

# 온톨로지구축을 위한 국어사전의 정의문분석과 어의분류

오승태  
한국의국어대학

1. 들어가기
2. 사전적 정의와 온톨로지
3. 기존 국어사전의 정의문의 문제와 그 해결책으로서의 중심어미체계
4. 온톨로지구축을 위한 국어사전 정의문 활용의 문제와 해결방안
5. 국어사전의 어의분류와 온톨로지구축을 위한 제안
6. 나오기

## 1. 들어가기

이 논문은 처음 조평옥 외(1999) 등과 같이 사전의 정의문의 중심어에 근거하여 명사의 의미계층체계를 구축한 후 이를 온톨로지 구축에 활용하려는 목적으로 작업을 하는 가운데 드러난 문제점을 제시하고 그 해결 방안을 모색하려는 의도에서 시작되었다.<sup>1)</sup>

사전의 정의문에 기반하여 의미체계를 구현하는 모든 방법론에 근본적인 문제가 내재되어 있음을 지적하지 않을 수 없다. 이는 기존 사전의 정의문 자체의 문제이기도 하고 사전의 정의문을 활용하여 온톨로지를 구축하는 방법의 문제이기도 하다.

어의분류문제는 사전학의 혹은 사전편찬의 문제이지 온톨로지 구축의 문제는 아니다. 그러나 사전의 정의문을 기반으로 온톨로지를 구축할 경우에는 이에 대해 신중히 고려하지 않으면 안 된다. 기존의 사전기반의 의미분류체계를 구축하는 대부분의 연구에서 동형어의 구별 정도에서 그치고 있다는 것은 중대한 문제이다.

온톨로지는 기본적으로 단어의 일반적인 의미에 기반 하는 것도 사전의 단어의 하위 부류인 어의에 기반 하는 것도 아니다. 사전의 정의문은 단어가 갖고 있는 어의에 대한 뜻풀이이다. 그리고 하나의 단어는 하나의 어의만으로 구성되는 경우보다는 둘 이상의 어의로 구성되는 것이 일반적이다. 더 나아가 하나의 어의는 대부분 하나의 개념을 갖지만 두 개 이상의 개념으로 구성되는 경우도 종종 있다.

사전에서는 개념간의 차이는 유의미하게 구별되지 않으나 어의간의 차이는 엄밀하게 구별된다. 반면에 온톨로지에서는 개념간의 차이는 엄밀하게 구별되어야 하나 어의간의 차이는 구별되어야 할 필요가 없다.

온톨로지 구축에 활용하기 위한 기초자료로 볼 때 사전의 동형어의어의 의미구별만으로는 정확한 의미구별이라고 할 수 없는 것은 당연한 것이다. 이 논문에서는 동형어의어의 의미구별에 기반하고 있는 기존의 온톨로지구축 연구에 대해 비판하고 기존 연구의 문제에 대한 대안으로 사전정의문의 동형어의어 구별뿐만 아니라 어의구별과 개념구별까지도 온톨로지 구축을 위한 기초자료로 활용할 것을 제안한다.

1) 이 논문은 국어의 일반명사 정의문에 한정하여 사전의 정의문을 언급할 것이다.

2절에서 언어사전의 정의문(뜻풀이)의 일반적 속성을 알아보고 3절에서는 기존 국어사전의 정의문을 분석해보고 일반적인 사전의 정의문의 문제점을 지적하고 그 해결방안을 제시하고자 한다. 4절에서는 온톨로지구축을 위한 국어사전 정의문의 문제점과 그 해결방안을 모색해보고 5절에서는 온톨로지구축을 위해 국어사전의 어의분류와 온톨로지의 개념구조와의 관계를 연결하기 위한 제안을 해 본다.

## 2. 사전적 정의와 온톨로지<sup>2)</sup>

사전적 정의란 어떤 단어가 사회에서 폭넓게 통용된 이후 그 사회의 구성원들이 약속으로 받아들이고 있는 의미를 정의한 것이다(김광혜, 1998). 사전적인 정의는 단어라고 하는 언어의 기호에 관한 것이다. 사전적 정의를 통하여 기호에 대해서만 설명할 수 있을 뿐이고 그 기호가 가리키는 대상에 대해서는 설명할 수 없다. <감>이라는 단어가 있다면 사전에는 <감>이라는 기호의 의미만을 기술할 뿐이지 실제 세계 속에서의 <감> 그 자체에 대해 기술하는 것은 아니다. 실제 세계 속에서의 <감> 그 자체에 대해 기술하는 것은 언어사전적 정의 보다는 백과사전적 정의에 가깝다. 언어사전은 낱말의 뜻풀이이고 백과사전은 사물, 사실에 대한 지식을 전달하기 위한 도구이다.

기호에 대한 정의와 대상에 대한 정의는 구별하지 않으면 안 된다. 단어의 뜻풀이는 기호에 대한 정의이므로 모든 사전의 정의가 형식적으로 갖는 표제어 A와 그 뜻풀이 B의 관계는 <A는 B이다>의 관계가 아니라 <A라는 말은 B를 의미 한다>로 바꾸어 표현하는 것이 옳다.

정의란 근본적으로 언어를 다른 언어로 설명하는 행위이다. 사전의 정의문은 같은 언어의 다른 어휘나 표현으로 동일한 의미를 갖도록 말 바꾸기 한 것이다. 사전의 정의문의 개념은 개념 그 자체가 아니라 특정언어에 반영된 개념이다. 그리고 개념망이라고도 할 수 있는 온톨로지는 다양한 경로를 통한 인간의 세상지식이 반영된 개념의 연결망이다. 그러므로 사전의 정의문을 기반으로 온톨로지 구축을 위해서는 언어적인 혹은 언어특이적인 개념의 부분을 걸러내지 않으면 안 된다.<sup>3)</sup>

또한 언어사전은 논리적, 철학적 순수성을 이상으로 삼지 않는다(이상섭, 1997). 논리적 정의는 학문적 이론적인 목적으로 사용된 정의이지만, 언어사전적 정의는 사전의 독자인 일반인을 위한 정의라고 할 수 있다. 그러므로 사전의 정의를 논리적인 정의와 백과사전적 정의와 구별하여야 할 것이다.

### (1) 논리적 정의 >> 언어사전적 정의 >> 백과사전적 정의

논리적 정의는 학문적 이론적인 목적으로 사용된 정의이고, 언어사전적 정의는 일반인의 언어 사용을 목적으로 하는 정의이고, 백과사전적 정의는 다양하고 세부적인 세상지식을 나타내기 위한 정의이다.

2) 국어사전의 경우 정의문이란 용어 보다는 뜻풀이란 용어를 사용하는 것이 더 적절한 것으로 보인다. 많은 국어학자들과 사전편찬자들 사이에서도 정의문이라는 용어 보다는 뜻풀이란 용어를 널리 사용하고 있음을 언급해 둔다.

3) 동의어, 반의어 관계는 언어적 어휘적 관계(lexical relation)이지 개념적 관계(conceptual relation)가 아니다(Fellbaum, 1999).

이 논문은 단어의 개념과 일반 개념은 다르다고 전제할 것이다. 3절에서 언급할 개념 중심어체계는 단어의 개념망이고, 온톨로지는 일반 개념망이다. 그러므로 언어사전에 기반을 둔 개념중심어체계와 언어화자의 공통적 세상지식에 기반을 둔 온톨로지를 구별하여야 하는 것은 당연하다.

### 3. 기존 국어사전의 정의문의 문제와 그 해결책으로서의 중심어의미체계

이 절에서는 온톨로지 구축을 목적으로 국어사전의 정의문을 활용하기 위해 우선 온톨로지와 독립적으로 기존 국어사전의 정의문을 분석한다. 정의문 중에서도 특히 중심어에 초점을 맞추어 정의문을 분석할 것이다. 이는 4절에서 언급될 정의문 중심어의미를 최대한 활용한 온톨로지 구축의 기본적인 토대를 마련하기 위한 것이다.

김광해(1998: 5-6)는 “우리나라의 국어사전이 안고 있는 중요한 문제점들이 수없이 많지만, 그 중의 하나는 사전의 뜻풀이가 불완전하다는 것이다. [...] 이러한 문제점은 최근에 출간된 최신 국어사전인 <<연세한국어사전>>(1998)에서도 그대로 이어지고 있다.”고 지적하고 있다.

이 논문은 온톨로지를 구축하는 것이 근본 의도이기는 하나 국어사전의 어의 구별과 정의문 기술 방식을 향상시키기 위한 제안을 함으로써 국어사전의 정의문 기술 발전의 틀을 마련하는 데에 기여하고자 한다.

