

대등관계 설정의 확장 지침에 관한 연구*

A Study on the Extend Guideline for the Equivalence Relationship in Thesaurus

남 영 준(Young-Joon Nam)**

초 록

시소러스가 적절한 검색효율을 유지하기 위해서는 등재된 디스크립터에 대한 지속적인 유지관리가 필수적이다. 시소러스의 규모를 적절하게 유지하기 위해서는 신규용어와 기존용어를 우선적으로 대등관계로 처리해야 한다. 따라서 전통적인 동의어 이외에도 여러 관계에 대한 새로운 대등관계설정기준이 요구된다.

본 연구는 기존의 대등관계 설정기준을 좀 더 세분하여 보다 효율적인 시소러스 역할을 수행할 수 있는 방안을 제시하였다. 대등관계로 설정할 수 있는 기준을 전통적인 동의어 관계를 비롯하여, 유사동의어, 전거어, 대립어 등도 대등관계로 설정될 수 있음을 제안하였다. 본 연구에서 제안한 대등관계는 여섯가지 국면(facet)으로 기준을 정하였으며, 이 기준은 검색효율을 위해 용어 관계의 재정립을 위한 지침으로 활용될 수 있다.

ABSTRACT

For the guarantee of retrieval ratio, thesaurus's maintenance for descriptors are necessary. To maintain the optimum scale of thesaurus, new terms and existing terms should be structured to the equivalence relationship. Therefore, equivalence relationships are needed to new standard. This study proposes new standard of the equivalence relationships, which is more specified for better guarantee of retrieval ratios. The relationships have seven facets. These six facets will be used as new knowledge-base, which could be reestablished between the descriptors.

키워드: thesaurus, equivalence relation, thesaurus guideline, thesaurus construction, thesaurus maintenance

* This work was supported by grant No. (R01-2003-000-11588-0) from the Basic Research Program of the Korea Science & Engineering Foundation.

** 중앙대학교 문과대학 문헌정보학과 부교수(namyj@cau.ac.kr)
Associate Professor, Dept of Library & Information Science, Chung-Ang University

- 논문접수일자 : 2004년 4월 1일
- 게재확정일자 : 2004년 6월 1일

1. 서론

시소러스의 유지보수 작업은 채택된 디스크립터에 대해 실제 개념과 사용 용어를 일치시키는 작업이다. 대부분의 시소러스가 역동 시소러스를 천명하는 이유는 용어에 대한 개념과 디스크립터의 관계를 실시간적으로 반영하여 최적의 검색효율을 얻기 위함이다.

한편 시소러스의 디스크립터간 관계 설정 가운데 계층관계나 연관관계에 대한 수정은 기존의 검색지식에 대한 전면적인 수정이 필요하기 때문에 이용자의 혼란을 초래할 수 있다. 이에 비해 대상용어를 기존의 등재된 디스크립터에 대한 비디스크립터로 처리하고 그 용어를 적절한 관계로 사후에 설정하는 방식이 이용자의 혼란을 최소화할 수 있다. 즉 등재 후보 디스크립터에 대한 검증을 거친 후에 디스크립터로 전환하는 방식이 보다 안정적인 시소러스 구축에 유리할 것이다.

본 연구에서는 이 점에 착안하여 시소러스 구축에 있어 대등관계를 유연하게 설정할 수 있는 기준의 새로운 지침을 제안하고자 한다. 유연한 대등관계는 디스크립터화할 수 있는 많은 용어들을 특정 개념어의 비디스크립터로 수용할 수 있기 때문에 시소러스 규모의 증가없이 신규용어나 개념을 시소러스에 반영시킬 수 있다. 한편 디스크립터의 유지보수 차원에서 기존 용어를 비디스크립터로 처리하는데 따른 혼란을 최소화할 수 있다. 즉, 기존 디스크립터를 일방적으로 삭제할 경우에 대체어나 관련어를 찾기 위한 이용자 검색의 혼란을 최소화할 수 있다. 이러한 장점 때문에 국내외의 주요 시소러스 개발지침은 기본적으로 대등관계로 설

정될 수 있는 용어 관계를, 일반 동의어를 비롯하여 계층어로 설정될 수 있는 '업워드 포스팅(upward postings)', 유사동의어 등으로 제한하고 있다. 한편 유사동의어나 유의어의 경우에도 상황에 따라 대등관계 혹은 연관관계로도 처리할 수 있음을 권고하고 있다. 이와 같이 여러 지침들은 상황에 따라 용어간의 관계를 임의적으로 구조화할 수 있도록 제한하고 있다.

본 연구에서는 시소러스 개발을 위한 주요 지침에서 제안하는 대등관계의 지침이외에 보다 포괄적으로 정의되어야 할 대등관계 지침을 제시하고자 한다. 또한 이러한 대등관계 설정 지침이 연관관계나 계층관계의 용어를 설정할 수 있는 지에 대한 이론적 관점을 제시하고자 한다. 궁극적으로 본 연구에서는 시소러스의 관계설정에 따른 애매성을 극복할 수 있는 확장된 새로운 대등관계 설정기준을 제시한다.

2. 시소러스 관계 설정의 지침

시소러스의 관계 설정에 대해서는 국제 및 국가 표준으로 개발지침에 대한 권고안의 형태로 제안되어 있다. 또한 국내의 시소러스는 이러한 권고안에서 제안하는 지침의 일부만을 수용하고, 많은 부분을 개발 주체기관에서 필요한 특성에 따라 자의적으로 반영하고 있다. 따라서 전자의 경우를 준거적 지침이라면, 후자의 경우는 실제적 지침이라 할 수 있다.

