

대화 시스템에서의 조응어 해석

조은경, 서정연

서강대학교 컴퓨터학과 자연어처리연구실
jek@nlpzodiac.sogang.ac.kr, seojv@ccs.sogang.ac.kr

Anaphora Resolution for Dialogue System

Eunyoung Jo, Jungyun Seo
Department of Computer Science, Sogang University

요 약

조응어 해석(Anaphora Resolution)은 여러 응용 분야에 적절히 적용됨으로써 자연어 이해에서의 양적, 질적 향상을 가져 올 수 있다. 구어에서 조응어(anaphor)는 화자와 청자간에 공유된 정보를 간략하게 나타내는 형태이다. 특히, 구어를 주된 사용 언어로 하는 대화 시스템에서의 조응어의 양상을 살피고 해석하는 문제는 인간과 기계간의 의사소통(man-machine communication)을 온전히 하는 과제이다. 이 논문에서는 인간과 기계간의 대화 시스템(man-machine dialogue system)에서 흔히 쓰일 수 있는 조응 표현에 중점을 두어 그 유형을 살피고 해석의 방식을 제시하도록 한다.

1. 서 론

한국어에서 'Anaphor'는 조응어, 조응사, 대용어, 대용사 등으로 불리지만 장석진(1984), 양명희(1998)의 정의²⁾를 따라 조응어라고 부르기로 한다[3]. 대용이 아니라 조응이라고 한 것은 전산적으로 처리 대상으로 하는 경우, 문맥적으로 혹은 Knowledge Base 내에 선행어(antecedent) 혹은 지시대상(referent)이 존재하는 것을 전제로 한다는 측면에서이고, 조응사가 아니라 조응어라고 한 것은 하나의 단어 단위 표현뿐만 아니라 구 단위 표현으로도 조응 표현이 된다는 측면에서이다.

구어 혹은 대화시스템에서의 조응어(anaphor)란 화자와 청자간에 공유된 정보가 있을 때에 이 공유된 정보를 간략하게 나타내는 표현이라고 할 수 있다.

조응어 해석이 갖는 의의는 실로 크다. 정보추출, 질의응답시스템, 기계번역, 문서요약, 에이전트시스템 등의 응용분야에서 조응어 해석은 자연어 의미 이해 및 자연어 생성에 매우 중요한 역할을 하기 때문이다 [9,10,12,16]. 동일한 선행어를 나타내는 여러 가지 조

응 표현을 해석한다는 것은 자연어를 더 잘 이해하고, 적합 문서를 더 많이 제대로 찾아낼 수 있게 하며, 문서 내에서의 내용을 보다 유기적으로 이해할 수 있게 할 뿐만 아니라 이에 따른 적절하고 자연스러운 표현 생성을 가능하게 한다.

조응어 해석에 일반적으로 이용되는 파서(parser)를 이용하는 방식이나 센터링 이론을 적용하는 방식은 구어 기반의 대화 시스템에서의 조응어 해석에 다소 부족한 면이 있으며, 때로는 부적절할 수도 있다. 특히, 앞한 두 문장까지를 한계 범위로 하여 선행어 후보 목록(antecedent candidate list)을 만들고, 여기에 제약조건(linguistic constraints)을 적용하여 후보들을 걸러낸 다음, 선호도(preference)나 현저성(Salience) 휴리스틱을 적용하여 보다 가까운(recent) 후보를 택하는 일반적인 방식이 구어 기반 대화시스템에서 쓰이는 조응어를 처리하는 데 있어서는 적용될 수 없을 때가 종종 있다. 영어의 'One-anaphora'³⁾에 대응되는 한국어의 '것 조응어'의 선행어로는 최근성이 있는 가까운 문장의 담화 개체나 선호순위가 높은 담화 개체와는 상관없는 것들이 종종 있으며, 대화 상황의 정보를 필요로 하는 상황 조응어(deictic anaphora)나 문맥적으로 드러난 담화개

1) 본 연구는 한국과학재단 목적기초연구 (R01-2003-000-11588-0) 지원으로 수행되었음.

