

예제 기반 대화 시스템을 위한 양태 분류*

김민정(고려대), 홍금원(고려대), 송영인(고려대), 이연수(고려대),
이도길(고려대), 임해창(고려대)

<차 례>

- | | |
|----------------------------|--------------------------|
| 1. 서론 | 4. 규칙 기반 양태 분류 |
| 2. 관련 연구 | 4.1. 규칙 추출 및 실험에 사용한 말뭉치 |
| 2.1. 양태 범주 및 분류 체계를 다룬 연구 | 4.2. 규칙 기반 양태 분류 시스템 |
| 2.2. 대화체 생성을 위해 양태를 이용한 연구 | 4.3. 양태 분류 규칙 |
| 3. 대화 시스템을 위한 양태 분류 체계 | 5. 실험 및 오류 분석 |
| | 6. 결론 및 향후 연구 |

<Abstract>

Modality Classification for an Example-Based Dialogue System

Min-Jeong Kim, Gumwon Hong, Young-In Song, Yeon-Soo Lee,
Do-Gil Lee, Hae-Chang Rim

An example-based dialogue system tries to utilize many <user utterance - system response> pairs which are stored in a dialogue database. The most important part of the example-based dialogue system is to find the most similar utterance to user's input utterance. Modality, which is characterized as conveying the speaker's involvement in the propositional content of a given utterance, is one of the core sentence features. For example, the sentence "I want to go to school." has a modality of hope.

In this paper, we have proposed a modality classification system which can predict sentence modality in order to improve the performance of example-based dialogue systems. We also define a modality tag set for a dialogue system, and validate this tag set using a rule-based modality classification system. Experimental results show that our modality tag set and modality classification system improve the performance of an example-based dialogue system.

* Keywords: Modality classification, Dialogue system.

* 이 연구에 참여한 연구자의 일부는 '2단계 BK21사업'의 지원비를 받았음.

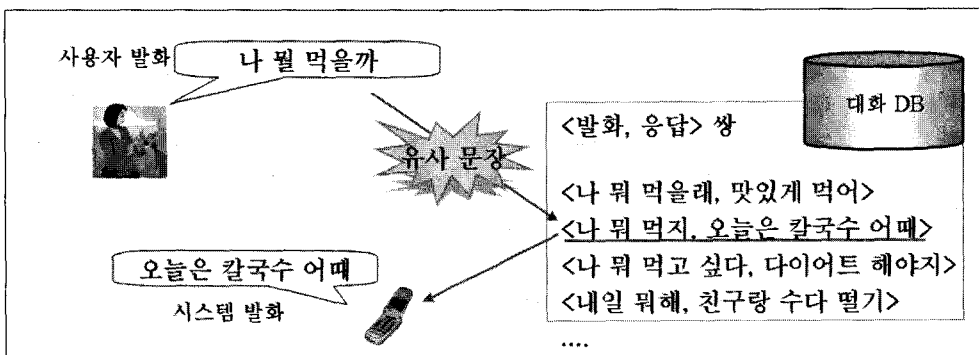
1. 서론

대화 시스템이란 사람의 발화를 분석하고 그에 적합한 응답을 생성하여 자연어로 사람과 대화할 수 있는 시스템이다. 대화 시스템의 응답 생성을 위해 여러 가지 방법을 사용할 수 있는데, 예제 기반 응답 생성 방법은 자연스러운 응답 생성을 위해 미리 구축된 <사용자 입력 발화, 시스템 응답> 쌍 데이터베이스(DB)에서 사용자 발화와 가장 유사한 발화를 검색하여 해당 발화의 응답을 시스템 응답으로 사용하는 것이다. 예제 기반 응답 생성 방법은 실제 대화의 응답을 그대로 사용함으로써 자연스러운 응답을 할 수 있다는 장점이 있다. <그림 1>은 이러한 예제 기반 대화 시스템에서 이루어질 수 있는 대화의 예이다[1].

예제 기반 대화 시스템에서 가장 중요한 것은 사용자의 입력 발화와 대화 DB 내에서 가장 유사한 발화를 찾는 것이다. 예를 들어, <그림 1>에서 “나 뭘 먹을까”라는 문장이 사용자 입력으로 들어왔을 때, DB에 있는 문장 중 완벽히 일치하지 않더라도 가장 유사한 문장인 “나 뭐 먹지”를 찾는 것이다. 시스템이 유사 발화를 잘못 찾아냈을 경우에는 대화가 부자연스럽게 된다. 예를 들어, 시스템이 “나 밥 맛있게 먹었다.”가 가장 유사한 발화라고 찾아냈을 경우 “응 나도 막 점심 먹었어.”와 같은 응답을 내보내게 될 것이다.

- (1) 먹고 싶다.
- (2) 먹지 그래?

대화 시스템에서는 위 예문 (1)과 예문 (2)가 사용자 입력으로 들어왔을 때 다른 답을 할 수 있어야 한다. 예문 (1)의 경우 “점심때가 다 되었네. 맛있게 먹어.”나 “늦은 밤이니까 참아.”와 같은 답변이 적합할 것이고, 예문 (2)의 경우 “혼자 먹기 싫어서 조금 더 있을래.”, “그렇까?”와 같은 답변이 적합할 것이다. 이렇게 다른 답변을 하기 위해서는 두 문장의 차이를 구분 지을 수 있는 것이 필요하며,



<그림 1> 예제 기반 대화 시스템에서의 유사 발화 검색

이는 본용언만을 통해서 알 수 없는 발화의 어감에 대한 인식이 필요함을 의미한다. 대화체에서는 뉴스나 기사와 같은 텍스트와는 달리 정보전달 이외의 목적을 가지는 문장들이 많이 나타난다. 이러한 정보나 사실과 같은 명제 이외의 문장을 구성하는 요소를 ‘발화에 나타난 화자의 의도’라고 정의한 것이 화행이며[2], 문장에 나타난 ‘화자의 심리적인 태도’라고 정의한 것이 양태이다[3].

(3) 우리 같이 영화 보러 갈까?

(4) 그 정도는 해도 돼.

예문 (3)의 경우 양태로 해석하면 화자가 청자와 함께 영화를 보러 가고 싶다는 의지를 나타내는 것이므로 ‘의지’ 양태가 된다. 하지만 이것을 화행으로 해석하면 화자가 청자와 함께 영화를 보러 가고 싶다는 것과 함께 청자의 뜻을 물어보는 ‘제안’ 화행이 된다. 예문 (4)의 경우 양태로 해석하면 문장의 주어인 청자가 어떠한 일을 해도 된다는 ‘능력’ 양태가 되며, 화행으로 해석하면 화자가 청자에게 어떠한 일을 허락한다는 ‘허락’ 화행이 된다. 이와 같이 양태는 해당 문장 내에 나타난 의미만을 담고 있으나 화행은 문장 내에 나타난 의미에 덧붙여 발화를 청자에게 전달함으로써 청자에게 어떠한 변화를 줄 수 있는 화자의 의도를 담고 있다.

