

동사의 애매성 해소를 위한 시소러스의 이용과 한계

송영빈, 최기선

전문용어언어공학연구센터 첨단기술정보센터 한국과학기술원

{ybsong,kschoi}@world.kaist.ac.kr

Using Thesaurus for Disambiguation and it's limit

Young-Bin Song, Key-Sun Choi
KORTERM, AITRC, KAIST

요 약

동사의 애매성 해소는 언어학의 여러 부문 중에서도 가장 실체가 불명확한 의미를 다루는 것이기 때문에 언어학뿐만 아니라 자연언어처리에 있어서도 가장 해결하기 어려운 문제 중에 하나이다. 애매성은 언어학에서 말하는 동음이의어와 다의어를 동시에 포괄하는 개념으로 정의된다. 단일어를 대상으로 한 이와 같은 분류는 비교적 명확한 반면 두 개의 언어 이상의 다국어어를 대상으로 하는 기계번역용 사전과 같은 대역사전에 있어서는 동음이의어와 다의어의 구별은 경계가 불명확하여 의미에 기반한 대역어의 작성에 도움이 되지 않는다. 그 원인은 의미를 구성하는 세 가지 요소인 [실체], [개념], [표현]의 관점에서 [실체]와 [개념]은 어느 언어를 막론하고 보편적인 반면 [실체]와 [개념]을 최종적으로 실현하는 현대인 [표현]의 경우 각각의 언어에 따라 그 형태가 다르게 표출된다고 하는 사실 때문이다. [나무]라는 [실체]가 있다고 할 때 [나무]에 대한 [실체]와 [개념]은 언어를 초월해서 공통적이라고 할 수 있다. 한편, [개념]을 표현하는 실체인 [표현]은 언어에 따라 [namu](한국어), [ki](일본어), [tree](영어) 등과 같이 언어에 따라 자의적으로 [개념]을 표현하고 있다. [namu], [ki], [tree]가 같은 뜻을 나타낸다고 인식할 수 있는 것은 [개념]이 같기 때문이지 이들 각각의 [표현]이 의미적 연관성을 갖고 있기 때문은 아니다. 지금까지 의미를 다루는 연구에서는 이와 같은 관점이 결여됨으로 인해 의미의 다양성을 정확히 파악하는 데 한계가 있었으며 애매성 해소에 관한 연구도 부분적 시도에 그친 면이 적지 않다. 본고에서는 다국어어를 대상으로 한 대역사전의 구축에 있어서 다의어와 동음이의어에 대한 종래의 분류의 문제점을 지적하고 나아가 애매성 해소의 한 방법론으로 활발히 이용되고 있는 시소러스의 분류체계의 한계를 지적한다. 나아가 이의 해결책을 한국어와 일본어의 대역사전의 구축에서 얻어진 경험을 바탕으로 제시한다.

1. 머리글

본 논문은 다음과 같이 구성된다. 2장에서는 다의어와 동음이의어의 구별과 애매성 해소에 대해, 3장에서는 시소러스의 정의와 종류에 대해, 4장에서는 시소러스의 분류체계에 대해 5장에서 7장까지는 구체적 시소러스의 분류에 대해 언급하고 8장에서는 한일번역에 있어서의 시소러스의 한계를 밝힌다.

예) 의미1 눈이 빨강다

의미2 눈이 멀다

의미3 눈이 가다

:

의미 n

= {의미1(시각기관), 의미2(시력), 의미3(시선)···의미 n}

2. 다의어와 동음이의어

애매성이란 다의어와 동음이의어를 포괄하는 개념으로 정의할 수 있다.

동음이의어: [의미1]부터 [의미 n]까지의 의미적 연관성이 없다

예) 의미1 공을 치다

의미2 양을 치다

의미3 가지를 치다

다의어: [의미1]부터 [의미 n]까지의 의미적 관련성이 인정됨

:

의미 n

= 의미1(접촉), 의미2(사육하다), 의미3(자르다)

다의어의 경우, [의미1]부터 [의미 n+1] 사이에 의미적 연관성이 인정이 됨으로 형식적으로는 {의미1, 의미2, 의미3...의미 n}과 같이 동일한 단어의 다른 의미로 해석할 수 있다. 한편, 동음이의어의 경우, [의미1]과 [의미 n+1]의 각각의 의미의 의미적 연관성을 인정하기 힘들기 때문에 [의미1]과 [의미 n+1]은 서로 다른 단어로 취급을 하여 사전에서도 각각 다른 표제어로 등재하는 것이 일반적이다.

