

An analysis of lexis in English-medium songs popular in Japan

Brian R. Morrison

Abstract

Songs are a common resource for learning English. In self-study modules and learning skills classes for freshman and sophomore students delivered through the Self-Access Learning Centre at Kanda University of International Studies, students regularly identify songs as a resource for learning, particularly in relation to improving vocabulary for listening or speaking to general conversation. Nevertheless, no research has established the suitability of songs as a primary resource for vocabulary building. Therefore, when teachers and learning advisors are confronted with learners choosing songs as their main resource, they are unable to offer advice based on sound principles. This research aims to investigate the suitability of song use by analyzing the lexis in a corpus of 500 English-medium songs popular in Japan.

Context

Students at Kanda University of International Studies are encouraged to build on the content covered in their language classes by continuing to learn language outside the classroom. As well as specific homework tasks, students often have speaking and reading journals, which afford them more choice over the content they cover. In addition, there are a variety of self-directed learning modules (SDLM) offered through the self-access learning centre (SALC) with feedback and guidance provided by learning advisors (LA). The SDLM have at their core the development and implementation of individual learning plans (ILP) and a

philosophy Morrison (2012) refers to as The Autonomy Approach.

For students taking an SDLM who focus either on speaking or listening, songs and films are commonly selected as resources in their ILP. A typical comment taken from a recent SDLM is:

I think a song words are used [in] daily conversation so I can learn natural vocabulary through it.

The recent work of Webb and Rodgers (2009) provides a clear indication of the linguistic coverage of popular film by analyzing 318 film scripts from a wide variety of genres from British and American cinema. The analysis showed that knowledge of the first 3000 most frequent words plus proper nouns and marginal words gave over 95% coverage, while this rose to 98% coverage with an understanding of the 6000 most frequent words plus proper nouns and marginal words. This suggests that films can potentially be a resource for students of English; however, similar research has not been carried out on songs.

The benefits of using songs in the classroom have been written about in detail and can be best summed up by Murphey (1992a: 7):

Songs in general... use simple, conversational language, with a lot of repetition, which is just what many language teachers look for in sample texts. The fact that they are affective makes them many times more motivating than other texts.

There is no controversy over using songs as a classroom resource as long as they are selected and used in ways that complement the syllabus and fulfill students' needs. As well as the advantages of using songs in the classroom mentioned in Murphey's (ibid) quote, they provide examples of:

- Authentic oral communication
- Syntax patterns and grammar used in context
- Stress patterns often mirrored in the rhythm of the music
- Representations of target-culture identities

Thus teachers can select songs on the basis of the specific language and the way they are being used to communicate particular ideas. In other words, songs can be selected for a focus on language or a focus on topics and themes of interest.

However, for self-directed learning it is unlikely that students choose songs based on sound pedagogic principles and more likely that they will choose songs they enjoy but have not yet become lyrically familiar with. It is therefore important for educators to understand more about the lexical range and therefore appropriateness of popular songs for self-directed language learning purposes. By analyzing the lexical coverage of popular songs and comparing it to frequency lists of vocabulary, it is possible to draw conclusions regarding the efficacy of using song lyrics as input material. Results of this study can inform learning advisors and teachers about the relative usefulness of songs and whether they contain a rich enough variety of words to be considered as principle resources of lexis for oral communication purposes. In turn, learning advisors and teachers can inform learners about the likely lexical coverage of songs.

In order to consider the lexical coverage of songs, it is important to first consider high-frequency vocabulary and lexical coverage of texts. The Oxford English Dictionary contains 250 000 word families (Brown and Culligan, 2008). However, research has shown that only a very small proportion of these are used on a daily basis. For example, McCarthy and O'Dell (2001) report that only around 2000 different word families are spoken by native speakers in a day. At the same time Nation's (2001) research into texts found that the 2000 most frequent words in English account for between 81% and 85% of all words in general texts. However, Coxhead's (2000) research into spoken and written academic language found the first 2000 high-frequency vocabulary accounted for 80% of academic discourse across disciplines, slightly less than in general texts yet still a significantly high proportion.

The relevance of these percentages is determined by the percentage of vocabulary required to function effectively within discourse communities. Laufer (1992) postulated that 90% was the minimum requirement to deal with written texts while Hirsch and Nation (1992) put this much higher and suggested that without at least 95% coverage, students would be unable to read fluently.

