

한국어 대용과 생략 해결을 위한 센터링 이론의 적용

차 건희, 송 도규, 박 재득
시스템공학연구소
{chakh, dgsong, jdpark}@seri.re.kr

Centering for Korean Anaphora and Ellipsis Resolution

Keon-Hoe Cha, Do Gyu Song, Jay-Duke Park

Systems Engineering Research Institute

대용(anaphora)과 생략(ellipsis)을 포함한 지시 해결(reference resolution)은 자연어 처리에 있어서 꼭 해결하고 넘어가야 할 문제이며 지시어가 가리키는 피지시어를 올바르게 찾아내는가의 여부가 자연어처리 시스템의 성패를 가름하는 중요한 요소가 된다해도 과언이 아니다. 본 논문에서는 한국어 대용과 생략을 해결하기 위한 기제로서 센터링 이론이 적합한가를 검토하고 나아가 언어에 따라서 다르게 제안된 Cf 목록의 순서를 한국어의 언어특성에 맞게 결정하는 기법을 제안한다. 센터링 이론에서 각 개별언어 간의 차이가 가장 두드러지게 나타나는 부분은 Cf 목록을 구성하는 요소들의 순서이다. 여기에서는 한국어와 언어적인 특성이 유사한 일본어를 대상으로 제안된 순서를 한국어에도 그대로 적용될 수 있는지 검토하고 문제점을 지적함과 동시에 개선 방안을 아울러 살펴보고자 한다.

1. 서론

언어학적으로 설득력 있는 이론을 만드는 데 있어서 뿐만 아니라 자연어 처리 분야에서도 까다롭고 어려운 문제 중의 하나가 지시 해결(reference resolution)이다. 자연언어의 특성 중의 하나가 동일한 대상을 다양한 방법으로 지칭할 수 있다는 데에 있으며 바로 이것이 담화나 텍스트의 일관성과 계속성을 유지시켜 주고 언어의 사용을 용이하게 한다. 새로운 지시 대상을 담화 상황이나 텍스트 안으로 도입하고 또 이미 도입된 지시 대상들을 가리키는 역할을 제안어 하게 된다. 여기서 다루기 어려운 점은 한 지시 대상이 여러 가지 지시어로 나타내어 질 수 있고 반면에 한 지시어가 여러 지시대상들을 가리킬 수 있다는 것이다. 이를 가장 잘 보여주는 대표적인 예가 바로 대용(代用, anaphora)과 생략(略辭, ellipsis)이다. 이들의 성립 매커니즘을 규명하려는 노력이 다양한 방법론적인 접근을 통해서 행해지고 있지만 아직까지도 대명사류의 대용어에 대해서조차 완전하고 설득력 있는 이론이 제시되지 못하고 있는 실정이다. 그러나 10여년 전부터 Grosz 와 Sidner 를 위시한 학자들이 지시 해결을

위해 센터링 이론(centering theory)을 제안한 이후 여러 학자들에 의해 발전적인 논의가 진행되면서 다양한 자연어처리 시스템에서 적용되고 있다. 본 논문에서는 먼저 센터링 이론의 근본 매커니즘을 살펴보고 한국어에 적용하는데 있어서의 몇가지 문제점을 살펴본 후에 Cf의 우선순위를 결정하기 위하여 베이시안 확률을 적용하는 기법을 제안한다.

2. 센터링 이론

센터링 이론은 Grosz 와 Sidner 의 연구에서 유래했으며[1, 2, 3] Grosz, Joshi, Weinstein 에 의해 발전된 이론이다[4, 5, 6]. 그 후에 Brennan, Friedman, Pollard 가 제약(constraints)과 규칙(rules)들을 추가하면서 센터링 이론은 골격을 갖추게 되었다[7, 8]. 담화 해석의 계산 모델로서 센터링 이론의 근본적인 목표는 추론(inference)을 통제하기 위해 담화내의 'ATTENTIONAL STATE'를 모형화하는 데에 있다[4].

