

한국어 복합문에서의 제로 대용어 처리를 위한 분해 알고리즘과 복원규칙

(Splitting Algorithms and Recovery Rules for Zero Anaphora Resolution in Korean Complex Sentences)

김 미 진 [†] 박 미 성 [†] 구 상 옥 [†] 강 보 영 ^{**} 이 상 조 ^{***}
(Mi-Jin Kim) (Mi-Sung Park) (Sang-Ok Koo) (Bo-Yeong Kang) (Sang-Jo Lee)

요 약 한국어에는 복합문에서 영 대용어 빈번하게 발생하여 해석을 어렵게 한다. 따라서 본 논문에서는 한국어 영 대용어 처리를 위해 복합문 분해 알고리즘과 복합문 영 대용어 복원 규칙을 제안하고, 해결방법을 제시한다. 본 논문은 신문 기사의 복합문 중에서 보조용언 내포문을 제외한 인용문, 접속문, 내포문을 처리 대상으로 한다. 복합문 분해를 위해서는 복합문 구성에 관여하는 어미들의 어미 분류표를 이용하고, 영 대용어 복원을 위해서는 생략될 때 적용된 통사규칙을 역으로 이용한다. 인용문은 주어 인칭제약에 따른 동일 명사구 탈락규칙을, 명사화 내포문은 동일 명사구 탈락규칙을, 관형화 내포문은 관계 명사구 탈락규칙을 그리고 접속문은 접속 삭감규칙을 역으로 이용하여 처리한다. 제안한 방법을 이용한 결과 전체 영 대용어 중 83.53%가 해결 가능하며 11.52%는 부분적으로 해결 가능하다.

키워드 : 영 대용어, 복합문 분해알고리즘, 영 대용어 복원규칙, 어미 분류표, 통사규칙

Abstract Zero anaphora occurs frequently in Korean complex sentences, and it makes the interpretation of sentences difficult. This paper proposes splitting algorithms and zero anaphora recovery rules for the purpose of handling zero anaphora, and also presents a resolution methodology. The paper covers quotations, conjunctive sentences and embedded sentences out of the complex sentences shown in the newspaper articles, with an exclusion of embedded sentences of auxiliary verb. We manage the quotations using the equivalent noun phrase deletion rule according to subject person constraint, the nominalized embedded sentences using the equivalent noun phrase deletion rule, the adnominal embedded sentences using the relative noun phrase deletion rule and the conjunctive sentences using the conjunction reduction rule in reverse. The classified table of the endings which relate to a formation of the complex sentences is used for splitting the complex sentences, and the syntactic rules, applied when being omitted, are used in reverse for recovering zero anaphora. The presented rule showed the result of 83.53% in perfect resolution and 11.52% in partial resolution.

Key words : zero anaphora, splitting algorithms, recovery rules, The classified table of the endings, the syntactic rules

1. 서 론

[†] 비 회 원 : 경북대학교 컴퓨터공학과

jean321@hanafos.com

mspark@knu.ac.kr

tomato@sejong.knu.ac.kr

^{**} 학생회원 : 경북대학교 컴퓨터공학과

comeng99@hotmail.com

^{***} 종신회원 : 경북대학교 컴퓨터공학과 교수

sjlee@bh.knu.ac.kr

논문접수 : 2002년 2월 20일

심사완료 : 2002년 8월 9일

자연언어에는 한 문장 속에 동일한 구조가 둘 이상 있을 때 이 중에서 하나만을 남기고 나머지 것은 생략을 하거나 더 짧은 다른 표현으로 대체하는 과정이 있다. 이러한 생략과 대체는 같은 유형의 언어 현상으로써, 대체는 어휘 문법적 견지에서 볼 때 단어와 구와의 관계요, 생략은 이러한 대체의 특수한 경우로써 영(null)대치라 볼 수 있다[1]. 따라서 생략은 기존 문장에 실현되거나, 있어야 할 것으로 상정된 성분이 표면구조에 나타나지 않은 언어적 현상으로서 많은 연구에서 대용의 한 범주로 다루어지고 있다[2,3]. 이러한 영 대용(생략) 현상 해결을 위해 국

어학, 언어학 분야[1,3,4,5,6,7,8]의 이론적인 연구와 전산학적[9,10,11,12,13,14,15,16,17,18]으로 많은 연구가 되어왔다. 국어학 분야에서는 변형규칙에 의하여 국어의 생략현상과 생략의 조건을 밝히거나, 변형생성문법의 틀 안에서 생략과 대용에 따른 제약조건을 제시하는 연구가 있었고[1], 언어학 분야에서는 중심화이론(Centering)의 개념과 확대 정보포장 이론을 접목시켜서 영역이 제한된 대화인 호텔 예약대화에 나타나는 한국어의 영 대명사의 선행사를 찾는 연구가 있었다[7,8,15]. 한편 전산학 분야에서 국내 연구로는 중심화 이론과 개념 그래프를 이용하여 체언의 대용과 생략을 시도한 연구가 있었고[9,10], 센터링 이론이 한국어 대용과 생략을 해결하기 위한 기제로서 적합한가를 검토하고, cf 목록순서를 한국어의 언어특성에 맞게 결정하는 기법[11] 및 HPSG 파서에 기반하여 한국어 조용대용어를 해결한 연구도 있었다[12,13]. 국외에서는 Slot Grammar를 이용하여 대용과 삭제 해결을 위한 3가지 경우의 알고리즘을 제안하고[14], 중심화이론에 기반한 일본어에서의 담화해석과 구문적 단서간의 연관관계를 밝힌 연구[15,16,17] 및 문서구조와 의미, 실용적 제약을 사용하여 일어 영 대명사를 해결한 연구도 있다[18]. 그러나 국외에서의 영 대용 연구의 활발함에 비해 국내에서의 연구는 아직도 독립된 부분으로 영 대용을 다룬 연구가 거의 없는 실정이다.

한국어에서의 영 대용은 문장과 문장이 이어지는 경우나 안기는 경우에 빈번하게 일어나는 보편적 현상이므로 자연어 처리의 한 부분인 기계번역 시스템이나 자동 문서 요약 시스템 구축을 위해서는 영 대용어 처리능력이 필수적으로 요구된다.

본 논문에서는 위에서 기술한 언어 현상 중에서 문장에서 야기된 생략되어 있는 성분을 대용어 대치현상 중의 하나인 영 대용으로 보고, 한국어 복합문에서의 영 대용 현상을 처리하는 시스템을 구축한다. 영 대용어 해결을 위해 생략될 때 적용된 생략 규칙을 역으로 이용하고, 복합문 구성에 관여하는 어미에 따라 각각 다른 복원 알고리즘을 수행하여 생략된 성분을 복원한다.

