인지과학, 제21권 제1호
Korean Journal of Cognitive Science
2010, Vol. 21, No. 1, 157～175

감정 온톨로지의 구축을 위한 구성요소 분석*

윤 애 선 권 혁 철†
부산대학교

의사소통에서 대화자 간 감정의 이해는 메시지의 내용만큼이나 중요하다. 비언어적 요소에 의해 감정에 관한 더 많은 정보가 전달되고 있기는 하지만, 텍스트에도 화자의 감정을 나타내는 언어적 표지가 다양하고 풍부하게 녹아들어 있다. 본 연구의 목적은 인간언어공학에 활용할 수 있는 감정 온톨로지를 구축할 수 있도록 그 구성요소를 분석하는 데 있다. 텍스트 기반 감정 처리 분야의 선행 연구가 감정을 분류하고, 각 감정의 서술적 어휘 목록을 작성하고, 이를 텍스트에서 검색함으로써, 추출된 감정의 정확도가 높지 않았다. 이에 비해, 본 연구에서 제안하는 감정 온톨로지는 다음과 같은 장점이 있다. 첫째, 감정 표현의 범주를 기술 대상과 방식에 따라 6개 범주로 분류하고, 이들 간 상호 대응관계를 설정함으로써, 멀티모달 환경에 적용할 수 있다. 둘째, 세분화된 감정을 분류할 수 있되, 감정 간 차별성을 가질 수 있도록 24개의 감정 명세를 선별하고, 더 섬세하게 감정을 분류할 수 있는 속성으로 강도와 극성을 설정하였다. 셋째, 텍스트에 나타난 감정 표현을 명시적으로 구분할 수 있도록, 경험자・기술 대상과 방식・언어적 자질에 관한 속성을 도입하였다. 넷째, 본 연구의 감정분류는 Plutchik의 분류와 호환성을 갖고 있으며, 언어적 요소의 기술에서 국제표준의 태그세트를 수용함으로써, 다국어 처리에 활용을 극대화할 수 있도록 고려했다.

주제어 : 감정 온톨로지, 텍스트 기반 감정 처리, 감정 분류, 다국어 멀티모달 환경, 애니메이션 합성

---

* 이 논문은 2007년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2009-0083761). 이 논문의 초고는 2009년 '한글 및 한국어 처리 학술대회에서 발표했던 내용을 보완한 것임.

† 교신저자: 권혁철, 부산대학교 공과대학 정보컴퓨터공학부/일반대학원 인지과학협동과정, 연구 분야: 인간언어공학
E-mail: hckwon@pusan.ac.kr

1. 서 론

의사소통 연구자들에 의하면, 인간 간 의사소통에서 메시지의 60-95%는 발화된 내용이나 글과 같은 언어적(verbal) 요소보다 얼굴 표정, 눈빛, 강세, 억양, 목소리의 톤, 동작, 자세, 옷차림 등 비언어적(non-verbal) 요소로 전달된다고 한다[1]. 특히 화자(speaker)의 감정(emotion)을 전달하는 데는 비언어적 요소의 중요성이 한층 증가한다. 마임(mime)에서 볼 수 있듯이 비언어적 요소 자체만으로도 상당히 구체적인 메시지 전달이 가능하며, (ex1)과 (ex2)의 밑줄친 부분처럼 동일한 언어적 요소에 아주 다양한 감정을 실을 수도 있고, 심지어 언어적 요소의 문자적 내용과 상반되는 메시지를 나타내기도 한다.

(ex1) (a) 희영: (볼이 잔뜩 부은 채) 내신이 어떻게 돼도 괜찮다는 건 아니야.
(b) 언니: (비아냥 거리며) 아까 고등학교 안 간댔잖아.
(c) 엄마: (눈물을 글썽이며) 뭐라구?
(d) 아빠: 자자, 진정해. 진정하라고!

(ex2) (a) 아아; 죄송해서요...;;; 그래도 다행이네요^ㅁ^;;
(b) 미안~~나중에 보게되면별다방한번대접하마ㅋㅋ안녕~~
(c) 나잠깐 친구만나서 걸을라꼬ㅋㅋㅋ-_-ㅅㅂ

더욱이 언어적 요소 중에서도 글로 전달되는 텍스트를 사용하는 경우, 그 분절적인 특성(segmental feature)에 힘입어 매우 상세하고 추상적인 메시지를 전달할 수 있지만, R. Jacobson([2])이 분류한 의사소통의 감정적/표현적 기능(emotive/expressive function)을 수행하는 데는 부족한 부분이 많다.

정보전달을 목적으로 하는 신문기사, 책, 논문 등의 텍스트와는 달리, 일상적인 대화가 오가는 SMS나 채팅 텍스트, 화자의 주관이 매우 뚜렷한 댓글 등 CMC (Computer-Mediated Communication) 환경에서 (ex2)에서 볼 수 있는 것과 같은 다양한 이모티콘(emoticon)을 개발하여 사용하는 것은 텍스트에 결여된 감정의 전달, 텍스트에 나타난 감정의 중의성 해결, 텍스트로 전달되는 문자적 내용의 부정과 같은

기능을 보완하려는 언중들의 무의식적인 노력이다. 지난 10여 년간 꾸준히 진행되어 온 MMI(Man-Machine Interaction) 환경에서도 자동차나 로봇에 인간과 같은 감정이 드러나도록 얼굴 표정을 생성하게 하거나, 인간의 감정을 분석하도록 하는 시도도 이같은 맥락에서 이해할 수 있다[3, 4].

이렇듯 비언어적 요소에 의해 감정에 관한 훨씬 더 많은 정보가 전달되고 있기는 하지만, 텍스트에도 화자의 감정을 나타내는 언어적 표지가 다양하고 풍부하게 녹아들어 있다.