When considering spoken English, the number of words required to reach these percentages appears to be lower than in written texts or spoken academic discourse. Schonell et al. (1956) had suggested that the 2000 most frequent word families contributed to 99% of all oral communication and there appeared to be a general consensus that these 2000 words were the target for fluent oral communication. However, with Adolphs and Schmitt's (2003) analysis of the

CANCODE spoken corpus, these numbers were revised down. Their findings discovered that the 2000 most frequent words accounted for less than 95% of spoken words and that for 96% coverage of spoken discourse, 3000 word families are required. While this discovery does not diminish the importance of high-frequency lexis, it does increase the number of words that English language learners need to know in order to communicate with ease when conversing.

Although there has been a lot of research conducted on corpora and high-frequency word families, there have been relatively few studies into the word knowledge of EFL and ESL students. Shillaw (1995) and Barrow et al. (1999) investigated the lexical knowledge of Japanese undergraduates who had received between 800 and 1200 hours of class-based English instruction. The results showed an average of 1700 and 2300 words respectively. However, there was no indication of which words were actually known and how these words were spread across frequency lists.

A recent investigation by Browne and Culligan (2008) sought to uncover more about Japanese undergraduates' vocabulary knowledge and where in the frequency range their vocabulary could be placed. Their findings were similar to Shillaw's (ibid) and Barrow et al.'s (ibid) in terms of vocabulary numbers but closer analysis found that out of 2430 words recognized by a typical student, there were significant gaps in high-frequency vocabulary knowledge, for example, from the first 2000 high-frequency words, 630 were unknown. This gap would undoubtedly be disabling when attempting to use English for any form of extended communication.

If this gap in high-frequency vocabulary is common across Japanese undergraduates, then clearly it needs to be addressed so that learners of English at university can use the language more effectively. While Webb and Rodger’s (2009) research into the lexical coverage of films identifies these as suitable resources, it is unlikely that freshman students will be able to cope with movies without considerable support. As a result films may not be appropriate for self-directed learning until the students have a greater range of lexis. However, if songs cover a large proportion of high-frequency vocabulary, their brevity could make them ideal as a linguistic resource for vocabulary building.

Table 1. Vocabulary profiles

Songs	1,000 Word Level	2,000 Word Level	Amalgamated percentage
<i>Just a Kiss</i> Lady Antebellum	89.9%	5.90%	95.8%
<i>Only Time</i> Enya	99.1%	0%	99.1%
<i>Top of the World</i> Carpenters	96.39%	1.03%	97.42%
<i>Stand by Me</i> Ben E. King	92.81%	1.2%	94.01%
<i>Your Song</i> Elton John	94.44%	3.7%	98.14%

(Adapted from Lieb 2011)

Lieb (2011) used 15 songs as part of an integrated skills course with an emphasis on vocabulary learning. Her analysis of the vocabulary profiled in each of the songs (see table 1) is very dependent on the first 2000 most frequent words. However, although we can see that over 90% of the words in songs come into this category, there is no indication of how many of the first 2000 most frequent words

are represented in songs. It is therefore the purpose of this study to investigate how many of the 2000 most frequent words are represented in a corpus of song lyrics.

Research question

Do songs provide the required lexical richness to be a primary source of input for learners with goals related to speaking or listening?

Methodology

The criteria for selecting songs for the corpus was heavily influenced by a small action research project I carried out in 2009 as part of a materials development task related to song worksheets. Questionnaires were distributed to 30 freshman and sophomore students who were either using the SALC or taking an SDML and had identified songs as a resource. The results influenced the development of the song worksheets (Cooke, 2010). They also made clear that these KUIS students had an eclectic approach to music. They were just as likely to listen to the Carpenters and the Beatles as they were to listen to Aerosmith and Backstreet Boys or more contemporary artists such as Avril Lavigne and Lady GaGa. It therefore seemed inappropriate to focus on commercial music downloads or other charts related to top 40 purchases if students were listening to music that spanned the last 5 decades. This presented a dilemma until considering that karaoke in Japan spans multiple decades of musical creativity and tends to have lyrics which can be heard and sung relatively easily. It was also probable that frequency lists of songs would provide a much wider range than a commercial chart, and perhaps a greater indication of the variety of music students in Japan are likely to be exposed to and to expose themselves to.