사전의 정의문의 중심어들을 살펴보면 사전 정의문의 체계를 전반적으로 확인 분석할 수 있기 때문에 사전의 총체적인 뜻풀이의 문제점이 드러난다. 또한 서로 다른 사전들의 정의문을 중심어를 통해 비교해 봄으로써 정의문의 전반적인 차이를 쉽게 비교 할 수도 있다.

비교의 기준을 마련하기 위하여 우선 아래의 그림에서와 같이 하나의 표제어 L의 의미를 포함하는 개념범주들 A, B, C가 있고 각각 A의 의미는 B에 포함되고, B의 의미는 C에 포함된다고 할 경우, L의 정의문의 중심어로서는 C보다는 B가, B보다는 A가 적절하다. 그 이유는 L의 정의문의 중심어 개념범주의 범위가 적을 수록 L이 속한 상위의 개념들 속에서 표제어 L의 의미를 같은 상위범주에 공통적으로 속한 다른 개념들로부터 분리하기 위한 변별자로서의 속성 기술이 줄어들 것이기 때문이다.

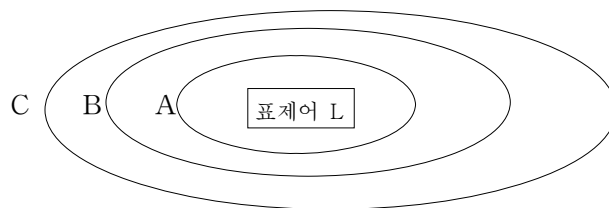


그림-1 중심어 설정을 위한 표제어 L의 의미를 포함하는 상위개념의 상대적 범위

이 후의 하위절에서 언급될 상향식(bottom-up 방식)으로 만든 위계구조를 이러한 기준에 비추어 사전의 전체 뜻풀이 구조를 확인 평가해 볼 것이다.

### 3.1. 기존 국어사전들 간의 중심어체계의 차이

기존 국어사전들의 중심어를 범주포함관계로 표현하여 앞에서 제시한 기준에 비추어 평가하기 위한 목적으로 사전의 정의문들의 중심어를 연결하여 상하위관계망을 구축하는 절차는 다음과 같다.

첫째, 임의로 하나의 단어를, 예를 들어 <여자>를, 선정한다.

둘째, 특정한 사전에서, 예를 들어 [연세]사전에서, 그 단어의 정의문을 찾는다.

(2) 표제어 <여자>의 정의문<sup>4)5)</sup>

[연세] 여자(女子) 성이 여성인 사람을 일반적으로 가리키는 말.

정의문 중 밑줄 친 부분과 같은 정의문의 중심어를 찾는다. 명사 표제어의 중심어는 명사이고, 동사 표제어의 중심어는 동사이고, 형용사 표제어의 중심어는 형용사이고, 부사 표제어의 중심어는 부사이다. 이것은 명사의 뜻풀이는 명사구로 표현하고, 동사의 뜻풀이는 동사구로 표현하고, 형용사의 뜻풀이는 형용사구로 표현하고, 부사의 뜻풀이는 부사구로 표현한다는 말과 일치하는 것이다.

셋째, 표제어 <여자>의 중심어인 <사람>를 다시 같은 사전을 이용하여 이를 표제어로 찾아 그 표제어의 중심어를 꼬리에 꼬리를 무는 방식으로 계속 찾아 기록한 후에 표제어와 중심어들을 서로 연결하여 본다.

(3) 중심어를 다시 표제어로 반복하여 찾아 기록함 [연세]

사람 1. 모든 생물 중에서 가장 지능이 높고 언어와 도구를 사용할 수 있으며 문화를 창조하고 생각하는 능력이 있는 존재. 2. 일정한 수준의 인격을 갖춘 존재."

존재(存在) 1. 실제로 있는 것. 2. 실제로 있는 사물이나 사람.

동물(動物) 1. 짐승, 새, 벌레, 물고기 따위의 생물. 2. [사람을 다른 동물과 구별하여] 사람 이외의 동물.

생물(生物) 생명이 있는 동물과 식물.

물체(物體) 구체적인 꼴을 가지고 존재하는 것.

표제어 <존재>의 경우 뜻풀이가 단순히 한자말을 푸는 수준인 것은 이상섭(1997)도 문제가 있음을 지적하고 있듯이 뜻풀이라고 볼 수 없다. 표제어 <생물>의 경우 중심어가 다른 정의문들은 상위어인 것과는 달리 하위어로 구성되어 있다. 이는 정의문 기술의 비밀관성의 한 요인이 된다. 표제어 <물체>의 경우 중심어를 <것>이 아니라 <존재>로 보아야

4) 이 후 <<연세 한국어 사전>>은 [연세]로, <<표준 국어 대사전>>은 [표준]으로, <<우리말 큰 사전>>은 [큰사전]으로 간략히 표기한다.

5) 이 후 모든 사전의 내용은 논의의 편의를 위하여 사전의 내용 중 관련된 부분만을 선별하여 적고 관련되지 않은 부분은 생략한다.

한다. <것>이라는 단어는 명사의미체계의 하나의 범주로 보기 힘들기 때문이다.<sup>6)</sup>

이러한 [연세]사전의 표제어-중심어 관계를 연결해 보면 다음과 같이 연결이 가능해진다.

(4) 존재(存在) > 물체(物體) > 생물(生物) > 동물(動物) > 사람 > 여자(女子)

이러한 결과에 비추어 볼 때, [연세]사전에서 표제어 <사람>의 경우 중심어가 지나치게 상위어가 설정되었다. 이를 바로 잡을 필요가 있다.

넷째, 한 사전의 결과를 다른 사전과 비교해 보기 위해 또 다른 사전으로도, 예를 들어 [표준]으로, 위의 절차를 반복하여 수행한다.

(5) [표준]의 <여자>의 뜻풀이와 그 중심어의 표제어들<sup>7)</sup>

여자02 (女子) 여성으로 태어난 사람.

사람 「1」 생각을 하고 언어를 사용하며, 도구를 만들어 쓰고 사회를 이루어 사는 동물. 「2」 어떤 지역이나 시기에 태어나거나 살고 있거나 살았던 자.

존재 (存在) 「1」 현실에 실제로 있음. 또는 그런 대상. 「2」 다른 사람의 주목을 끌 만한 두드러진 품위나 처지. 또는 그런 대상.

동물 (動物) 「1」 『생』 생물계의 두 갈래 가운데 하나. 현재 100만~120만 종이 알려져 있고 그 가운데 약 80%는 곤충이 차지한다. 원생동물부터 척추동물까지 23개 문(門)으로 분류된다. 주로 유기물을 영양분으로 섭취하며, 운동, 감각, 신경 따위의 기능이 발달하였다. 소화, 배설, 호흡, 순환, 생식 따위의 기관이 분화되어 있다. 「2」 사람을 제외한 길짐승, 날짐승, 물짐승 따위를 통틀어 이르는 말. 「참」 식물 02.

생물01 (生物) 생명을 가지고 스스로 생활 현상을 유지하여 나가는 물체. 영양·운동·생장·증식을 하며, 동물·식물·미생물로 나뉜다. ≡유생물. 「참」 무생물.

물체 (物體) 「명」 「1」 구체적인 형태를 가지고 있는 것. 「2」 물건의 형체. 「3」 『철』 정신이나 의식이 없는 유형물. 삼차원적인 공간을 차지하고 있는 형성물로서, 공간적 부피를 가진 것을 이른다.

표제어 <존재>의 경우 중심어를 동사의 전성명사어 <있음>으로 대신하였다. 이 또한 명사의 정의문의 중심어로서 자격을 충분히 갖추지는 못 한다. <또는 그런 대상>의 경우도 마찬가지로 중심어로서 자격이 없다. 표제어 <동물>의 경우 중심어로 제시되고 있는 <하나>는 중심어라고 할 수 없다. 이 경우 중심어 위치에 있지 않고 다른 위치에 있는 <생물>이 상위개념 즉 중심개념이다. 또한 이탤릭체로 표시된 부분은 백과사전적인 정의이다. 백과사전적인 정의를 줄이거나 없애고 언어사전적인 정의를 수정 보완하여야 했다. 이상섭(1997)이 지적하듯이 국어사전에서는 언어사전적 정의가 백과사전적 정의에 우선하여야 하고

6) 확대 중심어가 <N하는 것>일 경우 N을 개념중심어로 설정한다.

7) 이탤릭체로 표시된 부분은 백과사전적인 정의이다.

지나친 백과사전적 정의는 언어사전의 정의로서 부적합하다. 표제어 <물체>의 경우 <것>을 중심으로하고 있으니 이 또한 특정한 범주로 볼 수 없기 때문에 명사의 정의문의 중심으로 자격을 갖추지 못 한다. 이러한 중심어로는 명사의미체계를 구축할 수 없다.<sup>8)</sup>

이러한 [표준]사전의 표제어 중심어관계를 연결하여 이를 [연세]사전의 중심어관계를 비교해 볼 수 있다.