2) 조응 '그 자체 독립적 지시 능력이 없으면서 선행어와 공지시 관계를 맺는 지시 표현'대용: 어떤 사물이나 사건 등을 좀더 추상적인 용어로 대신하는 것으로 선행어를 구체화할 수 없는 경우도 포함한 지시 표현.

3) Daniel Jurafsky & James H. Martin, 『Language and Speech Processing』, Prentice-Hall, 2000.

체와 의미적 연관성이 있는 것으로 문맥적으로 드러나지 않는 지시대상을 가리키는 간접조용어(indirect anaphora) 등이 있기 때문이다.

2. 앞선 연구

조용어 해석의 초기 단계에 상당히 높은 성능을 보였던 방식의 하나가 문법이나 파서를 통해 조용어를 해석하는 것이었는데 이러한 방식은 파서에 대한 의존도가 컸기 때문에 실제로 수작업 파싱 결과에 의존하거나 파싱 결과를 매우 성공적인 것으로 간주한 다음에 이루어지는 문제점이 있었으며, 처리 대상도 일부 대명사에 한정되어 있었다[1,19].

조용어 해석에 가장 일반적으로 쓰이는 방식은 센터링 이론이다. 센터링 이론은 구현이 용이하다는 측면에서, 담화 구조나 조용어에 대한 강력한 설명력을 갖는다는 측면에서 자연어처리뿐만 아니라 자연어생성에서도 널리 쓰이는 조용어 해석 및 생성 방식이며, 나아가 센터링 이론의 규칙이 갖는 한계점이나 문제점이 몇 연구자들에 의해 지적되어 이에 대한 수정 및 확장이 지속적으로 이루어지고 있다[11,13,15]. 하지만, 이 이론으로 해석되는 조용어는 일부 대명사-1,2인칭 대명사를 제외한 대명사류(pronominal)와 한정 명사구(definite noun phrase)에 한정된다. 한국어의 경우, 소설이나 기타 문어 텍스트에서 주로 쓰이는 ‘그’, ‘그녀’ 등의 대명사와 ‘그 사람’, ‘그 분’ 등과 이들의 생략형에 한정되어 처리된 바 있고[4,5], 용언조용어(VP anaphora)를 비롯하여 문어와 구어 모든 영역의 조용어를 해결하려는 연구가 있긴 했지만, 다소 문어투의 대화를 대상으로 한 감이 있으며 보다 다양한 조용어 유형을 처리 못했다[6].

또, WordNet이나 특정 분야의 World knowledge를 사용하여 간접 조용(indirect anaphora)이나 생략(ellipsis)을 해결하는 방식도 있었다[2,7,12].

최근 기계학습 방법에 의해 조용어를 해석하는 방식이 나오고 또, 이것의 성능을 높이는 방법에 대한 것들이 나오고 있으나[14,17,18], 이들은 모두 그다지 많지 않은 데이터에 의존해 있는 한계점이 있는데, 이는 조용어를 해석하기 위해 여러 가지 정보를 주석하기가 매우 어려운 문제점 때문이다. 또, 구어(spoken language)에 대해 실험한 결과도 있었는데 이것도 그다지 큰 성능을 보인다고 볼 수 없다[14].

3. 처리 대상 조용어 유형

자연어처리에서 일반적으로 처리 대상이 되는 조용어

의 유형은 아래와 같고, 대화시스템에서 쓰일 만한 예와 함께 제시하도록 한다.

(밑줄 친 부분은 조용어이고, 기울인 부분은 조용어가 해결된 것이다.)

• 대명사류(pronominal) : 인칭 대명사, ‘그’, ‘그녀’, ‘그들’ 등이 있으나 주로 문어 표현으로 쓰인다. 보다 구어적인 표현으로는 ‘걀’, ‘그애’ 정도의 표현이 있겠다.

(1) user : 오후에 약속 있나?

agent : 네, 오후 3시에 조은경과 미팅이 있습니다.

user : 걀 전화번호 좀 알려줘.

agent : 네, 조은경의 전화번호는 000-0000입니다.

