

# 『표준국어대사전』을 이용한 시소러스 구축

## Construction of Thesaurus Using 『The Korean Standard Dictionary』

한 상 길(Sangkil Han)\*

### < 목 차 >

- |                                   |                   |
|-----------------------------------|-------------------|
| I. 서론                             | 3. 관련관계의 적용       |
| II. 시소러스와 『표준국어대사전』의 용어관계         | 4. 부가관계의 적용       |
| 1. 시소러스의 용어관계                     | IV. 용어관계 정의의 문제점  |
| 2. 『표준국어대사전』의 내용                  | 1. 관련어휘 정보 기술의 문제 |
| 3. 『표준국어대사전』 내용과 『우리말 시소러스』의 용어관계 | 2. 용어의 분리         |
| III. 『표준국어대사전』 정의항을 이용한 시소러스 구축   | 3. 동등관계의 설정       |
| 1. 동등관계의 적용                       | 4. 계층관계의 설정       |
| 2. 계층관계의 적용                       | 5. 한정어의 사용        |
|                                   | 6. 북한어 관계         |
|                                   | V. 결론 및 제언        |

### 초 록

시소러스 작성에 있어서 가장 현실적인 어려움은 용어의 수집에 있다. 용어사전 뿐만 아니라 어휘사전도 시소러스의 중요한 용어 수집원으로 사용될 수 있다. 『표준국어대사전』은 어문 규정을 충실히 반영한 우리나라의 대표 사전이다. 또한 『표준국어대사전』은 단순히 표제어에 대한 용어 정의뿐만 아니라 해당 용어에 대한 다양한 정보를 체계적으로 담고 있기 때문에, 이를 시소러스 사전 구축에 활용할 수 있을 것이다. 본 연구는 『표준국어대사전』이 갖고 있는 다양한 용어관계 정보를 시소러스 용어관계로 정의하는 방안을 모색하였다. 또한 용어의 분리, 동등관계와 계층관계의 설정, 한정어의 사용, 북한어 관계 등 시소러스 구축에서의 문제점과 해결 방안을 제시한다.

키워드: 시소러스, 한국어 시소러스, 표준국어대사전, 용어관계, 한정어

### ABSTRACT

Collecting terms in thesaurus construction work is the most difficult. A dictionary is thesaurus can be used as an excellent term acquisition. Reflect faithfully the provisions of Korean literary "Standard Korean Dictionary" is a standard dictionary of the Korea. The "Standard Korean Dictionary" is simply the definition of entry, as well as a wide range of information about the term because it contains a systematic, it can be used to build a thesaurus. In this study, the "Standard Korean Dictionary" has the relevant information using a variety of terms, it is defined as the thesaurus term relationship. In addition, the separation of the term, equal relationships and hierarchical set of relationships, the use of qualifiers, North Korean issues, the issue presented in thesaurus construction, and suggest ways to solve the problem.

Keywords: Thesaurus, Korean thesaurus, Korean standard dictionary, Term relationships, Qualifiers

\* 대림대학교 문헌정보학과 교수(skhan@daelim.ac.kr)

• 논문접수: 2013년 11월 20일 • 최초심사: 2013년 11월 29일 • 게재확정: 2013년 12월 13일

## I. 서론

시소러스는 정보검색시스템에서 색인용어를 통제하여 검색의 효율성을 높이는 중요한 도구로 활용되고 있다. 시소러스는 색인 시에는 동일한 개념이 각기 다른 용어로 색인되는 것을 방지하고, 검색 시에도 동등관계, 계층관계, 연관관계 등 각종 용어 간의 관계를 표시하여 이용자가 빠른 시간 내에 정확하게 정보를 검색할 수 있도록 하는 기능을 수행한다. 그러나 실제 시소러스를 구축하는 일은 많은 시간과 노력이 필요한 노동집약적인 일이다.

시소러스 작성에 있어서 가장 현실적인 어려움은 용어의 수집에 있다. 용어 수집원은 주로 문헌이 되는데 기본적으로 참고할 문헌은 ① 시소러스, 분류표 등의 기본 어휘집, ② 백과사전, 사전, 어휘사전(lexicon) 등의 사전류, ③ 전문분야의 용어집, ④ 색인지나 초록지, 기타 출판물의 색인, ⑤ 편람, 목록, 교재, 디렉토리, 규격 등의 기타 자료집을 들고 있다<sup>1)2)</sup>. 그러므로 초기 시소러스를 구축하거나 이미 구축된 시소러스를 확장하는데 있어서 국어사전은 중요한 용어정보원으로 사용할 수 있을 것이다.

국어사전은 단어 하나하나에 어문 규정을 적용하여 구체적인 언어생활에 기준을 제시해 주는 도구이다. 국립국어연구원은 이러한 규범 사전의 필요성에 따라 1992년부터 1999년까지 8년 동안에 걸쳐 『표준국어대사전』을 발간하였다. 이 사전은 어문 규정을 충실히 반영하였을 뿐만 아니라 통일 이후 남북한 언어의 동질성 회복을 위해 다수의 북한어 표제를 담고 있으며, 옛말, 방언, 그리고 다수의 전문어를 싣고 있어 명실 공히 우리나라의 대표 사전이라고 할 수 있다. 이 사전은 2000년에는 CD-ROM으로 발간하였으며, 2003년부터 개정작업을 시작한 후 2008년에는 웹사전 형태<sup>3)</sup>로 인터넷을 통해 서비스 하고 있다.

이 사전에는 모두 509,076개의 표제어가 실려 있다. 이 중에서 주표제어가 440,594개이고, 부표제어가 68,482개이다. 또한 『표준국어대사전』은 단순히 표제어에 대한 용어 정의뿐만 아니라 해당 용어에 대한 다양한 정보를 체계적으로 담고 있기 때문에 이를 시소러스 사전 구축에 활용할 수 있다.

그러나 『표준국어대사전』은 전체 표제어 중 22.4%에 해당하는 124,254개가 동형이의어로 나타나고 있으며, 30개 이상의 동의어를 가진 단어가 20개나 되고 최대 46개의 동형이의어를 가지는 단어도 있어서 이들을 구별할 수 있는 방법이 필요하다. 대부분의 사전에서는 용어에 대한 어깨번호를 사용하여 구별하고 있지만, 용어간의 관계를 기본으로 하는 시소러스에서는 한정어를 사용할

1) J. Aitchison and A. Gilchrist, *Thesaurus Construction : a practical manual*, 2nd ed.(London : Aslib, 1987), pp.69-70.

2) D. Soergel, *Indexing Languages and Thesauri : construction and Maintenance*(Los Angeles : Melville, 1974), p.355.

3) 표준국어대사전(<http://stdweb2.korean.go.kr/>) [인용 2013. 10. 31].

수밖에 없기 때문에 한정어 사용에 관한 규칙이 필요하다. 또한 『표준국어대사전』에서는 동형어의 어가 아니라도 동일한 용어에 여러 가지의 의미를 복합적으로 제시하고 있는 경우가 많아서 시소러스 용어 구축에 사용하기 위해서는 이들 개념의 분리가 필요하다.

본 연구는 『표준국어대사전』이 갖고 있는 다양한 정보를 수작업 과정으로 실제 우리말 시소러스를 구축하면서 제기되는 문제점을 도출하고, 이를 해결할 수 있는 방안을 제시하였다.

## Ⅱ. 시소러스와 『표준국어대사전』의 용어관계

### 1. 시소러스의 용어관계

기본적인 시소러스 용어관계의 종류로는 ISO 2788:1986(E)<sup>4)</sup>에서 정의하고 있는데, 여기에는 계층관계(hyponymy: NT/BT), 동등관계(synonymy: USE/UF) 및 관련관계(association: RT)가 그 중심이 된다. 특히 ISO에서는 세 관계 중에서 계층관계만을 더 확장하고 있는데 전체부분관계, 사례관계 및 속관계가 그것이다. ISO 2788:1986(E)에서 정의하고 있는 용어관계는 동의관계를 UF(*used for or use for*: 우선어)와 USE(*use*: 비우선어로부터의 우선어 참조), 계층관계를 BT(*broader term*: 상위어)와 NT(*narrower term*: 하위어), 관련관계로 RT(*related term*: 관련개념어), 그리고 부가관계를 SN(*scope note*)으로 정의하고, 여기에 관련관계를 BTG(*broader term/generic*: 상위어/속), BTI(*broader term/instance*: 상위어/사례), BTP(*broader term/partial*: 상위어/부분)과 NTG(*narrower term/generic*: 하위어/속), NTI(*narrower term/instance*: 하위어/사례), NTP(*narrower term/partial*: 하위어/부분)관계로 확장하고 있다.

