

# 소설 속 인물의 감정 분석을 위한 감정 용언 사전 제안

김규희<sup>0,1</sup>, 이수린<sup>2</sup>, 김명재<sup>3</sup>

서울대학교

salgu@snu.ac.kr, lsrwhite@snu.ac.kr, kmjj9999@snu.ac.kr

## Emotion Verb Dictionary for Emotional Analysis on Characters in Novel

Kyu-Hee Kim<sup>0, 1</sup>, Surin Lee<sup>2</sup>, Myung-Jae Kim<sup>3</sup>  
Seoul National University

### 요약

감정 분석은 긍부정의 극성을 판단하는 감성 분석과 달리 텍스트로부터 구체적인 감정 유형을 분류해내는 과제이다. 본 논문에서는 소설 텍스트에 감정 분석을 수행하는 것을 새로운 과제로 설정하고, 이에 활용할 수 있는 감정 용언 사전을 소개한다. 이 사전에는 맥락과 상관없이 동일한 감정을 전달하는 직접 감정 표현과 맥락에 따라 다른 감정으로 해석될 수 있는 간접 감정 표현이 구분되어 있다. 우리는 이로써 한국어 자연어처리 연구자들이 소설의 풍부한 감정 표현 텍스트로부터 정확한 감정을 분류해낼 수 있도록 그 단초를 마련한다.

주제어: 감정 분석, 감정 어휘 사전, 소설

### 1. 서론

감정 분석(Emotional Analysis)은 텍스트의 감정을 여러 개의 감정 유형으로 분류하는 작업이다. 긍정, 부정, 중립의 극성(Polarity)을 판단하는 감성 분석(Sentiment Analysis)과 달리, 감정 분석은 기쁨, 슬픔, 두려움 등의 다양한 감정을 레이블로 사용한다.

이러한 감정 분석은 감정 어휘 사전이나 빅데이터를 통한 딥러닝 등을 활용한다. 그리고 주로 상품평이나 영화평 등 비교적 길이가 짧고 정형화된 패턴을 지닌 텍스트에 적용되어 왔다. 그러나 문학은 다양한 문체나 자유로운 형식으로 인해 데이터 수집 및 모델 학습에 어려움이 있다. 그중 소설은 직접적인 감정 표현뿐만 아니라 행위나 표정을 묘사하는 간접적 표현을 이용해 감정을 전달한다.

맥락과 상관없이 일정한 감정을 전달하는 직접 표현과 달리, ‘웃다’와 같은 간접 표현은 ‘기쁘게 웃었다’와 ‘슬프게 웃었다’처럼 맥락에 따라 여러 감정으로 해석될 여지가 있다. 그러나 기존의 감정 어휘 사전은 간접 감정 표현에 상식적으로 가장 어울리는 감정만 태깅했다는 한계가 있다. 즉, 사용된 맥락에 따라 ‘슬픔’으로 해석될 수도 있는 ‘웃다’에 ‘기쁨’이라는 감정만 태깅한 것이다. 이로 인해 소설의 간접 감정 표현에 대한 감정 분석의 정확도가 낮다는 것이 중요한 문제가 된다.

본고는 이 문제의 해결을 감정 표현의 종류를 구분한 감정 용언 사전을 구축하는 것에서 시작하고자 한다. 기존의 감정 어휘 사전을 검토 및 확장하여, 8가지 감정 범주로 태깅된 1,009개의 직접 표현과 가능한 감정 범주가 모두 태깅된 2,405개의 간접 표현을 가진 코퍼스를 제시할 것이다. 보완된 감정 코퍼스는 소설 텍스트의 풍부한 감정 표현에 대한 감정 분석의 정확도를 향상시키고, 나아가 소셜데이터 등 구어체 표현의 감정 분석에까지 널리 활용될 수 있을 것이다.

### 2. 관련 연구

감정표현 어휘의 범위를 정립하려는 시도로 여러 감정 어휘 목록들이 제시되었다[1-5]. 이 연구들에서 감정 어휘의 판단 기준을 확인할 수 있다. 공통적으로 사전이나 자료집에서 다량의 감정 관련 단어를 수집한 뒤 범위를 좁혀나가는 방식을 사용하나, 세부적인 과정에서 차이가 있다.