(6) [연세]와 [표준]의 표제어-중심어 관계 비교

[연세] 존재(存在) > 물체(物體) > 생물(生物) > 동물(動物) > 사람 > 여자(女子)  
 > 식물  
 > 무생물  
 [표준] 물체(物體) > 생물(生物) > 동물(動物) > 사람 > 여자(女子)  
 > 식물  
 > 미생물

[표준]의 경우 <물체>와 <존재>가 연결되어 있지 않은 반면에 [연세]의 경우 <물체>와 <존재>가 연결되어 있는 것이 서로가 다르다는 것을 둘의 비교를 통해 알 수 있다. [표준]의 경우 <물체>와 <존재>가 연결되어 있지 않는 것은 한편 [표준]의 <물체>의 뜻풀이에 문제가 있음을 시사한다. 또한 우리는 이러한 비교를 통해서 사전의 정의문은 중위나 하위 보다는 상위로 올라 갈수록 차이가 난다는 것을 알 수 있다.

다섯째, 두 개 이상의 기존의 사전을 사용하여 주어진 표제어-중심어 관계를 모두 연결하여 최종적인 명사의미체계 만든다.

(7)

> 무생물  
 존재(存在) > 물체(物體) > 생물(生物) > 동물(動物) > 사람 > 여자(女子)  
 > 식물

이러한 명사의미체계를 활용하여 우리는 사전 정의문의 결함을 보완할 수 있다. 예를 들어 [연세]의 <사람>의 중심어는 지나치게 상위어이기 때문에 중심어를 [표준]에서와 같이 <동물>로 제시하여 풀이하는 것이 보다 적절하다. 중심어로서 자격이 없는 모든 어휘는 중심어로서 합당한 범주로 수정하여야 한다. 다행히도 그 해당 또한 기존의 국어사전들 안에 있다. 이를 합리적으로 체계적으로 일관성 있게 처리하는 것이 급선무이다. 우리는 이러한 절차를 통하여 국어사전의 모든 어휘의 뜻풀이를 재정립할 필요가 있다.

3.2. 기존 국어사전의 중심어 범주의 차이

앞에서 언급하였듯이 중심어란 보통 사전의 정의문의 가장 나중에 표현되는 단어(어절)로서 하나의 의미범주(혹은 개념범주)를 나타낸다. 이 범주는 범주체계를 구축하는데 결정적인 도

8) 중심어로서 <것>과 같은 단어의 문제에 대하여 3.3.에서 다시 언급할 것이다.



이 경우에는 [큰사전]의 <물>과 [연세]의 <액체>는 그 범위를 비교 할 수 있으나, 이에 비하여 [표준]의 <분비물>은 다른 사전의 중심어와 비교하기 힘들다. [표준]과 [큰사전]의 정의에는 이탤릭체로 표시된 부분과 같이 다른 사전에서는 찾아볼 수 없는 백과사전적 정보가 부가되어 있다는 것이 사전의 정의로는 부적합하다. 백과사전적 정보가 지나치게 열거되어 있지 않다는 면에서 보면 [연세]사전의 뜻풀이가 적절함을 알 수 있다.

아래의 표는 [연세], [표준], [큰사전] 세 사전의 중심어를 비교하여 보여주고 있다.

	[연세]	[표준]	[큰사전]
감	과일	열매	열매
눈물	액체	분비물	물
꽃	식물의 부분	번식 기관	생식기관
경찰(警察)	임무, 국가 기관	일, 조직	행정
공	물건	운동 기구	운동기구
교수(教授)	직위, 교원	사람	선생
대통령	사람. 또는 직위	원수	최고 지도자
개	집짐승	포유동물	집짐승
글자	기호, 또는 체계.	체계의 부호	부호
돈	주화나 지폐	물건	물건
도	이름	계이름	음
담배	물건	기호품	따위를 통틀어 일컬음
고개	목을 중심으로 한 머리 전체	목의 뒷등이 되는 부분	목의 뒷등 부분
건물(建物)	구조물	집	집 따위의 물건
가치(價値)	의의나 중요성	쓸모	중요성이나 쓸모
논	밭	땅	땅
국가	집단	사회집단	사회
나라	조직	사회집단	사회

표-1 [연세], [표준], [큰사전] 정의문의 중심어 비교

이와 같이 세 사전의 중심어를 서로 비교해 봄으로써 우리는 사전들의 뜻풀이 방식이 전반적으로 사전마다 현저하게 다르다고 같은 사전 내에서는 일관성이 결여되어 있다는 것을 확인할 수 있다.

### 3.3. <것>/<일>-중심어류 처리의 문제

기존의 사전의 정의문의 중심어에는 그 의미가 분명치 않은 명사를 사용하는 경우가 있다 이는 의미 범주로 적합하지 않으니 수정하지 않으면 안 된다. 중심어가 <때>, <곳>, <것>, <일> 등과 같이 순수명사가 아닌 경우에 이를 <때>는 <시간>, <곳>은 <장소> 등으로 각각 바꾸어 볼 수 있으나 <것>과 <일>의 경우는 마땅히 바꾸어 볼 수 있는 범주가 없다.

많은 연구자들이 주장하듯이 <것>의 경우에는 중심어를 확대하여 <N하는 것>으로 보고 이 경우 <N>을 중심어로 설정하는 것은 적절하다. 기존의 연구자들 사이에 합의가 이루어지지 않은 <일>의 경우에는 중심어를 확대하여 <X하는 일>로 보고 이 경우 전성명사형 <X함>을 중심어로 설정할 것은 제안한다.

<것>의 경우 예를 들어 표제어 <물질>을 찾아보면 아래와 같이 중심어를 <존재하는 것>과 <가지고 있는 것>으로 확대 할 수 있을 것이다. [연세]의 <존재하는 것>의 경우 <존재>를 개념중심어로 보아야 할 것이다.



- (12) [연세] 물체(物體) 구체적인 꼴을 가지고 존재하는 것.  
 [표준] 물체(物體) 구체적인 형태를 가지고 있는 것.

그러나 [표준]의 <가지고 있는 것>의 경우 개념중심어로 볼 수 있는 범주가 없다. 그렇다고 이를 더 확대하여 <형태>를 개념중심어로 볼 수는 없는 일이니 이 경우 중심어를 찾는 데 어려움이 있다. 이것으로 보아 [표준]은 표제어 <물체>의 개념중심어를 표현하지 않은 것이다. 이와 같이 사전에는 그 개념중심어를 표현하지 않은 것으로 보아야 하는 뜻풀이들이 나타난다.

참고로, 조평옥 외(1999)는 다음과 같은 개념중심어로 보기 힘든 형식적 중심어를 다음과 같이 제시하고 있다.<sup>11)</sup>

말	(3650)	준말	(3183)	하나	(2876)	일	(2797)
총칭	(1150)	곳	( 943)	것	( 808)	가지	( 624)
뜻	( 369)	없음	( 202)	구용어	( 153)	아니함	( 150)
있음	( 125)	일컬음	( 87)	속칭	( 70)	못함	( 53)
따위	( 47)	무렵	( 47)	쌈	( 43)	같음	( 38)

표-2 조평옥 외(1999)의 금성사전 MRD를 분석한 개념중심어로 보기 힘든 형식적 중심어

<것>의 예를 더 찾아보고 <일>의 예도 몇 가지 예를 더 살펴보기로 하자. 표제어 <느낌>, <글>, <가정>의 세 사전의 뜻풀이는 아래와 같다.

- (13)

[연세] 느낌 느껴지는 것. 느낀 것.  
 [표준] 느낌 몸의 감각이나 마음으로 느끼는 기운이나 감정.  
 [큰사전] 느낌 느끼는 것.

[연세] 글 생각이나 말을 글자로 나타낸 것.  
 [표준] 글 「어떤 생각이나 일 따위의 내용을 글자로 나타낸 기록.  
 [큰사전] 글<sup>1</sup> 말을 글자로서 나타낸 적발.<sup>12)</sup>

[연세] 가정<sup>3</sup>(家政) 집안 살림을 꾸려 가는 일.  
 [표준] 가정<sup>05</sup>(家政) 집안을 다스리는 일.  
 [큰사전] 가정<sup>2</sup>[家政] 집안 살림을 꾸려가는 일.

<느낌>과 <글>의 경우는 한 사전의 문제를 다른 사전을 참고로 하여 해결할 수 있으나 <가정>의 경우는 모든 사전의 뜻풀이가 동일하게 범주를 설정하기 어렵게 기술되어

11) 일부 수정하여 제시한다. 괄호 안의 숫자는 사전에서 해당 단어가 중심어로 나타나는 표제어의 수이다.