2.1 준거적 지침

국내외 시소러스 구축지침안에서 제안하는

관계설정 방법에 관한 방법은 약간의 차이는 있지만 기본 원칙은 디스크립터의 특성을 고려하여 계층관계와 연관관계, 유사관계 등 세 개의 관계로 설정하는 것이 기본이다.

2. 1. 1 국내 표준

시소러스 구축에 대한 우리나라 국가 표준으로는 한국데이터베이스진흥센터(Korea Database Promotion Center)에서 시소러스 개발지침(한국데이터베이스진흥센터 2000)을 제시하였다. 이 지침에서는 ISO 2788에서 제안하는 관계설정의 유형과 특성을 고려하여 시소러스 관계를 다음과 같이 등가관계와 계층관계, 관련관계 등 세가지 요소로 구분하여 설명하고 있다.

1) 대등관계: 대등관계는 등가관계로도 표현되며, 이 관계는 색인작업시 복수의 용어가 동일개념을 나타낸다고 인정되는 경우에 우선어 및 비우선어간의 관계이다. 따라서 우선어와 비우선어는 등가어의 집합을 형성한다. 책자형 시소러스에서의 자모순 배열은 보통은 우선어와 비우선어를 인쇄 서체로 구별한다. 이에 속하는 것으로는 유사어와 유사동어, 업워드포스팅으로 구분한다. 이 가운데 동의어는 광범위한 문맥에서 의미가 동일하다고 인정되기 때문에 실제로 교환 가능한 용어로 정의된다. 동의어는 일상적인 언어에는 그다지 출현하지 않으나, 과학분야를 비롯한 주제분야 전체에 걸쳐서 자연언어보다 통제된 언어 쪽에 동의어가 많다. 이것은 통제언어에서도 의도적으로 용어의 의미를 제한하기 때문이다. 유사동의어는 근사동의어라고도 하며, 일반적으로

의미가 상이라고 인정되나, 색인작업에서는 동의어로 취급하는 용어라고 정의된다. 유사동의어에는 의미가 현저하게 중복되는 용어를 포함한다. 업워드 포스팅은 하위개념어를 상위개념어의 종(種)의 관계보다 등가 관계로 취급하는 방법이다. 그 목적은 어휘의 수를 감소시키고 동시에 하위개념인 특정용어에서 하위개념을 나타내기 위하여 이용하는 상위개념어에 액세스하기 위한 방법을 확보하는데 있다.

2) 계층관계: 계층관계는 시소러스 디스크립터 용어를 상·하위 개념으로 연결시켜주는 관계를 말한다. 상위개념은 클래스 또는 전체를 나타내며, 하위개념은 한 요소 또는 일부분을 나타내는 것이다. 계층관계는 상위 및 하위개념을 논리적으로 전개하는 순서를 나타낸다. 이에 속하는 관계는 ‘속(屬)관계’를 비롯하여 ‘계층적인 전체 - 부분관계’, ‘사레관계’ 등 세가지를 제시하고 있다.

3) 연관관계: 연관관계란 계층적이 아니지만 개념적으로 밀접하게 관련되어 있으나, 등가집합에는 포함되지 않는 용어간의 관계이다. 관련관계에 있는 용어 즉, 관련어는 색인작업과 탐색에 이용될 가능성이 있는 대체용어를 제시한다는 관점에서 볼 때, 용어간의 연결을 시소러스 중에서 명시하는 것이 보다 좋을 것이라는 심리적인 연상에 따라 만들어진 것이다. 이 관계는 상호보완적인 관계로 설명하고 있다. 이 관계에는 속하는 것으로는 ‘동일 범주에 속하는 용어’를 비롯하여 ‘상이한 범주에 속하는 용어’로 두 개의 경우를 열거하고 있다.

2. 1. 2 국제 표준

1) ISO 가이드라인

시소러스를 구조화하기 위한 대표적인 국제 가이드라인은 ISO에서 제정한 ISO 2788(ISO 1986)과 ISO 5964(ISO 1985)가 있다. 전자는 단일어로 구성되는 시소러스를 구성하는데 원용된 표준이며, 후자는 다국어 시소러스를 구성하는데 원용된 표준이다.

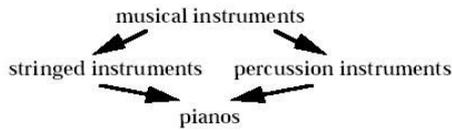
- 유사관계 : 이 관계는 동의어와 유사동의어로 구분할 수 있다. 이 관계는 색인어 선정에 있어 어느 것을 선정해도 무리가 없는 관계의 용어들을 군집화할 때 필요한 용어간 관계로 정의한다. 다국어 시소러스의 경우, 디스크립터로 선정되는 대표어와 이에 대응하는 외국어를 비디스크립터어로 선정할 경우가 이 관계에 해당한다.
- 계층관계 : 이에 속하는 관계로는 ISO 2788과 ISO 5964에서 제시하는 두가지 이론에 속하는 경우에만 활용한다. 하나는 속관계이며, 하나는 계층개념으로 전체와 부분관계로써 용어의 동일한 범주에 속하는 것이다. 예를 들면 지리적인 공간 개념과 신체와 신체 일부와의 관계 등이 이에 속한다.
- 연관관계 : ISO의 가이드라인에서도 연관 관계 설정이 다른 관계의 개념적 설정이 상대적으로 어렵다고 기술하고 있다. 따라서 연관관계에 대한 분명하고 주관적인 기준을 제시하기보다는 유사관계나 계층관계에 속하지 않는 용어간 관계들은 대부분 연관관계에 속한다고 정의하고 있다.