[1]과 [2]는 “나도 모르게 저절로 ~한 마음이 {생기다/들다}”로 표현가능해야 한다는 기준 같은 통사적 검증 기제를 설정하여, 각각 1,359개의 감정동사와 524개의 감정 용언을 제시한다. 그러나 최종 감정 어휘 목록에 파생어 등 중복되는 단어들 많고, 엄격한 검증으로 인해 일반적 상식에서 감정 어휘로 받아들여지는 것까지 배제되어, 실제 과제에 활용하기에는 적합하지 않다는 한계가 있다.

[3]은 설문조사를 통해 85% 이상의 피험자가 감정으로 평가한 총 213개의 단어를 감정동사 목록으로 제시했다. [4]는 [5]에 대해 대학생 설문조사를 진행하여 [5]의 어휘 목록을 압축하고 일부 어휘를 추가하여 총 265개의 정서 단어 목록을 제시했다. 그러나 설문조사는 피험자의 직관적 판단에 의존한다는 점에서 특정 어휘가 배제 혹은 포함된 이유를 알 수 없다는 한계가 있다.

[5]는 연구자들이 여러 차례의 합의를 거쳐 감정 단어로서의 적절성 여부를 평가했고 총 434개의 어휘를 제시했다. 전문 연구자들의 참여로 결과의 신뢰성이 높으나, 감정 표현에 해당하는 동사를 행위로 분류하여 배제하는 등 완전한 감정 어휘 목록이라고 볼 수 없다는 한계가 있다.

이에 [6]은 [5]를 비판 및 발전시켜, 배제되었던 감정 표현 단어들 포함 총 504개의 감정표현단어 목록을 완성했다. 그리고 이 목록에 기쁨, 슬픔, 분노 등 중성을 포함한 10개의 감정 범주를 추가하여 실제적인 감정

분석 과제에 활용가능한 감정 어휘 목록을 제시하였다. 각 어휘는 최대 3개의 감정 범주와 대응되며 한 감정 범주, 두 감정 범주처럼 대응가능한 감정 개수를 기준으로 3가지 층위로 나누어져 있다.

[7]은 감정 용언을 ‘감정 내용을 서술하는 용언’에 한정하고, 이를 제외한 용언들은 “감정과 관련된 영역의 일부분을 언급함으로써 화자가 특정 감정 상태에 있음을 추론할 수 있게 하는 환유적 표현”이라고 보았다. 그리고 4가지의 객관적 검증 기제(경험주 논항, 자극체 논항, 감정의 주관성, 마음 상태를 서술함)를 설정해 이 기준을 모두 통과하는 617개의 형용사 및 동사만을 감정 어휘 목록에 포함하였다. 이 목록은 감정 관련 행위나 감정에 동반되는 상태를 제외하고 감정 그 자체를 서술하는 용언만 선별했다는 장점이 있다. 또한 각 단어가 19가지의 감정 유형 중 하나의 감정으로 분류되어 각 단어가 표현하는 구체적 감정을 알 수 있다.

본 연구는 감정을 직접 서술한 용언과 환유적 표현, 즉 감정을 간접적으로 표현하는 용언을 구분한 [7]의 연구에서 착안하여, 맥락과 무관하게 감정을 전달하는 ‘직접 표현’과 맥락에 따라 해석이 열려있는 ‘간접 표현’을 구분한다. 그리고 오판 가능성을 줄이기 위해 여러 단계를 거쳐 감정 용언을 선정하고, 판단 기준의 모호성을 최소화하기 위해 각 단계별로 사용한 명확한 기준을 밝힐 것이다. 그 결과 감정표현 관련 용언들이 직접 표현과 간접 표현으로 구분된 감정 표현 코퍼스를 구축할 것이다.

### 3. 수집 과정

#### 3.1 재료

본고에서는 표준국어대사전과 우리말샘 사전을 감정 단어 목록의 기초로 삼았다. 이들 사전의 어휘가 감정 단어 목록을 만들기에 충분한지 살펴보기 위해, 7편의 선행연구[1-7]에서 제시한 감정 단어(2,557개)를 검토하였다. 선행연구에서 제시된 감정 단어와 표준국어대사전, 우리말샘사전에 등재된 어휘 간 차이가 발생하는 지점은 다음과 같다.

