

한-중 기계번역 시스템에서의 중간표현에 관한 연구

송 성대, 이 종혁, 이 근배
포항공과대학교 전자계산학과

A Study on Intermediate Representation for Korean-Chinese Machine Translation System

Seong-Dae Song, Jong-Hyeok Lee, Geunbae Lee
Department of Computer Science & Engineering
Pohang University of Science & Technology
Tel : (0562)279-5656 Fax : (0562)279-2299

한국어와 중국어는 상이한 언어적 특성을 보이는 언어쌍으로, 기계번역 시스템을 개발하기 위한 접근 방식으로 중간언어 방식이 타당하다. 본 논문에서는 한-중 기계번역 시스템의 개발과 관련하여, 의미표현 구조로서의 중간언어에 대하여 논한다.

기존의 중간언어들을 비교, 분석하여 문제점을 파악하고 개선하며, 한국어의 특성을 반영함으로써 한-중 기계번역 시스템을 위한 중간언어를 제안한다. 본 논문을 통하여 국내적으로 미진한 중간언어 연구를 활성화하며, 더욱 효율적인 중간언어를 설계하기 위한 원형을 마련한다. 앞으로 의미 체계의 보완, 개선 등을 통하여, 본 논문에서 제안된 원형은 다언어 기계번역 시스템을 위한 중간언어로 확장, 개선될 것이다.

1. 서론

한국어와 중국어는 상이한 언어적 특징을 나타내는 언어쌍이기 때문에, 한-중 기계번역 시스템의 개발을 위해서는 중간언어(Interlingua) 방식의 접근방식을 채택하는 것이 적절하다. 대역번역(Transfer Translation) 방식과 달리, 중간언어 방식에서는 번역의 목표가 되는 언어들과 독립적인 의미표현 구조를 매개로 번역이 이루어 지기 때문에 언어적 특성이 다른 언어간의 번역 시스템을 개발하는데 유용하기 때문이다.

중간언어 방식은 시스템의 개발이 번역대상이 되는 언어에 지역화(localized)된다는 장점이 있다. 즉, 중간언어와 자신의 언어 사이의 번역 시스템만을 개발하여 기존의 시스템에 첨가하면 원하는 기계번역 시스템을 구축할 수 있는 것이다. 그리고 지역화되어 개발되는 과정에서 획득되는 지식을 공유할 수 있다는 장점이 있다. 즉, 지식의 축적이 중간언어를 통하여 이루어지고, 개발되는 과정에서 서로 공유하게 되는 것이다.

그러나 무엇보다도 중간언어 방식의 장점은 다언어간 기계번역 시스템으로의 확장이 용이하다는 점이다. 다언어를 처리할 수 있는 시스템의 개발을 위해서는 효율적인 중간언어의 설계가 필수적으로 요구된다. 현재 선진국에서는 다언어간 번역 시스템으로의 확장을 위하여 중간언어에 대한 연구가 활발하게 이루어지고 있다. 그러나 분야 자체가 진보적이고 폭 넓은 언어학적인 지식을 요구하기 때문에 뚜렷한 성과가 나타나지 못하고 있다. 본 논문에서는 기존의 성과를 비교, 분석하고, 문제점을 파악, 개선하며, 한국어의 특성을 반영하여 한-중 기계번역 시스템을 개발하기 위한 중간언어를 제안한다. 다언어간 번역 시스템으로의 확장을 염두에 둔 원형을 제시함으로써, 차후의 수정, 보완을 통하여 더욱 세련되고 효율적인 중간언어를 설계할 수 있도록 한다. 아울러 국내적으로 미진한 중간언어에 대한 연구를 활성화하는 계기를 마련한다.