2) ANSI의 가이드라인

미국 국립표준원(American National Standard)에서 단일어 시소러스 구축을 위해 제안한 가이드라인으로써 1980년에 UNESCO 가이드라인의 기초가 된 ANSI/NISO Z39.19을 기반으로 작성한 것이다. 이는 2003년에 디지털 정보자료 검색을 위해 많은 부분을 수정하여 개정판(NISO 2003)에서 제안하였다. 이 가이드라인에서의 가장 큰 변화는 전체-부분 관계를 연관관계보다 계층관계적 시각으로 처리하도록 권고한 점이다. 기본적인 관계로는 다른 가이드라인과 같이 대등관계를 비롯하여 연관관계, 계층관계로 구분하고 있으나 해당 관계에 속하는 사례부분에서 의미적 연관성을 중시하여 다른 개념어와 아무런 관계가 없는 고립어를 인정하지 않고 있다.

- 대등관계 : 동일한 개념을 두 개 이상의 단어로 표현하는 용어간의 구조화에 사용되는 관계이다. 이 관계에 속하는 사례로써는 동의어관계와 어휘파생어(Lexical Variants)로써 이형동의어의 범주를 포함한 구체적인 관계, 유사어, 복합명사가운데 상호참조가 가능한 용어관계로 확대하여 제안하고 있다.
- 계층관계 : 이 관계에 속하는 것은 용어간 소속이나 포함범주에 의해 분명하게 상위 개념어와 하위개념어로 구분되는 관계이다. 이 관계에는 속관계를 비롯하여 전체와 부분관계,¹⁾ 사례관계, 다계층관계(Poly-hierarchical Relationships)로 제안하고 있다. 특히 다계층관계는 하나의 개념이 시소러스내에서 복수의 계층관계를 갖는 경우이다. 다음 <그림 1>은 피아노가

음악 시소러스내에서 현악기와 타악기라는 복수의 상위어를 가지며 현악기와 타악기는 악기라는 단일 상위어를 갖는 예이다.



〈그림 1〉 다계층관계의 예

- 연관관계 : 이 관계는 대등관계와 계층관계에 속하지 않는 관계로써 개념의 크기와 용어 수준이 대칭이 되는 관계에 속하는 용어간 관계가 이에 해당한다. 이 관계로 구조화하기 위한 우선 조건으로 다음과 같은 경우를 제안하고 있다.
 - 동일한 범주에 속해있는 용어간에 설정할 수 있다.
 - 다른 범주에 속해있으면서 동시에 출현하여 색인과 검색이 도움이 되는 용어간에 설정할 수 있다.
 - 관련어의 노드레이블(node labels)로 검색이나 색인에 연상이 되는 용어간에 설정할 수 있다.
 - 기계가독형 색인어집에서 특정 의미를 갖는 특정기호로 처리될 수 있는 용어들을 연관관계로 설정할 수 있다. 예를 들면, 국가코드가 붙는 용어는 해당 국가코드를 기준으로 서로 연관관계가 설정될 수 있다.

3) UNESCO 지침

UNESCO 시소러스 구축 지침은 SC/76/WS/555의 구축 원칙을 확대 개선한 것이다. 이 지침은 오스틴(D. Austin)과 데일(Peter Dale)이 ISO 2788을 개발하기 이전에 영국도서관에서 시소러스를 우선 구축할 수 있는 기준으로 활용하였다. 이 가이드라인에서는 디스크립터의 형태에 관한 것을 주로 제안하고 있으며, 해당 디스크립터 관계설정은 크게 다음과 같은 네가지 관계를 제안하고 있다(UNESCO 1981).

- 대등관계 : 이 관계는 우선어와 비우선어간의 관계로써 복수의 용어가 이 관계로 설정될 수 있다. 이에 속하는 관계로는 동의어를 비롯하여, 유사동의어, 업워드포스팅이 있다. 이 가운데 업워드 포스팅 관계는 색인어를 고려한 관계로써 검색상황에 따라 계층관계의 용어도 이에 속할 수 있음을 보여주는 것이다.
- 계층관계 : 이 관계는 비구조화된 용어군 가운데 가장 분명하게 계층관계를 설정할 수 있는 용어들의 관계이다. 이 관계는 최상위어를 선정하고 해당 최상위어에 속해있는 용어들을 파생하위어(subordinate terms)로 선정하는 방법을 택한다. 이에 속하는 관계로는 ‘속성관계’를 비롯하여 ‘계층적인 전체 - 부분 관계’, ‘사례관계(instance relationships)’를 제안하고 있다
- 연관관계 : 이 관계는 대등관계나 계층관계에 속하지 않는 용어 가운데 서로 근접화하여 색인이나 검색에 도움이 되는 관

1) 용어간에 계층의 개념이 뚜렷하게 존재하지 않아도 전체와 부분으로 표현될 수 있는 용어는 계층관계로 설정하고 있다.