먼저, 사전에는 ‘V-어 하다’와 ‘V-어 지다’가 붙은 활용형이 모두 등재되지 않았다. [1]에서는 ‘황홀하다’와 같은 형용사에 대해 ‘황홀해하다’라는 활용형을 코퍼스에 포함한 것과 달리, 그러나 ‘귀여워하다’, ‘예뻐하다’, ‘싫어지다’와 같이 자주 쓰이는 일부 용언을 제외하고 ‘V-어하다’와 ‘V-어지다’가 붙은 활용형은 사전에 등재되지 않았다. 이에 본고에서는 ‘-어하다’와 ‘-어 지다’가 생산성이 있는 어미라는 점을 감안해 코퍼스에 별도로 포함하지 않았다.

다음으로, ‘풀죽다’와 같이 단어와 구의 경계에 있는 동사들이 등재되지 않았다. ‘풀죽다’는 원형이 ‘풀(이) 죽다’로 구성단위가 단어가 아닌 구로 여겨진다. 그러나 이러한 구 표현들 중 일부는 띄어쓰기가 되지 않은 채 하나의 단어로 취급되는 용례가 많고, 고려대사전에서는 하나의 단어로 등재되어 있다. 이에 본고에서는

표준국어대사전 및 우리말샘 사전에 등재된 두 어절 이하로 구성된 서술어구 표현들을 검토하고, 감정표현에 해당하는 경우 띄어쓰기와 조사의 결합에 따라 가능한 모든 형태를 코퍼스에 포함했다. 표준국어대사전과 우리말샘 사전에 등재되지 않은 구 표현에는 ‘짜증(이)나다’와 ‘주눅(이)들다’ 등이 있었는데, 이 표현들의 의미적 핵심은 명사라고 판단하여 코퍼스에 포함하지 않았다. 본고에서 검토한 단어와 구의 개수는 표제어와 의미 쌍 기준으로 95,908개이다. 간접 감정 표현은 10명의 실험참여자로부터 감정과 관련된 동작, 행동 등의 구 표현을 묘사하도록 하여 수집했다. 이후 중복되는 문장과 조사와 어미, 단어 배열 순서의 차이만 존재하는 문장을 삭제하였다.

### 3.2 주석 과정

한국어 감정 표현 용언은 그 표현 방식에 따라 두 가지로 나뉜다. 1) 감정을 직접적으로 표현하는 것(예. 기쁘다) 2) 행위를 표현하여 그에 따른 감정을 유추하게 만드는 것(예. 웃다) 1)은 그 자체로 감정을 내포하지만 2)는 관련된 감정을 추론하는데 단서를 제공하는 역할을 수행한다. 본고에서는 1)을 직접 감정 표현, 2)는 간접 감정 표현으로 구분하였다. 소셜 데이터의 특성상 감정을 표현할 때 간접 감정 표현이 많이 쓰이므로 이를 코퍼스에 포함하되, 직접 감정 표현과 달리 여러 가지 감정과 대응될 수 있다는 점에서 주석 과정에 차이를 두었다.

#### 3.2.1 직접 감정 표현

표1. 직접 감정 표현 코퍼스 구조

	감정	검색 형태 (자소 단위로 분해)
경악스럽다	놀람, 부정	경악스럽 / 경악스러워/ 경악스러우
증오스럽다	혐오	증오스럽 / 증오스러워/ 증오스러우
기뻐하다	기쁨	기뻐하 / 기뻐해

이 용언들은 맥락에 따라 대응되는 감정이 바뀌지 않는다. 먼저 표준국어대사전과 우리말샘에서 추출한 용언들에 대해 감정과 관련되었는지 판단한 후 1차로 감정 해석이 제한된 용언들을 골라냈다. 이 감정 용언들은 1) 주관성이 있고 2) 감정과 상관없이 자연 발생 불가능하며 3) 비의도적으로 발생하는 마음 상태에 대해 서술하는 용언들이다. 만들어진 감정 용언 목록에 3명의 주석자가 감정 범주 태깅에 참여했다. 태깅은 Ekman의 기본 감정 6가지 범주[8]와 긍정, 부정 범주를 용언과 연결하는 방식으로 진행되었고, 용언이 복합 감정을 의미하는 경우

기본 감정의 집합으로 표현하였다. 3명의 주석자 중 2명 이상의 동의를 얻은 것은 데이터에 포함하였다.

