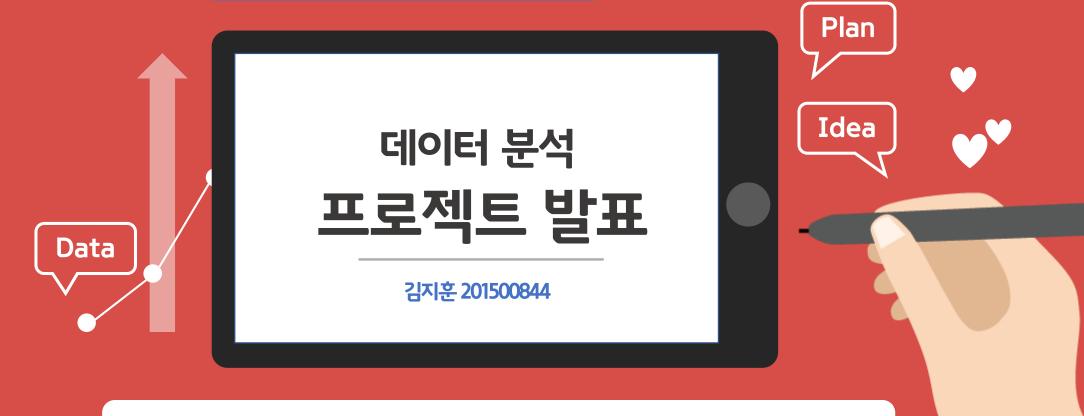
중국문화데이터포트폴리오



인기가요 가사 언어 분석

가사를 보면 시대상이 보인다. 어떤 가사의 노래가 대중으로부터 인기를 얻을까?

"

중국문화데이터포트폴리오 <mark>데이터 분석 프로젝트 빌</mark>



중국문화데이터포트폴리오 **데이터 분석 프로젝트 빌**



3. 대상

중복을 제외한 2000년 1월 ~ 2020년 10월까지의 멜론 월별 차트 10위 안의 국내음악의 가사

중복을 제외한 1964년 ~ 2020년까지의 멜론 연도별 차트 50위 안의 국내 음악 가사



4. 데이터 스키마

 \bigcirc

- · id
- title
- singer
- album
- genre
- year
- lyrics



- title
- singer
- year
- lyrics

AFTER

•

5. 방법



연도별로 분류된 국내음악 가사에 대한 언어 분석 (형태소 분석) 수행

1. 데이터 수집



2. 형태소 분석



3. 품사 분석 및 해석



5. 방법

데이터 수집: 2595개

→ 2276개 (중복제거)

	합체	형태소	품사	분류	연도 분류	카운트	상대빈도	시각화용
	!_Punctuation	!	Punctuation	1966	1960s	2	0.001002506	!!
	!_Punctuation	!	Punctuation	1987	1980s	14	0.002110977	11111111111111
	!_Punctuation	!	Punctuation	1992	1990s	4	0.000619387	1111
	!_Punctuation	!	Punctuation	1996	1990s	1	0.000106519	!
	!_Punctuation	!	Punctuation	2001	2000s	1	0.00010015	!
	!_Punctuation	!	Punctuation	2002	2000s	69	0.006545869	11111111111111
	!_Punctuation	!	Punctuation	2003	2000s	10	0.000871992	11111111111
	!_Punctuation	!	Punctuation	2004	2000s	17	0.001530429	11111111111111
	I Punctuation	!	Punctuation	2005	2000s	49	0.004206009	111111111111111
		!	Punctuation	2006	2000s	1	9.22679E-05	!
- 1\		!	Punctuation	2010	2010s	81	0.005581203	11111111111111
거) 🔛 💮		!	Punctuation	2020	2010s	21	0.001706346	111111111111111
	!!_Punctuation	!!	Punctuation	2002	2000s	2	0.000189735	!! !!
	!"_Punctuation	!"	Punctuation	2002	2000s	1	9.48677E-05	i
	!)_Punctuation	!)	Punctuation	2005	2000s	1	8.58369E-05	!)
형태소 분석 및 데이터 정제		!)		2010	2010s	1	6.89037E-05	
l 엉대소 눈꺽 5	!! 네이터 성세	!)		2020	2010s	1	8.12546E-05	,
<u> </u>	<u> </u>	n .	Punctuation	2002	2000s	1	9.48677E-05	
	\$_Punctuation	\$	Punctuation	2016	2010s	1	7.34107E-05	\$
	\$exy_CashTag	\$exy	CashTag	2016	2010s	1	7.34107E-05	\$exy
	&_Punctuation	&	Punctuation	1997	1990s	1	0.000122459	&
	&_Punctuation	&	Punctuation	2005	2000s	5	0.000429185	& & & & & &





