

## 초청논문

# 어의의 기술 – 중간언어의 관점에서

장석진

서울대학교

sjchang@snu.ac.kr

### 요약

중간언어(interlingua, IL) 구축을 지향하는 시각에서 단어의 다의성(polysemy) 문제를 각종 워드넷의 기술 방식과 용례를 대비하여 논의하고, 센스(어의) 번호를 부착한 세계 워드넷 (World Wordnet) 데이터베이스를 구상하고, 이 연장 선상에서 (Pustejvsky 식) 생성 어휘부와 TMR(텍스트의미표상)이론을 가미한 제약기반 통합문법(CUG)으로 IL 구축에 다리를 놓고, 이로부터 개별언어 기술로 변환한다.

## 1. 들어가기

형식과 의미의 측면에서 관찰할 때, 단어의 특성중의 하나는 같은 형태이면서 뜻이 다른 동형성(homonymy)과 다의성(polysemy)이다. 어떤 언어이든, 인공적인 언어가 아니라면, 수 천 년의 시간의 흐름 속에서 표현해야 할 사물이나 사태가 끊임 없이 늘어나는 상황에서, 다의성은 새 단어를 만들거나 외래어로 감당 못하게 되어 한정된 형태에 의미가 확산 심화되어 일어나는 자연언어 보편적인 현상이다. 동형성-이를 달리 불러, 동명성(同名性), 동형이의(同形異義), 동음이의(同音異義)-과 다의성을 엄밀히 분간하는 일은 어원과 민간전승 등이 관여하기 때문에 쉬지 않다. 우선, 발표하는 이 글의 제목-어의의 기술-을 예로 들어 생각해보자.

### (1) ‘어의의 기술’

a. 어의: 御衣 御醫 語義 / 語意 (금성사 사전 등재 순)

Ja: 御衣 きょい 語義(ごぎ gogi) / 語意(ごい goi)

b. 기술: 技術 奇術 記述 既述

Ja: きじゅつ kizyutu: 記述 奇術 既述

ぎじゅつ gizyutu: 技術

동형어는 어의<sup>1</sup>, 어의<sup>2</sup>, 어의<sup>3</sup>와 같이 별개의 등재어로 취급하고 다의어는 한 단어 안에 센스(sense) 별로 1, 2, 3 ...식으로 기재하는 것이 재래식 사전의 기술 방식이다. 위의 예는 다 전자에 속한다. 한자 사용을 생각하면 '語義/語意'는音が 같은 異形-同義인가 아니면 異形-異義인가? 일본어는 우리와 같은 한자를 쓰되 音が 달라 하나는 /gogi/, 하나는 /goi/이다. 그러니 두 단어의 뜻이 같다면 異形-異音-同義, 뜻이 다르다면 異形-異音-異義이다. '意'와 '義'의 뜻이 같을 수 없으니 한자를 일단 무시하고 의미를 상각하면 한 단어가 두 센스(즉, '語義'와 '語意')를 갖는 다의어이고, 일본어의 경우는 둘이 별개의 단어인데 의미는 같던 다르던 유사하니 類義語(synonym)라 하겠다. '기술'의 경우도 '어의'의 경우처럼 같은 형태-같은 音에-뜻이 다른 네 개의 동형어라 하겠다. 이 글의 취지는 단어의 다의성 문제를 단일 언어 내에서뿐 아니라, 교차언어적(cross-linguistic) 언어 간(inter-lingual)-'중간언어적'-시각에서 논의하는 것이다.

이 글은 다음 차례로 전개한다: 2 절에서 네 가지 워드넷을 조감하고, 지각동사 '보다'의 어의를 영어와 일어의 경우와 대비하고, 세계 워드넷(world wordnet, WWN)을 구상한다. 3 절에서 중간언어를 지향하는 사전과 문법을 제약기반적 통합문법 (Constraint-based Unified Grammar, CUG) 의 틀에서 설계하고 개별언어 문장으로 변환 생성하는 과정을 제시한다.

## 2. 개별 워드넷에서 세계 워드넷으로

워드넷(wordnet)은 데이터베이스 전자사전으로 프린스턴 대학의 Miller 교수 등이 개발한 WordNet(출처서, WN)이 그 효시이다. 이 WN과 유럽의 EuroWordNet (출처서, EWN)과 일본의 日本語語彙大系 (출처서, 日語大系), 그리고 현재 국내에서 연구개발 중인 21 세기 세종계획 전자사전 (출처서, 세종사전)-을 차례로 일람하고, 이들 워드넷과 여타의 워드넷을 통합한 세계 워드넷(world wordnet, WWN)을 구상한다.

### 2.1. WN

WN은 현재 WordNet 1.6판이 인터넷에 올라 있다(<http://www.cogsci.princeton.edu/wn/>). WN은 유의어 집합-신셋(synset < synonym set)-을 근간으로 하여 이밖에 상위어(hypernym) 하위어(hyponymy), 부분어(meronymy)와 같은 기본적인 의미 관계로 구축된 영어의 데이터베이스 전자사전이다. 명사, 동사, 형용사, 부사의 네 품사에 한정되었고, 수록된 WN(1.5)의 신셋 수는 91,591, 그 중 명사 신셋이 60,557이다. 온톨로지 입장에서 최상위에 명사의 유형(type) 11 가지를 설정하고 있다.<sup>1</sup> 이 유형은 다음 EWN에서 더 체계적으로 분류된다.

<sup>1</sup> 회상위 유형: entity, abstraction, group, phenomenon, location, act, state, event, psychological\_feature, possession, shape-form). Buitelaar (1998)의 CoreLex에서 28 유형을 추가하여 39 유형을 설정하여, 전체 부류를 1,648로 줄이고, 126 항의 의미 유형의 온톨로지를 설정하고 있다.

WN에서는 동형어와 다의어를 구분하지 않고 있다 이를테면, 등재어 'bank' 아래에 명사 센스 10, 동사 센스 7 이 등재되어 있고, 명사인 경우 센스-1이 '은행', 센스-2가 '뚝'의 뜻으로 되어 있다. 아래에 WN(1.6)에 등재된 기본적인 동사의 센스 수를 예시한다.

(2) 다의어의 예: 동사 (WN1.6)

*make* - 48    *give* - 45    *take* - 41    *get* - 37    *have* - 21    *do* - 13    *be*  
- 12

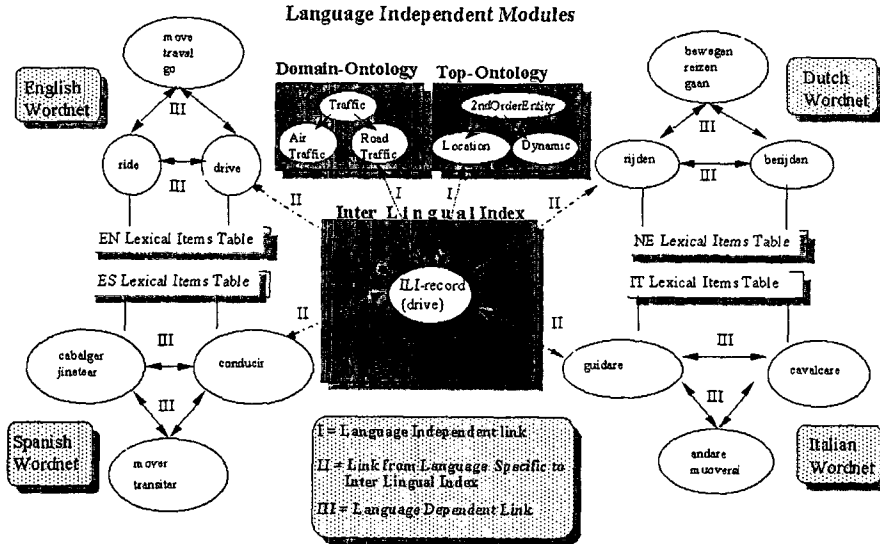
지각동사 'see', 'look' 와 관련된 WN의 예시는 2.6 절에서 지중적으로 다룬다.