다의어와 동음이의어에 대해 애매성이라는 관점에서 분류를 하면 전자는 상보적 애매성(complementary polysemies)으로, 후자는 대립적 애매성, 혹은 대조적 애매성(contrastive ambiguity)으로 분류할 수 있다¹⁾. 이와 같은 용어의 필요성은 다의어가 동일한 단어의 의미적 파생관계에 기반한다는 점에서 적극적인 의미기술의 대상이 되어 왔던 반면, 동음이의어는 형태가 동일한 다른 의미의 단어로 인식되어 사전 기술에서는 다른 표제어로 등재시켜 적극적인 의미기술의 대상에서 제외되어 왔다는 점에서 적어도 용어상으로는 양자를 애매성의 문제로 포괄적으로 인식을 한다는 점에서 중요한 의미를 갖는다.

James Pustejovsky²⁾는 상보적 애매성의 경우 애매성의 원인은 다음과 같은 의미교체 원리에 의해 해소될 수 있다고 설명한다.

- (a) Count/Mass alternations; lamb
 1. The lamb is running in the field
 2. John ate lamb for breakfast
- (b) Container/Containee alternations; bottle
 1. Mary broke the bottle
 2. The baby finished the bottle
- (c) Figure/Ground reversals; door, window
 1. The window is rotting
 2. Mary crawled through the window
- (d) Product/Producer alternations; newspaper, Honda
 1. John spilled coffee on the newspaper
 2. The newspaper fired its editor
- (e) Plant/Food alternations; apple, fig
 1. Mary ate a fig for lunch
 2. Mary watered the figs in the garden
- (f) Process/Result alternations; examination, merger
 1. The company's merger with Honda will begin next fall
 2. The merger will produce cars
- (g) Place/People alternations; city, New York
 1. John traveled to New York
 2. New York kicked the mayor out of office

단어에 따라서는 (b)[용기]와 [내용물], (d)[생산자]와 [생산물]과 같이 복수의 의미에 의해 설명해야 하는 경우도 있다. 예를 들어 [newspaper]의 경우, (d)의 의미는 물론 (b)의 [내용물]까지도 포함해서 복수의 의미로 해석을 해야 하는 경우가 있다.

John get angry at the newspaper.

상보적 애매성은 하나의 단어가 관련하는 다른 의미속성을 갖는 경우에 나타난다. 단어가 갖는 여러 의미를 기술하기 위해 의미속성의 부여와 같은 방법이 활발히 이용되었다. 여기서 문제가 되는 것은 이들 의미교체 원리가 영어에서는 유용하지만 한국어에서는 의미속성의 부여를 필요로 하지 않는 경우가 있다는 점이다. (a)의 1은 [양], 2는 [양고기]에 대응하고 있으며 (b)의 1은 [병], 2는 [우유]에, (c)의 1은 [신문], 2는 [신문사]에 각각 대응함으로써 동일한 형태를 문제로 삼는 상보적 애매성이 한국어의 이들 어휘에서는 관찰되지 않는다는 점이다.

대조적 애매성은 상보적 애매성에 비해 한국어와의 관계에서 대조적 애매성이 전혀 관찰되지 않는다.

- (a) Mary walked along the bank of the river
메리는 뚝을 따라 걸어갔다
- (b) HarberBank is the richest bank in the city
허버은행은 그 도시에서 가장 자산이 많은 은행이다
- (c) Drop me a line when you are in Boston
보스톤에 오면 나에게 편지를 해라
- (d) We built a fence along the property line
우라는 토지구획선에 따라 담을 쳤다

(a)와 (b), (c)와 (d)의 영어에 대해 한국어는 각각 다른 형태가 대응을 함으로써 영어에서와 같은 대조적 애매성이 성립되지 않고 있음을 볼 수 있다.

이상의 기술을 통해 단일언어에서의 상보적 애매성과 대조적 애매성의 분류는 다국어론 전체로 한 연구에서는 의미를 갖지 않는다는 것을 확인할 수 있다. 또한 동시에 의미속성의 부여에 의한 번역 시에 일어나는 애매성의 해소의 가능성을 확인할 수 있다. 그러나 의미속성의 부여만으로 이들 영어 예문의 단어에 대한 애매성이 해소될 수 있을 것인가에 대해서는 의문이 남는다. 즉 (c)와 (d)의 경우 [π 끈지]와 [토지구획선]이라는 대역어는 영어의 [line]과 공기하는 요소들에 의해 구별이 되는 것이지 의미속성을 부여한다고 해서 해소되는 것은 아니다. 또한 상보적 애매성의 예에서도 (b)의 경우, 공기하는 동시에 따라 의미속성은 달라지는 것이지 단어의 의미속성에 의해 대역어가 달라지는 것은 아니다.