기존의 음성 인식, 기계 번역, 대화 시스템 등에서는 이러한 발화의 분석에 있어서 양태보다는 화행을 많이 연구하였다. 양태는 음성 합성, 기계 번역에서 목표 언어(target language)의 생성과 같은 영역에서 자연스러운 발화의 생성을 위해서만 사용되었다[4][5]. 예를 들어, “비가 온다.”는 명제에 대해 “비가 오는 것 같아.”, “비가 오는지도 몰라.”, “비가 오더라.” 등 추측이나 경험 등과 같은 화자의 주관적인 생각이 반영되어 보이는 보다 자연스러운 발화를 생성하는 목적에 이용되었다.

화행을 분류하기 위해서는 이전 발화까지 고려해야 한다. 항공기 예약 시스템이나 일정 관리 시스템과 같은 특정 도메인에 한정된 대화 시스템에서는 ‘확인 요구(Ask-confirm)’, ‘소개(Introducing-oneself)’와 같은 화행이 유용할 수 있지만, 일상 대화에서는 이러한 화행이 많이 출현하지 않는다. 또한 화행에서는 ‘질문’ 화행 뒤에 ‘긍정대답’, ‘부정대답’ 혹은 ‘진술’이 오거나, ‘제안’ 화행 뒤에 ‘수락’, ‘거절’ 화행이 오는 것과 같은 인접 쌍(adjacency pair)이 있는데, 이런 인접 쌍은 도메인 한정적인 대화에서 많이 나타나며 대화의 흐름에 중요한 역할을 한다. 하지만 대화 시스템의 경우 항공기 예약이나 일정 관리와 같이 특정한 도메인만을 다루는 것이 아니고, 일상적인 대화가 주를 이루기 때문에 이전 발화와의 관계가 뚜렷하지 않은 경우가 많다. 또한 이전 발화의 정보도 사용하기 용이하지 않은 경우가 많고 이전 발화의 정보를 사용함으로써 오류가 전파되어 오히려 화행을 사용하는 것이 성능을 떨어뜨릴 수 있다.

양태는 발화에 담긴 화자의 심리적인 태도를 분석하는 것으로, 이전 발화를 고려하지 않고 분류가 가능하다. 이렇게 해당 발화만으로 양태를 분류했을 경우, 다음 발화 생성시 이전 발화의 양태를 고려하면 적절한 응답 생성이 가능하게 된다. 다음 예문 (5)의 경우, 양태로 분류하면 ‘의지’와 ‘의무’ 두 가지 양태로 분류할 수 있다. 이는 어느 정도의 모호성을 남겨놓는 것이 되나, 답변을 생성할 때에는 필요에 따라 그 이전 발화의 양태까지 고려하도록 하면 되므로 대화 시스템에서 사용하기에는 적절한 의미 분석이라고 할 수 있다.

(5) 운동 해야지.

기존의 한국어 양태 분류에 관한 연구는 언어학과 국문학에서 분류 체계를 정립하고 양태에 관한 표현을 수집하는데 집중되어 있었으며, 자연어처리나 음성처리 분야에서 양태를 사용한 연구는 자연스러운 대화체 생성을 위해 양태 표현 패턴이나 DB를 사용한 연구만 있었다. 이에 본 논문에서는 기존 양태 분류 체계를 정리하고 대화 시스템에서 필요한 감정 및 응답 등의 체계를 덧붙여 확장하여 한국어 대화 시스템을 위한 양태 분류 체계를 정의하고, 이러한 양태 분류 체계가 유용한지를 검증하기 위하여 규칙 기반 양태 분류 시스템을 이용하여 양태 분류를 수행한다. 또한 실제 대화 시스템에 양태를 사용한 결과를 통해 그 유용성을 입증하고자 한다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2장에서는 양태 분류와 관련된 기존 연구를 살펴본다. 3장에서는 본 논문에서 정의한 대화 시스템을 위한 양태 분류 체계를 설명하며, 4장에서는 규칙 기반 양태 분류 방법을 기술한다. 5장에서는 4장의 양태 분류 시스템의 실험 및 평가를 다룬다. 6장에서는 결론을 맺는다.

2. 관련 연구

본 장에서는 한국어 양태와 관련된 연구를 살펴본다. 우선 언어학 및 국문학에서 양태의 범주 및 분류 체계를 다룬 연구를 다루고 언어학, 국문학 및 자연어처리와 음성처리에서 양태를 이용한 연구를 살펴본다.

2.1. 양태 범주 및 분류 체계를 다룬 연구

한국어 양태는 선어말어미 혹은 보조 용언, 종결 어미 등을 포함한 서술부에 나타난다. 이를 예문 (6)으로 설명하면 다음과 같다[6].

(6) 할아버지께서는 벌써 가셨겠다.

양태를 가장 좁은 의미로 사용하는 경우에는 선어말어미로 표현하는 명제에 대한 양태적 태도와 종결어미로 표현하는 청자에 대한 태도를 구분 짓고, 선어말어미로 표현되는 것만 양태로 본다. 이러한 관점으로 보면 예문 (6)에서 양태는 ‘시’와 ‘겠’에 나타나는 주체높임법과 추측이나 가능성을 나타내는 것만이 양태가 된다. 양태를 넓은 의미로 규정하면 [7]에서와 같이 문장의 종결부에 실현되는 모든 문법 범주를 양태로 본다. 따라서 예문 (6)에서 나타나는 주체높임법, 시제, 문장 유형, 청자대우법 등을 모두 양태의 일종으로 파악한다. 본 연구에서는 [8]과 같이 양태와 문장 유형 범주는 구분하되 선어말어미뿐 아니라 종결어미도 양태를 표현한다고 본다. 예를 들어, “할아버지께서는 벌써 가셨겠네.”라는 문장은 종결어미 ‘네’가 추측의 양태를 표현하기 때문에 예문 (6)과 다른 양태를 가진다고 본다.

한국어 양태 분류 체계를 다룬 연구는 의미론적 측면으로 나눈 연구와 통사론적 측면으로 나눈 연구가 있다. 의미론적 측면으로 양태를 접근하는 연구는 영어 조동사의 의미 분석에 양태를 이용하면서 본격적으로 논의되기 시작하였다. 의미론적 측면에서 양태 분류 체계를 나눌 경우에는 명제의 내용을 정보 내용으로 간주하느냐, 행위의 내용으로 간주하느냐에 중점을 두어 인식 양태(혹은 사실 양태, epistemic modality)와 행위 양태(혹은 의무 양태, deontic modality)로 크게 나누었으며, [9], [10] 그리고 [11]에서는 이 두 가지 외에도 감탄과 같은 화자의 감정적 태도와 관련한 양태를 평가 양태(혹은 감정 양태, evaluative modality)라고 하여 양태의 하위 범주로 설정하기도 하였다. 하지만 많은 양태 관련 연구에서 평가 양태를 고려하지 않았다. <표 1>은 [6]의 양태 분류 체계이며 인식 양태와 행위 양태만을 다루었다.