표 1. 감정의 기술 대상 및 방식

대상 \ 방식	표현적	서술적	도상적
비언어적	①	②	③
언어적	④	⑤	⑥

표 1에서 볼 수 있듯이, ①이 앞에서 언급한 비언어적 요소를 통해 즐겁고, 슬프고, 놀라고, 싫은 감정 등을 직접적으로 표현(expressive)하는 방식이라면, ②-⑥은 이에 대응되는 감정을 언어적으로 실현하는 여러 변주이다. ②는 (ex3)처럼 ①을 언어적으로 서술하거나(descriptive) 기술하는 방식이고, ③은 (ex4)처럼 ①을 도상적으로(iconic) 재현하는 이모티콘 범주다. ④는 (ex5)처럼 언어적 요소를 이용하여 감정을 직접적으로 표현하는 방식이며, ⑤는 (ex6)처럼 언어적 요소로 감정을 정의하거나 서술하는 범주로서 메타 언어적인 특성을 갖는다. ⑥은 (ex7)처럼 언어적 요소의 일부를 도상적으로 사용하는 범주다.[1)]

(ex3) (a) 입이 귀에 걸리다 [기쁨]
(b) 도끼눈을 치켜 뜨다 [화남]

1) 언어적 금기(linguistic taboo)에 의해 욕설 등이 'ㅈㄹ(지랄), ㅅㅂ(시발), 18(시팔)'처럼 ⑥과 같은 방식으로 전달된다. 이때, 욕설의 본디 형태를 사용하는 경우가 화자의 화난 감정을 있는 그대로 전달하는 것과는 달리 ⑥과 같은 두자어(acronym)나 음성적 유사성을 가진 단위를 사용하는 경우의 발화 상황은 화자와 청자 간 친밀한 관계에서 감정의 강도가 훨씬 약화된다.

	(c) 눈물이 그렁그렁하다	[슬픔]
	(d) 눈길을 피하다	[싫음]
(ex4)	(a) ^^	[기쁨]
	(b) :-(	[화남]
	(c) *ㅡ_ㅡ*	[부끄러움]
	(d) *_*	[기대]
(ex5)	(a) 야, 신난다!	[기쁨]
	(b) 어머나!	[놀람]
	(c) 큰일났어!	[당황]
	(d) 짱나!	[화남]
(ex6)	(a) 기쁘다	[기쁨]
	(b) 화나다	[화남]
	(c) 슬프다	[슬픔]
	(d) 싫어하다	[싫음]
(ex7)	(a) ㅋㅋ	[기쁨]
	(b) 18	[화남]

지금까지 인간언어공학 기술에 기반하거나 이에 활용하고자 하는 텍스트 기반 감정처리 분야의 연구는 언어적이며 서술적인 기술 방식(⑤)을 주된 연구 대상으로 삼았다. 본 연구의 목적은 연구 대상을 확장하여 '텍스트에 나타난 감정 표현을 표준화된 방식으로 분석하고 태깅함으로써, 표 1의 ①-⑥ 범주처럼 이질적 감정표현 범주 간 상호 대응관계 설정할 수 있는 감정 온톨로지를 구축하기 위한 구성요소를 분석하는 데' 있다. 본 연구에서 제안하게 될 감정 온톨로지는 한국어 처리에 국한되지 않고, 다국어 처리 및 멀티모달 환경에 활용할 수 있도록 확장성을 고려한다. 2절에서는 선행연구의 현황을 살펴보면서, 감정 온톨로지를 구성할 수 있는 자질을 검토한다. 3절에서는 본 연구진이 설계하는 감정 온톨로지의 구성요소와 속성 및 그 값의 필요성 및 적용 범위를 제안한다. 4절에서는 향후 연구 계획과 활용 방안을 살펴본다.

2. 선행연구 검토

텍스트 기반 감정처리 분야의 기존 연구는 크게 두 방향에서 진행되고 있다.

하나는 이른바 의견 추출(Opinion Mining) 및 감성 분석(Sentiment Analysis)으로 특정한 주제에 대해 긍정 또는 부정적 평가 지표를 제공한다[5]. 새로운 정책, 영화, 상품, 인물 등에 대한 대중의 선호도를 실시간으로 측정하는 데 활발히 사용되고 있다. 이때 사용하는 감정의 분류는 2-3가지이며, 따라서 감정의 입자(grain size)가 매우 크고 분류가 거칠다. 이를 위한 언어자원으로는 개별언어의 내용어(content word)를 대상으로 어휘 또는 어의 수준에서 중립/긍정/부정 평가를 태깅한 것을 사용하며, 가장 대표적인 것은 그림 1처럼 영어 WordNet 2.0의 어의(word sense)에 3가지 값을 태깅한 SentiWordNet이다[6].[2] 이를 기반으로 핵심어 검색(Keyword spotting) 방식이 사용되었으나 정확성이 떨어지는 점이 문제로 지적되었고, 이를 개선하기 위해 문장 분석 정보나 감정의 다른 자질을 이용하는 방식이 활발히 연구되고 있다[7, 8, 9, 10].