2. 기존의 연구

기존의 기계번역 시스템 개발 현황에서, 중간언어 방식을 채택한 연구는 국내보다 일본이나 유럽, 미국 등 선진국에서 활발하게 이루어져 왔다. 프랑스 그레노블(Grenoble) 대학의 CETA 프로젝트, 일본의 ATLAS(Fujitsu), PIVOT(NEC),

Interlingua(CICC), 그리고 미국 카네기 멜론 대학의 KBMT 프로젝트 등의 연구가 있었다.[1][5] 이 중에서 PIVOT(NEC)와 Interlingua(CICC)의 경우에는 자료를 접할 수 있었다.

기존 연구에서 제안하는 중간언어들을 개념이나 의미를 대표하는 개념소(concept)와 이러한 개념소간의 관계를 대표하는 관계기호(relation symbols)를 통하여 문장의 의미를 표현한다는 공통점을 가지고 있으나 개념소의 체계나 관계기호의 운영, 표현된 구조 등에서 차이를 보인다.

국내에서는 영-한 기계번역과 관련하여, 1989년 과학재단의 과제인 "자연언어처리의 기초연구"의 제 2 세 부 과체에서 한국어의 의미에 대한 분류 작업이 있었다.[4] 한국어의 특성을 반영하고 있다는 의의가 있으나, 너무 세분한 경향이 있어, 다소 중복되는 부분이 많았다. 그리고 시스템공학연구소의 인공지능 연구부에서 기계번역 시스템 개발(1993)과 관련하여 중간언어에 대하여 발표한 바 있다. [6]

2.1 PIVOT(NEC)

PIVOT의 특징은 크게 두 가지를 들 수 있다. 관계기호가 단순한 기호가 아니라 하나의 노드 형태로써 필요에 따라 하위 구조를 가질 수 있다. 예를 들면 장소라는 관계를 나타내는 LOC의 경우 필요에 따라, in이나 at 혹은 on 등의 하위 구조를 가질 수 있다. 또 관계기호가 단순히 개념소간의 관계를 나타낼 뿐, 새로운 개념을 형성하지 않는다. 즉, 합성개념(compound concept)을 취급하지 않는다.[1]

2.2 Interlingua(CICC)

Interlingua의 특징도 PIVOT과 대비하여 두 가지를 들 수 있다. 관계기호가 단순한 기호로써 내부적으로 특정한 구조를 가지지 않는다. 예를 들면 장소라는 관계를 나타내는 LOC은 단순한 기호가 되며, in이나 at 혹은 on 등의 관계가 세부적으로 요구되면 특수 개념소로 설정하고 기존 개념소에 MOD 관계를 설정하여 합성개념(compound concept)로서 처리한다. 또 바로 앞에 언급한 것처럼 관계기호는 단순한 관계 설정의 기능을 할 뿐 아니라 합성개념을 형성하게 한다.[1]

2.3 기존 중간언어들의 문제점

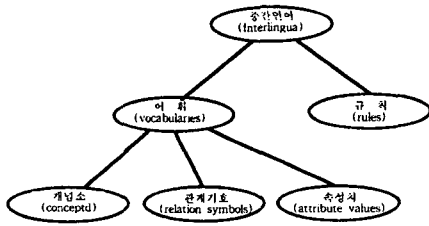
기계번역 시스템의 성능은 번역된 문장이 번역되는 문장의 의미를 반영하는 정도를 가지고 평가할 수 있다. 이런 관점의

본 연구는 한국과학재단 국제공동연구과제 "한-중 기계 번역 시스템"의 지원에 의한 것임.

평가에서 시스템의 성능을 저하되는 것은 주로 각 언어의 개념소들이 그 범주나 혹은 의미가 미치는 범위가 정확하게 일치하지 않기 때문이다. "아주머니"라는 개념소와 "aunt"라는 개념소를 비교하면, "aunt"가 더 폭넓은 의미 범위를 가지고 있다. PIVOT, Interlingua 등을 포함한 기존의 중간언어들은 이런 장애를 극복할 수 있는 효과적인 해결책을 제시하지 못하고 있다. PIVOT의 경우, 새로운 개념소를 첨가하는 방법으로 해결을 시도한다. Interlingua의 경우, 속성값을 부여하는 방법으로 해결을 시도한다. PIVOT의 방법은 중간언어의 어휘를 크게 증가시키는 문제점이 있고, Interlingua의 경우에는 속성치의 체계나 수가 그런 장애를 극복할 수준에 이르지 못하고 있다.