12) [큰사전] 적-발<sup>1</sup>: 적바림하여 놓은 글발. 적-바림: 간단히 적어 두는 일. 또는, 간단히 적어 놓은 것.

있다.

기존의 사전에는 동사의 전성명사형을 중심으로 하고 있는 표제어도 있다. 개념중심 어로서 적합하다고 할 수는 없다. 그 예로 표제어 <노력>을 살펴보기로 하자.

(14)

[연세] 노력<sup>1</sup>(努力) [좋은 일을 이루려고] 힘을 들이고 애를 쓰는 것.

[표준] 노력01 (努力) 목적을 이루기 위하여 몸과 마음을 다하여 애를 쓰.

[큰사전] 노력<sup>1</sup>[努力] 무엇을 이루려고 애를 쓰고 힘을 들임.

이 경우에는 <것>의 사용보다는 동사의 전성명사형을 사용하는 것이 좀 더 나아 보이나 의미범주체계의 관점에서 볼 때 명확한 범주가 아니기 때문에 여전히 문제가 있는 것이 사실이다. 그러나 사전성 혹은 서술성 명사의 경우에는 이와 같이 동사의 전성명사형이나 형용사의 전성명사형의 사용을 허용하여야 할 것이다.<sup>13)</sup> 개념범주로 전성명사형을 허용한다면 이 경우 <쓰> 보다는 <애를 쓰> 혹은 <애쓰>으로 <들임> 보다는 <힘을 들임> 혹은 <힘들임>으로 범주를 표현하는 것이 바람직하다. 긴 표현 보다는 짧은 표현이 범주표현으로 적합하나 각각 <애쓰>과 <힘쓰>을 사용하고 또한 어의 면에서 뿐만 아니라 개념적으로도 둘을 구별할 필요가 없으므로 그 의미가 더 쉽게 이해되는 <힘쓰>으로 통일하여 사용하기로 정할 수 있을 것이다.<sup>14)</sup> 위의 <가정>의 경우 <꾸려가는 일>은 <꾸림>으로, <다스리는 일>은 <다스림>으로 중심어를 정할 수 있을 것이다.

(15) 힘쓰<sup>15)</sup> > 노력

참고로, 조평옥 외(1999)는 금성사전 MRD를 분석하여 전성명사형 중심어를 다음과 같이 제시하고 있다.

없음 (230)	있음 (125)	됨 (61)	쓰 (43)
같음 (38)	흐름 (33)	어려움 (27)	아름다움 (26)
많음 (23)			

표-3 조평옥 외(1999)의 금성사전 MRD를 분석한 전성명사형 중심어

세 사전에서 나타나는 그 외의 다른 표제어들과 그 중심어를 제시해 본다.

13) 전성명사형이 <-로>만 있는 것이 아니고 다양한 전성명사어미의 기능이나 의미를 고려해서 결정해야 하지만 이 논문에서는 이에 대한 논의를 미루고 사전에서 일반적으로 사용하는 로-전성명사형을 사용하기로 한다.

14) [연세] 애2: [‘창자’를 뜻하는 옛말로] ‘애가 달다/타다/터지다’나 ‘애를 먹다/쓰다/태우다’와 같은 관용 표현에 쓰임. 애를 쓰다: 마음과 힘을 다하여 힘쓰다.

15) [연세] 힘쓰다: [I] (무슨 일에) 힘을 다하고 노력하다. [II] 1. 기운을 세게 발휘하다. 2. (어떤 일을 하는 데에) 폭력이나 권력 등을 사용하다.

	[연세]	[표준]	[큰사전]
느낌	것	기운이나 감정	것
걸음	것	동작	동작
글	것	기록	적발
기회(機會)	때	시기나 경우	고비 <sup>16)</sup>
노력(努力)	것	쓸	힘
가정(家政)	일	일	일
눈치	힘	것	힘
과거(過去)	때, 날	때	때
공장(工場)	곳	곳	곳

표-4 [연세], [표준], [큰사전] 정의문의 <것>/<일>-중심어류 비교

이와 같은 경우에도 앞에서 제시한 방법대로 중심어 확대를 통해서 개념중심어를 찾는 과정을 거쳐야 할 것이다.

여기에서 언급하고 갈 것은 <가지>의 경우이다. 이때 <가지>가 어떤 의미인지는 언급하고 있지 않다. <가지>에는 여러 가지 의미가 있기 때문이다. 동형어의어로서의 구별과 어의로서의 구별과 개념으로서의 구별 모두를 알 수 없기 때문이다.<sup>17)</sup>

#### 4. 온톨로지구축을 위한 국어사전 정의문 활용의 문제와 해결방안

이 절에서 우리는 기존의 국어사전의 정의문에 일관성이 결여되어 있음을 살펴보고 그러한 정의문의 비일관성의 몇 가지 요인에 대하여 살펴보았다. 이러한 기존 사전의 비일관성은 특히 표제어-중심어관계를 통해 비교적 간단히 드러난다. 그러나 표제어-중심어관계를 통해 여러 사전을 비교함을 통하여 그리고 서로를 보완함으로써 기존의 정의문의 수정가능성 또한 기존 사전에서 볼 수 있었다는 것은 긍정적이라고 할 수 있다.

사전의 뜻풀이 편찬의 비일관성으로 인해 사전의 뜻풀이만으로는 의미계층체계를 만들 수 없다. 의미계층체제로 사전의 뜻풀이를 수정보완하지 않으면 안 된다. 이를 위해서 김광해(1998, 유의어비교를 통한 정의내리기), 이기동(1989) 등은 그 방안을 제시하고 있다. 그러나 정의문의 일관성을 위해 전체적인 틀을 잡아주는 방법으로 개념기반 중심어의미체계를 활용하는 것이 최우선의 해결책이다.

이 논문에서는 처음 단어의 정의문의 중심어를 근거하여 명사의 의미계층체계를 구축하려하였고 이는 조평옥 외(1999)와 Ide & Véronis(1995)는 국어사전의 정의문의 중심어를 표제어의 상위어로 하여 상향식(bottom-up 방식)으로 구축된 명사의미체계의 문제점을 다음과 같이 지적하고 있다.

첫째, 사전편찬자의 비일관적인 정의문의 기술과 사전 마다 다른 정의문으로 인해, 구축된 국어명사의미체계가 일관성을 지니지 못한다.

둘째, 정의문의 순환성으로 인해 상위개념으로서의 중심어가 명확하지 못한 명사들이 상당수 있다.

셋째, 사전의 정의문만을 이용하여 구축하였기 때문에 서로 다른 계층으로 분류되어

16) [큰사전] 고비<sup>1)</sup>: 일이 되어 가는 데의 가장 요긴한 단계나 대목.

17) 5절에서 사전 뜻풀이의 세부 구별에 대해 논의한다.

야 할 명사들이 같은 계층으로 분류되거나, 같은 계층으로 분류되어야 할 계층의 명사들이 다른 계층으로 분류되는 경우가 있다.

이러한 사전정의문의 문제점에 대한 해결책으로, 첫째, 형식적 중심어체계 대신에 개념적 중심어 의미체계를 활용한 정의문을 구축해야 한다는 것이다. 이를 위해서는 사전뜻풀이편찬자들은 우선 중심어 의미체계를 구축하는 것이 우선 되어야 한다. 중심어 의미체계 없이 정의문을 기술하는 것은 일관성이 없는 정의문을 구축하게 되는 가장 근본적인 원인이 되기 때문이다.

둘째, 이러한 체계에 근거하여 어의 구별에 필요한 최소한의 속성을 사용하여 정의문이 구축되어야 한다는 것이다. 이러한 속성기술 간소화는 중심어 의미체계에 기반 한다면 해결 될 수 있는 이차적인 것이다. 중심어 의미체계에 충실하여야 이 단계의 실현이 가능하다.

셋째, 앞에 언급한 두 단계를 충실히 이행한 후에야 비로써 유의어비교를 통한 혹은 Pustejovsky(1995)의 의미속성구조를 통한 정의문의 구축 등도 의미 있는 것이다.<sup>18)</sup> 그러나 위의 유의어비교나 의미속성구조등이 지나치게 많은 속성을 언급한다는 데에는 문제가 있다. 백과사전식정의가 아닌 언어사전식 정의를 하는 사전의 정의문에서 지나치게 많은 속성을 언급하는 것은 언어사전적의 정의로서의 적합성을 떨어뜨리는 역할을 할 것이기 때문이다.

