

존대등분 계산법과 사례기반추론을 활용한 상황 인식형 모바일 인터페이스 시스템*

권오병
경희대학교 국제경영학부
(lobkwon, serveLord@khu.ac.kr)

최석재
고려대학교 국어국문학과
(lingua@korea.ac.kr)

박태환
경희대학교 국제경영학부
(serveLord@khu.ac.kr)

모바일 서비스의 수용성에는 사용자 인터페이스가 중요한 요소이다. 특히 모바일 인터페이스에서 청자(listner)인 인간에게 화자(speaker)인 기계가 어떻게 시의적절한 대화를 하는가는 수용성에 중요한 요소임에도 불구하고 아직까지 이에 대한 본격적인 연구가 진행되지 못했다. 따라서 본 연구의 목적은 사용자의 상황을 인식한 존대등분 계산법을 제안하여 이에 근거한 시의적절한 대화를 지원하는 상황 인식형 모바일 인터페이스를 설계하도록 하는 것이다. 다만 존대등분 계산은 문화별 및 언어별로 차이가 날 수 있으므로 한국어를 대상으로 계산법을 제안하려고 한다. 유비쿼터스 환경과 같은 분산 환경에서 사생활 정보를 보호하면서 사례기반 추론을 수행하기 위해 기존의 Minkowski aggregation 방법을 수정한 Nested Minkowski aggregation 방법을 사용하였다. 또한 본 방법론의 사용가능성을 증명하기 위하여 드라마 사례를 가지고 시뮬레이션 기반의 성능 비교를 수행하였다.

논문접수일 : 2007년 04월

게재확정일 : 2007년 09월

교신저자 : 권오병

1. 서론

모바일 단말기를 통한 정보 서비스가 활발하게 제공되고 있는 중이다. 이러한 모바일 정보 서비스에서 중요한 이슈로 등장하는 것이 모바일 상호작용이다. 모바일 상호작용은 사용자의 사용성을 증진시키기 위한 설계, 구현 및 평가에 관심을 가지는 인간 기계 상호작용(Human Computer Interaction, HCI)에서 기인하며 궁극적으로 인간과 기계의 협력 성과의 극대화라는 목표를 가진다. 이러한 목표를 달성하기 위하여 감성적이고 자연스러우며 직관적인 인터페이스를 제공하기 위한 연구

가 활발히 진행 중이다(Mignonneau and Sommerer, 2005). 이러한 인터페이스의 핵심은 인간에게 의도된 감성을 정확하고 효과적으로 전달하기 위한 방법의 개발에 있다. 2001년도에 개발된 Mobile Feelings 디바이스 개발 프로젝트는 그 대표적인 연구라고 볼 수 있다(Sommerer and Mignonneau, 2003).

최근 RFID나 GPS 등을 활용하여 위치 기반 등 상황 인식형 모바일 서비스가 더욱 진보적인 서비스 모델로 관심을 받고 있다. 상황 인식형 모바일 서비스는 이동하는 개인에게 개인화된 서비스를 제공하려는 목적을 가진다. 이는 개인마다의 선호

* 본 연구는 21세기 프론티어 연구개발 사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅 및 네트워크원천기반기술 개발사업의 지원에 의한 것임.

도가 차이가 나며 그 선호도는 상황에 의하여도 영향을 받는다는 가정 하에 의미 있는 서비스로 받아들여진다.

상황에 따른 사용자의 선호도 변화는 서비스 인터페이스에도 반영되어야 한다. 예를 들어 사용자의 연령이나 성별에 따라서 그들이 선호하는 대화의 방식이 차이가 날 수 있다. 그러나 현재의 일반 인터페이스는 이러한 개인적 차이를 고려하지 않고 불특정 다수에게 일정한 인터페이스를 제공하는 서비스가 대부분이다. 다수인의 다양한 모바일 의사소통을 이해하는 하나의 인터페이스로 구성하는 것은 복잡한 정보를 소규모의 디스플레이에 제공하는 문제와 함께 모바일 의사소통에 있어서 중요한 이슈로 받아들여진다(Väänänen-Vainio-Mattila and Haataja, 1997; Jones, 1999; Väänänen-Vainio-Mattila and Ruuska, 1999; Nielson and Sndergaard, 2000).

이중에서 본 연구에서는 이러한 상황인식형 모바일 인터페이스에서 청자(listner)인 인간에게 화자(speaker)인 기계가 어떻게 시의적절한 대화를 하는가에 초점을 두려고 한다. 특히 같은 대화 내용이더라도 사용자에게 따라서 적절한 존대 등분을 갖추어 표현하는 것이 필요하다. 컴퓨터가 사용자의 기대 이하로 존대 등분을 낮추어 말하면 사용자는 불편함을 느끼고 발화문에 대하여 예의 바르지 못하다고 느끼게 될 것이다. 반면에 사용자는 현재 사용하고 있는 컴퓨터가 자신과 가깝고 친밀한 관계에 있다고 생각하는데 컴퓨터는 지나치게 높은 존대 등분으로 자신에게 이야기하면 거리감을 느끼게 되고 심지어는 자신을 무시한다고 여길 수 있다. 그러나 개인에게 시의적절한 인터페이스를 제공하기 위해서 상황 정보에 기반을 두는 본격적인 연구가 미흡하다.

따라서 본 연구의 목적은 사용자의 상황을 인식

한 존대등분 계산법을 제안하여 이에 근거한 시의 적절한 대화를 지원하는 상황 인식형 모바일 인터페이스를 설계하도록 하는 것이다. 다만 존대등분 계산은 문화별 및 언어별로 차이가 날 수 있으므로, 한국어를 대상으로 계산법을 제안하려고 한다.

본 연구의 구성은 다음과 같다. 제 2장에서는 기존의 국어학 혹은 국어정보학에서 제안되어 온 존대등분 영향 요인들을 비교적 자세히 검토해 본다. 제 3장에서는 존대등분 계산법을 제시하고, 제시된 존대등분 계산법에 상황인식을 추가적으로 고려한 프레임워크와 사례기반추론방법을 제안하고 제4장에서 실험을 통하여 방법론의 우수성을 증명하였다. 그리고 마지막 제 5장에서는 본 연구의 공헌과 추가적인 연구방향을 기술하였다.

2. 문헌에서 본 존대 등분 선택에 영향을 주는 요소

국어학 분야에서는 존대 등분의 현상은 물론, 존대 등분에 영향을 주는 요소에 대한 연구가 이어져 왔다. 존대 등분 현상에 관한 연구로는 어말 어미 형태 종류의 파악, 말의 높임 수준인 화계의 설정, 화계의 넘나들 등에 대한 연구가 있어왔고 존대 등분에 영향을 주는 요소에 관한 연구로는 특정 높임말이 어떠한 관계에 있는 사람들에 의해서 쓰이는가가 있었다. 초기의 연구는 자신보다 사회적 지위가 높은 사람에게는 존댓말을, 낮은 사람에게는 낮춤말을 쓰며 어떠한 형태가 쓰인다는 것을 밝히는 수준이었으나 이후 연구가 진행되어 가면서 화자와 청자의 연령 등에 따라 즐겨 쓰는 화계가 달라지고, 사회적 지위 외에도 화자와 청자의 관계, 제 3자의 성격과 참여 여부 등도 존대 등분 선택에 영향을 미친다는 점이 알려지게

되었다. 이러한 점은 우리가 사용하는 존댓말이 단일한 요인에 의하여 선택되는 것이 아닌 보다 복합적인 요인들에 의하여 이루어진다는 것을 말한다. 여기에서는 그 동안 연구되어 온 상황 변인에 관한 대표적인 연구를 언급하도록 하겠다.

2.1 화자의 연령

화자의 연령은 존대 등분 선택에 가장 결정적인 영향을 미치는 것으로 알려져 왔다. 그 중 박영순(1976)은 녹음과 설문지 조사에 의하여 이를 실증하였다. 그의 연구에서 가장 주목할만한 점은 연령대에 따라 자주 사용하는 종결어미가 다름을 밝혀냈다는 것이다.

<표 1>은 각 연령대에 따라 쓰이는 종결어미 형태, 즉 화계의 빈도를 보여주고 있다. 예를 들어 60대의 경우에는 ‘하십니까’ 형태를 가장 많이 쓰고, 다음으로는 ‘하느냐/하니’ 형태를 쓴다. 그리고 ‘해’ 형태의 경우에는 20대 이하에서 많이 쓰는 것을 볼 수 있다. 상대방을 높이고자 할 때 ‘~하세요’ 형으로 말을 끝맺는 것은 쉽게 생각해 볼 수 있는 바이나 그것이 주로 청장년층에 집중되는 현상이고 노년층과 유년층에서는 다르게 나타남은 유의해서 볼 만하다. 노년층은 가장 격식을 갖춘

‘하십니까’ 체를 이용함으로써 상대방을 더욱 높일 뿐만 아니라 자신의 품위도 유지하려는 것으로 보이나 유년층에서는 아직 어법이 충분히 숙지되지 않아 존댓말을 잘 사용하지 못하고 가장 단순한 형태인 ‘해’ 체로써 대부분의 말을 하는 것이다. 우리가 기계를 통하여 문장을 생성시킬 때에는 이러한 면이 고려되어야 보다 자연스러운 대화가 이루어질 수 있을 것이다.