3. 중간언어 설계

일반적으로 언어는 어휘 부분과 어휘를 다루는 문법 부분으로 구성된다. 중간언어도 일종의 메타언어(meta-language)이므로 어휘부분과 문법에 해당하는 규칙부분(rules)으로 나눌



[그림 1] 중간언어의 구성

다.[그림 1]

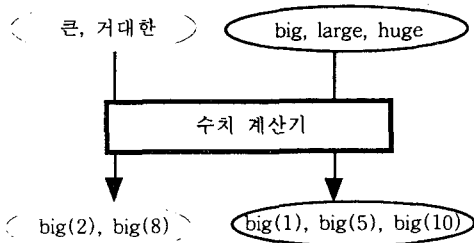
3.1 설계 원칙

- 단어어간 기계번역 시스템으로의 확장을 염두에 두고 설계한다. 특히 기존의 문제점을 해결하기 위하여 개념소에 대한 설계를 강화한다.
- 한국어의 특성을 반영하기 위하여, 한국어 구문 분석기와 연계하여 설계한다.

3.2 개념소(concepts)

언어마다 독특한 개념체계가 존재한다. 중간언어의 개념소 체계를 설계함에 있어서 중요한 것은 각 언어의 독특한 개념체계가 야기하는 의미표현의 갭(gap)을 극복할 수 있는 수단의 제공 여부이다.

- 의미 범위의 차이를 보이는 개념간의 의미 갭을 극복하기 위하여, 대표적인 개념을 중간언어의 개념소로 정하고 의미 범위를 수치화 한다. [그림 2]에서 "큰, 거대한"과 "big, large, huge"를 중간언어의 개념소로 표현하는 방법을 보이고 있다. "big"이라는 대표개념을 중간언어의 개념소로 정하고 이 개념소에 대한 상대적인 범위를 수치로 표현하였다. 수치는 중간언어로 변환하는 시스템에서 부여하도록 한다. 단지 최고값과 최



[그림 2] "큰"과 "large"의 표현

저값 사이에서 문장에서 개념소가 미치는 의미 범위를 수치로 나타내도록 한다. 이처럼 의미 범위를 수치화함으로써 번역의 정확성을 향상시킬 수 있다.

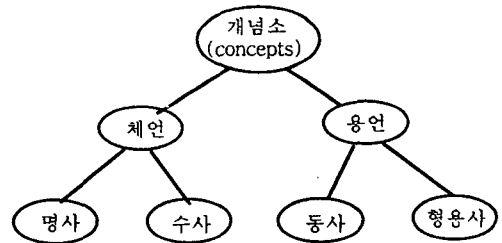
- 대표개념과 수치를 이용하는 방법이 항상 적절한 것은 아니다. "aunt"는 "아주머니", "이모", "고모" 등의 의미로 쓰일 수 있다. 이런 경우에는 각 개념을 한정하는 속성치를 부여하여 구별이 될 수 있도록 해야한다. Interlingua에서 속성으로 개념을 한정하는 방법으로는 이런 문제점을 해결할 수 없다. 이런 문제점을 해결하기 위해서는 속성치에 대한 정밀한 체계가 요구된다.