• 한정 명사구(definite noun phrase) : 구정보(old information)임을 나타내는 관형사와 명사가 나란히 쓰여 한정 명사구를 이루는 표현으로 문맥 내에서 선행어를 찾을 수 있는 것들. ‘그것’, ‘그 사람/분’, ‘그때’, ‘그날’

(2) user : 오후에 약속 있나?

agent : 네, 오후 3시에 조은경씨와 미팅이 있습니다.

user : 그 사람 전화번호 좀 알려줘.

agent : 조은경의 전화번호는 000-0000입니다.

• 상황 조용어(deictic anaphora)⁴⁾ : ‘나’, ‘너’, ‘여기’, ‘저기’와 같은 대명사나 ‘이’, ‘저’와 같은 지시사가 결합하여 쓰인 것으로 문맥적으로 선행어를 찾을 수 없고, 주어진 상황을 이루는 시간, 공간, 담화 참여자 등에 대한 정보를 통해 해석이 되는 것들이다.

(3) user : 오늘, 내가 참석해야 할 회의가 몇 개지?

agent : 오늘 user가 참석할 회의가 3개 있습니다.

user : 여기서 하는 회의가 몇시지?

agent : 네, 이곳 AS812호에서 하는 회의는 오후 2시입니다.

• 수사 조용어(surface count anaphora) : 서수사 ‘첫째/첫번째’, ‘둘째/두번째’ 등이나 양수사 ‘하나’, ‘둘’ 등의 표현이 대용의 기능을 하는 것이다.

4) 엄밀한 의미에서 언어학적으로 담화 상황에 의해 주어지는 시간, 공간, 대화참여자 등에 관한 지시 표현은 직시어(deixis)라고 하여 조용어와 구분되는 개념이지만, 여기서는 담화 층위에서 해석 대상이 되는 지시 표현으로서 자연어처리에서 일반적으로 통용되는 용어인 조용어(anaphora)의 한 유형으로 처리한다.

(4) user : 오늘, 내가 참석해야 할 회의가 몇 개지?
 agent : 참석할 회의가 3개 있습니다.
 user : 첫번째가 몇 시지?
 agent : 첫번째 회의는 오후 1시 30분입니다.

(5) user : 대장금 언제 하지?
 agent : 월요일 밤 10시와 화요일 밤 10시에 방영합니다.
 user : 둘 다 녹화해줘.
 agent : 네, 월요일 10시와 화요일 밤 10시에 녹화해 두겠습니다.

- 것조응어(one anaphora) : '것'은 영어의 '~ one'에 대응되는 것으로, 한국어의 경우에 더 다양한 표현의 조응어가 있다.
 - 시간 기준 : '아까 것', '지금 것', '처음 것', '조금 전 것', '나중 것', '먼저 것' 등과 같은 표현으로 두 발화자의 대화의 흐름상 자주 등장할 수 있다.
 - 공간 기준 : '뒤에 것', '앞에 것', '옆에 것', '위에 것', '아래 것' 등과 같은 표현으로 두 발화자간에 공유되는 시각적 공간이 있을 때 쓰일 수 있다.
 - 비교 표현 : '큰 것', '작은 것'
 - 선택 표현 : '다른 것', '아무것/거(나)'

아래의 (6)에서 보이는 조응어 '아무것', '다른 것', '처음 것', '아까 것' 등이 최근성이나 현저성 기준이 무시되며 선행어 탐색 범위를 대화 전체로 보게 하는 대표적인 것들이다. '다른 것'은 문맥에 드러난 담화개체와 동일한 속성을 가진 것으로 DB에서 찾도록 해야 한다.

(6) user : 뉴스 좀 보자.
 agent : MBC, KBS 뉴스가 있습니다. 어느 것으로 보시겠습니까?
 user : 아무거나 보자.
 agent : 네, 그럼 MBC 뉴스를 틀겠습니다.
 user : 아니, 그거 말고.
 agent : 네, 그럼 KBS 뉴스를 틀겠습니다.
 (.....)
 user : 음, 다른 걸로 돌려봐.
 agent : 네, 그럼 SBS 드라마를 틀겠습니다.
 (.....)
 user : 에이, 처음 꺼/아까 꺼 보자
 agent : 네, 그럼 다시 MBC 뉴스로 돌리겠습니다.