2.2 화자와 청자의 연령

박영순(1976)이 화자의 연령대만 고려하였다면 엄경옥(2002)는 청자의 연령대도 같이 고려하였다. 사실 경어법 현상은 두 사람의 대화에 의해 이루어지는 것이므로 청자의 연령이 같이 고려되어야 함은 당연한 일이다. 20살을 하나의 연령대로 묶어 처리한 이 연구의 결과를 보면 아래 <표 2>와 같다.

표는 각 연령대의 화자가 자신보다 순위, 동년배, 손아래 청자 중 어떠한 사람을 만나는가에 따라서 화계가 달라짐을 보이고 있다. 이에 따라 결과가 다소 복잡하게 나타났는데 예를 들어 ‘해체’는 주로 동년배 이하에 대해서 쓰였지만 손윗 사람에게 대해서도 쓰일 수 있음이 보인다. 이것은 우

<표 1> 박영순(1976)의 종결어미의 사용빈도

| | 하십니까 | 하세요 | 해요 | 하오 | 하는가 | 하나 | 해 | 하느냐/하니 |
|-------|------|-----|----|----|-----|----|----|--------|
| 60대 | 48 | 34 | 6 | 4 | 2 | 3 | 17 | 32 |
| 50대 | 32 | 48 | 14 | 2 | 1 | 1 | 32 | 16 |
| 40대 | 36 | 51 | 12 | 1 | 2 | 25 | | 20 |
| 30대 | 36 | 46 | 16 | 1 | | 1 | 30 | 14 |
| 20대 | 20 | 50 | 10 | | 2 | | 42 | 24 |
| 14~19 | 14 | 40 | 8 | | | | 52 | 32 |
| 9~13 | 8 | 38 | 18 | | | | 50 | 32 |
| 6~8 | 2 | 6 | 10 | | | | 90 | 36 |

<표 2> 엄경옥(2002)의 청자대우법 화계 구분

| 대우 정도 | 구분 | 유소년기 (초등학생이하) | 청년기 (10~20대) | 장년기 (30~40대) | 중년기 (50~60대) | 노년기 (70대 이후) |
|----------|-----|------------------|----------------------|--------------------------|--|----------------------------|
| 높임 | 순위 | 하세요체 해요체 | 하십시오체 하세요체 해요체 | 하십시오체 하세요체 | 하십시오체 하세요체 | 하십시오체 |
| | 동년배 | 해체 | 해요체 해체 | 하세요체 해요체 하게체 해체 | 하세요체 하시오체 하오체 하게체 하시게체 해체 | 하시오체 하오체 하시게체 하게체 |
| | 손아래 | 해체 | 해체 | 해체 | 하시게체 하게체 해체 | 하시게체 하게체 해체 |
| 안높임 | 순위 | 해체 | 해체 | 해요체 | 하세요체 하시오체 해요체 | 하시오체 하오체 |
| | 동년배 | 해체 | 해체 해라체 | 하게체 해체 | 하오체 하게체 하시게체 해체 | 하시게체 하게체 |
| | 손아래 | 해체 | 해라체 | 해체 | 하게체 해체 | 하시게체 해체 |

리가 손윗 사람이라고 하더라도 반드시 높이는 것이 아니라 친밀도에 따라 높이지 않는 경우가 있음도 고려한 것이다. 이경우(2001, 2003, 2004)에서는 청자가 가족과 같은 가까운 관계에 있는 사람일 경우에는 손윗 사람이라고 하더라도 낮춤말을 쓸 수 있다는 것을 드라마 대화 내용의 분석을 통하여 밝혔다. 친밀도가 높은 경우에 높임말만을 고집한다면 가족 사이의 대화가 아니라 은행에서 직원과 하는 대화처럼 들릴 수도 있어 오히려 부자연스러운 대화가 되는 것이다. 이렇게 높임의 상황에서 쓰인 낮춤말은 '친숙형'이라고 할 수 있을 것이다.

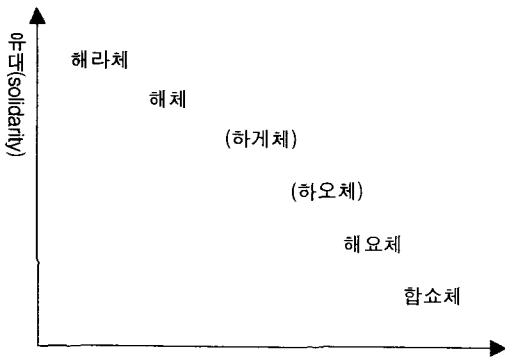
청자가 어떠한 사람인지, 나와서의 관계가 어떠한지에 따라서 높여야 되는 경우에 낮출 수도 있고, 낮춰야 되는 경우에 높일 수도 있다는 것은 흥미로운 결과이다. 기존의 연구에서는 주로 나이를

중심으로 하여 화자보다 손윗 사람에게서는 높임말을, 손아래 사람에게서는 낮춤말을 쓴다고 간단히 생각하였는데 이제 그 관계가 보다 복잡해진 것이다. 이렇게 생각해 보면 화계를 결정하는 요인은 보다 다양할 수 있을 것이다. 화자와 청자의 개인적 정보뿐만 아니라 장소, 감정상태, 활동상황 등도 어떠한 정도로 말을 높일 것인지에 영향을 줄 수 있을 것이다.

2.3 힘과 친밀감

Brown and Gilman(1960)은 청자대우법의 수준을 결정하는 요소로 힘(power)뿐만 아니라 유대관계(solidarity)도 관여한다고 하였다. 앞서 살펴본 엄경옥(2002)에서도 '대우 정도'라는 표현을 통하여 자신보다 높은 지위의 사람이더라도 의도적

으로 낮출 수 있음을 보였는데 Brown and Gilman (1960)의 연구에서는 이를 결정하는 요소가 유대 관계라고 보다 분명히 지적한 것이다. 이것은 높임말 선택에 있어 인적 사항과 같은 비교적 알기 쉬운 요소뿐만 아니라 ‘관계’라는 추상적이고 객관적 측량이 난해한 요소도 영향을 준다는 것을 말하기도 한다. 유송영(1993)은 이를 우리말에 적용시켜 다음과 [그림 1]과 같은 그래프를 도출하였다.



[그림 1] 유송영(1993)의 국어 불특정 청자 대우법 체계

[그림 1]은 화자와 청자의 친밀도가 높을수록 ‘해라체’, ‘해체’를 쓰는 경향이 있고, 순위 관계가 분명할 때에는 ‘해요체’, ‘합쇼체’를 쓰는 경향이 있음을 보이고 있다. 상대방의 지위 정도와 나와 친밀도 둘을 놓고 어느 쪽에 더 비중을 주어 말할 것인지를 결정해야 하는 것이다. 그러므로 상대방을 분명 높여 말해야 되는 상황이지만 나와 친밀도도 무시할 수 없다면 ‘합쇼체’보다는 ‘해요체’, ‘해요체’보다는 ‘해체’를 써야 한다. 여기서 ‘하계체’와 ‘하오체’는 화자와 청자의 연령에 제약이 따르기 때문에 괄호로 표시되어 있다.

또한 [그림 1]은 그 동안 잘 고려되지 않아왔던 유대 관계의 중요성을 강조하고 있다. 만약 ‘유대’

를 무시하고 ‘힘’만을 고려하여 말을 한다면 X축만을 생각하여 ‘해요체’ 또는 ‘합쇼체’를 쓸 것이다. 그러나 이는 관계를 무시한 것이기에 청자는 거리감을 느끼게 되고 둘의 관계에 실망할 수 있다. 둘 사이가 가깝다고 여겨지면 보다 많은 자신의 정보를 내놓을 수 있는데 마음이 닫히게 되면 형식적이고 격식적인 태도를 유지하면서 최소한의 정보만을 내놓게 된다. 반면에 ‘유대’만을 생각하여 화계를 지나치게 낮추면 청자는 자신이 충분히 대접받지 못한다고 여기고 불쾌하게 생각할 수도 있다. 이는 상대방을 무시하는 것이며 격식을 갖추어 말하는 것보다 더 안 좋은 결과를 초래할 수 있다. 그러므로 우리는 대화를 할 때 ‘유대’와 ‘힘’이라는 두 축의 긴장관계를 고려하여 적절한 수준으로 대화를 해야 한다. 이러한 점은 대화를 어렵게 만드는 측면도 있지만 상대방이 나에게 항상 관심이 있다는 느낌을 가지게도 한다.