의미론적 측면에서 나눈 양태 분류 체계는 지나치게 세분화 되어 있으며, 추상적이고 관념적인 분류 체계가 포함되어 있다. 이러한 양태 분류 체계를 적용시키려면 발화에 간접적으로 드러나는 어감까지 분석을 해야 하므로 실제 대화 시스템에서 직접 사용하기에는 적합하지 않다. 예를 들어 <표 1>의 ‘정보의 확실성에 대한 판단’이나 ‘정보의 획득 방법’에 관한 양태는 자연어 처리 기법으로 처리하기에는 무리가 따른다.

대부분의 양태 체계가 의미론적 측면이 강한데 비해 [12]는 표현에 중점을 두어 양태 분류 체계를 정리한 연구이다. <그림 2>는 [12]에서 정의한 양태 분류 체계이다. [12]에서는 양태 표현과 서법을 구분 짓고, 양태 표현은 다시 ‘의무, 인식, 존재, 능력, 의지, 자기 반성’의 여섯 가지 세부 양태로 나누었다. 이 양태 체계의 경우 양태 분류 체계를 너무 적게 나누어서 대화 시스템에서 사용하려면 오히려 세분화 과정을 거쳐야 한다. 또한 문형을 제외한 여섯 가지 양태로 대화 시스템에서 출현이 가능한 모든 발화를 분류할 경우, 여섯 가지 중 어느 것에도 속하지 않

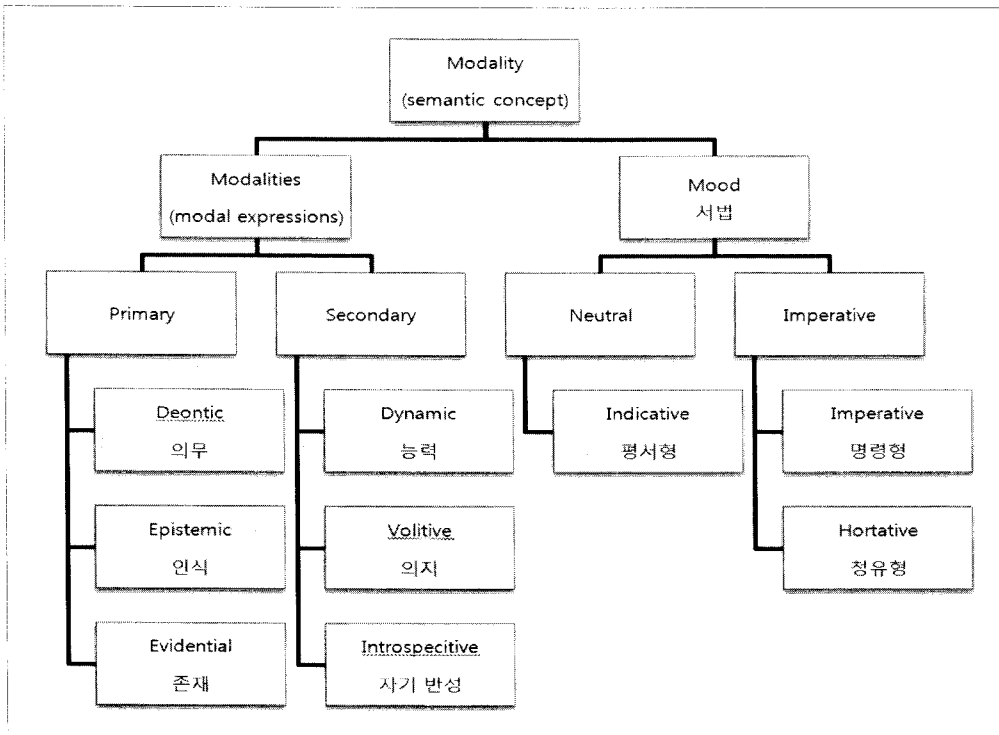
<표 1> [6]의 양태 분류 체계

대분류	중분류	소분류	소분류에 대한 설명
인식 양태	정보의 확실성에 대한 판단	확실성 판단	화자가 명제의 확실성에 대해서 완전한 확신을 가짐
		개연성 판단	화자가 명제의 확실성에 대해서 완전하지는 않지만 적어도 오십 퍼센트 이상의 확신을 가짐
		가능성 판단	화자가 명제의 확실성에 대해서 오십 퍼센트 미만의 확신을 가짐
	정보의 획득 방법	지각	해당 명제가 표현하는 정보가 감각적 경로에 의하여 획득된 것임
		추론	해당 명제가 표현하는 정보가 화자의 추론에 의하여 획득된 것임
		전언	해당 명제가 표현하는 정보가 다른 사람으로부터 들은 것임
	정보의 내면화 정도	이미 앎	해당 명제가 표현하는 정보가 이미 화자의 지식 체계에 내면화된 것임
		새로 앎	해당 명제가 표현하는 정보가 화자의 지식 체계에 미처 내면화되지 못한 것임
	청자 지식에 대한 화자의 가정	기지가정	해당 명제가 표현하는 정보가 참임을 청자도 이미 알고 있다고 화자가 가정함
		미지가정	해당 명제가 표현하는 정보가 참임을 청자가 아직 모르고 있다고 화자가 가정함
행위 양태	대타적 조건 부과	명령	화자가 청자에게 해당 명제가 표현하는 사태를 성립시킬 것을 요구함
		제안	다른 동작주가 해당 명제가 표현하는 사태를 성립시키는 것이 좋겠다고 화자가 의견을 말함
		기원	다른 동작주가 해당 명제가 표현하는 사태를 성립시키기를 화자가 바람
	재귀적 조건 부과	약속	화자가 청자에게 자신을 동작주로 하는 사태를 성립시키겠다는 의지를 선언함
		의도	화자 자신을 동작주로 하는 사태를 성립시키겠다는 의지를 표명함
		소망	화자가 스스로를 동작주로 하는 사태가 성립되기를 바람
	한국어 특유의 행위 양태 의미 영역	청유	명령의 일종
		경계	기원/소망의 일종
		후회	소망의 일종
		결심	약속의 일종

는 발화가 지나치게 많으므로 대화 시스템에서 양태를 이용하고자 한 목적을 달성할 수 없게 된다.

이와 같이 언어학 또는 국문학에서 정의한 양태 분류 체계는 대화 시스템에서 그대로 적용하기에는 여러 문제가 있다. 특정 몇몇 예제로 나와 있는 발화에서는 드러난 양태가 실제 대화에서는 매우 적게 출현하는 경우도 있으며, <표 1>의 ‘확실성 판단’, ‘개연성 판단’, ‘가능성 판단’과 같이 추상적이고 관념적인 분류의 경우, 대화 시스템에서 출현하는 짧은 길이의 발화를 처리하기에는 무리가 있을 수 있다. 또한 대화에서 많이 발생하는 감정을 다룬 발화를 배제하였다는 점도 문제가 된다. 예를 들어, “수고가 많다.”나 “생일 축하해.”와 같은 발화는 인식 양태나 행위 양태 등에 속하지 않지만 대화의 흐름에는 중요한 역할을 하기 때문에 다른 발화와 구분 지을 수 있어야 한다. 하지만 이를 <표 1>의 양태 분류 체계에 따르면 ‘의도’가 되고 <그림 2>의 양태 분류 체계에 따라 분류하게 되면 ‘의지’에 가깝게 된다. 따라서 이는 “내가 널 도와줄게.”와 같은 ‘의지’ 혹은 ‘의도’ 양태를 띄는 문장과 “생일 축하해.”와 같은 문장이 같은 분류를 가지게 되는 것으로, 대화 시스템의 응답이 부자연스러워질 수 있다.