3. 시소러스의 정의 및 종류

시소러스는 단어가 갖는 개념적, 의미적 특성을 중심으로 많은 양의 단어를 분류한 것이다. 시소러스는 그 형태에 따라 어의를 기술한 사전형식의 것을 유의어사전이라 하고 단어만을 분류한 것을 시소러스라고 한다. 시소러스는 본래 작문 시에 동일한 단어의 사용을 회피하기 위한 것으로 작성이 되었

1) Weinreich, U.(1964) 「Webster's Third: A Critique of its Semantics」 『International Journal of American Linguistics 30』 pp.405-409

2) James Pustejovsky(1995) 『The Generative Lexicon』 MIT press, pp.31

으나 단어를 의미적인 관점에서 분류했다는 점에서 언어학의 연구 도구로, 혹은 자연언어처리를 위한 도구로 사용되게 되었다. 시소러스는 그 종류에 따라 다음과 같이 나눌 수 있다.

1. 유의어 집합을 이용한 것
2. 의미역할에 기반한 것
3. 의미속성을 이용한 것

1.의 대표적 사전으로는, Roget's의 시소러스(약 200여종), 角川類語新辭典(1981), 분류어휘표(1996) 등이 있다. Roget's의 시소러스는 의미를 이해하는 기준은 매우 다양한 요소들의 결합에 의해서 의미를 파악하고 이해한다. 또한 상위개념과 하위개념으로 단어를 분류하고 구조의 최하위부에는 개개의 단어가 위치한다. 최하위부의 단어는 의미적으로 관련성을 갖고 문장작성에 있어서의 동일어의 반복을 회피하기 위한 목적으로 작성이 되었다. 상위개념의 분류는 Roget's의 이름을 단 시소러스가 약 200개가 넘는 것 처럼 다양하다. 가장 일반적인 것으로는 약 1,000개의 개념에 의해 단어를 분류하고 있다. [깊이]의 항목에는 다음과 같은 다양한 품사가 대응하고 있다.

[Depth]

N. depth, drop, fall: deepness etc.
 Adj. deep, steep, plunging, profound: abysmal, yawning etc.
 Vb. be deep, - profound etc.
 Adv. deeply, profoundly: deep down, beyond ones depth

다양한 품사를 망라하고 있다는 점에서 어의에 의한 분류보다는 개념 중심의 분류가 두드러진다는 점이 특징이다.

3. 시소러스에 있어서의 계층구조

[분류어휘표]는 우선 품사에 따라 단어를 분류하고 이를 [추상적 관계], [인간활동의 주제], [정신 및 행위], [생산물 및 도구], [자연물 및 자연현상]의 5가지의 의미분야로 분류하고 있다. 수록어휘는 약 67,000단어로 비교적 일본어에서 중요한 단어가 망라되어있다. 의미분류보다 품사분류를 우선하고 있는 데 이것은 어떤 의미분야에 어휘가 집중되어 있는 가를 조사하기 위해 [분류어휘표]가 작성되었다는 이유에서 이다.

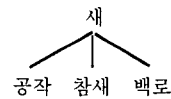
[角川類語新辭典]은 단어를 [자연계에 관련되는 것], [인간에 관련되는 것], [인간이 만든 것]의 세 종류로 나누고 그 이하의 하위 분류에 대해서는 도서관 분류법을 참고로 나누어 분기점을 10개로 통일하고 있는데 특징이 있다. [분류어휘표]가 품사분류에서 시작하고 있는데 비해 [角川類語新辭典]은 의미에서 출발하고 있다는 차이가 있다. 이들 유의어의 집합을 이용한 사전은 언어를 설명하기 위한 목적에서 편찬된 것으로 자연언어처리의 입장에서 다음과 같은 비판이 있다³⁾.

- (1) [분류어휘표]의 경우 분류의 관점이 불명확한 경우가 있고 [상위-하위관계]나 [전체-부분 관계]가 의미식별자 없이 존재하거나 한다.
- (2) 여러 연상관계를 이용함으로써 단어의 배열에 통일성이 없다.
- (3) [角川類語新辭典]의 경우 도서10진법에 의거 의미분류를 고정적인 관점에서 분류하고 있는데 이는 자연스러운 것이 아니다.