2.4 제 3자 참여자 효과

이정복(1994, 1996)은 화자와 청자 이외에 제 3자가 있을 경우의 경어법 사용 양상에 관하여 연구를 하였다. 제 3자가 있을 경우에는 그렇지 않은 경우에 비하여 경어법 사용에 다른 양상이 있는 점을 파악한 것이다. 이 연구에서는 다음과 같은 세 가지 조건에 따라 경어법 사용 전략이 바뀐다고 보았다.

여기서 말하는 ‘주체’는 대화 중 언급되는 주인공을 말하는데 다른 말로 대화에 관계되는 제 3자라고 할 수 있다. 대화는 기본적으로 화자와 청자 사이에 일어나는 것이지만 제 3의 인물에 대해서도 종종 대화가 진행되는데 제 3자를 얼마나 높여서 말할 것이냐 하는 문제는 화자와의 관계만으로 생각할 수 없고, 청자와의 관계도 같이 고려되어

<표 3> 제 3자 참여자 효과

| 효과명 | 설명 | 효과 | 특이사항 |
|------------|---------------------------|--|------------------------------|
| 주체인물현장성 효과 | 주체가 현장에 존재하고 있는가 여부 | 주체가 현장에 있는 경우 더 높게 대우한다. | 청자의 가족 내 지위가 낮을 때 변화폭이 더 크다. |
| 참여자 지위차 효과 | 주체와 청자 간의 지위차 | 주체와 청자가 비슷한 지위일 때 주체가 가장 높이 대우 된다. | |
| 참여자 범주차 효과 | 주체 인물과 청자가 같은 범주의 인물인가 여부 | 화자와 청자가 다른 범주의 인물일 경우 주체에 대한 높임의 정도가 더욱 커진다. | 화자와 청자의 지위차가 클수록 변화폭이 크다. |

야 하므로 복잡한 양상을 띄게 된다. 이러한 제 3자가 대화의 현장에 있을 때와 없을 때는 그에 대한 경어법 사용 양상이 달라지고, 또 그 제 3자가 화자 및 청자와 어떠한 관계에 있는지에 따라 다시 달라진다는 것이다. ‘관계’라는 추상적 요소는 분석하기가 복잡한 것인데 <표 3>은 이 중 영향력이 큰 요소를 파악해 놓았다는 점에서 의의가 크다고 하겠다.

이상과 같이 국어학에서 언급되어 온 상황적 변인들에 대한 문헌 조사를 하였다. 경어법 현상을 이루는 요소로는 기본적으로 화자와 청자의 연령이 지적되어 이에 대한 연구가 활발하였다. 그러나 화자와 청자의 나이 차이가 화계를 결정하는 절대적인 요인은 아니며 화자는 자신의 연령대에서 즐겨 쓰는 화계가 있고, 이는 또 청자와의 관계에 따라 다양하게 변할 수 있었다. 자신보다 손윗사람이라고 하더라도 관계에 따라서 적절한 수준으로 낮출 수가 있었는데 언급된 주요 관계는 유대관계, 즉 친밀도였다. 친밀도가 높은 사이이면 적절한 수준으로 낮추어 말하는 것이 보다 자연스러운 관계를 유지할 수 있는 것이다. 이에서 예상할 수 있는 바는 만약 기대보다 경어 수준을 너무 낮추어서 말을 하면 청자는 충분히 대우받지 못했다고 여기고 불쾌한 감정을 가지리라는 것이다. 이 외에 더 언급될 수 있는 요소로는 제 3자의 존재 여부 등이 있다. 제 3자가 존재하고 있는 상황

에서 그와 관계된 대화를 하고 있다면 경어 수준에 상당한 변화가 생길 수밖에 없다. 이러한 면들을 포함하여 경어 수준을 결정하는 요인들을 다 밝혀내는 것은 그 추상적 속성 때문에 어렵겠지만 결정적인 영향을 미치는 중요한 요소들은 명시적 기술이 가능함을 문헌들은 보여주고 있었다. 이에 따라 다음 장에서는 이러한 연구 결과들을 취합하고, 아직 다루어지지 않아왔던 요소들을 더 추가하여 상황 변인들이 만들어내는 존재 등분을 확인해 볼 것이다. 그 결과는 이상의 연구 결과가 실증적으로 얼마나 타당한지를 보여줄 것이며 동시에 우리가 경어 수준을 결정하기 위해서는 어떠한 요소들이 필요한지를 알려줄 것이다. 이러한 작업의 결과는 향후 상황 정보만 입력된다면 화자인 기계는 청자인 사람을 불쾌하게 하지 않는 수준에서 친밀감을 발휘하며 자연스러운 대화를 이어갈 수 있게 할 것이다.

3. 상황인식형 존재등분 계산 방법론

3.1 문장의 경어법 점수 산출

본 시스템은 기본적으로 상황 정보 인식에 의하여 존재 등분을 결정하는 것이지만 사용자가 입력한 문장을 통하여 존재 수준을 보다 정밀하게 보정할 수 있다. 상황이라는 것은 일반적으로 적

용되는 것이므로 개인과 대화 시의 상황이 입력된다면 그 개인에게 쓰이는 화계는 일반적으로 결정되어 고정될 것이다. 그러나 사용자가 기계와 대화를 진행해 나가면서 기계와의 대화가 보다 친밀하게 혹은 예의를 갖추어 대화하도록 바뀌기를 원할 수 있다. 일반적으로는 대화가 진행된 시간이나 횟수가 많아질수록 친밀감이 발생하여 존대 수준이 낮아지겠지만 둘 사이에 오해 등이 발생하여 관계를 재정립하는 것이 필요하다고 생각된다면 다시 경어 수준을 올려 거리를 두려고 할 것이다. 이러한 관계는 기계에 의해 주도되기보다는 사용자에 의해 이끌릴 것이므로 컴퓨터는 사람의 반응을 지속적으로 점검하며 존대 수준을 결정해야 한다. 그러나 이런 실시간적인 변화는 상황에 의해서만은 알 수 없기 때문에 사용자의 문장을 분석해야 하는데 문장의 완벽한 파싱(parsing)이 어려운 현재의 상황에서는 객관적으로 파악이 가능한 몇 가지 형태적 특징을 관찰하여도 경어법 수준이 결정될 수 있다.

경어법 점수가 필요한 또 다른 이유는 실험의 결과를 검증하기 위해서이다. 본고의 실험에서 사용된 존대 수준이 올바른지를 확인하기 위해서 사람이 직접 경어법 점수를 매길 수도 있으나 이 과정은 주관적이어서 일관성을 잃기 쉽다. 이보다는 문장의 형태를 파악해 도출된 경어법 점수를 산출하여 비교하는 것이 보다 신뢰할 수 있는 검증을 이룰 수 있을 것이라고 여겨진다.

3.1.1 존대 등분 계산에 영향을 주는 요소

경어법 요소는 크게 어휘적 요소, 형태론적 요소, 문체론적 요소로 구분할 수 있다. 오미정(2005)에서는 용언과 체언의 어휘적 존대표현에 관하여 표준국어대사전의 어휘를 대상으로 분석하였다. 이

연구에 따르면 ‘밥’에 대한 ‘진지’, ‘떡다’에 대한 ‘잡수시다’와 같은 어휘적 존대 표현은 그 수가 매우 적고 대부분은 ‘-님’과 같은 체언의 접사나 ‘-시-’와 같은 용언의 선어말어미에 의하여 이루어진다고 한다. 이것은 존대 표현을 가진 단어 형성에서 어휘적 요소는 분포 상 큰 역할을 하지 못하고 형태론적 요소가 대부분을 차지한다는 것을 말한다.

성기철(1996)은 국어 경어법 체계에는 문체론적 요소도 있음을 언급하였다. 비록 이 논문에서 문체론적 요소는 양상이 매우 복잡하여 객관적인 등급 설정이 어렵다고 하였지만 제시된 예문을 살펴볼 때 통제될 수 있는 요소들이 관찰되었다. 이 요소들을 분석해 보면 존대 등분은 일차적으로는 ‘해라체-해체-하계체-하오체-해요체-합쇼체’로 구분되는 화계와 깊은 관계가 있는 것으로 파악된다. 화계와 함께 높임의 등분에 관여하는 것으로 관찰되는 요소는 의문형의 종류이다. 같은 화계에서는 단순의문형을 사용한 경우보다는 부정의문형을 사용한 경우가 더 높은 등분을 갖는 것을 볼 수 있는데 그 이유는 부정의문형이 단순의문형보다 청자에게 더 많은 선택권을 부여하기 때문이다. ‘~하시겠습니까?’라는 표현보다는 ‘~하지 않으시겠습니까?’라는 표현은 화자가 하지 않을 것을 전제하고 질문하는 것이기 때문에 청자는 그만큼 덜 부담을 가지고 자신의 의사를 선택할 수 있게 된다.

