

# 다중모드 대화 시스템에서 이중 캐시 모델의 센터링 알고리즘을 이용한 명사 대응어구 처리

## (Resolution of Anaphoric Noun Phrases using a Centering Algorithm with a Dual Cache Model in a Multimodal Dialogue System)

김학수<sup>†</sup> 서정연<sup>††</sup>  
(Harksoo Kim) (Jungyun Seo)

**요약** 다중모드 대화에서 나타나는 대응어는 언어만을 사용하는 대화에서 나타나는 것과 비교하여 매우 다른 형태와 특징을 가진다. 그것은 행위나 시각이 대응 행위로 사용될 수 있기 때문이다. 본 논문에서는 터치스크린 인터페이스를 이용한 홈쇼핑 가구집 영역의 다중모드 대화 시스템에서 나타나는 다양한 대응어의 처리 방법을 알아본다. 먼저, 화면 대응어와 참조 대응어를 정의하여 다양한 형태의 대응어를 분류한다. 그리고 각 대응어를 처리할 수 있는 두 가지의 일반적인 방법을 제안한다. 하나는 지시 행위를 수반하거나 생략한 채 발화되어 현재 화면에 나타나 있는 아이템을 참조하는 대응어를 처리하는 단순한 매핑 알고리즘이다. 다른 하나는 다중 모드 대화 시스템을 위해 워커(Walker)의 센터링 알고리즘을 확장한 이중 캐시 구조의 센터링 알고리즘이다. 확장된 센터링 알고리즘은 발화와 시각 정보 그리고 화면 전환 시간을 유지할 수 있기 때문에 다중모드 대화에서 발생하는 다양한 대응어를 처리하기에 적합하다. 실험에서 제안된 시스템은 40개의 대화에서 나타난 402개의 대응어(발화당 0.54개)중에서 387개를 처리하여 96.3%의 정확도를 보였다.

**Abstract** Anaphora in multimodal dialogues have certain distinct properties compared to the anaphora in language-only dialogues. They often refer to the items signified by gestures or visual means. In this paper, we propose anaphora resolution algorithms for a multimodal dialogue system as applied to home shopping domain, where a user purchases pieces of furniture using Korean utterances with pointing gestures on a touch screen. We define two kinds of anaphora, screen anaphora and referring anaphora, and propose two general methods to resolve these anaphora. One is a simple mapping algorithm that can find items referred to with/without pointing gestures on a screen. The other is the centering algorithm with a dual cache model, in which Walker's centering algorithm is extended for a multimodal dialogue system. The extended algorithm makes an adequate resolution to resolve various anaphora in a multimodal dialogue since it keeps utterances, visual information and screen switching-time. In the experiments, the system correctly resolved 387 anaphora out of 402 anaphora in 40 dialogues (0.54 anaphora per utterance) showing 96.3 percent correctness.

### 1. 서론

본 연구는 과학재단 특정과제(과제번호:97-0102-03-01-3)의 지원으로 이루어진 것입니다.

† 비회원 : 서강대학교 컴퓨터학과  
hskim@nlpzodiac.sogang.ac.kr

†† 종신회원 : 서강대학교 컴퓨터학과 교수  
seojoy@ccs.sogang.ac.kr

논문접수 : 1999년 10월 20일  
심사완료 : 2000년 7월 27일

인간의 대화에서 나타나는 주요 특징 중의 하나는 대화 채널의 다양성이다. 즉, 대화는 발화와 함께 제스처, 표정들이 섞여서 진행된다. 다중모드 대화 시스템(multimodal dialogue system)이란 사용자와 시스템이 발화만을 통하여 상호 작용하는 단일모드 대화 시스템(language-only dialogue system)과는 달리 발화, 표정, 제스처 등의 다양한 입력 채널을 갖는 대화를 처리할 수 있는 시스템이다. 이러한 다중모드 대화 시스템은

사용자에게 보다 편리하고 자연스러운 인터페이스를 제공하지만 정확하게 사용자의 의도를 파악하기 위해서는 보다 다양하고 복잡한 대응어를 처리해야만 한다. 본 논문에서는 터치스크린 인터페이스를 이용한 홈쇼핑 가구점 영역의 다중모드 대화 시스템[1]에서 나타나는 다양한 대응어들의 처리 방법을 알아본다. 다중모드 대화 시스템의 입력은 발화를 받아들이는 청각 채널(auditory channel)과 터치스크린(touch screen)을 이용하여 지시 행위(pointing gesture)를 받아들이는 시각 채널(visual channel)로 나뉘어 진다. 그러므로 사용자는 [그림1]의 (4)번 발화와 같이 화면에 제시된 *모델200*을 가리키며, “이건 얼마예요?”라고 물을 수 있다. 그리고 사용자에게 제시된 화면에 빨간 물건이 유일할 경우에 가리키는 행위를 생략하고 (8)번 발화와 같이 “빨간 건 얼마예요?”라고 할 수도 있다. 또한, 이전에 언급했던 물건을 다시 지칭하기 위해서 (10)번 발화의 *아까 그거*와 같은 대응어를 사용할 수 있다.

