

가사로 분석하는 사회

- 시대별 가사 변화 양상

류재원 주어진 김민정 유건욱 이용하 임지우 정성균

INDEX



01 주제 발표

02 분석 방법

03 기본 분석

04 웹 소개

05 결론

06 Q&A

주제 발표

주제 → 주제 선정 계기 → 데이터

“

시대별 유행했던 노래의 가사를 당시 시대상과 연관지어 분석하고 시각화

”

주제 선정 계기

- 음악이나 시 등 글로 쓰여진 예술은 그 당시의 사회상을 그 시대 이후에도 비교적 실감나게 느낄 수 있게 하는 매체
- 인기가 높았던 노래 가사는 모든 세대가 공유했던 만큼 사회 모습을 잘 드러냄



약 50~55년간 국내/해외 히트곡 가사를
시대별/성별 등의 기준으로 분석하고
시각화하여 당시의 시대상을 알아봄

데이터(Data)

국내

- 55년간(1964-2018) 빅스차트 연간 1-100위



해외

- 51년간(1965-2015) 빌보드차트 연간 1-100위



제목, 가수, 가수 형태(그룹/듀오/솔로), 성별(가수), 가사

분석 방법

KoNLPy → TF-IDF → H-P-point

KoNLPy

: Korean NLP in Python

- Python으로 한국어 정보 처리를 할 수 있는 package
- 해당 package에서 제공되는 형태소 분석기를 이용하여 품사에 따른 분류 및 분석 가능



KoNLPy

KoNLPy에서 제공하는 한글 형태소 분석기 비교

Okt

- 속도 : 가장 빠름
- 분석 품질 : 상대적으로 낮음

실시간으로 처리할 필요가 있을 때
사용하는 것이 유용

Kkma

- 속도 : 느림
- 분석 품질 : 우수

고유명사 추출할 때 사용함

Komorán

- 속도 : 빠름
- 분석 품질 : 보통

Hannanum

- 속도 : 빠름
- 분석 품질 : 보통

TF-IDF

: Term Frequency – Inverse Document Frequency
(어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치)

$$W_{x,y} = \text{tf}_{x,y} \times \log \left(\frac{N}{\text{df}_x} \right)$$

Term(x) Frequency within document(y)

TF : 특정한 단어가 특정 문서 내에서 얼마나 자주 등장하는지를 나타내는 수치

DF : 특정 문서군 내에서 특정한 단어가 등장하는 문서의 수를 나타내는 수치

IDF : DF의 역수. 값이 높을 수록 해당 문서군 내에서 그 단어가 흔하게 쓰이지 않음을 나타냄
(등장하는 문서 내에서 단어가 특별한 의미를 가지게 됨)

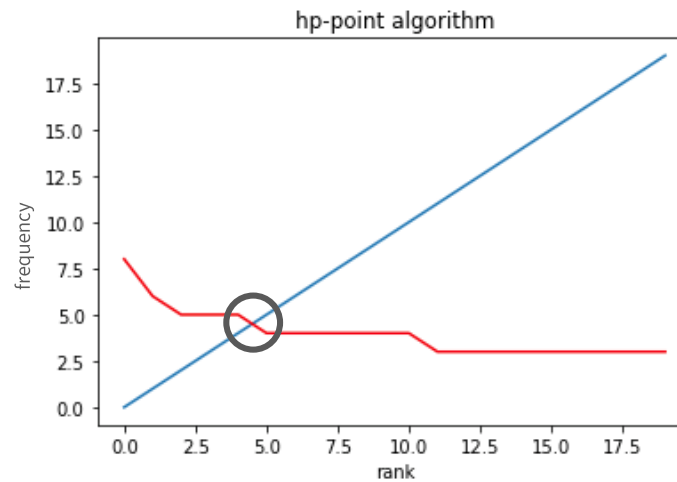


TF와 IDF를 곱하여 TF-IDF라고 하며,

이 값이 높을수록 문서군 전체에서 흔하지 않으면서 해당 문서 내에서는 중요한 단어들(키워드)을 뽑아낼 수 있음

H-P-point

: Hirsch-Popescu-point (Lexical richness를 나타내는 척도)