이와 같은 점을 생각해 볼 때 ‘서술-약속-의문-명령-청유’법으로 구분되는 서법 역시 존대 등분을 계산 시 고려할 수 있는 요소이다. 예를 들어 ‘저를 좀 도와주시겠습니까?’라는 의문형 요청 구문은 ‘나를 좀 도와줘’와 같은 명령형으로도 만들 수 있는데 전자에 비하여 후자가 보다 낮은 경어법 점수를 받게 된다. 명령형은 의문형에 비하

여 청자에게 선택권을 거의 주지 않기 때문이다. 그러나 서법 및 의문형의 종류는 모든 경우에 적용되기는 어렵다. 위와 같이 ‘요청’ 구문에서는 의문형이 쉽게 쓰이므로 가능하지만, ‘나는 어제 학교에 갔다’와 같은 단순 정보 전달형 구문은 다른 서법으로 만들기가 거의 불가능하기 때문이다. 그러므로 의문형 종류와 서법은 ‘요청’ 구문과 같이 특수한 경우에 한하여 생각해야 한다.

이상에서 우리는 존대 등분 계산에 관여하는 요소로 일부의 어휘적 요소, ‘-님’, ‘-시-’를 중심으로 하는 형태론적 요소, 통계 가능한 몇몇 문체론적 요소가 있음을 보았다. 뒤에서는 이러한 요소들을 가지고 적절한 경어법 점수 체계를 마련해 볼 것이다.

3.1.2 문헌에서 본 경어법 점수의 설정 사례

경어법 점수 산출을 최초로 시도한 것은 이정복(1993)에서이다. 경어법 요소로 ‘-님, -께서, -시-’를 들고 이들을 통하여 아래의 <표 4>와 같이 경어법 산출 체계를 마련하였다.

<표 4> 이정복(1993)의 존대등분 계산표

| 문장 유형 | 높임의 경어법 형식 | 경어법 점수 |
|-------------|--------------|--------|
| 과장님께서 오십니다. | -님, -께서, -시- | 10 |
| 과장님께서 옵니다. | -님, -께서, Ø | 6 |
| 과장님이 오십니다. | -님, Ø, -시- | 7 |
| 과장님이 옵니다. | -님, Ø, Ø | 3 |
| 과장께서 오십니다. | Ø, -께서, -시- | 7 |
| 과장께서 옵니다. | Ø, -께서, Ø | 3 |
| 과장이 오십니다. | Ø, Ø, -시- | 4 |
| 과장이 옵니다. | Ø, Ø, Ø | 0 |

‘-시-’에는 4점을, ‘-님’과 ‘-께서’에는 각각 3점을 배당하여 표현된 요소가 합하여져 경어법 점수가 산출되었다. ‘-시-’가 다른 요소보다 1점이 더 높은 것은 용언부에 속하여 존대 효과에 더욱 크

게 기여하기 때문으로 보인다. 이 체계에 따라 ‘과장이 옵니다’는 아무 경어 표현이 없이 단순 사실만을 전달하는 문장이므로 0점이 부여되었고, ‘과장님께서 오십니다’는 ‘-님, -께서, -시-’의 세 요소를 모두 갖추고 있으므로 10점이 부여되었다. <표 4>를 전체적으로 보았을 때 존대 등분 관계는 무리가 없어 보인다.

그러나 위의 체계는 한계를 갖고 있다. 위의 예문들은 주체경어법의 경어법 점수를 산출한 것인데 사실 주체경어법은 대화 상에서 그리 자주 사용되지 않는다. 대화는 기본적으로 화자와 청자 사이에서 이루어지는 것이기 때문에 경어법 사용 빈도로 보았을 때는 문장의 주어를 존대하는 주체경어법이나 행위가 미치는 대상을 높이는 객체경어법보다 대화를 듣고 있는 청자를 높이는 청자경어법이 월등히 많이 사용되게 된다. 또한 앞의 두 경어법은 대화 내용에 관한 것이지만 청자경어법은 대화 주체에 관한 것으로 범주가 다른 부분이 있어 청자경어법은 종종 주체경어법을 포괄하는 경우가 있다.

다른 또 하나의 문제는 위 예문의 상당수가 실제로는 쓰이기 어렵다는 점이다. 제시된 ‘과장님께서 오십니다, 과장께서 오십니다, 과장께서 옵니다, 과장이 오십니다’와 같은 예문은 실제로는 잘 쓰이지 않는다. 가능성 면에서는 생각해 볼 수 있지만 실제 대화에서 이러한 문장을 쓴다면 의도적으로 주체를 낮추려는 의도가 있다고 해도 매우 어색한 문장이 되는데 이는 호응을 지키지 않아서이다. 최현배(1937), 김민수(1973)에서는 주부(主部)와 술부(述部) 사이에 높임의 정도가 서로 맞아야 한다고 하였다. 술어부에서는 존대 표현을 사용하였는데 주어부에서는 낮춤의 표현을 사용하면 호응이 맞지 않아 문법성이 결여된다는 것이다. 위의 예문들은 주부를 높였다면 술부도 같이

높여주고, 주부를 낮추었다면 술부도 낮춰 주어야 하는데 이를 지키지 않아 발생하기 매우 어려운 예문들이 되었다.

이러한 문제점이 있지만 이정복(1993)의 연구는 경어법 점수 체계가 형태적 분석을 통하여 가능함을 보여주었고 점수 설정의 대략적인 윤곽을 지어 주었다는 의의가 있다.

3.1.3 경어법 점수의 설정 방법

경어법 점수를 설정은 두 가지로 나누어 볼 필요가 있다. 앞에서 문체론적 요소 중 서법과 의문형의 종류는 요청 구문과 같은 특수한 경우에만 발생한다고 하였는데 이러한 경우는 사람과 기계 사이의 대화에서 자주 일어나는 것이므로 제외시키기 어렵다. 그러므로 경어법 점수는 서술문을 중심으로 하는 일반경어법 점수체계와 요청구문을 중심으로 하는 특수경어법 점수체계로 이원화하여 살펴보는 것이 바람직할 것이다.

첫째, 일반경어법 점수체계에서 고려되는 요소는 높임의 선어말어미 '-시-'와 '해라체-해체-하계체-하오체-해요체-합쇼체'로 구분되는 화계이다. 먼저 화계의 점수 설정 방법부터 알아보자.

화계는 청자경어법의 대표적인 존대 등분 판별 요소로 어떠한 형태를 쓸 것이냐에 따라서 청자에 대한 존대 수준은 크게 달라지게 된다. 그런데 '하계체'와 '하오체'는 이익섭(1997)에서 언급한 바와 같이 화자와 청자가 중년층 이상의 연령이 되어야 쓰일 수 있다. 화자나 청자가 너무 어릴 때는 이 화계를 쓰는 것이 매우 어색한 것이다. 더군다나 최근에 들어서는 중년층 이상에서도 이 화계들은 잘 쓰이지 않고 있으며, 본고가 목표로 하고 있는 기계에 의한 경어법 문장의 자동 생성에서는 쓰일 일이 더욱 없을 것이다. 두 화계는 화자인 컴퓨터

를 높이면서 청자를 어느 정도 낮추는 말투가 되기 때문에 중년층 이상의 사람이 이 말을 들었을 때는 쉽게 감정이 상할 것이기 때문이다. 그러므로 이 두 화계는 점수 계산에서 배제하도록 한다. 화계 당 점수의 배당은 이정복(2001)의 연구를 참고하여 1점씩으로 하여 보면 다음 <표 5>와 같이 할 수 있다.

<표 5> 화계의 경어법 점수

| 화계 | 경어법 점수 |
|-----|--------|
| 해라체 | 0 |
| 해체 | 1 |
| 해요체 | 2 |
| 합쇼체 | 3 |

이에 따라 높임선어말어미 '-시-'의 경어법 점수 만점을 화계를 따라 3점으로 잡을 수 있다. 화계는 대화를 듣고 있는 청자를 높이는 것이고, '-시-'는 문장의 주체를 높이는 것이므로 둘의 범주는 서로 다르지만 모두 사람을 높이는 것이므로 만점 배당 기준을 같게 할 수 있기 때문이다. 이에 따라 일반경어법 점수체계를 정리하면 다음 <표 6>과 같다.

<표 6> 일반경어법 점수체계

| 주체 존대 | 청자 존대 | | | |
|-------|-------|-----|----|-----|
| | '-시-' | 해라체 | 해체 | 해요체 |
| 3 | 0 | 1 | 2 | 3 |

둘째, 특수경어법 점수체계에서 특수 경어법 요소에 속하는 서법과 의문형의 종류는 일반적인 상황에서는 적용이 되지 않다가 요청 구문이 발생하였을 때만 반영이 된다. 서법은 일반적으로 '서술법-약속법-의문법-명령법-청유법'의 5개가 있는 것으로 알려

저 있는데 이들은 화계와는 달리 서열 관계를 기본 속성으로 갖지는 않는다. 다만 발화의 종류가 어떠한 것이냐에 따라서 선택하는 서법의 종류가 달라지며 하나의 문장이 다른 서법으로 전환하는 것은 쉽게 일어나지 않는다. 그러나 상술한 바와 같이 요청과 관계된 구문에 있어서는 명령법과 의문법 두 가지 서법이 모두 사용이 가능한데 요청을 명령법으로 하게 되면 화자의 의지는 매우 강하게 전달되고 청자에게는 거의 선택권을 주지 않게 된다. 반면에 의문법을 쓰게 되면 청자의 의향을 묻는 형식을 띠게 되므로 화자의 강압적인 측면은 상당히 감소하고 청자의 선택권은 높아지게 된다.