그러나 ISO 2788:1986(E)에서 제시하고 있는 용어관계는 동의, 계층, 관련관계로 단순화하여 용어관계가 부족하기 때문에 시소러스 표준에서 결정된 관계를 확장할 필요성이 제기 되었다. 한상길은 기본용어관계 이외에 최상위어(TT: top term), 분류 및 주제구분(CC: classification code or category code), 이력주기(HN: history note), 반의어(AT: antonym term)와 동등관계 확장을 위해 UF를 외래어, 방언, 이전용어, 약어, 이형표기로 구분할 것과 범위주기 확장을 위해 용어정의(DF: definition notes), 범위주기(SN: scope note)를 제안<sup>5)</sup>했고, 최석두는 반의어(ANT: antonym), 유의어(NST: near-synonym), 줄임말(ABT: abbreviation), 상품명(BNT: brand name), 학명(SNN: scientific name) 등을 제안<sup>6)</sup>하였다.

4) ISO 2788:1986(E), *Documentation - Guidelines for the establishment and development of monolingual thesauri*.

5) 한상길, 시소러스 용어관계 확장에 관한 연구, 박사학위논문, 중앙대학교 대학원, 1999.

6) 최석두, 한글 시소러스 구축 기준에 관한 연구, 박사학위논문, 연세대학교 대학원, 2002.

이외에도 지금까지 연구에서 나타난 용어관계로는 이전용어와 이후용어(PT: *prior term*, LT: *later term*), 남한어와 북한어(NK: *North Korean*, SK: *South Korean*), 분류기호(KDC, DDC, UNSPSC, 국가과학기술표준분류, Agrovoc 분류 등), 대응 외국어, 저자와 표제명(AUT: *author*, TIT: *title*), 연도(YEA: *year*)<sup>7)</sup> 용어범주, 패싱, 용어의 출전, 링크, 이미지, 속성명/속성값 등이 있다<sup>8)</sup>. 저자/작품/연도의 관계, 품사(현재는 명사와 형용사가 주임), 출전, 용어의 정의, 용어의 수준, 용어의 범주, 그리고 각종 외국어 관계가 있다. 이때 사용하는 외국어 코드로 ENG(영어), ESP(스페인어), FRA(프랑스어), GER(독일어), GRE(그리스어), ITA(이탈리아어), JPN(일본어), LAT(라틴어), MON(몽골어), RUS(러시아어), CHN(중국어) 등<sup>9)</sup>으로 정의하고 필요시 확장 하도록 하고 있다.

## 2. 『표준국어대사전』의 내용

사전은 관련어들로 이루어진다. 사전의 가장 큰 목적이 단어의 뜻풀이라고 했을 때, 뜻풀이의 본질적 성격은 해당 단어와 관련 있는 다른 ‘단어’를 제시하는 것이다. 또한 관련된 단어가 ‘뜻풀이’ 항목에 제시되든, 예문에 제시되든, 길잡이 말이나 작용역(영역), 관련어 항목에 제시되든 해당 표제어와 의미적 관련성을 보여 준다는 점에서 모두 중요한 정보이다<sup>10)</sup>.

국어연구원에서 발행한 『표준국어대사전』에 대한 분석자료<sup>11)</sup>를 이용해 시소러스 용어 사전에 적용할 수 있는 내용을 요약하면 다음과 같다.

### 가. 표제어

다른 사전과 마찬가지로 『표준국어대사전』도 표제어가 가나다순으로 나열되어 있고 각 표제어에 담긴 정보들은 일정한 형식에 따라 제시되어 있는데 이들은 원어, 발음, 뜻풀이, 관련어휘와 참고 어휘, 어원, 속담과 관용구어의 순서로 실려 있다. 표제어의 범위는 다음과 같다<sup>12)</sup>.

- (1) 표준어를 비롯하여 북한어, 방언, 옛말을 두루 수록하였으며 흔히 쓰는 비표준어도 올렸다. 이두는 수록하지 않았다.
- (2) 일반어뿐만 아니라 전문어, 고유 명사도 수록하였다. 전문어와 고유 명사는 수가 많기 때문에 일부만 선별하여 수록하였다.

7) 최석두, 한상길, “국립중앙도서관을 위한 시소러스형 주제명표목표의 개발,” 한국시소러스학회 포럼(2003. 12), pp.65-66.

8) 최석두 등, 과학기술분야 시소러스 구축 연구 2007(대전 : 한국과학기술정보연구원, 2007), p.3.

9) 최석두 등, “대단위 우리말 온톨로지 구축을 위한 시소러스의 개발,” 정보관리학회지, 제23권, 제4호(2006. 12), pp.155-156.

10) 김진혜, “표준국어대사전의 관련어 정보와 어휘관계 기반 사전 기술,” 한국어의미학, 제24권(2007), p.24.

11) 이운영, 표준국어대사전 연구 분석(서울 : 국립국어원, 2002).

12) 국립국어연구원, 표준국어대사전(서울 : 국립국어연구원, 1999), p.5.

(3) 단어 이하의 단위만 수록하는 것이 원칙이나 전문어와 고유 명사는 구도 수록하였으며 관용구와 속담도 수록하였다.

(4) 북한어는 《조선말 대사전》(1992)에 수록된 단어 가운데 남한에서 쓰임이 확인되지 않은 단어와 어문 규정의 차이로 달리 표기하는 단어를 편찬 원칙에 따라 선정하여 수록하였다. 남한에서 쓰는 단어라도 북한에서만 쓰는 용법이 있다면 북한어 뜻을 덧붙였다.

한편, 품사에 따른 표제어는 명사가 335,057개로 가장 많고, 이는 전체 사전 표제어의 65.82%에 달하는 수치로 국어 어휘에서 명사가 얼마나 많은 비중을 차지하는가를 단적으로 보여 주는 것이다. 다음으로는 동사가 68,394개로 13.43%의 비율을 보이고 있고, 무품사 표제어를 제외하면 부사(3.52%)와 형용사(3.41%)가 거의 비슷한 수치로 그 다음을 차지하고 있다.

또한『표준국어대사전』에 실린 표제어의 42.5%는 순수하게 전문 어휘라는 의미이다. 이렇게 전문어 비율이 높은 것은 『표준국어대사전』이 종합 사전적인 면모를 갖고 있다는 것을 단적으로 보여 준다.

전문어는 다시 53개의 전문 분야로 나누어진다. <표 1>은 전문어 주표제어를 분야별로 알아본 것이다. 한 표제어가 여러 분야에 속하는 경우가 있는데 이럴 때에는 중복하여 계산되기 때문에 아래 표의 계는 전체 전문어 수보다 많게 나타난다.

<표 1> 전문어 주표제어의 분야별 분류

가톨릭	1,448	법률	9,350	전기	3,003
건설	6,215	불교	10,029	정치	2,119
경제	8,542	사회	2,329	종교	1,012
고적	2,398	생물	4,243	지리	5,961
공업	3,695	수공	2,275	천문	2,095
광업	3,338	수산	910	가톨릭	1,448
교육	1,479	수학	4,127	철학	2,065
교통	1,552	식물	13,127	출판	1,464
군사	4,979	심리	1,613	컴퓨터	1,500
기계	2,369	약학	1,487	통신	1,077
기독교	1,173	언론	558	한의	5,077
논리	718	언어	4,208	항공	891
농업	4,732	역사	20,858	해양	1,132
동물	11,695	연영	1,731	화학	8,865
문학	3,788	예술	1,295	인명	10,327
물리	7,905	운동	4,962	지명	6,996
미술	1,516	음악	6,939	책명	2,051
민속	4,912	의학	10,501	고유명사	498
법률	9,350	전기	3,003	계	229,129

<표 1>에서 볼 수 있듯이 전체 229,129의 전문어 중에서 전문어 주표제어가 219,044개이므로 10,085개가 둘 이상의 전문 분야에 속함을 알 수 있다.