## 2.2. EuroWordNet(EWN)

유럽의 8 개국-영국, 프랑스, 독일, 네덜란드, 스페인, 벨기에, 체크, 에스토니아-이 WN(1.5)를 모체로 하여 공동으로 개발하고 있는 EWN(<http://www.hum.uva.nl/ewn/>)은 신셋에 중간-언어-색인(Inter-Lingual-Index, ILI)을 구성하여 이 ILI에서 가장 가까운 개념을 찾아 등가 관계를 설정한다. 이 ILI는 WN(1.5)의 모든 신셋을 포함하고 더 나아가 신셋 간의 정밀한 등가 관계를 설정하기 위한 개념들로 확장되어 있다. 이 ILI의 등가 관계를 통해 한 워드넷이 다른 워드넷으로 옮겨갈 수 있고, 또 언어 간의 신셋과 그 관계를 비교할 수 있다. 언어의 모듈은 각각 신셋 간의 언어 내적 관계들의 자율적이고 특유한 시스템을 나타낸다. 다른 언어들과의 신셋과 WN 간의 등가 관계가 ILI를 통해 표현되므로 동일한 ILI-기록에 연결된 언어 개별적인 신셋은 모든 언어 사이에서 등가이다. 다음 그림에서 보듯이 ILI는 ILI-기록으로 구성되며, 이들은 언어 내적 모듈의 단어의 의미들, 최상위 개념과 영역 등을 제시되어 있다. 모듈들의 연계는 세 가지이다-I: 언어 독립적 연계; II: 개별 언어에서 ILI 로의 연계; III: 언어 의존적 연계.

(3) EuroWordNet 의 구조

Architecture of the EuroWordNet Data Structure



ILI는 구조화되지 않은 의미들의 리스트로서 그 안에 모든 ILI-기록은 신셋과, 의미를 기술하는 풀이말(gloss)과, 그 출처에 대한 참조로 구성되어 있다. 따라서, 언어 개별적인 등가 관계를 多-대-多의 관점에서 고려할 필요가 없다는 장점이 있다. 형태/통사적 차이, 스타일의 차이, 화용적 차이를 가리지 않고 같은 범위의 실체들을 나타낼 때 의미는 등가이다. 따라서 EWN에서 다음 예는 각각 동일한 신셋을 형성한다.

- (4) a. {people, folk}
- b. {cop, pig, policeman, police officer}<sup>2</sup>

또한 EWN에서는 WN과는 달리 품사 간의 유의성(cross-part-of-speech, XPOS\_NEAR\_SYNONYMY)을 허용한다.

- (5) a. 명사 ↔ 동사
  - movement N XPOS\_NEAR\_SYNONYM move V
  - move V XPOS\_NEAR\_SYNONYM movement N
- b. 상태성 명사 ↔ 동사
  - sleep N XPOS\_NEAR\_SYNONYM sleep V
  - sleep V XPOS\_NEAR\_SYNONYM sleep N

<sup>2</sup> WN에는 {cop, pig, bull, copper, fuzz}가 신셋을 이루고 그 상위어(hypernym) 신셋은 {policeman, police offer, officer}로 형성되어 있다. EW와 EWN의 차이를 보이는 예이다.

## c. 상대성 명사 ↔ 형용사/부사

poverty N XPOS\_NEAR\_SYNONYM poor A  
 poor A XPOS\_NEAR\_SYNONYM poverty N

EWN의 ILI-기록을 거쳐 보강된 워드넷-WN<sup>+</sup>-은 기타 개별언어의 워드넷-일본어 워드넷(JaWN), 한국어 워드넷(KoWN)-등으로 보완되어, 궁극적으로 세계 워드넷(WWN) 구축의 변칙과 데이터베이스를 제공한다.

## 2.3. 日本語語彙大系

日本語語彙大系(줄여서, 日語大系)는 日本電信電話株式會社(NTT) 커뮤니케이션 과학연구소에서 이케하라(池原) 교수 팀이 10 여년 연구 개발한 日-英 기계번역시스템 ALT-J/E(Automatic Language Translator-Japanese to English)의 번역사전 중에서 일본어 의미사전에 관한 부분을 추출하여 사람이 쓸 수 있게 편찬한 사전이다(5 권; 岩波書店, 1997). 본래의 ALT-J/E의 번역사전은 일본어 단어의 문법 정보를 수록한 『日本語文法辭典』, 의미 정보를 수록한

『日本語意味辭典』, 日-英 단어의 어휘 대응 관계를 수록한 『日英對照辭典』, 영어의 문법 정보를 수록한 『英語辭典』의 4 종류의 사전으로 구성되어 있고, 이 日語大系는 그 중 두 번째의 『日本語意味辭典』부분이다. 기계번역용의 의미사전에 수록된 단어 수는 30 만이며 명사, 고유명사, 용언으로 대별된다. 일반 명사의 의미속성체계에는 2,710 항이 의미속성이 설정되어 있고, 최상위 1 단에서 최하위 12 단까지의 깊이로 계층화되고, 의미속성에 일련 번호를, 최상위 좌측에서 아래로 위로 매기는 분류법을 쓰고 있다. 고유명사의 의미속성체계에는 130 항, 용언의 의미속성체계에는 36 항의 의미속성이 上下 관계의 나무 구조로 표시된다. 용언의 의미속성별 문형 패턴에는 용언의 의미속성에 그 의미속성을 갖는 용언과 日-英의 문형 패턴을 열거하고, 일반 문형 1 만, 관용 표현의 문형 6 천, 도합 1만 6천의 문형을 영어 문형과 대응시키고 있다. 이 日語大系가 바탕이 되어 보완될 일본어 워드넷(JaWN)를 가정한다. 이 JaWN이 WWN이 입력된다.

日語大系の 예시는 2.6 절에서 지각동사 ‘みる[miru]’를 들어 다의어 대응 관계를 살핀다.

## 2.4. 21세기 세종계획 전자사전

2000년말로 1 단계(3 년간: 1998~2000) 연구를 마친 <21세기 세종계획 전자사전>(줄여서 세종사전) 연구보고서(문화관광부 2000)에 따르면, 이 연구는 6 개의 소분과-체언사전 및 통합관리, 용언사전, 부사사전, 조사(어미 및 관형사사전, 고유명사사전, 통합전자사전-로 이루어졌다. 이 1 단계의 어휘 규모는 표제어 총 194,100로 체언(고유명사를 포함) 163,500, 용언 25,500, 부사 5,000; 조사/어미/관형사 1,600이다. 3 단계까지 지속될 이 세종사전의 연구 결과를 예견하면서 한국어 워드넷을 대표할 워드넷

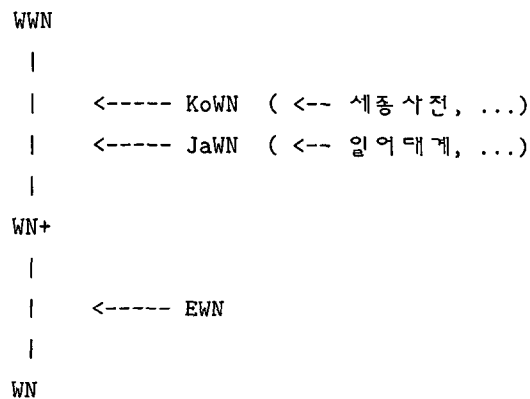
KoWN을 가정한다. 이 KoWN이 WN, EWN, JaWN과 더불어 WWN 구축에 입력된다.

2.6 절에서 세종사전에 근거하여 지각동사 '보다'의 다의성을 WN, JaWN의 대응어들과 대비한다.

## 2.5. 세계 워드넷

WWN은 WN을 모체로 하여 EWN 으로 보강되고, 이 보강된 WN+ 는 다시 KoWN, JaWN등으로 보완된다.

### (6) WWN



WWN 구축에 선행해야 할 조건은 개별언어의 워드넷-EWN, JaWN, KoWN 등-이 (1) 세분된 유의어 집합 (즉, 신셋)의 데이터베이스, (2) 온톨로지적 개념의 유형화된 데이터베이스, (3) 관용 표현을 포함한 영어와의 이중어(bilingual) 대역 사전을 구비하는 것이다. 이들 조건을 갖춘 WWN은 데이터베이스 전자사전으로서 다음과 같은 특성을 지닌다: (1) 온톨로지적 개념 위계에 기반을 둔다; (2) 어휘화된 신셋 군으로 구성된다; (3) 영어가 메타 언어로 사용된다; (4) 4 품사-명사, 동사, 형용사, 부사-는 어휘 등재에 어의 번호(sense-id #)가, 신셋에는 신셋 번호(synset-id #)가 부착된다; (5) 어휘 등재에는 EN, WN에서처럼 어의별 개념의 정의나 풀이(gloss), 온톨로지적 개념 유형의 위계에 따른 상·하위어(hypernym-hyponym), 유의어(synonym), 반대어(antonym), 전체어(holonym), 부분어(melonym) 등의 의미 관계의 정보가 등재된다.

