

## 중심어 주도 단방향 차트 파싱을 이용한 문맥 대응어 해결

°김정해\*, 조준모\*\*, 이상국\*\*\*, 이상조\*\*  
상지전문대학 전산정보처리과\*, 한국MJL(주)\*\*\*,  
경북대학교 컴퓨터공학과\*\*

### A Resolution of Text Anaphora using Unidirection Chart Parsing in HPSG

Jung-Hae Kim\*, Jun-Mo Jo\*\*, Sang-Kook Lee\*\*\*, Sang-Jo Lee\*\*  
Dept. of Computer Information Processing, Sangji Junior College\*,  
MJL Co.\*\*\*, Dept. of Computer Engineering, Kyungpook Nat'l Univ.\*\*

#### < 요약 >

대응어(anaphor)는 한 문장이나 문장과 문장간에 같은 요소가 되풀이될 때 언어 사용의 경제성(language economy)을 위하여 잉여적 표현을 제거하는 방법으로, 좀 더 간략한 언어 표현으로 대체하여 쓰는 현상이다. 따라서 본 논문에서는 중심어 주도의 단방향 활성 차트 파싱을 이용하여 한국어 문장내에서 야기되는 문맥 대응어의 해결 방안에 대해 제안한다. 이는 자연어를 입력으로 하는 실용목적의 자연어처리 시스템 구축에 있어 필수적으로 요구되는 부분이다.

대응어 해결을 위해 먼저 전산학적인 대응어 정의를 내리고, 대응어와 선행어사의 의미 분류 및 대응어 해결 과정에 필요한 처리 조건등을 설정하였다. 또한 파서내에 대응어 처리를 위해 사전내 자질구조로 ANAPMAJ, ANAPMIN, PERSON, NUM, INDEX자질을 추가하였고, 대응어 해결을 위한 알고리즘을 제안하였으며, 기존에 개발된 HPSG 파서가 처리하는 모든 문장에서 야기된 문맥 대응어를 해결하여 파서이후의 응용 시스템에서 이용할 수 있는 내적표현을 보다 분명하게 형식화하였다.

#### I. 서론

대응어(anaphora)는 자연어 대화에서 만연된 현상이면서도 대응의 범위가 매우 다양하고 미묘하게 발생된다는 점에서 오랫동안 이론 및 계산 언어학 분야에서 중요한 주제로 연구되어 왔다 [1-3, 8, 9, 11-15]. 특히 최근에 대화형 컴퓨터 시스템에서 자연어 인터페이스에 대한 상당한 관심이 대두되었고, 이러한 자연어 인터페이스 개발의 주요 동기는 사용자가 특별한 인공언어를 배우는 대신에 자연어로 컴퓨터 시스템과 더 자연스럽게 쉽게 대화할 수 있다는 신뢰성때문이었다. 그러나 컴퓨터상의 자연어 인터페이스가 자연어에 대두되는 여러 가지 현상 특히 화용적인 현상인 대응, 생략 등을 다루어야 하기 때문에 자연어를 컴퓨터의 내부형태(internal form)로 분석하기 위한 능력은 자연스럽게 효과적인 인터페이스를 만들기에는 충분하지 못했다.

본 논문에서는 한국어의 다양한 대응어 현상중에서 하나의 문장내(intra-sentential)에서 야기되는 문맥 대응어의 해결로 조응대응 중심으로 그 선행어(antecedent)를 찾는 것에 중점을 두었다. 이러한 처리는 체언 대응의 해결에 중점을 둔 [2]의 논문을 확장한 것으로 통합기반 문법의 한 유형이며 한국어의 특징을 잘 나타낼 수 있는 문법이론인 HPSG를 이용한 중심어 주도의 단방향 활성 차트 파서를 구현하였다. 따라서 본 논문의 시스템은 기존에 개발된 HPSG 파서가 처리하는 모든 문장에서 야기된 대응어를 해결할 수 있도록 파서의 처리 범위를 확대하여 자연어 처리에 필수적인 화용처리에 보다 접근한 방법으로 대부분의 실용적인 국어 문장에 대해서 분석할 수

1) 선행어는 되풀이된 요소를 대응어로 바꾸는 작용을 한 요소이고, 대응어으로써 대체되는 요소를 대응어의 분디말(full form)이 된다.