따라서 『표준국어대사전』에 실린 표제어 중 65.82%인 335,057개의 명사를 이용한 시소러스 사전 구축이 가능하고, 필요할 경우 형용사 사전이나 동사사전의 구축이 가능하다. 또한 이 사전은 전문용어를 수록하고 있고, 이들을 주제별로 구분하고 있어 이를 이용한 전문용어 시소러스를 구축이 가능하며, 종합 시소러스일 경우는 주제 카테고리 사용하여 용어를 구분할 수 있을 것이다.

#### 나. 원어정보

한자어와 외래어는 표제어 다음에 그 원어를 밝혔다. 만일 고유어와 결합한 표제어이면 고유어 음절수만큼 '-'를 사용하여 표시하였다.

〈예〉 가터-뜨기 (garter--), 나비-넥타이 (--necktie), 다락-방 (--房)

원어의 표기를 위하여 사용한 문자는 한자와 로마자로 국한하였다. 그 밖의 문자는 로마자화하여 표기하였다. 원어의 언어명을 약호를 사용하여 제시하였다. 또한 중국어나 일본어를 원음대로 차용한 경우는 로마자로 표기하고 그 뒤 '[' ] 안에 한자를 제시하였다.

〈예〉 기모노(〈일〉kimono[着物]), 시안 비림(Xián[西安]碑林), 콩소메(〈프〉consommé)

'한자 병기'와 '외래어 병기'는 해당 표제어에 대한 원어가 둘 이상이어서 이를 병기한 경우이다. 이 때에는 가능한 원어를 모두 써 주되 각각을 '/'로 구분한다. 한자어에서 이렇게 둘 이상의 원어가 나타나는 표제어는 모두 2,045개로, 이 중에서 47개는 세 개의 원어가 병기되었고 나머지는 둘이 병기되었다. 외래어에서 둘 이상의 원어가 나타나는 표제어는 모두 49개로, 모두 두 개의 원어가 병기되어 있다.

언어명은 영어를 제외하고 모두 30개가 제시되어 있다. 각 언어의 첫 자를 ○ 안에 넣어서 해당 언어를 밝히고 있는데, 언어명이 표시되어 있는 주표제어 4,143개에 달한다.

따라서 시소러스 구축에 있어서 외국어 표기를 가능하게 한다. 그러나 영어와 한자 혹은 우리말이 함께 있는 복합 명사의 경우에는 용어에 대한 완전한 외국어를 기입할 수 없는 것이 한계점이다.

#### 다. 뜻풀이

『표준국어대사전』은 표제어로 실린 어휘만을 사용하여 뜻풀이를 하고 있다. 뜻풀이는 기본적으로 정의항과 설명 부분으로 되어 있으며, 정의항은 기본적으로 '종차+유개념'로 이루어지고 표제어의 품사와 같은 형식으로 끝맺게 된다. 품사와 같이 끝맺는 것이 어려울 때는 상위 언어적(meta-linguistic) 뜻풀이를 하고 있다. 이러한 용어들에는 '강조하여, 겸손하게, 구어적으로, 낮잡아, 놀림조로, 높여, 달리, 비속하게, 속되게, 아올러, 예스럽게, 일상적으로, 전문적으로, 통틀어, ○○의 은어로' 등등이 있다. 뜻풀이 맨 앞이나 뒤에 이러한 용어를 사용함으로써 뜻풀이의 전체적인

성격 등을 표시하고 있다.

그러나 주표제어가 북한어, 방언, 옛말, 비표준어, 본말, 준말, 높임말, 낮춤말, 순화대상어, 이칭, 은어 등일 때는 직접 뜻을 풀이를 하지 않고 기본표제어를 참고 하도록 하고 있다.

북한어의 경우 직접 뜻을 풀이를 한 것이 51,180개인데 이는 남한 대응어로 돌린 것보다 두 배 이상 많은 수치이다. 이러한 결과는 남한의 표준어와 단순한 표기나 표현상의 차이를 보이는 북한어보다 실제로 의미나 어휘 자체가 남한과 다른 북한어가 상당히 많다는 것을 보여 주는 것이라 하겠다.

『표준국어대사전』은 모두 20,503개의 방언표제어가 실려 있어 전체 표제어의 4.0%에 해당한다. 또한 현대어 외에 12,488개의 옛말(고어)을 실었고, 9,464개의 비표준어도 등재하였다.

또한 품사 등의 약호에는 「명사」, 「의존명사」, 「대명사」, 「수사」, 「동사」, 「보조동사」, 「형용사」, 「보조형용사」, 「관형사」, 「부사」, 「조사」, 「감탄사」, 「어미」, 「접사」, 「본(본말)」, 「준(준말)」, 「비(비슷한말)」, 「반(반대말)」, 「낮(낮춤말)」, 「참고어휘」, 『방언』, 『옛말』, 『북한어』, 관용, 속담의 26개가 있다.

따라서 표제어의 정의항을 이용해서는 시소러스의 상위개념과 하위개념을 적용할 수 있으며, 이들 관계 중 「본(본말)」, 「준(준말)」, 「비(비슷한말)」, 「반(반대말)」, 「낮(낮춤말)」, 「참고어휘」, 『방언』, 『옛말』, 『북한어』 등은 시소러스의 용어관계로 활용할 수 있다.

#### 라. 동의어

『표준국어대사전』에서는 의미가 동일한 표제어가 여러 항목이 있을 경우에는 표제어 기독교인의 뜻풀이에 “기독교를 믿는 사람. ≡ 기독교인, 기독교도, 기독교자, 크리스천”과 같이 하나의 표제어에만 직접 뜻을 풀이를 하고 그 밖의 표제어는 동의어로 처리하고 있다.

뜻풀이 뒤에 동의어를 제시한 표제어는 50,488개이고, 뜻을 풀이를 하지 않고 기본 표제어의 동의어로 돌린 표제어는 70,340개이다. 『표준국어대사전』에 실린 약어는 모두 398개이며, 약어는 “원말을 줄여 이르는 말”과 같은 방식으로 뜻을 풀이하고 있다. 따라서 이들 용어는 시소러스의 동등관계 정의에 활용할 수 있다.

#### 마. 관련어휘와 참고어휘

『표준국어대사전』에서는 해당표제어와 관련이 있는 어휘들을 용례 뒤에 관련어휘로 제시하여 표제어의 이해를 돕고 있다. 관련어휘에는 본말, 준말, 비슷한말, 반대말, 높임말, 낮춤말 등이 있다. 또 어떤 표제어의 의미를 이해하는 데 도움이 될 수 있는 표제어를 관련 어휘 뒤에 「참고 어휘」 표시로 보여 주었다. 셋 이상의 대립 관계가 형성되는 동류 개념, 상위 범주어, 하위 범주어, 동의 관계에 있는 일반어와 전문어, 별칭이나 이칭, 모음이나 자음의 차이에 따라 그 의미가 미세하게 달라지는 말, 대립적인 성격을 띠기는 하지만 반의 관계로 볼 수 없는 말, 표제어의 의미를 이해하는 데 참조가 되는 말, 구별해야 하는 말, 환경에 따라 형태가 달라지는 어간이나 어미 따위가 이에 해당한다.

의미의 차이가 거의 없어 동의 관계에 있는 표제어들은 어느 하나를 기본 표제어로 삼아 뜻풀이를 하고 나머지에서는 ‘=’ 기호를 사용하여 기본 표제어로 뜻풀이를 돌렸다. 기본 표제어에서는 뜻풀이 다음에 ‘≡’ 기호로 해당 동의어를 모두 보여 주고 있다.

『표준국어대사전』에는 표제어와 명시적인 관계에 있는 어휘들은 관련 어휘 항목에 제시한다. 관련 어휘 항목에는 본말, 준말, 비슷한 말, 반대말, 높임말, 낮춤말 등이 있다.

따라서 이들의 정보 중 비슷한말과 높임말, 준말은 시소러스의 동등관계에 반대말, 참고어휘는 관련관계 정의에 이용할 수 있다.