WWN은 다음 단계에서 생성어휘부(Generative Lexicon, GL) 이론을 원용한 언어 독립적인 중간언어사전(Interlingual Dictionary, ILD)을 구축하는 과정에서 데이터베이스로 제공되고, 이 ILD는 중간언어문법(Interlingual Grammar, ILG)과 더불어

어 중간언어(interlingua, IL)를 구성한다. 이 IL은 의미/화용의 기술이 중심이고 제약기반적 통합문법 (CUG -장석진 1994, 2000a,b)의 틀에서 유형계층적 자질구조인 AVM(attribute-value matrix) 도식으로 표상된다. 중간언어의 의미표상과 개별언어로의 변환은 3 절에서 예시한다.

## 2.6. 다의어의 언어 간 대응

여기서는 WN, JaWN, KoWN의 데이터베이스를 이용하려 한국어(Ko)의 동사 ‘보다’와 이에 대응하는 일본어(Ja)와 영어(En) 단어들의 다의 현상을 대조한다.

- (7) a. (Ko) 보다  
 b. (Ja)みる[miru]: 見る 看る 觀る 診る  
 c. (En) see look (watch view ...)

한국어(Ko) ‘보다’와 일본어(Ja) ‘みる’는 겹으로 1-대-1 대응이 성립한다. 그러나 Ja에서는 발음은 같으면서(다 [miru]) 한자를 써서 적어도 네 가지로 달리 쓸 수 있다. 이 가운데서 ‘見る’가 기본이다. 영어(En)는 ‘see’와 ‘look’ 둘이 대충 ‘보다’, ‘見る’에 대응한다. 이밖에 ‘watch’, ‘view’ 등도 부분적으로 대응한다. 이 사정은 다른 지각동사 ‘듣다’의 경우도 비슷하다.

- (8) a. 듣다  
 b. きく[kiku]: 聴く 聞く 訊く  
 c. look listen

그러나 이들 지각동사의 용법은 화맥에 따라 같지가 않다.

- (9) a. 돌아다 보았지만, 아무것도 보지 못했다.  
 b. 回りを見たが何も見なかった。  
 c. I looked around, but I saw nothing.  
 d. 視之不見
- (10) a. 들어보았지만 아무것도 안 들렸다.  
 b. 聴いてみたが何も聞えなかった。  
 c. I listened but I heard nothing.  
 d. 聽之不聞

‘보다’에 한정해서 먼저 영어의 경우 WN(1.6)의 등재사항을 들여다 보자. ‘see’는 동사의 센스가 21 항(명사 센스는 하나)이며 그 중 (빈도수에 따른) 처음 기재된 3 항목만 아래에 옮겨 실고 그 전체는 부록에 수록한다. ‘look’은 동사의 센스가 9 항(명사 4 항)이다. 이 경우도 동사 센스 3 항만 실고 전체 리스트는 부록에 수록한다. 어휘 등재는

동일형의 품사별로 기재되고 빈도(frequency), 상위어(hypernom), 하위어(hyponym) (동사의 경우 WN에서는 'troponym'이라 부름), '문틀(sentence frame)'의 정보가 들어 있다. 등재 사항은 센스 별로, 먼저 해당 동사의 신셋의 구성 단어들-단일어로 된 'singleton' 신셋의 경우에는 해당 표제어 하나('see'의 센스-1의 경우)가 나열하되, 그 다음에 그 신셋의 개념 풀이(gloss)가 괄호 ( ) 안에 명시되고, 끝으로 예문(들)이 따옴표(" ") 안에 제시된다. 이 외에 유의어, 상·하위어, 반대어 등 관계에 대한 어휘 정보가 제공된다. 아래 (11)에 상위어(hypernomy) (EWN에서는 'hyperonymy' 이라 부름) 관계를 '→' 다음에 표기한다 (하위어 → 상위어).

(11) *see* (동사 21 센스 중 3 - WN1.6)

Sense 1

*see* - (perceive by sight; "You have to be a good observer to see all the details"; "Can you see the bird in that tree?")

⇒ perceive, comprehend - (to become aware of through the senses; "I could perceive the ship coming over the horizon")

Sense 2

understand, realize, *see* - (perceive mentally, as of an idea; "Now I see!"; "I just can't see your point")

Sense 3

witness, find, *see* - (perceive with any or all of one's senses; "We found Republicans winning the offices";

"You'll see a lot of cheating in this school"; give rise to or be characterized by; "The 1960 saw the

rebellion of the younger generation against established traditions")

⇒ experience, undergo, go through - (go or live through; "We had many trials to go through")

⇒ participate, take part - (be involved in something)

⇒ act, move - (perform an action; "think before you act"; "We must move quickly")

(12) *look* (동사 9 센스 중 3 - WN1.6)

Sense 1

*look* - (perceive with attention; direct one's gaze towards; "She looked over the expanse of land";

"Look at your child!" "Look-a deer in the backyard!")

Also See→ look back#1; look away#1; look after#1; look on#1; look up to#1; look for#1;

look into#2; look back#2; look to#1; look up#1

Sense 2

*look*, appear, seem - (give a certain impression or have a certain outward aspect; "She seems to be sleeping";

"This appears to be a very difficult problem"; "This project looks fishy"; "They appeared



like people

who had not eaten or slept for a long time")

⇒ be - (have the quality of being; (copula, used with an adjective or a predicate noun);

"John is rich"; "This is not a good answer")

Sense 3

look - (have a certain outward or facial expression; "How does she look?" "The child looks unhappy";

"She looked pale after the surgery")

'see' 와 'look'의 사용빈도수에 따른 상위의 어의는 '지각('perceive by sight')'으로 풀이되고, 'see'의 경우 센스-2는 신셋 {understand, realize, see}을 이루는 '이념의 심적 감지/이해 행위 (perceive mentally)', 센스-3은 신셋 {witness, find, see}을 이루는 '감각적 인지 행위(perceive with any or all of one's senses)'로 풀이되고 있다. 'look'의 경우는 시각 행위는 센스-1 ('perceive with attention /direct one's gaze toward') 뿐이고, 센스-2는 {look, appear, seem}의 신셋으로 '외형적으로 어떤 인상을 주다(give a certain impression or have a certain outward aspect)', 센스-3은 '외적 표현(certain outward or facial expression)'의 뜻으로 풀이 되고 있다. 상위 유형 관계로는 'see'의 센스-1, 센스-3은 각각 다음과 같다. (센스-2는 상위 관계가 미명되어 있지 않음.)

(13) see의 상위 유형 관계

a. 센스-1

⇒ {perceive, comprehend}

{see}

b. 센스-3

⇒ {act, move}

⇒ {participate, take part}

⇒ {experience, undergo, go through}

{witness, find, see }

WN에는 이밖에 '문틀'이라는 검색에서 문형(sentence pattern)이나 예문이 제시되어 있다. 'see'의 경우, 센스 1-3의 문형은 다음과 같다. 소잡한 하위범주화나 선택제약의 정보이다.

(14) (En) see의 문형 정보 (센스 1-3) 3

a. Sense 1

see - (perceive by sight; "You have to be a good observer to see all the

details”; ”Can you see  
the bird in that tree?”)

EX: Sam and Sue see the movie

b. Sense 2

understand, realize, see - (perceive mentally, as of an idea; ”Now I see!”;  
”I just can’t see your  
point”)

EX: They see that there was a traffic accident

c. Sense 3

witness, find, see - (perceive with any or all of one’s senses; ”We found  
Republicans winning

the offices”; ”You’ll see a lot of cheating in this school”; give rise to or  
be characterized by;

”The 1960 saw the rebellion of the younger generation against established  
traditions”)

\*> Somebody -s somebody to INFINITIVE

\*> Somebody -s that CLAUSE

일본어의 ‘みる’는 (7b)에서 언급하였듯이 발음은 같으면서 최소 네 가지 다른 한자로 표기할 수 있으며 ‘見’이 무표의 한자이고 (‘환자를 보다’) ‘診る’는 ‘見’으로 대치못하지만 ‘看る’, ‘觀る’는 ‘見る’로 대치할 수 있다. 동사의 등재사항은 그 동사의 문형 패턴의 번호를 [ ]안에 기재하고 이어서 용언의 의미속성 번호와 속성명을 기재하고, 그 우측에 다음 행의 문형 정보에 대한 변형정보를 ( ) 속에 기재한다. 변형정보는 진행형과 피동형 (‘受身(ukemi)’의 변환가능 여부에 대한 것이다.