단어 빈도수 순위와 빈도수가 일치하는 지점을 나타냄

예를 들어 10번째로 빈도수가 높은 단어가 특정 문서에 출현한 횟수가 10번이면 H-P-point가 10
이걸 기준으로 10보다 크면 반복적, 작으면 덜 반복적인 텍스트로 판단

H-P-point

	year	title	singer	lyrics	hpoint	hp-point rule
0	2018	사랑을 했다 (LOVE SCENARIO)	IKON	사랑을 했다 우리가 만나 지우지 못할 추억이 됐다 불만한 멜로드라마 관찰은 걸말 그...	7	됐다
1	2018	그날처럼	장덕철	참 많은 시간이 흘러가고 넌 어떻게 사는지 참 궁금해 날 걱정하는 사람들에게 다 잊...	5	그날
2	2018	선물	멜로망스 (MeloMance)	빛이 들어오면 자연스레 뜨던 눈 그렇게 너의 눈빛을 보곤 사랑에 눈을 떴어 항상...	5	항상
3	2018	모든 날, 모든 순간 (Every day, Every Moment)	폴킴(Paul Kim)	네가 없이 웃을 수 있을까 생각만 해도 눈물이나 힘든 시간 날 지켜 준 사람 이제는 ...	5	의
4	2018	뚜두뚜두 (DDU-DU DDU-DU)	BLACKPINK	착한 얼굴에 그렇지 못한 태도 가녀린 몸매 속 가려진 Volume은 두 배로 거침없...	6	do
5	2018	뽕뽕	모모랜드 (MOMOLAND)	설레어 너와 나의 랑데뷰 (랑데뷰) 내 마음을 들었다 놔다 해 맘대루 (맘대루) 지...	6	네
6	2018	You	멜로망스 (MeloMance)	어떻게 난 해야 해 니가 떠난다고 말하면 난 혼자 남겨지는 게 더 두 려울 텐데 이유...	6	있는
7	2018	그때 헤어지면 돼	로이킴	나를 사랑하는 법은 어렵지 않아요 지금 모습 그대로 나를 꼭 안아 주세요 우리 나중에...	4	우리
8	2018	Roller Coaster	청하	I remember very first time in love I remember ...	7	need



노래 '사랑을 했다(LOVE SCENARIO)'에 '됐다'라는 단어가 7번 출현하였으며, 그 출현 빈도의 순위가 7번째로 높은 것

기본 분석

키워드 분석 → 가수 형태 비중 변화 → 성별 비중 변화 → 장소 단어 → 단어의 반복성

국내

1973
[토요일, '주오', '조잘', '아아아아아아', '잘가세요', '아니올까', '잘있어요', '소낙비', '잊게', '잊겠어요', '아아아', '당신만', '그건', '이들을', '했었네']
1972
[딩동댕, '못살아', '세실리아', '뜨는', '달은', '사모해', '아리랑', '뜬다', '임마', '단골', '동동', '별은', '짓을', '스타트라폼바', '세노야']
1971
[거짓말이야, '잘했군', '랄랄', '썸머와인', '랄랄랄랄', '괴롭힐', '잘했어', '어이하라고', '루루루루', '마오', '라랄라', '루루루루루', '루루루', '나는', '안녕']
1970
[네로, '두글자는', '가시내야', '난올라', '검은고양이', '범띠', '봄비', '축제의', '싫어지면', '마음', '달콤하고', '사나이', '울었다네', '랄랄랄랄랄랄랄', '길고도']
1969
[김상사, '숨길', '사랑', '마리아', '아느냐', '그런집을', '없어요', '찾아라', '서울에서', '한번만이라도', '꽃사시오', '꿀뚜기짓', '돌아', '에라', '지어요']
1968
[까투리, '나간다', '했더라요', '크로바', '네잎', '까투리사냥을', '총각님', '어서가자', '서울이며', '보았다', '별야', '아아아아', '어찌하오', '가지를', '이라']
1967
[못살아, '어허', '사모해', '여자지', '사랑', '김일병', '빛과', '왔네', '브라보', '울음', '팔도강산', '몽땅', '당신을', '좋아해', '그림자']
1966
[살짜기, '읍서예', '에헤야', '미스터', '수덕사에', '고향', '홀러가듯', '추풍령', '천리길', '남남으로', '가지를', '지울까', '모래', '사랑했기에', '키다리']
1965
[트위스트, '프레이보이', '울렁대는', '이름은', '자장', '원더풀', '울산', '삼돌이가', '부라보', '배나', '서울', '울렁울렁', '동동', '소녀', '애기']
1964
[리라꽃, '아가씨', '구두', '베사메우초', '베사메', '사나이', '못하네', '하오', '울어야만', '용두산아', '베사메우초야', '똑똑똑', '흐른다', '영등포의', '부르리까']