사전의 정의문을 그대로 온톨로지구축에 활용할 수는 없다. 정의문의 중심어를 개념범주가 되도록 수정하고 이를 전체적으로 조절하여야 할 것이다. 3절에서는 형식적 중심어의 문제점에 대하여 지적하고 그 해결방안으로 문제가 되는 형식적 중심어 대신에 중심어 확대를 통한 개념중심어를 찾기를 제안하고 그 실례를 제시하였다. 사전 정의문의 모든 형식적 중심어만을 열거하는 것만으로는 의미가 없고 실질적인 중심어 즉 개념중심어를 찾아야 하고 이 중심어가 개념범주로서 합당하가를 판단하여야 할 것이다.

지금까지 언급한 순수한 중심어체계를 연결해 보면 조평옥 외(1999)에서 제시하고 있는 것과 같은 표-5의 관계망을 구축 할 수 있다.<sup>19)</sup>

가곡a	< 노래a	< 말e	< 소리a	< 청각a	< 감각a	< 느낌	< 감정	< 정신a
가마d	< 탈것a	< 도구a	< 기구a	< 물건a	< 무생물체a	< 물체a	< 사물b	
가마니a	< 자루a	< 주머니a	< 물건a	< 무생물체a	< 물체a	< 사물b		
가뭇a	< 날씨a	< 상태a	< 형편a	< 모양a				
가정부a	< 여자a	< 사람a	< 생물체a	< 물체a	< 사물b			
가죽a	< 짚질a	< 켄a	< 부분a	< 것a				
가지a	< 줄기a	< 식물부분a	< 부분a	< 것a				
간판a	< 물건a	< 무생물체a	< 물체a	< 사물b				
간호사a	< 사람a	< 생물체a	< 물체a	< 사물a				
갈대a	< 여러살이해풀a	< 풀a	< 식물a	< 생물체a	< 물체a	< 사물b		
감a	< 음식a	< 물건a	< 무생물체a	< 물체a	< 사물b			

표-5 정의문의 중심어로 구축한 명사의미체계

또한 사전을 통해 중심어 의미체계를 구축해 보면 최상위 노드가 하나만 나오는 것이

18) 유의어 비교를 통한 정의문 구축을 제안하는 김광해(1998)와 Pustejovsky의 의미속성구조를 통한 정의문 구축을 제안하는 이해윤(1997)을 참고 할 것.

19) 조평옥 외(1999; 589). 일부 수정함.

아니다. 조평옥 외(1999)는 **사물, 공간, 관계, 모양, 것, 기호, 동안, 때, 방법, 양, 힘, 정신, 조건**이라는 13개의 최상위 노드가 만들어 진다고 한다. 그러므로 이를 하나로 수렴하기 위해서는 사전 이외의 개념체계가 필요할 것이고 이는 사전의 정의문체계를 수정하는 데도 활용될 수 있을 것이다. 이러한 해결책은 상위 온톨로지에서 찾을 수 있다.

일반적으로 온톨로지는 상위온톨로지와, 중위온톨로지, 하위온톨로지로 구분한다. 상위온톨로지는 대단히 추상적인 개념들로 구성되어 철학자들에 의해, 중위온톨로지는 일반적 세계의 개념들로 구성되어 인지과학자들과 언어학자들에 의해, 하위온톨로지는 전문분야의 개념들로서 특정분야의 전문가들에 의해서 구축되는 것이 바람직한 것으로 알려져 있다.<sup>20)</sup>

사전의 정의문을 통해 온톨로지를 구축하고, 온톨로지를 사전의 정의문 기술의 틀로 활용한다는 것이 얼핏 순환적인 것으로 느껴질 수 있으나 이 둘의 관계를 상호보완적 관계로 이해하는 것이 옳을 것이다. 사실 온톨로지는 언어외적인 것이고 사전의 정의문은 언어내적인 것이다. 2절에서도 언급하였듯이 사전의 단어가 갖고 있는 개념은 엄밀히 말해서 순수한 개념 그 자체라기 보다는 언어에 반영된 개념 언어가 갖고 있는 개념인 것이다.

## 5. 국어사전의 어의분류와 온톨로지구축을 위한 제안

이 절에서는 기존 사전의 어의 분류의 문제점을 지적하고 잠정적인 해결 방안을 제시하고자 한다. 또한 사전의 언어사전이 고유하게 갖고 있는 분류방식과 온톨로지의 분류방식을 각각 유지하면서 어떻게 이들을 연결하는 것이 합리적인가를 논의하고자 한다. 언어사전과 온톨로지와의 관계가 먼저 정립되어야 그 후에 비로소 이를 연결할 방법을 결정할 수 있을 것이다.

5.1에서는 사전의 표제어의 어의분류와 개념분류의 방식에 대해 논하고 5.2에서는 언어사전의 분류방식과 온톨로지의 분류방식을 각각 유지하면서 어떻게 이들을 연결할 수 있는가를 논의할 것이다.

### 5.1. 국어사전 어의분류의 문제점

기존의 사전들을 살펴보면 사전들마다 개념들을 어의로 묶는 방식이 다르고, 어의를 하나의 표제어로 묶는 방식이 다르고, 동형이의어와 다의어를 구별하는 방식이 다르다는 것을 알 수 있다.

사전들 마다, 하나의 같은 표제어가 갖는 어의가 개수가 다르고, 하나의 같은 어의가 갖는 개념의 수도 다르다. 이는 분류의 기준이 명확하지 않기 때문이다. 이를 검증하기 위한 여러 가지 방법들이 제안되어 왔으나 사실상 확고한 검증방법은 아직 존재하지 않는다. 다만 여러 가지 방법들을 사용해서 가장 적절한 결과를 선택할 수 있을 뿐이다.

일반적으로 사전의 구성단위를 표제어(lexeme), 동형이의어(homonym), 어의(sense), 개념(concept)으로 구별할 수 있다. 이들을 각각 구별하여 그 순서대로 나열 하면 다음과 같다.

(16) 표제어 구별 > 동형이의어 구별 > 어의 구별 > 개념 구별

---

20) Hovy(2005) 참조.

이러한 구별을 기준으로 임의의 사전 A와 또 다른 임의의 사전 B의 관계를 분류하여 다음과 같은 표로 나타낼 수 있다.

	표제어	동형이의어	어의	개념
첨가	표제어 첨가	동형이의어 첨가	어의 첨가	개념 첨가
삭제	표제어 삭제	동형이의어 삭제	어의 삭제	개념 삭제
통합	표제어 통합	동형이의어 통합	어의 통합	개념 통합
분리	표제어 분리	동형이의어 분리	어의 분리	개념 분리

표-6 사전A와 사전B의 관계분류표

그 각각의 내용을 어의와 개념의 차원에서만 예를 들며 설명해 보기로 한다.

● 어의 첨가/어의 삭제<sup>21)</sup>

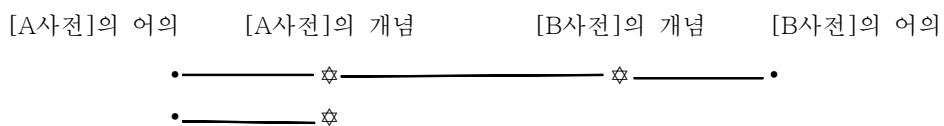
한 사전인 [A사전]에는 없는 어의가 다른 사전인 [B사전]에 새로운 어의가 나타날 경우를 어의 첨가의 경우라고 한다. 물론, 똑 같은 경우를 [B사전]의 입장에서 보면 어의 삭제가 된다. 이 경우 개념의 개수도 변해야 (증가하여야) 한다.

(17) 어의 첨가<sup>22)</sup>



한 사전인 [A사전]에는 있는 어의가 다른 사전인 [B사전]에는 어의로 나타나지 않을 경우를 어의 삭제의 경우라고 한다. 물론, 똑 같은 경우를 [B사전]의 입장에서 보면 어의 첨가된다. 이 경우 개념의 개수도 변해야 (감소하여야) 한다.

(18) 어의 삭제



21) 편의상 첨가와 삭제에 대해 하나의 예만을 제시할 것이다. A사전에서 B사전으로의 첨가는 B사전에서 A사전으로의 삭제로 볼 수 있기 때문이다.

22) 아래에서 •는 어의를 표시하고, ☆는 개념을 표시한다.

(19)

[연세]가정2(假定) (사실이 아니거나, 사실인지 아닌지 분명하지 않은 것을) 어떤 논리를 펴 나가기 위하여 사실인 것처럼 받아들이는 것.

[표준]가정07 (假定) 「1」 사실이 아니거나 또는 사실인지 아닌지 분명하지 않은 것을 임시로 인정함. 「2」 「논」 결론에 앞서 논리의 근거로 어떤 조건을 내세움. 또는 그 조건이나 전제. 「3」 「수」 정리에서, 어떤 조건을 임시로 내세움. 또는 그 조건.