(15) 문형 정보에 따르는 구문 변형 정보

- |              |                |
|--------------|----------------|
| a. (動作)      | 진행형 가, 피동형 가   |
| b. (狀態)      | 진행형 불가, 피동형 가  |
| c. (動作 受身不可) | 진행형 가, 피동형 불가  |
| d. (狀態 受身不可) | 진행형 불가, 피동형 불가 |

문형에는 N1(주어), N2(목적어) 등의 하위범주 정보와 이어 ( )속에 해당 N의 명사의 의미속성 번호가 기재된다. 문형마다 대응하는 영어문형이 기재된다. JaWN에는 용언의 의미속성 36항목, 일반명사 의미속성 2,710, 고유명사 의미속성 130 항목으로 각각 구성되어 있다. 동사 ‘見る[miruru]’는 문형 25 형이 등재되어 있고 빈도순에 의거한 문형 번호 1-5의 등재 내용은 다음과 같다. 문형 [1]의 의미자질은 22번 ‘결과’이고 이 문형

은 N2 자리에 특정 명사가 오는 관용 표현이다.<sup>3</sup>

(16) (Ja) 見る(みる) [miru] (25 문형 중)

[1] 22結果 (動作 受身不可)

N1가 馬鹿を 見る N1 make a fool of N1-self

[N1(4人)]

[2] 23身體動作 (動作)

N1가 N2를 見る N1 visit N2

[N1(4人) N2(407博物館)]

[3] 26生成 狀態 受身不可)

N1가 完成を 見る N1 be completed

[N1(533具體物)]

[4] 30知覺動作 (動作)

N1가 N2를 N3에 N4에 見る N1 see N2 on N3 in N4

[N1(3主體 535動物) N2(“畫面/モニター/顔/脈わい/映畫/映像/芝居)

N3(“表面”) N4(\*)]

[5] 30知覺動作 (動作)

N1가 N2를 N3에서 見る N1 watch N2 on N3

[N1(4人) N2(“テレビ” 970通信機器 1552放送 1120目錄) N3(“テレビ/テレビジョン”)]

위의 문형[2]는 ‘신체동작’, [3]은 ‘완성’, [4]-[5]는 ‘지각동작’의 의미속성이 각각 기재되어 있다. 영어 대역어는 [4], [5]에 ‘see’와 ‘watch’가 각각 기재되어 있다. WN이 의미속성인 센스에 입각해 있고 JaWN은 의미속성에 하위범주, 선택제약을 가미한 문형에 입각하고 있기 때문에 양자의 센스 설정의 기준이 같지 않지만 (E)WN과 JaWN의 대응을 위한 기술의 편의상 문형 번호를 센스 번호에 대응시킨다.

이제 세종사전의 ‘보다’의 등재 내용을 살펴보기로 한다. 전자사전의 등재 항목으로 동사에는 문자, 형태, 방언, 형태, 통사, 격틀, 선택제약, 영어 대역 등 많은 것이 논리적인 틀로 들어 있다.<sup>4</sup> 우선 ‘보다’는 하위범주(‘격틀’)에 따라 17 항으로 나뉘어

<sup>3</sup> 주어나 목적어등 명사 자리-N1, N2,...-에 특정 명사 (‘馬鹿[baka]’)들이 오면 관용 표현으로 간주한다.

<sup>4</sup> 틀이 틀이 있을 뿐 아직 채워지지 못한 곳이 많다. 여기서 검색하는 지각동사 ‘보다’, ‘듣다’, ‘보이다’ 등의 (영어) 번역 <trans> 항이 비어 있다.

등재되어 있고, 각 격들은 선택제약(selection restriction, selRst)과 의미역(theta role, thtRol)에 따라 다시 세분되어 있다. 이 점에서 JaWN과 흡사하지만, JaWN의 명사나 용언의 의미자질의 유형 별 분류 정보가 없고 영어 문형이나 어휘의 대역도 결여되어 있다. 이 선택제약과 의미역에 따라 문형[1]에는 23항의 하위 문형이 들어 있다. (17)의 문형 [1]-[5]와 [1]의 첫 5항의 하위 문형을 간소화하여 여기에 제시한다.<sup>5</sup> 격틀에서 주어 자리에 N0을 사용하고, JaWN에서 N1을 사용하고 있는 점도 표기 상의 차이점이다.

(17) (Ko) '보다'의 문형[1]-[5] (17 문형 중에서)

- [1] N0 N1-을 V
- [2] N0 N1-와 서로 V
- [3] N0 서로 V
- [4] N0 N1-을 ADV V
- [5] N0 S고1 V

(18) 문형[1]의 첫 5 하위 문형 (23 항 중에서)

[1] N0 N1-을 V

1. N0=인물 N1=인물

N0=AGT N1=THM

'기환이는 그를 보러 회사로 찾아갔다.'

2. N0=인물 N1=인물—추상(처지—사정—이름—성격)

N0=AGT N1=THM

'제발 나를 봐서 그 문제는 더 이상 거론하지 말아 나오.'

3. N0=인물 N1=(땅—집—물건—부동산)

N0=AGT N1=THM

'어머니는 땅을 보러 과천에 갔다.'

4. N0=인물 N1=(책—신문)

N0=AGT N1=THM

'영희는 집에서 하루 종일 소설책을 봤다.'

5. N0=인물 N1=(사전)

VN0=AGT N1=THM

'모르는 단어가 있으면 사전을 봐야지 어휘 실력이 늘지.'(=찾다, 찾아보

다)

첫 문형은 주어-목적어로 구성된 타동사 구문이다. 문형[1]은 의미역 정보가 행위자(agent, AGT)와 대상(theme, THM)으로 다 같으며 첫 [1.1]의 경우가 사람이 사람

<sup>5</sup> (98년도 연구 결과물에 포함되어 있는) '보다'의 기재에는 첫 격틀(caseFrame)의 일부가 두번째 것 하고 일부 중복되어 있다.

을 보는 경우이고, 사물(thing)은 그 대상에 들어 있지 않다. [1.2]이하는 대상이 특정한 대상으로 좁혀져 있다. 이런 경우 WN 식의 신셋 구성은 보다-1.1-1.5에 한정해서 살필 때 다음 (19a)와 같이 또 '사다'의 센스 번호를 매기면 (19b)와 같은 부분적인 신셋을 형성한다. 물론 '사다, '읽다', '찾아보다'도 그들 나름의 신셋 번호를 부착해야 한다.

(19) '보다'의 부분적 신셋

- a. {보다, 사다,...}, {보다, 읽다,...}, {보다, 찾아보다,...}
- b. {보다-1.3, 사다,...}, {보다-1.4, 읽다,...}, {보다-1.5, 찾아보다,...}

이제 WWN의 입장에서 개별 언어간의 단어와 센스의 다의 대응 관계를 '보다'의 예로 살펴보자. 앞 (9)의 예로 돌아가자.

(20) =(9)

- a. 돌아다 보았지만, 아무것도 보지 못했다.
- b. 回りを見たが何も見なかった \*<sup>6</sup>
- c. I looked around, but I saw nothing.
- d. 視之不見

WWN의 등재는 대문자로 표기한다. SEE, LOOK 등은 (기술의 편이상) WN의 센스 번호를 부착한다.

(21) '보다'의 어의 대응관계

[lex]	WWN	WN	JaWN	KoWN	한자
a.	SEE-1	see-1	みる-4	보다-1	見
b.	LOOK-1	look-1	みる-4	보다-1	視

같은 지각 행위이지만 영어에서는 그 대상에 따라 TV 같으면 'watch'를 쓰는데, JaWN에서는 이에 맞추어 'みる'의 문형을 별도로 설정해 문형[5]을 배정하고 있다 (<みる-5>). KoWN는 따로 문형을 설정하지 않고 있다.

- (22) a. I watch TV for one hour.  
 b. 一時間テレビを見ます?  
 c. 한 시간 TV를 봅니다.

(23)        WWN        WN        JaWN    KoWN        한자  
 WATCH-1   watch-1   みる-5    보다-1    視 / 觀

<sup>6</sup> 자연스러운 표현은 '... 何も見えなかった' (Ko: '...아무것도 안 보였다')이다. 이 경우, 보다-보이다(: 見る-見える)의 의미 관계도 살피는 것이 적절하겠으나, 여기서는 '보다(見る)'의 센스 안에서 논의한다.