2018
[you, 'love', 'oh', 'likey', 'it', 'know', 'ba', 'really', 'me', '내가', 'bum', 'my', 'banana', '너를', 'yeah']
2017
[knock, 'you', 'me', 'rookie', 'na', 'it', 'we', 'oh', 'yeah', 'love', '내가', 'baby', 'really', 'my', 'blue']
2016
[you, 'me', 'ooh', 'na', 'baby', 'pick', 'like', '내가', 'up', 'good', 'luck', 'yeah', '너의', 'my', 'love']
2015
[you, 'me', 'bad', 'baby', 'love', 'it', 'oh', 'booty', 'my', 'girl', 'dumb', 'that', '아예', 'yeah', 'shake']
2014
[you, 'eh', 'holler', 'love', 'me', 'my', 'it', 'baby', 'body', 'oh', '너를', '내가', 'back', 'lonely', 'hellow']
2013
[oh, 'na', 'you', 'me', 'love', 'ooh', 'yeah', 'du', 'to', 'boy', 'bad', '호옹', 'it', '으르렁', '니가']
2012
[you, 'na', 'uh', 'illa', 'oh', 'cream', 'baby', 'my', 'love', 'me', 'ring', 'up', 'sexy', 'heaven', '내가']
2011
[oh, 'it', 'put', '거울아', 'baby', 'hot', 'you', 'shake', 'come', 'boy', 'up', 'yeah', 'eh', 'high', 'lonely']
2010
[oh, 'you', 'no', 'uh', 'it', 'me', 'love', 'huh', 'my', 'girl', 'change', 'baby', 'so', 'wake', '니가']

- 최근으로 올수록 영어단어가 키워드로 많이 잡히는 경향을 보이며 아이돌 노래 가사에 나올 만한 단어들이 많음
- 예전 노래들은 거의 한국어 단어를 사용하며 단어의 스펙트럼이 넓은 편. 생소한 단어들도 자주 등장

해외

1965

['rhonda', 'help', 'her', 'cara', 'pussycat', 'ill', 'wooly', 'down', 'sloopy', 'mia', 'well', 'he', 'doggone', 'bully', 'about']

1966

['barefootin', 'submarine', 'coal', 'down', 'good', 'writer', 'bop', 'hazel', 'girl', 'monday', 'excitations', 'hooray', 'paperback', 'ill', 'her']

1967

['lifting', 'la', 'girl', 'little', 'georgy', 'hey', 'too', 'down', 'bit', 'miles', 'her', 'world', 'fa', 'as', 'carousel']

1968

['gurdy', 'mony', 'surry', 'hurdy', 'tighten', 'din', 'na', 'simon', 'telegram', 'ill', 'think', 'her', 'hey', 'dig', 'judge']

1969

['he', 'dig', 'la', 'loneliest', 'girl', 'back', 'down', 'day', 'little', 'world', 'mine', 'grazin', 'where', 'night', 'her']

1970

['julie', 'chh', 'chhchh', 'dat', 'gimme', 'llola', 'down', 'buh', 'well', 'rainin', 'lola', 'ready', 'fadin', 'ya', 'little']