단어에는 예를 들어 [학교]의 경우 [조직], [건물], [장소]와 같은 다의성이 존재하는데 이들 사전은 이와 같은 다의어 기술에 있어서 미흡하다는 지적이다.

4. 단어의 분류체계

단어는 의미적인 관점에서 상위관계와 하위관계로 나눌 수 있다. 예를 들어 [새]와 [공작], [참새], [백로]라는 단어가 있었을 때 다음과 같이 표현할 수가 있다.



이와 같은 상하위 관계에 있어서 [새]는 [공작], [참새], [백로]라는 것을 대표하는 전의어(holonym)가 되고 [공작], [참새], [백로]는 [새]의 분의어(meronym)가 된다. 그러나 다음과 같은 경우 상위개념과 하위개념이 항상 전의어와 분의어 관계에 있다고 할 수는 없다. 예를 들어 [장미]는 [꽃]의 하위개념으로 [꽃]에 대해 분의어 관계에 있지만 [엔진]은 [자동차]의 상위 개념이 아니고 [자동차]를 구성하는 한 부분이다. 이 차이는 다음을 보면 명확해 진다.

- (1) ?장미는 꽃의 일부분이다.
- (2) 엔진은 자동차의 일부분이다.
- (3) 장미는 꽃의 일종이다.
- (4) ?엔진은 자동차의 일종이다.

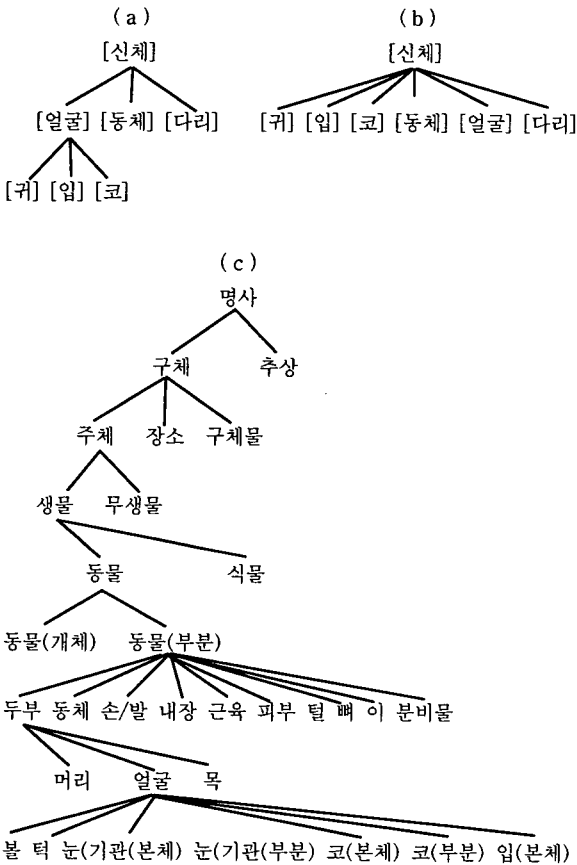
예를 들어 A, B, C라는 단어가 있을 경우, A가 B의 하위개념을 나타낸다고 하면 A는 C의 하위 개념이 된다고 하자. 위의 예에서 말하면 [장미]는 꽃의 하위개념이고 [꽃]은 [식물]의 하위개념이라고 하자. 그러면 [장미]는 [식물]의 하위개념이 된다. 이런 의미에서 [꽃]은 [장미]에 대해 전의어 관계에 있다고 볼 수 있다. 한편, [엔진]은 [자동차]의 분의어이며 [자동차]는 [교통수단]의 분의어 관계에 있지만 [엔진]이 [교통수단]의 분의어가 되지는 않는다. [장미], [꽃], [식물]의 관계는 분류라는 관점에서 일관성이 있지만 [엔진], [자동차], [교통수단]은 분류라는 관점에서 일관성이 없다. [자동차]와 [교통수단]은 [이동수단]이라는 관점에서의 분류이지만 [자동차]와 [엔진]은 [전체]와 [부분]이라는 관점에서의 분류이다. [장미], [꽃], [식물]이라는 관계에 [꽃잎]을 추가하면 [식물]과 [꽃잎]사이엔 분의관계가 성립하지 않는 것과 같은 것이다.