있는 파서를 구현하였다.

본 논문의 구성은 II장에서는 전산학적인 대응어의 정의를 내리고, 이 정의를 바탕으로 대응어의 선행어를 찾기 위한 대응어와 선행어사이의 의미 분류 및 대응어 해결 과정에 필요한 처리 조건등을 설정하였다. III장에서는 대응어의 해결을 위해 추가된 자질 구조, 알고리즘 및 처리예문을 보였다. IV장에서는 연구 결과와 앞으로 과제를 제시하였다.

## II. 전산학적인 대응어 현상

### 1. 대응어의 정의

대응어의 전산학적인 정의를 위해 먼저 국어학에서의 대응어 정의, 분류, 유형은 [2]에서 논의하였고 또한 국어학적으로 지시사 「이, 그, 저」 표현의 특성이 이미 [2,9]에서 논의한 것과 같이 전술문과 조응하여 사용될 때 대응어에 쓰이는 대응어는 지시사 「그」+대용사 「이, 너, 남자, 것, 어기, ...」 등에 의해 결합된 조응대용만을 말하므로 본 논문에서는 다음을 전산학적인 대응어로 정의한다.

<<정의 : 전산학적인 체언류의 대응어>>	
대용이란 말의 경제성에 따라 반복하여 나타난 용어에 대해 대신하여 쓰는 것으로 '지시사+대용사'의 구성을 지니며, 특히 「지시사「그」 + 대용사「이, 너, 늑, 남자, 것, -어기」로 된 경우로 언어적 문맥에서 선행어와 대응어와의 관계로써 나타나는 조응대용을 전산학적인 대응어이다. 따라서 대응어의 구조는 다음과 같다.	
<ul style="list-style-type: none"> <li>■ 대응어의 구조 = '지시사 「그」' + '대용사'</li> </ul>	

한국어는 어휘의 특성상 대치되는 대응어의 종류를 여러가지 형태로 정의할 수 있으나, 선행어의 의미에 따라 체언류 대응어중에서도 사람 대응인 경우 '자기, 그, 그사람, 그분, 그녀, 그남자, 그이'가 쓰이고, 사물 대응인 경우 '그것, 그'가 쓰이며, 위치·방향 대응인 경우 '그리, 거기'가 쓰이며, 관형어 대응어인 경우 '그런'이 쓰인다.

### 2. 대응어와 선행어의 의미분류 관계

대응어는 본래 고유한 실질적 의미가 없는 것을 특징으로 하며, 항상 선행어에 의해 그 뜻이 해석된다. 또한 대응어는 선행어가 가지는 구체적인 뜻대신에 그것을 형식적으로 가리키기만 하는 부류적 의미를 지닌 것이 특징이다. 따라서 대응어와 선행어사이의 의미관계를 다음과 같이 나타내었다.

선행어의 의미분류		대응어
체언류 대응	사람-재귀	자기
	사람	그, 그남자, 그녀, 그이
	사물	그것, 것
	방향, 위치	그리, 거기
수식어 대응	관형어	그런
	부사어	그리

### 3. 대응어의 해결 조건

앞 절에서 나타난 대응어의 정의와 그에 따른 대응어의 종류 및 선행어와의 의미관계에 의해 선행어가 대응어에 대한 가능성 있는 선행어가 되기 위한 조건은 [1,9,12]를 근거로 [2]에서 정리로 나타내었고, 다음의 각 절에서 선행어의 선정 범위와 대응어의 처리 조건을 설정하였다.

1) 선행어의 선정 범위

대용어가 지칭하는 선행어의 범주는 '그'에 의한 조용대용에서 그 지시대상의 유일성, 특정성이 언어적 문맥에 의해서만 보증됨을 나타내므로 '그'가 붙은 명사구의 지시대상이 선행하는 언어적 문맥에 의해 특정되는 대상에서 핵심성분이 된다. 즉, '수식어+명사' 구조가 대용되었을 경우 이 구조의 핵심성분인 명사가 대용어의 선행어가 된다. 따라서 III장의 처리 예문에서 보여줄 체언-사람-재귀화 예문 「한 소년이 자기의 책을 읽는다」에서 대용어 「자기」의 선행어는 명사구에서 핵심성분인 「소년」이 된다.