센터링 모델은 매우 단순해서, 센터링 이론에서는 명사구에 의해 실현되는 문장 구성 성분으로 그 논의를 제한한다. 그러나 결과 문장도 센터가 될 수 있는 만큼 앞으로의 이론적인 발전

에 따라 절이나 문장 단위에 까지 확장될 수 있으리라 기대한다. 센터링 이론에 따르면 담화내의 모든 발화는 두 가지 구조를 가지고 있다. 먼저, 각 발화는 FORWARD-LOOKING CENTERS(이후 약해서 Cf)라는 담화 요소의 목록에 연관되어 있으며 두 번째는 이 Cf내에 BACKWARD-LOOKING CENTER(이후 약해서 Cb)라는 특별한 요소가 있다. Cb는 하나의 발화에 있어서 가장 중심적인 담화 요소이고 이것은 보통 주제(theme)에 해당된다. 이 Cb는 현 발화에 선행하는 담화에 연결된다. Cf 목록의 요소들은 일정한 기준에 의해서 순서가 매겨지는데 그들 중 가장 우선 순위의 담화 요소가 PREFERRED CENTER(이후 약해서 Cp)로 설정된다. 이 Cp가 Cf 목록의 요소 중 다음 발화의 Cb가 될 가능성이 비교적 가장 높다는 것이 센터링 이론의 기본 전제이다. 대개의 경우 이 Cp가 담화의 앞부분의 주제, 즉 Cb가 되기도 하나 항상 그런 것은 아니다. 이 Cb와 Cf의 구분이 센터링 이론의 핵심이다.

센터링 이론은 이 Cb와 Cf 두 구조에다가 일련의 제약과 규칙을 두고 있다. 여기에서의 제약은 엄격히 지켜져야 하는 반면에 규칙은 절대적으로 지켜져야 하는 것은 아니다. 먼저 센터링 이론에서 도입하고 있는 제약을 보자.

◆ 제약

담화내의 각 발화에 대하여:

1. 각 발화에 하나의 Cb가 있다.
2. 각 발화의 Cf 목록의 모든 요소는 반드시 이 발화 안에서 실현되어야 한다.
3. 각 발화의 Cb는 이 발화에서 실현된 바로 전 발화의 Cf에서 가장 우선 순위가 매겨진 담화요소이다.

여기서 Cf 목록의 순서는 일반적으로 언어에 따라 조금씩 다르다. 먼저 영어에 주어진 순서는 대개 다음과 같다.

주어 > 직접목적어 > 간접목적어 > 보어 > 수식어구 [9, 10]

이에 반해 일본어를 대상으로 제안된 Cf 목록의 순서는 다음과 같다.

주제(Topic) > 강조된 요소(Empathy) > 주어 > 간접목적어 > 직접목적어 > 여타의 성분 [11, 12]

Cb는 주로 생략되거나 대명사화된다. 그리고 담화 성분 간의 전이는 몇 가지 유형이 있는데 이들 간에는 일관성(coherence)의 차이가 있다. 센터링 이론에서 일관성은 담화 해석시에 담화를 해석하는 사람의 추론 과정에 의해 측정된다.[4][5] 구체적인 예를 들면, 같은 내용을 계속 센터로 유지하는 담화 성분이 이 내용에서 저 내용으로 되풀이해서 센터를 옮겨가는 것보다 더 일관성이 있다. 이 관점이 다음 두 규칙으로 요약된다.

◆ 규칙

담화내의 각 발화에 대하여:

1. 앞 발화의 Cf의 어떤 요소가 현 발화에서 대명사화 되었다면, 현 발화의 Cb도 역시 그러하다.
2. 센터의 담화 성분 간의 전이 유형은 다음 순서로 적용된다.

CONTINUING > RETAINING > SHIFTING TO PREFERRED
> SHIFTING TO NONPREFERRED

다음은 이들 담화 성분 간의 센터 전이 유형을 요약 정리한 것이다.

앞, 뒤 발화의 Cb 관계 현 발화의 Cb와 Cf 관계	$Cb(U_i) = Cb(U_{i-1})$	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$	CONTINUING	SHIFTING TO PREFERRED
$Cb(U_i) \neq Cp(U_i)$	RETAINING	SHIFTING TO NONPREFERRED

Cb(U_i): 현 발화의 Cb
Cb(U_{i-1}): 앞 발화의 Cb
Cp(U_i): 현 발화의 Cp
(표 1 센터의 전이 유형)

위 센터의 전이 유형을 나타내는 용어 중에 'CONTINUING' (CONTINUATION)과 'RETAINING' (RETENTION)은 일반적으로 그대로 널리 쓰이나 'SHIFTING TO PREFERRED'와 'SHIFTING TO NONPREFERRED' [10]는 이들의 개념 자체가 여러 학자들에 의해 점진적으로 발전, 성립되었기에 여러 용어가 뒤섞여 쓰이고 있다. 이들은 각각 'SMOOTH-SHIFT'와 'ROUGH-SHIFT'[8,9] 또는 'SHIFTING-1'과 'SHIFTING'로 불리기도 한다[7, 10].

