2015/index.html

Toth, C. (2008). J-Pop and performances of young female identity: Music, gender and urban space in Tokyo. *Young*, *16*(2), 111-129. https://doi.org/10.1177/110330880801600201

Appendix

Fig. 10
Summary Stats – Tokens Per Song

Avg English Tokens	49.8
Avg Japanese Tokens	253.0
Avg Total Tokens	302.4
Avg English Types	15.4
Avg Japanese Types	120.3
Avg Total Types	135.9

Fig. 11
English as a % of Types per Song,
Avg by Year, 2012 – 2021

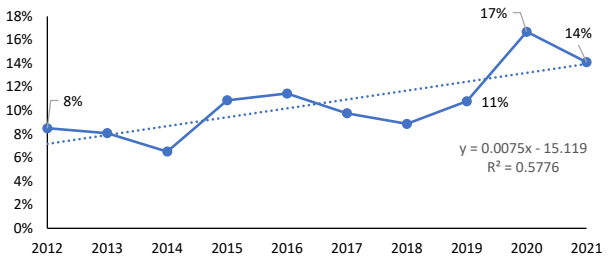


Fig. 12
Linear Model Details

```
Call:
lm(formula = e_token_pct ~ year + gender_int + group + kpop +
    rank, data = jpop2012-21)


Residuals:

     Min       1Q   Median       3Q      Max

-0.39295 -0.11348 -0.05414  0.06388  0.83123

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.558e+01  5.651e+00  -2.756  0.00606 **
year         7.715e-03  2.802e-03   2.753  0.00612 **
gender_int   1.020e-01  1.765e-02   5.778 1.34e-08 ***
group        1.104e-01  3.581e-02   3.084  0.00216 **
kpop         1.778e-01  3.014e-02   5.900 6.76e-09 ***
rank        -7.768e-04  6.090e-04  -1.275  0.20274
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1799 on 494 degrees of freedom
Multiple R-squared:  0.1519,    Adjusted R-squared:  0.1434
F-statistic:  17.7 on 5 and 494 DF,  p-value: 3.774e-16
```