A corpus of the lyrics from 500 songs was compiled. The 500 songs were selected using frequency data of the 500 most popular English-medium songs in the year prior to 11th November 2011. This frequency data was provided by Daiichikosho Music Entertainment, the largest supplier of songs to the Japanese karaoke industry.

Prior to analysis, the corpus was proofread and edited so that irregular spelling, contractions, and other features of connected speech were modified to be consistent with the BNC/COCA corpus. Transcriptions of phonological features common in this medium such as *la la la* and *oh oh oh* were omitted to prevent this frequency. This gave a corpus of approximately 150 000 tokens i.e. an average of 300 words per song.

Analysis

The RANGE and FREQUENCY software packages (Nation and Heatley 2002) and Cobb's (2012) Compleat Lexical Tutor webware were used to analyse the corpus. These programs allowed analyses to be done that compared the corpus against the BNC/COCA. The BNC/COCA is a corpus of 10 000 000 running words with a 50-50 split between English from US and UK/NZ sources and is comprised of 60% spoken and 40% written English (Nation, 2012).

Results and discussion

A quick analysis of the top 20 songs from the popular karaoke songs (see table 2) showed a striking resemblance to the artists mentioned regularly by students in the short survey.

Table 2. The top 20 English-medium karaoke songs

1 - 10	11 - 20
1. Poker Face: Lady GaGa	11. Judas: Lady GaGa
2. My Heart Will Go On: Celine Dion	12. I Was Born To Love You: Queen
3. Bad Romance: Lady GaGa	13. Let It Be: The Beatles
4. A Whole New World: Peabo Bryson and Regina Belle	14. Yesterday Once More: Carpenters
5. I Don't Want To Miss A Thing: Aerosmith	15. Born This Way: Lady GaGa
6. Happy Birthday To You: Traditional	16. I Will Always Love You: Whitney Houston
7. I Want It That Way: Backstreet Boys	17. Telephone (feat. Beyonce): Lady GaGa
8. Top Of The World: Carpenters	18. It's My Life: Bon Jovi
9. Paparazzi: Lady GaGa	19. We Are The World: USA For Africa
10. Baby feat. Ludacris: Justin Bieber	20. Dancing Queen: ABBA

The overlap between artists students referred to and the whole list indicated that this corpus was indicative of the types of songs KUIS students were interested in learning English from.

Table 3. Most frequent vocabulary

1 - 10	11 - 20
I	NOT
YOU	AM
THE	IN
TO	DO
AND	THAT
IS	LOVE
IT	ARE
ME	ON
A	WILL
MY	YOUR

A frequency analysis of the vocabulary found that 25% of the text was made up of only 9 words, a finding similar to Murphey's (1992b) analysis of the top 50 songs in English from Music and Media's Hot 100 Chart in September 1987. In that analysis, Murphey found that 25% of the corpus was made up of 10 words and included gonna and love (replacing *it* and *to* in this corpus). The 20 most frequent words made up almost 38% of the text in this corpus and content words only start to appear in the second ten most frequent word types.

If prepositions, pronouns, articles and auxiliary verbs are removed, the list gives a clearer idea of the frequency of content lexis (see table 4).

Table 4. Most frequent content vocabulary

Word Type	Rank	Freq.	Word Type	Rank	Freq.
LOVE	16	1486	COME	53	512
KNOW	26	1014	GO	54	505
WANT	31	816	SAY	55	494
BABY	34	745	WAY	56	488
LIKE	35	717	MAKE	57	463
GOT	36	711	HEART	60	445
GOING	45	588	SEE	61	437
LET	46	579	NEVER	64	423
TIME	49	562	(YEAH)	(65)	(414)
GET	52	519	TAKE	66	409

All the words that can be identified as verbs in this list only contain one syllable. One-syllable verbs in English are indicative of vocabulary that has derived from Anglo-Saxon (Ellis, 2012), which points towards spoken language rather than the multi-syllable Latinate verbs more common in written texts. Therefore, there is a clear indication that although songs are composed and are not spontaneous speech, they do contain a high frequency of words common in spoken English.