2) 대용어의 처리 조건

II 장 1절과 앞 절에서의 정리 및 선행어의 선정 범위에 의해서 하나의 문장이 아래와 같은 성분으로 이루어져 있을 때 대용어(NPb)에 대해 선행어(NPa)를 찾기 위한 조건을 다음과 같이 나타내었다. 여기서 S는 하나의 문장을 의미하고 문장 S의 성분으로 NPa는 선행어, NPb는 대용어, X와 Y는 각각 임의의 문장 성분요소가 된다.

$$S : NPa - X - NPb - Y$$

[대전제조건1] 대용어 NPb의 왼쪽으로 선행어의 조건을 만족하는 가장 가까운 명사가 대용어 NPb의 선행어이다.

[대전제조건2] 대용어 NPb의 의미자질 값과 대상 선행어의 의미자질 값을 비교하여 동일한 값을 지니는 것을 대용어의 선행어로 한다.

만약 NPb가 '자기'라는 대용어가 쓰인 경우 NPa가 선행어가 될 조건

[조건1] NPa는 문장의 주어로써 3인칭 단수이다.

[조건2] NPa는 사람이거나 유정명사의 의미부류에 속해야 한다.

만약 NPb이 '대명사'의 대용어가 쓰인 경우

[전제조건] 대용어 NPb의 선행어로 결정되는 명사는 명사구의 핵심이어야 한다.

① NPb가 대용어 '그, 그사람, 그분, 그남자, 그녀,그이'인 경우 선행어의 조건

[조건1] 대상 선행어인 NPa는 3인칭 단수, 유정명사라야 한다.

[조건2] 대상 선행어인 NPa는 체언-사람을 지칭한다.

② NPb가 대용어 '그것, 것'인 경우 선행어의 조건

[조건1] 대상 선행어인 NPa는 무정명사라야 한다.

[조건2] 대상 선행어인 NPa는 체언-사물을 지칭한다.

③ NPb가 대용어 '그리, 거기'인 경우 선행어의 조건

[조건1] 대상 선행어인 NPa는 장소의 자질을 갖는 명사라야 한다.

[조건2] 대상 선행어인 NPa는 체언-방향 또는 체언-위치를 지칭하는 명사이다.

만약 NPb가 '관형어'나 '부사어' 대용어인 경우

[대전제조건] 대용어 NPb의 왼쪽으로 선행어의 조건을 만족하는 가장 가까운 관형어나 부사어가 대용어 NPb의 선행어이다.

① 만약 NPb가 '관형어'로 NPb의 대용어가 '그런'인 경우 선행어의 조건

[조건] 대용어 NPb의 선행어로 결정되는 어휘는 관형어 자질을 갖는 수식성분이다.

② 만약 NPb가 '부사어'로 NPb의 대용어가 '그리'인 경우 선행어의 조건

[조건] 대용어 NPb의 선행어로 결정되는 어휘는 성분 수식 또는 문장 수식을 하는 부사어 자질을 갖는 수식성분이다.

### III. 대용어의 해결

#### 1. 중심어 주도의 차트 파싱

본 논문의 파서에서는 대용어의 해결을 위해 문법은 HPSG를 채택[10]하고 중심어 주도의 단방향 활성 차트 파서[4,5]를 이용한다. 이는 한국어의 경우 선행어는 항상 대용어 앞에 존재하기 때문에 위에서 좌로의 단방향 파서를 이용할 경우 선행어를 찾는 데 유리하기 때문이다.

따라서 본 논문에서 대용어의 해결을 위해 고려한 대상은 문맥 대용어를 담고 있는 모든 종류의 한 문장이지만, 실제 실험 대상은 본 논문의 파서에 의해 파싱 가능한 문장으로 한다. 본 논문 파서의 파싱 범위는 보어-중심어, 수식어-중심어, 중심어-중심어 구성의 거의 모든 실용적인 문장에 대해 분석가능하고, 파서의 입력은 형태소 분석이 이루어진 결과라고 가정하며 출력은 입력 받은 문장에 대해서 통사 및 의미를 통합적으로 처리하고 입력문장에서 추출한 정보를 파서이후의 응용 시스템에서 이용할 수 있는 내적표현으로 형식화하였다.