이에 본 논문에서는 양태를 ‘대상이나 명제에 대한 화자의 태도’로 조금 폭넓게 정의하고, [6]와 [12]의 분류 체계를 기반으로 대화 시스템에 필요한 선호 및 비선



<그림 2> [12]의 양태 분류 체계

호, 감정, 문장 구조를 나타내는 분류를 추가하여 양태 분류 체계를 확장시켰다. 본 연구에서 사용하는 양태 분류 체계는 3장에서 다루도록 한다.

2.2. 대화체 생성을 위해 양태를 이용한 연구

자연어처리나 음성처리 분야에서 양태를 이용한 연구로는 기계 번역이나 대화 시스템에서 자연스러운 대화체를 생성하기 위해 양태를 사용한 연구가 있다[4][5].

[5]는 일·한 기계 번역에서 정확하고 자연스러운 한국어 술부 생성을 위해 양태를 사용한 연구로서 한국어와 일본어에서의 술부 표현의 차이를 극복하기 위하여 고안한 양상 자질 테이블을 사용하여 한국어 술부를 생성한다. 이 양상 자질 테이블은 한국어 술부의 언어학적인 지식과 한국어 술부를 생성할 때의 경험에 의해 만들어졌다. 이 연구에서 사용한 양태 분류 체계는 [13]에서 정의한 한국어 보조 용언의 19가지 체계인 ‘사역, 피동, 강조, 봉사, 완료, 시도, 중지, 희망, 의도, 허식, 진행, 가능, 부정, 당위, 승낙, 이유, 추정, 일념, 인용’ 중 일한 번역에서 직접 대응이 가능한 양태인 ‘강조, 시도, 허식, 승낙’을 제외하고, 선어말 어미가 가지는 양상 정보 ‘존경, 시제, 겸양, 회상’을 사용하였으며, 여기에 번역시의 경험에 의해 필요하다고 판단한 ‘지정, 필연, 불가능, 금지, 3자 희망, 결과/보존’의 체계를 추가하였다. 이 논문에서 사용한 양태 분류 체계는 일한 번역에서 직접 대응이 가능한 부분은 삭제하고 ‘3자 희망’과 같이 번역에 필요한 양태를 추가하였기에 대화 시스템에서 사용하기에는 부적합하다.

자연스러운 대화체를 합성하기 위해 약 27만 발화의 대규모 말뭉치에서 관례적으로 많이 쓰는 화행이나 양태가 포함된 표현을 추출한 연구가 있다[4]. 화행 및 양태는 동사나 형용사, 명사와 같이 어휘적인 요소에 의해 표현되는 경우와 종결어미, 보조용언 등의 문법적 요소에 의해 표현되는 경우가 있다. 이 연구에서는 어휘적인 형태로 표현되는 경우에는 수동으로 표현을 추출했으며, 문법적인 형태로 표현되는 경우에는 양태 표현 형태의 구성에 쓰이는 특정한 보조용언, 보조사, 의존명사, 종결어미 등을 선정하고 말뭉치의 형태소 부착 결과로부터 그 결합 형태를 추출하였다. 이 논문에서는 사용한 양태 체계를 밝히지 않았으며, give-information이나 request-information, request-action 화행에 속하는 다양한 양태 표현을 주로 추출하였다.

이와 같이 본 논문의 양태 분류 체계 및 양태 분류 시스템을 이용한 예제 기반 대화 시스템[1]을 제외하고는 한국어 양태의 분류에 관한 연구가 없었으며, 대화체 생성을 위한 연구에서도 보조 용언과 선어말어미, 종결어미의 체계에서 필요한 부분만을 추려 사용하였다. 이에 본 연구에서는 자연어처리에 적합하도록 기존 언어학 및 국어학에서 정립한 한국어 양태 분류 체계를 정리하고, 대화 시스템에 적용이 가능하도록 이를 확장한 양태 분류 체계를 정의한다. 그리고 대화체 발화를 정

의한 양태 분류 체계로 분류할 수 있는 규칙 기반 양태 분류 시스템을 구축하여 양태 분류 체계의 유용성을 검토하도록 한다.

3. 대화 시스템을 위한 양태 분류 체계

본 연구에서는 [6]과 [12]에서 정의한 양태 분류 체계에서 중복되는 것과 지나치게 추상적인 양태를 삭제한 후, 대화 시스템에서 필요한 양태를 추가한 양태 분류 체계를 정의하였다. 본 연구에서 정의한 분류 체계는 <표 2>와 같다. ‘서술’은 다른 양태로 분류되지 않은 발화를 의미한다.

양태 분류 체계와 관련된 기존 연구에서는 주로 인식 양태와 행위 양태만을 양태로 정의하였다. 그리고 감정 양태를 고려한 경우에도 실제 대화에서 나타나는 감정을 모두 다룬 것이 아니고 ‘감탄’과 같은 일부만을 다루었다. 그러나 대화 시스템에서는 자연스러운 대화의 생성을 위해 다양한 좀 더 많은 감정과 문장 구조, 관용적인 표현들을 다룰 수 있어야 한다.

감정을 나타내는 발화는 대화 시스템에서 자연스러운 대화를 생성하는데 중요하다. 예를 들어, “생일 진심으로 축하한다.”(축하)나 “수고가 많다.”(칭찬) 뒤에는 “고맙습니다.”(감사)와 같은 발화가 와야 대화가 자연스럽게 이어질 수 있다. 이런 식으로 다음 발화의 생성에 결정적인 영향을 미칠 수 있는 감정 발화로 ‘칭찬, 축하, 감사, 사과’가 있을 수 있다. 또한 ‘선호, 비선호, 기쁨, 슬픔’과 같은 감정을 나타내는 발화의 경우, 대화 시스템의 응답이 적절히 호응을 하지 않으면 사용자의 기분을 상하게 할 수도 있다. “나 지금 울고 싶어.”(희망, 슬픔)과 같은 문장을 ‘희망’ 양태로만 분류하고 “하고 싶으면 해.”라고 하는 것보다는 “슬퍼 보인다.”(추측)이나 “잘 되길 바래.”(희망)과 같은 문장이 오는 것이 적절할 것이다.