### 3. 『표준국어대사전』 내용과 『우리말 시소러스』의 용어관계

『우리말 시소러스』는 의미 정보와 용어 간의 관계를 갖는 우리말 용어의 메타데이터를 구축한다는 목표 아래 1993년부터 용어를 수집하고 이를 표준 시소러스 구조로 기술하고 있다. 현재 약 한글 용어 약 133만 영어(한글 용어 82만 용어, 외국어 51만 용어)를 등재하고 용어관계를 정의하고 있다. 등재 용어의 대상으로는 주제명, 인명, 기관/단체명, 지명, 동물명, 식물명, 법률명, 문헌명, 기념물명, 상품명, 회의/집회명, 사건명 등의 명사류, 그리고 동사와 형용사를 포함시키고 있다. 아직 충분하지는 않으나 이제 현장의 비구조화 데이터를 어느 정도 처리할 수 있는 양의 데이터를 구축하였다고 생각하여 지금까지의 결과를 공개하고 인터넷을 통해 공개하고 있다<sup>13)</sup>. 매월 1만 용어 이상의 새로운 용어를 추가하고 있으며, 이 작업은 수정 작업과 함께 계속되고 있다.

『우리말 시소러스』의 용어관계는 2.1에서 언급한 대부분의 용어관계를 포함하고 있으며, 필요하면 얼마든지 추가할 수 있도록 설계되어 있다. 여기에서는 다양한 용어관계 중에서 『표준국어대사전』이 갖고 있는 정의항의 내용을 이용하여 『우리말 시소러스』를 구축할 수 있는 명사를 중심으로 다룬다.

『표준국어대사전』내용을 시소러스의 용어관계에 적용하면 다음과 <표 2>와 같이 나타낼 수 있다. 동의관계는 방언, 유의어, 줄임말, 남한어/북한어, 이전/이후관계, 학명, 외래어 관계를 사용하여 정의할 수 있으며, 계층관계는 표제어의 정의항에 포함된 상위어와 하위어를 이용할 수 있다. 마찬가지로 관련관계는 표제어의 정의항에 포함된 참고어휘와 반대말, 표제어의 정의항에 포함된 저자와 저작을 이용하여 관계를 설정할 수 있다. 부가관계에서 정의는 표제어의 정의항 내용을, 분류 및 주제구분은 229,129개에 달하는 주표제의 주제 구분을 이용하여 용어관계를 설정할 수 있다.

13) 우리말 시소러스 <<http://www.ontorus.net>> [인용 3013, 10, 31].



〈표 2〉 『표준국어대사전』내용과 『우리말 시소러스』 용어관계

관계내용	시소러스의 용어관계		『표준국어대사전』의 내용
	용어관계	관계기호	
동의관계	동의관계 <방언>	USE/UF	동의를 제시한 50,488개의 표제어와 동의어로 돌린 70,340개의 표제어 68개의 높임말과 29개의 낮춤말 20,503개의 방언 표제어
	유의어	USE/UF	18,646개의 비슷한 말
	줄임말	USE/UF	1,011개의 본말과 2,462개의 준말
	남한어/북한어	NK/SK	20,883개의 남한 대응어
	이전/이후관계	LT/PT	12,488개의 옛말
계층관계	상위어	BT(BTP,BTI)	표제어의 정의항에 포함된 상위어
	하위어	NT(NTP,NTI)	표제어의 정의항에 포함된 하위어
관련관계	관련관계	RT	표제어의 정의항에 포함된 참고어휘
	반의어	RT	5,834개의 반대말
	저자/저작	AUT/TIT	표제어의 정의항에 포함된 저자와 저작
부가관계	정의	DF	표제어의 정의항 내용
	분류	CC	229,129개의 전문어 주표제의 구분
	학명	SNN	표제어의 정의항에 포함된 학명
	외래어관계	ENG, JPN 등	4,143개의 언어정보

### Ⅲ. 『표준국어대사전』 정의항을 이용한 시소러스 구축

용어는 특정 주제분야에서 특정 주제개념을 표현하는 하나 이상의 단어로 구성된 의미단위로서, 지식의 단위요소인 개념을 언어기호로 표현한 것이다. 정의를 내리는 일은 용어의 기본적인 의미를 제시하는 과정, 다시 말해 용어의 의미를 확정짓는 행위라고 한다면 이 행위의 결과가 용어로 표현된다. 그러므로 “이 정의를 통해 특정개념이 속한 개념체제와 이 체계 속에서 특정 개념을 고유하게 식별하고 분류하는 기능을 지닌다.”<sup>14)</sup>고 할 수 있다. 따라서 역으로 이러한 사전의 정의항을 이용하면 시소러스 구축에 많은 도움을 얻을 수 있다.

#### 1. 동등관계의 적용

##### 가. 동의 및 유의관계

14) 김태수, “용어정의를 도입한 시소러스 개발 연구,” 정보관리학회지, 제18권, 제2호(2001. 6), p.233.

『표준국어대사전』을 이용한 동등관계의 정의는 정의항의 내용을 이용할 수 있다. 여기에는 동의어로 지시한 50,488개의 표제어와 동의어 안내로 돌린 70,340개의 참조어를 사용할 수 있다. 이 경우 동등관계 지시에 나타난 용어를 기준 용어의 UF관계로 사용할 수 있다. 한색(寒色)의 경우 정의항에 “『명사』 『미술』 차가운 느낌을 주는 색. 파란색 계통의 색이다”라고 설명하고 ≡냉색·찬색과 같이 설명하고 있으므로 한색의 UF로 냉색[冷色], 찬색[--色]을 사용할 수 있다.

이외에도 정의항에 사용된 높임말, 낮춤말, 속어, 음역어, 잘못, 북한어, 방언, 무당용어, 궁중용어, 호 등과 같은 설명들은 동의관계로 사용할 수 있다. 이때 정의항에 설명된 용어를 디스크립터로, 표제어를 비디스크립터로 하여 동등관계를 정의할 수 있다.

유의어의 경우 책자형 『표준국어대사전』에는 표시되어 있으나 웹상 『표준국어대사전』에는 표시되어 있지 않다. 따라서 유의어도 동등관계의 동의 관계로 취급하는 것이 바람직할 것이다. 또한 실제로는 동등관계의 용어이면서 사전상의 표기에는 전혀 관계가 제시되지 않은 용어들은 시소러스의 편찬목적에 따라 동등관계로 정의해야 할 것이다.

예를 들면, 펀치(punch)의 경우 그 하나의 설명에는 『명사』 『1』종이나 기차표, 공작물 따위에 구멍을 뚫어 표를 내는 공구라고 설명하고 있다. 또한 펀치-기(punch器)의 경우에도 『명사』서류철이나 팸플릿 따위를 철할 때에 작은 구멍을 뚫는 기구로 설명하고 있으므로 이 둘은 동의관계로 취급하는 것이 바람직할 것이다. 마찬가지로 퍼린 빛(다소 탁하고 어둡게 푸른 빛)은 푸른빛(맑은 가을 하늘이나 깊은 바다, 풀의 빛깔과 같이 맑고 선명한 빛)으로, 퍼린색[--色](다소 탁하고 어둡게 푸른 색)은 파란색[靑色](가을 하늘과 같이 맑고 선명한 푸른색)으로 동등관계로 연결해 줘야 할 것이다.

#### 나. 줄임말

『표준국어대사전』에는 약 1,011개의 본말과 2,462개의 준말이 있다. 그러나 특활의 경우 정의항에는 특활[特活] “『교육』 ‘특별 활동’을 줄여 이르는 말”이라고 정의하고 있으나 특별~활동(特別活動)은 『교육』“학교 교과 과정에서 교과 학습 이외의 교육 활동. 학생의 개성을 신장하고 건전한 취미와 특수 기능 및 민주적 생활 활동을 육성하기 위한 것으로 학급 활동, 학생회 활동, 클럽 활동, 학교 행사 따위가 있다.”고 설명만 되어 있다. 이 경우는 특별히 본말과 준말의 관계기호를 제시할 수도 있지만, 동의 관계로 처리할 수도 있을 것이다.

#### 다. 남한어/북한어

『표준국어대사전』에는 약 20,883개의 북한말에 대한 국어 대응어가 있다. 가로골 “『북』 『지리』 ‘횡곡(橫谷)’의 북한어”와 같은 경우는 가로골 SK 횡곡[橫谷]으로 정의할 수 있다. 그러나 많은 용어의 경우 북한어라고 표시되어 있지만 남한에서도 함께 사용하는 용어가 있기 때문에 주의가 필요하다.

## 라. 이전/이후용어

이전용어와 이후 용어는 LT/PT로 정의 할 수 있다. 선형 그래프[扇形graph] “『수학』 ‘원그래프’의 전 용어”의 경우는 원그래프[圓graph] PT 선형 그래프[扇形graph]로 서이리안 주[西Irian州] “『지명』 ‘이리안자야 주’의 전 이름”의 경우는 서이리안 주[西Irian州] LT이리안자야 주’와 같이 나타낼 수 있다.