다른 하나는 세분화된 감정 분류(fine-grained classification of emotion)에 기반한다. 이 방향의 연구는 Plutchik[11, 12, 13], Russell[14], Ekman[15] 등 심리학을 중심으로 한 인문사회 분야의 성과에 힘입은 바 크다. 특히 초기 연구가 [14, 15]에 기반하여 소위 기본 감정을 소수(4개-8개)의 범주로 구분하여 이를 인식하거나 표현하고자 하고,[3] MMI와 CMC에서 효용성을 확인하였다. 미 MIT대학 Media Lab의 감정

---

2) 영어 WordNet은 어휘 또는 다어절 표현을 어의별로 구분하고, 동일한 의미를 가진 어의를 동의어 집합인 신셋(synonym set, 약칭 synset)으로 묶은 후, 신셋을 단위로 상하위어, 유사어, 반의어, 함의 등의 다양한 의미 관계를 설정한 어휘의미망이다. 따라서 그림 2의 'heavy'라는 어휘는 명사로 2개 어의와 형용사로 5개 어의를 가지며, 각 어의의 긍정/부정/중립에 해당하는 값은 상이하게 나타날 수 있다. SentiWordNet의 경우, 소수의 핵심 신셋만 사람이 감정 값을 부여하고, 나머지에 대해서는 자동으로 값을 부여했기 때문에, 후자의 경우 감정값의 신뢰도가 많이 떨어지고, 동일 신셋에 나타나는 어의라도 다른 감정값을 갖는 경우에 적절한 감정값 부여가 이루어지지 않았다는 단점이 있다.

3) 기본 감정에 대한 명세는 동서양을 막론하고 수 천 년 전부터 많은 연구자들에 의해 '喜, 樂, 怒, 哀, 悲, 懼, 憂, 愛, 好, 惡, 憎, 慾'이나 'Joy, Happiness, Anger, Rage, Sadness, Grief, Fear, Surprise, Acceptance, Disgust, Anticipation, Desire, Love, Admiration'처럼 제안되었다.

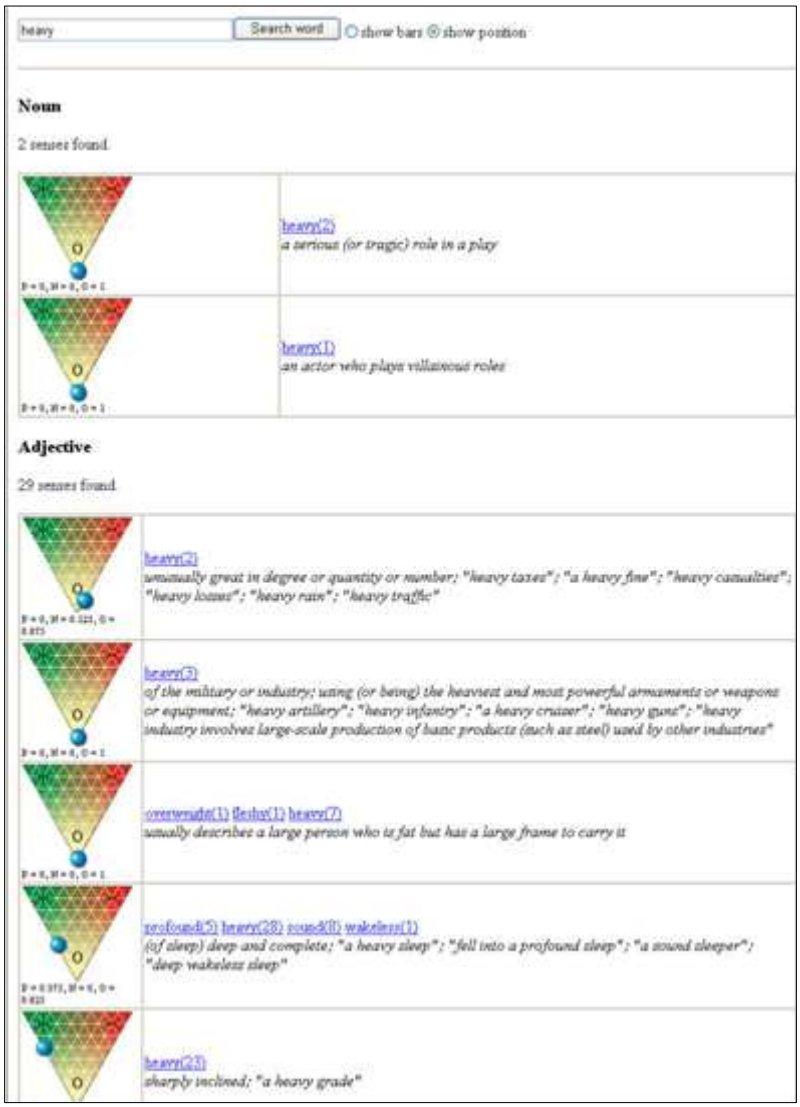

그림 2. SentiWordNet이 제공하는 어의별 감정 값: 'heavy'의 예

표현 자동차 등을 비롯한 일련의 감성 연구[3]와 국내에는 심리학 분야에서 수행한 얼굴 표정 자동생성에 관한 연구가 있으며[16, 17], 텍스트 기반 감정 처리 분

야로는 9개-12개 정도의 감정을 분류한 연구가 수행되었다[18, 19].

인간 간 의사소통에서 감정을 전달하는 데 비언어적 표현의 장점은 언어적 표현을 훨씬 능가하는 세심함과 미묘함이다. 하지만, 감정처리 초기 연구에 사용되었던 감정 범주는 지나치게 크고 거칠었으므로, 감정 분류를 더욱 세분화할 필요성이 대두하였다.

이러한 필요에 부응하여 세분화된 감정 분류가 다양하게 제안되어 감정 처리 분야에 응용되었다[20, 21, 22], 이중 최근 10년간 세분화된 감정 구분의 모델이 된 것은 R. Plutchik과 동료들의 분류이고[11, 12, 13], 이를 도식화한 것이 그림 3의 '플러칙의 감정 팽이(Plutchik's Wheel of Emotions)'이다. 예를 들어, 기본 감정 'joy'는 'sadness'와 대립쌍을 이루며, 'anticipation'과 'trust'에 인접해 있다. 인접 감정인 'joy+anticipation', 'joy+trust'의 결합은 각각 ''optimism'과 'love'라는 조합 감정을 구성한다. 또한 감정의 강도에 따라 'ecstasy(강)-joy(중)-serenity(약)'으로 구분한다.