본 논문의 구성은 다음과 같다. 2장에서는 한국어 복합문의 구성과 통사규칙을 살펴보고 복합문 구성에 관여하는 어미들을 분류한다. 3장에서는 복합문에서의 영 대용어 해결을 위해 복합문 판별과정과 복합문을 EE(Elementary Event)로 분리하는 분해과정, 그리고 제안한 복합문 복원 규칙과 알고리즘에 대하여 설명하고, 그 해결 과정을 예문을 통하여 보인다. 그리고 4장에서는 실험결과 및 분석을, 5장에서는 연구 결과와 향후 과제를 제시한다.

2. 영 대용어 해결을 위한 한국어 복합문 고찰

이 장에서는 본 논문에서 처리할 한국어 복합문을 정의하고 생략 시에 적용되는 통사규칙과 복합문 구성에 관여하는 어미들을 분류한다.

2.1 복합문

서술기능을 한 번 수행하는 문장구성을 '단순문'이라고 하고, 서술기능을 두 번 이상 수행하는 문장구성은 '복합문'이라 한다[19]. 다음의 표는 복합문 구성체계를 보여주고 있다.

표 1 복합문 구성체계



본 논문에서는 [표 1]의 구분에서 보조용언 내포문(★)을 제외한 복합문만 다루며 인용구문 내포문은 '인용문'으로, 명사구 내포문은 '내포문'으로 접속문은 그대로 '접속문'으로 하고 다음과 같이 정의한다.

(1) 접속문

접속문은 선행절과 후행절로 구분되고, 선·후행절의 의미관계가 대등한가, 아니면 후행절이 의미의 중심이 되는가에 따라 '대등 접속문'과 '종속 접속문'으로 나눈다.

(2) 내포문

내포문은 상위문과 내포절로 구분되고, 명사화 어미 'ㄱ, 음, 기'와 함께 사용되는 명사화 내포문은 명사화 어미에 의해 술어에 연결되고, 관형화 어미 '은, 는, 을'과 함께 사용되는 관형화 내포문은 관형화 어미에 의해 피수식어(명사 또는 명사구)에 연결된다.

(3) 인용문

인용문은 상위문과 인용절로 구분되고, 직접 인용어미 '-(이)라고' 또는 '-하고', 간접 인용어미 '-고'에 의해서 술어에 연결된다.

2.2 복합문 어미 분류

한국어는 어미에 의해 문장의 문법적인 관념들이 다양하게 실현된다. 복합문에는 접속문 어미, 내포문 어미, 인

표 2 어미 유형에 따른 어미분류 테이블

문장 유형		어미 유형	어미	
접속문	대등 접속문	접속어미	면서, 거나, 든지, 고, (으)며, 으나, 지만, 는데, 는게	
	종속 접속문	접속어미	느라고, 니까, 으러, 으려고, 거든, 면(서), 며, 고(서), 어/아(서), 마자, 으나, 아도, 더라도, 러 지라도, 아야, 도록, 러, 러(고), 고자, 므로, 다고, 된들, 자(마자), 게, 염표(,)	
내포문	관형화 내포문	관형화 어미	-은, -(다/라/냐)는, -을, -ㄴ, -ㄹ	
	명사화 내포문	명사화 어미	-음, -기, -ㅁ, -ㄴ/ㄹ 것	
인용문	직접·간접 인용문	인용어미	-고, -(이)라고, -하고	
		종결어미	명령형	-(으/아/어)라
			의문형	-(으/느)냐, -(느)ㄴ가, 나
			청유형	-자
			서술형	-(는/ㄴ/겠)다, -라, -(는)구나(감탄형), -(으)마(약속형) -ㄴ 것

용문 어미가 관여하는데, 인용문은 종결어미를 온전히 갖춘 채로 내포되는 구성이어서 문장 종결어미의 제약도 받는다. 아래의 [표 2]는 복합문 구성에 관여하는 어미들의 분류[5]에 본 논문 처리에 필요한 어미 몇 가지를 더 첨가하여 완성한 어미분류 테이블이다. ‘...명사류, 명사류’와 ‘~등’ 내부의 **염표**는 접속어미에서 제외시킨다.

본 논문은 [표 2]의 복합문 구성에 관여하는 어미에 따라서 서로 다른 분해 알고리즘을 적용시키게 된다.

2.3 변형(통사) 규칙[4]

국어의 생략현상을 다루는 변형규칙으로는 다음의 규칙들이 있다.

(1) 접속 삭감 규칙(Conjunction Reduction Rule)

접속문이 형성될 때 접속성분의 일부(동일 성분)가 탈락되는 규칙이다.

예) 영화는 자전거를 타고, 철수는 \emptyset 밀었다. (\emptyset =자전거를)

(2) 관계 대명사 탈락 규칙(Relative Noun Phrase Deletion Rule)

관계절의 수식을 받는 명사구와 관계절 내의 어떤 명사구와의 사이에 상호지시성이 성립될 때 적용되는 규칙이다.

예) 순경이 [\emptyset 보석을 훔친] 도둑놈을 잡았다. (\emptyset =도둑놈이)

(3) 동일 명사구 탈락 규칙(Equivalent NP Deletion Rule)

일정한 조건 아래서 상위문의 주어와 하위문의 주어를 삭제시키는 통사규칙이다.

조건 1) 상위문장의 주어와 보문의 주어가 같을 때

예) 영화는 [\emptyset 성실한 사람이 되기를] 원한다. (\emptyset =영화, 자신이)

조건 2) 상위문장의 간접목적어(여격어)와 보문의 주어가 같을 때

예) 선생님이 학생들에게 [\emptyset 조용히 하라]고 말했다. (\emptyset =학생들이)

본 논문에서는 생략될 때 적용된 위의 규칙들을 역으로 문장의 생략성분을 복원하는 정보로 사용한다.

3. 복합문 영 대용어 해결 시스템

3.1 전체 시스템 구성도

복합문 영 대용어 해결을 위한 본 논문의 전체 시스템 구성도는 [그림 1]과 같다. 전체 과정은 전처리 과정과 영 대용어 해결 과정으로 나눌 수 있다. 입력으로는 태깅된 문서를 문장 단위로 받아들인다. 전처리 과정에는 본 논문의 대상인 복합문을 선택하기 위한 복합문 판별 과정과 복합문으로 판별된 문장을 EE(Elementary Event)단위로 분해하는 복합문 분해 과정, 그리고 생략된 성분인 영 대용어를 찾는 과정이 포함된다. 다음 단

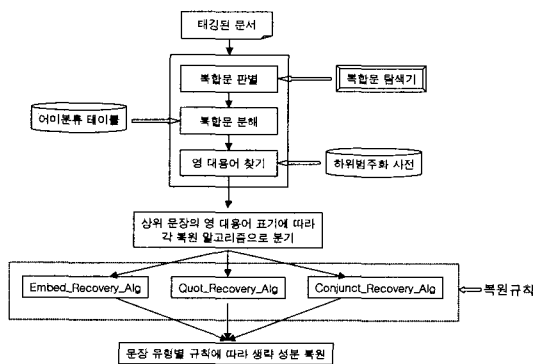


그림 1 영 대용어 해결을 위한 전체 구성도

제인 영 대용어 해결과정에서는 상위문장의 영 대용어 표기에 따라 각 복원 규칙을 수행하는 과정과 문장 유형별 규칙에 따라 생략 성분을 복원하는 과정이 있다. 이 둘 과정에서 사용되는 정보는 명사 의미 사전, 어미 분류 테이블, 용언의 하위범주화 사전, 통사규칙, 영 대용어 복원 규칙이 있다.