중간언어가 실용성을 확보하기 위해서는 개념소가 체계를 이루고 있어야 한다. 개념소의 체계를 통하여 개념간의 관계기호를 설정하는 작업이 가능하기 때문이다. 본 논문에서 제안하는 중간언어는 대체로 과학재단의 "자연언어처리의 기초연구"의 의미역 분류를 개념소의 체계로 수용하고 있다.[그림 3]

실질어와 기능어의 관점으로 어휘를 구분할 때, 개념소를 이루는 것은 실질어가 된다. 실질어는 체언과 용언으로 구성된다. 체언은 다시 명사, 대명사, 수사 등으로 나누어 지는데, 주로 명사에 대하여 체계를 나눈다. 대명사의 경우는 그 체계가 명사의 개념을 한정하는 속성치의 체계와 유사한 체계를 가지기 때문에, 구체적인 체계를 나누는 대신 속성을 부여하여 그 의미를 한정하고 표현하는 방법을 채택하였다. 용언의 경우는 동사와 형용사로 나눌 수 있다.

명사, 수사, 동사, 형용사 등의 세부적인 분류 체계는 생략하겠다.

개념소의 분류체계는 다양한 분류 관점 중에서 하나를 택하게 된다. 그러나 다른 관점의 반영이 요구되기도 하며, 때로는 분류선이 애매모호한 경우도 발생한다. 이런 경우의 처리를 위



[그림 3] 개괄적인 개념소의 분류체계

해 의미표지를 둔다. 의미표지는 변동의 소지가 많은 부분을 분류체계에서 독립시켜 표현할 수 있는 수단도 제공하여 추후의 수정,보완에 도움을 준다. 다음과 같은 의미표지를 두나, 응용분야나 확장등의 과정을 거치면서 더욱 세밀하게 설계할 필요가 있다.[4]

- [Abstract] : 추상적인 어휘가 가지는 자질
- [Collective] : 집합명사가 가지는 자질
- [Countable] : 가산성 명사가 가지는 자질
- [Human] : 인간의 행동이 가지는 자질
- [Wh] : 의문사가 가지는 자질

3.3. 관계기호(relation symbols)

관계기호는 기능어의 역할을 하는 중간언어의 기본 어휘라고 할 수 있다. 개념소와 개념소간의 의미적인 관계나, 문장과 문장간의 의미적인 관계를 나타내는 역할을 한다. 기능적 측면을 고려하여 같은 기능의 관계는 통합하고, 관련 개념들이 가지는 의미적 특성을 고려하여 관계의 의미가 설정되도록 한다. 본 논문은 관계기호의 구조로 Interlingua의 방식을 채택하여 세부적인 하위구조를 갖지 않는 단순한 기호의 형태를 취한다.

요구되는 세부구조는 각 개념소의 속성으로 처리된다.

다음은 본 중간언어의 관계를 나열한 것으로, 간단한 예문을 첨가하여 이해를 도왔다. 주의할 점은 관계가 대체로 서술어와 타 성분간의 의미적인 관계를 의미한다는 점이다.[1][3][4]