- 
- (1)시스템: 무엇을 도와드릴까요?  
 (2)사용자: 책상을 좀 보고 싶는데요.  
 (3)시스템: (*모델200*과 *모델250*을 보여주며)  
     이런 것들이 있습니다.  
 (4)사용자: (*모델200*을 가리키며)  
     이건 얼마예요?  
 (5)시스템: 150,000원입니다.  
 (6)사용자: 의자도 좀 보고 싶는데요.  
 (7)시스템: (*모델100*과 *모델150*을 보여주며)  
     이런 것들이 있습니다.  
 (8)사용자: *빨간 건* 얼마예요?  
 (9)시스템: 80,000원입니다.  
 (10)사용자: (*모델100*을 가리키며)  
     *아까 그거*랑 *이거* 주세요.
- 

그림 1 다중모드 대화의 예1

다중모드 대화 시스템에 관한 기존의 연구들은 지시 행위와 지시 표현(deictic expression) 사이의 연관 관계를 찾는 것[2,3,4,5,6], 또는 미리 정의된 심볼과 명령을 매핑하는 것[7]이었다. 그러나 그것 중에 어느 것도 (8)번 발화의 *빨간 것*과 같은 대응어의 처리에 대해서는 다루지 않았다. 또한, 이전에 언급한 발화나 화면에 있던 참조 대상을 다시 지칭하는 대응어를 고려하지 않았다. 그러나 다중모드 대화 시스템에서 사용자의 의도를 정확하게 파악하고 적절한 대답을 생성하기 위해서는 위에서 언급한 대응어들을 모두 처리하는 것이 필수

적이다.

[그림1]에서 지시 행위와 함께 나타난 (4)번 발화의 *이것*이나 지시 행위가 생략된 (8)번 발화의 *빨간 것*과 같은 표현의 지시 대상을 찾기 위해서는 지시 행위의 개수와 대응어의 개수가 유용하게 이용될 수 있다. 만약 지시 행위의 개수와 대응어의 개수가 같다면 1대 1로 매핑하여 간단하게 대응어를 처리할 수 있다. 만약 그렇지 않다면 매핑을 위한 몇 가지 휴리스틱이 제안될 수 있다.

참조 표현(referring expression)을 처리하기 위한 방법 중에 가장 잘 알려진 것이 Grosz, Joshi, Weinstein의 센터링 이론(centering theory)[8]이다. 센터링 이론은 Brennan, Friedman, Pollard에 의해서 대명사의 지시 대상을 찾기 위한 센터링 알고리즘(centering algorithm)[9]으로 발전되었고, 워커(Walker)[10]에 의해서 개선되었다. 그러나 기존의 스택이나 캐시 구조의 센터링 알고리즘은 다중모드 대화의 대응어를 처리하기에는 부족한 면이 많다. 또한, [그림1]의 (10)번 발화 “*아까 그거*랑 *이거* 주세요.”에 나타난 *아까 그거*를 처리하기 위해서는 ‘*아까*’라고 하는 수식어가 갖는 시간적인 기준 즉 화면이 전환된 시점을 알아야 하지만 기존의 센터링 알고리즘에서는 그것이 모델링되어 있지 않다. 본 논문에서는 이런 종류의 대응어를 처리하기 위해서 워커(Walker)의 센터링 알고리즘을 확장한 이중 캐시 구조의 센터링 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 먼저, 2장에서 접근 방법을 보이고, 3장에서는 다양한 형태의 대응어들을 처리하기 위한 2가지 방법을 제안한다. 4장에서 제안된 방법의 유용성을 평가하기 위한 실험을 한 후, 5장에서 결론을 맺는다.

## 2. 접근 방법

본 논문에서는 대응어를 화면 대응어(screen anaphora)와 참조 대응어(referring anaphora)로 분류한다. 화면 대응어는 지시 행위나 시각 채널을 통해 현재 사용자가 보고 있는 화면 위의 대상을 지칭하는 대응어이다. 예를 들어, [그림1]의 (4)번 발화 “*이건 얼마예요?*”에 나타난 *이것*이나 (8)번 발화 “*빨간 건 얼마예요?*”의 *빨간 것* 등이 여기에 속한다. 참조 대응어는 이전의 발화에서 언급했거나 이전의 화면에 나타났던 대상을 다시 지칭하는 대응어이다. 예를 들어, (10)번 발화 “*아까 그거*랑 *이거* 주세요.”의 *아까 그거*와 같은 것이 여기에 속한다.

화면 대응어 처리 알고리즘은 사용자가 행한 지시 행

위의 개수와 발화에 나타난 대응어의 개수를 계산하고, 그것들을 비교한다. 만약, 개수가 같으면 1대 1 매핑을 통하여 대응어를 처리한다. 그렇지 않으면, '이것'이라는 대응어는 지시 행위를 꼭 수반하며, '그것'은 수반하지 않을 수도 있다는 휴리스틱과 문장 내에서의 발화된 순서를 이용하여 대응어에 우선 순위를 부여하고, 그 순위에 따라 처리한다.