3.1.1 입력

입력으로 세종 표지[20]를 이용하여 태깅된 문서를 문장단위로 받아들인다. 본 논문에서 필요로 하는 어미들을 위해서는 재 정의하거나 수동으로 세분화하였다.

세종표지	재정의	본 논문에서 사용한 posTag
마침표, 물음표 : SF	⇒	마침표 : SS.
따옴표, 괄호.. : SS	⇒	따옴표 : SS"
관형화 전성어미 : ETM	⇒	관형화 어미 : adn
명사형 전성어미 : ETN	⇒	명사화 어미 : nom
연결어미 : EC	⇒(새분화)	접속어미 : conj
인용어미 : 없음	⇒(생성)	인용어미 : quot

[표 3]은 아래의 예제 문장을 태깅한 결과이다.

3.1.2 복합문 판별

입력된 문장을 복합문 탐색기로 복합문인지를 판별한다. 문장 내의 서술어 수를 검사해 2개 이상이면 복합문으로 간주한다.

3.1.3 복합문 분해

1) 복합문 분해 정의

복합문에서 하나의 서술어는 하나의 사건을 나타내고 볼 수 있으므로, 이것을 기본적인 사건(Elementary Event)으로 본다[22]. 본 논문에서 다루는 생략은 두 개 이상의 EE가 접속되거나 내포되면서 발생하는 현상이므로 생략된 성분 복원을 위해서는 두 개 이상의 EE를 각각의 EE로 분리하여 처리할 필요가 있다. 이 과정

을 복합문 분해라고 한다. 복합문 분해는 하위범주화 사전 참조 회수를 줄이고 복원 규칙 적용을 용이하게 한다. 일반적으로 접속문에서는 후행절과 선행절, 내포문에서는 내포절, 그리고 인용문에서는 인용절에 생략현상이 일어나기 때문에 분리된 문장의 내포된 절의 서술어만 조사하면 된다. 복합문 분해 시에 위의 [표 2]에서 분류한 어미 분류 테이블을 참조한다.

2) 복합문 분해 알고리즘

복합문 분해 순서는 가장 상위문장부터 시작해서 내포된 문장으로 내려가며 수행한다. 분해 알고리즘에서 분해는 대괄호([' '])로 표시하고, 각 문장유형에 따른 문장성분들과 어미는 다음과 같이 축약하여 표기한다.

인용문 QS(Quotation) : 인용어미 quot(quotative endings)	
상위문 H	주어 Hsub, 서술어 Hpred, 목적어 Hobj 간접목적어 Hobj1
인용절 Q(Quotative clause)	주어 Qsub 목적어 Qobj 서술어 Qpred
종결어미	청유형 S(Suggestion), 서술형 D(Declarative) 명령형/의문형 : Im/In(Imperative/Interrogative)
내포문 ES (Embedded Sentence)	
명사화 내포문 NES(Nominalized ES) : 명사화 어미 nom(nominalized endings)	
명사화 내포절 NE (Nominalized Embedded Clause)	주어 NEsub 목적어 NEobj 서술어 NEpred
관형화 내포문 AES(Adnominal ES):관형화 어미 adn(Adnominal endings) 피수식어 MODee(Modifice)	
관형화 내포절 AE (Adnominal Embedded Clause)	주어 Aesub 목적어 AEobj 부사어 AEadv 서술어 Aepred
접속문 CS(Conjunctive S.) : 접속어미 conj(conjunctive endings)	
선행절 A(Antecedent clause)	주어 Asub 목적어 Aobj 서술어 Apred
후행절 C(Consequent clause)	주어 Csub 목적어 Cobj 서술어 Cpred

표 3 예문 태깅 결과

한국투자신탁은 거래소시장 활성화 대책이 중소형 우량주에 유리한 환경을 제공해 거래소 우량 중소형주들의 주가 상승에 도움이 될 것이라고 덧붙였다			
어절	태깅 결과	어절	태깅 결과
한국투자신탁은	한국/NNG+투자/NNG+신탁/NNG+은/JX	거래소	거래소/NNG
거래소시장	거래소/NNG+시장/NNG	우량	우량/NNG
활성화	활성/NNG+화/XSN	중소형주들의	중소형주/NNG+들/XSN+의/JKG
대책이	대책/NNG+이/JKS	주가	주가/NNG
중소형	중소형/NNG	상승에	상승/NNG+에/JKB
우량주에	우량주/NNG+에/JKB	도움이	도움/NNG+이/JKS
유리한	유리/NNG+하/XSV+ㄴ/adn	되	되/VV
환경을	환경/NNG+을/JKO	ㄴ 것이라고	ㄴ 것/EF+이라고/quot
제공해	제공/NNG+하/XSV+아/conj	덧붙였다.	덧/XPN+붙이/VV+있/EP+다/EF+./SS

다음은 복합문 분해 알고리즘이다.

```

/* 분해 알고리즘 */
Sentence sent;
Function Sentence Split_Sentence(sent)
  If !IsComplexSentence(sent) Then /* 복합문 판별 */
    return Sentence
  /* 인용문, 명사와 내포문, 접속문, 관형화 내포문 순으로
  문장유형을 체크하여 문장을 분해한다. */
  Else
    Sentence subsent;
    If IsQ_Sent Then /* 인용문 판별 및 분해 */
      subsent = Q_Split_Algorithm(sent);
      Split_Sentence(subsent)
    If IsNE_Sent Then /* 명사와 내포문 판별 및 분해
    */
      subsent = NE_Split_Algorithm(sent);
      Split_Sentence(subsent)
    If IsC_Sent Then /* 접속문 판별 및 분해 */
      subsent = C_Split_Algorithm(sent);
      Split_Sentence(subsent)
    If IsAE_Sent Then /* 관형화 내포문 판별 및 분해 */
      subsent = AE_Split_Algorithm(sent);
      Split_Sentence(subsent)
  End If
End Function

/* 명사와 내포문 분해 알고리즘 */
/* 한 어절씩 읽어가며 주어와 명사화 어미를 찾고,
  주어 다음 어절부터 명사화 어미까지를 분해한다. */
Function NE_Split_Algorithm(Sentence sent)
  Vector ejuls = sent.ejul
  For ( i=1; i<ejuls.size; i++ )
    If (sent.ejul[i] == IsSubject_JOSA)
      start = i+1
    Else If (sent.ejul[i] == IsNE_Ending)
      end = i-1
    End If
  End For
  return Make_SubSentence(sent, start, end)
End Function

/* 접속문 분해 알고리즘 */
/* 한 어절씩 읽어가며 접속어미를 찾고,
  문장의 처음부터 접속어미까지를 분해한다.*/
Function C_Split_Algorithm(Sentence sent)
  Vector ejuls = sent.ejul
  For ( i=1; i<ejuls.size; i++ )
    If (sent.ejul[i] == IsConj_JOSA)
      start = i
    End If
  End For
  return Make_SubSentence(sent, 1, start-1)
  + "+" + Make_SubSentence(sent, start+1,
ejuls.size)
End Function

```

3) 복합문 분해 알고리즘 적용 예
위의 복합문 분해 알고리즘에 따라 분해하는 예는 아래와 같다.