영어는 일련의 시각 행위에서 see → look → watch 의 방향으로 집중력이 강해지는 (more attentive) 행위라고 하겠다.

- (24) WN: watch-1 (look attentively)  
 look-1 (perceive attentively)  
 see-1 (perceive by sight)

WN 나 EWN에 없는 외래어, 신조어 등이 WN+ 나 WWN에 새로 등재될 경우에는 EWN의 ILI (색인)에 개념을 풀이한 기록으로 인덱스를 추가하거나 WWN에 의미 속성에 따라 온톨로지를 검색하여 추가한다. 후자를 택하여, '김치'를 <KIMCHI> 의 형태로 다음과 같은 의미 속성으로 WWN에 수록한다. 이 바탕은 WN1.6)

- (25) <KIMCHI>

Sense 1

kimchi, pickle - (vegetables, especially cabbage, cucumber and radish)

⇒ relish - (spicy or savory condiment)

⇒ condiment - (a preparation (a sauce or relish or spice) to enhance flavor or enjoyment)

... [중간 상위 온톨로지 기재 생략]

⇒ food, nutrient - (any substance that can be metabolized by an organism to give energy

and build tissue)

⇒ substance, matter - (that which has mass and occupies space)

⇒ object, physical object - (a physical (tangible and visible) entity)

⇒ entity, something - (anything having existence (living or nonliving))

즉, <김치>는 영어 <pickle>과 동일한 신셋으로 개념을 파악한 것이다. 아니면 <pickle>을 <김치>의 일종(is-a)으로, 아니면 그 반대로 <김치>을 <pickle>의 일종으로 개념화할 수 있겠다. 한국사람은 'pickle'을 '김치'의 일종-즉, '서양김치'-으로, 서양사람은 '김치'를 'pickle'의 일종-즉, 'Korean pickle'-이라고 부를 것이다. 개념화의 차이를 볼 수 있다.

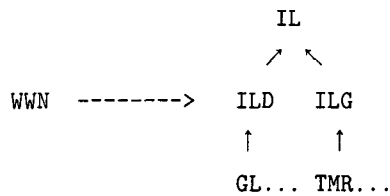
### 3. 월드워드넷에서 중간언어로

월드워드넷(WWN)에서 중간언어(interlingua, IL)로 가는 과정은 워드넷의 데이터베이스를 형식화하여 언어간 공통의 사전과 문법을 구축하여 그 사전과 문법의 틀에서 개별언어의 문법과 사전을 통한 개별언어의 문장으로 변환하는 과정이다. 이 점에서 국제어로 알려진 에스페란토 어<sup>7</sup>와는 다르다. 특수한 기능을 갖는 점에서 아라비아

<sup>7</sup> 폴란드의 Ludwik Lejzer Zamenhof(1859-1917)가 1887에 여러 (로맨스계) 유럽어를 모체로 하여 만든 국제어. Esperanto는 그의 필명으로 '희망하는 자'의 뜻.

숫자, 음악 악보, IPA 발음기호 등이 중간언어의 범주에 들 수 있겠다. 여기서 구상하는 IL은 WWN의 데이터베이스를 바탕으로 하여 어휘표상언어(lexical representation language) 로 영어를 사용한다. IL의 구성부인 사전(ILD)과 문법(ILG)의 두 부문에서, 전자에는 Pustejovsky(1995)의 생성어휘부 이론 (Generative Lexicon, GL)-특히 켈리아(qualia) 구조-를 도입하여 체계적으로 일어나는 다의성(systematic polysemy)을 자질구조로 기술하고, 후자는 Nirenburg 등의 텍스트의미표상(Text Meaning Representation, TMR) 이론 으로 가공된 제약기반 통합문법(CUG)의 틀에서 기술한다.

## (26) WWN에서 IL로



이렇게 해서 구축되는 중간언어의 사전과 문법은 상술한 GL과 TMR외에 HPSG와 같은 제약기반문법의 일반 특성-특히, 유형이론(Carpenter 1992)에 입각한 (1) 유형화된 자질구조, (2) 자질의 당연치(default value)의 계승과 그 무효화(defeasibility), (3) 계층화된 어휘부 (Koenig 1999)-등을 보유한다.

## (27) ILD

ILD는 GL의 켈리아 구조의 구성을 일부 수용한다. WN에서 세분된 센스 중 체계적인 다의성을 복합(complex; dotted) 유형으로 기술한다. 이를테면, 영어 'book'은 'physical object'와 'information'의 두 센스를 복합한 어휘개념 파라다임(lexical conceptual paradigm, lcp) 유형 'information·physobj' 으로 기술된다.<sup>8</sup>

## (28) BOOK (Pustejovsky 1995:116)

$$\left[ \begin{array}{l}
 \text{ARGSTR} = \left[ \begin{array}{l}
 \text{ARG1} = x: \text{information} \\
 \text{ARG2} = y: \text{phyobj}
 \end{array} \right] \\
 \text{QUALIA} = \left[ \begin{array}{l}
 \text{info-phyobj\_lcp} \\
 \text{FORMAL} = \text{hold}(y,x) \\
 \text{TELIC} = \text{read}(e,w, x\cdot y) \\
 \text{AGENT} = \text{write}(e', z, x\cdot y)
 \end{array} \right]
 \end{array} \right]$$

<sup>8</sup> 김윤신(2001) 참조. WN(1.6)에는 'book'에 최상위 유형으로 '실체(entity)'와 '추상(abstraction)'으로 나뉘어 그 하위 유형들을 달리하는 8가지의 센스가 기재되어 있다. 좁쌀알 같은 WN식 센스를 묶어서 센스와 신셋의 수를 줄이는 장치가 필요하지만 GL식의 다의성 처리에는 '체계적'인 다의에 한정되어 있다.

‘소설(novel)’은 BOOK-1의 켈리아 자질을 계승(<) 한다. 그러나 ‘사전(dictionary)’은 켈리아의 용도(telic) 자질의 값이 ‘참조하다(refer\_to)’로, ‘책’이나 ‘소설’ 용도의 값 ‘읽다(read)’를 덮어쓰운다(overwrite)-즉, ‘읽다’를 무효화하는 당연치로 ‘refer\_to’를 슬래시('/') 뒤에 매긴다(Copestake and Briscoe 1996:23).<sup>9</sup>

(29) *dictionary 1*

<QUALIA> < book \_1 <QUALIA>  
<QUALIA TELIC PRED> = /refer\_to

부연하면, *dictionary-1*의 특질(qualia) 값은 *book-1*의 특질(qualia) 값을 계승하고, 그 켈리아의 용도(telic)의 PRED는 ‘refer to’가 당연치이다. 즉, 사전은 ‘참고하는 것(refer-to)’이지 책이나 소설처럼 ‘읽는 것(read)’이 아니라는 정보를 어휘 구조에 수록하는 것이다.

어휘 해체(lexical decomposition)의 본보기로 여겨온 ‘kill’은 <CAUSE-BECOME-NOT-ALIVE> 식 어휘해체(McCawley 1971)를 지양하고 GL에서는 다음과 같은 자질 구조로 기술하고 있다.

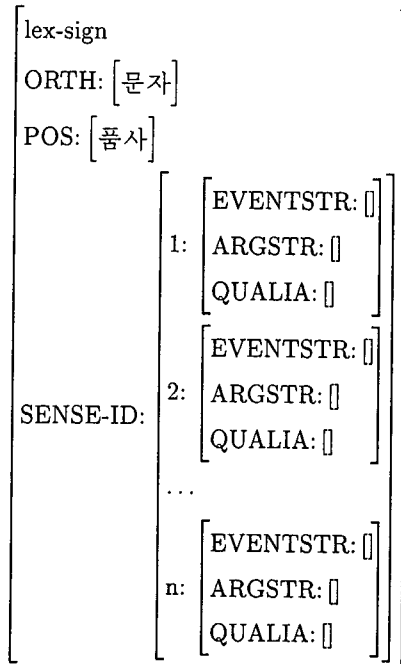
(30) KILL (Pustejovsky(1995:102))

$$\left[ \begin{array}{l} \text{kill} \\ \text{EVENTSTR} = \left[ \begin{array}{l} E_1 = e_1:\text{process} \\ E_2 = e_2:\text{state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = e_1 \end{array} \right] \\ \text{ARGSTR} = \left[ \begin{array}{l} \text{ARG1} = \textcircled{1} \left[ \begin{array}{l} \text{ind} \\ \text{FORMAL} = \text{physobj} \end{array} \right] \\ \text{ARG2} = \textcircled{2} \left[ \begin{array}{l} \text{animate\_ind} \\ \text{FORMAL} = \text{physobj} \end{array} \right] \end{array} \right] \\ \text{QUALIA} = \left[ \begin{array}{l} \text{cause\_lcp} \\ \text{FORMAL} = \text{dead}(e_2, \textcircled{2}) \\ \text{AGENTIVE} = \text{kill\_act}(e_1, \textcircled{1}, \textcircled{2}) \end{array} \right] \end{array} \right]$$

어휘부는 단어의 센스 별로 자질구조-사건구조(event structure, EVENTSTR), 논항구조(argument structure, ARGSTR), 특질(QUALIA)-구조로 구성된다.