## 2. 계층관계의 적용

『표준국어대사전』의 정의항을 이용한 계층관계는 ‘특별 정기 예금(特別定期預金)’의 정의가 ‘예치 조건이 보통 예금과 다른 정기 예금. 무기명 정기 예금 따위가 있다’고 되어 있기 때문에 ‘정기예금’의 NT는 ‘무기명 정기예금’이 되고 ‘보통예금’은 RT로 정의 된다.

또한 “팔천[八賤] 『명사』 『역사』 조선 시대에, 노비이거나 신분은 양인이지만 천역에 종사하던 여덟 천민. 사노비, 중, 백정, 무당, 광대, 상여꾼, 기생, 공장(工匠)을 이른다.”의 경우에는 팔천[八賤]의 하위개념으로 사노비, 중, 백정, 무당, 광대, 상여꾼, 기생, 공장을 정의할 수 있다.

이외에도 “아래-팔 『명사』 『의학』 팔꿈치부터 손목까지의 부분. ≡ 전박02(前膊) · 전완02(前腕) · 하박02(下膊).”의 경우에는 팔(생물기과) NTP 아래팔과 같이 전체부분관계를 적용할 수 있다. 마찬가지로 절피의 정의가 ‘오리를 먹일 수 있도록 실로 감은 활시위 부분’ 이므로 활시위 NTP 절피가 된다.

그러나 곡분(穀粉)의 경우 정의에는 “『명사』 곡식을 빻거나 갈아서 만든 가루. 쌀가루, 보릿가루, 밀가루 따위가 있다.”라는 정의가 있다면 곡분의 하위개념으로 이들을 사용할 수 있다. 하지만, 곡분에는 이 외에도 메밀가루, 미수가루, 옥수수가루, 완두가루, 잡분, 호밀가루 등의 가루가 있기 때문에 『표준국어대사전』만으로 모든 계층관계의 정의는 불가능하다.

계층관계에서 사례관계의 작성에도 사용할 수 있다. 음수(陰樹)의 정의에는 “음수02(陰樹) 『명사』 『식물』 그늘진 곳에서도 잘 자라고 번식할 수 있는 나무. 전나무, 비자나무, 팔손이나마, 가문비나무, 누운주목, 목란 따위를 이른다. ‘음지나무’로 순화. ≡ 음지나무.”로 되어 있기 때문에 음수의 NTI로 이들 용어를 사용할 수 있다.

## 3. 관련관계의 적용

### 가. 관련관계

『표준국어대사전』의 반의 관계나 정의항의 설명을 이용하여 관련관계를 정의할 수 있다. 황모필[黃毛筆: 족제비의 꼬리털로 맨 붓. ≡ 황필(黃筆)]과 황모[黃毛: 족제비의 꼬리털. 뾰뾰한 세필(細筆)의 붓을 만드는 데 쓴다]는 관련 관계를 갖는다.

### 나. 저자와 저작권계

저자와 저작의 관계는 아래의 카뮈의 용어설명과 카뮈의 소설에서 보는 것과 같이 정의할 수 있다.

카뮈(1913~1960:작가)[Camus, Albert]

정의 『인명』 프랑스의 소설가·극작가(1913~1960). 평론 〈시시포스의 신화〉에서 부조리의 철학을 논하여 실존주의를 더욱 심화시켰고, 전후(戰後)의 사상과 문학에 크게 영향을 끼쳤다. 1957년에 노벨 문학상을 받았다. 작품에 〈이방인〉, 〈페스트〉 따위가 있다.

UF 카뮈, 알베르(1913-1960)[Camus, Albert]

YEA 1913-1960

BTI 소설가

BTI 극작가

TIT 이방인(소설:카뮈)[異邦人]

TIT 페스트(소설:카뮈)[La Peste]

RT 노벨 문학상[Nobel文學賞]

FRA Albert Camus

FRA Camus, Albert

## 4. 부가관계의 적용

### 가. 학명

『표준국어대사전』에는 동식물명에 대한 학명이 포함되어 있다. 이것을 이용해 학명의 관계기호를 사용할 수 있다.

〈예〉 아네모네(anemone) 『명사』 『식물』 미나리아재빗과의 여러해살이풀. 덩이줄기 식물로 높이는 30cm 정도이며, 잎은 우상 복엽이다. 4~5월에 줄기 끝에서 붉은색, 자색, 청색, 흰색 따위의 꽃이 피고 관상용으로 재배한다. 지중해 연안이 원산지이다.(Anemone coronaria)

이 경우 아네모네에 대한 용어관계는 다음과 같이 생성할 수 있다.

아네모네[anemone]

BT 미나리아재빗과[--科]

ENG anemone

SNN Anemone coronaria

## 나. 외래어

『표준국어대사전』에는 표제어 다음에 병기된 한자어를 제외하고도 순수 외래어가 2,527개에 이른다. 이를 이용하여 용어 관계를 표현할 수 있다.

예를 들면 영어인 가이드라인(guide--line)은 가이드라인 ENG guide line과 guide line KEN 가이드라인, 독어인 드라마투르기(°독Dramaturgie)는 드라마투르기 GRE Dramaturgie와 Dramaturgie KGR 드라마투르기, 프랑스어인 데가주망(°프dégagement)은 데가주망 FRA dégageement와 dégageement KFE 데가주망, 라틴어인 레보도파(°라levodopum)는 레보도파 LIT levodopum와 levodopum KLI 레보도파, 이탈리아어인 렐리지오소(°이religioso)는 렐리지오스 ITA religioso와 religioso KIT 렐리지오스, 러시아어인 루바슈카(°러rubashka)는 루바슈카 RUS rubashka와 rubashka KRU 루바슈카와 같은 방법으로 외래어의 용어관계를 정의 할 수 있다.

그러나 외래어 병기에서 라오콘 군상(Laocoon群像)이나 라인 조직(line組織), 몰리브덴비료(°독Molybdän肥料)와 같이 한자어와 외래어가 병기된 경우와 나트륨 램프(°독Natrium°영lamp)과 같이 한자 없이 다른 외래어로 병기된 경우는 외래어 관계적용이 어렵다.

이외에도 『표준국어대사전』이 포함하고 있는 정의항과 분류 및 주제구분을 이용하여 정의와 분류코드를 작성할 수 있다.

## IV. 용어관계 정의의 문제점

### 1. 관련어휘 정보 기술의 문제

『표준국어대사전』은 다국어사전, 분류사전과 같은 특수목적 사전이 아니라 일반사전 혹은 규범사전이다. 김진해는 이 사전의 기본은 인쇄형 사전으로 편집되었기 때문에 편찬 당시에는 단어들의 어휘관계 정보를 사전에 어떻게 체계적으로 기술할 것인지에 대해 깊이 고민하지 않았다고 주장하고 관련어휘 정보 기술의 문제점을 다음과 같이 지적하고 있다<sup>15)</sup>.

① 하위 유형 분류의 비체계성으로 관련어휘를 동의어, 관련어휘, 참고 어휘의 3가지 유형으로 하위분류하고 있는데, 이 하위분류가 형태적 관련과 화용적 관련 및 기타 참조 정보가 혼재되어 있어 체계적이지 못하다.

② 관련 어휘 정보의 편재성으로 하위유형 분류의 비체계성 때문에 3가지 유형의 관련정보가 제시되는 위치가 각기 다르다는 점이다. 특히 뜻풀이에 부분-전체 관계나 동위관계, 상하위 관계에

15) 김진해, 전개서, pp.28-32.

있는 단어가 제시되면 관련어휘는 제시되지 않기 때문에 그 결과 관련 어휘는 뜻풀이와 연동되게 되어 독립성과 체계성을 갖지 못하게 된다.

③ 상호 참조 기능의 미흡으로 일반적으로 상위어의 표제에 대한 하위어 관련어는 뜻풀이에 제시되는 반면 하위어의 경우에는 상위어의 뜻풀이에, 동위어는 참고 어휘에 제시되어 상하위 관계를 이용자 스스로 추측해야 한다.

결국 『표준국어대사전』은 관련 어휘 정보에서 발견 하는 문제점들은 주로 인쇄 사전이 갖는 내적 한계에 기인한다고 볼 수 있을 것이다. 그러므로 여기에서는 이러한 문제들을 기반으로 해서 시소러스 구축시 문제가 되고 있는 내용을 중심으로 기술한다.