한국 투자 신탁은 거래소시장 활성화 대책이 중소형 우량주에 유리한 환경을 제공해 거래소 우량 중소형주들의 주가 상승에 도움이 될 것이라고 덧붙였다.

입력된 문장에서 활용 어미(Endings)를 찾는다.

(1) Endings = 'quot'이면 Q_Split_Algorithm 수행

한국 투자 신탁은 [거래소시장 활성화 대책이 중소형 우량주에 유리한 환경을 제공해 거래소 우량 중소형주들의 주가 상승에 도움이 될 것]이라고 덧붙였다.

quot '-이라고'를 찾았으므로 Hsub '한국 투자 신탁은' 뒤와 quot '_이라고' 앞에서 분해.

(2) Endings = 'nom' 이면 NE_Split_Algorithm 수행

입력된 문장에 'nom'이 없으므로 (3) 수행

(3) Endings = 'conj'이면 C_Split_Algorithm 수행

한국 투자 신탁은 [[거래소시장 활성화 대책이 중소형 우량주에 유리한 환경을 제공해] [거래소 우량 중소형주들의 주가 상승에 도움이 될 것]]이라고 덧붙였다.

conj '-아'를 전후하여 A(선행절)과 C(후행절)로 분리

(4) Endings = 'adn'이면 AE_Split_Algorithm 수행

한국 투자 신탁은 [[거래소시장 활성화 대책이 [중소형 우량주에 유리한] 환경을 제공해] [거래소 우량 중소형주들의 주가 상승에 도움이 될 것]]이라고 덧붙였다.

'유리한'의 adn '-ㄴ' 뒤에서 분리.

(5) 'adn'이 더 이상 없으므로 분해 중지

3.1.4 영 대용어(생략 성분) 탐색

복합문을 활용어미에 따라 분해 한 후에 용언의 하위 범주화 사전을 이용하여 생략된 성분인 영 대용어를 찾는다. 하위범주화 사전은 용언의 필수적 정보와 각각의 격 내에 올 수 있는 명사 의미 자질로 구성된다[22]. 하위범주화 사전은 경북대학교 자연어 처리 LAB에서 만든 사전을 이용했으며, [표 4]는 용언 '제공하'에 대한 예이다.

표 4 용언 '제공하'에 대한 하위범주화 사전

```

<Title>제공하</Title>
<CaseFrame>
  <Frame>N0가이 N1을( N2에) V</Frame>
  <SynSem>
    <SelRst>N0=사람|단체</SelRst>
  </SynSem>
</CaseFrame>

```

다음은 앞 단계에서 분해가 끝난 복합문에 대해 각 문장 유형별 생략 성분을 찾아 표시하는 방법을 보인다. 심볼 'Ø'는 영 대용어를 나타낸다. 영 대용어는 가장 상위문장부터 시작하여 내포된 문장으로 내려가며 찾는다.

1) 인용절이 conj로 분리되어 있으므로 앞뒤 절의 용언 '제공하'와 '되'의 하위범주화 성분을 조사하여 결여 성분을 표시한다.

한국 투자 신탁은 [[거래소시장 활성화 대책이 중소형 우량주에 유리한 환경을 제공해] [QØsub₁ 거래소 우량 중소형주들의 주가 상승에 도움이 될 것]]이라고 덧붙였다.

⇒ 용언 '되'의 '주어'가 결여 성분이므로 QØsub₁으로 영 대용어를 표시

2) 관형화 내포문의 용언 '유리한'의 하위범주화 성분의 결여성분을 찾아 표시한다.

한국 투자 신탁은 [거래소시장 활성화 대책이 [AEØsub₁ 중소형 우량주에 유리한] 환경을 제공해] [QØsub₁ 거래소 우량 중소형주들의 주가 상승에 도움이 될 것]]이라고 덧붙였다.

⇒ 내포절의 결여성분도 '주어'이므로 AEØsub로 표시한다.

3) 가장 상위에 표시된 QØsub₁의 Q를 보고 다음 단계인 인용문 복원루틴으로 분기한다.

3.2 복합문 영 대용어 복원 규칙

아래의 복원 규칙들에서 '>'는 왼쪽 성분이 오른쪽 성분으로 대체됨을, 규칙 반복 표기 시에 '+'는 한 번 이상의 반복을, '|'는 생략이 가능함을, 그리고 '|'는 '또는'을 의미한다.

3.2.1 접속문(Conjunctive Sentence)

'접속 삭감 규칙'을 역으로 이용하여 생략 성분 복원 규칙을 생성한다.

(1) 접속 삭감 복원 규칙

접속 삭감 복원 규칙은 [표 5]와 같고, C₁ ~ C₆는 접속문에서의 접속삭감 복원규칙들이다. 접속어미 '-고'는 conj①로 표기한다. 본 논문에서는 지면관계 상 [Asub Aobj Apred]를 [Ante]로, [Csub_n |Cobj_n| Cpred_n]을 [Cons_n]으로 약기한다. 규칙에서 [Cons₀]는 [Ante]를 나타내고 n의 범위는 0 < n < 3 이다.

규칙 C₁, C₂, C₃는 순행 생략이 된 경우에 적용 가능한 규칙으로 주어, 목적어, 주어와 목적어가 생략된 경우에 적용되고 C₄, C₅, C₆는 역행 생략이 된 경우에 적용 가능한 규칙으로 주어, 서술어, 목적어와 서술어가 함께 생략된 경우에 각각 적용된다. C₅와 C₆의 Cpred_n 복원 시에는 conj①과 함께 복원된다.

표 5 접속 삭감 복원 규칙

[Ante] conj [Cons _n]	
C ₁	Csub _n 생략 시 Csub _n => Csub _{n-1}
C ₂	Cobj _n 생략 시 Cobj _n => Cobj _{n-1}
C ₃	Csub _n & Cobj _n 생략 시 Csub _n & Cobj _n => Csub _{n-1} & Cobj _{n-1}
C ₄	Csub _n 생략 시 Csub _n => Csub _{n-1}
C ₅	Cpred _n 생략 시 Cpred _n => Cpred _{n+1} , conj => conj①
C ₆	Cobj _n & Cpred _n 생략 시 Cobj _n => Cobj _{n+1} , Cpred _n => Cpred _{n+1} , conj => conj①

3.2.2 내포문(Embedded Sentence)

'관계 명사구 탈락 규칙'과 '동일 명사구 탈락 규칙'을 역으로 이용하여 생략 성분 복원 규칙을 생성한다.