3. 센터링 이론의 한국어에의 적용

여기에서는 한국어 대용과 생각을 해결하기 위한 기제로서 센터링 이론이 적합함을 검토하고 나아가 언어에 따라서 다르게 제안된 Cf 목록의 순서를 한국어의 언어특성을 고려해 통계적 기법을 사용하여 결정하는 방식을 제안하려 한다.

앞에서 이미 언급한 대로 각 개별언어 간의 차이가 가장 두드러지게 나타나는 부분은 Cf 목록을 구성하는 요소들의 순서이다. 여기에서는 한국어와 언어적인 특성이 유사한 일본어를 대상으로 제안된 순서를 한국어에도 그대로 적용할 수 있는지 검토하고 문제점을 지적함과 동시에 개선방안을 아울러 살펴보기로 한다.

일본어에 대해서는 [11, 12]에 심도높은 논의가 개진되어 있고 그것을 간략히 요약하면 다음과 같다. 일본어에서도 여러 가지 문법 기능들이 격형태소로 주어지는데 그 중에서 'TOPIC'은 형태 표지자(morphological marker) "wa"에 의해 표시되고 주어는 "ga", 직접목적어는 "o", 간접목적어는 "ni"에 의해 나타내어진다. 특히 "wa"와 "ga"를 구별하여 다음 예문을 통하여 전자는 'TOPIC'을 추자는 평이한 주격 성분을 나타낸다.

(1)

Un:
Hanako wa tosyokan de benkyoositeita.
Library in studying-was
"Hanako was studying in the library."
Cb: HANAKO
Cf: [HANAKO]

Un+1:
Taroo ga Hanako o tetudatte-kureta.
help-gave
"Taroo gave Hanako a favor in helping her."
Cb: [HANAKO]
Cf: [HANAKO, TAROO]
Empathy subj

Un+2:
Tugi no hi o o eiga ni sasotta.
Next of day SUBJ OBJ movie to invited
"Next day she invited him to a movie."
Cb: HANAKO
Cf: [HANAKO, TAROO] CONTINUING
subj obj

(2)

Un:
Taroo wa kooen o sanpo-siteita
SUBJ park walk-around
"Taroo was walking around the park."
Cb: TAROO
Cf: [TAROO, PARK]
subj obj

Un+1:

Hanako ga o yatto mituketa
SUBJ finally found
"Hanako finally found (him)."
Cb: TAROO
Cf1: [TAROO, HANAKO] (C)
TOPIC/obj subj
Cf2: [HANAKO, TAROO] (R)
subj obj

Un+2:

O O yotei o setumeisita
SUBJ OBJ schedule explained
He explained the schedule to her. (Cf1)
She explained the schedule to him. (Cf2)
Cb1: TAROO
Cb2: HANAKO
Cf1: [TAROO, HANAKO] (C)
subj obj
Cf2: [HANAKO, TAROO] (S-1)
obj subj

이렇게 해서 이들은 Cf 목록의 순서를 다음과 같이 제안하였다.

주제(Topic) > 강조된 요소(Empathy) > 주어 > 간접목적어 > 직접목적어 > 여타의 성분

이제 이 논의를 한국어에 적용시켜 보자. 먼저 검토해야 할 사항은 일본어를 대상으로 제안된 Cf 목록의 순서대로 과연 직접목적어보다 간접목적어를 우선해야 하는가 하는 점이다. 다음 일련의 발화를 예로 들어보자.

(3)

- a. 철수가 영희에게 명자를 소개했다.
- b. 그는 그녀를 좋아하고 있었다.(그녀는 그를 좋아하고 있었다.)
- c. 그녀도 그에게 호감을 가지고 있었다.
- d. 다음 날 영화에 초대했다.
- e. 하루 종일 아무 일도 하지 못했다.