6. 분석 및 해석

인기가요 노래 주제 사랑

명사로 본 차이

명사 '사랑'이 출현한 상대빈도의 총 합계 1.463%

6. 분석 및 해석

화자와 대상

3.412%

0.510%

'나' 관련 대명사 및 '우리' 관련 대명사 상대빈도 총 합계 비교





6. 분석 및 해석

의별에 대한 자세 잊지 못한다

사랑하는 이와 이별에 대해 이별을 극복하기 보다는 상대방을 잊지 못하고, 혼자 남아 아파하고, 하염없이 기다릴 것이라는 태도 : 이별로 인한 아픔에 대하여 노래 가사를 통해 공감과 위안을 받고자 함

감사합니다





사랑_Noun	1.525%	1.417%	2.073%	1.531%	1.572%	0.757%	1.463%
마음_Noun	0.888%	0.927%	0.960%	0.491%	0.226%	0.159%	0.581%
가슴_Noun	0.632%	0.326%	0.453%	0.217%	0.248%	0.120%	0.308%
눈물_Noun	0.343%	0.284%	0.403%	0.304%	0.345%	0.159%	0.301%
밤_Noun	0.470%	0.332%	0.449%	0.180%	0.119%	0.274%	0.291%
속_Noun	0.357%	0.186%	0.419%	0.373%	0.265%	0.169%	0.287%
때_Noun	0.321%	0.249%	0.216%	0.352%	0.282%	0.201%	0.264%
생각_Noun	0.197%	0.284%	0.330%	0.299%	0.176%	0.181%	0.248%
눈_Noun	0.195%	0.297%	0.289%	0.196%	0.218%	0.227%	0.240%
모습_Noun	0.145%	0.184%	0.371%	0.370%	0.160%	0.092%	0.227%
꿈_Noun	0.348%	0.305%	0.299%	0.223%	0.139%	0.098%	0.225%
세상_Noun	0.236%	0.169%	0.197%	0.377%	0.237%	0.111%	0.220%
맘_Noun	0.079%	0.182%	0.092%	0.151%	0.289%	0.311%	0.204%
길_Noun	0.438%	0.345%	0.186%	0.152%	0.074%	0.061%	0.191%
바람_Noun	0.163%	0.241%	0.420%	0.112%	0.077%	0.086%	0.185%
지금_Noun	0.154%	0.162%	0.192%	0.275%	0.126%	0.180%	0.184%
하늘_Noun	0.193%	0.316%	0.259%	0.142%	0.109%	0.088%	0.182%
꽃_Noun	0.614%	0.236%	0.144%	0.055%	0.031%	0.056%	0.176%
님_Noun	0.338%	0.278%	0.141%	0.035%	0.022%	0.010%	0.174%
노래_Noun	0.123%	0.313%	0.241%	0.115%	0.048%	0.118%	0.162%
오늘_Noun	0.124%	0.150%	0.166%	0.128%	0.177%	0.195%	0.161%
시간_Noun	0.122%	0.084%	0.125%	0.237%	0.134%	0.183%	0.159%
곁_Noun	0.105%	0.120%	0.196%	0.199%	0.151%	0.098%	0.150%
소녀_Noun	0.713%	0.174%	0.192%	0.044%	0.047%	0.009%	0.135%
혼자_Noun	0.181%	0.184%	0.082%	0.156%	0.113%	0.110%	0.134%
손_Noun	0.197%	0.176%	0.151%	0.097%	0.104%	0.102%	0.128%
추억_Noun	0.108%	0.113%	0.207%	0.141%	0.127%	0.065%	0.126%