(1) 관계 명사구 탈락 복원 규칙

관계 명사구 탈락 복원 규칙은 [표 6]과 같고 E₁ ~ E₅는 내포문에서의 관계명사구 탈락 복원규칙들이다. [Aesub_n |AEobj_n| |AEadv_n| |Aepred_n]은 [AE_n]으로 약기한다. [AE_n]은 접속어미에 의해 1회 이상 반복될 수 있고 제안한 규칙이 적용되지 않을 때는 접속문 복원규칙을 적용시킨다.

표 6 내포문에서의 관계 명사구 탈락 복원 규칙

	[Hsub] [[AE _n] _{adn}]	MODEe Hpred
E ₁	Aesub 생략 시	Aesub => MODEe
E ₂	AEobj 생략 시	AEobj => MODEe
E ₃	AEadv 생략 시	AEadv => MODEe
E ₄	Aesub & AEobj 생략 시	Aesub => Hsub, AEobj => MODEe
E ₅	Aesub _n & Aesub _{n+1} 생략 시	Aesub _n & Aesub _{n+1} => MODEe

규칙 E₁는 내포절 주어, E₂는 내포절 목적어, E₃는 내포절 부사어, E₄는 내포절 주어와 목적어가 함께 생략되었을 때 적용되는 규칙이고, 규칙 E₅는 관계절 내의 문장이 접속어미로 연결되어 MODEe와 공지시 관계를 이룰 때 적용 가능한 규칙이다.

(2) 동일 명사구 탈락 복원 규칙

동일 명사구 탈락 복원 규칙은 [표 7]과 같고 E₆는 내포문에서의 동일 명사구 탈락 복원규칙이다. [NEsub_n |NEobj_n| |NEpred_n]은 [NE_n]으로 약기한다. [NE_n]은 접속어미에 의해 1회 이상 반복될 수 있고, 제안한 규칙이 적용되지 않을 때는 접속문 복원규칙을 적용시킨다.

표 7 내포문에서의 동일 명사구 탈락 복원 규칙

Hsub [(NE _n _nom] Hpred		
E ₆	NEsub _n 생략 시	NEsub _n => Hsub

규칙 E₆은 명사화 내포문의 경우 단일 내포절이거나 내포절이 접속어미로 연결될 경우, 모든 주어가 생략되었을 때 적용되는 규칙이다.

3.2.3 인용문(Quotation)

본 논문에서는 인용 어미 '-고, -(이)라고, -하고'와 함께 쓰이는 문장을 인용문으로 취급하고, 직접·간접 인용문을 다 포함한다. 인용문에서의 생략 성분 복원 규칙 생성을 위하여 '인용문에서의 주어 인칭 제약'과 '동일 명사구 탈락 규칙'을 이용한다.

(1) 주어 인칭제약에 따른 동일 명사구 탈락 규칙 (Equivalent NP Deletion)

인용절의 종결어미는 명령형, 의문형, 청유형, 서술형으로 나뉜다. 인용문의 일반화는 서술형인 경우에만 인정되므로, 의문형, 청유형, 명령형의 경우 인용절 주어는 인칭 제약을 받는다[1,19,23,24]. 그러므로 명령형과 의문형은 인용절 주어와 상위문 간접 목적어와, 청유형은 상위문 주어+간접 목적어와, 서술형은 상위문 주어와 일치하여야 한다. 이렇게 일치하는 명사구가 상위문장과 인용절에 각각 나타날 때, 인용절에 있는 명사구가 생략된다.

(2) 주어 인칭제약에 따른 동일 명사구 탈락 복원 규칙

주어 인칭제약에 따른 복원 규칙은 [표 8]과 같고, QT₁~QT₄는 인용문에서의 주어인칭 제약에 따른 생략 성분 복원 규칙들이다. [Qsub_n |Qobj_n |Qpred_n]은 [Q_n]으로 약기하며 [Q_n]은 접속어미에 의해 반복될 수 있다.

표 8 인용문에서의 주어인칭 제약에 따른 생략 성분 복원 규칙

서술형(약속형 포함)		
Hsub [Hobj ₁] [[Q _n] _D]_quot Hpred		
QT ₁	Qsub _n 생략 시	Qsub _n => Hsub
[QS] conj [QS]		
QT ₂	Hpred _n , Hsub _{n+1} 생략	Hpred _n => quot & Hpred _{n+1} , Hsub _{n+1} => Hsub _n
청유형		
Hsub [Hobj ₁] [[Q _n] _S]_quot Hpred		
QT ₃	Qsub _n 생략 시	Qsub _n => Hsub & Hobj ₁
명령형 의문형		
Hsub [Hobj ₁] [[Q _n] _{Im In}]_quot Hpred		
QT ₄	Qsub _n 생략 시	Qsub _n => Hobj ₁

규칙 QT₁~QT₂는 인용절 종결어미가 서술형일 경우 QT₁은 인용절이, QT₂는 인용문(QS)이 접속어미로 연결될 때, 적용되는 규칙으로 conj 앞에 quot와 Hpred_{n+1}도 함께 복원된다. 규칙 QT₃는 청유형일 때, 규칙 QT₄는 명령형이거나 의문형일 때 적용되는 규칙들이다. 인용절 내에 접속문이나 내포문이 내포될 때는 내포된 각각의 규칙을 적용시킨다.

다음은 복합문 영 대용어 처리를 위한 인용문 복원 알고리즘이다.

Quot_Recovery_Alq()

```

1. 종결어미 유형을 체크한다 : 서술형, 청유형, 명령형, 의문형
2. 종결어미 유형별 복원 규칙을 적용한다.
If Endings = "Suggestion" then apply QT3
  else if Endings = "Imperative" or
  Endings="Interrogative" then apply QT4
  else if Endings = "Declarative" and Flag = 1 then
  apply QT2
  else apply QT1; /* 인용문의 접속어미
  연결여부에 따라* /
    
```

3.3 영 대용어 처리 예제

다음은 본 논문에서 제안한 영 대용어 복원 규칙을 신문기사문에 적용시켜 영 대용어를 처리하는 과정이다.

유니텔과 하나은행은 새로운 수익원 창출을 위한 전략적 업무제휴를 맺고 금융솔루션을 공유하고 공동마케팅에 나서기로 했다.