문장 내에 ‘가정’이나 ‘이유’가 포함되어 있는 경우와 과거에 있었던 일에 대한 회상을 하는 경우 역시 구분 짓는 것이 대화 시스템에 유용하다. 예를 들어, “기분이 나빠졌어.”(슬픔) 문장과 “너 때문에 기분이 나빠졌어.”(이유, 슬픔) 뒤에 올 발화는 달라져야 한다. “사귀어 봤어”(경험) 문장과 “사귀고 있어”(서술) 문장 뒤에 올 발화 역시 달라져야 한다. 이 외에도 ‘인사’와 ‘감탄’, ‘긍정(대답)’, ‘부정(대답)’은 대화 내에서 다른 발화가 하지 못하는 역할을 하므로 따로 양태 분류 체계로 두었다.

이런 이유로 본 연구에서는 감정 등을 나타내는 15가지 분류를 기존의 양태 분류 체계에 추가하였다. <표 3>은 특히 화자의 주관적인 태도 중 호불호가 명확하게 드러나는 양태를 정리한 것이다. ‘희망’과 ‘의지’는 호불호가 반영되나, 긍부정과 서술어를 고려하여 화자의 태도가 긍정적 혹은 부정적인지를 결정하는 양태이다.

<표 2> 예제기반 대화 시스템을 위한 양태 분류 체계

대분류	양태 이름	양태 정의	발화의 예
-	서술	특정 양태를 띄지 않는 발화.	
인식 양태	지각	과거나 현재에 알거나 모름 또는 인지함을 표현하는 발화.	전화번호 까먹었다. 밥 먹네.
	추측	과거, 현재, 미래에 대한 추측을 표현하는 발화.	엄청 귀찮아 보인다.
	전언	화자가 이전에 다른 사람에게 들은 말을 전하는 발화.	우리 엄마 수술 하신대.
행위 양태	희망	어떤 일을 이루어지는 것이나 좋은 결과를 바라는 발화.	비가 왔으면 좋겠다. 네가 오길 바래.
	금지	어떤 행동을 하지 말라고 제한하는 발화.	그만 가.
	의무	어떤 행동을 해야 함을 표현하는 발화.	형이 입원해서 간호하러 가야 돼.
	의지	화자가 어떤 행동을 하고자 하는 의지를 표현하는 발화.	요가 배워 볼까. 헬스 다녀야지.
	능력	어떤 행동을 할 수 있는지 여부를 나타내는 발화.	내가 할 수 있어. 먹을 줄 몰라.
	질문	화자가 모르는 것을 상대방에게 물어 보는 발화.	너 왜 말 안 해?
서술어를 통해 화자의 주관, 판단이 개입된 양태	선호	대상이나 상대방의 외향이나 성격, 특성 등에 대해 긍정적인 평가를 내리는 발화	재밌다. 비 내리는 거 좋아해?
	비선호	대상이나 상대방의 외향이나 성격, 특성 등에 대해 부정적인 평가를 내리는 발화	비 오는 거 싫어해? 너무 더럽다.
	칭찬	대상이나 상대방의 행동에 대해 긍정적인 평가를 내리는 발화	잘 생각했어. 수고가 많다.
	축하	어떤 사람의 행동이나 상태, 상황에 대해 축하하는 발화	생일 진심으로 축하한다.
	감사	화자가 청자에게 고마움을 표현하는 발화	놀아줘서 고마워.
	사과	화자의 미안한 감정을 표현하는 발화	자는데 미안하다.
	기쁨	좋은 기분, 감정을 표현하는 발화	난 지금 행복해.
	슬픔	안 좋은 기분, 감정을 표현하는 발화	내 신세야.
명제를 기술할 때 이유 나 조건을 제시한 양태	가정	특정 조건을 가정하고 그러할 경우의 행동, 소망 등을 표현하는 발화.	할 가치가 있는 일이면 제대로 해.
	이유	어떠한 일의 이유를 설명하는 발화	아파서 기분이 나빠.
발화 자체가 어떠한 표현을 담고 있는 양태	긍정(대답)	긍정적인 대답이나 반응을 표현하는 발화	응. / 알았어.
	부정(대답)	부정적인 대답이나 반응을 표현하는 발화	아니.
	인사	청자와 하는 인사를 나타내는 발화	잘 자. / 안녕.
	감탄	화자의 놀람, 느낌, 부름, 응답에 의한 발화	헉! / 아이고.
과거 경험을 표현	회상	과거에 했던 일에 대한 회상을 하는 발화	사귀어 봤어?

<표 3> 화자의 주관적인 태도(호불호)가 반영되는 양태

화자/주어가 양태를 통해 주관적인 태도를 취하는 대상	양태
어떤 일의 결과, 상태	희망
어떠한 행동	의지, 칭찬, 금지
대상이나 상대방의 외향, 성격, 특성	선호, 비선호
청자의 이전 행동이나 청자가 겪은 어떤 일의 결과	축하, 감사
사람의 기분, 감정	기쁨, 슬픔
긍정, 부정을 고려하여 화자의 태도를 구분해야 하는 양태	
희망, 의지	

4. 규칙 기반 양태 분류

3장에서 정의한 양태 분류 체계의 유용성을 검증하기 위하여 본 연구에서는 규칙 기반 양태 분류 시스템을 구축한다. 규칙 기반으로 양태 분류를 할 경우 국어학, 언어학의 양태 표현에 관한 연구를 적용해볼 수 있으며, 직관적으로 규칙 작성 및 결과 분석이 가능하다.

4.1. 규칙 추출 및 실험에 사용한 말뭉치

본 연구에서는 양태 분류 시스템에서 사용할 규칙 추출 및 시스템 평가에 사용하기 위한 말뭉치를 구축하였다. 말뭉치는 총 4,650개의 발화로 되어 있으며, 구어체 문장 및 양태 관련 서적의 예문을 위주로 수집하였다. 이 중 4,000개의 발화에서 규칙을 추출하였으며 규칙 추출에 사용하지 않은 650개의 발화를 대상으로 실험을 시행하였다. 말뭉치를 구축하는 과정에서는 한 개의 발화가 여러 양태로 분류될 수 있음을 고려하여 말뭉치 구축 시 최대 3개의 양태까지 부착하도록 하였다. 또한 한 발화에 여러 문장이 나올 경우 마지막 문장만을 고려하여 양태를 부착하도록 하였으며, 문장을 구분하는 기준은 문장부호로 하였다. 발화에 양태를 부착하면서 각 양태를 결정할 때 중요한 역할을 하는 부분을 원문에 표기하도록 하여 규칙 추출에 사용하도록 하였다. 규칙 추출 말뭉치 내 양태의 분포는 <표 4> 및 <표 5>와 같다.