단순의문형을 쓸 것이냐 부정의문형을 쓸 것이냐도 선택권에 영향을 주게 된다. 단순의문형에 비해 부정의문형은 상대방이 요청을 거절할 것을 염두에 두고 하는 질문이므로 듣는 사람에게 부담을 덜 주게 된다. Brown and Levinson(1987)은 부정형을 쓰면 상대방에게 덜 위협적인 요청을 할 수 있다고 하였다.

서법과 의문형의 종류는 서로 다른 범주에 속하는 것이지만 존대 또는 예의바름(politeness)의 관점에서는 연속적인 관계에 있다고 할 수 있다. 즉 명령법 보다는 의문형이 높고, 의문형 중 단순의문형보다는 부정의문형이 더 높은 존대등분(politeness level)을 갖는 것이다. 또한 이들은 청자 경어법의 존대 등분에 관여하는 요소이므로 화계와의 비교가 가능한데 화계가 청자경어법에서 가장 결정적인 영향을 미치는 요소라는 점을 생각하면 서법과 의문형의 종류는 그보다는 낮은 영향력을 가질 것이다. 화계가 각 단계마다 1점씩을 담당 받았으므로 서법과 의문형 종류는 이의 절반 정도의 영향을 가지도록 0.5점씩을 담당하면 <표 7>과 같이 특수경어법 점수체계가 구성된다.

<표 7> 특수경어법 점수체계

| 특수경어법 요소 | 점수 |
|----------|-----|
| 명령법 | 0 |
| 의문법 | 0.5 |
| 부정의문법 | 1 |

3.1.4 존대 등분 계산

위에서 도출된 결과에 따라 일반 경어법 점수 체계를 갖추고 있는 서술법 문장의 예를 만들어 보면 <표 8>과 같다.

<표 8> 일반 경어법 체계에 따른 서술법 문장의 예

| 점수 | 문장 | 화계 | '-시-' 유무 |
|----|------------|-----|----------|
| 0 | 일을 잘 했구나 | 해라체 | |
| 1 | 일을 잘 했어 | 해체 | |
| 2 | 일을 잘 했어요 | 해요체 | |
| 5 | 일을 잘 하셨어요 | 해요체 | 있음 |
| 3 | 일을 잘 했습니다 | 합쇼체 | |
| 6 | 일을 잘 하셨습니다 | 합쇼체 | 있음 |

'-시-'는 높임말에만 붙을 수 있으므로 해라체와 해체에는 적용되지 않는다. 또 합쇼체의 경우에는 명령법과 청유법에서 '-시-'를 제외한 형태를 상정하기 어렵다. 다음 <표 9>는 명령문의 경우의 경어법 체계이다.

<표 10>은 요청 구문에 대한 경어법 점수 체계이다. 요청 구문의 경어법 체계는 서법 상 명령법과 의문법만을 활용하며, 의문법의 종류로 일반의 문문과 부정의문문을 이용한다.

특수 경어법 요소의 점수 배당은 경어법에서 더 큰 영향력을 가지는 것으로 여겨지는 화계보다 낮게 설정되어 있기 때문에 전반적으로 화계의 질서를 깨뜨리지 않는다. <표 9>에서는 해요체와 합쇼체 사이에서 '-시-'의 유무에 따라 화계 질서가 지켜지지 않은 점수 배정도 보이는데 이 경우

<표 9> 일반 경어법 체계에 따른 명령법 문장의 예

| 점수 | 문장 | 화계 | '-시-' 유무 |
|----|-----------|-----|----------|
| 0 | 일을 잘 해라 | 해라체 | |
| 1 | 일을 잘 해 | 해체 | |
| 2 | 일을 잘 해요 | 해요체 | |
| 5 | 일을 잘 하세요 | 해요체 | 있음 |
| 3 | | 합쇼체 | |
| 6 | 일을 잘 하십시오 | 합쇼체 | 있음 |

에는 오히려 '-시-'가 있는 해요체가 '-시-' 없는 합쇼체보다 높은 존대 등분을 가지는 것이 직관적으로 옳아 보인다. 또한 두 요소가 합쳐졌을 때는 1점을 가지게 되어 있어 한 화계가 올라갔을 때의 효과를 보이게 되었다. 이것은 서로 다른 형식을 가지는 두 문장이 존대 정도가 비슷한 경우를 계량적으로 포착할 수 있게 해주었다.

이상으로 경어법 점수 체계를 마련하고 예문의 존대 등분을 산출함으로써 귀납적 검증을 시도하

였다. 제4장에서는 이러한 점수 체계를 근거로 하여 실험을 하도록 하겠다.

3.2 대화에서의 상황 차원의 고려

일반적으로 사람이 대화를 할 때 존대등분을 결정함에 있어서 영향을 줄 것으로 예상되는 상황 요소 후보는 다음과 같다.

- **화자의 특성:** 화자의 나이, 성별, 직업이나 신분, 성격 등은 존대 등분 결정에 영향을 줄 것으로 예상된다. 예를 들어 박영순(1976)은 화자의 성별에 따른 격식체 사용 양상을 조사하였다. 우리말의 화계는 단선적으로 생각할 수도 있지만, 해체 해요체는 격식성(formality)이 부족한 말투로, 해라체-하계체-하오체-합쇼체는 격식성을 갖춘 말투로 2원화해서 생각할 수도 있다. 위 연구에서 남자 사이에

<표 10> 특수 경어법 체계에 따른 문장의 예

| 점수 | 문장 | 화계 | '-시-' 유무 | 서법 | 의문문 |
|-----|-------------------|-----|----------|-----|-----|
| 0 | 나를 도와 줘라 | 해라체 | | 명령법 | |
| 0.5 | 나를 도와 주겠나? | 해라체 | | 의문법 | 일반 |
| 1 | 나를 도와 주지 않겠나? | 해라체 | | 의문법 | 부정 |
| 1 | 나를 도와 줘 | 해체 | | 명령법 | |
| 1.5 | 나를 도와 주겠니? | 해체 | | 의문법 | 일반 |
| 2 | 나를 도와 주지 않겠니? | 해체 | | 의문법 | 부정 |
| 2 | 나를 도와 줘요 | 해요체 | | | |
| 2.5 | 나를 도와 줄래요? | 해요체 | | 의문법 | 일반 |
| 3 | 나를 도와 주지 않으래요? | 해요체 | | 의문법 | 부정 |
| 5 | 나를 도와 주세요 | 해요체 | 있음 | 의문법 | |
| 5.5 | 나를 도와 주실래요? | 해요체 | 있음 | 의문법 | 일반 |
| 6 | 나를 도와 주지 않으실래요? | 해요체 | 있음 | 의문법 | 부정 |
| 6 | 나를 도와 주십시오 | 합쇼체 | 있음 | | |
| 6.5 | 나를 도와 주시겠습니까? | 합쇼체 | 있음 | 의문법 | 일반 |
| 7 | 나를 도와 주지 않으시겠습니까? | 합쇼체 | 있음 | 의문법 | 부정 |

서는 낮은 격식체의 말투가, 남자와 여자 사이의 대화에서는 높은 격식체의 말투가 사용된다고 하였으므로 이러한 점을 존대등분 계산에 반영할 수 있다.

- **시간(chronos)**: 아침, 점심, 저녁 등 대화 당시의 시간이 어떠한지가 영향을 줄 수 있다. 특히 지속적으로 만나는 사람이라고 할 지라도 그날 처음 만나는 아침 시간과 헤어지는 저녁 시간에는 인사말을 나눈다. 인사말은 일반적인 존대 수준보다 높은 합쇼체를 쓰는 경향이 있으므로 각 시간 대의 첫 대화와 마지막 대화는 조사의 대상이 될 수 있다.
- **위치**: 현재의 위치로서 강의실, 거실, 회사, 병원 등등 대화의 장소가 존대등분 결정에 영향을 줄 수 있다. 예를 들어 공중이 있는 곳에서는 사적인 감정을 드러내기가 어려우므로 낮춤말 보다는 존중하는 말을 하기가 쉬울 것이다. 또한 위치가 그의 직장이라고 한다면 사적인 대화를 나누듯이 대화를 하지 않을 것이며 나이가 어리다고 하더라도 존대 수준을 높일 것이다.
- **감정상태**: 평상시, 기쁨, 슬픔, 격분, 긴장 등 대화 당시의 화자 및 청자의 감정 상태가 존대 등분에 영향을 줄 수 있다. 예를 들어 위기 상황에서는 높임말보다는 단호함을 전달할 수 있는 평서문으로 명령을 해야 질서를 유지할 수 있는 것으로 알려져 있다. 그런데 본 연구에서는 인간과 기계사이의 인터페이스에 관심을 가지므로 인간으로부터 명령을 받는 기계의 감정 상태는 중요하지 않아 청자인 인간의 감정 상태에 대해서만 관심을 가지기로 한다.
- **활동**: 운동, 강의, 근무, 레저활동, 휴식, 식사,

데이트 등 화자가 현재 하고 있는 활동이 존대등분의 결정에 영향을 줄 가능성이 있다. 공적인 활동에서는 격식체가 많이 사용될 것이고, 사적인 활동에서는 비격식체가 많이 사용될 것이다.