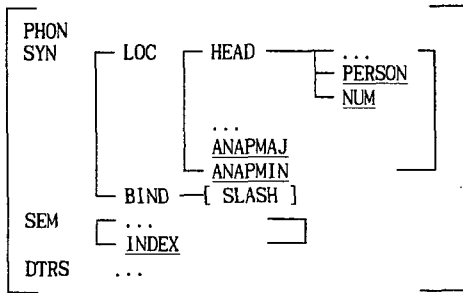


그림1. 자질 구조

(그림1) 자질구조의 자세한 설명은 [2,4,5]에서 나타내었고, 여기서는 추가한 자질만을 기술하였는데, 이 자질구조를 이용해서 언어정보를 형식화하였다. 먼저 PERSON자질은 인칭을 나타내는 자질로서 1인칭(first), 2인칭(second), 3인칭(third)의 값을 갖는다. NUM자질은 수를 나타내는 자질로서 단수(single), 복수(couple)의 값을 갖는다. ANAPMAJ자질은 체언류와 수식어의 대용어 유형중 '1'은 체언류의 대용, '2'는 관형어 대용, '3'은 부사어 대용의 값을 나타낸다. ANAPMIN자질은 체언류 대용어 유형을 나타내는 자질로서 1은 '자기'의 재귀화 대용, 2는 '그것' 등의 사물 대용, 3은 '그리, 거기' 등의 방향, 위치 대용, 4는 '그, 그사람, 그이, 그분' 등의 대명사화 대용을 나타낸다. INDEX자질은 대용어의 경우 선행어 노드를 지칭하는 자질이다.

#### 2. 알고리즘

본 논문에서는 대용어의 해결을 위한 파서로 중심어 주도의 상향식 차트 파서[5]를 사용한다. 이 파서의 대표적인 자료구조는 다음과 같으며, 여기서 각 edge의 구조는 시작노드 번호, 종료노드 번호, 자질구조 및 다음 노드에 대한 포인터들로 구성된 구조를 가진다.

- INPUT : 입력 edge들의 모임
- PEND : 처리가 완료되지 않은 edge들의 모임
- CHART : 처리가 완료된 edge들의 모임

(그림2)는 본 논문에서 제안한 대용어 해결의 흐름도를 기술한 것으로 HPSG의 전체 파서 알고리즘에서 [2]의 흐름도를 수정·보완하여 간략히 나타냈다.