상위개념과 하위개념은 언어에 의해서도 달라진다. 한국어에서 [고

3 池原悟他(1997) 『日本語語彙大系 1 意味体系』 岩波書店, pp. 23-24

기]라고 했을 경우는 [닭고기], [소고기], [돼지고기] 등이 다 포함되지만 영어의 경우 [닭고기]가 포함되지 않는 경우가 있다. 마찬가지로 중국과 한국에서는 [해바라기씨]를 [음식]의 하위개념으로 분류할 수 있는데 일본에서는 [음식]으로 [해바라기씨]를 먹지 않는다.

상위개념과 하위개념에 속하는 단어들을 어휘적인 관점에서는 상위어와 하위어로 나눈다. 이때 문제가 되는 것은 상위어와 하위어를 몇개의 계층구조로 표현하느냐는 점이다. 계층구조의 표현에는 여러 가능성이 있다. 예를 들면 [얼굴]의 경우 [코], [입], [볼] 등이 있는데 이를 다음과 같이 여러 개의 계층구조로 나눌 수가 있다.



또한 [분류어휘표]의 [턱]을 보면 다음과 같이 5개의 계층구조에 의해 표현되고 있다.

- 1.0000 명사
- 1.5000 자연물 및 자연현상
- 1.5700 신체
- 1.5710 머리, 눈코, 얼굴
- 1.5712 턱

이들 각 계층구조에 있어서의 차이는 시소러스 구축의 목적의 차이가 하나의 원인으로 작용하고 있다. [분류어휘표]는 일본어의 어휘통계

를 위해 작성된 것으로 거시적인 통계를 위해서 품사분류를 우선하고 있고 (c)의 경우 [일본어어휘대계]의 분류인데 이는 일영번역시스템에서의 응용을 위한 분류이다. 기계번역의 경우, 보다 어휘의미에 접근한 분류체계가 사용되고 있음을 알 수 있다.

5. IPAL사전

IPAL사전은 일본의 정보처리진흥사업회기술센터에서 작성된 사전으로 [계산기용 일본어 기본 동사 사전]과 [계산기용 일본어 기본 명사 사전] 및 [계산기용 일본어 기본 형용사 사전]의 세 권으로 구성되어 있다. 이중 [명사사전]이 일종의 시소러스로 볼 수 있다. [명사사전]을 일종의 시소러스라고 표현한 것은 의미기술이 다른 시소러스와는 달리 동사와 공기하는 명사의 의미역할을 중심으로 명사를 분류했다는 점이다. [명사사전]에서는 58개의 의미역할로 명사를 분류하고 있는데 예를 들어 [병원]의 경우 다음과 같이 복수의 의미역할에 의해 의미를 분류하고 있다.

- 병원을 개축하다<CON>
- 병원에 실려가다<INT>
- 병원에 달려가다<LOC>
- 병원을 개원한다<PRO>
- 병원에 입원하다<ORG>
- 병원이 논다<APO>

의미역할에 의한 기술은 표면상 나타나지 않는 문장 내부에서의 의미역할을 동사와의 관계에서 기술한다는 점에서 격조사를 쓰지 않는 영어와 같은 인도유럽어족에 속하는 많은 언어와의 번역에서 유효하다. 또한 IPAL명사사전에서는 위와 같은 속성들을 개개의 동사의 문형에서 기술하는 한편 명사사전에서는 다음과 같이 닷타입표현으로 기술한다.

병원:[CON · INT · LOC · PRO · ORG · APO]

이와 같은 의미역할의 기술은 한국어와 일본어처럼 한자어를 공유하고 문법적으로도 격조사를 사용하는 언어에 있어서는 유용한 방법이 아니다. 예를 들어 심층격을 주장하는 사람들은 표층에는 나타나지 않는 다음과 같은 문장에서의 (1)과 (2)의 [철수]의 의미적 차이를 심층격을 기술함으로써 설명할 수 있다고 한다⁴⁾.

- (1)철수는 화재로 집을 태웠다.
- (2)철수는 가스로 생선을 구웠다.