계를 의미한다. 이 관계는 다른 용어를 의미적으로나 계층적으로 대체하지 못하는 관계이다. 이 관계에 속하는 것으로는 '동일한 카테고리에 속해 있는 용어'와 '다른 카테고리에 속해 있는 용어'로써 서로 관련성이 있는 경우를 제안하고 있다. 이 관계를 일련의 제안된 형태로 표현하는 것은 어려움이 있기 때문에 이 가이드라인에서는 9가지의 구체적인 예를 들어 설명하고 있다.

- 상호범주 관계 (Inter-category relationships) : 이 관계는 일련의 용어들이 속하게 되는 범주가 복수에 해당하는 용어들이 이에 해당한다. 이러한 관계가 설정될 수 있는 것으로는 다음 두가지²⁾를 설정하고 있다.
 - 특정 주제 분야나 기본이론들 사이에서 기본이 되는 대상
 - 패싯체계에서 기본이 되는 대상

2. 2 실제적 지침

실제적 지침은 국가 수준의 준거적 지침이라 할 수는 없으나 주요 기관에서 자체적으로 활용한 시소러스의 관계설정 방법이다. 본 연구에서는 국내에서 개발된 법원도서관 시소러스와 고속철도건설공단의 시소러스를 구축하는데 적용한 실제적 지침을 조사한다.

2. 2. 1 법원도서관

대법원 도서관은 국내 주요 법률정보를 효

율적으로 검색하기 위한 지식베이스로써 1998년에 법원도서관 시소러스(한국정보관리학회 1998)를 개발하였다. 이 시소러스의 개발에 적용할 디스크립터에 대한 관계 설정방법은 ISO의 권고안 가운데 기본적인 관계인 계층과 연관, 대등관계로 단순화한 방법을 사용하였다. 법률 문헌의 특성상 한글 디스크립터와 대응되는 영어, 불어, 독일어, 일본어 등과 같은 외국어 디스크립터를 대등관계로 설정하여 비우선어로 채택하였다. 다음은 대법원 시소러스 구축에 채택한 관계설정의 기준 방법을 설명한 것이다.

1) 계층관계: 계층관계는 선정된 디스크립터간 상·하위 개념이 분명한 것을 대상으로 하였다.

2) 연관관계: 연관관계는 선정된 디스크립터간에 상·하위적인 관계는 아니나 용어간 관계가 분명한 것을 대상으로 하였다. 특히, 상·하위적인 관계가 분명하지 않은 것은 연관관계로 표시하였으며, 이에 대한 최종적인 판정은 전문위원회와 별도의 법원 전문가탐에 의해 결정하도록 하였다.

3) 대등관계: 대등관계에 설정된 용어는 동일한 개념이나 다른 용어로 표기 또는 사용되는 디스크립터를 대상으로 하였다. 특히 법률 자료의 특성상 한글에 대응되는 외국어 단어를 비우선어로 처리하였다. 이 지침에서 대등관계의 수준을 가장 일반적인 관계로 정의하여 특정 용어간 대치가 이루어져도 이용자에게 개념적으로 잡음을 최소화하는 용어들간의 관계만을 대등관계로 처리하였다. 따라서 색인작업시

2) - primary organization into fields of disciplines
- primary organization by facets

복수의 용어가 동일개념을 나타낸다고 인정되는 경우에 우선어 및 비우선어간의 관계로 하였다.

2. 2. 2 한국고속철도건설공단

한국고속철도건설공단(KTX: Korean Train Express)은 2002년에 해외의 기술이전 자료와 자체 생산된 문헌 및 보고서, 문서 등 고속철도건설사업과정에서 생산된 각종 자료를 효과적으로 검색할 수 있도록 KTX 시소러스를 개발하였다. 특히 동 시소러스는 건설 및 토목, 전기, 기계 등과 같은 주요 분야의 자료를 모두 감당할 수 있는 수준으로 개발되었다. 동 시소러스의 관계설정 방법은 다음과 같이 국내외의 준거적인 지침에서 제안하는 기본적 원칙을 사용하였다(쓰리소프트 2002).

1) 계층관계: 일련의 유사 용어군에서 용어간 개념 유사성을 유지하고 있으며, 개념의 크기가 서로 차이가 있고, 그 차이를 계층으로 구분하여 열거함으로써 이용자의 검색효율성에 도움이 될 수 있는 관계이다.

2) 연관관계: 연관관계는 실질적으로 용어의 개념간 관련성을 확인할 수 있으나 계층관계나 혹은 대등관계에 속하지 않은 관계를 의미한다. 특히 상위어가 같은 경우에는 해당 용어는 연관관계에 속할 수 있으며, 이 때에는 반드시 용어의 수준이 동일해야 한다.

3) 대등관계: 대등관계에 속할 수 있는 개념으로는 동의어(이형동의어 포함)와 유사어, 계층관계와 연관관계로 설정하기 힘든 관련어들이 모두 포함될 수 있다. 특히 유사어의 개념은 전거어와 매우 유사하다. 그러나 전거어는 전거파일에 속한 용어의 경우에는 거의 지

속적으로 전거파일에 속할 수 있으나 시소러스에서 사용되는 대등어사전(혹은 유사어 사전)은 주제적 특성과 정보의 생산성에 따라 비디스크립터에서 정식 디스크립터로 변환이 가능하며, 모용어(母用語)와 별도의 관계가 설정될 수 있다. 따라서 유사한 개념을 갖고 있으면서 상하관계나 연관관계로 표기하기 부적합한 개념의 용어들도 대등관계로 표기할 수 있으며, 이 관계는 언제든지 변화될 수 있는 관계이다.