<sup>9</sup> 이 무효화할 수 있는 당연치 표기는 Sag and Wasow(1999), CUG에서도 사용하고 있다. Copestake and Briscoe(1996:24)는, Sag and Wasow(1999), Koenig(1999) 등과는 달리, 유형(type)의 확산이 바람직하지 않다는 입장에서 어휘 자체를 유형으로 정의하지 않는다.



(31) ILD<sup>10</sup>의 자질 구조

## 3.1. ILG

ILG 구성을 위해 도입한 TMR의 틀을 먼저 소개하면, TMR은 자연어 문장의 의미를 언어 중립적으로 표현하는 일종의 중간언어이다. 입력 문장에 대해 온톨로지와 사전을 이용하여 형태소, 구문, 의미 분석을 함으로써 얻어지는 결과물이며, 개별 언어 생성 단계의 입력으로 사용된다. 즉, 개별언어의 생성 모듈로 의미를 전달하는 중간 의미 표현이다. 입력 문장의 의미를 나타낼 수 있는 온톨로지의 부분들, 즉 하위범주화된 격들을 선택하여 어휘화하고 구조화한 후에, 상(aspect), 태도(attitude), 양태(modality)와 같은 부가적인 언어 정보를 추가한다. TMR은 다음 7 부분으로 되어 그 전체가 본래 텍스트의 총체적 의미가 된다.

## (32) TMR의 구성

- (1) 목차(table of contents): 술어(predicate), 관계(relation), 문체(style)등의 일람.  
[이 부분이 제일 마지막 채워진다]
- (2) 서술(statement): 텍스트의 범위, 화자/필자, 청자/독자, 발화/문서 작성의 시간 등 포함.

<sup>10</sup> 이 AVM에서 자질 명과 값 사이에 편이상 콜론(:) 을 둔다. 유형(type)은 이탤릭 체, 자질 값은 소문자로 표기.

- (3) TMR 본문(body): 여기서 문장들이 IL로 표기. 절(clause) 단위로 본문에서 IL로 번역되는 부분.
- (4) 태도(attitude): 의도 등 머리 부분에서 채워짐.
- (5) 시간관계(temporal relation): 절과 절 사이의 시간 관계.
- (6) 영역 관계(domain relation): 통사 성분 사이의 연계 관계<sup>11</sup>
- (7) 공지시부(coreference section): 텍스트 본문에 나타난 동일 개체나 사상의 공지시 관계 기재.

TMR의 이 IL 표기는 유형화된 자질구조 AVM의 형식과 유사하여 우리의 중간언어표상 체계에 일부 원용된다. TMR의 시간관계(temp\_rel)의 유형에는 AFTER, AT, DURING 세 가지가 있다. TMR은 개별 언어를 영어로 먼저 번역한다. 번역은 절(clause) 단위로 차례로 자리-채우기(slot-filler) 형식의 틀에서 IL로 번역된다. 이 번역에는 단순한 명제의 번역뿐 아니라 문체(style), 초점(focus), 태도(attitude), 양태(modality), 담화관계(discourse relation)와 같은 화용적 특성들을 취급하는 자리(slot)가 마련되어 있다. ILG에서는 자리-채우기를 AVM 자질 구조로 대치한다. 아래에 TMR을 도입한 ILG의 담화/텍스트 단위의 자질구조를 제약기반 통합문법(CUG)의 틀에서 종합적으로 제시한다.

<sup>11</sup> 'domain'은 EWN의 전문영역의 온톨로지와는 다른 개념으로 두 문장을 잇는 Sowa(1984)의 담화 표지(discourse marker)와 유사하다. TMR에서 설정한 도메인: causal, volitional, non-volitional, reason, enablement, purpose, condition, result, conjunction, addition, enumeration, contrast, adversative, concessive, comparison, particular, representative, reformulation, inclusive-or, exclusive-or, definition, equivalence, arg<sub>1</sub>, arg<sub>2</sub>, value (optional decimal value between 0 and 1), scope (optional).

## (33) ILG (담화/텍스트 단위)

DS_ID:	$u_1$
SPEAKER:	<i>speaker</i>
HEARER:	<i>hearer</i>
SP_TIME:	<i>speech_time</i>
SP_MANR:	<i>plain   polite   casual ...</i>
MODE:	<i>assert / question / request ...</i>
TEMP_REL:	$\left[ \begin{array}{l} \textit{at\_rel}   \textit{after\_rel}   \textit{during\_rel} \\ \text{ARG1: } t_i \\ \text{ARG2: } t_j \end{array} \right]$
SPACE_REL:	$\left[ \begin{array}{l} \textit{on\_rel}   \textit{in\_rel}   \textit{at\_rel}   \textit{under\_rel} \dots \\ \text{ARG1: } x_1 \\ \text{ARG2: } x_2 \end{array} \right]$
EVENTSTR :	$\left[ \begin{array}{l} \text{E}_1: \left[ \begin{array}{l} \text{ARGSTR: } \left[ \begin{array}{l} \text{ARG1: AGT: } x \\ \text{ARG2: THM: } y \end{array} \right] \\ \text{SEM: } \left[ \begin{array}{l} \text{INDEX : } \textit{ref}   \textit{event} \\ \text{RESTR : } \langle \textit{psoa-list} \rangle \end{array} \right] \\ \text{ASPECT: } \textit{durative}   \textit{perfective}   \textit{iterative}   \dots \\ \text{TOPIC : } \langle x, \dots \rangle \\ \text{FOCUS : } \langle y, \dots \rangle \end{array} \right] \\ \text{E}_2: [] \\ \dots \end{array} \right]$
REF :	$\left[ \begin{array}{l} x_1: [] \ x_2: [] \ y: [] \dots \\ t_i: [] \ t_j: [] \dots \end{array} \right]$

약어: AGT agent, ARGSTR argument-structure, DS\_ID discourse segment identification, REF referent, RESTR restriction, SEM semantics, SP\_TIME speech time, SP\_MANR speech manner, TEMP\_REL temporal relation, THM theme, e(vent), psoa(parameterized state of affairs), rel(ation), u(tterance)

이 ILG의 자질구조에서는 화자(SP), 청자(HR), 화식(SP\_MANR) 등 화용 정보가 들보인다. 이런 정보를 표출시키지 않는 언어에서는 명시하지 않는다. 또한 이 구조에서 TMR의 자리-채우기(slot-filler) 방식을 따라 시간관계(TEMP\_REL), 공간관

계(SPACE\_REL), 지시(REF)의 값은 한 화단(discourse segment)의 다른 값들이 다 채워진 다음에 충당된다.

### 3.2. 중간언어에서 개별언어로

언어중립적인 의미표상에서 개별언어로 변환하는 과정을 앞에서 본 '보다' 예문들이 ILG에서 변환 생성된 과정을 생각해보자. 먼저 개념이 형성되고 어휘화한다.<sup>12</sup> 개념의 구성이 다음과 같은 명제의 형식으로 이루어진다고 가정하자: (1) 화자가 주위를 돌아다본다; (2) 그 사람은 아무것도 보지 못한다. 이 두 사건은 화자의 발화 시간 이전에 이루어졌다. 이 두 사건( $E_1, E_2$ )의 '보다'는 앞 (20)에서 검정했듯이 WWN에서 각각 SEE-1, LOOK-1으로 어휘화한다.

- (34) [1]  $E_1$ : SEE-1  
[2]  $E_2$ : LOOK-1

이 두 사건을 MRS(Minimal Recursion Semantics, 최소귀환의미론 -Copestake et al. 1996) 식 단층 논리형으로 표상하면 대충 다음과 같다.