▶ 분해 알고리즘 수행

- 1) Endings ≠ 'quot' then goto 2)
- 2) Endings = 'nom' then NE_Split_Alq 수행
 ⇒Hsub '유니텔과 하나은행은'과 nom '-기' + 격조사 '-로' 다음에 분해
 유니텔과 하나은행은 [[새로운 수익원 창출을 위한 전략적 업무제휴를 맺고 금융솔루션을 공유하고 공동마케팅에 나서기로] 했다.
- 3) Endings = 'conj' then C_Split_Alq 수행
 ⇒'및고'의 conj '-고'와 '공유하고'의 conj '-고'의 앞뒤로 분해
 유니텔과 하나은행은 [[새로운 수익원 창출을 위한 전략적 업무제휴를 맺고][금융솔루션을 공유하고][공동마케팅에 나서기로]] 했다.
- 4) Endings = 'adn' then AE_Split_Alq 수행
 ⇒'새로운'의 adn '-은' 다음에 분해.
 유니텔과 하나은행은 [[새로운] 수익원 창출을 위한 전략적 업무제휴를 맺고][금융솔루션을 공유하고][공동마케팅에 나서기로]] 했다.
- 5) Endings = 'adn' then AE_Split_Alq 수행

⇒한 splitting 내에서 반복되므로 이전 adn ('새로운'의 adn '-은')과 현재 adn('-위한'의 adn '-ㄴ') 다음에 분해

유니텔과 하나은행은 [[[[새로운] [수익원 창출을 위한] 전략적 업무제휴를 맺고] [금융솔루션을 공유하고] [공동마케팅에 나서기로]] 했다.

6) 더 이상의 'adn'이 없으므로 분해 중지

▶ **연 대용어 탐색** ⇒ 가장 상위 문장부터 수행

1) 명사화 내포문

유니텔과 하나은행은 [새로운 수익원 창출을 위한 전략적 업무제휴를 맺고 금융솔루션을 공유하고 공동마케팅에 나서기로] 했다.

2) 명사화 내포문(접속문)

유니텔과 하나은행은 [[NEØsub₁ 새로운 수익원 창출을 위한 전략적 업무제휴를 맺고] [NEØsub₂ 금융솔루션을 공유하고] [NEØsub₃ 공동마케팅에 나서기로]] 했다.

3) 명사화 내포문(접속문(관형화 내포문))

유니텔과 하나은행은 [[NEØsub₁ [AEØsub₁ 새로운] 수익원 창출을 위한 전략적 업무제휴를 맺고] [NEØsub₂ 금융솔루션을 공유하고] [NEØsub₃ 공동마케팅에 나서기로]] 했다.

4) 명사화 내포문(접속문(관형화 내포문+관형화 내포문))

유니텔과 하나은행은 [[NEØsub₁ [AEØsub₁ 새로운] [AEØsub₂ 수익원 창출을 위한] 전략적 업무제휴를 맺고] [NEØsub₂ 금융솔루션을 공유하고] [NEØsub₃ 공동마케팅에 나서기로]] 했다.

▶ **규칙 적용**

1) 명사화 내포문(내포절이 접속어미로 연결)

Hsub [[NE _n]_nom] Hpred		
E ₆	NEsub _n 생략 시	NEsub _n => Hsub

2) 관형화 내포문

[Hsub] [[AE _n]_adn] MOD _{ee} Hpred		
E ₁	AEsub 생략 시	AEsub => MOD _{ee}

▶ **복원 결과**

1) 명사화 내포문(규칙 E₆ 적용)

유니텔과 하나은행은 [유니텔과 하나은행은 [AEØsub₁ 새로운] [AEØsub₂ 수익원 창출을 위한] 전략적 업무제휴를 맺고] [유니텔과 하나은행은 금융솔루션을 공유하고] [유니텔과 하나은행은 공동마케팅에 나서기로] 했다.

2) 관형화 내포문(규칙 E₁ 적용)

유니텔과 하나은행은 [유니텔과 하나은행은 [수익원

창출이 새로운] 수익원 창출을 위한] 전략적 업무제휴를 맺고] [유니텔과 하나은행은 금융솔루션을 공유하고] [유니텔과 하나은행은 공동마케팅에 나서기로] 했다.

3) 관형화 내포문(규칙 E₁ 적용)

유니텔과 하나은행은 [유니텔과 하나은행은 [[수익원 창출이 새로운] 전략적 업무제휴가 수익원 창출을 위한] 전략적 업무제휴를 맺고] [유니텔과 하나은행은 금융솔루션을 공유하고] [유니텔과 하나은행은 공동마케팅에 나서기로] 했다.

위의 예)는 명사화 내포문의 내포절이 두 개의 접속어미 '-고'에 의해 연결되었고 내포절의 모든 주어가 생략되었다. 이런 경우, 규칙 E₆에 의하여 내포절의 생략된 주어는 모두 선행절 주어로 복원된다. 그러므로 NEØsub₁ ~ NEØsub₃는 상위문 주어인 '유니텔과 하나은행은'으로, AEØsub₁ ~ AEØsub₂는 피수식어인 '수익원 창출이'와 '전략적 업무제휴가'로 각각 복원된다.

4. 실험 결과 및 분석

4.1 실험 대상 문장

본 논문에서 처리한 실험 대상 문장은 신문 기사문 130편(총 어절 수 26000여 어절)으로서 대상 문서는 매일 경제, 한겨레 신문에서 증권, 금융, 부동산, 정치, 사회, 문화 분야별로 발췌하였다. 실험문장은 전체 1328문장 중 단문이 302문장이고 복합문이 1026문장이었다.

다음의 [표 9]는 1026문장에서의 각 문장 유형별 출현빈도 및 복합문의 인식 정확도와 분해 정확도를 나타낸다. 1026문장에서 내포된 문장유형은 모두 1718번 출현하였다. 즉, 한 문장 당 1.68개의 내포된 문장을 가지고 있는 것으로 나타났다. 실험 결과, 복합문의 인식 정확도는 91.56%, 분해 정확도는 96.76%로 매우 높은 것으로 나타났다. 특히 인용문과 명사화 내포문은 인식 정확도와 분해 정확도가 모두 100%로 신문기사 도메인에서는 이 논문에서 제시한 분해 알고리즘이 타당함을 보여준다.

복합문의 인식 정확도와 분해 정확도는 다음 식에 의해 구하였다.

$$\text{인식정확도} = \frac{\text{시스템이 인식한 내포된 문장유형의 수}}{\text{말뭉치에 출현한 각 문장유형의 수}}$$

$$\text{분해정확도} = \frac{\text{시스템이 정확하게 분해한 문장유형의 수}}{\text{시스템이 인식한 각 문장유형의 수}}$$

따라서 [표 9]에서 각 내포된 문장유형별 인식 정확도는 (출현회수-인식오류개수)/(출현회수)가 되고, 분해 정확도는 (출현회수-인식오류개수-분해오류개수)/(출현회수-인식오류개수)가 된다.