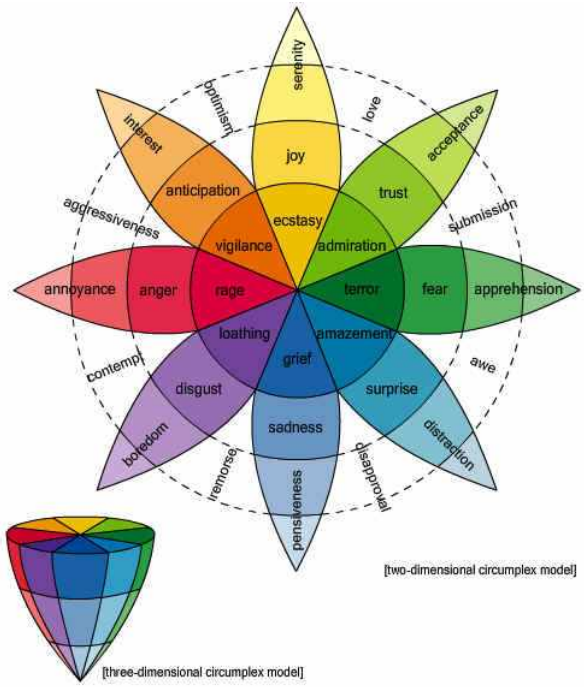

그림 3. Plutchik's Wheel of Emotions

플러칙의 감정 팽이가 ① 인간의 감정을 8개의 기본 감정군으로 분류하고, ② 이것을 4개의 대립 쌍으로 설정하며, ③ 강도에 따른 변화형을 제시하고, ④ 조합 감정을 기본 감정으로 환원할 수 있다. 이러한 분석적인 설명은 감정의 공학적 처리를 가능성을 높이는 데 크게 기여하였다. 특히 다양한 감정을 기본 감정으로 환원할 수 있다는 가능성은 그림 4와 같은 2차/3차 조합 감정과 서로 상반되는 감정의 목록을 제시한다[23].

한국어를 대상으로는 [11, 12, 13]을 이용한 텍스트 기반 감정 처리는 없었고, Pennibecker[20]의 분류를 토대로 한 심리상태에 관한 연구가 수행되기도 했다[24].

3. 감정 온톨로지 설계를 위한 자질 분석

플러칙의 감정 팽이와 이를 활용한 감정 분류는 감정 온톨로지를 구성하는 데 중요한 좌표를 제공할 수 있다. 하지만 본 연구에서 목적으로 다양한 기술 대상과 방식 간 대응 관계를 설정할 수 있는 감정 온톨로지를 구축하기 위해서는 다음 두

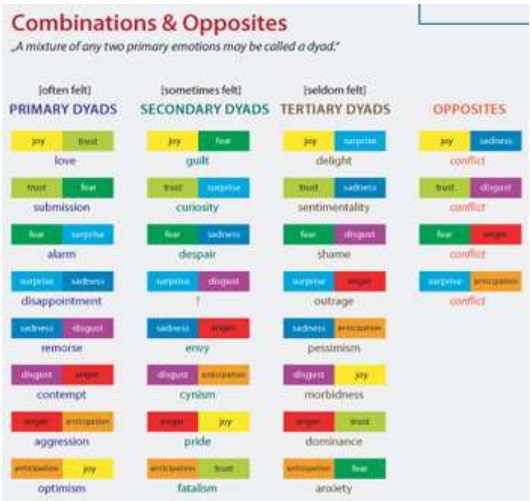

그림 4. Plutchik's Combinations of Emotions proposed by Markus Drews

측면에서 보완하거나 더욱 심도 있게 연구해야 할 부분이 있다.

도식적인 측면이 지나치게 강하여 재검토가 필요한 그림 3과 그림 4의 감정 분류 및 감정 명세는 3.1에서 살펴본다. 다음, 텍스트 기반 감정 처리에 활용할 수 있는 감정 온톨로지를 구축하려면, 언어적 요소와 관련된 특성에 관한 연구가 필요하다. 이것은 3.2에서 상세히 논의한다.

3.1 감정의 분류

감정 온톨로지 구축에서 가장 선행해야 할 것은 세분화된 감정의 종류를 구분하고 이를 메타용어로 명세화하는 것이다.

3.1.1 감정 명세

다양하고 많은 수의 명세는 감정을 세분화하여 기술하는 데 도움이 될 수 있으나, 자칫하면 메타용어가 특정한 개별 언어의 감정 특성에 경도될 수 있으며, 인접 감정과의 구분이 어렵다. 본 연구의 기반이 되는 [11, 12, 13](이하, Plutchik)의 감정 명세는 다음과 같은 문제점이 발견된다.

첫째, 기본 감정의 강도 변화형으로 제시된 'joy[중] – ecstasy[강], disgust[중] – boredom[약]'의 관계는 'trust[중] – admiration[강], trust[중] – acceptance[약]' 등의 관계와는 달리 더 큰 강도 조정이 필요하다.

둘째, 그림 4에서 볼 수 있는 1차/2차/3차 조합 감정 중 다수는 기본 감정의 강도 변화형으로 정의할 수 있다. 예를 들어 'delight'나 'outrage' 등은 'joy'나 'anger'의 강도 변화형으로 기술하는 것이 더 타당하다.

셋째, 감정 명세 중 잘못된 부분이나 감정 명세를 설정할 수 없는 부분이 있다. 예를 들어, 'fatalism'을 'anticipation + trust'의 결합으로 제안하지만 이러한 감정의 조합은 오히려 'gratefulness'에 상응한다.