### 3.2.2 간접 감정 표현

이들 용언들은 감정 자체를 표현하기보다 감정과 관련된 반응 혹은 감정 추론의 단서가 되는 행위를 표현하며, 감정과 일대다의 대응 관계를 보인다. 주석자에게 행위에 대한 용언을 보여주고 해당 용언의 사용 맥락을 세 가지(긍정적인 해석, 부정적인 해석, 감정 외 요소)로 제시한다. 간접 감정 표현이 갖는 해석의 다양성을 반영하기 위해, Ekman 기본 감정으로 분류하지 않고 용언이 어떤 감정을 내포하는지 Open text 형태로 진술하게 하였다.

따라서 기본 감정보다 세밀하고 정확한 표현이 가능하며, 대응된 해석을 구분하기에도 용이하다. 신체적 상태 맥락은 주석자가 해석의 다양성을 고려하도록 도움을 주기 위한 장치로, 감정 분석 과제에서는 쓰이지 않는다. 완성된 데이터는 다시 3명의 주석자가 교차검증을 실시하여, 2명 이상의 동의를 얻으면 데이터에 포함하였다.

표2. 간접 감정 표현 코퍼스 형태

	긍정 해석	부정 해석	감정 외 요소
얼굴이 시뻘겋게 달아오른다	썩스럽다	수치스럽다, 화나다, 분노하다	
눈빛이 이글거린다	기대하다, 욕망하다	화나다, 분노하다	단호하다
눈물이 그렇그렇하다	감동받다, 감사하다	우울하다, 고뇌하다, 상처받다	감성적이다

## 4. 코퍼스 유형별 통계

최종 데이터는 기본 감정의 집합으로 해석되며 맥락의 영향을 받지 않은 직접 감정 표현과 기본 감정보다 세밀한 감정 범주를 요하며 맥락에 따라 두 가지 이상의 범주로 해석될 수 있는 간접 감정 표현(행위 표현)의 두 가지 층위로 구분된다. 직접 감정 표현은 Ekman에 해당하는 6가지 감정 범주에 긍정, 부정을 더해 8가지 감정 범주로 이루어져 있고, 각 범주별 통계는 부록에 제시되어 있다. 간접 감정 표현 2,405개의 행위 표현에 대해 특정 맥락에서 해석 가능한 감정 범주를 모두 태깅하여 제시된다.

## 5. 실험 및 결과

평가 과제는 소설 속에서 감정을 드러내는 표현을 찾아내어 감정 범주를 판단하는 것이다. 실험에 쓰인 평가 데이터셋은 위키문헌<sup>1</sup>에서 수집한 한국 근대 단편 소설 25편이며, 주석자들은 장면 단위로 나뉜 소설에 대해 감정을 드러내는 구를 표시하고, Ekman의 6가지 감정 및 긍, 부정 중에서 감정 범주를 선택하였다. 이때 구의 감정 범주는 중복 태깅이 가능하도록 하였다.

표3. 감정 태깅을 한 한국 근대 소설 데이터

감정 태깅된 문장	감정 범주
병주는 일과의 하나를 거저 넘긴 것 같이 부정 섭섭한 생각이 났다.	부정
이상스럽게도 오늘 밤에는 창경원 안 고자리 꿇듯 움직이는 많은 사람 중에서 아는 사람은 하나도 만나지 못한 것이 그를 더욱 쓸쓸하게 하였다. 슬픔, 부정	슬픔, 부정
그는 깜짝 놀라 머리를 돌이켰다. 놀람	놀람

표4. 평가 데이터의 감정 분포

	emo	count
0	슬픔	546
1	기쁨	138
2	혐오	152
3	놀람	195
4	긍정	546
5	부정	1,053
6	두려움	201
7	분노	472
8	총계	2,833