## 2. 용어의 분리

『표준국어대사전』의 용어는 동형이의어가 있을 경우 어깨번호로 이들을 구분하고 있다.

<예>

소곰-바치01 『명사』 『방언』 소금을 파는 사람(제주).

소곰-바치02 『명사』 『방언』 ‘사마귀01’의 방언(제주).

소곰-바치03 『명사』 『방언』 ‘노래기01’의 방언(제주).

이 경우는 명확하게 구분되기 때문에 각각의 의미에 따라 동등관계를 정의할 수 있다. 소곰-바치01는 소곰장수로, 소곰-바치02는 사마귀로, 소곰-바치03은 노래기로 USE관계를 사용하여 정의할 수 있다. 그러나 동일한 용어를 번호를 이용해 설명하고 있는 용어에 대해서는 용어의 분리와 한정어 사용이 필요하다.

<예> 손-장난[-짱-] 『명사』

「1」 쓸데없이 손을 놀려서 하는 장난. 「2」 손을 놀려 잔재주를 부리는 간단한 요술.

「3」 ‘노름01’을 달리 이르는 말. 「4」 ‘수음02(手淫)’을 속되게 이르는 말.

「5」 『방언』 ‘소꿍장난’의 방언(경남).

이 경우는 예 「1」 쓸데없이 손을 놀려서 하는 장난은 장난의 NT로, 「2」 손을 놀려 잔재주를 부리는 간단한 요술은 요술의 NT로, 「3」 ‘노름01’을 달리 이르는 말은 USE 노름으로 「4」 ‘수음02(手淫)’을 속되게 이르는 말은 USE 수음으로, 「5」 『방언』 ‘소꿍장난’의 방언(경남)은 USE 소꿍장난으로 용어관계를 정의해야하며, 이 때 각각은 한정어를 사용해서 의미를 구별해야 한다.

또한 단일 용어 설명에서 2가지 이상의 개념을 설명한 경우도 많이 나타나고 있는데, 이 경우에도 개념의 분리가 필요하다. 세이블(sable)의 경우 정의항에 『명사』 검은 담비. 또는 그 털기죽이나 그 털로

만든 화필로 정의하고 있기 때문에 이들은 검은 담비와 화필로 구분해서 용어관계를 정의할 필요가 있다.

세이블(담비)[sable]

1) 검은 담비.

USE 검은담비

세이블(화필)[sable]

1) 검은담비 그 털가죽이나 그 털로 만든 화필.

BT 화필(붓)[畫筆]

RT 검은담비

그러나 전공04(前功)「명사」 「1」 이전에 세운 공로나 공적. ≒전훈02(前勳) 「1」, 「2」 옛사람이 세운 공로나 공적. ≒전훈02 「2」.와 같이 동일한 개념의 용어임에도 분리하여 설명한 경우가 있기 때문에 이 경우는 동일한 용어로 취급할 필요가 있다.

### 3. 동등관계의 설정

『표준국어대사전』에는 동일한 의미이나 유의어 기호를 사용하지 않고 별도의 용어로 취급해서 각각 정의를 하고 있는 용어가 있다. 이들 용어의 경우는 정보검색을 위한 시소러스의 본래 목적을 고려할 때 동등관계로 연결시키고 별도의 정의를 하는 것이 타당할 것이다.

임금의 마음을 나타내는 용어는 용심(龍心: 임금의 마음을 비유적으로 이르는 말), 구중신금(九重宸襟: 『북』 구중궁궐에 있는 마음이라는 뜻으로, 궁중에 있는 임금의 심정을 이르는 말), 룡심(龍心: 『북』 ‘용심’의 북한어), 상심(上心: 임금의 마음), 상의(上意: 임금의 마음. ≒상정), 상정[上情: = 상의(上意)], 성심(聖心: 『북』 임금의 마음을 높여 이르는 말), 신금(宸襟: 임금의 마음), 신려(宸慮: 임금의 뜻. 또는 임금의 마음. ≒신의(宸意)·신지(宸旨)], 신려(神慮: 임금의 마음), 신의[宸意: = 신려(宸慮)], 신지[宸旨: = 신려(宸慮)], 신충(宸衷: 임금의 마음이나 고충), 어심(御心: 임금의 마음을 이르던 말)과 같이 다양하게 나타나 있다. 이 중에서 사전에서 동의어 관계를 제시한 용어는 상정=상의, 신의=신려, 신지=신려와 같이 나타나고 있으며 다른 용어들도 미세한 의미 차이는 있으나 동일한 용어로 볼 수 있음에도 불구하고 별도로 정의하고 있다. 특히 구중신금(九重宸襟)과 룡심(龍心)은 북한어로 표기 되어 있지만 동일한 의미로 볼 수 있을 것이다. 따라서 이 경우는 동등관계로 취급하는 것이 바람직 할 것이다.

마찬가지로 방랑자와 관련된 용어는 표박자, 방랑자, 방랑아와 같은 용어와 『표준국어대사전』에는 없지만 다른 국어사전에 있는 방랑인과 함께 시소러스의 동등관계정의 할 수 있을 것이다.

또한 창갈(槍--: 창과 칼을 아울러 이르는 말)과 검궤(劍戟: 칼과 창을 아울러 이르는 말. ≒검삭(劍槊))과 같은 정의가 포함되어 있기 때문에 이들의 용어관계를 동등관계로 설정할 필요가 있다. 벼이삭의 경우도 나락 모가지[『방』벼의 이삭(충남)], 나락 모개[『방』벼의 이삭(전남)], 벧 모가지[『방』벼의 이삭(충남)], 벧 모개[『방』벼의 이삭(충남)], 회수(禾穗: 벼의 이삭)과 같이 별도로 정

의 되어 있지만 동등관계로 처리할 필요가 있다. 검술사(劍術師)와 검선(劍仙)[검술이 뛰어난 사람], 복수 다리와 복쉬 다리[『방』엉치등뼈가 부서져 상한 말이나 소(제주)]와 같이 동일한 정의를 갖고 있지만 동의 표시를 하지 않고 별도의 정의를 준 용어가 많이 있다.

#### 4. 계층관계의 설정

일반적으로 계층관계는 의미적인 계층관계와 조어적인 계층관계로 나눌 수 있다<sup>16)</sup>. 의미적인 계층관계는 의미를 분석해서 계층관계를 설정하는 것이므로 어려움이 많이 따르지만 복합용어의 조어적 계층관계는 비교적 쉽게 이용할 수 있는 방법이다. 조어적 계층관계의 경우 많은 경우 용어 자체로 계층관계를 정의 할 수 있으며, 의미적 계층관계도 『표준국어대사전』의 용어 정의를 이용해서 정의할 수 있다.

그러나 용어 정의를 이용해 계층관계를 설정할 경우에 발생하는 문제가 많이 나타난다. 가스 지지미[*gas*°일chijimi: 가스사로 짠 쫄쫄한 직물]의 경우 용어 정의에서 나타나는 직물을 상위 개념으로 사용하게 되면 포괄적 정의가 되어 버린다. 따라서 지지미[°일chijimi: 가스사로 짠 면직물의 하나. 신축성이 좋으며 여름옷의 속옷감으로 흔히 쓰는 일본산 베이다. ‘쫄쫄이’로 순화]를 상위개념으로 설정하면 지지미의 정의가 면직물임을 알 수 있다. 이 경우 가스지지미의 상위 개념은 직물이 아니라 지지미이며, 다시 지지미의 상위 개념이 면직물이 된다. 면직물의 상위 개념이 직물(織物)이 되어 정확한 상하위 관계를 형성할 수 있게 된다.

마찬가지로 다체 웅예[多體雄藥: 『식물』 합성 웅예의 하나. 네 개 이상으로 된 수술이다]의 경우 상위 개념이 웅예가 아니라 합성 웅예[合成雄蕊]가 되어야 한다. 또한 상위 개념이면서 어비통[魚肥桶: 『북』물고기를 썩혀 만든 거름을 담는 통]과 같이 거름통을 그름을 담는 통과 같이 표현한 것은 인위적으로 ‘어비통’의 상위관계를 ‘거름통’으로 기술할 수밖에 없다.