- 간접 조응어(indirect anaphora) : 'bridging' 현상이라고도 불리는 것으로 언급된 적이 있는 담화 개체를 통해 표면적으로 언급된 적이 없는 지시 대상을 추론해야 하는 조응 표현으로 '문맥외조응어(exphora)'라고도 한다. 간접 조응어는 개체명 명사가 의미적으로 다른 명사의 속성일 경우에 인식된다. 아래 (9)의 '주인공'이라는 명사는 의미적으로 'XX드라마'라는 명사를 요구하며, 이것이 있어야 사용자의 질문에 대답을 할 수가 있다.

(7) John entered a large dining room. The chandelier hung by an imported gold chain.

(8) 어제 여기서 스테이크를 먹었는데, 맛이 별로더라.

(9) user : 대장금 보자
 agent : 네, 대장금을 틀겠습니다.
 user : ∅ 주인공이 누구지?
 agent : 대장금의 주인공은 '이영애'입니다.

- 영조응어/생략(zero anaphora/ellipsis) : 조응 표현의 가장 극단적인 형태로서 너무나 명백한 지시 대상이나 가정된 상황 등을 의미하는 것으로 문맥적으로 드러나지 않는 조응 표현이다. 대개 명사구 생략(NP ellipsis)은 논항(argument)이 생략된 것으로 조응어임을 인식하고, 용언 생략(VP ellipsis)는 서술어가 없는 것으로 조응어임을 인식한다.

(10) user : 뉴스 좀 틀어 봐
 agent : 지금은 뉴스를 하지 않습니다.
 user : 그럼, 언제 ∅ 하지?
 agent : 뉴스는 저녁 8시, 9시에 합니다.

(11) user : MBC 뉴스 틀어봐.
 agent : 아직, 뉴스를 하지 않습니다.
 user : KBS도 ∅? / KBS는 ∅?
 agent : 아뇨, KBS는 지금 뉴스를 시작했습니다. KBS 뉴스를 틀까요?

- (비생략)용언 조응(VP anaphora) : 한국어는 '그렇다', '그러다' 등 용언이 서술어나 부사어⁵⁾ 등의 활용 형태로 쓰이는데, 구어인 경우는 앞 문장 전체를 의

5) 영어의 경우는 'so'와 같은 표현의 adverbial anaphora가 따로 있다.

미하는 경우가 많다.

- (12) user : MBC 뉴스 틀어봐.
 agent : 아직, 뉴스를 하지 않습니다.
 user : KBS도 Ø?
 agent : 아뇨, KBS는 지금 뉴스를 시작했습니다.
 KBS 뉴스를 틀까요?
 user : 그렇게 해.

- (13) user : 짱구야, 뉴스 좀보자
 agent : 네, 지금은 SBS에서 뉴스를 합니다. SBS 뉴스를 틀겠습니까?
 user : 음, 그러지 말고 드라마나 보자.

4. 조응어 해석을 위한 가정

1. 조응 표현은 잘 인식되어 들어온다.
2. 대화 시스템에서 쓰인 모든 조응 표현은 선행어/지시대상을 갖는다.
3. 시스템은 조응어를 생성하지 않는다.

첫 번째 가정은 어떤 자연어 표현이 조응어이나 아니냐, 혹은 선행어를 갖는 조응 표현으로 볼 수 있는냐 하는 문제점을 논외로 한다. 특히, '~것' 표현은 조응 표현이나 아니냐를 구분하는 것 자체가 어려운 문제점이며, 간접 조응이나 영조응, 수사 조응어 등도 조응어로 인식해 내기가 다소 어려운 점이 있고, 이러한 문제들 자체가 각각이 이슈가 되기 때문이다.