기존 감정 어휘 사전과의 성능 비교를 위하여 어휘의 감정 범주가 명시된 [6]와 [7]의 데이터를 이용하였다. 그리고 본 연구의 데이터와 감정 범주를 통일하기 위해 Ekman의 6가지 기본 감정 외의 범주는 다음과 같이 수정하였다. [6]의 10가지 감정 범주 중 4가지에 대하여 ‘흥미’는 ‘긍정’으로, ‘지루함’과 ‘통증’은 ‘부정’으로 바꾸었으며, ‘중성’은 배제하였다. [7]의 19가지 감정 범주 중 13가지에 대해서는 ‘감동’, ‘감사’, ‘수용’, ‘이완’, ‘상쾌함’을 ‘긍정’으로, ‘동정’을 ‘슬픔’으로, ‘고뇌’, ‘고독’, ‘미안함’, ‘수치심’, ‘허무’, ‘후회’, ‘욕구’를 ‘부정’으로

<sup>1</sup> <http://ko.wikisource.org>

바꾸었다.

성능 평가를 위해 소셜 데이터와 사전의 감정 어휘를 매칭하는 과정에서는 형태소 분석이 아니라 문장을 자소 단위로 분해하는 방식을 사용하였다. 각 감정 어휘 사전에 사용 빈도가 낮은 감정 어휘가 다수 존재하여, 형태소 분석기로는 일부 감정 표현이 매칭되지 못하였기 때문이다. 이에 소셜 문장의 어절을 자소 단위로 분해한 다음, 그 안에서 감정 어휘의 어근 및 불규칙 활용형의 자소를 찾는 방식으로 감정 표현을 탐색하였다. 단, ‘짜증이 나다’ 처럼 2어절로 구성된 구는 두 번째 어절이 첫 번째 어절로부터 4어절 이내로 떨어져 위치한 경우에만 감정 표현으로 인정하였다.

평가를 위한 실험은 각 코퍼스에 대해 두 번 진행된다. 1)Ekman 6가지 감정 범주에 대해 탐색하도록 하는 것, 2) 긍정/부정에 대해 탐색하도록 하는 것이다. 긍정, 부정 범주를 만들기 위해 Ekman의 6가지 범주 중 ‘기쁨’을 ‘긍정’으로, ‘슬픔’, ‘분노’, ‘두려움’, ‘혐오’를 ‘부정’으로 분류하고 긍/부정 판단과 거리가 먼 ‘놀람’은 제외하였다. 또한 Ekman의 6가지 범주로는 설명할 수 없는 감정들을 긍/부정 범주에 포함하여(예. 긍정: 사랑, 편안함; 부정: 절망, 허무) 폭넓은 감정을 검색할 수 있도록 하였다.

한 문장이 여러 가지 감정을 내포하고 있는 경우가 있기에 이 과제는 다중 레이블에 대한 다중 클래스 분류 문제로 환원된다. 따라서 다중 클래스 다중 레이블 평가 척도로 가장 최근에 고안된 MLCM[9]을 사용해 결과를 분석한다. 문장이 포함하는 감정 범주를 제대로 찾았다면 해당 범주에 대한 TP(True Positive)로 간주한다.

### 5.1 기본 감정 분류

표5. 기본 감정 분류 F1 점수

감정범주	손선주	홍중선, 정연주	Ours
기쁨	0.22	0.38	0.37
슬픔	0.26	0.17	0.24
분노	0.12	0.07	0.13
두려움	0	0.1	0.2
놀람	0.09	0.01	0.39
혐오	0.13	0.13	0.15
NTL	0.84	0.9	0.89
micro avg	0.71	0.81	<b>0.8</b>
macro avg	0.24	0.25	<b>0.34</b>
weighted avg	0.71	0.76	<b>0.77</b>