1971

['her', 'la', 'da', 'chirpy', 'cheep', 'bon', 'sweet', 'yoyo', 'home', 'back', 'he', 'bring', 'toke', 'wheres', 'em']

1972

['awimoweh', 'coconut', 'day', 'marie', 'nana', 'lime', 'ill', 'bang', 'levee', 'good', 'well', 'down', 'said', 'call', 'hunk']

1973

['willy', 'he', 'louie', 'his', 'mash', 'wo', 'her', 'down', 'man', 'ill', 'yes', 'world', 'stir', 'darn', 'well']

1974

['down', 'boogie', 'beach', 'eres', 'waterloo', 'mighty', 'her', 'boogity', 'rock', 'said', 'hey', 'locomotion', 'thing', 'keep', 'wolfman']

1975

['down', 'bom', 'good', 'lucy', 'blitz', 'ballroom', 'satisfied', 'swearin', 'he', 'only', 'life', 'why', 'girl', 'been', 'fame']

2006

['laffy', 'taffy', 'lean', 'rompe', 'ya', 'girl', 'wanna', 'down', 'back', 'right', 'yo', 'dat', 'keep', 'ha', 'booty']

2007

['bay', 'her', 'da', 'wanna', 'aye', 'girl', 'eh', 'ay', 'walk', 'ba', 'right', 'rock', 'shawty', 'why', 'around']

2008

['lollo', 'wee', 'her', 'ay', 'bum', 'wanna', 'damaged', 'girl', 'ya', 'life', 'keep', 'from', 'shawty', 'bedum', 'could']

2009

['na', 'down', 'bum', 'girl', 'hey', 'boom', 'wanna', 'right', 'halo', 'muh', 'her', 'diva', 'ima', 'bedum', 'could']

2010

['imma', 'down', 'right', 'aah', 'wanna', 'back', 'bottoms', 'blah', 'eh', 'girl', 'chick', 'dougie', 'her', 'only', 'na']

2011

['la', 'wanna', 'girl', 'boom', 'tonight', 'back', 'bottoms', 'right', 'put', 'down', 'life', 'had', 'na', 'her', 'night']

2012

['rack', 'girl', 'down', 'gangnam', 'bitch', 'good', 'ill', 'turn', 'op', 'ooh', 'lets', 'back', 'woahohohoh', 'had', 'wanna']

2013

['wanna', 'night', 'down', 'uoeno', 'gangnam', 'ill', 'girl', 'op', 'our', 'ooh', 'from', 'here', 'heart', 'back', 'nigga']