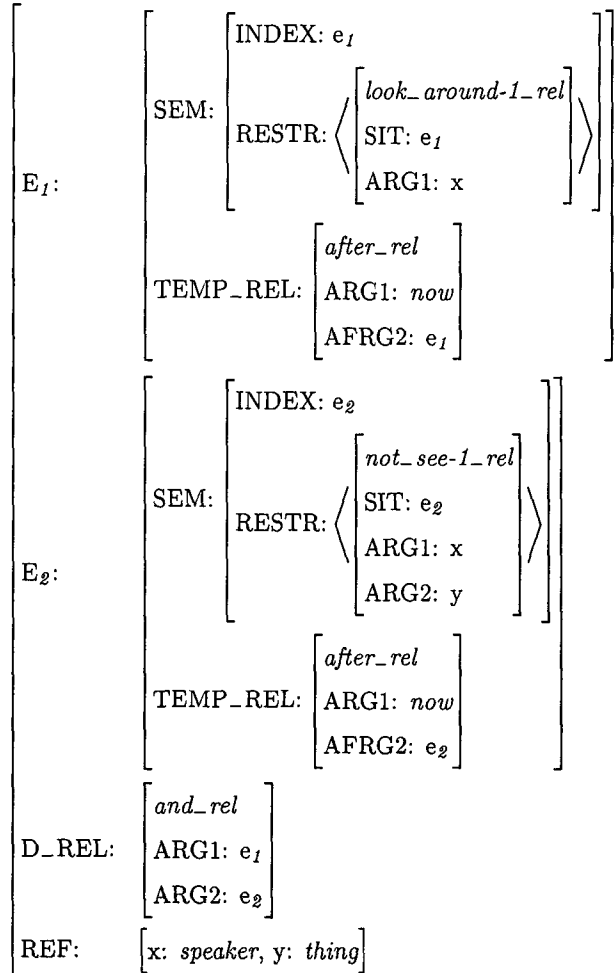
- (35) a.  $E_1$ : LOOK-1( $e_1, x, y$ )  $\wedge$  PAST( $e_1$ )  
b.  $E_2$ : NOT\_SEE-1( $e_2, x, y$ )  $\wedge$  PAST( $e_2$ )<sup>13</sup>

이 어휘 정보가 충당된 IL의 의미표상은 다음과 같다.

- (36) (35)의 IL 표상

<sup>12</sup> Sowa(1984, 1996)의 개념구조론은 개념의 도식(conceptual graph)으로 명제나 상황의 개념화를 잘 표상해 준다. 서상욱(1987), 구유선(1999) 참조.

<sup>13</sup> 'see'의 부정을 NOT(SEE( $e_2, x, y$ ))대신, 술어만의 부정('不見')으로 기술한다.



위의 IL 표상에 개별언어의 사전과 문법에 따라 형태통사적 구조를 취하고 음운구조를 취하고 음성화한다.

(37) = (20) = (9)

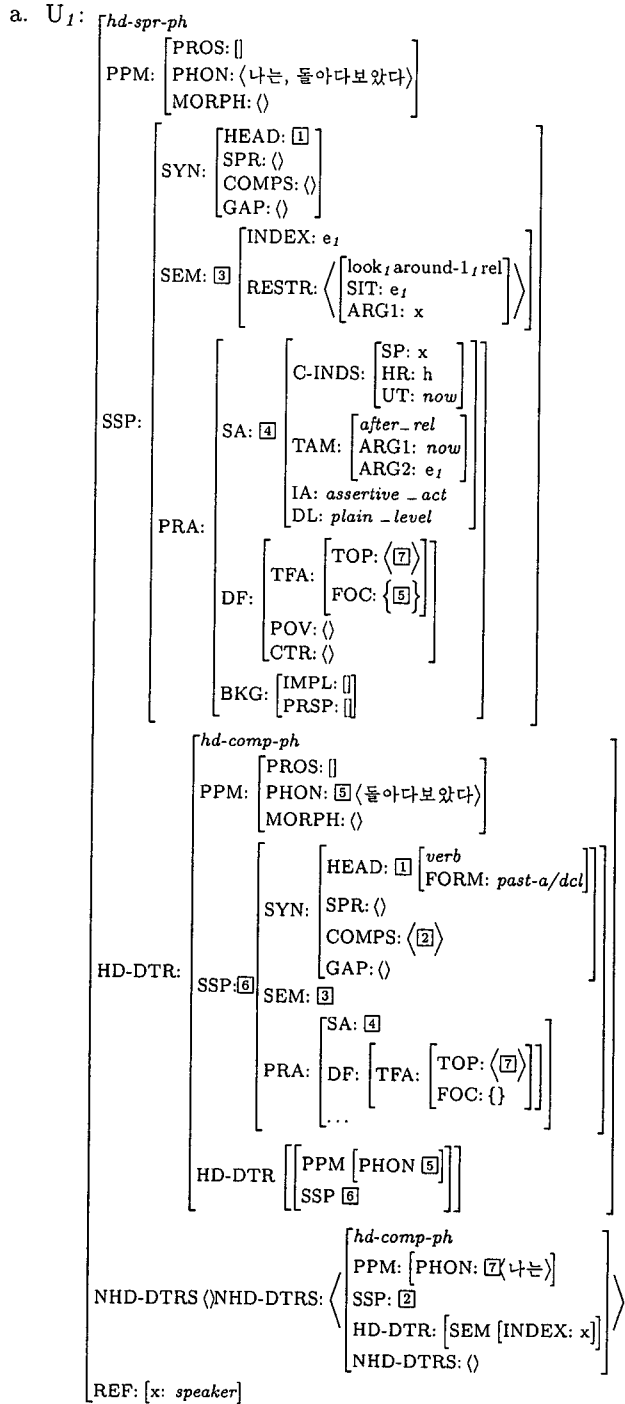
- a. 돌아다 보았지만, 아무것도 보이지 않았다.
- b. 回りを見たが何も見えなかった.
- c. I looked around, but I saw nothing.
- d. 視之不見<sup>14</sup>

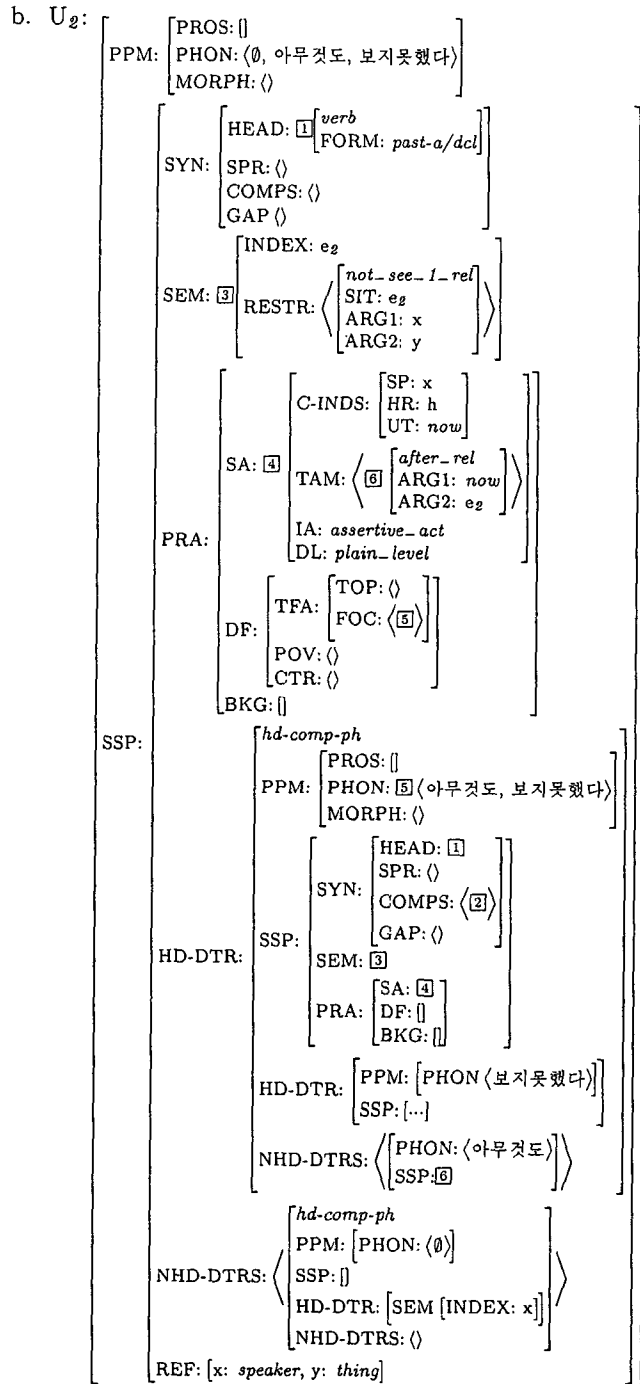
<sup>14</sup> 老子(14장)에 나오는 視之不見(名曰夷)와 聽之不聞(名曰希)는 '보(視) 아도 안보(不見)이고, 들(聽) 어도 안들린(不聞)다'로 해석하는 것이 (도올 김용옥을 포함해서) 일반적인데, 이경숙(2001:104 6)은 안보(不見)이는 것을 보(視)고, 안들리(不聞)는 것을 들(聽)는다'로 해석하고, 더 나아가 '視'와 '聽'은 수동적으로 '보고', '들'는 것이고, '見'과 '聞'은 적극적으로 '보'고, '들'는 것이라고 부연하고 있다. 그러나 '視聽'이 '見聞'보다 더 집중적(more attentive) 지각 행위임을 'look'와 'see'의 경우를 통해서 이미 살핀 바이다.