따라서 본 연구에서는 Plutchik의 연구를 수용하되, 다음과 같은 기준으로 표 2처럼 24개 감정 명세를 선별하였다.

1) Plutchik이 제안한 8개 기본 감정의 명세를 수용한다. 단, 메타용어의 선택에

표 2. 본 연구가 제안하는 감정 명세

	감정명	대립 감정명
기본감정	JOY_HAPPYNESS	SADNESS
	ANGER	FEAR
	SURPRISE	DESIRE_ANTICIPATION
	TRUST	DISGUST
조합감정	LOVE	REMORSE
	DISAPPOINTMENT	PRIDE
	CONTEMPT	SHAME
	AGGRESSION	FATALISM*
	OPTIMISM	PESSIMISM
	GUILT	GRATEFULNESS**
	CURIOSITY	CYNISM
	ENVY	SENTIMENTALITY

서 Plutchik의 용어 외에 광범위하게 통용되는 등가적 용어는 병기한다. 예를 들어, 많은 연구에서 기본 감정으로 분류되는 'happy'나 'desire'가 Plutchik의 감정 명세에는 없다. 본 연구에서는 이를 병기하여 <JOY-HAPPINESS>, <DESIRE_ANTICIPATION>로 사용하거나, 2개 이상 메타용어 간의 등가성을 미리 정의한다.

2) Plutchik이 제안한 기본 감정의 강도 변화형은 메타용어를 따로 설정하지 않고, 기본 감정과 강도 및 극성으로 구분한다. 예를 들어, 기본 감정의 강도 변화형인 'acceptance'는 <TRUST>보다 강도가 낮은 감정으로 정의한다.

3) 1차/2차/3차 조합 감정 명세 24개 중 [20, 21, 22] 등의 연구와의 비교와 대립쌍 개념을 바탕으로 14개를 선택한다. 이때 선택되지 않은 명세는 선별된 기본 감정이나 조합 감정 명세를 이용하여 정의할 수 있어야 한다. 예를 들어 'submission, delight, outrage'는 각각 <TRUST>, <JOY>, <ANGER>의 속성값을 달리하여 기술할 수 있다.

4) 감정 명세는 '감정의 경험자(experiencer) = 화자' 원칙에서 선별한다. 따라서

‘alarm’, ‘blame’ 등 특정한 감정을 느끼는 화자의 행위가 청자에게 특정한 감정을 유발하는 것을 목적으로 하는 ‘발화 행위(speech act)’는 감정 명세에서 배제한다. ‘발화 행위’는 경험자 속성을 도입함으로써 본 연구에서 일관된 방식으로 기술할 수 있다.

5) Plutchik이 제안한 조합 감정 중 ‘fatalism’은 본 연구에서 <AGGRESSION>의 대립쌍으로 재정의하였으며, <GUILT>의 대립적 자질을 갖는 감정 명세로 <GRATEFULNESS>를 추가한다. 표 2에서 두 감정명은 별표(*)로 표시하였다.

6) 다국어 처리에 사용할 수 있도록 본 연구의 감정 분류 메타용어는 < > 안에 대문자로 된 영어의 명사형을 사용한다.

3.1.2 강도(intensity)와 극성(polarity)

감정 분류를 세분화하기 위해 본 연구에서는 강도와 극성이라는 2가지 속성을 도입한다.

감정의 강도는 Plutchik에서도 제안된 속성이다. 다만 Plutchik은 강도 변화를 강/중/약으로 구분하고, 8개의 기본 감정의 변이형 간 강도의 차이 두지 않았으며, 조합 감정에 대해 강도 변이형을 제시하지 않았다. 본 연구에서는 표 2에 소개한 감정의 강도를 ‘0’으로 설정하고, 기본 감정 8개에 대해서는 ‘+2～-2’의 값을 가질

표 3. Plutchik과 본 연구에서 제안하는 분류와의 비교 1: 기본 감정의 강도 변화형 (발췌)

Plutchik			본 연구		
기본 감정	‘중’ 감정	강도	기본 감정	강도	극성
serenity	joy	약	<JOY_HAPPINESS>	-1	Pos
rage	anger	강	<ANGER>	+1	Neg
grief	sadness	강	<SADNESS>	+1	Neg
terror	fear	강	<FEAR>	+1	Neg
amazement	surprise	강	<SURPRISE>	+1	Pos
ecstasy	joy	강	<JOY_HAPPINESS>	+2	Pos
boredom	disgust	약	<DISGUST>	-2	Neg

표 4. Plutchik과 본 연구에서 제안하는 분류와의 비교 2: 조합 감정 (발췌)

Plutchik			본 연구		
조합 감정	기본감정 1	기본감정 2	기본/조합 감정	강도	극성
submission	trust	fear	<TRUST>	+2	Neg
alarm	fear	surprise	<SURPRISE>	+2	Neg
delight	joy	surprise	<JOY_HAPPINESS>	+1	Pos
outrage	surprise	anger	<ANGER>	+2	Neg
despair	sadness	fear	<DISAPPOINTMENT>	+1	Neg
morbidness	disgust	joy	<LOVE>	+1	Neg
dominance	anger	trust	<PRIDE>	+1	Neg

수 있도록 변화형의 폭을 넓혔으며, 조합 감정에 대해서도 '+1~-1'의 강도 변이형을 허용하였다.