3. 대등관계의 유형

시소러스는 실제 이용자들이 정보를 검색하는 도구이기 때문에 이용자 수준과 인지능력이 반영되어야 한다. 따라서 디스크립터간의 논리적 대등관계와 실제 사용되고 있는 용어사이에 실제적 대등관계를 수용할 필요가 있다. 따라서 관계설정의 기준은 시소러스의 특성과 함께 디스크립터가 내포하고 있는 의미를 직관하는 이용자나 색인자의 관점에 따라 변화될 수 있다. 또한 사람들이 인지하고 있는 수준과 환경에 따라 동일하다는 개념의 범주와 범위가 변화될 수 있다.

따라서 시소러스의 관계 설정가운데 대등관계를 설정하는 기준과 연관관계 및 계층관계를 설정하는 기준을 구별하기가 용이하지 않다. 왜냐하면 ISO를 비롯하여 각종 기준으로 제시한 관계 설정 방침에서 대등관계 설정 기준과 연관관계 설정 기준이 애매한 경우가 있기 때문이다. 예를 들면, ISO 2788에서 연관관계에 속하는 범주를 다음과 같이 제시하고 있다.

같은 범주에 속하는 용어는 연관관계를 설정

한다.³⁾ 연관관계에 대한 설명 예는 ‘달 것(vehicles)’에 배정된 배(ships)와 보트(boats)를 제시하였다. 한편 대등 관계에서 유사동의어의 경우에도 동일한 개념 하에 두개의 부속물로 이루어진 복합세탁기(washing equipment)의 세탁기(washer)와 건조기(dryer)는 앞의 연관관계로도 다음과 같이 설정이 가능하다.

대등관계	연관관계
washing equipment	washer
uf washer	bt washing equipment
dryer	rt dryer

즉 대등관계를 설정하는 것과 연관관계를 설정하는 것, 확장하여 처리하는 것도 고려할 경우에는 계층관계까지도 정확한 관계설정 기준을 제시하는 것이 불가능할 수 있다.

이러한 애매성이 발생하는 것은 용어간의 관계설정은 단순한 외형적인 형태만으로 용어를 구조화하는 것이 아니라 용어가 갖고 있는 의미로 구조화하는 작업이기 때문이다. 즉, 같다는 개념을 단순히 동일한 물체에 대해 이형적인 표현으로 한정할 경우에는 애매성이 급격히 사라지지만 유사물체에 유사개념을 모두 같다는 것으로 처리하는 것은 개념의 다의성과 다의적인 해석 때문에 이를 정확하게 분류하지 못하게 된다.

3. 1 인지적 대등관계

언어학적으로 대등어나 동의어, 유사어, 유사동의어는 의미적으로 최소한 차이가 있다.

대등어는 용어 수준이 동일하다는 것을 의미하고, 동의어는 동일하거나 유사한 개념을 나타내는 단어들의 쌍으로 규정되고 있다. 이 때 동의어의 의미는 문맥을 통해 파악되는 양상으로 이해하는 것이 일반적이다. 특히 구체적인 사물에 대한 지시적 차원에서는 어느 정도 의미적 동의성을 인정할 수 있지만, 추상적인 개념의 차원에서는 절대적 동의성이 인정되지 않는다(이종열 2000). 인지적 과정에서 대등이라는 개념은 외부적 환경에 따라 절대적 동의어와 상대적 동의어로 구분이 될 수 있다. 전자의 경우는 완전한 동의어로 표현될 수 있으며, 후자의 경우는 부분적 동의어로 표현될 수 있다.

3. 1. 1 완전한 동의어 관계

완전한 동의어는 둘 이상의 단어가 동일한 대상을 지시할 경우에 해당한다. 이러한 것은 외부적 환경과 문화적 환경에 따라 어원의 변화에 의해 이루어지는 관계가 이에 해당한다. 따라서 환경의 변화에 따라 의미의 차이가 발생할 수 있기 때문에 완전한 대등관계라는 것은 현실적으로 불가능하다(이종열 2000, Seetharama 1975). 이는 동일한 용어에 대해서도 사용자의 환경에 따라 느낌이 달라질 수 있기 때문에 서로 다른 형태의 용어간 완전한 동의어관계는 상대적으로 어렵다는 의미이다. 완전한 동의어에 대해 가장 대표적인 것은 외래어 혹은 대응 외국어 등이 있다. 국내에서 개발된 대부분 시소러스는 한글 디스크립터와 함께 한글표기 디스크립터와 외국어 디스크립터를 병기하고 있어 대표적인 완전한 동의어 관

3) ISO 2788: 17에서 Terms belonging to the same category

계라 할 수 있다. 또한 인지적 판단에서 완전한 동의어 관계를 형성하는 것으로 ‘존대말과 평어, 비속어’를 비롯하여, ‘표준어와 방언’, ‘현대어와 고어’ 등의 관계를 완전한 동의어 관계로 간주할 수 있다.