표 9 복합문 인식 정확도 및 분해 정확도

	출현 회수	출현빈도 (%)	인식 오류	인식정확도 (%)	분해 오류	분해정확도 (%)
직접 인용문	312	18.26	0	100	0	100
간접 인용문	289	16.82	0	100	0	100
명사화 내포문	109	6.34	0	100	0	100
접속문	537	31.25	26	95.17	2	99.68
관형화 내포문	470	27.36	79	83.16	34	91.22
계	1718	100	105	91.56	36	96.76

복합문 분해 알고리즘을 통해 분해된 문장유형 중 영
대용어가 야기된 문장 수는 1172개였다. [표 10]은 문
장유형에 따른 영 대용어 출현빈도로서 접속문, 인용문,
내포문(명사화와 관형화)이 각각 전체 영 대용어 출현
문장의 31.06%, 44.71%, 24.23%를 보이고 있다.

[표 11]은 문장 간-문장 내 영 대용어 분포를 보여주
는데 문장 간 영 대용어 5.46%의 적은 분포를 차지하
지만 이로 인해 해결 불가능한 문장과 부분적으로 해결
가능한 문장의 비율을 높이는 결과를 낳게 된다. 신문
기사문에서는 문장 간 영 대용어의 경우 이전 문장의 주
어가 영 대용어로 복원되는 것이 거의 대부분이다. 그러
므로 본 논문에서도 문장 간 영 대용어를 처리하게 되
면 해결 가능한 문장이 6.14% 더 증가하게 된다.

표 10 문장 유형별 영 대용 분포

문장유형 영 대용	접속문	인용문	명사화 내포문	관형화 내포문	전체
문장 수	364	524	100	184	1172
분포(%)	31.06	44.71	8.53	15.70	100

표 11 문장 간내 영 대용 분포

문장 간내 영 대용	문장 간 영 대용	문장 내 영 대용	전체 영 대용
문장 수	64	1108	1172
분포(%)	5.46	94.54	100

[표 12]는 제안한 규칙을 이용하여 전체 문장에서의
영 대용어 처리 결과를 보여주는데, 해결된 문장 수는
979문장이고 부분 해결 문장은 135문장, 해결 불가능한
문장은 58문장으로 각각 83.53%, 11.52%, 4.95%를 차
지한다.

[표 13]은 문장 유형별 영 대용어 처리 결과인데 해
결 가능한 문장 수는 관형화 내포문, 인용문, 접속문, 명
사화 내포문의 순으로 감소한다.

표 12 전체 문장에서의 영 대용어 처리

영 대용어	문장 수	빈도(%)
해결	979	83.53
부분 해결	135	11.52
해결 불가능	58	4.95

표 13 문장 유형별 영 대용어 처리

문장유형 영 대용어	접속문 (%)	관형화 내포문(%)	명사화 내포문(%)	인용문 (%)
해결	82.42	91.30	80	86.26
부분 해결	7.69	8.70	16	11.45
해결 불가능	9.89	0	4	2.29

4.2 분석 및 고찰

영 대용어 해결에 문제가 있는 경우는 주로 복합문
분해 문제, 복원 규칙과 관형화 내포문에서의 MODec
범위 문제, 문장 간 영 대용어 미 처리, 기타 (격조사
생략, 어순도치...) 등으로 인하여 발생한다. 다음의 예문
① ~ ③은 부분적으로 해결이 가능한 문장들이고 ④
~ ⑥은 완전하게 처리가 어려운 문장들이다.

① [이는 [이 달말 개설될] [이북 5도 홈페이지를 북
한정부의 인터넷 사이트인] 인포뱅크와 연계해] [남북의
이산가족들이 인터넷에 사연을 올려] [잃어버린 가족을
찾도록 한다는 것.]

② [3일 예정된] [창립기념행사에 참석하기 위해 내
한한] 이트레이드 인터내셔널의 CEO 주디 밸린트는 2
일 코스닥이나 거래소의 상장계획은 미정이지만 ~

③ [애초 유럽연합은 [[우리 정부에 구조조정이 상업
적 기준에 따라 이뤄지는 것을 보장해 줄 것을] 요구했
으나,] [이는 받아들이지 않았다]]고 최 국장은 전했다.

④ [황 애널리스트는 ["삼우통신은 [한국통신에 대한]
매출이 전체의 40~50%를 육박한다"]며] [""[적정마진
이 보장되는] 매출처이기 때문에 안정적 수익기반을 갖
출 것"]이라고 평가했다.]

⑤ [위법행위를 정당별로 보면] [민주당이 837건으로
가장 많았고] [이어 한나라당 512건, 청년진보당 3건 등
이었으며] [무소속 후보자의 경우도 294건에 달했다.]

⑥ [자민련은 이날 이한동 총재 주재로 당선자 간담
회를 열고] [""[16대 국회의원의 정수가 15대보다 줄어든
만큼 교섭단체 구성 요건도 완화해야 한다]"]며] [""[이같은
의견을 모았다고] 이규양 부대변인이 밝혔다.]

실험문장에서 제시한 예문 ①~⑥까지는 다음과 같은
추가적인 정보와 연구가 필요하다. 예문 ①은 분해문제
로서 AE_Split_Alg에 따라 분해 할 경우 '이북 5도 홈

페이지를' 앞에서 분해하게 되어 생략성분 복원 시에 이상한 문장을 형성하게 된다. 그러므로 분해 조건을 추가할 필요가 있다.

예문 ②는 복원문제로서 관형화 내포문에서의 MODee 범위 문제이다. adn뒤에 'NP1의 NP2'가 올 경우, 관형화 내포문의 생략성분은 75%정도가 NP1이 복원된다. 그러나 예문 ②에서의 옳은 복원(MODee)값은 NP1의 NP2인 '이트레이드 인터내셔널의 CEO 주디 벨린트'이다. 명사의 의미자질 표현을 좀더 세분화하고, 용언의 하위범주화 정보를 세분화된 명사 의미자질 표현에 맞게 구체적으로 제시한다면 적절한 MODee범위('NP1' 또는 'NP1의 NP2')를 찾을 수 있을 것이다.

예문 ③은 '어순 도치'로 인한 복원 문제로 인용문 상위문 주어('최국장은')가 문미에 위치함으로써 문두에 위치한 '유렵연합은'이 상위문 주어로 인식되어 인용절 내에 생략된 주어가 '유렵연합은'으로 잘못 복원된다. 이 경우는 전처리 과정에서 어순 정렬을 하면 해결이 가능하다.

예문 ④는 복원규칙 문제로 인용문 규칙 QT1에 의해 인용절 내의 생략된 주어가 상위문 주어인 '황 애널리스트는'으로 잘못된 값을 복원시키게 된다(옳은 값: '삼우통신은'). 이 경우엔 복원규칙을 좀 더 추가할 필요가 있다.

예문 ⑤는 '무주어'와 '격조사 생략'으로 인한 복원 문제로 접속문 규칙 적용으로 후행절 주어인 '민주당이'가 선행절의 주어로 복원되는 오류가 일어나게 된다. 그리고 '한나라당도' 뒤의 격조사 생략으로 인해 용언 '이 있었으며'의 주어를 찾을 수가 없다. 이 경우는 전처리 과정에서 생략된 격조사를 복원시킨 후 처리를 하면 옳은 복원값을 찾을 수 있을 것으로 본다.