An analysis of the range of vocabulary to identify the coverage of high-frequency vocabulary by songs, however, shows that there are gaps in the corpus (see Table 5)

Table 5: Range of lexis

WORD LIST	TOKENS	TYPES	FAMILIES	MISSING FAMILIES (CUMMULATIVE)
1000	133391 (89.40%)	1802 (34.61%)	845	155 (155)
2000	7378 (4.94%)	1007 (19.34%)	631	369 (524)
3000	537 (0.36%)	202 (3.88%)	152	848 (1372)
offlist	7896 (5.29%)	2195 (42.16%)	?	N/A
Total	149202	5206	1628	N/A

The gaps in the lists even in the first 1000 most frequent words clearly point to songs being a rather impoverished resource to learn lexis, and this is confirmed by considering that only 1476 of the 2000 most frequent words from the BNC/COCA were present in this song corpus. Given the emphasis researchers put on the 2000 most high-frequency words for effective participation in conversation, the missing 26% is likely to have a severe impact on the ability of students to take part in oral communication if they focus on songs for their learning. Therefore self-directed learners of English who choose songs as their principle vocabulary learning materials should be encouraged to expand their resources to beyond lyrics to encompass materials that cover the range of vocabulary that is needed for effective communication.

It is of interest to note the unexpectedly large percentage of offlist types. There are obviously some words, perhaps neologisms, that prevent the software from calculating the number of families present from the offlist types. An investigation into the number of word families and the frequency of the offlist words is beyond the scope of this research but could form the basis of further research.

Conclusion

Learners of English choose songs to learn English from because they are fun, memorable and motivating, and perhaps because so many learners carry music around with them in their phones and other digital devices. While clearly the medium of song can help students to learn about the language, the topics, and something of the identities portrayed by singers, this research strongly suggests that songs are not lexically rich enough and must be supplemented with other materials. In other words, in spite of the many sound pedagogic reasons for using songs for learning, this investigation into this corpus leads to the conclusion that more resources are required in order to cover the range of vocabulary that is needed for effective communication.

ACKNOWLEDGMENTS

Thank you to the Research Institute of Language Studies and Language Education for providing the funding for this research and to Daiichikosho Music Entertainment (DAM) who kindly supplied the data of the 500 most popular English-medium karaoke songs, which allowed the corpus to be built.

References

- Adolphs, S., and Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24, 425–438.
- Barrow, J., Nakanishi, Y., and Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, 27, 223–247.
- Browne, C. and Culligan, B. (2008). Combining technology and IRT testing to build student knowledge of high frequency vocabulary, *The JALT CALL Journal*, 4(2), 3–16.
- Cobb, T. (2012). Compleat Lexical Tutor [software]. Online from <http://www.lex tutor.ca/>
- Cooke, S. (2010). Developing song worksheets for a SALC. *Studies in Self-Access Learning Journal*, 1(2), 129-138.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- Ellis, R. (2012). *Consciousness-raising tasks for grammar teaching*. Paper read at PAC@FEELTA 2012, Learning languages – building communities.
- Hirsh, D., and Nation, I.S.P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689–696.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Béjoint and P. Arnaud (Eds.), *Vocabulary and applied linguistics*, 126–132. London: Macmillan.
- Lieb, M. (2011). English Through Popular Songs: Reinforcing High-frequency Vocabulary and Enhancing All Four Skills. In B. Jones and R. Silver (Eds) *Teachers Helping Teachers. JALT. 19-32*.
- McCarthy, M. J., and O'Dell, F., (2001). *Basic vocabulary in use. American English*. New York: Cambridge University Press.
- Morrison, B.R. (2012). Learning lessons: Implementing the Autonomy Approach. *IATEFL 2011 Conference Selections*, 73-75. Canterbury: IATEFL

- Murphey, T. (1992a) *Music and Song*. Oxford: Oxford University Press.
- Murphey, T. (1992b). The discourse of pop songs. *TESOL Quarterly*, 26(4), 770-774.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P., and A. Heatley (2002). Range: A program for the analysis of vocabulary in texts [software]. Downloadable from <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>
- Nation, I.S.P. (2012). Information on the BNC/COCA word family lists. Available from: [http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC COCA-word-family-lists.pdf](http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC-COCA-word-family-lists.pdf) [accessed 1st November 2012]
- Schonell, F.J., Meddleton , I.G., and Shaw, B.A. (1956). *A study of the oral vocabulary of adults*. Brisbane, Australia: University of Queensland Press.
- Shillaw, J. (1995). Using a word list as a focus for vocabulary learning. *The Language Teacher*, 19 (2), 58-59.
- Webb, S. and Rodgers, P.P.H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30 (3), 407-427.