또한, 같은 강도를 갖는 동일한 감정이라도 긍정/부정적 평가에 따라 구분되어야 하는 경우가 많다. 예를 들어, '부러움'과 '질투'는 모두 <ENVY Intensity="0">이나, 전자는 긍정적, 후자는 부정적인 평가를 나타낸다. 표 3과 표 4는 Plutchik의 분류와 본 연구에서 제안하는 분류를 비교한 것이다. 표 3에서 'ecstasy'와 'boredom'의 강도 변화를 볼 수 있고, 표 4에서는 각기 다른 명칭를 부여했던 조합 감정 중 상당한 수를 표 2에서 제시한 감정명과 강도 및 극성 구분으로 기술할 수 있음을 보여준다.

3.2 텍스트 기반 감정 처리를 위한 언어적 속성

기존의 텍스트 기반 감정 처리 연구에서는 감정을 분류한 후, 각 감정에 해당하는 서술적([표 1]의 범주 ⑤) 어휘 목록을 작성하고 이것을 이용하는 데 그쳤다. 하지만 텍스트 기반 감정 처리를 하기 위해 고려해야 할 언어적 속성은 범위가 훨씬 더 다양하다.

첫째, 감정의 경험자를 구분해서 기술할 수 있어야 한다. 둘째, 표 1의 범주 ②-⑥의 특성에 따라, 텍스트에 나타난 감정 표현을 명시적으로 구분해야 한다. 셋째,

표 1의 ②-⑥ 범주로 실현되는 감정 표현의 언어적인 자질을 기술하여, 감정 온톨로지의 활용도를 높여야 한다.

3.2.1 경험자

감정을 겪고 이를 기술할 수 있는 경험자는 기본적으로 화자이지만, 텍스트를 통해 기술된 감정 표현은 청자나 제3자와 관련된 경우가 많다. 단순한 핵심어 검색이 감정처리의 정확도를 낮추는 이유 중 하나이고, 따라서 구문 분석을 필요로 한다.

(ex8)	(a) 내가 영희를 사랑해.	<LOVE Experiencer="Speaker">
	(b) 네가 영희를 사랑하는 거냐?	<LOVE Experiencer="Hearer">
	(c) 철수가 영희를 사랑해.	<LOVE Experiencer="Third">
(ex9)	(a) 아이, 창피해.	<SHAME Experiencer="Speaker">
	(b) 창피한 줄 알아라.	<SHAME Experiencer="Hearer">

(ex8-a)에서 사랑하는 감정(<LOVE>)의 경험자는 화자이나, (ex8-b, c)에서 사랑하는 감정의 경험자는 각각 청자와 철수다. (ex8-b, c)에서 화자의 감정은 슬플 수도, 화가 날 수도, 의기소침할 수도 있다. 또한, (ex9-a)에서 창피한 감정(<SHAME>)의 경험자는 화자이지만, (ex9-b)라는 '비난(blame)'에서 창피한 감정의 잠재적 경험자는 청자이다. 이렇게 경험자 속성을 설정하고 이를 명시하는 것이 화자와 청자 간의 감정 전이가 일어나는 비난과 같은 발화 행위를 더 효과적으로 기술할 수 있다.

3.2.2 기술 대상과 방식

감정은 언어적(verbal)으로 나타낼 수도, 음성이나 신체 표현처럼 비언어적(non-verbal)으로 나타낼 수도 있다. 텍스트에 나타난 감정 표현은 이 두 종류를 기술 대상으로 하며, 각각 3가지 기술 방식으로 나타낼 수 있다.

첫째, 표현적 방식은 언어적/비언어적 표현이 특정한 감정을 직접적으로 전달한다. '내가 현재 여기에서'로 규정되는 발화상황('I-Now-Here' enunciation)에서 '감정의 경험자=화자'의 등식 관계가 항상 성립한다. 언어적인 요소로는 감탄사가 대표

적이고, 비언어적 요소는 음성이나 신체표현 자체가 표현적 방식에 해당한다.

둘째, 서술적 방식은 특정한 감정을 나타내는 언어적/비언어적 표현을 서술하거나, 메타적으로 정의하는 것이다. 이때, 발화상황에서 '감정의 경험자=화자'의 등식이 언제나 성립하는 것은 아니다. 언어적 기술 대상의 경우, 특정 감정을 정의하거나 설명하는 용어 또는 표현을 들 수 있다. 비언어적 요소의 경우, 해당 감정을 나타내는 음성/동작/얼굴 표정/자세 등의 특성을 설명하는 표현을 들 수 있다.

셋째, 도상적 기술 방식인데, '감정의 경험자=화자'라는 등식이 적용된다. 대표적인 예는 비언어적 요소를 표상한 이모티콘의 사용이다. 언어적 요소를 표상화하는 경우(표 5의 ⑥)는 제한적으로 사용되는데, 대화체에서 매우 빈번한 감정 명세인 <JOY_HAPPINESS>를 표현하는 'ㅋㅋ', 'ㅎㅎ'이나, 언어적 금기(linguistic taboo) 때문에 사용하는 'ㅅㅂ', 'ㅈㄹ' 등이 대표적이다.

표 5는 실제 텍스트에서 발견할 수 있는 <SURPRISE>에 해당하는 예이다.[4)]

표 5. 감정 표현의 기술 대상 및 방식: <SURPRISE>의 예

대상 \ 방식	표현적	서술적	도상적
비언어적	① 음성 (억양, 강세) | 신체 표현 (얼굴 표정, 동작, 눈빛, 자세)	② '움찔하다' | '눈이 휘둥그레지다' | '목소리가 떨리다'	③ !!(O O)!! | ~(⊙ㅅ⊙~)
언어적	④ '허걱' | '놀래라!' | '깜짝이야!'	⑤ '놀라다' | '깜짝 놀라다' | '화들짝 놀라다'	⑥ (cf. 'ㅋㅋ' | 'ㅎㅎ')

3.2.3 언어적 자질

선행 연구에서 주목한 표 1 및 표 5의 범주 ⑤에 해당하는 어휘가 대화체 텍스트에서 화자의 감정을 표현하는 경우는 드물고, 오히려 범주 ④에 해당하는 표현이 더 다양하고 사용 빈도도 높다.