- AFF = AFFected : 어떤 행동을 받는 대상
 - * 영화가 **감자를** 삶았다.
- AFO = Artificial FORce : 무의지 주체
 - * **이러한 비교가** 그런 결론을 가능하게 한다.
- AGT = AGenT
 - : 행위 동사의 행위 주체(의지적 주체로 AFO와 구별)
 - * **영화가** 감자를 삶았다.
- BEN = BENefactive
 - : 수혜자(인간에 한하며 DAT와 구별 포함)
 - * 그는 **그녀에게** 선물을 주었다.
- CAU = CAUse : 원인
 - * **철수는** **앞으로** 죽었다.
- CHA = CHAracterized
 - : 형용사나 명사에 의하여 그 성격이 규정되는 대상
 - * **그 작업은** 용이하지 않다.
- CMP = Comparison :
 - : 비교의 상대가 되는 대상(비특정 대상인 CRI와 구별)
 - * **영회는** **그녀보다** 아름답다.
- CON = CONdition
 - : 조건(가정적인 의미가 배제되어 SUP와 구별)
 - * 그곳은 **신분증이 있어야** 출입이 가능하다.
- CRI = CRIterion : 기준(CMP와 구별)
 - * 그녀는 **나이에 비해** 조속하다.
- CSE = CAusEe
 - : 피사동주(사역의 의미로 HRR과 구별)
 - * 우리가 **철수에게** 노래를 시켰다.
- CTE = ConTEnt
 - : 거의 필연적인 내용을 가지는 목적어, (동사가 그 목적어를 동사화하는 정도의 역할을 함)
 - * 우리는 **사흘 동안** 그 작업을 했다.
- DAT = DATive : 여격(BEN과 구별)
 - * 그는 망치를 **내게** 건네 주었다.
- DEG = DEGree : 정도
 - * 나는 그보다 **50m나** 크다.
- DIR = DIRection : 방향
 - * 철수가 **남쪽으로** 향했다.
- DIS = DIStance : 거리
 - * 우리들은 **오십 리나** 걸었다.
- DUR = DURation : 지속된 시간
 - * 우리는 **사흘 동안** 그 작업을 했다.
- EFF = EFFected : 결과된 사물(결과와 다름)
 - * 그가 **그 음악**을 작곡했다.
- ELM = ELeMent
 - : 집단이나 큰 구조물 또는 화합물 따위를 이루는 요소
 - * 인간의 몸은 수십 조의 **세포**로 구성된다.
- EXI = EXIstent : 존재하는 대상
 - * **신은** 존재한다.
- EXP = EXPeriencer
 - : 정서의 경험자 또는 무의지적 정신활동의 주체
 - * **우리는** 그 소식에 놀랐다.
 - * **코끼리가** 자고 있다.
- FCS = FoCuS : 초점
 - * 우리는 **그 문제**에 대하여 논쟁했다.
- GOA = GOAl : 목적
 - * 철수는 **좋은 성적**을 위하여 열심히 공부한다.
- HRR = HeaReR : 청자(DAT와 구별)
 - * 철수가 **영희에게** 운동하자고 제안했다.
- INS = INStrument
 - : 도구(구체적인 개체로서, MEA와 구별)
 - * **영희가** 열쇠로 문을 열었다.
- LOC = LOCation
 - : 공간적인 장소(논리적 장소인 SCN과 구별)
 - * **그 화재는 도서관**에서 발생했다.

- MAN = MANner : 행동하는 태도
 - * 철수가 **논리적으로** 설명했다.
- MAT = MATerial : 재료
 - * 그 집은 **벽돌**로 지어졌다.
- MEA = MEAn : 방법(INS와 구별)
 - * **부당한 방법**으로 세금을 징수하였다.
 - * **버스**로 학교에 갔다.
 - * 그들은 **영어로** 말한다.
- MGL = Mental GOal
 - : 인간이 가지 어떤 정신적인 실재 (어떤 대상이 그것에 감응을 일으키게 되는 것)
 - * 그 책은 **철수의 취향**에 맞다.
- MTC = MaTChing entity
 - : 다른 대상과의 관련이 문제가 되는 대상
 - * 이것은 **저것과** 다르다.
- MTG = MaTChing GOal
 - : 어떤 대상이나 상태가 그 대상에 대하여 문제가 되는 바로 그 대상
 - * **그 옷은 너**에게 잘 어울린다.
- NAM = NAME : 명칭
 - * 그는 **거리의 천사**로 통한다.
- OBJ = OBJect : 행위 대상
 - * 그는 **연극**을 보았다.
- PAR = PARTner : 어떤 행동의 동반자
 - * 나는 **너와 함께** 사냥을 하겠다.
- POS = POSsessor
 - : 어떤 물건을 소유하고 있는 대상 (반드시 물건 자체의 소유주를 의미하는 것은 아님)
 - * **형은** 그 집을 소유했다.
 - * **내** 책을 빌려 주마.
- PSD = PoSsessed : 소유된 대상
 - * 철수는 **책**을 가지고 있다.
- QUA = QUAntity : 수량
 - * 이 책은 **3킬로**가 나간다.
- REA = REASon : 이유(서술어에 의해 CAU와 구별)
 - * 그는 **갑기 때문에** 결석했다.
- RUT = RoUTE : 중간 경로
 - * 철수는 **대전**으로 해서 부산에 도착했다.
- SCN = SCeNe : 논리적인 장소(LOC와 구별)
 - * 그는 **그 드라마**에 의사로 출연했다.
- SOR = SOURce : 이동이나 변화의 시작점
 - * 철수는 **집**에서 나왔다.
 - * 우리는 **10시부터** 공부했다.
 - * **철수는** 의사가 되었다.
- SUP = SUPposition : 가정
 - * 이 몸이 **새라면** 날아가리.
- STP = STAndPoint : 관점
 - * **심리학적으로** 영회는 정상이다.
- STA = STAtus : 자격
 - * 철수는 **회장**으로서 그 회의에 참석했다.
- TIM = TIME : 시간/시각
 - * **삼월에** 꽃샘 추위가 닥쳤다.
- TAR = TARget : 이동이나 변화의 목적 지점
 - * 우리는 **부산까지** 갔다.
 - * 그는 **10까지** 노래했다.
 - * 철수는 **의사**가 되었다.