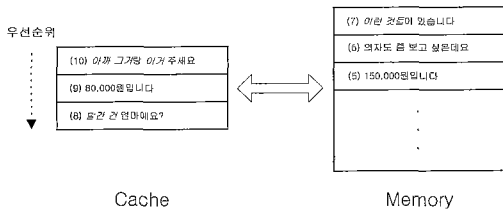


그림 2 워커(Walker)의 캐쉬 모델

참조 대응어 처리 알고리즘은 워커(Walker)의 센터링 알고리즘을 기반으로 한다. 센터링 알고리즘[8,9,11]은 Forward-looking Center, Preferred Center, Backward-looking Center라는 3가지 주요 요소로 구성된다. Forward-looking Center(이하 *Cf*)는 각 발화에서 나타난 명사구와 같은 담화 요소의 집합으로 문법적 역할(grammatical role)에 따라 우선 순위가 결정된다. 즉 주체에 해당하는 명사구는 주어보다 우선 순위가 높고, 주어는 목적어보다 높다. 예를 들어, 사용자가 “친구가 의자를 여기서 샀습니다.”라고 발화를 했다면, ‘친구’는 주어에 해당하고 ‘의자’는 목적어에 해당하므로 ‘친구’가 높은 우선 순위를 가진다. *Cf*중에서 가장 우선 순위가 높은 것을 Preferred Center(이하 *Cp*)라고 한다. 위의 예에서 ‘친구’라는 담화 요소가 그 발화의 *Cp*이다. Backward-looking Center(이하 *Cb*)는 현 발화에 선행되는 담화에 연결되는 요소로서 이전 발화의 *Cf*들 중에서 현재 발화에서 실현된 가장 높은 우선 순위의 담화 요소이다. 예를 들어, 사용자가 “친구가 의자를 여기서 샀습니다.”라고 발화를 한 후에 “그는 50,000원을 주고 샀습니다.”라는 했다면, 이전 발화의 ‘친구’를 의미하는 ‘그’가 바로 *Cb*이다. 본 논문에서는 *Cf*를 추출하기 위하여 구문 분석기의 결과로부터 명사구를 추출한다. 센터링 알고리즘은 *Cf*, *Cp*, *Cb* 외에도 제약 조건과 규칙 그리고 전이 유형(transition state)으로 이루어진다. 제약 조건, 규칙, 전이 유형은 현재 발화의 *Cb*와 *Cf*를 결정하여 대응어를 처리할 수 있도록 한다. 일반적인 기존의 센터링 알고리즘은 센터를 하나의 담화 세그먼트

(discourse segment) 안에서 나타나는 현상[9,12,13]으로 생각하여 스택을 기반으로 구현되었다. 그러나 워커(Walker)는 스택 기반의 모델을 수정하여 캐시 모델 기반의 센터링 알고리즘을 제안하였다. 워커(Walker)가 제안한 캐시 모델의 요점은 다음과 같다. 모든 발화는 하나의 담화가 끝났어도 다시 참조될 때를 대비하여 일정한 크기만큼의 캐시와 메모리에 유지되어야 하며, 필요에 따라 검색될 수 있어야 한다[10]. [그림2]는 워커(Walker)가 제안한 캐시 모델에 [그림1]의 대화를 적용해 본 결과이다. 캐시와 메모리에는 각 발화의 센터들이 들어가도록 되어 있지만 [그림2]에서는 이해를 쉽게 하기 위해서 발화로 대체하였다. [그림1]의 (10)번 발화 “아까 그거랑 이거 주세요.”에 나타난 아까 그것의 지시 대상을 찾기 위해 워커(Walker)의 캐시 모델을 그대로 적용한다면, (8)번 발화인 “빨간 건 얼마예요?”의 센터들을 참조하게 되어 올바른 결과인 ‘모델200’이 아니라 ‘모델100’이 된다. 그것은 수식어 ‘아까’에 대한 명확한 시간적 기준(time base)이 없기 때문이다. 본 논문의 다중 모드 대화 시스템은 사용자가 현재 화면에 없는 다른 상품을 요구할 때 현재 화면을 새로운 화면으로 대체한다. 그러므로 아까 그것이라는 표현은 현재 화면에 있는 것이 아닌 이전의 화면에 있는 것들 중에 하나로 정해진다. 이렇게 ‘아까’에 대한 시간적 기준이 더 명확해지기 때문에 센터링 알고리즘을 다중모드 대화 시스템에 적용하기 위해서는 새로운 캐시 구조가 필요하다. 본 논문에서는 시간적 기준을 명확히 하기 위해서 시스템과 사용자 사이의 발화들을 시스템이 제공한 화면 단위로 분류하여 저장하는 이중 캐시 모델(dual cache model)을 제안한다. 이중 캐시 모델은 각 발화의 센터들과 시각 정보(visual information) 그리고 화면 전환 시간(screen switching time)을 유지한다. 시각 정보는 화면에 나타난 목록들의 모델 번호, 색상, 크기, 모양 같은 시각적 특징을 의미하며, 화면 전환 시간은 시스템이 이전의 화면을 현재 화면으로 바꾼 시간을 말한다. [그림3]은 [그림1]의 대화를 이중 캐시 모델에 적용한 결과이다. [그림2]와 마찬가지로 센터들은 발화로 대체하였다. [그림2]에서 캐시와 메모리에 저장되는 각 발화 단위의 센터들을 살펴보면 나중에 발화된 문장에 속한 센터들이 먼저 발화된 문장에 속한 것들 보다 항상 상위에 위치하지는 않는다는 것을 알 수 있다. 본 논문에서는 각 발화들 사이의 우선 순위를 결정하기 위해서 ‘더 많은 정보를 가지고 있는 발화가 더 높은 우선 순위를 가진다.’는 휴리스틱을 제안한다. 이것은 다중모드 대화에서 나타나는 담화 요소가 가지는 정보량의 차이를