실험 1에서 F1 값의 macro avg와 weighed avg 모두 손선주[6]와 홍중선, 정연주[7]보다 본고에서 제안한 코퍼스가 더 높았다. [6]은 ‘두려움’ 범주에 대해 하나도 탐색하지 못한 것으로 나타났고, ‘기쁨’과 ‘슬픔’ 범주에 대해서는 높은 F1 점수를 기록했지만 다른 범주에 대해서는 0.13 이하의 점수를 기록했다. 이는 [6]이 ‘기쁨’, ‘슬픔’ 범주에 치우쳐 있기 때문이다. [7]은 분노, 놀람을 제외한 범주에서 [6]보다 F1 점수가 높았으나, 코퍼스의 범위가 작아 감정이 없는 것으로(NTL, Not True Label) 탐색되는 문장이 많았다. 본고에서 제안한 코퍼스는 감정 범주가 고르게 분포되어 있고 코퍼스의 범위가 넓어 F1 값의 평균이 가장 높았다.

### 5.2 긍/부정 분류

표6. 긍/부정 분류 F1 점수

감정범주	손선주	홍중선, 정연주	Ours
긍정	0.39	0.23	0.32
부정	0.3	0.23	0.32
NTL	0.82	0.83	0.83
micro avg	0.7	0.72	<b>0.72</b>
macro avg	0.51	0.43	<b>0.49</b>
weighted avg	0.68	0.66	<b>0.68</b>

실험 2에서 F1 점수의 macro avg는 [6]이 가장 높았다. 그러나 [6]은 다른 코퍼스에 비해 긍/부정 범주간 F1 점수 차이가 크게 나타나 데이터의 편향성을 다시 확인하였다. [7]과 본고에서 제안한 코퍼스는 범주별로 고른 F1 점수 분포를 보였고, 본고에서 제안한 코퍼스가 [7]에 비해 F1 점수가 0.09 높아 탐색 기능이 보완되었다.

### 6. 결론 및 향후 방향

본 연구에서는 소셜 텍스트에서의 감정 탐색을 위해, 한국어에서 감정을 표현하는 용언들을 모아 직접 감정 표현과 간접 감정 표현으로 나누어 제시하였다. 한국어 감정 용언 사전 중에서 가장 다양한 어휘를 포함하고 있으며, 표현들을 두 가지 층위로 나누어 제시했다는 점에서 의의가 있다. 본 연구는 이 감정 용언 사전을 통해 소셜 데이터의 직접 감정 표현에서 감정을 분류해보았다. 그 결과 기존 감정 어휘 사전보다 성능이 향상되었다. 다만 소셜에서 찾아낸 감정에 대해 감정의 주체를 선택하는 것과, 소셜에 사용된 간접 감정 표현이 전달하는

감정에 대해 여러 후보 중 알맞은 감정을 선택하는 것은 추후 과제로 해결되어야 할 것이다.

**참고문헌**

[1] 김은영, “국어 감정동사 연구”, 전남대학교 대학교 국어국문학과, 박사학위논문, 2004.  
 [2] 신은주, “국어 감정용언의 양상 고찰: 국어교과서를 중심으로”, 한국교원대학교 교육대학원, 국어교육학, 석사학위논문, 2008.  
 [3] 안신호, 이승혜, 권오식, “정서의 구조: 한국어 정서단어 분석”, 한국심리학회지: 사회 및 성격, 제7권, 제1호, pp. 107-123, 1993.  
 [4] 이준용, 송현주, 나은경, 김현석, “정서 단어 분류를 통한 정서의 구성 차원 및 위계적 범주에 관한 연구”, 한국언론학보, 제52권, 제1호, pp. 159-167, 2008.  
 [5] 박인조, 민경환, “한국어 감정단어의 목록 작성과 차원 탐색”, 한국심리학회지: 사회 및 성격, 제19권, 제1호, pp. 109-129, 2005.  
 [6] 손선주, 박미숙, 박지은, 손진훈, “한국어 감정표현 단어의 추출과 범주화”, 감성과학, 제15권, 제1호, pp. 105-120, 2012.  
 [7] 홍종선, 정연주, “감정동사의 범주 규정과 유형 분류”, 한국어학, 제45권, pp. 387-420, 2009.  
 [8] Ekman, P. “Universals and cultural differences in facial expressions of emotion”, Proceedings of the 1971 Nebraska Symposium on Motivation, pp.207-283, 1971.  
 [9] M. Heydarian, T. E. Doyle and R. Samavi, "MLCM: Multi-Label Confusion Matrix," in IEEE Access, vol. 10, pp. 19083-19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