(1)의 문장에서의 [철수]는 실수로 집을 태운 것이기 때문에 [경험자격]이 되며 (2)의 경우는 목적을 갖는 의도적인 행위를 나타내기 때문에 [철수]는 [동작자격]이 된다는 것이다. 또한 [화재로]의 경우는 [원인격]이 되고 [가스로]의 경우는 [도구격]이 된다. 그러나 다음의 일

4國立國語研究所(1997) 「日本語における表層格と深層格の對應關係」三省堂、pp.7

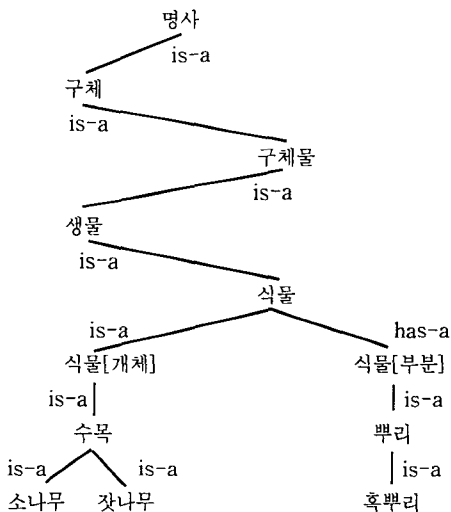
본어와 같이 심층격의 기술은 한국어와 일본어의 번역에 있어서는 의미를 갖지 않는다.

- (1)' 太郎は火事で家を焼いた
- (2)' 太郎はガスで魚を焼いた

한국어 (1), (2)에서는 [태우다], [굽다]가 일본어 (1)', (2)'에서는 모두 [焼く]가 되어 있는 것을 알 수 있다. 이러한 대응관계를 설명하기 위해서는 심층격의 기술보다는 목적격의 위치에 오는 명사와 용언의 결합관계를 기술해 주는 것이 명시적이다.

6. [일본어어휘대계]

일본어와 영어의 번역을 위해 작성된 [일본어어휘대계]는 일영번역을 위해 최소한 3,000개의 의미속성이 필요하다고 주장한다. 의미속성은 IPAL사건과 같은 평면적인 기술이 아니라 앞서 3장에서 본 바와 같이 계층구조로 표현이 된다. 이와 같은 표현구조는 기본적으로 [is-a관계]와 [has-a관계]에 의해 분류된다⁵⁾.



이와 같은 계층구조를 이용하는 이유에 대해 [일본어어휘대계]에서는 다음과 같이 밝히고 있다.

- (1) 용언의 문형패턴에 있어서의 의미 제약 조건을 기술할 시에 필요에 따라 상위의 의미속성을 이용해서 자세히 기술하거나 상위의 의미속성을 써서 거칠게 기술할 수가 있다.
- (2) 임의의 의미속성의 바로 위에 있는 속성을 하나의 노드에 의해 표현함으로써 신속한 검색이 가능하다.
- (3) 상위의 의미속성(노드)의 성질(속성)을 하위의 의미속성(노드)에 계층할 수 있게 하여 하위의 의미속성을 정의하기 위한 기술을 삭감할 수 있다.

(4) 상위의 기본적인 분류체계를 흐트러트림 없이 필요에 따라 최하위의 분류를 세분화하는 것에 의해 계층적인 의미속성을 확장시킬 수 있다.

(1)의 기술은 실제의 대역사전 작성에 있어서 매우 중요한 정보라고 할 수 있다. 각 최하위 노드에는 명사목록이 따르는데 동사에 따라 명사목록의 범위가 달라진다. 간단한 예로 주격에 오는 [인간명사]가 있다고 했을 때 이를 기술하기 위해 [인간명사]를 나타내는 상당량의 의미속성을 일일이 기술할 필요는 없다. 예를 들어 [경찰이 도둑을 잡았다]와 같은 문장이 있다고 할 때, 문형을 시소러스의 분류체계에 맞추어서 다음과 같이 작성했다고 하자.

[경찰]이 [도둑]을 잡다

[경찰]과 [도둑]에는 여러가지 상위 개념이 있어서 각각 다음과 같은 계층구조에 의해 이를 표현할 수 있다.

경찰:인간>인간(직업/지위/역할)>인간(직업)>인간(보안직)>경찰
 도둑:인간>인간(직업/지위/역할)>인간(직업)>인간(자유직)>도둑

이 경우 앞서 제시한 문형은 과연 효과적인 문형인가 하는 문제가 있다. 예를 들어 [도둑이 경찰을 잡다]와 같은 문장도 성립이 되기 때문에 최하위 노드인 [경찰]과 [도둑]을 문형에서의 의미속성으로 기술하는 것은 효과적인 방법이 되지 못한다. 따라서 이 경우, 다음과 같은 의미속성을 보다 상위개념인 [인간]을 써서 표현하는 것이 효과적이다.

[인간]이 [인간]을 잡다

한편 [드리다] [주다], [여쭙다][묻다]와 같이 경어적인 차이가 나는 동사의 경우 공기명사의 제한이 필요하다.