고려하고자 하는 것이다. 만약, 같은 양의 정보를 포함하고 있으면 나중에 발화한 것이 높은 우선 순위를 가진다. 예를 들어, [그림1]의 (8)번 발화 “빨간 건 얼마예요?”는 (9)번 “80,000원입니다.”보다 먼저 발화하였지만

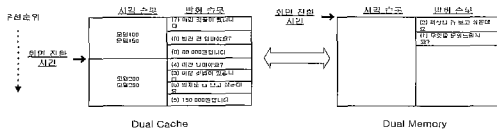


그림 3 이중 캐시 모델

(8)번이 시각 정보를 수반하므로 [그림3]에서 보는 것과 같이 보다 높은 우선 순위를 가진다. (7)번 발화 “이런 것들이 있습니다.”의 경우는 지시 행위를 수반하므로 (8)번이나 (9)번 보다 높은 우선 순위가 된다. 지금까지 살펴본 것을 바탕으로 [그림1]의 (10)번 발화 “아까 그거랑 이거 주세요.”에 나타난 아까 그것의 지시 대상을 찾기 위해서 [그림3]을 적용하면, 아까 그것에 대한 시간적 기준이 명확해 진다. 그러므로 지시 대상은 (4)번 발화 “이건 얼마예요?”에 나타난 모델200이 된다.

### 3. 대용어 처리 알고리즘

#### 3.1 화면 대용어 처리 알고리즘

화면 대용어 처리는 사용자가 발화를 할 때 나타난 대용어와 행위 정보를 매핑하거나 사용자에게 제공된 시각적 정보 중의 하나와 매핑하는 것이다. 예를 들어, 사용자가 의자를 가리키며 “이것과 빨간 의자를 주세요.”라고 한다면, 발화 중에 나타난 이것은 사용자가 가리킨 화면 위의 물건을 의미한다. 그리고 빨간 의자는 사용자에게 제공된 화면 위에 존재하는 빨간 색상을 가진 의자이다. 이렇게 발화를 할 때 사용자에게 제공되고 있는 정보들과 대용어 사이의 관계를 살펴서 지시 대상을 찾는 것을 화면 대용어 처리라고 한다. 화면 대용어 처리 과정은 발화에 포함된 대용어의 개수와 지시 행위의 개수에 따라 다음과 같이 3가지 종류로 분류된다.

가. 대용어의 개수와 지시 행위의 개수가 같을 때

이런 경우가 다중모드 대화 시스템에서 가장 빈번히 관찰된다. 대용어의 개수와 지시 행위의 개수가 일치하는 경우에는 발생한 대용어의 순서에 따라 화면 대용어를 지시 행위로 대체한다. 예를 들어, [그림1]의 (4)번 발화 “이건 얼마예요?”에서 이것은 모델200으로 간단히 결정된다.

나. 대용어의 개수가 지시 행위의 개수보다 많을 때

발화에 포함된 대용어의 개수가 지시 행위의 개수보다 많은 경우는 시각적 정보의 영향을 받아 행위를 생략하거나, 이전 발화의 지시 대상을 참조할 때 발생한다. 시각적 정보의 영향을 받아 지시 행위를 생략한 경우는 화면에서 지시 대상을 쉽게 결정할 수 있다. 그러나 [그림 1]의 (10)번 발화 “아까 그거랑 이거 주세요.”에 나타난 아까 그것을 해결하기 위해서는 이전의 발화들을 참조해야 한다. 이런 종류의 대용어는 3.2절에서 살펴볼 참조 대용어 처리 알고리즘에서 처리한다.

다. 대용어의 개수가 지시 행위의 개수보다 적을 때

발화에 포함된 대용어구의 개수가 지시 행위의 개수보다 적은 경우는 대용어구의 생략에 의해 발생한다. 대용어구의 생략은 발화의 일부가 생략된 경우와 전체가 생략된 경우로 나뉜다. 발화의 일부가 생략된 경우에는 필수격의 생략 여부를 검사하여 격을 보충해 준다. 그리고 전체가 생략된 경우에는 “이것이요.”나 “이것들이요.”를 생략했다고 보고 해당하는 새로운 의미 분석 결과를 생성한다. 이렇게 갱신된 의미 분석 결과는 (가)처럼 처리한다. 발화 중에 자주 생략되는 격은 행위의 대상을 지칭하는 대상격(affected entity)이나 성격을 규명하는 주제격(theme)이다. 예를 들어, 사용자가 모델100을 가리키며 “얼마예요?”라고 했다면, 이것은 주제격인 이것이 생략되어 있는 문장이다. 그러므로 이것을 보충하여, “이것이 얼마예요?”와 같은 의미 분석 결과를 생성하고 (가)의 방법에 따라 처리한다.