두 번째 가정은 사람과 기계간에 공유되고 있는 정보를 간략하게 표현하는 측면에서만 조응어를 사용한다고 보기 때문에, 선행어나 지시대상을 갖지 않는 대응 표현은 해석의 대상이 되지 않는다고 본다. 이를테면 아래의 (14~15)에서와 같이 인간간의 대화에서 흔히 쓰일 수 있는, 직접적인 표현을 피하거나 막연한 의미로 쓰는 대응 표현은 없다고 가정한 것이다. 이는 애초에 대응어(substitute)가 아니라 조응어라는 용어적 정의와도 관계가 있다.

- (14) 좀 전에 내가 먹은 거 사실은 개밥이었어.
 (15) 야, 거시기 있잖아.

세 번째 가정이 필요한 것은, 인간이 발화한 내용이 잘 전달되었는지를 확인하는 기계의 응답 발화로는 조응어가 쓰이지 않는 것이 바람직한 면이 있고, 한편으로는, 아직 그다지 수준이 높지 않은 인간과 기계간의

대화 시스템에서 기계의 응답 발화로서 조응어를 쓰는 것은 부정확한 의미 전달이 일어날 소지도 있고, 때로는 기계가 조응어를 쓰는 것은 다소 불성실한 발화로 보일 수도 있기 때문이다.

- (16) user : 대장금 보자
 agent : 네, 그것을/Ø 틀겠습니까.
 (17) user : 뉴스 좀 보자
 agent : 네, MBC와 KBS에서 뉴스를 합니다. 어느 것을 보시겠습니까?
 user : 아무거나 보자.
 agent : 네, 그럼 아무거나/그것으로 틀겠습니까.

5. 조응어 해석 방법

조응어 해석을 위한 알고리즘은 대개 다음과 같은 세 단계로 이루어지고, 두 번째와 세 번째 단계는 순서가 바뀌어 적용되기도 한다.

1. 선행어 후보 모으기
2. 선행어 후보 거르기 : 제약 조건 및 일치 적용
3. 후보 우선 순위 매기기

두 번째 단계를 먼저 적용한 후에 선행어 후보의 우선순위를 매기는 과정으로 진행을 한다고 할 때, 세 번째 단계는 두 번째 단계를 적용한 후에도 선행어 후보가 둘 이상인 경우에 필요하다. 여기서는 현실적인 응용을 전제로 하여 구축된 시스템이기 때문에 시스템에 대해 협조적인 사용자를 가정한 인간과 기계간 대화시스템(Man-machine dialogue system)이고 여기서 쓰이는 구어 발화는 다소 짧고 단순하여 조응어 유형별 제약 조건 및 일치만 적용한 후에 남는 선행어 후보가 거의 하나⁶⁾이므로 후보 순위 매기기의 과정을 생략할 수 있다.^{6) 1)}

만약, 순위 매기기 과정을 사용하게 된다면 센터링 이론에서 제시한 선행어 후보들(잠재 초점들 : forward-looking centers)의 선호 순위를 한국어 특성에 맞게 조정할 것[4,6]이나 서술어(동사)의 의미적 성격에 따라 매겨진 의미역들의 위계[8]를 이용할 수 있다. 선행어

6) 대화 예 (1)을 다시 보면, 밑줄친 조응어 '개'의 선행어 후보로는 '오후3시', '조은경', '미탕' 등이 있지만 일치 조건을 적용하고 나면, '개'가 취할 수 있는 선행어로는 '조은경'만 남게 된다.
 (1) user : 오후에 약속 있냐?
 agent : 네, 오후 3시에 조은경과 미탕이 있습니다.
 user : 개 전화번호 좀 알려줘.
 agent : 네, 조은경의 전화번호는 000-0000입니다.

후보의 우선 순위를 매기는 것에 관해서는 센터링 이론 내의 여러 연구에서 수 차례 논의된 바 있고, 어떤 새로운 대량의 말뭉치 분석 결과를 바탕으로 한 새로운 순위가 제시되지 않는 한 이미 어느 정도가 정리가 된 순위를 사용하면 된다.

5.1. 선행어 후보 수집

선행어 후보 수집에 있어서는 우선 후보를 찾는 윈도를 결정하는 일이 요구된다. 대부분의 조응어 해석 방식에서 담화는 응집성(coherence)을 가지려 한다는 원칙적 가정 하에 하나의 담화 세그먼트 내에서만, 그것도 앞 두 문장까지 선행어 후보를 수집하고, 바로 앞 문장에 있는 후보에 대해 최근성(recency) 점수를 더 부여한다.