- 예) 한글과 외국어 : 도서관 - library
 존대말과 평어 : 아버지 - 아빠
 표준어와 방언 : 명계 - 우렁챙이
 현대어와 고어 : 식사 - 수라

3. 1. 2 부분적 동의어 관계

부분적 동의어는 명사와 같이 대상물이 정확하게 지칭되는 경우보다 지칭하기 어려운 추상적 대상물에 대해 이루어진다. 개념을 표현하는 대부분의 용어들은 대상물 자체가 추상적이기 때문에 인지적인 측면에서 완전하게 동의 관계가 성립할 수 없다. 따라서 유사개념의 구조로써 모호성과 유사성을 동시에 갖는 대상물을 지칭하는 경우가 부분적 동의어 관계로써 이에 해당한다. 이 관계는 전적으로 해당 용어를 받아들이는 인간의 인지능력에 좌우되기 때문에 인지자의 경험이나 지식 등의 차이에 따라 대상물의 의미가 크게 좌우될 수 있다. 즉, 문맥이나 환경과 같은 외부적 요인에 기인하여 완전한 동의어간에도 ‘개념화자(색인자)’와 ‘개념 해석자(이용자)’, 그리고 해당 용어가 지시하는 ‘대상 혹은 개념의 세가지 요소에 따라 최소한 의미 변이가 발생할 수 있다.

시소러스구축과정에서 연결되는 대등관계는 완전한 동의어 관계보다 대부분 부분적 동의어 관계로 연결될 수 있는 용어들이 대부분이다. 다음 예는 형태가 다르고 의미의 차이가 유사

한 경우와 형태는 같으나 의미의 차이가 있는 용어쌍으로써 부분적 동의어 관계의 예이다.

예) 용어 형태의 차이 :

- 철수가 곶을 먹었다.
 - 철수가 밀감을 먹었다.
- 용어 의미의 차이
- 바위 틈으로 물이 샌다
 - 면회할 틈도 주지 않는다

따라서 개념이 부분적으로 일치하는 용어를 동의어로 인지하여 복수개의 단어를 동일한 범주에 배정하는 것도 인간의 감정이 포함된 판단에 의해 이루어질 수 있다. 시소러스의 대등관계설정에서 인간의 유사 혹은 동일하다고 인지하는 감정을 최소화하고 객관적인 관계설정이 필요하다.

3. 2 실제적 대등관계

실제 시소러스 구축과정에서 대등관계가 설정될 수 있는 것은 일반적인 대등 관계를 포함하여 유사관계 혹은 동의어 관계, 전거어 관계 등과 같이 검색 시스템의 상황에 따라 포함범위가 결정된다.

3. 2. 1 전거어 관계

전통적으로 정보검색에 있어 중요한 정보자원 가운데 전거어 파일은 이용자의 검색 효율성을 향상시키는 주요 검색도구 가운데 하나이다. 전거어 사전의 종류는 크게 인명을 비롯하여, 지명, 서명, 주제명으로 구분이 되며, 지식을 조직하고 서지의 질서를 확보하여 궁극적으로

목록의 일관성을 극복하기 위한 것이다(오롬정보 1997). 이는 접근점 표기의 다양화로 인해 재현율의 저하를 보전하는 도구인 것이다. 인명(단체명 등 포함)의 경우에는 전거어 형태로 이루어지는 대등관계에 대한 개념설정은 다음과 같다.

1) 축약형: 동일한 개념을 표현하는 것으로써 특별한 원칙을 갖고 있지 않으며, 일반적인 보편성에 기반하여 이루어지는 것을 근집화하는 용어들이 이에 해당한다. 예를 들면, 이화여자대학교를 지칭하는 단어로써 이화여대, 이화여자대학, 이대 등과 같은 축약형이 이에 해당한다. 보편성에 근거하는 것은 국민대학교의 경우 국민대학, 국민대라고 용어 표현은 가능하나, 국대라고 표현하지 않는다. 국대라는 용어가 국민대학교를 지칭하는 대치 용어로 사용되기 위해서는 보편적으로 일반인들에게 이 관계가 통념적으로 인식될 경우에만 허용되는 것이다. 즉 축약형의 경우 원형과 전거어로 처리되는지의 여부는 외부 환경이나 시간적 차이보다는 사회 통념적인 허용정도에 따라 파생되는 관계이다. 따라서 축약형의 경우에는 시간과 상황의 변화에 따라 매우 역동적으로 변화될 수 있다.

2) 외국어 번역형: 외래어나 외국어의 경우 한글 대응어와 원 용어간에 이루어지는 형태이다. 이는 보편성보다는 원칙에 근거하는 것으로써 한글 번역이나 한글표기된 데이터의 영문 번역도 이에 속한다. 특히 이러한 전거형이 나타나는 것은 외국어와 대응외국어의 용어열이 상대적으로 긴 경우가 이에 해당한다. ALA의 경우 American Library Association이며, 이에 대응되는 한글 표목은 미국도서관협회로

번역되기 때문에 전거어 사전이 필요한 경우이다. 또한 축약형과 외국어 번역형이 동시에 나타나는 경우가 있다. 예를 들면, 텔레비전에 대해 TV라는 용어를 허용한다. 이 형태는 해당 용어를 사용하는 영미권에서 허용하는 것을 국내에서도 축약형으로 사용하는 것이다. 또한 테레비라는 용어는 일본에서 사용되는 축약형을 일본어 번역과 함께 축약형을 사회 통념적으로 허용한 경우이다.

3) 변화형: 축약형과 번역형에 속하지 않고 초기에 사용되는 명칭이 시간이 경과함에 따라 변화된 용어들을 관리할 때 적용되는 형태이다. 표현대상이 되는 물리적 개념은 변하지 않았으나 이를 지칭하는 용어가 변화된 경우가 이에 해당한다. 예를 들면 경성제국대학과 서울대학교의 경우에 해당 대학교라는 물리적 개념은 변화가 없었으나 시대의 흐름에 따라 용어의 형태가 달라진 경우이다. 이는 하나의 개념을 대체나 혹은 부분집합의 형태가 아니라 물리적 속성은 정확하게 일치한 경우이다. 전거제어의 경우 총서명과 연속간행물의 표제가에 이에 해당한다. 또한 주제명의 전거제어도 이에 속한다.