수식관계의 설정을 위하여 'MOD'라는 관계기호를 설정한다. 관형사나 부사 등, 수식언어와 개념소간의 관계를 나타내게 된다. 대체로 수식언어는 수식하는 개념소의 속성치로 처리되나 아래의 예문처럼 개념을 가질 수 있을 경우의 처리를 위하여 설정된 것이다.

- MOD = MODifier
 - * **같은** 양념으로 맛을 냈다.

문장이 하나의 합성개념으로 취급되기 때문에, 복합문처럼 문장간의 관계를 표현하기 위한 관계기호는 개념소간의 관계를 표현하는 관계기호를 그대로 수용한다. 예를 들어 "밥을 먹

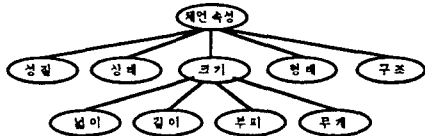
으면서 음악을 들었다"와 같은 복합문의 경우, "밥을 먹으면서"와 "음악을 들었다" 사이에 "PAR" 관계가 설정될 수 있다. 기타 문장간의 관계만을 설정하기 위한 관계기호로는 다음과 같은 것이 있다.

- SUN : 순접관계
 - * 하늘을 보고 힘차게 웃었다.
- YUK : 역접관계
 - * 하늘은 흐렸으나, 비는 오지 않았다.

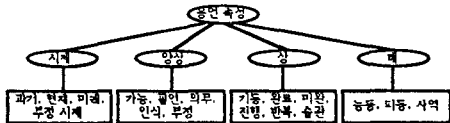
서술어를 찾을 수 없는 문장의 시제나 양상 등의 정보를 위해 'SMOD' 관계를 설정하여, 문장의 마지막 문장성분에 할당한다. 이를 위해서 "SEN"이라는 개념소를 둔다. "SEN"에는 시제, 양상, 태 등과 같은 문장차원의 속성치가 가해진다.

3.4 속성치(attribute values)

속성은 개념소의 의미에 제한을 가하는 중간언어의 어휘이다. 속성은 항상 개념의 의미를 제한하여, 개념이 실제로 쓰일



[그림 4] 체인의 속성 체계



[그림 5] 용언의 속성 체계

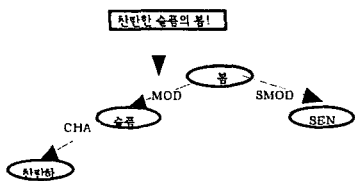
때는 속성이 가해진 가공 개념이 쓰인다. 개념소 부분에서 언급한 것처럼 여러 언어의 개념소가 야기하는 의미범위의 차이를 극복하기 위하여 세밀하게 설계해야 할 어휘이다. 체인과 용언이 문법적인 차이를 보이기 때문에 이에 대응하여 제한을 가하는 속성치도 서로 다른 특성을 보인다. 따라서 그 분류체계도 차이가 있다. [그림 4]와 [그림 5]에 대략적인 체계가 나타나 있다.