하나의 IL 표상이 형태통사적으로 여러 다른 문장-이문(異文, allosentence) 뿐 아니라 한 문장을 두 세 문장으로 만들어낸다. 한 의미(meaning)가 여러 형(form)으로 풀이(paraphrase) 되는 사례이다. 예를 들면, (39)은 (37a) 뿐 아니라 (38)의 '그러나'로 연결된 두 문장(a-문( $U_1$ ), b-문( $U_2$ ), 그리고 c( $U_1$  과  $U_2$ 의 연결))을 만들어 낸다.<sup>15</sup>.

(38) 나는 돌아다 보았다. 그러나 아무것도 보지 못했다.

<sup>15</sup> 이 CUG의 틀은 통사부분에서 Sag and Wasow(1999)의 모형을 수용한다. 최근의 CUG에 대해서는 장석진(2000a,b) 참조







c.  $U_1$  와  $U_2$  의 연결

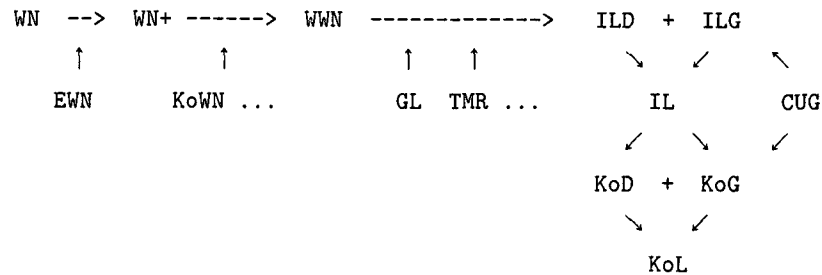
$$D\_REL: \begin{bmatrix} and\_rel \\ ARG1: u_1 \\ ARG2: u_2 \end{bmatrix}$$

이런 CUG 의 틀로 일본어, 영어의 문장도 변환 생성된다. 개별언어의 형태/통사적, 의미/화용적 제약을 받아 구현된다. 화용부의 담화 기능(discourse function, DF)인 TOPIC, FOCUS 등 TFA 정보도 일본어 영어에 구현되고, 화형(speech act, SA)의 DL(discourse level, 화계)은 영어에는 미명시로, 일본어는 공대(polite)와 평대(plain)의 두 화계로 된다. 통사부(SYN)의 FORM 자질도 언어에 따라 값이 달라짐은 물론이다.

## 4. 마무리

이상 개별언어들의 워드넷을 바탕으로 중간언어 사전과 문법을 구축하고 개별언어 문장으로 변환 생성하는 과정을 살펴보았다. 다의의 처리는 센스 번호를 부착하여 변별하는 방법을 채택하고, 어휘부의 자질구조의 구성을 위해 생성어휘부(GL) 이론을 일부 원용하여 제약기반적 통합문법에 접목하는 모형을 제시하였다. 중간언어의 설계와 구축은, 건축에서 그러하듯이, 필요 충분한 자재와 자료-데이터베이스-가 언어 별로 준비되어 있어야 한다. 다행히 WordNet의 개발과 그 범용성으로 미루어 이제 세계 워드넷(WWN)의 구축이 가시화되고 있다. 국내에서 추진중인 21세기 세종 계획에 의한 한국어 전자사전의 연구 성과와 서울대, 한국과학기술원(KAIST), 포항공대 등에서 이미 개발한 데이터베이스, 기계번역 시스템 등의 연구 성과가 집약되고, 통합되고, 표준화된 한국어 워드넷(KoWN)이 구축되어, WordNet, EuroWordNet, 『日本語語彙大系』와 더불어 WWN 구축에 한 몫을 하고, 나아가 중간언어 구축에 기여할 수 있는 날이 도래하기를 기대한다. 여기에 제시한 ‘제약기반 중간언어(CIL)’의 구상은 그날을 위한 하나의 작은 길잡이가 되었으면 한다.

(39) 정리:  $WN \rightarrow IL \rightarrow KoL$



## 참고문헌 및 사이트

- 구유선 1999. 영어 담화 표지의 개념구족적 기술. 충남대학교 박사학위논문.
- 김윤신. 2001. 파생동사의 어휘의미구조: 사동화와 피동화를 중심으로. 서울대학교 박사학위논문.
- 문화관광부. 2000. 21세기 세종계획 전자개발 연구보고서.
- 서상옥 1996.(편) 英語概念構造研究. 서울: 형설출판사.
- 이경숙. 2001. 노자를 웃긴 남자 [1 2 권]. 서울: 자인.
- 장석진. 1994. 統畝文法論. 서울: 서울대학교 출판부.
- 장석진. 2000a. '통합문법: 주제·초점과 시점' <언어와 정보 사회> 1:145-175. 서강대학교 언어정보연구소.
- 장석진. 2000b. 'Information Unpackaging: A Constraint-based Unified Grammar Approach to Topic-Focus Articulation,' *Japanese-Korean Linguistics* 10 (Stanford, CSLI Publications).
- 장석진. 2001. '자연언어 이해를 위한 중간언어 표상,' <학술원논문집(인문·사회과학편)> 제40집.
- 홍재성 외. 1997. 한국어 동사 구문 사전. 서울: 두산동아.
- Buitelaar, Paul. 1998. *Corelex: Systematic Polysemy and Underspecification*. Ph.D Dissertation, Brandeis University.
- Carpenter, Robert. 1992. *The Logic of Typed Feature Structures*. Cambridge: Cambridge University Press.
- Copestake, Ann and Antonio Sanfilippo. 1993. 'Multilingual Lexical Representaton,' in Dorr, ed. *Building Lexicons for Machine Translation*, Proceedings of the AAAI Spring Symposium, Stanford, CA.
- Copestake, Ann and Ted Briscoe. 1996. 'Semi-productive Polysemy and Sense Extension,' in Pustejovsky and Boguraev (1996), 15-67.
- Copestake, Ann, Dan Flickinger, Ivan A. Sag, Carl Pollard. 1999. 'Minimal Recursion Semantics: An Introduction,' ms. Stanford: CSLI, Stanford University.
- Fellbaum, Christiane, ed. 1998. *Wordnet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Koenig, Jean-Pierre. 1999. *Lexical Relations*. Stanford, CA: CSLI Publications.
- McCawley, James. D. 1971. 'Prelexical Syntax,' *Monograph Series in Language and Linguistics* 24, 19-33. Georgetown University.

- Nirenburg, Sergei, ed. 1987. *Machine Translation*. Cambridge: Cambridge University Press.
- Nirenburg, Sergei, Victor Raskin and Allen Tucker. 1987. 'The Structure of Interlingua in TRASLATOR,' in Nirenburg 1987:90-113.
- Pollard, Carl. and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Stanford: CSLI, Stanford University.
- Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky, James and Branimir Boguraev, eds. 1996. *Lexical Semantics: The Problem of Polysemy*. Oxford: Clarendon Press.
- Sag, Ivan A. and Thomas Wasow. 1999. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications.
- Sowa, John F. 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley.
- Sowa, John F. 1993. 'Lexical Structures in the Lexicon,' in Pustejovsky and Bergler, eds. *Lexical Semantics and Knowledge Representation. Lecture Notes in Artificial Intelligence*, 577-613. Berlin: Springer Verlag