- **제 3자의 특성**: 화자와 청자 사이에서 대화와는 직접 이해관계가 없으나 함께 듣고 있는 제 3의 청자의 특성이 대화 중 사용하는 존대등분에 영향을 줄 수 있다. 예를 들어 제 3자가 화자나 청자보다 높은 최고위자일 경우 두 사람은 화계 선택에 보다 민감하게 된다.

이와 같은 내용을 포함한 사례의 구조는 다음과 <표 11>과 같이 나타날 것이다. 여기서 화자와 청자의 특징으로는 나이, 직업, 성별, 중요도를 고려하였다. 또한 제 3자의 특징은 집단 여부, 나이, 직업, 성별, 중요도 등을 고려할 수 있다. 여기서 중요도란 기계 혹은 그 기계를 소유하는 서비스 제공자의 입장에서 얼마나 중요한 상대인지 하는 것이다. 단골, 명성을 가진 자 등이 이에 해당될 것이다. 그러나 본 연구에서는 단순화를 위해 제 3자의 존재 여부만으로 분석하기로 하였다.

사례의 수집은 드라마 대본에 의한 내용 분석 방법을 사용하였다. 드라마 대본을 활용한 방법은 국어정보학에서 즐겨 사용하는 자료 수집 방법이며, 특히 본 연구의 관점에서 보더라도 드라마 대본에는 극중 출연진에 대한 개인적 정보를 비교적 정확하고 충실하게 확보할 수 있으면서도 사례의 제 3자로의 활용에서 발생할 수 있는 사생활 침해 우려나 개인 정보의 익명화 등 사전 처리에 따른 정확도 감소나 비용 발생 등의 문제가 없어 좋은 측면이 많다. 또한 대본 상에는 상황 정보를 쉽게 파악할 수 있는 지문 등이 있어 편리하다. 또한 실질적으로 사례 수집 시 재현이 불가능하여 찰나의

<표 11> 사례베이스의 구조

| Category | Item Set | Item | Type |
|---------------|----------|------------------------|--------------------------------------|
| Input feature | Identity | 화자의 특징 (나이, 성별, 직위) | String array |
| | | 청자의 특징 (나이, 성별, 직위) | String array |
| | | 제3자 존재유무 | Binary |
| | Time | 대화 시간 | Time |
| | Location | 대화 위치 | String |
| | Emotion | 대화 시 청자 감정 | String |
| | Activity | 대화 시 활동 | ['public' 'protected' 'private'] |
| Result | | 종결어미 | ['해라' '해' '해오' '합쇼'] |
| | | -시- | Binary |
| | | 의문형 | Binary |

대화를 정확하게 반영하기 위해서는 동영상 촬영 등의 방법을 동원해야 하나 드라마 대본의 경우에는 이미 동영상에 준비되어 있고 드라마 대본이 확보된 관계로 따로 문자적 형태로 작성해야 하는 불편함도 없다. 또한 이미 웹사이트 등을 통해 많은 방송사들이 다양한 드라마의 대본을 동영상과 함께 공개하고 있으므로, 연구자는 다양한 상황 습득을 위해 복수개의 드라마에 다수의 장면들을 자유롭게 접근하여 볼 수 있다. 결국 드라마 대본에 의한 사례 수집법은 본 연구에 가장 최적의 방법으로 판단된다. 예를 들어, KBS 모 드라마의 179회분 씬4에 다음과 같은 대화가 있다.

<씬4> 진국 서재(밤)
 둘 잔잔하게 마주보다가
 희수(30대 주부, 평상): 아버님께 말씀 드려야
 하지 않을까요?
 진국(30대 남편, 평상): 아시겠지!

여기에서 희수의 대화는 다음 <표 12>와 같은 사례로 입력될 수 있다.

3.3 상황인식에 근거한 존대등분 결정

같은 존대등분의 문장이라고 하더라도 그 문장을 보는 사람의 상황에 따라 그 사람이 실제로 인지하는 존대등분은 달라질 수 있다는 가능성에 대해 살펴보았다.

인터페이스가 상황인식에 근거한 존대등분을 결정할 때 두 가지 대안을 제안할 수 있다. 첫째는 사례기반추론에 의한 방법이며, 두 번째는 상황인식에 의한 계산 보정방법이다. 본 논문에서는 이 중 사례기반추론에 의한 방법에 근거한 존대등분 계산법을 채택하였다.

사례기반추론에 의한 방법에 근거한 존대등분 계산법은 이는 사용자 프로필과 같은 비상황정보와 상황정보를 모두 고려한 사례들이 이미 존재한다고 가정할 때 사례기반추론을 활용하여 가장 유사한 사례를 추출하고 그에 맞게 존대등분을 결정한 후에 대화할 문장을 생성하여 사용자에게 제공하는 방법이다.

사례베이스에서 현재 상황에 가장 적합한 사례를 탐색하기 위한 방법으로 본 연구에서는 Minkowski

aggregation 방법을 응용하였다. Minkowski aggregation 방법은 가중평균방법에 비하여 p값이 1보다 큰 경우 유사하지 않은 항목의 영향을 더욱 부각시키는 일종의 가중평균법의 일반형이라고 볼 수 있다. p값이 1보다 작고 0보다 큰 경우는 당연히 차이가 나는 항목에 대한 영향력을 작게 하려는 경우이다. 본 연구에서는 이를 근간으로 하되 네 가지의 주요 평가 차원 수준에서 한번 적용하고, 그 차원 내에서도 한번 더 적용하는 이른바 Nested Minkowski aggregation 방법을 고안하였는데 계산 방법은 다음과 같다.

단계 1: 각 모듈별로 Minkowski aggregation 방법을 사용하여 모든 사례에 대해 유사도 계산(식 (2) 참조).

단계 2: 유사도 계산 결과 및 사례의 결과에 해당하는 부분을 최종사례 선정 주제인 존재등분 관리 에이전트에 전달

단계 3: 다시 Minkowski aggregation 방법을 사용하여 그룹별 유사도 계산한 것을 가지고 최종 유사도 계산하여 사례 선정(식 (1) 참조).

Nested Minkowski aggregation 방법의 유사도 계산 함수는 다음 식 (1)과 같다. 식 (1)에 존재하는 두 가지 부분은 각각 화자와 청자에 대한 유사도($s_{i,1}, s_{i,2}$) 및 대화 당시 상황 정보인 대화시간, 대화 장소, 대화 시 감정, 대화 중 활동, 제 3자의 존재에 대한 유사도($s_{c,1}, s_{c,2}, s_{c,3}, s_{c,4}, s_{c,5}$)를 의미한다. 화자와 청자에 대한 프로파일 정보 및 상황 정보를 모바일 환경에서 별도의 센서 혹은 모듈에서 관리한다고 보고 이들 각각에 대해서 유사도를 구한 후에 유사도 정보만을 가지고 존재등분 결정 에이전트에서 최종 유사도를 구하는 형식으로 계산하는 것이 Nested Minkowski aggregation

방법의 핵심이다. 즉, Nested Minkowski aggregation 방법을 새로이 개발한 이유는 유비쿼터스 컴퓨팅 환경이 기본적으로 분산화된 컴퓨팅 환경을 기반으로 하고 있고, 이에 따라 유사도 분석에 연관된 정보도 또한 기기나 서버에 따라 분산되어 있을 가능성이 크기 때문이다. 이러한 경우 어떤 한 기기 혹은 서버에서 사용자의 프로파일이나 상황 관련된 정보를 모두 수집하여 유사도 계산을 한다고 했을 때 개인 정보 침해의 우려가 높으며, 따라서 이러한 서비스는 수용되지 못할 것이다. 따라서 유사도 계산을 요청한 후에 각 기기 혹은 서버에서 자신이 가지고 있는 내용을 토대로 유사도 계산을 하게 하고, 그 후에 계산된 결과를 가지고 와서 최종 유사도를 계산하도록 하여 이러한 문제를 피하도록 한 것이다.

$$\Psi(s_{i,1}, s_{i,2}, s_{i,3}, s_{i,4}, s_{i,5}) = \left(\sum_{j=1}^2 w_j s_j^p \right)^{\frac{1}{p}} \quad (1)$$

한편 식 (1)에 존재하는 두 가지 유사도는 일반 Minkowski aggregation 방법을 따라 다음 식 (2)와 같은 함수로 유도된다. 단 w는 가중치이다.

$$s_1 = \left(\sum_{k=1}^2 w_{i,k} s_{i,k}^{p_1} \right)^{\frac{1}{p_1}} \quad (2)$$

$$s_2 = \left(\sum_{k=1}^5 w_{c,k} s_{c,k}^{p_2} \right)^{\frac{1}{p_2}}$$

4. 실험

본 연구에서 제안한 상황인식형 개인화된 인터페이스 전략의 유용성을 보이기 위해 실험실 실험을 하였다. 먼저 한국에서 2006년도 하반기에 방영되었던 드라마들 중에서 현대극으로 5개의 드

라마를 선정하고 각 드라마 안에서 10개의 방영분을 임의로 추출하였으며 그 방영 분에 등장하는 시의적절한 문장을 사례로 수집하였다. 여기서 시의적절한 문장이란 드라마 상에서 화자의 말에 대해서 청자가 분노 등 이상적 반응을 보이지 않은 문장을 의미한다. 이렇게 수집된 사례는 총 2,000건이었다.