예문 ⑥은 복원 & 분해 문제로 접속문과 인용문이 이어지고 인용문의 상위문 주어가 도치되어 상위문 주어를 찾을 수 없고 분해하기가 매우 복잡하다. 전처리 과정에서 미리 어순을 정렬하고 분해 순서에 예외사항을 추가한다면 해결되리라 생각된다.

5. 결과 및 향후 과제

본 논문에서는 제안한 복합문 분해 알고리즘과 복합문 영 대용어 복원 규칙을 기반으로 하여 한국어 복합문에서 빈번히 발생하는 영 대용어 현상에 대한 해결 방법을 제시하였다. 영 대용어는 복합문에서 흔히 일어나는 보편적 현상이므로 기계번역이나 자동 문서요약 시스템 구축에 있어 문장 내의 영 대용어 처리는 필수적이다. 영 대용어 해결을 위해서 복합문 판별과정, 복합

문 분해과정, 영 대용어 탐색과정 그리고 생략성분 복원 과정을 거치며, 이 때 복합문 구성에 관여하는 어미들의 어미 분류표와, 생략될 때 적용된 통사규칙을 역으로 이용한다. 본 논문은 신문 기사문을 대상으로 복합문 중에서 보조 용언 내포문을 제외한 인용문, 접속문, 내포문을 처리하였다.

제안한 방법을 이용한 결과 전체 영 대용어 중 83.53%가 해결 가능하며 11.52%가 부분적으로 해결된다. 문장 간 영 대용어 처리가 이루어지면 88.99%까지 복합문 영 대용어 처리가 가능하다.

앞으로 보조 용언 내포문에서의 생략 성분 복원에 대한 연구가 더 필요하며, 현재 관형화 내포문에서의 적절한 MODee 범위 결정, 규칙의 추가, 문장 간 영 대용어 처리 등에 관한 연구는 계속 진행되고 있다.

참고 문헌

- [1] 한재현, "생략과 대용현상 -한국어와 영어를 중심으로-", 한신문화사, 1984.
- [2] 김일용, "우리말 대용어의 연구", 부산대학교 박사학위논문, 1982.
- [3] 양명희, "현대국어 대용어에 대한 연구", 서울대 박사논문, 1996.
- [4] 남기심, "언어학 개론", 탑출판사, 1985.
- [5] 김옥, "국어 접속문의 생략에 관한 고찰", 전남대학교 교육학과 국어교육전공 석사학위논문, 1998.
- [6] Ik-Hwan Lee, Minhaeng Lee (2000a), "Anaphora Resolution and Discourse Structure: A Controlled Information Packaging Approach," In: Language and Information 4(1), 67-82, 2000, 6[in English]
- [7] Ik-Hwan Lee, Minhaeng Lee (2000b), "On the Discourse Analysis in Korean Dialogues," In : Proceedings in PACLIC14. pp. 207-218, 2000 [in English]
- [8] Ik-Hwan Lee, Minhaeng Lee, "On the Anaphora Resolution in Korean Dialogues," In : Harvard Studies in Korean Linguistics Vol. 7, 1999.
- [9] 한승연, 송만석, "개념 그래프를 이용한 대용어 해결 시스템", 한국정보처리학회 추계 학술발표논문집, 제2권 제2호, pp. 844-851, 1995.
- [10] 한승연, "지식 기반을 이용한 대용어 해결 시스템", 연세대학교 전산학과 석사학위논문, 1995.
- [11] 차건희, 송도규, 박재득, "한국어 대용과 생략 해결을 위한 센터링 이론의 적용", 한글 및 한국어정보처리 발표 논문집, pp347-352, 1997.
- [12] 김정해, 이상국, 이상조, "한국어 문장내 체언류 조용 대용어의 해결방안", 대한 전자공학회 논문지, 제33권 B편 제 4호, pp. 183-190, 1996.
- [13] 김정해, "HPSG 파서에 기반한 한국어 문맥 조용대용어의 해결", 경북대학교 박사학위 논문, 1996.

- [14] S. Lappin and M. McCord, "Anaphora Resolution in Slot Grammar," Computational Linguistics, Vol. 16, No.4, pp. 197-212, 1990.
- [15] Walker, Marilyn, "Centering, Anaphora Resolution, and Discourse Structure," Centering in Discourse, eds. Marilyn A. Walker, Aravind K. Joshi and Ellen F. Prince, Oxford University Press, 1997.
- [16] Walker, Marilyn, Masayo Iida & Sharon Cote, "Centering in Japanese Discourse," Proceedings of the 13th International Conference on Computational Linguistics(Coling 90), 1990.
- [17] Walker, M.; M. Iida and Cote, S. "Japanese Discourse and the Process of Centering," Computational Linguistics, 20, pp. 193-232, 1994.
- [18] Nakaiwa, H., "Automatic Extraction of Rules for Anaphora Resolution of Japanese Zero Pronouns from Aligned Sentence Pairs," Proc. of ACL-97/EACL-97 Workshop on Operational Factors in Practical, Robust, Anaphora Resolution for Unrestricted Texts, Madrid, Spain, pp. 22 - 29, 1997.
- [19] 권재일, "국어의 복합문 구성연구", 집문당, 1985.
- [20] 김홍규, "국어 기초 자료 구축", 21세기 세종계획 연구 보고서, 문화관광부, 2000.
- [21] Saliha Azzam, "Resolving Anaphor in Embedded Sentences," 1997.
- [22] Young-Giu Jung, "Implementation of a parser for news summarization," IC-AI, 2001.
- [23] 남기심, "국어 완형보문법 연구", 한국학연구총서(叢書) (1), 계명대학교 출판부, 1983.
- [24] 이필영, "국어의 인용구문 연구", 탐 출판사, 1993.



구 상 옥

2001년 경북대학교 컴퓨터공학과 졸업(학사). 2001년 ~ 현재 경북대학교 대학원 컴퓨터공학과 석사과정 재학. 관심분야는 자연어처리, 정보검색, 정보이론

강 보 영

정보과학회논문지 : 소프트웨어 및 응용 제 29 권 제 8 호 참조

이 상 조

정보과학회논문지 : 소프트웨어 및 응용 제 29 권 제 8 호 참조



김 미 진

1979년 한양대학교 영어영문학과 졸업(문학사). 1991년 경북대학교 대학원 컴퓨터공학과(공학석사). 1997년 ~ 현재 경북대학교 대학원 컴퓨터공학과(박사과정 수료). 관심분야는 정보검색, 문서요약, 음성인식 후처리, 자연어처리



박 미 성

1990년 창원대학교 전자계산학과 졸업(이학사). 1992년 경북대학교 대학원 컴퓨터공학과(공학석사). 2000년 경북대학교 대학원 컴퓨터공학과(공학박사). 1992년 ~ 현재 경북대학교 도서관 전산실 근무. 관심분야는 자연어처리, 음성인식, 전자도서관, 데이터베이스