

한국어정보처리

2014130069 박찬희

SynKDP 사용 결과 보고서 장범준 1-3집

1. 활용 텍스트

지난번 과제로 사용했던 장범준 1-3집 가사를 코퍼스로 활용했다.
빈도수를 기준으로 하여 상위 10개 항목들을 추려서 분석해보았다.

2. 결과 분석

1) 어절 빈도 결과

어절	빈도 ▼	백분율	누적율
이	50	1,44802%	1,44802%
난	36	1,04257%	2,49059%
내가	36	1,04257%	3,53316%
없는	31	0,89777%	4,43093%
너무	30	0,86881%	5,29974%
그대로	28	0,81089%	6,11063%
내	28	0,81089%	6,92152%
그	27	0,78193%	7,70345%
더	27	0,78193%	8,48538%
그대는	24	0,69505%	9,18042%

- 1차 분석 결과와 유사하게 1인칭 대명사 '나'가 포함된 '난'/'내가'/'내' 가 상위 10개 항목 안에 포함되어 있는 모습을 볼 수 있다. 세 개 항목의 총 지분은 전체의 2.89603%였다.
- 같은 맥락으로 2인칭 대명사 '그대'가 포함된 '그대는'/'그대로' 도 마찬가지로 높은 빈도를 보였다. 다만 이 경우에 '그대로'는 '그대'에 격조사 '로'가 더해진 경우와 부사 '그대로'로 쓰인 경우 모두 집계되었다. 두 항목의 총 지분은 1.50594%였다.
- 관형사 '이'/'그'도 도합 2.22995%로 많은 지분을 차지하고 있다. 다만 '이'는 주격 조사로 쓰인 경우도 감안해야 한다.

2) 음절 빈도 결과

음절	빈도 ▼	백분율	누적율
이	309	3,52659%	3,52659%
그	239	2,72769%	6,25428%
는	215	2,45378%	8,70806%
머	211	2,40813%	11,11618%
가	201	2,29400%	13,41018%
마	169	1,92878%	15,33896%
메	166	1,89454%	17,23351%
요	146	1,66629%	18,89979%
다	144	1,64346%	20,54325%
사	136	1,55216%	22,09541%

- 주격 조사로 쓰이는 '은'/'는'/'이'/'가' 중 '은'을 제외한 나머지가 상위 10개 항목 안에 포함되어 있는 것이 상당히 흥미롭다. 종결어미의 대표격인 '다'의 빈도수와 종합해봤을 때 (음절 각각이 쓰이는 다양한 맥락과 어절을 감안하더라도) 장범준의 가사에서 'A가 B이다'와 같은 문장 구조가 많이 쓰였다는 것을 어느 정도 짐작해볼 수 있는 부분이다. 세 개 항목의 총 지분은 8.27437%였다.
- 또한 이 경우에 끝 어절로만 쓰인 보조사 '요' 역시 총 146회 사용되었는데, 이는 존대어 표현을 많이 사용하는 장범준 가사의 특징이라고도 할 수 있다.
- 마지막으로 10위 항목인 '사'는 지난번 91회의 빈도수로 가장 많이 언급된 단어 '사랑'의 영향이 큰 것으로 보인다.

3) 음소 빈도 결과

음소	빈도	백분율 ▼	누적율
중성-ㅏ	2237	10,84764%	10,84764%
초성-ㅇ	2085	10,11056%	20,95820%
중성-ㄴ	1226	5,94511%	26,90331%
중성-ㅡ	1128	5,46989%	32,37319%
초성-ㄱ	1009	4,89283%	37,26603%
중성-ㅣ	995	4,82494%	42,09097%
중성-ㄷ	994	4,82010%	46,91107%
초성-ㄴ	881	4,27214%	51,18320%
중성-ㅓ	814	3,94724%	55,13044%
중성-ㄹ	726	3,52051%	58,65096%

- 상위 2개 음소가 총 20%의 빈도수를 차지하고 있는 모습이 눈에 띈다.
- 나머지 항목들은 대개 3-5%대로 비교적 유사한 분포를 보이고 있다.
- 초성 'ㅇ'와 중성 'ㅏ'가 많이 쓰인 이유에 대해 다음과 같은 가설을 세워볼 수 있다.
 - 1) 모음으로 시작되는 어절에는 음가 없는 초성 'ㅇ'가 수반되기 때문이다.
 - 2) 초성 'ㄱ'와 'ㄴ'이 많이 쓰였다는 점과 함께 유추해보자면, 중성 'ㅏ' 역시 모음의 대표격으로 많은 단어에 사용되기 때문이다.
 - 3) 장범준이 '아아'와 같은 감탄사의 사용을 선호하기 때문이다.
- 흥미롭게도, 소설창작연습의 과제로 제출한 필자의 소설의 원고로 동일하게 음소 빈도를 통계 냈더니 상당히 유사한 결과가 발견되었다.

음소	빈도 ▼	백분율	누적율
초성-ㅇ	2693	9,80057%	9,80057%
중성-ㅏ	2663	9,69139%	19,49196%
중성-ㅣ	1728	6,28867%	25,78062%
중성-ㅡ	1545	5,62268%	31,40330%
중성-ㄴ	1534	5,58265%	36,98595%
초성-ㄱ	1482	5,39341%	42,37936%
중성-ㄷ	1395	5,07679%	47,45615%
초성-ㄴ	1160	4,22156%	51,67771%
중성-ㄹ	1073	3,90494%	55,58265%
중성-ㅓ	1056	3,84307%	59,42572%

- 초성 'ㄴ'과 초성 'ㄷ'을 제외하곤 상위 10개 항목이 정확하게 일치한다. 이를 통해 어쩌면 자주 사용하는 음소의 유형이 정해져 있다고 가정해볼 수 있을지도 모르겠다.