#### 5. 한정어의 사용

언어의 중의성은 동형이의어나 다의어와 같이 단어의 형태적 동일성으로 발생하는 것으로서, 동형이의어는 두 개 이상의 단어가 형태는 같으나 의미가 완전히 다른 것을, 다의어는 하나의 단어가 두 개 이상의 의미를 가지는 것을 말한다. 『표준국어대사전』에 대한 분석 결과를 보면, 전체 509,076개의 표제어 중에 22.4%에 해당하는 124,254개가 동형이의어로 나타나고 있다. 30개 이상의 동형이의어를 가지는 단어가 20개나 되며, 심지어 46개의 동형이의어를 가지는 단어도 있다<sup>17)</sup>.

16) 최석두 등, 과학기술분야 시소러스 구축 연구 2007, p.83.

17) 이운영, 전개서, p.30.



시소러스에서와 같이 어원으로 용어를 구분하지 않고 의미(meaning과 sense)에 따라 용어를 구분한다면 동형이의어의 수는 더 늘어나게 된다.

김태수와 최석두는 한정어는 '설명부분의 내용을 대표할 수 있는 가장 간결한 한글용어를 선정하여 한정어로 사용할 것'을 제안하였는데, 여기에서는 '한정어 사용원칙을 상위개념어(BT: broader term)를 사용한다. 쉽게 의미를 파악할 수 있어야 한다. 한정어는 다의성이 없거나 적어야 한다. 다의성이 적은 것 중 가능한 한 짧은 것으로 한정어를 삼는다'<sup>18)</sup>고 주장하고 있다. 그러나 『표준국어대사전』의 경우 앞서 언급한 것처럼 동형이의어가 많기 때문에 간결한 상위개념어로 구분이 어려울 경우가 발생할 수 있다. 그러므로 경우에 따라서는 정의 전체를 한정어로 사용하는 것도 하나의 방안이 될 수 있을 것이다.

예를 들면, 소경(자그마한 경치)[小景], 소경(화창한 봄 경치)[韶景], 소림(쓸쓸하고 적막한 숲)[蕭林], 소림(나무가 듽성듬성 들어서 있는 숲)[疏林], 소우(뚝뚝 성기게 내리는 비)[疏雨], 소우(잠시 동안 조금 내리는 비)[小雨], 소촌(분산되어 있는 마을)[疏村], 소촌(작은마을)[小村], 중욕(많은 사람이 모인 곳에서 모욕을 줌)[衆辱], 중욕(아주 심하게 모욕을 줌)[重辱], 진취(극적으로 나아가서 일을 이룩함)[進取], 진취(일을 차차 이루어 감)[進就], 진심(마음을 다함)[盡心], 진심(속세의 일에 더럽혀진 마음)[塵心], 진심(왈각 성내는 마음)[嗔心]와 같은 것들이다.

그러나 정의가 길어 한정어로 사용하기 곤란한 경우에는 중점(가운데점)[中點], 중점(이등분)[中點], 석룡자(도룡농)[石龍子], 석룡자(도마뱀)[石龍子]과 같이 상위개념의 간략한 용어를 사용할 수 있을 것이다.

또한 수식문[數息門: 육묘문의 첫째 단계. 나고 드는 호흡을 마음속으로 세면서 마음을 고요히 하는 관문이다]과 수식문[隨息門: 『불교』 육묘문의 둘째 단계. 고요한 마음으로 숨을 들이쉬고 내쉬는 것을 의식하며, 마음이 산란하지 않게 하는 관문이다]의 경우와 같이 용어가 동일한 경우에는 수식문(육묘문1)[數息門], 수식문(육묘문2)[隨息門]과 같은 형태로 나타낼 수 있다.

도둔[逃遁: 몰래 달아나 숨음. ㄴ도은(逃隱)·도타(逃躲)·찬닉(竄匿)·찬복(竄伏)·포찬(逋竄)과 같이 동의어가 여러 개 있을 경우 이들의 한정어로는 도은(도둔)[逃隱], 도타(도둔)[逃躲], 찬닉(도둔)[竄匿], 찬복(도둔)[竄伏], 포찬(도둔)[逋竄]처럼 동일한 한정어를 사용할 필요가 있을 것이다.

## 6. 북한어 관계

북한어는 《조선말 대사전》(1992)에 수록된 단어 가운데 남한에서 쓰임이 확인되지 않은 단어와 어문 규정의 차이로 달리 표기하는 단어를 편찬 원칙에 따라 선정하여 수록하였다. 『표준국어대사

18) 김태수, 최석두, "동형이의어 구별을 위한 한글한정어 사용에 관한 연구," 정보관리학회지, 제14권, 제1호(1997, 3), pp.110-119.

전』에는 약 20,883개의 북한말에 대한 국어 대응어가 있다. 가로골 “『북』 『지리』 ‘횡곡(橫谷)’의 북한어”와 같은 경우는 가로골 SK 횡곡[橫谷]으로 정의할 수 있다.

그러나 많은 용어의 경우 북한어라고 표시되어 있지만 남한에서도 함께 사용하는 용어가 있을 뿐만 아니라 북한과 남한의 문법적 차이에 의한 용어가 많기 때문에 단순히 남북한 용어의 관계인 NK와 SK로만 정의하기 어려운 경우가 많다. 따라서 이 경우 남한에서도 사용하고 있는 용어는 동등관계로 취급하여 처리하는 것이 효과적일 것이다.

또한 려과기[濾過器]처럼 『표준국어대사전』에서 북한어로 정의하고 있는 용어 가운데 우리나라에서도 사용되는 용어가 많이 있으므로 이를 우리말로 취급하여 정의하는 것이 필요하며, 들가방[『북』 들고 다니게 된 가방]처럼 북한어로 정의되어 있지만 우리말로도 사용 가능한 용어는 손가방[손에 들고 다니는 작은 가방]의 동등관계로 처리하는 것이 효과적일 것이다.

## V. 결론 및 제언

시소러스 사전의 구축은 많은 시간과 노력을 필요로 하는 끝이 없는 작업이다. 문헌에 사용되고, 이용자가 탐색어로 사용할 수 있는 용어를 가능한 한 많이 모으는 일이 무엇보다도 필요하다. 그러나 시소러스는 단순히 용어를 모으는데서 그치는 것이 아니라 용어들 간의 어의 관계를 구축하는 작업이기 때문에 용어 수집이 무엇보다도 중요하다. 그리고 용어간의 관계를 설정하기 위해서는 무엇보다도 잘 정의된 용어 사전이 절대적인 필요성을 갖는다.

국어사전은 단어 하나하나에 어문 규정을 적용하여 구체적인 언어생활에 기준을 제시해 주는 도구이다. 국립국어연구원은 이러한 규범 사전의 필요성에 따라 1992년부터 1999년까지 8년 동안에 걸쳐 『표준국어대사전』을 발간하였다. 이 사전은 어문 규정을 충실히 반영하였을 뿐만 아니라 통일 이후 남북한 언어의 동질성 회복을 위해 다수의 북한어 표제를 담고 있으며, 옛말, 방언, 그리고 다수의 전문어를 실고 있어 명실 공히 우리나라의 대표 사전이라고 할 수 있다. 이러한 의미에서 국립국어원에서 우리나라 표준국어사전으로 출판한 『표준국어대사전』은 어떤 다른 용어사전보다 중요하며, 시소러스 구축의 기본용어사전으로 사용할 수 있다.

그러나 『표준국어대사전』은 많은 동형이의어를 가지는 단어도 있어서 이들을 구별할 수 있는 방법이 필요하며, 동형이의어가 아니라도 동일한 용어에 여러 가지의 의미를 복합적으로 제시하고 있는 경우가 많아서 시소러스 구축 시에는 이들 개념의 분리가 필요하다.

따라서 본 연구에서는 『표준국어대사전』에서 설명하고 있는 정의항을 이용하여 시소러스를 구축하는 방법을 모색했다. 기본적으로는 많은 부분이 시소러스 용어관계 설정이 그대로 이용될 수 있지만, 많은 부분에서 단순히 『표준국어대사전』의 정의만으로는 용어관계 설정에 문제가 있는 것

으로 나타났다. 특히, 동일한 용어를 번호를 이용해 설명하고 있는 용어에 대해서는 용어의 분리와 한정어 사용이 필요하며, 동일한 의미이나 유의어 기호를 사용하지 않고 별도의 용어로 취급해서 각각 정의를 하고 있는 용어는 동등관계로 연결시키고 별도의 정의를 할 필요가 있다. 또한 계층관계 정의, 한정어 사용, 북한어 문제 등도 고려해야할 문제점이 있는 것으로 나타났다.