4) 다만 ⑥의 경우는 <SURPRISE>의 실제 예를 찾지 못해 <JOY_HAPPINESS>의 예를 제시하였다.

(ex10) 놀라다 | 깜짝 놀라다 | 화들짝 놀라다 | 질색하다 | 대경질색하다

(ex11) (a) 놀랐어(요)!| 놀래라! | 놀랍군! | 놀라운데(요)!

(b) 깜짝 놀랐어(요)!| *깜짝 {놀래라! | 놀랍군! | 놀라운데(요)!}

(c) *화들짝 {놀랬어(요)! | 놀래라! | 놀랍군! | 놀라운데(요)!}

(d) 질색이야!| † 질색이군! | 질색인데(요)!

(e) *대경질색이야!| *대경질색이군! | *대경질색인데(요)!

(ex12) 깜짝이야! | 어머나! | 진짜? | 정말(로)? | 웬일이니? | 뭐라고? | 말도 안 돼!

예를 들어, '놀람'을 나타내는 범주 ⑤에 해당하는 어휘가 (ex10)처럼 제시될 수 있다. 하지만 이 어휘의 모든 활용형이 범주 ④처럼 사용되는 것은 아니다. (ex11-a)처럼 '놀라다'의 일부 종결어미 활용형이 화자의 감정을 전달할 수 있는 반면, (ex11-b, c, d, e)에서 '깜짝 놀라다 | 화들짝 놀라다 | 질색이다 | 대경질색이다'의 경우는 '깜짝 놀랬어(요)!'나 '질색이야!' 등 (ex11-a)의 어미 중에서도 매우 제한적인 범위만이 화자의 감정을 표현하는 데 사용될 수 있으며, 다른 활용형은 비문이거나 서술적으로만 사용된다.[5] 또한 (ex11)의 정문 빈도는 높지 않은 데 비해, (ex12)와 같이 (반)고정 표현(semi-frozen expression)과 다어절(multi-word) 표현이 빈번하다. 따라서 품사, 기본형, 활용형, 공기 어휘 정보 등 언어적 자질을 기술하여, 텍스트에서 분석과 생성의 효율성을 높여야 한다. 이를 기술하고 동시에 다국어 처리의 효율성을 기하기 위해, ISO와 TEI의 국제표준 규격인 ISO 24610-1(2006)[25]에서 제안한 태그세트를 도입하였다.[6]

5) (ex11-d)의 '질색이군!'은 경험자=화자가 아닌 경우 가능한 발화체다. 이를 표시하기 위해 본고에서 임의적으로 '† '를 사용한다.

6) ISO 24610-1(2006)은 기술표준원에 의해 한국산업규격으로 채택되었으며, 2007년 이 국제표준이 한국어로 번역되었다.

4. 향후연구와 활용방안

이상에서와 간이 본 연구에서는 텍스트를 기반으로 한 감정 온톨로지를 구축할 수 있도록 구성요소를 분석하였다. 텍스트 기반 감정 처리 분야의 선행 연구가 감정을 분류하고, 각 감정의 서술적 어휘 목록을 작성하고, 이를 텍스트에서 검색함으로써, 추출된 감정의 정확도가 높지 않았다.

이에 비해, 본 연구에서 제안하는 감정 온톨로지는 다음과 같은 장점을 갖는다.

첫째, 감정 표현의 범주를 기술 대상과 방식에 따라 6개로 분류하고 전체를 연구 대상으로 확장함으로써, 이질적 특성을 갖는 6개 범주 간 상호 대응관계를 설정하였다. 이러한 특질은 멀티모달 환경에 적용할 수 있는 활용도를 높인다.

둘째, 세분화된 감정을 분류할 수 있되, 감정 간 차별성을 가질 수 있도록 24개의 감정 명세를 설정하고, 더 섬세하게 감정을 분류할 수 있는 속성으로 강도와 극성을 설정하였다.

셋째, 텍스트에 나타난 감정 표현을 명시적으로 구분할 수 있도록, 경험자・기술 대상과 방식・언어적 자질에 관한 속성을 도입하였다.

넷째, 본 연구의 감정분류는 Plutchik의 분류와 호환성을 갖고 있으며, 언어적 요소의 기술에서 국제표준의 태그세트를 수용함으로써, 다국어 처리에 활용을 극대화할 수 있도록 고려했다.

본 연구에서 제안한 감정 온톨로지 설계를 기반으로, 한국어의 서술형 텍스트와 대화형 텍스트를 분석 및 태깅하여 감정표현 라이브러리를 구축하고 있다. 다른 연구팀에서 본 연구 결과를 도입하여, 한국어 TTA(Text-To-Animation) 시스템을 개발하고 있으며, 한국어에서 성공하는 경우, 영어를 대상으로 한 감정표현 라이브러리 구축과 TTA 시스템을 개발할 계획이다.

참고문헌

[1] Birdwhiste, R. (1972). “Paralanguage twenty-five years after Sapir.” *Communication in face to face Interaction*, ed. by J. Laver/S. Hutchenson, Penquin.

[2] Jakobson, R. (1963). *Essais de linguistique générale*, Les Éditions de Minuit.

[3] Affective Computing(http://www.media.mit.edu/ research/groups/affective-computing).

[4] Global Robot Academia at Waseda Univsersity(http://www.rt-gcoe.waseda.ac.jp/index.html).

[5] Liu, B. (2006). "Opinion Mining", ch. 11, in *Web Data Mining*, Springer.