2014

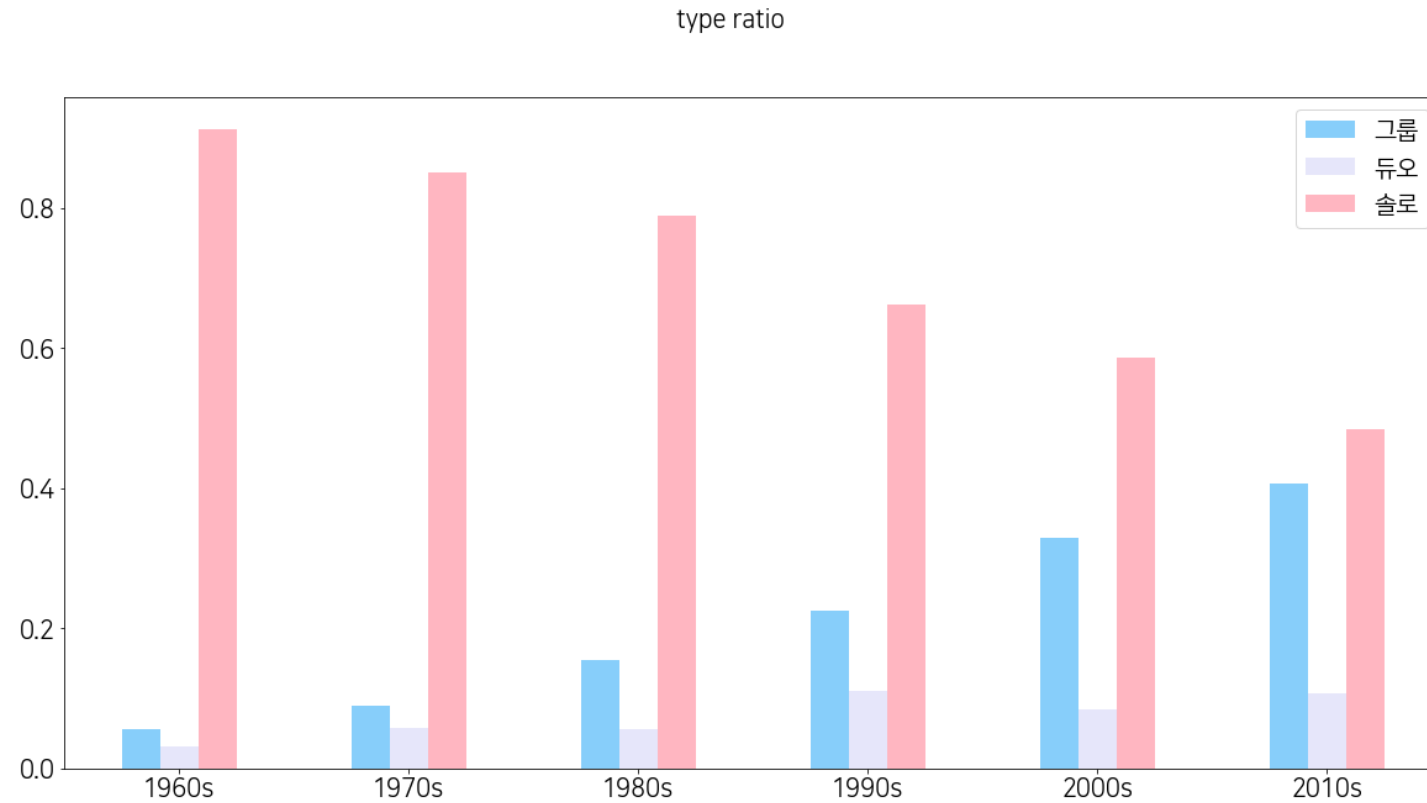
['hitta', 'down', 'mediocre', 'na', 'la', 'her', 'girl', 'wanna', 'nigga', 'hittas', 'em', 'ill', 'been', 'right', 'night']

2015

['wanna', 'down', 'coco', 'fuck', 'her', 'bitch', 'quan', 'watch', 'only', 'hey', 'girl', 'money', 'ill', 'nae', 'back']