두 개나 세 개의 양태로 분류한 발화는 다음과 같은 경우가 있었다. 우선, 한 문장에 여러 양태가 같이 나타나는 경우가 있다. 이는 한 문장이 여러 양태로 동시에 해석될 수 있는 경우로, “너무 추워서 냉방병 걸리겠다.”는 ‘이유’와 ‘추측’이 같이 분류가 되며, “그 옷 입으면 좋아.”는 ‘가정’과 ‘선호’가 한 발화에 같이 나타

나는 경우이다. 그리고 주어나 문맥에 따라 다른 양태가 나타나는 경우가 있는데 이는 모호성이 남아있는 것으로 볼 수 있다. 예를 들어, “집에 가야지.”라는 문장은 ‘의무’와 ‘의지’로 해석될 수 있으며, “골프 잘 해.”는 ‘명령’, ‘능력’, ‘칭찬’이 한 문장에서 같이 나타나는 경우이다. 마지막으로, 문장 분리가 안되어서 나타나는 경우가 있다. 이는 ‘긍정 대답’, ‘부정 대답’, ‘감탄’ 등이 들어가 있는 문장이 특히 많다. 예로 “응 그거 꽤 잔인하더라.”라는 문장을 들 수 있다.

<표 4> 단일 양태가 부착된 발화의 분포

단일 태깅된 양태	개수	비율(%)
서술	2240	58.947
선호	279	7.342
비선호	272	7.158
의지	246	6.474
지각	239	6.289
추측	101	2.658
희망	89	2.342
이유	53	1.395
회상	44	1.158
금지	38	1.000
능력	38	1.000
가정	35	0.921
의무	35	0.921
슬픔	22	0.579
전언	17	0.447
인사	15	0.395
사과	8	0.211
감탄	7	0.184
긍정	6	0.158
칭찬	6	0.158
부정	4	0.105
기쁨	3	0.079
축하	2	0.053
감사	1	0.026
총합계	3800	100

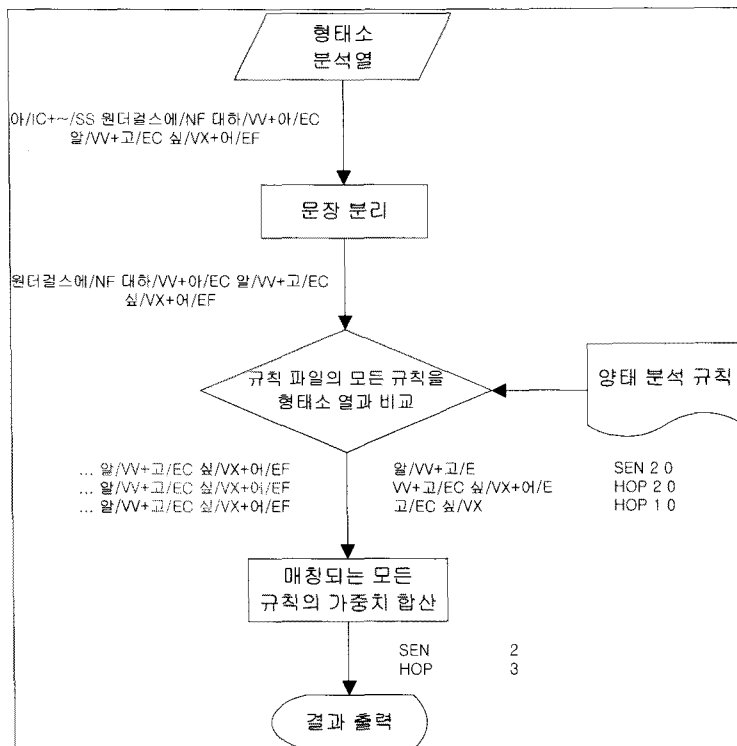
<표 5> 여러 양태가 부착된 발화내 양태 조합의 수

	복수를 태깅된 발화의 수	양태 조합의 수
양태 2개에 태깅	196	71
양태 3개에 태깅	4	4

4.2. 규칙 기반 양태 분류 시스템

본 논문에서 제안하는 규칙 기반 양태 분류 시스템은 <그림 3>과 같으며, 형태소 분석된 문장을 입력으로 하고, 출력은 최대 3개의 양태로 한다. 이때 사용하는 품사 집합은 세종 품사 집합[14]을 사용하며, 형태소 분석은 [15]의 형태소 분석기를 사용하였다. 문장 분리는 문장 부호를 기준으로 하며, 간단한 규칙으로 문장 분리를 시행하였다. 이 규칙 기반 양태 분류 시스템은 규칙 간에 독립이며, 어느 한 규칙의 형태소 열이 다른 규칙의 형태소 열에 포함이 되어도 각 규칙은 독립적으로 적용이 된다. 모든 규칙을 비교한 결과 매칭된 양태가 없을 경우 기본적으로 '서술'을 출력하도록 하였다. 이 시스템에서 사용한 규칙의 형태 및 각 요소의 역할은 4.3절에서 설명한다.

우선, 형태소-품사열을 입력으로 받아 문장 부호를 기준으로 간단하게 문장 분리를 수행한다. 그리고 양태 분석 규칙 사전에 있는 규칙들을 입력으로 들어온 형태소 열과 비교하여 규칙이 형태소 열에 포함이 될 경우 규칙에 부여되어 있는 가중치를 해당 양태에 합산한다. 그리고 모든 규칙을 비교한 후, 가중치가 0 이상인 양태 중 가중치가 높은 순서대로 세 개까지 결과로 출력한다. 본 연구에서 사용한 규칙의 형태 및 각 요소의 역할은 다음 4.3절에서 설명한다.



<그림 3> 규칙 기반 양태 분석 시스템 흐름도

4.3. 양태 분류 규칙

한국어에서 양태가 실현되는 방식은 다양하다. 한국어에서 양태표현을 용언에 한정하여 말할 때, 일반적으로 이를 어미를 통해 실현되는 경우, 보조용언을 통해 실현되는 경우, 본용언을 통해 실현되는 경우로 분류한다. 국문학에서 이러한 양태 표현을 집중적으로 다룬 연구로 다음과 같은 연구가 있다.

[6]에서는 ‘더’나 ‘겠’과 같은 어미를 다루었다. 이 연구에서는 ‘양태 어미’를 양태적 의미를 표현하는 활용 형식을 가리킨다고 정의하고, 한국어에서 ‘명제에 대한 화자의 태도’를 표현하는 활용 형식을 연구 대상으로 삼았다. [16]에서는 보조용언과 일반용언을 통해 양태가 실현되는 경우를 중점적으로 다루었다. 한국어에서는 그 의미가 양태와 연관되면서도 양태 의미의 담지자가 화자나 청자, 혹은 문장의 주어 지시 대상인 요소가 있다. 양태용언을 화자 중심 양태용언과 주어 중심 양태용언으로 나누고 이를 다시 인식 양태와 의무 양태(혹은 행위 양태)로 나누어 각 부류별 용언의 통사적 특성을 살펴보았다. [12]에서는 품사를 고려하지 않고 양태가 나타나는 표현을 다루었다. 이 때 다른 연구들과는 달리 품사를 고려하지 않고 ‘-수 밖에 없다’, ‘-지 모르다’와 같이 표현을 그대로 사용하였다. 본 논문에서는 위 연구들에서 다룬 양태 표현을 참고하여 규칙을 작성하였다.