(3.b)의 "그녀"가 앞 발화의 "영희"와 "명자" 중에 어느 것을 가리킬까? 일본어를 대상으로 Walker[11, 12]등이 제안한 바에 따르면 "영희"가 우선적으로 가리켜져야 한다. 그러나 한국어에서는 꼭 그래야 할 이유가 없을 뿐만 아니라 오히려 한국어 화자의 직관상 "명자"를 가리킴이 더 적합할 것 같다. 두 가지 해석에 따른 Cb와 Cf의 변화과정을 보자.

(4)

a. 철수가 영희에게 명자를 소개했다.

Cb: [?]
Cf: [철수, 영희, 명자]
Cf: [철수, 명자, 영희]

b. 그는 그녀를 좋아하고 있었다. (그녀는 그를 좋아하고 있었다.)

Cb: [철수] CONTINUING
Cf: [철수, 영희]
Cf: [철수, 명자]

c. 그녀도 그에게 호감을 가지고 있었다.

Cb1: [영희] SHIFTING TO NONPREFERRED
Cb2: [명자] SHIFTING TO NONPREFERRED
Cf: [영희, 철수]
Cf: [명자, 철수]

d. 다음 날 영화에 초대했다.

Cb: [철수] SHIFTING TO NONPREFERRED
Cf: [철수, 영희, 영화]
Cf: [철수, 명자, 영화]

e. 하루 종일 아무 일도 하지 못했다.

Cb: [철수] CONTINUING
Cf: [철수]
Cf: [철수]

위 일련의 Cb와 Cf 중 1의 표지가 붉은 흐름과 2의 표지가 붉은 흐름 중에 어느 것이 더 자연스러울까 하는 것이다. b의 발화를 괄호안에 있는 “그녀는 그를 좋아하고 있었다”로 바꾸어 보면 더 분명해 진다. 여기서 “그녀”는 피지시어로 “명자”를 선호할 것이기 때문이다. 따라서 한국어에서는 간접목적어를 직접 목적어보다 우선에 놓아야 할 아무런 이유가 없으며 어느 것을 우선해야 할 것인가를 결정하기에 충분하고 남득할 만한 기준을 세우기가 어렵다.

두 번째는 격형태소에 따른 주제의 선정 문제이다. 앞의 논의에서 본대로 일본어에서는 주격을 나타내는 격형태소에서도 “wa”와 “ga”에 따라 주제와 비강세주어가 구분된다. 예문 (1)에서 “Hanako wa”는 Cp가 되는 반면에 “Taroo ga Hanako o tetudatte-kureta”의 “Taroo ga”는 Cp가 되지 못하고 오히려 그 자리를 “Hanako”에게 넘겨주고 있다. 한국어에 대한 이 주제의 논의는 [13]에 자세히 개진되어 있으며 Walker 등과 비슷한 결론에 도달한다. 그에 따르면 한국어도 조사에 따라 주제가 달라지는데 가장 특징적인 주제 표지자가 한국어 격형태소 ‘은/는’이다. 따라서 ‘은/는’(그는 이를 topic markers 라고 부른다)이 붉는 체언이 주격조사 ‘이/가’나 목적격조사 ‘을/를’(그는 이것들을 case markers 라고 부른다)이 붉는 담화요소보다 그 발화의 Cb며 Cp

가 될 가능성이 높은 담화요소라는 것이다! 그러나 다음 예문을 보자.

(5)

1. 영희가 철수에게 명자를 소개했다.
2. a. 그녀는 그를 좋아했다.
- b. *그녀가 그를 좋아했다.
- c. 그녀는 그를 좋아했었다.
- d. 그녀는 그를 좋아하게 되었다.
- e. 그녀는 곧 후회했다.
- f. *그녀가 곧 후회했다.

(6)

1. 영희는 철수에게 명자를 소개했다.
2. a. 그녀는 그를 좋아했다.
- b. *그녀가 그를 좋아했다.
- c. 그녀는 그를 좋아했었다.
- d. 그녀는 그를 좋아하게 되었다.
- e. 그녀는 곧 후회했다.
- f. *그녀가 곧 후회했다.