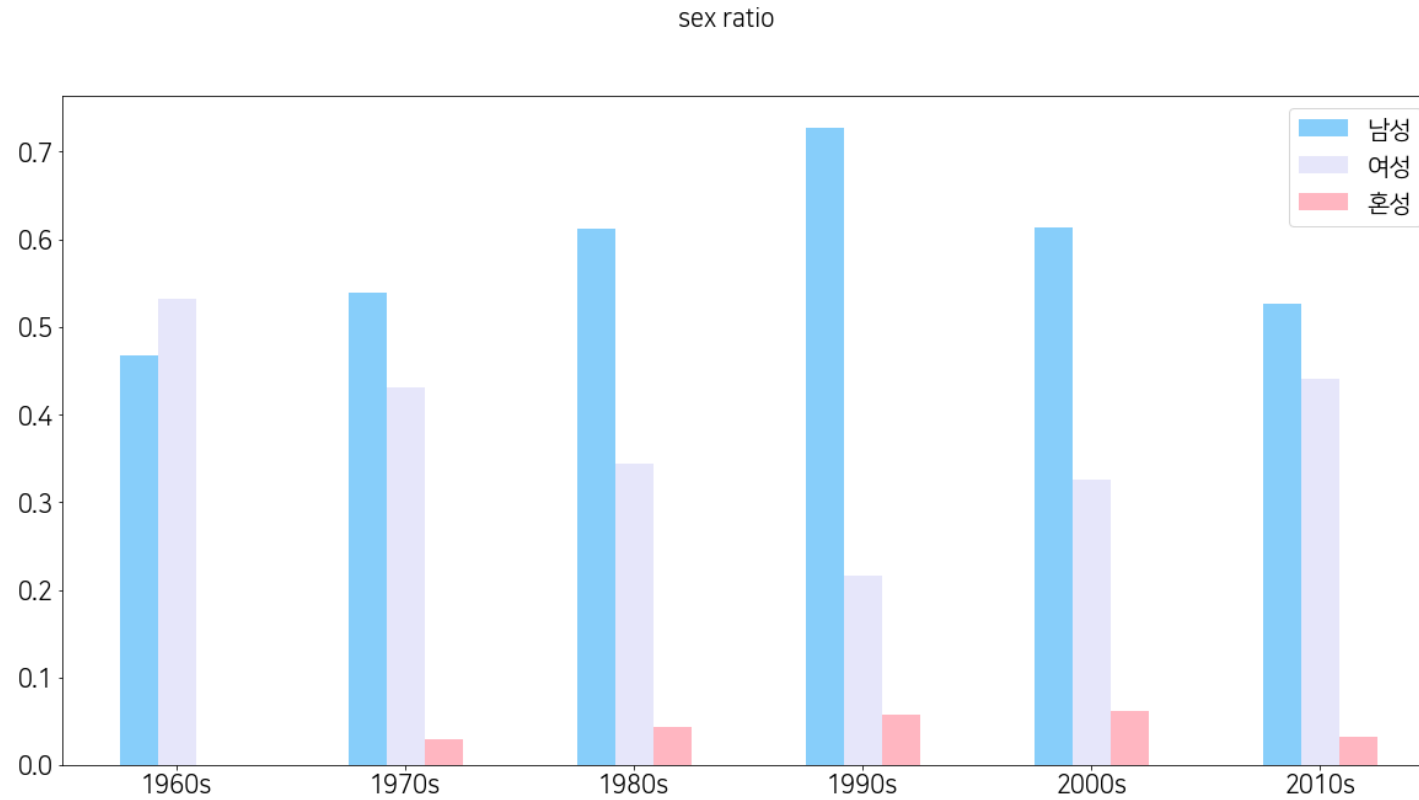
- 예전 노래에서 조금 더 생소한 단어들이 많이 등장
- 전체적으로 크게 달라지지 않은 양상을 보임

가수 형태 비중 변화



- 전체적으로 솔로 가수의 비중은 감소하고, 그룹의 비중은 증가
- 그룹 비중의 증가는 아이돌의 등장과 유행의 영향을 받은 것으로 보임(전문적인 엔터테인먼트의 등장)
- 듀오의 비중이 항상 가장 작는데, 이는 그룹 내의 멤버 수가 점점 증가하는 것의 영향을 받은 것으로 보임