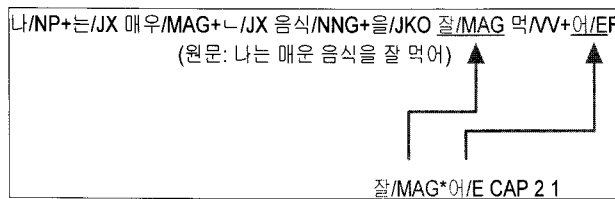
<그림 4>와 같이 본 연구에서 사용한 규칙은 형태소-품사 열, 양태 아이디/코드, 가중치, 와일드카드(wildcard) 사용 여부로 구성되어 있다. 양태 아이디는 ‘서술’을 100으로 하여 차례대로 부여한 세 자리 숫자이며, 양태 코드는 양태 명을 영어로 번역하여 세 글자의 영문으로 표기한 것이다. 형태소-품사 열을 규칙에서 사용할 경우 언어학이나 국문학에서 양태 표현에 관해 한 연구를 직접 규칙으로 적용이 가능하며, 규칙을 작성하거나 결과를 분석하는 과정에서 사람이 이해하기 쉽다. 본 연구에서 규칙을 작성할 때에는 말뭉치 구축시 양태 결정에 도움이 되는 부분을 원문에 표기하도록 하여 이를 수용할 수 있는 형태소-품사 열을 사용하였다. 가중치는 규칙 추출 말뭉치 4,000건에 규칙들을 적용하여 카이 제곱 통계량[17]을 사용하였는데, 수식은 다음과 같다.

$$\chi^2(r, m) = \frac{(A+B+C+D) \times (AD - CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \quad (1)$$

- A: 규칙 r 을 포함하고 양태가 m 인 발화 수
- B: 규칙 r 을 포함하고 양태가 m 이 아닌 발화 수
- C: 규칙 r 을 포함하지 않고 양태가 m 인 발화 수
- D: 규칙 r 을 포함하지 않고 양태가 m 이 아닌 발화 수

(형태소-품사열)	(양태ID/code)	(가중치)	(wildcard 사용 여부)
모르/VV+겠/EP	101(SEN)	96	0
모르/VV+겠/EP+네/E	101(SEN)	11	0
것/NNB 같/VA	102(GUS)	669	0
것/NNB 같/VA+나/E	102(GUS)	30	0
번/EC*르래/EC	106(HYP)	33	1
잘생기/VA*NNG	112(LIK)	0	1

<그림 4> 양태 규칙 파일의 예제



<그림 5> Wildcard 규칙 적용의 예

<그림 4>의 맨 마지막 규칙과 같이 가중치에 0을 준 규칙은 예외를 처리하기 위한 규칙이다. 이러한 규칙과 매칭될 경우에는 그 이전까지 합산한 가중치에 0을 곱하도록 해서 최종 출력 결과에서 제외하도록 하였다. 예외 처리 규칙은 규칙 사전의 가장 하단에 위치하도록 하였다. 예를 들어, “안 졸려.”(서술)의 경우 부사 ‘안’ 뒤에 동사가 오기 때문에 ‘의지’로 해석될 수 있지만 이 발화는 현재 화자가 졸리지 않다는 사실을 나타낸다. 이런 경우 예외 처리를 하여 정확률을 높일 수 있다.

와일드 카드는 규칙의 적용률을 높이기 위해 사용하였으며 이 규칙을 적용한 예는 <그림 5>와 같다.

5. 실험 및 오류 분석

이 장에서는 4장에서 설명한 규칙 기반 양태 분류 시스템이 얼마나 양태를 잘 분류하는지와 그 결과에 대한 오류를 분석한다. 실험 결과에 대한 평가 척도로는 식 (2)에서 식 (4)와 같은 정확률과 재현율을 사용하였다.

$$\text{정확률}(\%) = \frac{\text{맞은 양태 수}}{\text{시스템이 반환한 양태 수}} \times 100 \tag{2}$$

$$\text{재현율}(\%) = \frac{\text{맞은 양태 수}}{\text{해당 발화에 태깅되어 있는 정답 양태 수}} \times 100 \quad (3)$$

$$F\text{-measure} = \frac{2 \times \text{정확률} \times \text{재현율}}{(\text{정확률} + \text{재현율})} \quad (4)$$

실험 결과는 <표 6>과 같다. ‘선별한 규칙’은 카이제곱통계량 값이 1 이상인 규칙 880개만 사용한 결과이고, ‘모든 규칙’은 규칙 추출 과정에서 추출한 1,664개의 모든 규칙을 사용하되, 카이제곱통계량이 1 미만으로 나온 경우 가중치로서 1을 부여하여 실험한 결과이다. 추출한 모든 규칙을 모두 사용하였을 때 정확률과 재현율이 모두 높게 나왔는데, 이는 규칙 추출에 사용한 4,000건 발화에서 추출한 규칙에서 어미 등을 빼고 일반화시킨 규칙이 실험 집합에서는 적용되었음을 나타낸다.

<표 6> 양태 분류 실험 결과

	정확률(%)	재현율(%)	F-measure
선별한 규칙	95.82	94.49	95.15
모든 규칙	98.20	98.56	98.38

오류 분석을 통해 규칙 기반 양태 분류 시스템의 한계점을 몇 가지 발견할 수 있었다. 우선, 안긴문장과 명사를 수식하는 구 때문에 발생하는 오류가 있었다. 예를 들어, “하고 싶은 대로 모두 할 수는 없어.”라는 문장의 경우 ‘능력, 금지’로 분류되어야 하나 “하고 싶은”이라는 구 때문에 ‘능력, 금지, 희망’으로 잘못 분류되었다. 이러한 오류는 수동으로 구축한 규칙에 예외 처리가 되어 있지 않은 경우 발생할 수 있다.

서술어 생략으로 인한 오류도 발생하였는데, 예를 들어 “겨울이 되어도 따뜻하길.”이라는 문장의 경우 ‘희망’으로 분류되어야 하나 ‘바란다’ 등의 서술어가 생략되어 있으므로 양태가 올바르게 분류되지 않았다. 문장 내 도치로 인하여 발생한 오류도 있었는데 “몸을 좀 움직여봐 그럼.”이란 문장의 경우 강조의 의미로 쓰인 ‘그럼’ 때문에 ‘긍정(대답)’으로 분류되었다. 이런 서술어의 생략이나 문장 내 도치로 인한 오류 역시 규칙으로 모두 구축하는 것은 불가능하다.

이 외에도 형태소의 중의성이나 의미적 중의성으로 인한 오류가 발생하였다. “네가 좋아서 그래.”에서 ‘아서’는 ‘이유’의 단서로 사용되었으나, “집에 가서 하면 더 잘 돼.”에서 ‘아서’는 ‘가정’의 단서로 사용되었다. 이런 경우 많이 발생하는 양태에 맞게 규칙을 작성하는 등의 정책이 필요할 것이다. 또한 ‘선호’, ‘비선호’의 경우 말뭉치 구축에 참여한 사람의 주관에 반영이 될 수 있다. 예를 들어 ‘비싸

다.'라는 말이 있을 경우 이것을 '비선호'로 할 것인가 '서술'로 할 것인가를 결정해야 한다. 이런 경우 양태 분류 시스템을 사용하고자 하는 용도에 따라 지침을 일관되게 적용해야 할 것이다. 이러한 한계에 불구하고 수동으로 구축한 1,664개의 수동 구축 규칙을 사용한 시스템의 결과는 95.15의 F-measure를 보였다.