명사로 본 차이

- 다음의 단어들은 1960년대부터 지금까지 공통적으로 빠지지 않고 인기가요 가사에 나오던 단어들이다.
- 이는 사람들에게 인기 있는 가요들의 주제 또는 제재가 '사랑, 눈물, 생각, 마음' 등의 서정적인 것들과 '밤, 세상, 바람, 길, 하늘, 꽃' 등의 자연 풍경을 묘사하는 것들이 주를 이룬다고 할 수 있다.
- '사랑'은 분석전에 예상했듯이, 역사상 한 번도 빠지지 않고, 가장 많이 사용된 단어이다. 결국 인기 가요 가사에서 가장 많이 나타나는 주제가 사랑이라는 것임을 알 수 있다. 그만큼 사람들의 공감을 많이 얻을 수 있고, 또 대중적으로 선호하는 노래 주제 역시 사랑 관련 노래임을 알 수 있다.
- 그러나 2010년 이후 '사랑'의 사용 빈도 비율이 그 전 시기에 비해 절반 가량 줄었다. 이는 오늘날 노래 주제가 '사랑' 외에도 다른 주제가 인기를 얻으며 다양화되고 있다고 추측할 수 있다.
- '마음'의 단어의 경우에는 시간이 지남에 따라 사용 빈도 비율이 줄어드는 것을 확인할 수 있었다. 이는 '마음'의 줄임말인 '맘'이 시간이 지남에 따라 사용 빈도가 늘어나는 것과 함께 보았을 때, 마음을 맘으로 줄이는 언어 양상에 의한 것이라 볼 수있을 것이다. 또한 이는 '눈물'이나 '가슴' 등의 단어 사용 빈도 역시 줄어드는 것과함께 보면, 시간이 지날수록 보다 가사에서 자신의 타인에 대한 생각이나 마음을 직설적으로 묘사하기 보다는 상황 자체를 묘사하면서 간접적으로 드러내려 하거나, 직접적으로 드러내는 것을 약간 민망하게 (오글거리게) 보기 때문이라 볼 수 있다.

I							
내_Noun	0.984%	1.157%	1.096%	1.375%	1.665%	1.370%	1.297%
나_Noun	0.569%	1.182%	1.421%	1.529%	1.187%	0.959%	1.178%
난_Noun	0.150%	0.235%	0.401%	0.650%	0.797%	0.663%	0.556%
I_Alpha	0.340%			0.096%	0.334%	0.556%	0.381%
							3.412%
You							
너_Noun	0.145%	0.539%	0.539%	1.780%	1.521%	1.452%	1.090%
그대_Noun	0.423%	0.580%	1.465%	0.982%	0.768%	0.310%	0.770%
널_Noun		0.078%	0.063%	0.454%	0.548%	0.446%	0.396%
당신_Noun	0.379%	0.580%	0.587%	0.098%	0.098%	0.041%	0.307%
네_Noun	0.043%	0.154%	0.125%	0.269%	0.219%	0.469%	0.257%
YOU_Alpha		0.158%	0.209%	0.091%	0.251%	0.329%	0.256%
							3.077%
We							
우리_Noun	0.253%	0.451%	0.580%	0.500%	0.277%	0.323%	0.409%
우린_Noun			0.137%	0.105%	0.079%	0.091%	0.101%
							0.510%
우리둘_Noun		0.074%	0.035%	0.034%	0.009%	0.012%	0.023%

대명사로 본 차이

전반적으로 화자 자신(I)에 대한 지칭대명사가 많이 나오는 것으로 보아, 화자가 말하는 대상보다는 화자 자신의 감정 이나 생각을 더 드러내려는 경향을 많이 보인다.