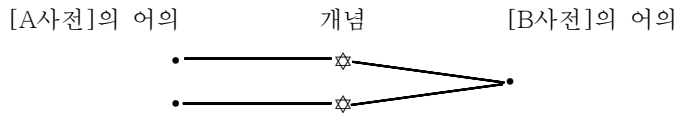
[큰사전] 가정9(假定) ① 사실이 아닌 것을, 또는 사실인지 아닌지 분명하지 아니한 것을 사실인 것처럼 여김. ② ((논)) 추리의 근거로서 내세우는 명제. ③ ((수)) 가설<sup>4</sup> ③.

위의 예에서는 논리학의 분야에서의 의미와 수학의 분야에서의 의미가 각각의 첨가된 의미로 주어져 있다. [연세]에서 [표준]으로의 어의 첨가는 아래의 가정(假定)의 경우에서 처럼 대부분 전문분야의 어의이다.

● 어의 통합/어의 분리<sup>23)</sup>

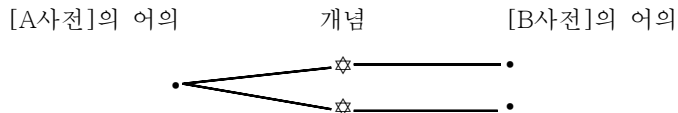
한 사전인 [A사전]에서 두 개 이상의 어의인 것이 다른 사전인 [B사전]에서 한 개의 어의로 나타날 경우를 어의 통합의 경우라고 한다. 물론, 똑 같은 경우를 [B사전]의 입장에서 보면 어의 분리가 된다. 이 경우 개념의 개수는 변하지 않는다.

(20) 어의 통합



한 사전인 [A사전]에서 하나의 어의인 것이 다른 사전인 [B사전]에서 두개 이상의 어의로 나타날 경우를 어의 분리의 경우라고 한다. 물론, 똑 같은 경우를 [B사전]의 입장에서 보면 어의 통합이 된다. 이 경우 개념의 개수는 변하지 않는다.

(21) 어의 분리



(22)

[연세] 가정<sup>1</sup>(家庭) 주로 한 장소에 모여 사는 아버지와 자녀들의 조직, 또는 그러한 생활의 장소.

[표준] 가정06 (家庭) 「1」 한 가족이 생활하는 집. 「2」 가까운 혈연 관계에 있는 사람들의 생활 공동체.

23) 편의상 통합과 분리에 대해 하나의 예만을 제시할 것이다. A사전에서 B사전으로의 통합은 B사전에서 A사전으로의 분리로 볼 수 있기 때문이다.

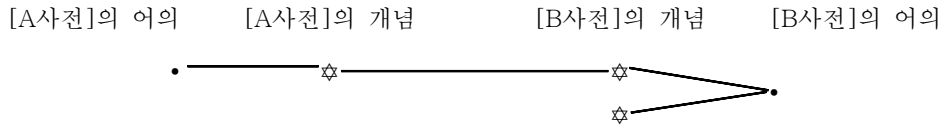
[큰사전] 가정3[家庭] 한 가족으로 이루어진 집안.

가정(家庭)의 경우에서 처럼 [연세]에서는 한 어의로 되어 있는 두 개의 개념이 [표준]에서는 두 개의 어의로 구분되어 있는 경우도 있다.

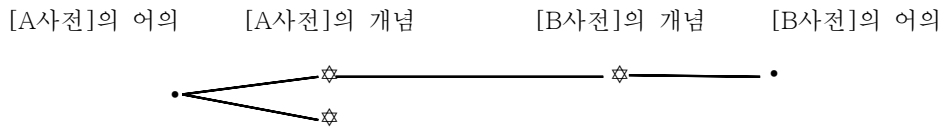
● 개념 첨가/개념 삭제

같은 어의에 대하여 A 사전에 없는 개념을 B 사전에서 첨가되거나, 같은 어의에 대하여 A 사전에 있는 개념을 B 사전에서 삭제하는 경우이다.

(23) 개념 첨가



(24) 개념 삭제



그 예는 다음과 같다.

(25)

[연세] 교수1(教授) 1. 대학 또는 그와 수준이 비슷한 교육 기관의 부교수 위의 직위. 또는 대학에서 가르치는 정식 교원.

[표준] 교수06(教授) 「2」 대학에서, 전문 학술을 가르치고 연구하는 사람. 부교수, 조교수, 전임 강사가 있는데 흔히 이들을 통틀어 이르는 말이다.

[큰사전] 교:수9[教授] ① 대학의 선생. ㉠ 대학에서 가장 높은 급의 선생. 부교수의 위. 정교수.

(26) 개념첨가

[연세] 가족(家族) 【명사】 대개 한 곳에 모여 생활하는, 부부와 그 자식들과 그 밖의 친척들의 집단.

[표준] 가족01 (家族) 「명」 「1」 부부와 같이 혼인으로 맺어지거나, 부모·자식과 같이 혈연으로 이루어지는 집단. 또는 그 구성원.

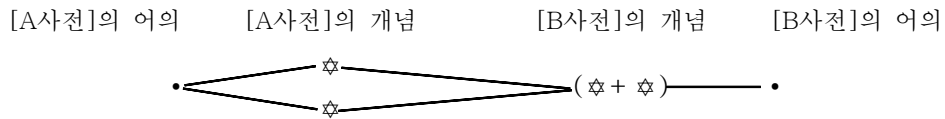
[큰사전] 가족1[家族] ① 아버지와 자식, 부부 따위의 혈연관계로 맺어져 한 집안을 이루는 사람들.



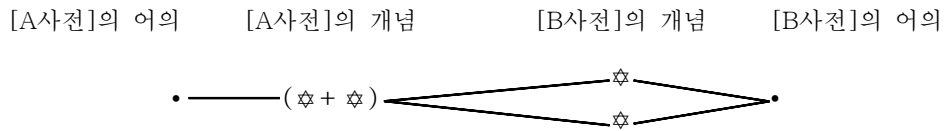
● 개념 통합/개념 분리

A 사전에서는 같은 어의에 대하여 두 개의 개념이 B 사전에서 하나의 개념으로 통합되거나, A 사전에서는 같은 어의에 대하여 하나의 개념이 B 사전에서 두 개의 개념으로 분리되는 경우이다.

(27) 개념 통합



(28) 개념 분리



그 예는 다음과 같다.

(29)

[연세] 대통령(大統領) 공화 정치 체제의 국가에서 최고의 직위에 있는 사람. 또는 그 직위.

[표준] 대통령(大統領) 『법』 외국에 대하여 국가를 대표하는 국가의 원수. 행정부의 실질적인 권한을 갖는 경우와 형식적인 권한만을 가지는 경우가 있는데 우리나라는 전자에 속한다.

[큰사전] 대통령[大統領] ((정)) 공화국의 최고 지도자. 국민에 의하여 직접 또는 간접으로 뽑히어, 일정한 임기 동안 나라의 전반에 걸친 행정을 통할하며 나라를 대표한다. 그 권한에 있어서 형식적 권한만을 가지는 경우도 있다.

국어사전의 뜻풀이말(정의문)을 이용하여 명사들의 상하위어를 규명하여 우리말의 의미계층계를 구축하려는 기존의 대부분의 연구는 단어와 어의를 구별하고 있지 않으며 어의 안에서 개념을 구별하고 있지 않는 것을 지적하지 않을 수 없다. 국어의 경우 특히 명사의 경우 한자어가 많기 때문에 같은 한글 철자의 다른 한자를 쓰는 동형이의어가 많은 관계로 동형이의어를 구별하는 수준에서 그치는 경우가 많다. 그리하여 하나의 단어 차원에서만 구별하고 어의 차원에까지 구별하지 않는다는 것이다. 이는 최근의 <<연세한국어사전>>구축을 위한 말뭉치를 이용한 빈도조사에서 어의를 고려한 것을 제외하고 어떤 말뭉치를 이용

한 빈도조사에서도 어의의 빈도를 말하고 있지 않다는 것에서 알 수 있다. 더군다나 하나의 어의 내의 개념까지도 구별하는 빈도조사는 전무하다. 이는 사전 자체를 위해서는 필요치 않은 빈도조사이기 때문이다.

온톨로지는 개념연결망이지 어의연결망이 아니다. 반면에 언어사전은 어의를 중심으로 의미가 배열되어 있는 것이지 개념을 중심으로 배열되어 있는 것이 아니다. 온톨로지 구축을 위해 국어사전을 사용하려면 우선 어의와 개념과의 관계를 명확히 할 필요가 있다.

어의 구별에 대해서는 사전의 어의 구별법에 대해 간략히 논의하고 기존사전의 어의 구별현황을 보이고 이의 문제점을 지적한 후 잠정적인 해결방안을 제시하고자 한다.