<표 7>은 본 논문에서 정의한 양태 분류 체계 및 양태 분류 시스템을 사용한 예제 기반 대화 시스템의 정확도 실험 결과이다[1]. 이 대화 시스템에서는 발화 집합과 그에 대한 응답 발화 쌍 DB를 구축하고, 사용자의 입력 발화를 발화 집합 중 가장 비슷한 발화를 검색하기 위한 유사도 계산의 자질로서 양태를 사용하였다. [1]에서는 문형, 시제, 긍/부정, 어휘 유사도와 양태를 함께 사용하여 발화의 유사도를 측정하였다. 아래 실험 결과에서 양태를 사용한 정확도가 양태를 사용하지 않은 정확도에 비해 약 11.2%의 성능 향상을 보였다. 이를 통해 양태가 예제 기반 대화 시스템에서 중요한 자질임을 알 수 있다.

<표 7> 예제 기반 대화 시스템에서 자질 사용에 따른 정확도 평가

사용한 자질	정확도(%)
양태를 제외한 모든 자질을 사용	66.52
모든 자질을 사용	73.97

6. 결론 및 향후 연구

본 연구에서는 언어학 및 국문학에서 이루어진 양태에 관한 연구를 정리하고 이를 대화 시스템에 적합한 양태 분류 체계로 정리하였다. 또한 이를 적용한 말뭉치를 사용하여 규칙 기반 양태 분류 시스템을 구축하고 실험 결과를 분석하여 향후 양태 분류 연구에 초석을 마련하고자 하였다.

본 연구에서는 형태소 분석 결과만을 사용하여 규칙 기반으로 양태를 분류하였으나 시제, 문형과 같은 정보나 구문 분석기를 활용할 경우 형태소 분석 결과만으로 분류하지 못한 발화를 올바르게 분류할 수 있을 것이다. 예를 들어, “못 해.”(능력), “못 해 봤어.”(회상)의 경우와 “너 그거 잘 해?”(능력), “너 그거 잘 해.”(능력, 칭찬)의 경우 시제 및 문형의 사용으로 양태 분류에 도움을 줄 수 있을 것이다. 구문 분석기를 활용할 경우 안긴문장과 명사를 수식하는 구에 생기는 오류를 감소시킬 수 있을 것이라고 기대한다. 이 외에 본 연구에서는 규칙을 수동으로 구축하여 양태 분류 시스템에 적용하였으나 향후에는 작성한 규칙을 활용하여 규칙 자동 추출이나 통계 기반 양태 분류 방법에 적용해볼 수 있을 것이다.

참 고 문 헌

- [1] 이연수, *다양한 의미 자질을 고려한 발화간 유사도 측정 방법*, 석사학위논문, 고려대학교, 2008.
- [2] J. A. Austin, *How to Do Things with Words*, Harvard University Press, 1962.
- [3] J. Lyons, *Semantics I · II*, Cambridge University Press, 1977.
- [4] 오승신, 김상훈, 이영직, “화행과 양태(modality)를 고려한 대화체 합성 DB의 구축”, *제20회 음성통신 및 신호처리학술대회 논문집*, 20권, 1호, pp. 97-100, 2003.
- [5] 김정인, 문경희, 이종혁, 이근배, “일한 기계번역에서 양상자질테이블 구동 방식에 의한 한국어 슬루 생성”, *정보과학회논문지(B)*, 24권, 11호, pp. 1284-1296, 1997.
- [6] 박재연, *한국어 양태 어미 연구*, 국어학총서 56, 태학사, 2006.
- [7] C. Fillmore, “The case for case”, in Bach and Harms Eds., pp. 1-88, 1968.
- [8] 장경희, “국어의 양태 범주의 설정과 그 체계”, *언어*, Vol. 20, No. 3, pp. 191-205, 1995.
- [9] P. R. Palmer, *Mood and Modality*, Cambridge University Press, 1986.
- [10] 박진호, 박병선, “서법의 정의 및 한국어 서법 범주의 체계화”, *제19차 한국어학회 전국학술대회 발표 요지*, 1999.
- [11] 이선용, “국어의 양태 체계 확립을 위한 시론”, *관악어문연구*, Vol. 26, pp. 317-339, 2001.
- [12] A. T. Wymann, *The Expression of Modality in Korean*, Ph.D. Dissertation, Universität Bern, 1996.
- [13] 고영근, 남기심, 박경조, *고등학교 문법 자습서*, 탑출판사, 1985.
- [14] 21세기 세종계획 국어 기초자료 구축 분과 연구 보고서, 2003.
- [15] 이상주, *자동 품사 부착을 위한 새로운 통계적 모형*, 박사학위논문, 고려대학교, 1999.
- [16] 김지은, *우리말 양태용언 구문 연구*, 한국문화사, 1998.
- [17] Y. Yand, J. O. Pefersen, “A comparative study on feature selection in text categorization”, *Proc. 14th International conference on Machine Learning*, pp. 412-420, 1997.

접수일자: 2008년 11월 10일

게재결정: 2008년 12월 6일

▶ 김민정(Min-Jeong Kim)

주소: 136-713 서울특별시 성북구 안암동

소속: 고려대학교 컴퓨터·전파공학과

전화: 02) 924-2054

FAX: 02) 929-7914

E-mail: mjkim@nlp.korea.ac.kr

▶ 홍금원(Gumwon Hong)

주소: 136-713 서울특별시 성북구 안암동

소속: 고려대학교 컴퓨터·전파공학과

전화: 02) 924-2054

FAX: 02) 929-7914

E-mail: gwhong@nlp.korea.ac.kr

▶ 송영인(Young-In Song)

주소: 136-713 서울특별시 성북구 안암동

소속: 고려대학교 컴퓨터·전파공학과

전화: 02) 924-2054

FAX: 02) 929-7914

E-mail: song@nlp.korea.ac.kr

▶ 이연수(Yeon-Soo Lee)

주소: 136-713 서울특별시 성북구 안암동

소속: 고려대학교 컴퓨터·전파공학과

전화: 02) 924-2054

FAX: 02) 929-7914

E-mail: yslee@nlp.korea.ac.kr

▶ 이도길(Do-Gil Lee)

주소: 136-713 서울특별시 성북구 안암동

소속: 고려대학교 민족문화연구원

전화: 02) 3290-5289

FAX: 02) 926-8385

E-mail: motdg@korea.ac.kr

▶ 임해창(Hae-Chang Rim) : 교신저자

주소: 136-713 서울특별시 성북구 안암동

소속: 고려대학교 컴퓨터·전파공학과

전화: 02) 924-2054

FAX: 02) 929-7914

E-mail: rim@nlp.korea.ac.kr