#### 3.2 참조 대용어 처리 알고리즘

참조 대용어 처리는 이전의 발화나 이전의 화면에서 지시 대상을 찾는 것이다. [그림1]의 (10)번 발화에서 사용자는 모델100을 가리키며, “아까 그거랑 이거 주세요.”라고 말을 했다. 이 때에 이것에 대한 지시 대상은 화면 대용어 처리 과정을 통해 모델100으로 결정된다. 그러나 아까 그것과 같은 대용어는 이전의 발화를 살펴 보지 않으면 처리할 수 없다. 참조 대용어 처리는 아까 그것과 같이 이전 발화나 화면의 정보를 이용해야 처리할 수 있는 대용어의 지시 대상을 찾는 과정이다.

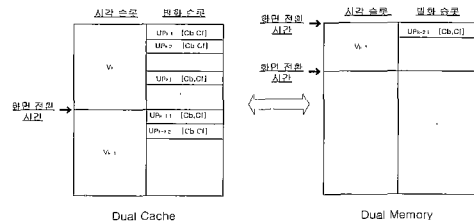


그림 4 이중 캐시 모델의 구조

본 논문에서는 센터링 알고리즘을 이용하여 [그림1]의 (10)번 발화 “*아까 그거랑 이거 주세요.*”에 나타난 *아까 그것*과 같은 참조 대응어를 처리할 수 있도록 [그림4]와 같은 이중 캐시 모델을 제안한다. 이중 캐시 모델은 발화 슬롯(utterance slot)과 시각 슬롯(visual slot) 그리고 화면 전환 시간으로 구성된다. 발화 슬롯은 각 발화의 센터들을 저장하며, 시각 슬롯은 화면에 보여진 목록들의 시각 정보를 담고 있다. 화면 전환 시간은 화살표로 표시되어 있으며, 이전의 화면이 현재 화면으로 전환된 시간을 유지한다. [그림4]에서  $V_k$ 는 k번째 화면의 시각 정보를 담고 있는 k번째 시각 슬롯을 의미한다.  $U_{P_{k,j}}$ 는 k번째 시각 슬롯에서 j번째 우선 순위를 갖는 발화를 의미한다. 센터링 알고리즘의 중심 요소인  $C_b$ 와  $C_f$ 의 쌍은 일반적으로 앵커(anchor)라고 불린다. 단일모드 대화 시스템의 경우, 앵커들은 스택의 운용 방식인 LIFO(Last-In-First-Out) 메커니즘(mechanism)으로 우선 순위가 결정된다. 하지만 다중모드 대화에서 발화는 수반하는 정보의 양이 서로 다르므로 LIFO를

논문에서는 이중 캐시 모델에 속하는 발화들의 우선 순위를 결정하기 위해서 [그림5]와 같은 우선 순위 규칙을 제안한다. 같은 우선 순위의 발화들은 LIFO 메커니즘을 따른다. 이중 캐시 모델에서 시각 정보를 유지하는 이유는 사용자가 선택을 하거나 발화를 한 적이 없는 이전 화면에 있던 물건을 선택하는 경우를 처리하기 위해서다. 즉, [그림6]의 (6)번 발화에 나타난 *아까 빨간 것*과 같은 종류의 대응어를 처리하기 위해 필요하다.

본 논문에서  $C_f$ 들의 우선 순위도 [그림5]를 따른다. 만약,  $C_f$ 들이 같은 우선 순위를 가진다면 알고리즘은 본 용언이 갖는 문장 성분들의 문법적 역할에 따라 주체, 주어, 목적어 순서로 우선 순위를 부여하고, 다른 문장 성분들은 문장에 나타난 순서대로 먼저 나타난 것에 높은 우선 순위를 부여한다[14,15]. 센터링 알고리즘은  $C_b$ 와  $C_f$  뿐만 아니라 제약 조건과 규칙으로 이루어진다. 본 논문에서는 캐시가 이중 구조로 바뀌었기 때문에 [그림7]과 같은 새로운 제약 조건과 규칙을 제안한다. [그림7]은  $C_A(U_{i-1})$ 이  $C_A(U_{P_{k,j}})$ 로 바뀐 것과 4번째 제약 조건이 추가된 것을 제외하면 기존의 센터링 알고리즘과 동일하다.  $C_A(U_{i-1})$ 이  $C_A(U_{P_{k,j}})$ 로 바뀐 것은  $C_b(U_i)$ 가 이전의 발화  $U_{i-1}$ 에서 실현되지 않기 때문이다. 예를 들어서, [그림1]의 (10)번 발화 “*아까 그거랑 이거 주세요.*”에 나타난 *아까 그것*의 지시 대상은 이전 화면에서 사용자가 발화한 것들 중에 있을 것이다. 그러므로  $C_b(U_i)$ 는  $C_A(U_{P_{k,j}})$ 에서 실현되어야 한다.  $C_A(U_{P_{k,j}})$ 에서 k-1은 이전의 시각 슬롯을 의미하고, j는 우선 순위 규칙에 의해 결정된 j번째 발화를 의미한다. 다시 말해서, 시스템은 *아까 그것*의 ‘*아까*’라는 수식어가 이전의 화면을 가리킨다는 것을 결정하고 그 다음에 우선 순위에 따라 처리한다. 4번째 제약 조건의 의미는 사용자의 현재 지시 행위가 이전의 발화와는 연관성이 없다는 것이다. 이것은 화면 대응어 처리 알고리즘에서 이미 처리된 대응어를 참조 대응어 처리 알고리즘에서는 대응어가 아닌 일반적인 명사구로 취급되도록 하기 위한 것이다.