대상(YOU)은 화자(I)보다 다양한 대명사가 활용되고, 시간이 지남에 따라 사용하는 대명사가 서로 다름을 알 수 있다. 1960년 대에는 대상을 '너'라고 지칭하기 보다는 '당신'이나 '그대' 등으로 보다 높여 말했다면, 시간이 지남에 따라, '당신'이나 '그대' 등의 대명사 대신 보다 대상을 낮춰 말하며, '너'의 사용빈도가 높아진 것으로 나타났다.

대체적으로 화자가 말하는 대상은 둘 이상보다는 하나를 지칭하는 경우가 많았다. 이는 We보다 You의 사용빈도가 더 높은 것에서 알 수 있으며, We의 사용빈도에 I와 You를 합친 We도 있는 것 역시 감안할 때, 대상이 둘인 경우는 하 나일 때보다 더 적을 것이라고 할 수 있다.





잊지_Verb	0.306%	0.073%	0.101%	0.046%	0.045%	0.027%	0.077%
마세요_Verb	0.090%	0.048%	0.104%	0.093%	0.013%	0.007%	0.073%
피는_Verb	0.171%	0.094%	0.077%	0.012%	0.014%	0.013%	0.073%
찾아_Verb	0.173%	0.084%	0.070%	0.056%	0.048%	0.024%	0.069%
따라_Verb	0.162%	0.079%	0.074%	0.024%	0.036%	0.055%	0.065%
두고_Verb	0.093%	0.096%	0.083%	0.045%	0.043%	0.036%	0.065%
떠나_Verb	0.030%	0.072%	0.084%	0.090%	0.056%	0.029%	0.065%
타는_Verb	0.155%	0.070%	0.072%	0.025%	0.010%	0.015%	0.063%
돌아_Verb	0.080%	0.167%	0.042%	0.052%	0.034%	0.021%	0.063%
젖어_Verb	0.096%	0.043%	0.060%	0.035%	0.015%	0.011%	0.040%
만나고_Verb	0.042%	0.148%	0.046%	0.036%	0.019%	0.033%	0.040%
떠날_Verb	0.139%	0.103%	0.046%	0.031%	0.014%	0.018%	0.040%
가고_Verb	0.141%	0.138%	0.085%	0.038%	0.029%	0.020%	0.071%
말아요_Verb	0.035%	0.075%	0.150%	0.067%	0.046%	0.037%	0.071%
울어_Verb	0.063%	0.131%	0.074%	0.019%	0.036%	0.020%	0.060%
봐도_Verb	0.092%	0.042%	0.051%	0.040%	0.070%	0.064%	0.059%
잊을_Verb	0.087%	0.132%	0.078%	0.065%	0.030%	0.018%	0.064%
떠난_Verb	0.118%	0.092%	0.097%	0.059%	0.046%	0.019%	0.064%
젖은_Verb	0.084%	0.116%	0.063%	0.045%	0.018%	0.008%	0.055%
흘러_Verb	0.116%	0.097%	0.052%	0.031%	0.039%	0.022%	0.054%
만나_Verb	0.139%	0.124%	0.046%	0.031%	0.024%	0.051%	0.051%
어린_Verb	0.110%	0.103%	0.043%	0.055%	0.018%	0.014%	0.051%
흘러서_Verb	0.074%	0.095%	0.051%	0.011%	0.022%	0.007%	0.050%
몰라요_Verb	0.039%	0.364%	0.041%	0.021%	0.011%	0.012%	0.050%
싶은_Verb	0.081%	0.105%	0.027%	0.041%	0.038%	0.036%	0.050%
떠나간_Verb	0.090%	0.105%	0.065%	0.032%	0.026%	0.022%	0.050%
흐르는_Verb	0.098%	0.057%	0.075%	0.037%	0.027%	0.012%	0.050%
가버린_Verb	0.082%	0.044%	0.060%	0.035%	0.008%	0.007%	0.049%
기다려요_Verl	0.210%	0.048%	0.035%	0.042%	0.018%	0.032%	0.049%
왔다가_Verb	0.066%	0.060%	0.037%	0.043%	0.013%	0.008%	0.049%
주고_Verb	0.030%	0.181%	0.032%	0.077%	0.019%	0.021%	0.049%
기다려_Verb	0.085%	0.123%	0.031%	0.048%	0.036%	0.029%	0.047%