[인간(하위자)]가 [인간(상위자)]에게/께 여쭙다

이와 같은 동사에 따른 명사의 공기제한을 위해서는 계층구조에 의한 시소러스는 매우 유용하다. 그러나 시소러스는 어휘적인 의미보다는 개념적인 분류가 앞설 수 밖에 없기 때문에 섬세한 번역, 혹은 대역사전을 구축할 시에 공기제한 정보로 사용하기에는 부족한 점이 많은 것도 사실이다.

7. Lexical Semantics의 어휘계층구조

James, Pustejovsky(1995)는 의미를 표현하는 것은 사전이라고 정의하고 다음과 같은 정보를 사전에서 기술할 것을 제안하고 있다.

- 항구조(AS): 단어가 요구하는 논항의 수와 의미속성
- 사전구조(ES): 단어가 나타내는 사건의 타입과 내부구조
- 특질구조(QS): 단어가 나타내는 개념고 인식양식

5池原悟·宮崎正弘(1997)「日本語語彙大系」岩波書店、

논항구조에서는 예를 들면 [칠하다]의 경우

X가 Y에 Z를 칠하다

와 같이 세 개의 논항을 요구한다⁶⁾. 이의 각각의 논항의 의미속성은 다음과 같이 표현된다.

X= human_ind

Y= physobj

Z= [material, FORMAL=mass]

사건구조는 (1)전체적으로 영향이 남는 경우와 (2)부분적으로 영향이 남는 경우로 나눌 수 있다.

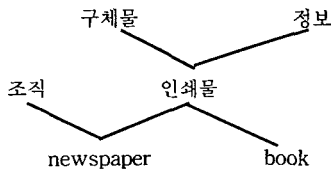
- (1) 철수가 벽을 페인트로 칠했다
- (2) 철수가 벽에 페인트를 칠했다

(1)의 경우는 [상태]에 중점이 놓인 표현이며 (2)의 경우는 [동작]에 중점이 놓인 표현이다.

특질구조는 단어의 개념과 인식의 방법을 기술하는 것으로 다음 4가지에 의해 구성된다고 한다.

- (1)구성역할: 재료, 재질, 성분, 무게 등(part of X)와 같은 내적 성질
- (2)형식역할: 구체물인가 추상물인가, 자연물인가. 인공물인가, 성별, 색, 크기 등의 외적 성질
- (3)목적역할: 대상물이 본래적으로 의도된 목적이나 기능
- (4)주체역할: 이를 생성해 내는 동작이나 원인, 성립, 출처

이는 명사의 속성과도 연관이 되는 것인데 Pustejovsky는 이를 어휘 계층구조(Lexical inheritance structure)라고 부르고 있다. 예를 들어 [book]과 [newspaper]에 대해 다음과 같이 표현하고 있다⁷⁾.



[newspaper]sms [조직]외에 [인쇄물]로서의 [구체물]과 [정보]라는 속성을 갖는데 [book]의 경우는 [조직]으로서의 의미속성을 갖지 못한다. 이는 다음과 같은 문장을 통해서도 확인할 수 있다.

- (1) The newspaper has just fired its sports editor.
- (2)*The book has just fired its sports editor.

여기서 문제가 되는 것은 어휘계층구조의 이론은 예시되어 있으나 그의 실천으로 시소러스가 제시되어 있지 못하다는 점이다. 또한 한국어나 일본어와 같은 경우, [newspaper]가 [조직]을 나타내는 경우는 [조선일보], [동아일보]와 같이 고유명사를 쓰거나 [신문사]와 같이 표현하며 [정보]와 [구체물]을 나타낼 때에는 [신문]이라 표현한다. 여기서도 영어만을 대상으로 한 어휘계층구조는 한계를 보인다.

8. 대역사건의 입장에서 시소러스 및 어휘기술

[일본어어휘대계]의 [칠하다]에 해당되는 [塗る]를 보면 영어와의 번역을 위해 다음과 같이 5가지 문형으로 나누고 있다.