[그림7]과 같은 이유에서 센터들 사이의 전이 유형은 [그림8]과 같이 정의된다.

제약 조건, 규칙, 전이 유형 등의 변화에 기초하여 참조 대응어 처리 알고리즘은 다음과 같은 과정을 따른다. 먼저 앵커  $C_b(U_{P_{k,j}})$ 와  $C_f(U_{P_{k,j}})$ 를 선택하고, 현재 발화  $U_i$ 에서 모든 가능한 앵커들을 구성한다.  $U_{P_{k,j}}$ 를 결정하기 위해서는 대응어의 수식어를 참조하여 k번째 시각 슬롯을 결정하고, 그 슬롯과 연결되어 있는 발화 슬

● 우선 순위 규칙(priority rule)

더 많은 정보를 수반하는 발화가 더 높은 우선 순위다. 즉, 여러 가지 정보가 동시에 수반된 어구가 앞으로 대화에서 초점으로 등장할 가능성이 높다. 그러므로 발화 사이의 우선 순위는 다음과 같이 결정된다.

행위 정보 수반 대응어 > 시각 정보 수반 대응어 > 기타 정보를 수반하지 않은 대응어

그림 5 발화 사이의 우선 순위

- (1)시스템: 무엇을 도와드릴까요?
- (2)사용자: 책상을 좀 보고 싶네요.
- (3)시스템: (모델100과 모델250을 보여주며)  
이런 것들이 있습니다.
- (4)사용자: 의자도 좀 보고 싶네요.
- (5)시스템: (모델100과 모델150을 보여주며)  
이런 것들이 있습니다.
- (6)사용자: (모델100을 가리키며)  
*아까 빨간 것*과 *이거* 주세요.

그림 6 다중모드 대화 예2

따르는 것이 적절하지 않다. 예를 들어, [그림1]의 (4)번 발화 “*이건 얼마예요?*”는 지시 행위를 수반하며, (8)번 발화 “*빨간 건 얼마예요?*”는 시각 정보를 수반한다. 본

를 우선 순위에 따라 탐색한다. 즉, [그림1]의 (10)번 발화 “*아까 그거랑 이거 주세요.*”에 나타난 *아까 그것의* 대응어를 찾기 위해서는 먼저 ‘*아까*’의 의미를 파악하여 바로 이전의 시각 슬롯을 결정하고, 그것과 연결된 발화

● 제약 조건

담화내의 각 발화에 대하여:

1. 각 발화  $U_i$ 는 하나의  $Cb$ 를 가진다.
2. 각 발화에 대한  $C\mathcal{A}(U_i)$ 의 모든 원소는  $U_i$ 에서 구현되어야 한다.
3. 각 발화의  $Cb(U_i)$ 는  $U_i$ 에서 구현된  $C\mathcal{A}(U_{p_{k,i}})$  중에서 가장 우선 순위가 높다.
4. 시각 정보나 행위 정보는 발화와 동시에 얻어지는 정보이므로 이전 발화와와의 연관성은 없다.

● 규칙

담화내의 각 발화에 대하여:

1.  $C\mathcal{A}(U_{p_{k,i}})$ 의 어떤 요소가  $U_i$ 에서 대명사로 구현되었으면,  $Cb(U_i)$ 도 역시 그렇다.
2. 센터의 담화 성분간의 전이 유형은 다음 순서로 적용된다.

CONTINUE > RETAIN > SMOOTH SHIFT > ROUGH SHIFT

그림 7 확장된 규칙 및 제약

	$Cb(U_i) = Cb(U_{p_{k,i}})$	$Cb(U_i) \neq Cb(U_{p_{k,i}})$
$C\mathcal{A}(U_i) = C\mathcal{A}(U_j)$	CONTINUE	SMOOTH SHIFT
$C\mathcal{A}(U_i) \neq C\mathcal{A}(U_j)$	RETAIN	ROUGH SHIFT

그림 8 확장된 전이 유형

다음의 두 가지 필터를 이용하여 적당한  $Cb$ ,  $Cf$  목록을 선택한다.