그런데, 문어 텍스트(monologue)가 아니라 대화(dialogue)를 처리 대상으로 할 경우, 구어에서 흔히 쓰일 수 있는 조응어 중에는 그 의미에 따라 담화의 응집성이나 최근성 원칙이 지켜지지 않는 경우가 있었다. 그래서, 대화의 시작에서부터 조응어가 발생하는 발화까지 선행어 후보가 될 수 있는 것들 모든 것들이 저장되어야 한다. 따라서, 본 시스템은 조응어가 발생했을 때 정해진 범위 내에서 선행어 후보를 수집하는 것이 아니라 대화의 시작에서부터 선행어 후보가 될 수 있는 모든 서술어(predicate)와 모든 서술어의 논항(argument)이 저장되는 방식이다. 다만, 상황 조응어의 경우만, 지시대상을 찾기 위해 청자, 화자, 시간, 공간 등의 대화 상황 정보를 참조하도록 한다. 그래서, 엄밀한 의미에서 선행어 후보 수집이라기보다 선행어의 범위를 폭넓게 잡아주는 것이다.

5.2. 선행어 후보 거르기 및 선행어 선택

선행어 후보를 거르는 과정은 대개 의미적 제약 조건(selectional restriction constraint)이나 일치 조건(agreement constraint)을 이용한다.

제약 조건(linguistic constraints)을 이용하여 후보를 걸러내는 과정은 서술어의 논항 정보를 파악함으로써 입력 조응어의 의미적 제약 조건과 성수(gender, number) 정보를 파악한 다음, 선행어 후보들 중 이 조건들과 일치하는 것을 골라내게 한다. 그런데, 한국어에서, 특히 단수 조응 표현은 수 일치를 잘 안 지키고 있는 것들이 많아 7) 수 일치 적용은 오히려 선행어를 걸러 버리는 문제가 있을 수 있으므로, 조응어가 복수 표현일 경우

에만 적용하도록 한다.

위 대화에서 '그거'가 가리키는 것은 'a,b,c인 메일들'이라는 복수 명사이다. '그것들 몽땅 지워버려'라고 하는 표현이 문법적으로 맞긴 하지만 그다지 잘 지켜지지 않고 구어에서는 오히려 다소 잉여적인 느낌을 준다.

5.2.1. 대명사류와 한정명사구 조응어 (pronominal and definite np anaphora)

대명사류와 한정명사구는 현재 발화의 이전 발화(agent 발화)와 그 이전 발화(user의 발화)에서의 명사구(np) 후보들 중 동일한 대상을 가리키는 것을 제외한 후 그 중 제약 조건이 일치하는 것만을 남긴다.

5.2.2. 상황 조응어 (deictic anaphora)

상황 조응어는 조응어가 발생한 현재 발화의 상황에 의해 해석된다.

가장 많이 쓰일 수 있는 표현을 예로 들면, '나'는 청자로, '너'는 화자로, '지금'은 현재 발화의 시간으로 '여기'는 대화 상황의 공간으로 해석한다.

이 외에도 '이번 주말', '다음 주말', '지난 주' 등 현재 발화 시간을 기준으로 한 여러 가지 시간 표현도 해석될 수 있다.

5.2.3. 수사 조응어(surface count anaphora)

수사 조응어는 현재 발화의 이전 발화까지의 np 후보들 중 복수 표현의 np나 ',', '와/과', '그리고' 등으로 연결된 np를 하나의 개체로 하는 집합 개체 중에서 조응어의 의미 조건과 일치하는 것만을 남긴다.

5.2.4. 것 조응어 (one anaphora)

'것 조응어'의 선행어는 조응어 각각의 개별적 의미에 따라 후보 수집의 범위가 다르다.

'처음 것' : 대화 시작에서부터 현재 발화의 이전 발화까지 만들어진 선행어 후보 중에서 조응어 논항의 제약 조건과 일치하는 것 중에서 가장 먼저 나온 것으로 해석한다.