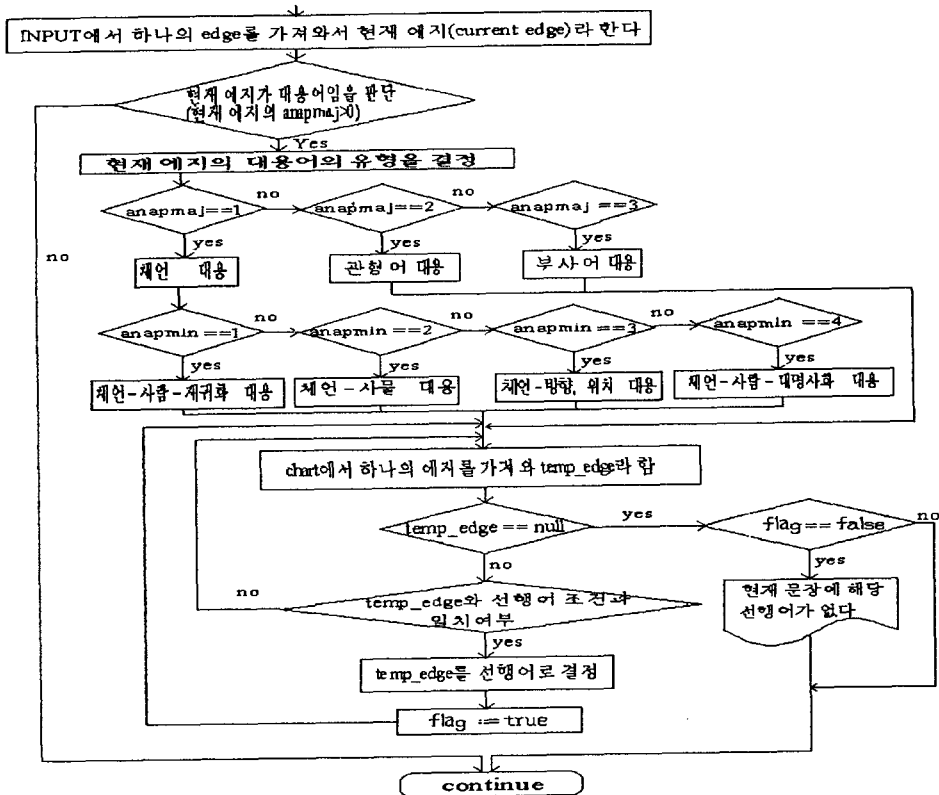


그림 2. 대응어의 해결을 위한 흐름도

### 3. 처리 예문

본 절에서는 대응어의 해결을 위해 구현한 알고리즘에 따라 처리 예문을 기술하였는데, 처리 예문 및 상세한 처리과정은 지면 관계상 몇가지에 대해서만 기술하였다.

#### 1) 체언-사람-재귀화 대응

##### (a) 예문

[ 한 소년이 자기의 책을 읽는다.]

##### (b) 선행어의 조건

- ① fs->sem.sm.anim == '+' (생물)
- ② fs->syn.loc.head.person == third(3인칭)
- ③ fs->syn.loc.head.num == single(단수)
- ④ fs->syn.loc.head.gr == subj(주어)

##### (c) 처리 과정

[ 한 소년이 자기의 책을 읽는다.]

edge1    edge2    edge3    edge4    edge5

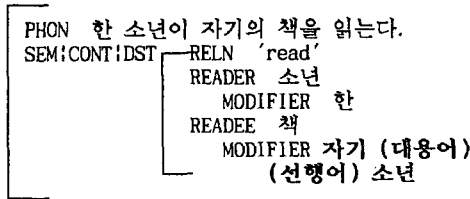
위의 예문이 입력되었을 경우 edge2까지 파싱된 결과로써 CHART에는 (edge1, edge2, edge1+edge2)가 존재한다. 그 다음 입력에서 하나의 edge를 가져와 처리를 계속할 때 edge3[자기의]가 현재예지(current edge)가 되고, 이 예지의 anapmaj자질값(anapmaj=1)에 의해 현재 예지가 대응어임을 알 수 있다. 그래서 앞에서 제시한 대응어 처리 알고리즘에 따라 anapmaj자질값이 1 이므로 이 대응어가 재귀화 대응어를 결정하고 지금까지 처리가 완료된 예지들이 들어있는 CHART

를 탐색해서 위의 선행어의 조건에 부합하는 에지를 이 에지의 선행어로 결정하게 된다.

이 예문의 경우 edge1과 edge2가 결합된 edge1+edge2[한 소년]이 선행어의 조건을 만족하므로 먼저 선행어 후보가 되고, 아직 CHART에 에지들이 남아 있으므로 탐색을 계속한다. 이때 edge2[소년이]가 또한 선행어의 조건을 만족하므로 선행어 후보로 등록된다. 본 논문의 대용어 선행어 선정 범위는 「수식어-중심어구조의 의미 합성」이 이루어지지 않는 단독 중심어 「LEX+」를 선행어로 선정한다. 이는 본 논문에서 사용하는 자질구조의 자질속성, LEX를 이용하여 처리할 수 있다. 즉, edge1+edge2는 fs->syn.loc.lex가 '-'값을 가지며 edge2는 '+'값을 갖는다. 따라서 edge2[소년이]가 선행어로 결정된다.

대용어 처리에 직접 관련되지 않는 다른 구성요소의 처리 및 본 논문의 처리 결과에 사용된 의 미표현 및 의미합성 원리는 중심어 주도의 단방향 활성화차트파서[5]의 파싱결과에 따르는 것이므로 본 논문에서는 생략하기로 한다.