한 단어에 대해 비록 시간이 흘러가면서 혹은 개인 마다 다를 수는 있어도 같은 시대에 존재하는 국어사전의 어의의 개수와 다르다는 것은 한 단어의 어의 분류결정의 일정한 법칙이 없다는 것을 의미한다. 최근 <<연세한국어사전>>에 이르기까지 기존의 모든 사전들이 어의분류결정기준에 대해 언급하고 있지 않다.

이는 개념의 문제가 아니라 어의의 문제이다. 한 언어에 단어가 몇 개가 있느냐는 큰 의견의 차이가 없다. 그러나 각각의 단어의 어의가 몇 개냐는 동형어의어나 다의어의나 문제와 더불어 사전학자 마다 사전편찬자 마다 의견이 다르다. 마찬가지로 단순한 개념의 차이로 봐야 할지 어의의 차이로 봐야 할지도 의견이 분분하다.

반면에 단어의 어의를 구성하고 있는 개념이 몇 개냐는 사전마다 개수가 다르기는 해도 일반적으로 있다는 것에 대해 큰 의견의 차이를 보이지 않는다. 사전은 일반적으로 사전의 읽는 독자의 수준에 맞추어 개념의 수를 확대하여 표현 할 수도 있고 줄여서 표현 할 수도 있다.

그러나 두 개의 개념이 하나의 어의에 속하는지 아니면 각각의 서로 다른 어의에 속하는지는 언어학적인 정보이다. 그러나 아직까지도 이에 대해 확답하고 있는 언어학의 어떤 이론도 존재하지 않는다. 이에 대한 선택은 국어사전의 독자가 누구냐에 의해 결정될 수 있는 것이 아니다. 물론 사전학자의 자의적 선택에 맡겨질 수도 없는 것이다.

온톨로지를 구축하는 우리의 입장에서 이 문제를 해결할 방법을 무엇인가? 절차를 제시 한다면, 첫째 최근에 만들어진 신뢰할 수 있는 국어사전을 모아, 둘째 그 사전들에 수록된 특정단어의 모든 개념을 추출하고, 셋째 이를 (최소한) 일관된 어의분류법칙 하에 어의를 변별하여야한다.

조평옥 외 (1999)의 중심어 자체의 동형어의어만을 구별하고 있다. 예를 들어 중심어로서 사용되는 <말>은 동형어의어가 사전 상 7개 있으며 이 중 중심어로 사용되고 있는 것은 4개뿐이라고 밝히고 있다. <말>이 중심어로 사용된 표제어는 3,650개이고 7개의 동형어의어중 말(語)이 3, 510개가 사용되고 말(馬)이 130개 사용되고 말(단위)이 3개, 말(장기)이 7개 사용되었다고 기술하고 있다. 그러나 명사의의미체계를 구축하는 데 있어서 단지 동형어의어만을 구별하는 것은 불충분하다.

## 5.2. 온톨로지구축을 위한 제안: 개념기반 중심어의의미체계

첫째, 개별언어의 사전에 기록되어 있는 개별언어의 어의목록에서 개념중심어를 추출하여 개별언어의 개별 중심어체계를 구축한다. 이러한 개별언어의 개념중심어체계는 여러 개의 언어를 구축한다.

둘째, 여러 개의 언어를 통해 구축한 개념중심어체계들의 공통점을 위주로 하여 언어

중립적인 하나의 온톨로지를 구축한다.

셋째, 개별언어의 어의목록과 개념중심어체계 온톨로지를 하나로 연결한다. 모든 언어를 이와 같은 연결망을 만들어 사용할 수 있다

예를 들어 표제어 <교수>를 [연세] 사전에서 찾아보기로 하자. 두 개의 동형이의어가 있고, 이 중 첫 번째 동형이의어는 두 개의 어의로 구성 되어 있으며, 이 중 첫 번째 어의는 두 개의 개념으로 구성되어 있다.

(30) [연세]

- 교수1(教授) 1. 대학 또는 그와 수준이 비슷한 교육 기관의 부교수 위의 직위. 또는 대학에서 가르치는 정식 교원. 2. 가르치는 일.  
 교수2(敎首) (형벌로 죄수를) 목을 매어 죽이는 것.

이러한 목록이 사전의 어의목록으로 기재되어 있다면 우리는 이것을 기반으로 개념중심어 찾기 과정을 거쳐 개념중심어체계를 구축한다. 그림-3에서와 같이 여러 언어들의 개념중심어체계를 참고로 하여 언어중립적 온톨로지를 구축하고, 그림-2에서와 같이 사전의 어의목록과 개념중심어체계와 온톨로지를 서로 연결하는 연결망을 구축한다.

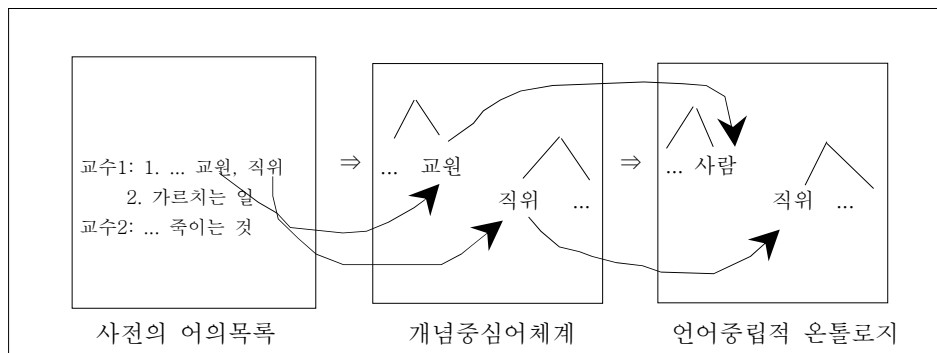


그림-2 사전, 개념중심어체계, 온톨로지의 연결

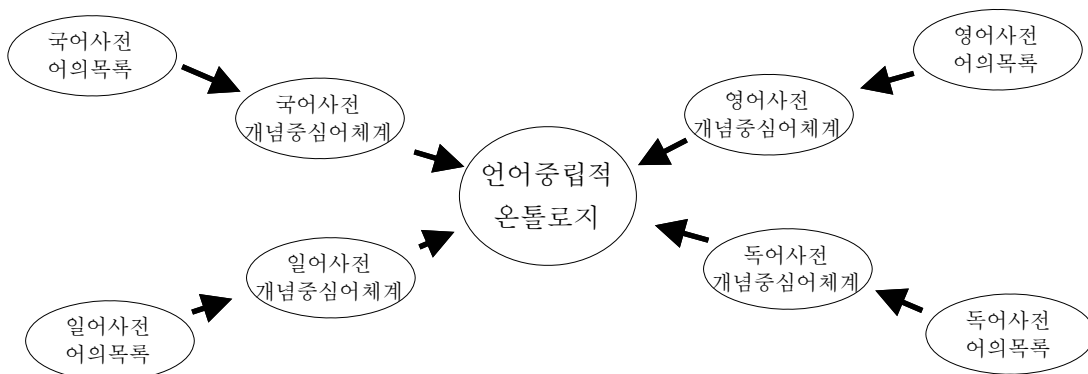


그림-3 개별 언어와 언어중립적 온톨로지의 연결망

이와 같이 사전과 온톨로지를 연결 하는 데에 Philpot etc.(2003)는 시시커밥(shishkebab)을 이용하여 어의와 개념의 차이를 설명하려고 한다. 시시커밥은 한 어의의 개념들을 묶는 또 하나의 언어적 장치이지 결코 온톨로지의 장치는 아니다. 물론 이 장치는 사전과 온톨로지를 연결할 경우만 의미가 있고 온톨로지 부분에서 보다는 사전 부분에서만 필요한 장치이다. 그림-2의 경우 시시커밥은 사전과 온톨로지 사이에 작용하는 것이 아니고 단지 한 언어 내에서 사전의 어의목록과 개념중심어체계 사이에서만 작용하는 것이다. 시시커밥이 순수 언어적인 장치라는 것을 말하고 있는 것이다. 이는 시시커밥이 사전과 온톨로지 사이에 작용하는 것이라는 Philpot etc.(2003)의 주장과는 대립된다.

2절에서 단어의 개념과 일반 개념은 다르다고 우리는 전제하였다. 개념중심어체계는 단어의 개념망이고, 온톨로지는 일반 개념에 가깝다. 그러므로 언어사전에 기반을 둔 개념중심어체계와 언어화자의 공통적 세상지식에 기반을 둔 온톨로지를 구별하여야 하는 것이다.