동사로 본 차이

- 다음의 단어들은 1960년대부터 현재까지 공통적으로 빠지지 않고 인기가요 가사에 나오던 동사이다. '잊다, 떠나다, 울다, 가버리다, 기다리다' 등의 이별과 헤어짐의 정서와 연관이 있는 동사들이 공통적으로 많이 나타나고 있다. 여기서 시대를 막론하고 공통적으로 나타나는 가사의 제재는 사랑하는 사람과의 이별이나, 사랑하는 사람과의 이별로 인해 나타나는 슬픔, 기다림 등의 정서일 것이다. 나아가 이러한 슬프고 아련한 정서를 잘 묘사할 수 있는 동사인 '젖다, 흐르다, 타다, 돌다, 만나다, 어리다' 등의 동사 역시 공통적으로 많이 활용되고 있음을 알수 있다. 눈물이나 빗물에 옷이나 몸이 젖어서 더욱 이별로 인한 슬픔을 강조하거나, 빗방울이나 물이 흐르고, 또 시간이 흐르거나 눈에 사랑하는 이의 모습이 어리는 등 슬픔과 기다림등을 강조하고 묘사할 수 있는 동사들이 꾸준하게 활용되고 있음을 알수 있다.
- 또한 다음의 동사들은 1960년대부터 2020년까지 빠지지 않고 사용된 단어이지만, 단어 사용비율에 있어서는 시대에 따라 차이를 보인다. 공통적으로 활용된 동사들은 대체적으로 시대가 흐름에 따라 사용 빈도가 줄어 들고 있음을 알 수 있다. 예를 들어 동사 '잊다'의 '잊지'의 경우에는 1960년대에는 0.306%의 사용 비율을 보였으나, 시간이 지나고 2010년대에는 0.027%로 0.279%p 줄어든 것으로 나타났다. 이 외에도 '마세요, 피는, 찾아, 따라' 등등 대부분의 동사가 1960년대에 쓰였던 것과 비교했을 때, 2010년대에 많이 쓰이지 않고 있음을 알수 있다. 이러한 현상이 나타나는 이유는 현대 노래 가사의 잦은 외국어 사용, 상대적으로 다양해진 어휘와 표현 사용, 다양해진 노래 주제로 인한 특정 어휘 사용의 감소 등으로 추측할수 있을 것이다.