(d) 처리 결과



2) 관형어 대응

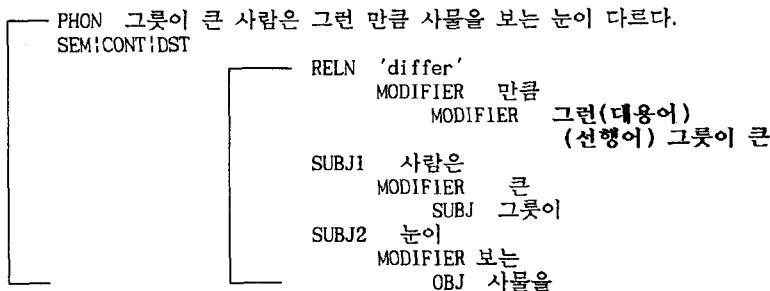
(a) 예문

[그릇이 큰 사람은 그런 만큼 사물을 보는 눈이 다르다.]

(b) 선행어의 조건

- ① fs->syn.loc.head.maj == verb
- ② fs->syn.loc.head.modal.vform == adj

(c) 처리 결과



IV. 결론

본 논문에서는 자연어 사용에 있어 빈번히 발생하는 대용어 현상에 대한 해결 방안으로 중심어 주도의 단방향 활성화 차트를 이용하였다. 대용어의 해결을 위해 먼저 전산학적인 대용어의 정의 및 선행어와 대용어의 관계, 처리조건을 설정하였다. 또한 시스템의 구현을 위해 국어와 같이 중심어 후행 언어에 대하여 효율적인 중심어 주도의 단방향 활성화 차트 파싱을 이용하여 문맥 조응 대용어 해결을 구현하였다. 따라서 본 논문에서는 문장내에서 발생하는 대용어에 대해 대용어가 지칭하는 내용인 선행어를 찾아 보다 명확한 문장의 분석을 통해 자연어 처리에 필수적인 화용처리에 보다 접근한 방법으로 대부분의 실용적인 국어 문장에 대해서 분석할 수 있는 파서를 구현

하였다.

현재 한 문장내 대용어 처리뿐만 아니라 문장과 문장간의 대용어 처리에 대해서도 계속 연구중에 있다.

< 參考文獻 >

- [1] 김일웅, 우리말 대용어의 연구, 부산대 박사학위논문, 1982
- [2] 김정해, 이상국, 이상조, "한국어 문장내 체언류 조용대용어의 해결방안", 대한전자공학회 논문지, 제33권 B편 제4호, pp.183-190, 1996
- [3] 유현경, "국어 접속문의 통사적 특질에 대하여", 한글 제191호, pp.77-104, 1986
- [4] 서영훈, 의미정보를 이용하는 중심어 주도의 한국어 파싱, 서울대 컴퓨터공학과 박사학위 논문, 1991
- [5] 이상국, 특수문형의 처리를 강화한 HPSG 한국어 파서의 구현, 경북대 컴퓨터공학과 석사학위 논문, 1993
- [6] 이상조, 한국어 자연어 인터페이스를 위한 사전 구성에 관한 연구(II), 한국전자통신연구소 위탁과제, 경북대 전자기술연구소, 1992
- [7] 이정현, 한국어 처리를 위한 구절 문법과 질의 응답 시스템, 인하대학교 박사학위논문, 1988
- [8] 허정, 국어 대용어해결 시스템의 설계 및 구현, 한국과학기술원 석사학위논문, 1985
- [9] 홍순성, 국어 대명사의 조용현상에 관한 연구, 영남대 박사학위 논문, 1986
- [10] C.Pollard and I.Sag, An Introduction to HPSG, 1987, Jan, Draft
- [11] James Allen, Natural Language Understanding, The Benjamin/Cummings Pub.Company, Inc., 1987
- [12] Joke Dorrepeal, "Discourse Anaphora", In Proceedings of Colling90, Vol. II, pp.95-99, 1990
- [13] J.G.Carbonell, R.D.Brown, "Anaphora Resolution : A Multi-Strategy Approach", In Proceedings of Colling Budapest, Vol.1, pp.96-101, 1988
- [14] Philip J. Hayes, "Anaphora for Limited Domain System", IJCAI-81, Vol.1, pp.416-422, 1981
- [15] T.Ukita, K.Sumita & S.Amano, "A Procedure Resolving Anaphoric Ambiguity by Finding the Most Informative Interpretation", ICOT TR-220, 1986, Dec.