- (1)  $Cb$ 는  $C\mathcal{A}(U_{p_{k,i}})$  목록 중에서 가장 높은 우선 순위를 가져야만 한다. (제약 조건 3)
- (2)  $Cf$  목록 안의 대응어로 실현된 것 중에 적어도 하나는 꼭  $Cb$ 와 일치하여야 한다. (규칙 1) 그러나 시각 정보나 행위 정보를 통해 처리된 대응어구는 제외된다. (제약조건 4)

그림 9 확장된 필터

슬롯을 탐색하는 것이다. 발화 슬롯을 탐색하는 동안에 알고리즘은  $C\mathcal{A}(U_i)$ 와  $C\mathcal{A}(U_{p_{k,i}})$  사이에 수와 성이 일치하는지 선택 제약을 만족하는지 등을 검사한다[9,10]. 그런 다음에 [그림9]와 같은 확장된 필터를 사용하여 적당하지 않은 후보 앵커들을 제거한다. 확장된 앵커는 [그림7]의 제약 조건과 규칙에 기초한 것이다. 만약, 하나의 앵커만이 남겨지면 그것이 현재 발화의  $Cb$ 와  $Cf$ 로 결정한다. 그렇지 않으면, [그림8]에 정의된 전이 유형의 우선 순위에 따라 결정한다.

4. 실험 및 평가

4.1 실험 데이터 분석

실험 데이터는 10명의 식사 과정 학생들이 가상의 다중모드 대화 시스템이 존재한다고 가정하고 사용자의 입장에서 시뮬레이션한 40개의 대화, 754개의 발화로 구성되어 있다. 각 대화는 사용자가 다중모드 대화 시스템과의 상호 작용을 통해 특정 가구의 모델을 선택하는 것을 주된 내용으로 하고 있다. 실험에 사용된 실제 데이터는 발화 정보 뿐만 아니라 지시 행위와 시각 정보가 수반된다. 40개의 실험 데이터에는 402개의 대응어가 포함되어 있었다. 이것은 대화당 10.05개, 발화당 0.54개의 출현 빈도에 해당하는 것으로 흡소평과 같은 영역의 다중모드 대화 시스템에서 대응어 처리의 중요성을 잘 보여준다. [표1]에서와 같이 화면 대응어는 328개로 전체 대응어의 81.6%를 차지해 참조 대응어보다 4배 정도의 출현 빈도를 보였다. 이는 사용자가 이전의 발화나 시각 정보를 참조하기보다는 지시 행위를 이용하여 화면의 물건을 직접 선택하는 경향이 강하다는 것을 보여준다.

4.2 실험 결과 분석

본 논문에서 제안한 두 가지의 다중모드 대응어 처리 알고리즘은 402개의 대응어 가운데 388개를 처리하여 96.5%의 정확률을 보였다. [표1]에서 보듯이 재현율은 402개의 대응어 중에 389개를 찾아내어 96.5%이며, 그 중에서 384개를 처리하여 정확률은 98.7%이다. 제안된 알고리즘을 이용한 다중모드 대응어 처리 시스템에서 처리에 실패한 것들의 원인을 살펴보면 다음과 같다.

● 시스템은 대응어의 범위를 형용사의 수식을 받는 명사구나 복합명사구로 제한하였기 때문에 “*아까 선택했던 빨간 의자로* 하죠.”와 같이 관형절의 수식을 받는 대응어를 처리하지 못하였다.

● 시스템은 “*두 모델의 가격 차이가 많이 있나요?*”와 같은 문장에서 사용자가 명시적으로 대응어라는 단

서(예를 들어, 지시 관형사)를 제공하지 않았기 때문에 두 모델이 대응어라는 것을 찾아내지 못하였다. 이것은 영어의 정관사 'the'와 같은 뚜렷한 언어 표지의 생략이 한국어에서는 자유롭기 때문이다.

표 1 실험 데이터의 구성 및 실험 결과

	화면 대용어	참조 대용어	합계
빈도수 (성분비)	328 (81.6%)	74 (18.4%)	402 (100%)
찾아진 대응어 수 (재현율)	317 (96.6%)	71 (95.9%)	388 (96.5%)
처리된 대응어 수 (정확률)	316 (99.7%)	68 (94.4%)	384 (98.7%)
(처리된 대응어 수/빈도수)*100	96.3%	91.9%	95.5%

● 참조 대용어 처리 알고리즘은 “*아까 의자* 다 주세요.”와 같은 문장에서 *아까 의자* *다*라는 대용어를 처리하지 못하였다. 이것은 한국어에서는 수의 표현이 자유로와 복수 표현인 ‘*아까 의자들*’이 단수 표현인 ‘*아까 의자*’와 ‘*다*’라는 부사로 표현될 수 있기 때문이다. 그러므로 대용어와 지시 대상 사이의 수일치 조건이 만족되지 않아 처리에 실패했다. 만약, ‘*다*’라는 단어의 의미를 파악할 수 있다면 이런 종류의 처리 실패는 해결할 수 있다.

대용어 처리 실패의 3가지의 원인 중에 첫 번째 원인이 전체의 73.3%를 차지하였다. 이는 관형질의 분석과 담화 지식의 필요성을 보여주는 단적인 예라고 할 수 있다. 실패의 원인들 중에 대부분은 한국어 대화의 고유한 특징 때문이었다.