그리운_Adjective	0.168%	0.192%	0.096%	0.038%	0.035%	0.020%	0.102%
같은_Adjective	0.090%	0.119%	0.136%	0.105%	0.141%	0.124%	0.123%
슬픈_Adjective	0.052%	0.124%	0.127%	0.163%	0.051%	0.032%	0.095%
아름다운_Adjective	0.110%	0.145%	0.175%	0.055%	0.042%	0.032%	0.094%
영원히_Adjective	0.091%	0.100%	0.096%	0.124%	0.053%	0.024%	0.076%
먼_Adjective	0.137%	0.155%	0.082%	0.075%	0.038%	0.009%	0.076%
외로운_Adjective	0.156%	0.083%	0.105%	0.048%	0.014%	0.011%	0.076%
좋아_Adjective	0.130%	0.081%	0.049%	0.050%	0.034%	0.080%	0.064%
고운_Adjective	0.113%	0.105%	0.067%	0.031%	0.018%	0.016%	0.064%
좋은_Adjective	0.040%	0.119%	0.049%	0.069%	0.086%	0.041%	0.063%
길은_Adjective	0.105%	0.194%	0.040%	0.034%	0.012%	0.010%	0.063%
하얀_Adjective	0.136%	0.164%	0.092%	0.065%	0.028%	0.017%	0.079%
이런_Adjective	0.061%	0.034%	0.032%	0.099%	0.079%	0.080%	0.078%
많은_Adjective	0.129%	0.078%	0.040%	0.155%	0.048%	0.033%	0.075%
조용히_Adjective	0.132%	0.103%	0.052%	0.019%	0.012%	0.014%	0.058%
젊은_Adjective	0.137%	0.110%	0.048%	0.024%	0.009%	0.015%	0.057%
머나먼_Adjective	0.068%	0.089%	0.033%	0.045%	0.007%	0.007%	0.056%
그리워_Adjective	0.110%	0.093%	0.059%	0.032%	0.031%	0.051%	0.054%
마라_Adjective	0.163%	0.047%	0.046%	0.015%	0.010%	0.031%	0.054%
있게_Adjective	0.091%	0.021%	0.115%	0.065%	0.048%	0.035%	0.051%
없다_Adjective	0.202%	0.104%	0.063%	0.020%	0.023%	0.029%	0.051%
아픈_Adjective	0.050%	0.052%	0.075%	0.034%	0.060%	0.031%	0.050%
있겠지_Adjective	0.100%	0.111%	0.039%	0.027%	0.027%	0.008%	0.049%
입니다_Adjective	0.061%	0.089%	0.098%	0.011%	0.027%	0.018%	0.049%
없어도_Adjective	0.030%	0.227%	0.041%	0.031%	0.053%	0.026%	0.048%

없는_Adjective	0.132%	0.178%	0.244%	0.318%	0.193%	0.176%	0.214%
없어_Adjective	0.130%	0.084%	0.125%	0.234%	0.164%	0.200%	0.169%
있어_Adjective	0.061%	0.161%	0.084%	0.205%	0.135%	0.120%	0.139%
있는_Adjective	0.087%	0.102%	0.131%	0.180%	0.098%	0.093%	0.120%
없어요_Adjective	0.196%	0.134%	0.099%	0.036%	0.020%	0.019%	0.070%
없네_Adjective	0.104%	0.103%	0.102%	0.017%	0.018%	0.014%	0.066%

형용사로 본 차이

- 다음의 단어들은 1960년대부터 지금까지 공통적으로 빠지지 않고 인기 가요 가사에 나오던 형용사들이다. 우선 '없다'와 '있다'를 비교했을 때, '없다'의 비율이 '있다'보다 더 많음을 알 수 있는데, 이는 노래 가사에서 대상에 대한 존재보다 대상에 대한 상실에 대해 더 많은 감정을 느끼고 이에 대해 더 얘기하고 싶어하는 경향을 보인다고 추측할 수 있다. 즉, 대상을 가지고 있는 것보다 대상의 부재로 인한 상실감이나 허탈함, 슬픔, 그리움 등의 감정이 사람들에게 더 많은 인기를 얻는다는 것이다. 이는 사람들이 대상의 부재로 인한 슬픔 등의 부정적인 감정에 대하여 노래를 통해 공감을 얻고 위로를 받고 싶어하기 때문이라고 볼 수 있을 것이다.
- 또한 '그립다, 슬프다, 아름답다, 영원하다, 외롭다, 좋다, 곱다, 젊다, 머나 멀다, 아프다' 등등의 형용사 역시 앞서 언급한 이별이나 그리움 등의 정 서를 잘 나타낼 수 있는 형용사로, 모든 시기의 노래가 서로 비슷한 주제 를 공유하고 있음을 알 수 있다.