보다 광범위한 개체명 인식과 다의성 해소를 위해서는 시소러스는 모든 분야의 용어가 망라적으로 수집될 필요가 있다. 이 작업은 새로운 용어가 등장하는 한 계속되어야 하며, 가능하다면 정교한 알고리즘에 의한 자동 구축이 필요할 수도 있다. 그러나 현재까지의 연구 결과로는 시소러스의 자동 구축은 현실적으로 어려운 것으로 나타나고 있다. 그러므로 사전과 같이 정의된 다양한 도구에 의해 도움을 받아 수작업으로 진행하는 것이 현재로서는 가장 바람직 할 것으로 보인다. 이러한 측면에서 『표준국어대사전』은 시소러스 구축의 기본 용어로 사용될 수 있을 것이다.

또한 『표준국어대사전』도 단순히 용어를 설명하는 도구에서 탈피하여 어휘관계를 하이퍼링크를 통해 연결시킨다면 보다 효과적인 용어 검색이 가능해질 것이며, 반대로 이 사전 전체 내용을 시소러스로 구축한다면 보다 정교한 우리말 시소러스가 될 수 있을 것이다.

## 참고문헌

- 국립국어원. 표준국어대사전. 서울 : 두산동아. 1999.
- 김지훈, 김태수. “용어정의와 관계추출을 통한 시소러스 확장에 관한 연구.” 한국문헌정보학회지, 제40권, 제1호(2006. 3), pp.293-314.
- 김진해. “표준국어대사전의 관련어 정보와 어휘관계 기반 사전 기술.” 한국어 의미학, 제24권 (2007), pp.23-50.
- 김태수, 최석두. “동형이의어 구별을 위한 한글한정어 사용에 관한 연구.” 정보관리학회지, 제14권, 제1호(1997. 3), pp.107-124.
- 김태수. “용어정의를 도입한 시소러스 개발 연구.” 정보관리학회지, 제18권, 제2호(2001. 6), pp.231-254.
- 남영준. “지식관리시스템을 위한 의미형 한영 시소러스 구축에 관한 연구.” 한국도서관·정보학회지, Vol.32, No.4(2001), pp.77-98.
- 서 휘. “클러스터링을 이용한 시소러스 브라우저의 설계에 대한 이론적 연구.” 한국도서관·정보학회지, Vol.30 No.3(1999), pp.427-456.
- 우리말 시소러스<<http://www.ontorus.net>> [인용 2013. 10. 31].
- 유영준. “SKOS를 이용한 신학 시소러스의 온톨로지로의 변환에 관한 연구.” 한국도서관·정보학회지, Vol.43 No.3(2012), pp.143-163.

- 이승민. “시소러스 통합을 위한 개념기반 패시 프레임워크 구축.” 한국도서관·정보학회지, Vol.41 No.3(2010), pp.269-290.
- 이운영. 표준국어대사전 연구 분석. 서울 : 국립국어원, 2002.
- 장인호. “『뉴스 코어 시소러스』의 구축 및 활용 방안에 관한 연구.” 한국도서관·정보학회지, Vol.44 No.3(2013), pp.489-512.
- 정호성. “〈표준국어대사전〉수록 정보의 통계적 분석.” 새국어생활, 제10권, 제1호(2000. 봄), pp.55-72.
- 최석두 등. “대단위 우리말 온톨로지 구축을 위한 시소러스의 개발.” 정보관리학회지, 제23권, 제4호(2006. 12), pp.147-164.
- 최석두 등. 과학기술분야 시소러스 구축 연구. 대전 : 한국과학기술정보연구원, 2007.
- 최석두, 한상길. “국립중앙도서관을 위한 시소러스형 주제명표목표의 개발.” 한국시소러스연구회 포럼(2003. 12), pp.63-76.
- 최석두. 한글 시소러스 구축 기준에 관한 연구. 박사학위논문, 연세대학교 대학원, 2002.
- 표준국어대사전<<http://stdweb2.korean.go.kr/>>[인용 2013. 10. 31].
- 한상길, 최석두. “한글시소러스에서 저자와 저작에 대한 관계 설정과 동형이의어의 기술.” 한국문헌정보학회지, 제45권, 제4호(2011. 12), pp.139-155.
- 한상길. 시소러스 용어관계 확장에 관한 연구. 박사학위논문, 중앙대학교 대학원, 1999.
- 한상길. “시소러스의 기본 용어관계에 관한 연구.” 한국도서관·정보학회지, Vol.30, No.4(1999), pp.107-136.
- Aitchison, J. and A. Gilchrist. *Thesaurus Construction : a practical manual*, 2nd ed. London: Aslib, 1987.
- ISO 2788:1986(E) *Documentation - Guidelines for the establishment and development of monolingual thesauri*.
- Soergel, D. *Indexing Languages and Thesauri : construction and Maintenance*. Los Angeles: Melville, 1974.

#### 국한문 참고문헌의 영어 표기

(English translation / Romanization of references originally written in Korean)

- Chang, Inho. “A Study on the Establishment and Applications of the ‘News Core Thesaurus.’” *Journal of Korea Library and Information Science Society*, Vol.44, No.3(2013), pp.489-512.

- Choi's Thesaurus Home page. <<http://www.ontorus.net>> [cited 2013. 10. 31].
- Choi, Sukdoo & Han, Sangkil. A Study on Development of National Library Subject Headings. *International Workshop of the Korean Thesaurus Association*(2003. 12), pp.63-76.
- Choi, Sukdoo et al. "Constructing a Grand-Scale Korean Thesaurus as Base for Ontology Development." *Journal of the Korean Society for Information Management*, Vol.23, No.4(2006. 12): pp.147-164.
- Choi, Sukdoo et al. *Thesaurus construction in the field of science and technology research*. Daejeon : KISTI, 2007.
- Choi, Sukdoo. *A Study on Guidelines for the Construction of Korean Thesaurus*. Doctoral Thesis, Yonsei University, 2002.
- Han, Sangkil & Choi, Sukdoo. A Study on the Description of Relationships and Homographs in Terms of Creator and Work in the Korean Thesaurus. *Journal of the Korean Society for Library and Information Science*, Vol.45, No.4(2011. 12), pp.139-155.
- Han, Sangkil. *A Study on the Expansion of the Thesaurus Term Relationship*. Doctoral Thesis, Chungang University, 1999.
- Jeong, Hosung. "Statistical analysis of the information contained in the Korean Standard Dictionary." *New Korean Life*, Vol.10, No.1(2000, Spring), pp.55-72.
- Kim, Ji-Hun & Kim, Tae-Soo. "A Study on Thesaurus Expansion through Definitions of Terms and Extraction of Relationships." *Journal of the Korean Society for Library and Information Science*, Vol.40, No.1(2006. 3), pp.293-314.
- Kim, Jin Hae. "Description of the relationship of lexical information and related words on the 'The Korean Standard Dictionary'", *Journal of the Society of Korean Semantics*, Vol.24(2007), pp.23-50.
- Kim, Taesoo & Choi, Sukdoo. "A Study on Hangul qualifier for homographic descriptors." *Journal of the Korean Society for Information Management*, Vol.14, No.1(1997. 3), pp.107-124.
- Kim, Taesoo. "A Study on the Development of Thesaurus Using Terminological Definitions." *Journal of the Korean Society for Information Management*, Vol.18, No.2(2001. 6), pp.231-254.
- Korean Standard Dictionary Home page. <<http://stdweb2.korean.go.kr/>> [cited 2013. 10. 31].
- Lee, Seungmin. "Construction of the Concept-Based Faceted Framework for Thesaurus

Integration.” *Journal of Korea Library and Information Science Society*, Vol.41, No.3(2010), pp.269-290.

Lee, yoonyoung. *A Study on the Korean Standard Dictionary*. Seoul : The National Institute of Korean Language, 2002.

The National Institute of Korean Language. 1999. *The Korean Standard Dictionary*. Seoul : Doosandong. 1999.

Yoo, Yeong Jun. “A Study on Converting the Theological Thesaurus to the Ontology by Using SKOS.” *Journal of Korea Library and Information Science Society*, Vol.43, No.3(2012), pp.143-263.