위의 예에서 1번 문장 다음에 2번의 각 문장이 뒤따른다고 가정하자. 일단 b와 f의 “가”가 붉은 발화는 1번 발화에 뒤따라 일어나는 발화로서는 적절하지 않다. 그 이유는 [17]에 잘 설명되어 있다시피 한국어에서 “가”는 주제와 결합할 수 없기 때문이다. 다시 말해서 발화의 주제인 “그녀”와 조사 “가”는 양립할 수 없으므로 당연히 “그녀”는 조사 “는”과 결합되어 발화의 주제로서의 역할을 하게 된다. 그러나 이 주제도 대응어의 피지시어를 찾는 데 있어 절대적인 우선 순위를 가질 수 없다는 것이다. 이제 그 이유를 논의해 보자. 2번의 각 발화에 따라서 “그녀”가 가리키는 피지시어가 “영희”와 “명자” 중에 어떤 것일까? 한국어 화자라면 2번의 a, c, d, e의 “그녀”에 각각 “명자”, “영희”, “명자”, “영희”를 할당하는 데에 이의가 없을 것이다. 특히 a와 c 문장의 “그녀”는 발화의 나머지 부분이 모두 같고 동사의 시제만 다른 상황에서 다른 피지시어를 가질 수 있다는 사실이 앞의 첫 번째 논의와 더불어 센터링 이론에서 제안한 Cf 목록의 순서가 적어도 한국어에서 만큼은 일반화될 수 없음을 잘 보여준다. 각 발화의 Cb와 Cf의 흐름을 보면 다음과 같다.

(7)

1. 영희가 철수에게 명자를 소개했다.
- Cb: [?]
- Cf: [영희, 철수, 명자]
- Cf: [영희, 명자, 철수]

1) 자세한 논의는 [17] 참조.

- 2.a. 그녀는 그를 좋아했다.
 Cb1:[영희] CONTINUING
 Cb2:[명자] SHIFTING TO NONPREFERRED
 Cf1:[영희, 철수]
 Cf2:[명자, 철수]
- b. 그녀는 그를 좋아했었다.
 Cb1:[영희] CONTINUING
 Cb2:[명자] SHIFTING TO NONPREFERRED
 Cf1:[영희, 철수]
 Cf2:[명자, 철수]
- c. 그녀는 그를 좋아하게 되었다.
 Cb1:[영희] CONTINUING
 Cb2:[명자] SHIFTING TO NONPREFERRED
 Cf1:[영희, 철수]
 Cf2:[명자, 철수]
- d. 그녀는 곧 후회했다.
 Cb1:[영희] CONTINUING
 Cb2:[명자] SHIFTING TO NONPREFERRED
 Cf1:[영희, 철수]
 Cf2:[명자, 철수]

각 발화의 대응어 “그녀”의 피지시어로 선호(preference)되는 해석을 굵게 표시했다. 여기에서 확인할 수 있는 바와 같이 ‘CONTINUING’과 ‘SHIFTING TO NONPREFERRED’간에도 어떤 절대적인 우선 순위를 부여할 수 없다.

다음으로는 주제(topic)와 주어(subject)에 관한 문제이다. 앞에서 본대로 한국어는 일본어와 동일하게 조사 ‘는’이 첨가되는 주어(topic)가 나타난다. Li & Tompson은 주어/주제에 관련된 언어 유형을 4가지로 분류하면서 일본어와 한국어가 주어와 주제가 두드러진 언어에 속한다고 하였다.[14] 따라서 이 분류에 따라 주어가 우세한 인구나 주제가 우세한 중국어에서와 같이 미리 정해진 Cf의 우선순위를 한국어나 일본어에 적용한다는 것은 문제가 있을 수 있다. 다음의 몇가지 예를 살펴보자.

- (8)
 a. 코뿔소는 뿔이 단단하다.
 b. 사람들은 그것을 좋아한다.
- (9)
 a. 가격은 크레도스가 50만원이 더 비쌉니다.
 b. 그것은 나에게 너무 비싸군요.

위의 예문 (8.b)에서 “그것”이 지시하는 대상은 ‘코뿔소’이거나 아니면 ‘뿔’이다.³⁾ 만약에 [11]에서와 같이 한국어에도 동일한 우선순위를 적용한다면 주어어가 되는 ‘코뿔소’를 가리켜야 할 것이다. 그러나 이 문장에서는 ‘뿔’을 가리킨다고 보는 것이 자연

³⁾ “그것”이 지시하는 대상이 (8.a)의 문장 전체 내용인 경우는 논외로 한다.