'아까 것/꺼/거' : 이전 발화에 있는 것과 이와 동일한 것을 지시하는 표현을 제외한 나머지 것들 중에서 제약 조건과 일치 하는 것을 남긴다.

'다른 것' : 처리 영역별 KB에서 제약 조건에 맞는 것을 골라온다.

'아무거/아무데/아무때'(나) : '아무데', '아무때'가 '것 조응어'는 아니지만, 제약 조건만을 달리할 뿐 처리 방식은 같다.

5.2.5. 간접 조응어 (indirect anaphora)

7) (18) user : 오후 2시부터 지금까지 새로 온 메일 있나?
agent : 네, 수신인이 a,b,c인 메일들이 새로 왔습니다.
user : 그거 몽땅 지워버려

간접 조응은 (7~9)에서처럼 어떤 담화 개체가 발생한 후 부대화(subdialogue) 혹은 부발화(subutterance)에서 주로 생기며, 그 선행어는 이전 발화의 목적어라는 경험적 가정 하에, 간접 조응어의 선행어는 이전 발화에 있는 목적어로 해석한다.

5.2.6. 생략(ellipsis)과 용언 조응 (VP anaphora)

명사구 생략(명조응어)은 이전 발화에 있는 np 후보들 중 제약조건에 일치하는 것으로 해석하고, 용언 생략은 이전 발화의 용언으로 해석한다.

비생략 용언 조응은 이전 발화의 뒷문장을 후보로 한다.

5.3. 조응어 해석 알고리즘

구어나 대화시스템에서의 조응어를 해결 대상으로 할 경우, 실험을 위해서 쓸만한 말뭉치가 매우 희귀하다는 문제점이 있다. 무엇보다 두 사람간의 대화만으로 구성된 데이터가 드물고, 기계와 사람간의 대화도 당연히 많지 않다. 그리하여, 본고에서 제시하는 해석 방식을 써서 조응어 해결을 한 것은 작은 양의 예제 대화에 불과하다.

그래서, 본고의 알고리즘을 적용하여 실험 및 평가와 개선은 추후 과제로 남기기로 하고, 위 절에서 정리한 여러 유형의 조응어를 해석하는 알고리즘을 간략하게 시하는 것으로 그친다.

```

발화 목록을 만든다.
While(입력 발화)
{
    1. 입력발화의 서술어를 파악.
        1.1. VP 조응어가 있으면,
            if(서술어 == 조응어)
            {
                1.1.1. 생략형이면, 이전 발화 서술어로
                1.1.2. 비생략형이면, 이전 발화 전체로
            }
        2. 서술어의 논항 목록을 만든다.
            while(!발화의 끝)
            {
                2.1. 논항과 그 의미정보를 파악한다.
                2.2. 논항목록에 논항을 저장한다.
            }
        3. NP조응어가 있으면,
            if(서술어의 논항 == 조응어)
            {
                3.1 조응어 논항의 의미 정보를 파악.
                3.2 조응어 유형별 선행어 후보를 만든다.
            }
    }
}
    
```

```

3.3 제약조건에 따라 후보 거르기 및 선택.
    // 필요시 순위 매기기 확장 가능.
3.4 조응어에 선행어/지시어 입력.
}
4. 발화 목록에 발화(서술어,논항) 저장.
}
    
```

7. 결론 및 남은 문제

지금까지 국내에서도 조응어 해석을 위한 여러 가지 접근 방식들이 있어 왔지만, 대개는 문어에 치우친 감이 있고, 센터링 알고리즘을 확장 적용하여 대화시스템을 고려한 조응어 처리 예제도 문어투의 표현과 크게 차이 나지 않는 것들이었다.

본고에서는 구어 대화에서 쓰일 수 있는 여러 유형의 조응어를 살펴 보았고, 이들을 해석하기 위해서는 가까운 범위의 담화개체만이 아니라 대화 전체에서 그 지시대상을 찾도록 하는 방법이 요구되며 특히, 조응어의 개별적 유형에 따라 선택의 범위나 해석의 방식이 다르게 적용되어야 함을 제시하였다.