2,000건의 사례 중에서 일부는 학습용으로, 일부는 테스트 용으로 나누어서 실험을 진행하였으며 화자는 컴퓨터, 청자는 인간으로 분류하였다. 그리고 사례기반추론법과 존대등분에 대한 기존 국어학에서의 연구 결과와의 비교를 위해 다음과 같이 네 가지 방법으로 존대등분 계산을 수행하였다.

- 방법 1:** 연령에 근거한 존대등분 계산(AGE)
- 방법 2:** 힘과 유대감에 근거한 존대등분 계산(POWER)
- 방법 3:** 상황인식형 개인화된 존대등분에 의한 계산(모든 사례 정보 활용, CONTEXT)
- 방법 4:** 상황인식형 개인화된 존대등분에 의한 계산(화자, 청자 특성만 활용, PROFILE)

이때 본 실험의 목적은 존대등분이 정확하게 계산되어 시의적절한 대화가 되었는지를 결정하는 존대등분의 정확도를 가지고 위의 네 가지 방법의 성능을 비교하는 것이다. 구현은 Java application으로 하였으며 이를 위해 Java SDK 1.4.x가 사용되었다. 사례베이스는 Microsoft Access 2003이 사용되었으며 결과도 같은 DB에 저장하였다. 본 실험은 2007년 4월과 5월 중에 실시되었으며 총 36,307회 반복 실험하였다.

성능 비교 평가를 위해서 사용한 두 가지의 성능 측정 기준은 각각 정확도(accuracy)와 일치도(hit ratio)였다. 정확도는 7점 만점의 존대 등분에

<표 13> 실험 방법 간 정확도에 대한 Paired samples t-test 결과

| | 평균 | 표준오차 | t-값 | 유의수준 |
|-----------------|------------|-----------|-----------|---------|
| AGE-POWER | -.00471845 | .01494526 | -60.158 | .000*** |
| AGE-CONTEXT | -.32714943 | .07837222 | -795.388 | .000*** |
| AGE-PROFILE | -.26231277 | .04509777 | -1108.306 | .000*** |
| POWER-CONTEXT | -.32243099 | .07887887 | -778.881 | .000*** |
| POWER-PROFILE | -.25759432 | .04547902 | -1079.246 | .000*** |
| CONTEXT-PROFILE | -.06483667 | .06164368 | 200.414 | .000*** |

<표 14> 실험 방법 간 일치도에 대한 Paired samples t-test 결과

| | 평균 | 표준오차 | t-값 | 유의수준 |
|-----------------|---------|--------|---------|---------|
| AGE-POWER | -.01080 | .32056 | -6.420 | .000*** |
| AGE-CONTEXT | -.08020 | .28658 | -53.328 | .000*** |
| AGE-PROFILE | -.00243 | .35258 | -1.312 | .190 |
| POWER-CONTEXT | -.06940 | .23046 | -57.382 | .000*** |
| POWER-PROFILE | .00837 | .28206 | 5.656 | .000*** |
| CONTEXT-PROFILE | .07778 | .25450 | 58.231 | .000*** |

대해서 실제 사례에서의 존대등분 점수와 추정된 점수 사이의 오차를 가지고 계산을 한 것이며, 일치도는 실제 사례에서의 존대등분 점수와 추정된 점수가 정확하게 일치하는 경우가 전체의 몇 %에 해당하는지를 계산한 것이다. 그리하여 다음과 같은 <표 13> 및 <표 14>와 같은 결과를 얻을 수 있었다.

<표 13>에 의하면 CONTEXT > PROFILE > POWER > AGE의 순서로 정확도가 우수하게 나온 것을 알 수 있다. 일치도에 의한 성능 비교도 <표 14>와 같이 비교적 비슷한 결과가 나왔는데, 재미있는 것은 방법 4(PROFILE)가 방법 3(CONTEXT) 뿐만 아니라 방법 1(AGE)이나 방법 2(POWER)에 대해서도 열등한 것으로 나타났다. 이 원인의 하나로 혹시 너무나 많은 학습용 사례를 가지고 학습하게 되는 이유로 overfitting되는 것이 아닌지

파악하기 위하여 우선 방법 3과 방법 4에 대해서 학습용 사례의 수에 따라서 성능이 어떻게 변경되는지를 파악하기 위해 학습용 사례를 40, 80, 120, 160, 200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800개로 선정하여 다음 [그림 2]와 [그림 3]과 같은 결과를 관측하였다.

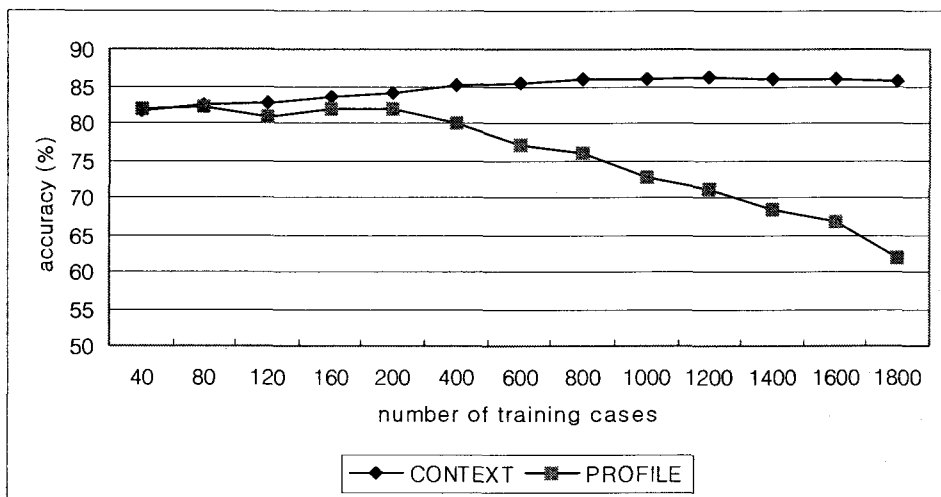
먼저 [그림 2]를 보면 방법 3(CONTEXT)의 경우 학습용 사례의 수의 증가에 따라서 정확도 또한 증가하는 것을 보여주고 있으나, 방법 4(PROFILE)의 경우 학습용 사례가 200개 정도를 넘어서면서 점차로 정확도가 낮아지는 것을 관찰할 수 있었다. 한편 [그림 3]에서는 두 방법 모두 학습용 사례의 수가 600개에 이를 때까지는 일치도가 증가하다가 그 이상이 되면 방법 3은 계속하여 일치도가 증가하는 데 반해 방법 4는 감소하기 시작하는 것을 알 수 있다. 이러한 관찰 결과를 토대로 할 때 방법 4는 학습용 사례의 수가 일정 규모 이상이 되면 오히려 성능이 떨어지는 이른바 overfitting 현상을 야기하고 있다. 이를 통계적으로 검증하기 위하여 학습용 사례의 수가 600미만인 경우에 대해

서만 Paired samples t-test를 검증을 하여 <표 15>와 같은 결과를 얻었다. <표 15>에서는 PROFILE이 AGE나 POWER에 비하여 통계적으로 우수한 성능을 보이고 있는 것을 알 수 있다.