3.5 규칙(rules)

개념소들과 관계기호를 이용하여 의미구조를 표현하는 규칙은 다음과 같이 간단하게 나타낼 수 있다.

meaning ::= term | compound | <empty>
 compound ::= (relation meaning meaning)meaning

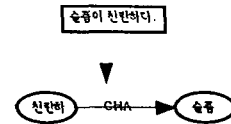
term은 concept에 속성치가 가해진 가공 개념이며 relation은 관계기호이다. meaning은 개개의 개념소가 나타내는 의미를 포함하여, 개념소와 관계기호로 구성되는 복합개념



[그림 6] '찬란한 슬픔의 봄!' 를 위한 예

(compound concepts)도 나타내고, 문장의 의미까지 표현하게 된다.

문장을 하나의 합성개념으로 처리하고 문장 차원의 속성, 즉 시제나 양상, 태 등의 정보가 서술어의 속성으로 나타낸다. 따라서 문장 차원의 속성은 용언의 속성 체계를 따르게 된다. 서술어가 존재하지 않는 문장의 속성은 "SEN"이라는 개념소에 나타낸다. "찬란한 슬픔의 봄!"과 같은 문장을 표현하면 [그림 6]과 같이 나타낼 수 있다.



[그림 7] "슬픔이 찬란하다"의 표현 예

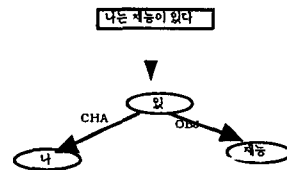
합성개념을 형성할 때, 항상 일정하게 개념소간의 지배관계가 결정되는 특징은 주목할 가치가 있다. 예를 들어 용언 개념소와 체인 개념소 사이에 "OBJ" 관계가 설정된다면, 용언 개념소가 체인 개념소를 지배하게 되는 것이다. 개념소간의 지배관계가 항상 결정적이므로, 관계기호에 방향성을 두어 새로운 의미를 표현할 수 있다. "찬란한 슬픔"이라는 합성개념에 대한 구조는 [그림 6]에 나타나 있다. "슬픔이 찬란하다"의 구조를 표현하면 [그림 7]과 같다. [그림 6]과 [그림 7]을 비교하면 "CHA"라는 관계기호의 방향만 다른 것을 알 수 있다. 이렇게 관계기호의 방향성을 통하여 내포문의 수식을 받는 개념소의 구조를 쉽게 표현할 수 있다.

문장의 속성이 서술어에 표현되므로, 복합문의 문장구조는 각 문장의 중심 서술어간의 관계 설정을 통하여 표현할 수 있다.

4. 한국어 문장의 특성 반영

한국어 문장의 특성을 반영하기 위하여 포함공대 지식 및 언어 공학 연구실의 한국어 구문 분석기와 연계한다. 특히 이 중주어 문장의 구문분석 결과로 두 개의 주어와 주어지는데, 한국어의 특성 반영의 측면에서 이중주어 문장의 처리방법을 제시하겠다.

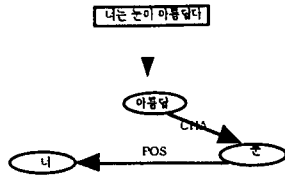
4.1 이중주어 문장의 처리



[그림 8] '나는 재능이 있다'의 표현 예

이중주어 문장은 서술어의 범주에 따라 다른 모습을 보인다. 서술어의 범주가 동사인 경우에, 두 개의 주어 중 하나는 서술어의 주어 역할을 하며, 나머지 하나는 목적어 역할을 하게 된다. 육하원칙에 적용했을 때, "who"에 해당하는 것이 주어의 역할을 하고, "what"의 역할을 하는 것이 목적어의 역할을 하게 된다. "나는 재능이 있다"라는 문장을 표현하고 있는 [그림 8]에서 "나는"은 "있다"의 주어 역할을 하고, "재능이"는 "있다"의 목적어 역할을 하고 있다.