4) 유사형: 유사형은 물리적 속성이 정확하게 일치하기보다는 많은 부분이 일치하여 두 개의 개념을 거시적으로 동일한 것으로 처리할 수 있는 경우가 해당된다. 전거제어의 경우에 있어 통일서명이 이에 해당한다. 집합으로 표시할 경우에 부분집합이나 완전집합으로 표기되지 않고 개념의 영역가운데 많은 부분이 겹쳐지는 경우가 이에 해당한다. 예를 들면, 고전인 '춘향전'과 근대소설인 '옥중화'는 스토리의 상당부분이 일치하나 등장인물의 이름이나 시

대적 공간에 차이가 있어 이 두 개의 자료가 완전하게 일치하지 않는 경우이다.

5) 오용형(誤用形) : 이 관계는 학계에서는 수용되지 않는 관계이거나 속어나 비표준어로써 실제 이용자들이 광범위하게 쓰이는 형태를 의미한다(류주희 2004). 한국역사용어 시소러스 구축에 있어 새로운 등가관계로 제시한 것이다. 예를 들면, ‘이씨조선’이나 ‘민비’와 같은 용어는 ‘조선’이나 ‘명성황후’와 같은 올바른 역사용어가 있음에도 불구하고 일반인들이 완전하게 오용인지를 인지하지 못하는 경우이다. 이는 외국어 번역형이나 축약형 등이 포함될 수 있으나 일반인들이 오용을 인지하지 못한다는 점에서 별도로 처리할 수 있다.

3. 2. 2 동의어 관계

시소러스의 대등관계에서 비우선어에 포함될 수 있는 것으로는 동의어와 유사동의어로 구분할 수 있다.

1) 일반 동의어관계

동의어는 넓은 의미로 동일한 범주에 속하는 용어으로써 각각의 용어를 대치시켜도 개념의 손상이 거의 없는 관계어를 의미한다(UNESCO 1981). 앞에서 설명한 것 가운데 완전한 일치에 속하는 용어간 관계가 이에 해당할 수 있다. 시소러스에서 동의어관계에 속하는 것으로는 다음과 같은 범주와 각각의 예를 확인할 수 있다(ISO 1976).

- 어원의 차이 : polygot와 multilingual
- 일반명과 학명 : aspirin과 acetylsalicylic acid
- 일반명사와 상표명 : vacuum flasks와

thermos flasks

- 신조어의 이형: hovercraft와 air cushion vehicles
- 신어와 구어 : radio와 wireless
- 이형철자표시어 : Romania와 Rumania, Roumania
- 문화차이에서 발생하는 용법의 차이 : flats와 apartments
- 약어와 완전명 : PVC와 Polyvinyl Chloride
- 복합어 표기의 차이에서 발생하는 이형동어 : Coal & mining과 Coal Mining

이상에서 확인할 수 있는 것은 동의어 관계는 동일한 하나의 개념을 다르게 표현하는 것을 의미하며, 각 용어들이 갖고 있는 개념은 완전하게 일치하는 관계를 나타낸다.

2) 유사동의어 관계

유사동의어는 엄격하게 구분할 경우에 서로 다른 개념적 크기를 갖고 있는 용어이면서 개념의 넓이가 상당부분 일치하는 용어들을 유사동의어라고 한다. 또한 개념의 넓이가 전혀 일치하지 않는 경우이지만 각각의 용어가 갖는 의미의 지향점이 일치할 경우에 이를 유사동의어로 간주한다. 예를 들면, 핸드폰 서비스 업체인 KTF와 LG 텔레콤이라는 회사는 서로 다른 회사이지만 동일한 방식(PCS 방식)의 유사 서비스를 수행하고 있기 때문에 이 두개의 회사를 PCS방식의 핸드폰으로 일치화할 수 있다. 또한 SK 텔레콤이라는 회사는 이와는 다른 방식으로 서비스를 제공하고 있어 기술적인 측면에서는 다른 서비스를 제공하고 있지만 인

지적인 측면에서 일반 이용자들은 이를 사용성이라는 측면에서 동일하게 인지하고 있기 때문에 이 세 개의 단어는 하나의 개념 하에 배정할 수 있다.

즉, 유사동의어는 개념간의 많은 부분의 넓이가 일치하거나 혹은 전혀 일치하지 않을 경우에, 이용자나 색인자의 판단에 따라 검색의 효율을 높일 수 있을 경우에 각각의 용어를 유사동의어로 선정하여 대등관계로 설정할 수 있다. 따라서 시소러스의 규모수준이 매크로일 때와 마이크로일 때에 따라 이 유사동의어에 속해 있는 개념들은 연관관계나 계층관계로 변화될 수 있다.