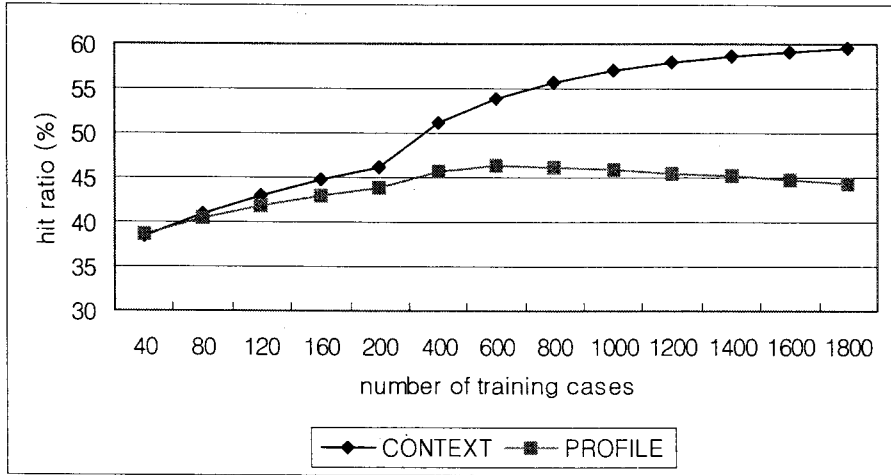
<표 15> 실험 방법 간 일치도에 대한 Paired samples t-test 결과

| | 평균 | 표준오차 | t-값 | 유의수준 |
|-----------------|---------|--------|---------|---------|
| AGE-POWER | -.00796 | .33135 | -3.291 | .001*** |
| AGE-CONTEXT | -.06630 | .28137 | -32.286 | .000*** |
| AGE-PROFILE | -.04805 | .31459 | -20.926 | .000*** |
| POWER-CONTEXT | -.05834 | .25075 | -31.879 | .000*** |
| POWER-PROFILE | -.04009 | .23814 | -23.064 | .000*** |
| CONTEXT-PROFILE | .01825 | .21018 | 11.900 | .000*** |

따라서 화자 및 청자의 나이, 성별, 지위 정도의 비교적 적은 상황 자료를 가지고 사례기반추론을 수행하려는 경우에는 적정규모의 사례의 수가 존재하며 그것은 당사자들의 프로파일 정보를 포함한 관련된 상황정보를 모두 고려하는 경우보다 적



[그림 2] 정확도 관점에서의 성능 비교



[그림 3] 일치도 관점에서의 성능 비교

은 수의 학습 사례만을 필요로 함을 알게 되었다. 또한 적절한 규모의 학습용 사례를 활용한다 하더라도 상황정보 전체를 고려하는 경우가 더 성능이 우수함을 알게 되었다. 결국 상황정보를 고려한 사례기반추론에 의한 존대등분 계산 방법은 기존의 국어학자들의 제안한 방법이나 사용자 프로파일 등을 활용한 방법 등 상황정보를 고려하지 않거나 제한되게 사용하는 경우보다 정확도 면에서 우수한 것으로 보인다.

한편 Minkowski aggregation 방법과 Nested Minkowski aggregation 방법은 기본 알고리즘이 동일하고 이를 분산적으로 수행한다는 면에서 차이가 나므로 정확도와 일치도 측면의 성능 상에서는 차이가 나지 않는다. 다만 유비쿼터스 센서 네트워크 및 무선망을 통하여 개인 정보가 기기들 사이에 송수신된다고 하는 사실을 시스템 사용자들이 원치 않는다면 Minkowski aggregation 방법은 사용할 수 없다. 즉, Nested Minkowski aggregation 방법은 성능 상의 문제가 아니라 사생활 정보에 대한 사용자 수용도 측면에서 일반 방

법보다 우수하다고 볼 수 있다.

5. 결론

본 연구는 HCI의 고유 연구분야에 국어정보학과 경영정보학 연구가 어떻게 접목될 수 있는지를 모바일 인터페이스라고 하는 영역에서 시도한 초기의 연구로서 학제적 공헌이 있다고 본다. 기존의 국어정보학에서 제안한 존대등분법 계산보다 상황 정보를 활용하게 되면 더욱 정확한 존대등분 계산이 가능함을 통계적으로 증명하였으며 이는 상황인식 서비스의 중요한 연구 분야로 향후 지속적인 연구가 필요함을 보여준다.

추후에는 다음과 같은 연구를 진행하려고 한다. 먼저 존대등분 선택 방법 알고리즘을 더욱 다양하게 검토할 것이다. 상황추론 방법 이외의 상황인식적 존대등분 계산법을 더 고려해 성능을 개선시킬 것이다. 둘째로 이모티콘, 배경색 등 정보디자인적 속성을 고려한 존대등분 보정이나 추론 방법

을 고려할 것이다.

참고문헌

- [1] 김민수, 국어문법론, 일조각, 1973.
- [2] 박영순, “국어 경어법의 사회언어학적 연구”, *국어국문학, 국어국문학회*, 72-73합권(1976), 47~65.
- [3] 성기철, “현대 한국어 대우법의 특성”, *외국어로서의 한국어교육*, 21권(1996).
- [4] 엄경옥, “현대국어 청자대우법 화계에 대한 고찰”, *어문론집, 중앙어문학회*, 30권(2002).
- [5] 오미정, “한국어의 존재 어휘 연구 -고유어를 중심으로-”, *한국어학, 한국어학회*, 27권(2005), 225~248.
- [6] 유송영, “국어 청자 대우법에서의 힘(power)과 유대(solidarity) (1) -불특정 청자 대우를 중심으로-”, *한국어학회 하계학술대회*(1993).
- [7] 이경우, “현대국어 경어법의 사회언어학적 연구(2)”, *국어교육, 한국어교육학회*, 106권(2001).
- [8] 이경우, “국어 경어법 변화에 대한 연구(1)”, *국어교육*, 110권(2003).
- [9] 이경우, “현대국어 경어법의 사회언어학적 연구(3)”, *국어교육*, 113권(2004).
- [10] 이익섭, 임홍빈, *국어문법론*, 학연사, 1997.
- [11] 이정복, “경어법 요소의 기능 부담량과 쓰임에 대하여”, *해양문학과 국어국문학-양천이용욱 교수 환력기념논총*, 형설출판사, 1993.
- [12] 이정복, “제 3자 경어법 사용에 나타난 참여자 효과 연구”, *국어학회*, 24권(1994).
- [13] 이정복, “대학생들의 제 3자 경어법 사용에 나타난 참여자 효과”, *한국문화, 서울대 한국문화연구소*, 18권(1996).
- [14] 이정복, *국어 경어법 사용의 전략적 특성*. 태학사, 2001.
- [15] 최현배, *우리말본*. 정음사, 1937.
- [16] Brown, P. and Levinson, S., *Politeness: Some universals in language usage*. Cambridge, Cambridge University Press, 1987.
- [17] Brown, R. and Gilman, A., “The Pronoun of Power and Solidarity”, *American Anthropologist*, Vol.4, No.6(1960), 252~282.
- [18] Jones, M., Marsden, G., Mohd-Nasir, N., and Boone, K., “Improving web interaction on small displays”, *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Vol. 31, No.11(1999), 1129~1137.
- [19] Mignonneau, L. and Sommerer, C., “Designing Emotional, Metaphoric, Natural and Intuitive Interfaces for Interactive Art, Edutainment and Mobile Communications”, *Computers & Graphics*, Vol.29, No.6(2005), 837~851.
- [20] Nielsen, C. and Sndergaard, A., “Designing for Mobility: an Integration Approach Supporting Multiple Technologies”, *Proceedings of the 1st Nordic Conference on Human-Computer Interaction(CD-ROM)* (2000).
- [21] Shirado, T. and Isahara, H., “Numerical Model of the Strategy for Choosing Polite Expressions”, *Proceedings of Computational Linguistics and Intelligent Text Processing, CICLing*, Vol.2004(2001), 98~109.
- [22] Sommerer C. Mignonneau L., “Mobile feelings,” *CODE-The Language of our Time, Ars Electronica*(2003), 258~261.
- [23] Väänänen-Vainio-Mattila, K. and Haataja, S., “Mobile Communication User Interfaces for Everyone”, *Advances in Human Factors /Ergonomics 21B, Design of Computing Systems: Social and Ergonomic Considerations* (1997), 815~819.
- [24] Väänänen-Vainio-Mattila, K. and Ruuska, S., “User Needs for Mobile Communication

Devices: Requirements Gathering and Analysis
Through Contextual Inquiry”, *Proceedings*

*of the First Workshop on HCI for Mobile
Devices, glasgow, 21st&22(1998).*

Abstract

Applying Polite Level Estimation and Case-Based Reasoning to Context-Aware Mobile Interface System

Kwon Ohbyung* · Choi, Sukjae** · Park Tae Hwan*

User interface has been regarded as a crucial issue to increase the acceptance of mobile services. In special, even though to what extent the machine as speaker communicates with human as listener in a timely and polite manner is important, fundamental studies to come up with these issues have been very rare. Hence, the purpose of this paper is to propose a methodology of estimating politeness level in a certain context-aware setting and then to design a context-aware system for polite mobile interface. We will focus on Korean language for the polite level estimation simply because the polite interface would highly depend on cultural and linguistic characteristics. Nested Minkowski aggregation model, which amends Minkowski aggregation model, is adopted as a privacy-preserving similarity evaluation for case retrieval under distributed computing environment such as ubiquitous computing environment. To show the feasibility of the methodology proposed in this paper, simulation-based experiment with drama cases has performed to show the performance of the methodology proposed in this paper.

Key Words : Politeness Level, HCI, Context-Aware Computing, Mobile Services, Case-Based Reasoning, Minkowski Aggregation Method

* School of International Management, Kyunghee University

** Department of Korean language and literature, Korea University