(a) 屬性變化 身體動作

N1가 N2를 N3에 塗る N1 color N2 N3

[N1(3主体) N2(具体) N3(2352色彩)]

(b) 屬性變化 身體動作

N1가 N2를 N3에 塗る N1 coat N3 with N2

[N1(3主体) N2(809塗料) N3(389施設 760人工物)]

(c) 身體動作

N1가 N2를 N3에 塗る N1 apply N2 to N3

[N1(3主体) N2(-803農藥 -804化粧品等 -806石鹼 -807齒磨き

-808香料 -809塗料

-810染料 -812火藥 800藥品 801藥品類(医用) 802藥品類(非医用)

805化粧品 811

接着劑) N3(4人 533具体物)]

(d) 身體動作(動作)

N1가 N2를 N3에 塗る N1 spread N3 with N2

[N1(4人) N2(838食料) N3(838食料)]

(e) 感情動作(動作)

N1가 泥을 N2에 塗る N1 disgrace N2

[N1(3主体) N2(3主体)]

(a)에서 (d)까지는 [무엇인가를 무엇에 칠하다]는 점에서 의미적 관련성이 있는 것으로 인정할 수 있다. (e)는 관용적인 것으로 여기서는 고찰 대상에서 제외한다. 일반적으로 (a)에서 (d)까지는 신체동작과 속성변화의 두 종류로 나눌 수 있다. 이는 Pustejovsky, 남승호(1999)가 [paint], [칠하다]를 두 개의 의미로 나눈 것과 일치한다. 그러나 [일본어어휘대계]에서는 영어로의 번역을 위해 더욱 세분화된 문형을 제시하고 있다는 것이 특징이다. 한편 한국어와의 관계를 보면 다음과 같이 두 개로 분류할 수 있다.

- (a) 太郎が赤を紙に塗る → 칠하다
- (b) 太郎がペイントを壁に塗る → 칠하다
- (c) 花子が軟膏を手に塗る → 바르다
- (d) 花子がパンにバターを塗る → 바르다

[연고], [버터]의 경우는 [바르다]가 되고 [빨강], [페인트] 등은 [칠하다]가 된다. 한편, [바르다]와 같은 동사는 [빠를 바르다]와 같은 뜻도 갖고 있다(이와 같은 경우는 일반사전에서는 다의어로 간주하지

6 남승호(1999)[한국어 다의 술어의 유형론과 의미 명세] 인지과학회 발표자료

7 影山太郎(1999)「形態論と意味」くろしお出版、pp.41

않고 동음이의어로 간주하여 따로 기술하고 있다. 그러나 한일대역구문사전에서는 형태가 같은 것은 모두 다의어로 간주하고 기술 대상으로 삼고있다. 또한 [바르다]의 경우 [벽지를 바르다]와 같이 표현하기도 한다. 이 경우 일본어 문형 (d)와는 또다른 의미가 된다. 또한 일본어의 경우 [벽지]가 오면 [塗る]가 아니라 [張る]가 된다. 즉 개별적 공기 명사에 의해 공기하는 동사가 모두 달라지는 것이다. 이는 대역사전에 있어서의 공기명사를 중심으로 하는 동사 대역어의 선정의 중요성을 보여주는 것으로 시소러스의 구축에 앞서 선행될 문제임을 뜻한다고 볼 수 있다.

9 결론

시소러스의 유용성과 이의 가장 구체적이고 체계적인 실천이라고 할 수 있는 [일본어어휘대계]의 예를 통해 시소러스의 한계를 살펴보았다. 효율적인 대역사전을 위해서는 동사에서 공기하는 명사를 보는 것도 중요하지만 명사에서 동사를 보는 것도 그에 못지 않는 다는 것을 간단한 예를 통해서도 확인할 수 있다.

참고문헌

- [1] 田中穂積監修(1999) [自然言語処理-基礎と応用]. 社団法人電子情報通信学会編
- [2] Pustejovsky, J.(1995)[The generative Lexicon] MIT press
- [3] 남승호(1999)[한국어 다의 술어의 유형론과 의미 미명세] 인자과학회발표자료
- [4] 강범모외(1999)[형식 의미론과 한국어 기술]한신문화사
- [5] 池原悟、宮崎正弘他(1997)[日本語語彙大系]岩波書店
- [6] 情報処理振興事業協会技術センター(1987)[計算機用日本語基本動詞辞書]IPAL
- [7] 情報処理振興事業協会技術センター(1996)[計算機用日本語基本名詞辞書]IPAL
- [8] 影山太郎(1999)「形態論と意味」くろしお出版

감사의 글

본 연구는 전문용어언어공학연구센터에서 수행한 과학기술부와 KISTEP의 핵심소프트웨어 사업 중 “대용량 국어정보 심층처리 및 품질관리 기술개발”과제의 일환으로 수행되었으며, 첨단정보기술연구센터를 통하여 과학재단의 지원도 받았음을 밝힌다.