**7. 부록**

표 7. 손선주 실험 1 Norm Confusion matrix

	기쁨	슬픔	분노	두려움	놀람	혐오	NPL
기쁨	66	5	1	0	0	1	26
슬픔	7	31	2	0	0	1	59
분노	5	7	9	0	1	3	76
두려움	6	6	2	0	1	2	85
놀람	3	6	1	0	6	0	85
혐오	11	9	13	0	2	12	53
NPL	7	7	2	0	1	1	83

표 8. 손선주 실험 2 Norm Confusion matrix

	긍정	부정	NPL
긍정	41	8	51
부정	7	23	71
NPL	5	7	88

표 9. 홍종선, 정연주 실험 1 Norm Confusion matrix

	기쁨	슬픔	분노	두려움	놀람	혐오	NPL
기쁨	32	2	0	0	0	1	65
슬픔	0	11	0	0	0	3	85
분노	0	1	4	1	0	5	90
두려움	0	1	0	15	0	2	82
놀람	1	0	0	2	1	1	97
혐오	3	2	0	0	0	15	80
NPL	1	1	0	1	0	2	96

표 10. 홍종선, 정연주 실험 2 Norm Confusion matrix

	긍정	부정	NPL
긍정	19	6	75
부정	2	17	81
NPL	3	4	94

표 11. Ours 실험 1 Norm Confusion Matrix

	기쁨	슬픔	분노	두려움	놀람	혐오	NPL
기쁨	35	2	0	0	1	1	62
슬픔	0	18	1	0	0	4	76
분노	0	2	7	1	0	6	83
두려움	0	2	1	15	1	3	79
놀람	1	0	0	2	29	1	68
혐오	3	3	0	0	1	19	74
NPL	1	2	0	1	0	2	94

표 12. Ours 실험 2 Norm Confusion Matrix

	긍정	부정	NPL
긍정	27	9	64
부정	3	23	74
NPL	4	6	91

제34회 한글 및 한국어 정보처리 학술대회 논문집 (2022년)

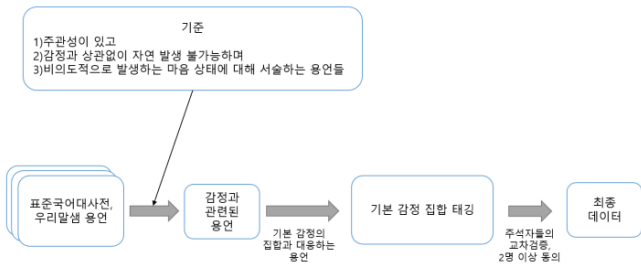


그림 1. 직접 감정 표현 태깅 과정

분노/부정	3
슬픔/분노	3
슬픔/긍정	3
놀람/슬픔	2
혐오/슬픔	2
부정/혐오	2
부정/기쁨	2
혐오/두려움	2
긍정/기쁨	2
슬픔/기쁨	1
두려움/혐오	1
두려움/긍정	1
기쁨/긍정	1
복합감정	1

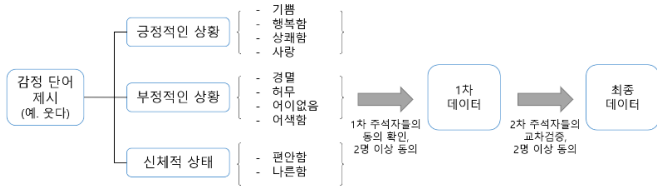


그림 2. 간접 감정 표현 태깅 과정

표 13. 직접 감정 표현 코퍼스 통계

기본 감정 집합	개수
부정	241
긍정	178
혐오	124
기쁨	92
슬픔	85
분노	65
두려움	48
슬픔/부정	28
놀람	24
놀람/부정	21
부정/슬픔	15
혐오/부정	9
긍정/부정	8
두려움/부정	6
긍정/슬픔	6
분노/혐오	5
부정/두려움	5
놀람/두려움	5
놀람/긍정	5
긍정/두려움	5
혐오/분노	4
분노/슬픔	4