서술어가 형용사인 경우에, 두 개의 주어는 하나의 주어와 다른 주어의 소유주로 나타난다. 이 경우에도 육하원칙의 "who"와 "what"의 관계로 파악을 할 수 있으며, 이때 "who"에 해당하는 개념소가 "what"에 해당하는 개념소의 소유주 역할을 한다. "너는 눈이 아름답다"는 문장을 표현하고 있는 [그림 9]에서 "너는"은 "눈이"의 소유주 역할을 하고,

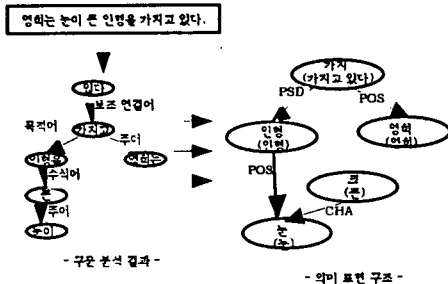


[그림 9] '너는 눈이 아름답다'의 표현 예

눈이"는 "아름답다"의 주어 역할을 하고 있다.

4.2 기타 한국어 문장의 처리

[그림 10]에는 "영희는 눈이 큰 인형을 가지고 있다"라는 문장이 표현되어 있다. 제시된 구문분석 결과는 포항공대 지식 및 언어 공학 연구실의 구문분석기가 분석한 결과이다. 의미 표현 구조의 각 노드에는 개념소와 그 아래에 속성이 가해진



[그림 10] "영희는 눈이 큰 인형을 가지고 있다"의 표현 예

실질적인 의미, 즉 가공 개념이 팔호속에 표현되어 있다.

5. 결론 및 향후 연구과제

본 논문에서는 기존의 중간언어들을 비교, 분석하고 문제점을 파악하여, 문제점을 개선하는 중간언어를 제안하였다. 각 언어 특유의 개념소 분류체계로 인하여 개념소간의 의미범위의 차이가 발생함으로써 생기는 문제점의 해결을 위하여, 대표개념과 의미범위를 가지고 개념소의 의미를 제한하는 방법을 제안하였으며, 한국어 구문 분석기와 연계함으로써 한국어의 특성을 반영하려고 시도하였다. 특히 이중주어 문장의 처리를 통하여 한국어의 특성 반영을 보였다.

효율적인 중간언어의 설계를 위하여 본 논문에서 제안된 원형은 지속적으로 수정, 보완의 과정을 거쳐야 한다. 향후 요구되는 연구과제로는 개념체계의 세밀한 분류 작업과 의미 범위의 갭을 극복하기 위한 속성치 설계작업, 그리고 미흡한 관형사 및 부사의 처리에 대한 보완 등이 있다.

6. 참고 문헌

[1] Uchida, Zhu, Interlingua(Final Edition), CICC, pp.1-47, 1993.
 [2] 고 영근, 남 기심. 고교문법자습서, 탐출판사, 1991.
 [3] Zhang Gui-Ping, Yao Tian-Shun, "The Chinese Analyzer for Multilingua Machine Translation," Computer Processing of Chinese & Oriental Languages, 1990.
 [4] 신 수송, 자연언어처리의 기초연구 : 한국어 문법의 정의와 파싱, 한국과학재단 보고서, pp.139-184, 1989.
 [5] Bonnie lean Dorr. "Machine Translation : A View from the Lexicon," The MIT Press, pp.233-281, 1993
 [6] 김 상국, 박 창호. "중간언어에 기반한 기계 번역 시스템의 설계," 제 5 회 한글 및 한국어정보처리 학술대회 발표 논문집, pp.521-525, 1993 .