언제나_Adverb	0.154%	0.108%	0.106%	0.158%	0.093%	0.043%	0.105%
언젠가_Adverb	0.050%	0.087%	0.097%	0.101%	0.046%	0.020%	0.064%
정말로_Adverb	0.081%	0.246%	0.028%	0.043%	0.022%	0.031%	0.061%
너무나_Adverb	0.094%	0.079%	0.052%	0.054%	0.064%	0.029%	0.058%
그만_Adverb	0.100%	0.101%	0.126%	0.028%	0.030%	0.045%	0.055%
외로이_Adverb	0.097%	0.069%	0.034%	0.019%	0.011%	0.009%	0.053%
어느_Adverb	0.135%	0.041%	0.079%	0.057%	0.038%	0.021%	0.052%
그래도_Adverb	0.090%	0.091%	0.065%	0.062%	0.037%	0.015%	0.051%
아무리_Adverb	0.092%	0.040%	0.051%	0.021%	0.045%	0.032%	0.047%
한없이_Adverb	0.055%	0.093%	0.048%	0.021%	0.039%	0.013%	0.046%
끝없이_Adverb	0.055%	0.237%	0.050%	0.020%	0.014%	0.008%	0.039%
어쩌다_Adverb	0.097%	0.078%	0.043%	0.011%	0.052%	0.009%	0.038%
살며시_Adverb	0.137%	0.052%	0.035%	0.017%	0.011%	0.009%	0.037%
오랜_Adverb	0.039%	0.158%	0.036%	0.020%	0.022%	0.039%	0.036%
빨리_Adverb	0.050%	0.157%	0.026%	0.030%	0.018%	0.018%	0.033%
옛_Adverb	0.061%	0.060%	0.035%	0.020%	0.009%	0.011%	0.032%
어딜_Adverb	0.085%	0.060%	0.027%	0.021%	0.017%	0.019%	0.030%
너무나도_Advert	0.050%	0.038%	0.056%	0.030%	0.019%	0.017%	0.027%
어디서_Adverb	0.052%	0.032%	0.026%	0.024%	0.014%	0.025%	0.026%

부사로 본 차이

- 다음의 단어들은 1960년대부터 지금까지 공통적으로 빠지지 않고 인기가요 가사에 나오던 부사들이다. '언제나, 언젠가, 어느, 오랜, 빨리, 옛' 등과 같이 시간과 관련된 부사들이 많이 사용되는데, 시간에 대한 화자의 인식은 과거에 대해서 회상을 하거나, 시간이 빠르게 흘러 가거나, 아니면 시간과 상관없이 항상 또는 언제나 그 상태를 유지하는 것과 같이한결 같음을 강조하고자 하는 것으로 보인다. 이는 '한없이, 끝없이, 아무리, 살며시, 너무나도, 너무나, 정말로' 등의 부사 역시 많이 쓰였다는 점과 함께 볼 때, 화자가 처한 상태를 보다 강조하고자 하면서도, 시간의제약을 받지 않고 태도를 굳건하게 유지하고자 하는 화자의 의지가 보인다고 할 수 있다.
- 공통적으로 나타나는 부사에서의 정서는 회상, 기다림, 절실함, 한결같음, 슬픔, 외로움 등이다. 이는 가장 많이 쓰인 명사가 사랑으로 전 시대를 통틀어 가장 많이 나타난 인기가요 주제가 사랑 노래라는 점과 함께 해 석하자면, 예전에 사랑하던 사람이 현재는 곁에 없어 외롭지만 사랑하는 마음을 버리지 않고 한결같이 가지고 끝없이 기다릴 것이라는 외로움과 기다림을 통한 외로움의 극복 의지를 보여준다고 할 수 있다.