스럽다. 또한 (9.a)에서는 “가격”(TOPIC, “크레도스”(subject1), “50만원”(subject2)의 순서로 Cf 목록이 정해지지만 (9.b)의 ‘그것’은 피지시어를 다양하게 가리킬 수 있다. 따라서 한국어의 담화분석에서 센터링이론을 적용할 때 미리 고정된 Cf의 우선순위를 갖는 것은 많은 문제점을 유발할 수 있음을 보았다.

지금까지의 논의에 따라 센터링 이론에서 제안하고 있는 센터의 전이 유형간의 우선 순위와 Cf 목록의 우선 순위가 한국어 대응과 생략 현상의 해결을 위해서 체계적이고 신뢰할 만한 기준으로 삼기에는 다소 무리가 있다.

4. 베이시안 확률(Bayesian Probability)을 이용한 Cf 우선순위 결정

본 논문에서는 이러한 문제점을 해결하기 위하여 베이시안 기법을 제안하고자 한다. 이에 따르면 주제어(topic), 주어(subject), 목적어, 동사구 등의 Cf 내에서의 순위가 각 확률값에 의해 정해진다. Mitkov는 영어의 담화분석에서 센터링 이론을 적용할 때 Cf의 순위를 베이시안 확률값으로 결정하는 기법을 제안했다.[15] 그는 문장의 각 성분들이 어떤 징후와 함께 나타났을 때 센터가 될 사후확률을 다음과 같이 정의한다.

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum P(H_i)P(A|H_i)}$$

여기서 P(H_k)는 어떤 문장 성분이 센터가 될 확률이고 P(A|H_k)는 그 문장성분이 센터일 때 징후(symptom) 또는 조건 A가 나타날 확률값이다. 보통 영어에서 주어, 목적어, 동사구가 센터가 될 확률은 각각 73%, 25%, 2%로 나타난다. 그리고 어떤 동사들에서는 목적어가 센터가 될 확률이 높는데 이런 동사들로서는 discuss, present 등이 있다. 또한 주어가 ‘chapter’, ‘section’, ‘table’ 이거나 대명사 ‘I’, ‘we’, ‘you’ 일 경우에도 역시 목적어가 센터가 될 확률이 가장 높다. 이런 조건들을 징후로 정하여 각 문장 성분들에 대하여 사전확률값과 그 징후들에 대한 조건부확률값들을 통해 징후에 대한 사후확률값을 계산하여 문장의 센터를 결정한다.

이러한 베이시안 기법은 주제어와 주어와 각기 센터가 될 수 있는 한국어에 적용할 수 있다. 주제어와 주어와 동시에 나타날 때 둘중의 하나를 센터로서 선호하는 조건들은 다음과 같다.

1. 주어가 대주어, 주제가 소유어역할을 하여 주제가 주어의 일부가 될때는[16] 주제가 센터가 될 확률이 높다.

2. 주제어의 표지로 쓰이는 조사 '는' 이 문두에 오지 않을 때는 대조의 의미가 강하기 때문에 이때는 주어가 센터가 될 확률이 높다. 그러나 문두에 올 때는 일반적으로 주어가 화자의 "유일한 관심 대상"이 된다는 견해[17]에 따르면 주제어가 센터가 될 확률이 높다.
3. 예문 2 예와 같이 수량사에 관련된 문장은 기저구조의 수량사가 추치변형으로 이루어진 것으로 본다. 즉 2(나)의 문장은 "크레도스가 50만원의 가격이 더 비싸다" 라는 문장에서 변형된 것으로 볼 수 있다는 것이다.[18] 따라서 이때는 하나의 주어가 있는 문장이며 주어가 센터가 될 확률이 더 높다.

이외에도 선호도를 결정짓는 조건들은 더 있을 것이며 그것을 찾아내는 것은 이후의 연구 주제가 될 것이다. 이제 이들 세가지의 조건들을 각각 A_i (여기서 $i=1,2,3$)라고 하고 H_k 를 문장성분 k 가 센터가 될 확률이라고 하자. 그러면 우리는 $P(H_k)$ 와 $P(A_i|H_k)$ 를 알고 있으며 베이시안 확률공식에 의해서 사후확률 $P(H_k|A_i)$ 를 계산해낼 수 있다. 즉 어떤 조건 A_i 를 갖는 문장에서 주제어나 주어와 같은 문장성분이 센터가 될 사후확률을 계산하여 그 값에 의해 C_f 의 우선순위를 결정할 수 있다.