한편, 최근 국외에서 통계적인 방법이나 기계학습에 의한 조응어 해석이 부각되고 있기는 하지만 [12,14,17,18], 국내외적으로 조응어 해석을 위해 보다 실험적인 접근 방식을 적용하기가 어렵고 또 적용했다고 하더라도 여러 다른 방식들을 비교하기 어려운 것은 실험에 사용할 수 있는 조응어 관련 정보가 태깅된 공통된 말뭉치(standard anaphorically tagged corpora)의 부족이 가장 큰 원인이다.

또, 앞으로 더 많은 연구의 필요가 있는 조응 표현은 의존명사가 결합한 조응 표현이나 수사가 결합한 조응 표현, 개체명 명사가 결합한 조응 표현 등이 있는데, 이들에 대해서는 선행-지시 관계라기보다 전체-부분 관계라는 측면에서 고찰되어 해석될 필요도 있겠다.

참고 문헌

[1] 김정해, "HPSG 파서에 기반한 문맥 조응 대응어의 해결". 경북대학교 박사학위논문, 1997

[2] 신효식, "한국어 백과사전에 등장하는 외조응적 영대명사의 복원에 관한 전산학적 연구", 연세대학교 컴퓨터산업시스템공학과 석사학위논문, 2002

[3] 양명희, 『현대국어 대응어에 대한 연구』, 국어학총서33, 국어학회, 태학사, 1998

[4] 이익환, 이민행. "한국어 대화에서의 대명사의 선행사 탐색 - 통제된 중심화 이론적 접근", 제11회 한글 및 한

- 국어 정보처리 학술대회. 1999
- [5] 양재형, “중심스택을 이용한 한국어의 대명사 지시 해결”, 정보과학회논문지(B) 제25권, 제12호, 1998.
- [6] 노현철, 이근배, 이종혁, 박재득. “한국어 담화 특성에 기반한 영역 독립 생략 및 대응 처리”, 정보과학회논문지(B) 제25권 제12호, 1998.
- [7] 윤철진, “제한된 영역의 대화에서 계획에 기반한 체언구의 생략 처리”, 서강대학교 컴퓨터학과 석사학위논문, 1998
- [8] 김현주, “한국어 재귀사 「자기」의 지시해석에 관한 연구 -의미 역할의 현저성 위계를 중심으로”, 서울대학교 언어학과 석사학위논문, 1997
- [9] Jess Peral et al, “Translation of Pronominal Anaphora between English and Spanish : Discrepancies and Evaluation”, *Journal of Artificial Intelligence Research* 18 (2003)
- [10] Jos L. Vicedo; Antonio Ferrndez, “Importance of Pronominal Anaphora resolution in Question Answering system”, *acl 2000 proceedings*
- [11] Laurel Fais, “Inferable Centers, Centering Transitions, and the Notion of Coherence”, *Computational Linguistics* vol.30, Number 2, 2004.
- [12] Massio Poesio et al, “Learning to Resolve Bridging References”, *acl 2004 proceedings*.
- [13] Michael Strube, Udo Hahn, “Functional Centering Grounding Referential Coherence in Information Structure”, *Computational Linguistics* vol25. num3. 1999
- [14] Michael Strube and Christoph Muller, “A Machine Learning Approach to Pronoun Resolution in Spoken dialogue”, *ACL 2003 proceedings*
- [15] Rodger Kibble, “A Reformulation of Rule 2 of Centering Theory”, *Computational Linguistics* vol27. num4. 2001
- [16] Rodger Kibble, “Cb or not Cb-Centering theory applied to NLG”, *ACL Workshop*, 1999
- [17] Xiaofeng Yang, Guodong Zhou, et al, “Coreference Resolution Using Competition Learning Approach”, *ACL 2003 proceedings*
- [18] Vincent Ng, “Learning Noun Phrase Anaphoricity to Improve Conference Resolution- Issues in Representation and Optimization”, *ACL 2004 proceedings*
- [19] Hobbs, Jerry R., 1978, “Resolving Pronoun References”, *Lingua*, Vol.44