[6] Esuli, A., & Sebastiani, F. (2006). "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining", *Proceedings of the 5th Conference on Language Resources and Evaluation(LREC'06),* 417-422.

[7] Al Masum, S. M., Prendinger, H. & Mitsuru, I. (2007). "SenseNet: A Linguistic Tool to Visualize Numerical-Valance Based Sentiment of Textual Data", *ICON-2007*, 147-152.

[8] Osherenko, A. (2008). "Lexical Affect Sensing: Word Spotting Revisited", Emotional Computer Systems and Interfaces. *Proceedings of The 3rd International Conference on Cognitive Science*(http://www.informatik.uni-augsburg.de/~osherenk/ Moskau_camera_ready.pdf).

[9] Boucouvalas, A. C., & Zhe, X. (2002). "Text-to-emotion engine for real time internet communication," *Proceedings. of the Int. Symposium on CSNDSP2002*, 164-1682.

[10] 황재원, 고영중 (2009). "문장 감정 강도를 반영한 개선된 자질 가중치 기법 기반의 문서 감정 분류 시스템", **정보과학회논문지: 소프트웨어 및 응용 제36권 제6호,** 491-497.

[11] Plutchik, R., & Kellerman, H. (1980). *EMOTION Theory, Reserch, and Experience*, Academic press.

[12] Plutchik, R. & Conte, H. R. (1997). "Circumplex models of personality and emotions", *American psychological association.* 89, 344-350.

[13] Plutchik, R. (2001). "The Nature of Emotions", *American Scientist*, 89, 344-350.

[14] Russell, J. A., Fernabdez-Dols, J. M., Mastead, A. S. R. & Wellenkamp, J.C. (1995). *Everyday Conceptions of Emotion: An introduction to the Psychology, Anthropology and Linguistics of Emotion,* Kluwer Academic Publishers.

[15] Ekman, P. (2002). *Emotions and Life: Perspectives form Psychology, Biology, and Evolution*, American Psychological Association.

[16] 반세범, 정찬섭 (2001). “얼굴 특징점의 지각적 위계구조에 기초한 표정인식 신경망 모형”, **인지과학, 12**(1, 2), 77-89.

[17] 박수진, 한재현, 정찬섭 (2002). “얼굴의 물리적 특징 분석과 얼굴 관련 감성어휘 분석”, **인지과학, 13**(3), 1-10.

[18] 문현구, 장병탁 (2001). “HMM을 이용한 채팅 텍스트로부터의 화자 감정상태 분석,” **한국정보과학회 추계학술발표논문집, 28**(2), 127-129.

[19] 임성수, 조성배 (2007). “Computing with Words 기반 SMS로부터의 사용자 감정 추론,” **한국정보과학회 한국컴퓨터종합학술대회 논문집, 34**(1), 69-70.

[20] Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves, *Annual Review of Psychology*, 54, 547-577.

[21] Subasic, P. & Huettner, A. (2001). “Affect analysis of text using fuzzy semantic typing,” *IEEE Transaction on Fuzzy System*, vol. 9, 483-496.

[22] Strapparava, C., & Valitutti, A. (2004). “WordNet-Affect: an Affective Extension of WordNet”, *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 1083-1086.

[23] http://incom.org/code/projekte/projekt_anzeigen.php?4, 483, 17,0,0,504

[24] 이창환, 심정미, 윤애선 (2005). “언어적 특성을 이용한 ‘심리학적 한국어 글분석 프로그램(KLIWC)’ 개발 과정에 대한 고찰”, **인지과학, 16**(2), 93-121.

[25] ISO 24610-1. (2006). *Language resources management: Feature Structure, Part 1: Feature structure representation*, (http://www.iso.org/iso/store.htm).

1 차원고접수 : 2009. 10. 22
2 차원고접수 : 2009. 12. 16
최종게재승인 : 2010. 1. 4

(*Abstract*)

Component Analysis for Constructing an Emotion Ontology

Aesun Yoon Hyuk-Chul Kwon
Pusan National University

Understanding dialogue participant's emotion is important as well as decoding the explicit message in human communication. It is well known that non-verbal elements are more suitable for conveying speaker's emotions than verbal elements. Written texts, however, contain a variety of linguistic units that express emotions. This study aims at analyzing components for constructing an emotion ontology, that provides us with numerous applications in Human Language Technology. A majority of the previous work in text-based emotion processing focused on the classification of emotions, the construction of a dictionary describing emotion, and the retrieval of those lexica in texts through keyword spotting and/or syntactic parsing techniques. The retrieved or computed emotions based on that process did not show good results in terms of accuracy. Thus, more sophisticate components analysis is proposed and the linguistic factors are introduced in this study. (1) 5 linguistic types of emotion expressions are differentiated in terms of target (verbal/non-verbal) and the method (expressive/descriptive/iconic). The correlations among them as well as their correlation with the non-verbal expressive type are also determined. This characteristic is expected to guarantees more adaptability to our ontology in multi-modal environments. (2) As emotion-related components, this study proposes 24 emotion types, the 5-scale intensity (-2～+2), and the 3-scale polarity (positive/negative/neutral) which can describe a variety of emotions in more detail and in standardized way. (3) We introduce verbal expression-related components, such as ‘experiencer’, ‘description target’, ‘description method’ and ‘linguistic features’, which can classify and tag appropriately verbal expressions of emotions. (4) Adopting the linguistic tag sets proposed by ISO and TEI and providing the mapping table between our classification of emotions and Plutchik's, our ontology can be easily employed for multilingual processing.

Keywords : Emotion ontology, HLT applications, Plutchik's wheel of emotion, Multilingual processing, Multimodal environment, Text-To-Animation