성별 비중 변화



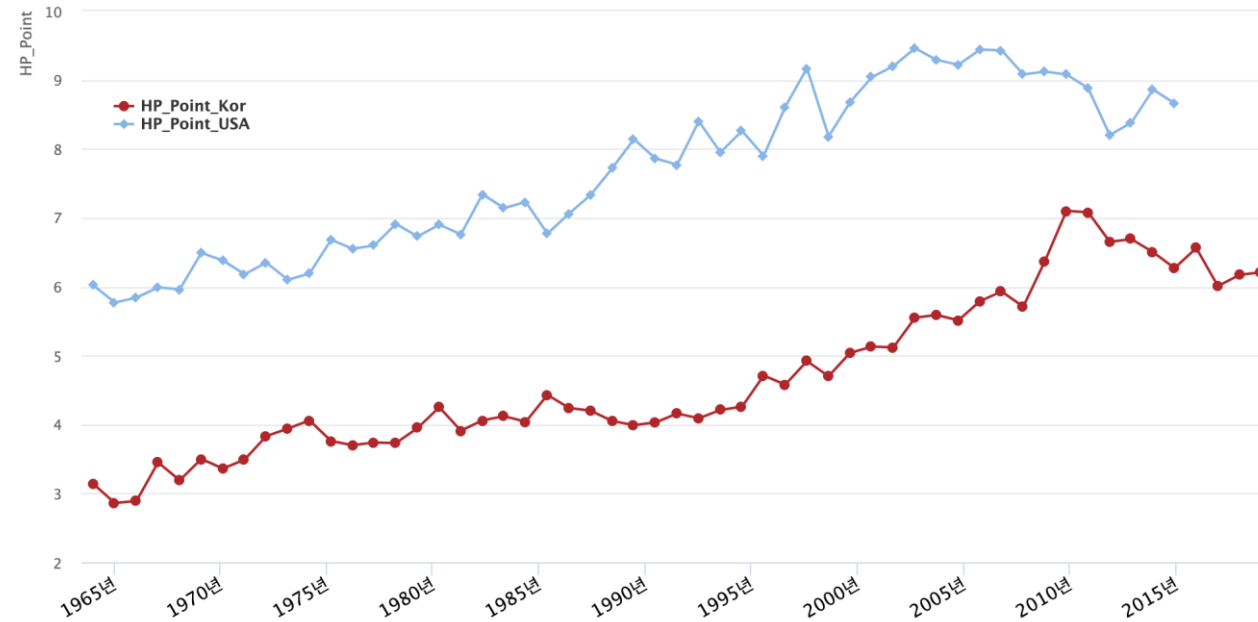
- 1990년대부터 힙합이라는 장르가 유행하기 시작(ex. 서태지와 아이들, 듀스, 현진영 등)
- 힙합과 락 장르의 유행을 남성 가수들이 선도하였으므로 1990년대에 남성 가수의 비중이 압도적으로 높아지고, 이에 따라 여성 가수의 비중은 낮아진 것으로 보임

년대별 빈도수 상위 20개 고유명사 중 장소 관련 단어

1960	1970	1980	1990	2000	2010
울릉도	명동	영동	개성	죽도	제주도
팔도	강원도	정동	평양	독도	한강
박달재	소양강	독도	대동	정동	서면
추풍령	지리산	을지	서면	제주도	강동
영산강	한라산	제주도	소양강	서면	미국
지리산	금산	남포동	한산도	신사동	진도
연평도	섬진강	명동	한강	미국	동양
목동	하동	대동강	칠갑산	명동	홍원
흑산도	금강산	섬진강	청해	경림	보성
진도	팔도	하동	혜화동	사동	용현동
평양	설악산	서면	명동	설악산	구월동
유달산	대관령	무진	두만강	동강	
계룡산	강릉	울릉도	지리산		
남산	울릉도	울릉군	광진		
남도	태백산	남면도동	강릉		
금강산	충주	미국			
	속리산	뉴욕			
	계룡산				

- 예전 노래에 ‘산’이 많이 등장하는 경향을 보임(자연을 더 가까이 하는 성향의 영향을 받은 것으로 보임)
- 예전 노래에 ‘북한’, ‘한반도’와 관련된 단어가 많이 등장하는 경향을 보임(6.25 전쟁 이후로 영향을 받은 것으로 보임)
- 1980년대부터 ‘미국’ 등 해외와 관련된 단어가 등장하기 시작함
- 최근으로 올수록 단어가 나타내는 장소의 범위가 작아지는 경향을 보임(ex. 도 -> 동)

연도별 H-P-point(평균) 추이



- 최근으로 올수록 H-P-point가 증가하는 양상을 보임
- 한국 곡의 H-P-point가 미국 곡의 H-P-point보다 항상 적음. 언어의 특성에 영향을 받았을 것으로 추정



최근으로 올수록 곡에서 반복되는 단어의 수가 증가하는 추세를 보인다

웹 소개

“ 시대별 유행했던 노래의 가사를 당시 시대상과 연관지어 분석하고 시각화한 결과를 Web으로 interactive하게 구현 ”

Q & A



THANK YOU

♡ 류재원 주어진 김민정 유건욱 이용하 임지우 정성균 ♡