5. 결론

자연언어의 내용과 생각을 포함한 지시 해결은 자연어처리 시스템의 완성도를 높이기 위해서는 물론이고 언어의 근본 메커니즘을 이해하기 위해서도 그 규명이 불가피한 언어 현상이다. 그런 이유로 많은 학자가 다양한 방법론을 동원해 그 해결을 모색하고 있지만 아직도 아찔다할 해결책이 제시되지 못하고 있다. 이런 가운데 Grosz, Sidner 등이 센터링 이론을 제안하여 학계에 활발한 논의가 진행되고 있으며 자연어처리 시스템에도 적극 활용되고 있다. 국내에서도 포항공대 등에서 지시 해결을 위해 센터링 이론을 도입하고 있다.

본고에서는 과연 이 센터링 이론이 한국어에 그대로 적합하게 적용될 수 있는지를 검토하고 영어와 일본어를 대상으로 제안된 기준이 한국어에는 적합하지 않음을 확인했다. 아울러 한국어의 내용과 생각 현상을 해결하기 위해서는 베이시안 확률을 이용해서 C_f 목록의 순서를 정함이 더 합리적이라는 것을 보았다. 보다 더 구체적이고 심도있는 논의는 앞으로의 연구를 통해 밝혀보려 한다.

참고 문헌

- [1] Grosz, Barbara J., 1977, 'The representation and use of focus in dialogue understanding', Technical Report 151, SRI International, Menlo Park.
- [2] Sidner, Candace L., 1979, *Towards a computational theory of definite anaphora comprehension in English discourse*, Doctoral dissertation, Artificial Intelligence Laboratory, MIT Cambridge.
- [3] Grosz, Barbara J. & Candace L. Sidner, 1986, 'Attentions, intentions and the structure of discourse', *Computational Linguistics*, 12, pp. 175-204.
- [4] Joshi, Aravind K. & Scott Weinstein, 1981, 'Control of inference: Role of some aspects of discourse structure-centering', *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 385-387.
- [5] Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein, 1983, 'Providing a unified account of definite noun phrases in discourse', *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pp. 44-50..
- [6] Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein, 1995, 'Centering: A Framework for Modeling the Local Coherence of Discourse', *Computational Linguistics*, 21-2, pp. 203-225.
- [7] Brennan, Susan E., Marilyn Walker Friedman & Carl J. Pollard, 1987, 'A Centering Approach to Pronouns', *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pp. 155-162.
- [8] Walker, Marilyn, 'Centering, Anaphora Resolution, and Discourse Structure', 1997, *Centering In Discourse*, eds. Marilyn A. Walker, Aravind K. Joshi and Ellen F. Prince, Oxford University Press.
- [9] Strube, Michael & Udo Hahn, 1996, 'Functional Centering', *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*.
- [10] Allen, James, 1995, *Natural Language Understanding*, The Benjamin/Cummings Publishing Company, Redwood City.
- [11] Walker, Marilyn, Masayo Iida & Sharon Cote, 1990, 'Centering in Japanese Discourse', *Proceedings of the 13th International Conference on Computational Linguistics(Coling 90)*.
- [12] Walker, Marilyn, Masayo Iida & Sharon Cote, 1994, 'Japanese Discourse and the Process of Centering', *Computational Linguistics*, 20-2, pp. 193-232.
- [13] Yongkyoon No, 1991, 'A centering approach to the *[CASE][TOPIC] restriction in Korean', *Linguistics*, 29, pp. 653-668.
- [14] Li, Charles & Sandra Thompson, 1976, 'Subject and Topic: A New Typology of Language', *Subject and Topic*(ed), N.Y.:Academic press.
- [15] Mitkov, Ruslan, 1995, 'Anaphora Resolution in Natural Language Processing and Machine Translation', IAI Working Paper.
- [16] 김영희, 1978, '접주어론', 한글, 162.
- [17] Yang, In-seok, 1973, 'Semantics of Delimiters in Korean', *Language Research*, 9-2.
- [18] 정인상, 1980, '현대국어의 주어에 대한 연구', 국어연구(서울대), 44.