온톨로지 구축을 위한 방안으로 사전의 정의문과 특히 그 중심어를 활용하는 연구의 문제에 대한 해결책으로는 첫째 동형이의어 차원에서나 혹은 어의 차원에서 의미를 구별하는 것보다는 반드시 개념차원에서 의미를 구별해야 하므로 사전의 정의문을 세분화하여 개념차원에서의 중심어를 활용하여야 한다. 그러므로 온톨로지 구축을 위해서는 사전의 어의 구별은 개의치 않아도 되고 오직 개념에 의하여 체계를 구축하여야 할 것이다. 의미태깅 빈도조사에 있어서도 동형이의어구별 통계나 어의구별 통계보다도 개념구별 통계를 구축하여 활용할 필요가 있음을 언급해 둔다.

둘째 온톨로지 구축을 위해서는 오직 사전의 개념만이 관심사이므로 사전 마다 정의가 다르다는 것의 문제는 개념이 정의문안에 있느냐 없느냐만이 문제가 될뿐이고 어의 구별은 온톨로지 구축에는 전혀 고려되지 않는 것이다. 온톨로지 구축자로서의 우리의 관심사가 대폭 줄어드는 효과가 있다.

셋째 사전의 어의와 직접 온톨로지의 개념을 연결하는 방법보다는 일단 사전의 정의문을 통해 중심어의체계를 구축한 후에 그 체계를 독립적으로 구축된 (최)상위 온톨로지와 연결하여야 한다. 물론 중심어의체계 자체는 언어에 연결된 개념이므로 독립적인 개념과 혼동해서는 안 될 것이다. 결과적으로 어의구별을 무시하고 언어개념구조를 만들고, 이를 언어 독립적 혹은 언어 중립적 온톨로지를 구축해야 한다는 것이다.

넷째 언어사전을 통해 온톨로지를 구축하는 것의 완벽한 것이 아님을 인정하고, 온톨로지 구축을 위한 다른 방안들과 상호보완적으로 작업을 하지 않으면 안 된다. 조평옥, 옥철영(1999) 등에서의 연구 결과가 말해주듯이 언어사전만으로는 온톨로지의 전체를 구축할 수 없고 온톨로지의 중간단계 즉 중위온톨로지의 자료로 활용할 수 있을 뿐이다. 그러므로 사전의 정의문 이외의 자료를 활용하는 온톨로지 연구와 사전을 활용한 온톨로지 연구는 계속 병행되어야 할 것이다.

## 6. 나오기

지금까지 우리는 사전의 정의문의 중심어에 근거하여 명사의 의미계층체계를 구축한 후 이를 온톨로지 구축에 활용하는 작업을 문제점이 근본적으로 사전정의문 자체에도 있고 이러한 불완전하고 일관성이 결여된 사전의 정의문을 활용하여 온톨로지를 구축하는 방법에도 문제가 있다는 것을 확인하였다.

사전정의문 자체의 문제점에 대한 해결책으로 중심어의미체계를 활용한 정의문을 구

축해야 한다는 것이다. 이를 위해서는 사전뜻풀이편찬자들은 우선 중심어의미체계를 구축하는 것이 급선무이다.

온톨로지 구축을 위한 방안으로 사전의 정의문과 특히 그 중심어를 활용하는 연구의 문제에 대한 해결책으로는, 첫째, 동형어의어 차원에서나 혹은 어의 차원에서 의미를 구별하는 것보다는 반드시 개념차원에서 의미를 구별해야 하므로 사전의 정의문을 세분화하여 개념차원에서의 중심어를 활용하여야 한다. 둘째, 온톨로지 구축을 위해서는 오직 사전의 개념만이 관심사이므로 사전 마다 어의분류가 다르다는 것은 문제가 되지 않고 오직 개념이 어떤 어의에든 간에 정의문안에 있느냐 없느냐 만이 문제가 될 뿐이고 어의 구별은 온톨로지 구축에는 전혀 고려되지 않는 것이다. 셋째, 사전의 어의와 직접 온톨로지의 개념을 연결하는 방법 보다는 일단 사전의 정의문을 통해 중심어의미체계를 구축한 후에 그 체계를 독립적으로 구축된 (최)상위 온톨로지와 연결하여야 한다. 넷째, 언어사전만을 통해 온톨로지를 구축하는 것은 온톨로지 구축에 있어 언어적인 면만을 보는 것이어서 완벽한 것이 아님을 인정하고 온톨로지 구축을 위한 다른 방안들과 상호보완적으로 작업을 하지 않으면 안 된다. 언어사전만으로는 온톨로지의 전체를 구축할 수 없고 온톨로지의 중간단계 즉 중위온톨로지의 자료로 활용할 수 있을 뿐이다. 그러므로 사전의 정의문 이외의 자료를 활용하는 온톨로지 연구와 사전을 활용한 온톨로지 연구는 계속 병행되어야 할 것이다.

이러한 제안에 대해서 기존 자료가 그것이 사전의 분류 작업이든 아니면 말뭉치의 의미태깅자료이든 현실적으로 개념까지 세분화하여 둔 경우는 없으므로 이를 기준으로 온톨로지를 구축할 상황이 아닌 것이다. 그러나 온톨로지는 개념망이고 그 이상도 그 이하도 아니므로 새로이 준비할 수 없다면 기존의 의미자료를 수정해 나아가야 할 것이고 이것이 최종적으로 온톨로지를 구축하는데 의미가 있을 것이고 우리의 국어사전의 뜻풀이의 발전에도 기여 할 것으로 본다.

우리는 사전을 활용한 온톨로지 연구를 통해 온톨로지 발전에 기여할 수 있을 뿐만 아니라 언어사전 즉 국어사전의 발전 특히 국어사전의 정의문기술의 발전에도 기여할 수 있음을 살펴보았고 이것이 계기가 되어 온톨로지뿐만 아니라 국어사전의 정의문(뜻풀이) 연구에 많은 이들이 관심을 갖게 되기를 바란다.

## 참고문헌

- 김광해(1998), <유의어의 의미 비교를 통한 뜻풀이 정교화 방안에 대한 연구>. 선청어문 제 26호 5-40.
- 서상규(2000), <<“한국어 교육 기초어휘 의미 빈도 사전의 개발” 사업보고서>>, 문화관광부 한국어 세계화 추진 위원회.
- 이기동(1989), <사전 뜻풀이의 검토>, 연세대학교 언어정보개발원.
- 이상섭(1997), <사전의 뜻풀이에 대한 소견-『연세 한국어 사전』과 기존 국어 사전들과의 비교를 곁하여>, 연세대학교 언어정보개발원.
- 이찬중(2008), <단어빈도가 단어정의에 미치는 효과>, ms.
- 이해윤(1997), <전문용어의 정의문 분석>, 독어학 제 7집.
- 조평옥, 옥철영(1997), <한국어 명사 의미계층구조 구축>, 한글 및 한국어 정보처리 학술대회.
- 조평옥, 옥철영(1999), <의미속성에 기반한 한국어 명사 의미 체계>, 정보과학논문지(B) 제

26권 제 4호.

최경봉 (1996) <명사의 의미분류에 대하여>, 한국어학 4, 한국어학회.

허정, 김준수, 이수광, 옥철영 (2000), <국어사전의 명사 뜻풀이말 Parser>, 한국정보과학회  
봄 학술발표논문집 Vol. 27. N0.1.

Ide, N. & J. Véronis (1995), <Knowledge Extraction from Machine-Readable Dictionaries:  
An Evaluation>, Lecture Notes in Computer Science, Vol. 898, 19-34.

Fellbaum, C. (Ed.) (1999). <<WordNet: An Electronic Lexical Database>>, Massachusetts: MIT Press.

Hovy, E. H. (2005), <Methodologies for the Reliable Construction of Ontological  
Knowledge>, In F. Dau, M.-L. Mugnier, and G. Stumme (eds.), Conceptual  
Structures: Common Semantics for Sharing Knowledge. Proceedings of the 13th  
Annual International Conference on Conceptual Structures (ICCS 2005), pp. 91-106.

Philpot, A., M. Fleischman, E. H. Hovy (2003), <Semi-Automatic Constuctuion of a  
General Purpose Ontology>. Proceedings of the International Lisp Conference. New  
York, N.Y.

Pustejovsky, J. (1995), <<The Generative Lexicon>>, MIT Press.

Vossen, P. (2004), <EuroWordNet General Document>, Version 3 Final, University of  
Amsterdam, 108p.

사전류

[연세] <<연세한국어사전>>, 연세대학교 언어정보개발연구원 편(1998), 두산동아.

<http://kordic.britannica.co.kr/>

[큰사전] <<우리말 큰 사전>>, <http://nlpweb.kaist.ac.kr/Urimal/>

[표준] <<표준 국어 대사전>>, [http://www.korean.go.kr/08\\_new/index.jsp](http://www.korean.go.kr/08_new/index.jsp)