### 5. 결론 및 향후 과제

다중모드 대화 시스템은 단일모드 대화 시스템과는 달리 입력 채널이 다양하기 때문에 서로 다른 특성을 갖는 여러 가지 형태의 대용어가 발생하며 정확한 의도 파악을 위해서는 이것들의 처리가 무엇보다도 중요하다. 본 논문에서는 이를 위하여 2가지 종류의 대용어를 정의하고 각각을 처리하기 위한 알고리즘을 제안하였다. 화면 대용어 처리를 위해서는 발화와 행위, 시각 정보를 이용한 매핑 알고리즘을 제안하였다. 그리고 참조 대용어 처리를 위해서는 워커(Walker)가 제안한 캐시 모델의 센터링 알고리즘을 다중모드 환경에 맞도록 수정한 이중 캐시 모델의 다중모드 센터링 알고리즘을 제안하였다.

본 논문에서는 담화 요소간의 우선 순위를 발화와 함께 수반된 정보의 양에 따라 부여하고, 동일한 정보의 양을 가진 담화 요소간의 우선 순위는 워커(Walker)가 일본어를 대상으로 제안한 방법[14]을 참조하였다. 그러

나 일본어의 우선 순위를 그대로 한국어에 적용하는 것은 많은 문제점[16]이 있다. 그러므로 동일한 정보의 양을 가진 담화 요소간의 우선 순위에 대한 한국어 측면에서의 연구가 필요할 것으로 보인다. 그리고 본 논문에서는 다중모드 시스템에서 발생하는 대용어의 분석적 측면만을 살펴보았으나, 시각적 정보를 이용한 대용어의 생성 측면에 대한 연구도 필요하다.

### 참 고 문 헌

- [1] 김학수, 서정연, “홈쇼핑 영역에서의 다중모드 사용자 인터페이스 시스템”, 1997년도 인지과학회 춘계 학술 발표 논문집, pp. 74-80, 1997.
- [2] Bolt R., “Put-That-There: Voice and gesture at the graphics interface,” *Computer Graphics*, Vol.14, No.3, pp. 262-270, 1980.
- [3] Neal J., Dobes Z., Bettinger K. and Byoun J., “Multi-Modal References in Human-Computer Dialogue,” *Proceedings of AAI-88*, pp. 819-823, 1988.
- [4] Salisbury M. W., Hendrickson J. H. and Lammers T. L., Fu C. and Moody S. A., “Talk and draw: bundling speech and graphics,” *IEEE Computer*, Vol.23, No.8, pp. 59-65, 1990.
- [5] Shimazu H., Arita S. and Takashima Y., “Multi-Modal Definite Clause Grammar,” *Proceedings of COLING94*, pp. 832-836, 1994.
- [6] Shimazu H. and Takashima Y., “Multi-Modal-Method: A Design Method for Building Multi-Modal Systems,” *Proceedings of COLING96*, pp. 925-930, 1996.
- [7] Johnston M., Cohen P., McGee D., Oviatt S., Pittman J. and Smith I., “Unification-based Multimodal Integration,” *Proceedings of 8th EACL*, pp. 281-288, 1997.
- [8] Grosz B. J., Joshi A. K., and Weinstein S., “Providing a unified account of definite noun phrases in discourse,” *Proceedings of the 21st Annual Meeting of the ACL*, pp. 44-50, 1983.
- [9] Brennan S., Friedman M. and Pollard C., “A Centering Approach to Pronoun,” *Proceedings of 25th ACL*, pp. 155-162, 1987.
- [10] Walker M., *Centering Theory in Discourse*, pp. 401-435, Oxford/UK: Oxford U.P., 1998.
- [11] Walker M., Iida M. and Cote S., “Centering in Japanese discourse,” *Proceedings of the 13th International Conference on Computational Linguistics*, pp. 1-7, 1990.
- [12] Grosz B. J. and Sidner C. L., “Attentions, intentions, and the structure of discourse,” *Computational Linguistics*, Vol.12, No.3, pp. 175-204, 1986.
- [13] Walker M., “Evaluating discourse processing

algorithms," *Proceedings of the 27th Annual Meeting of the ACL*, pp. 251-261, 1989.

- [14] Walker M., Iida M. and Cote S., "Japanese discourse and the process of centering," *Computational Linguistics*, Vol.20, No.2, pp. 193-233, 1994.
- [15] 김학수, 다중모드 대화 시스템에서의 명사 대응어구 처리, 서강대학교, 석사학위 논문, 1997.
- [16] 차진희, 송도규, 박재득, "한국어 대응과 생략 해결을 위한 센터링 이론의 적용", *제2회 한글 및 한국어 정보처리 학술대회 논문집*, pp. 347-352, 1997.

김 학 수

정보과학회논문지:소프트웨어 및 응용  
제 27 권 제 1 호 참조

서 정 연

정보과학회논문지:소프트웨어 및 응용  
제 27 권 제 1 호 참조