3. 2. 3 업워드 포스팅

업워드 포스팅은 동의어나 유사동의어 개념보다는 계층관계에 속하는 용어에 포함되는 용어를 시소러스의 규모나 수준에 따라 대등 관계로 처리하는 경우에 해당한다. 이는 개념적이거나 논리적인 방법보다는 시소러스의 용어의 총수를 적정 규모로 유지하기 위해 검색결과 건수가 적거나 혹은 시소러스 디스크립터로 너무 세분화된 계층 용어군에서 하위개념어들을 상위개념어의 동의어 처리를 하는 방법을 의미한다. 즉, 해당 계층용어군에서 최상위어를 디스크립터로 선정하고 그 용어 밑으로 전개된 하위어들을 모두 해당 용어의 비디스크립터로 처리하는 방법이다. 예를 들면, 암석이라는 상위어에 화강암과 석회암 등과 같이 하위어가 전개된 경우에 화강암이나 석회암과 같이 세분하여 처리할 필요가 없을 때 암석을 디스크립터로 선정하고 화강암과 석회암을 비디스크립터로 처리하여 하나의 계층 용어군을 유사

용어군으로 처리하는 방법이다.

3. 2. 4 대립관계

대립관계에 속하는 것으로 반의어와 대비어가 있다. 반의어는 의미적으로 서로 다른 용어를 뜻한다. 용어의 중첩도를 표현할 경우에 두 개의 용어가 외형적 형태나 의미적 내삽관계의 일치나 일부나 전부가 일치하지 경우가 반의어에 해당한다. 다만 각 용어간의 위치는 동위관계(同位關係)를 유지해야 한다. 대비어의 경우는 의미적으로 반대의 개념보다는 대립의 개념에 속하는 용어를 의미한다. 예를 들면 '절대선'과 '절대약'은 반의어 관계에 속한다. 이 관계는 시대나 사회의 외부적 환경의 영향에 따라 설정되는 관계이다. 이 관계는 대치관계로 표현하기도 하며, 의미적으로는 반의어로 확대 해석할 수 있다. 따라서 ISO 2788 과 5486 에서 제시하는 연관관계 기준에 포함될 수 있다. 그러나 ISO 에서 제시하는 기준은 같은 개념적 카테고리에 속하는 경우에 연관관계를 설정하도록 제시하기 때문에 이 관계설정 기준과는 구별할 수 있다(남영준 2003). 후자는 대비관계로써 '오디오'와 '비디오'의 관계에서 확인할 수 있다. 이 두 용어는 반대의 개념은 아니지만 의미적으로 서로 대비되는 특성을 갖고 있다. 이러한 관계에 속하는 용어는 각 용어를 연관관계로 처리할 수도 있다. 실제로 ISO 의 기준에도 반의어나 대비어는 연관관계에 속한다고 규정하고 있다. 또한 연관관계로 설정된 용어들은 서로 동일한 상위어를 가질 수 있다. 이와 같이 하나의 상위어에 반의어나 혹은 대비어로 구조화된 경우, 해당 용어가 필요에 따라 대등관계로 설정될 수 있다. 예를 들어 '동

영상파일'이라는 용어에 '오디오파일'과 '비디오파일'이라는 용어가 하위어로 설정이 가능하다. 이 용어를 연관관계에서 대등관계로 설정하면 재현율이 높아지고 정도율은 낮아지기 때문에 검색시스템의 관점에 따라 '오디오파일'과 '비디오파일'을 비디스크립터로 채택할 수 있다. 이 때 이 용어들을 서로 다른 의미를 갖는 대비어으로써 연관성을 갖지 않지만 '동영상파일'의 대등어인 비디스크립터로 처리한다. 이러한 관계는 시소러스를 활용하는 시스템의 특성이나 요구조건에 따라 해당 용어들이 독립적인 디스크립터로 사용될 수 있다. 특히 이 관계는 해당 용어간의 공기빈도(co-occurrence)에 의존하는 경우가 상대적으로 많다. 이를 그림으로 표현하면 다음과 같다.

즉, 오디오 파일과 비디오 파일이 시스템의 요구조건에 따라 상위어인 '동영상 파일'의 비디스크립터로 사용될 수 있으며, 상황에 따라서는 '오디오 파일'을 디스크립터로, '비디오 파일'이 비디스크립터로 활용될 수 있다 따라서 대등의 관계는 일치나 유사이외에 의미적으로 불일치하고 있는 용어간에 대등관계를 유지할 수 있다.

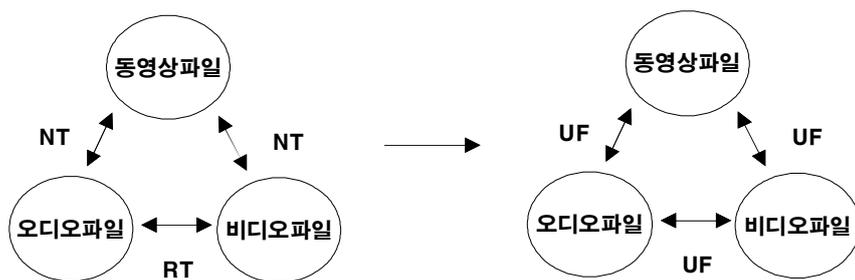
3. 3 새로운 대등 관계의 유형

3. 3. 1 시소러스 관계설정 조건

관계설정의 기준을 제시하기 위해서는 의소⁴⁾로서 디스크립터는 단일 개념과 단일 계층관계를 유지해야 한다. 즉, 개별 개념어의 의미는 복수개의 개념으로 해석되지 않아야 한다. 또한 다양한 해석을 최소화하기 위해 디스크립터의 형태는 명사(복합명사 포함)로 제한한다. 왜냐하면 디스크립터는 형용사와 같은 용어는 명사형에 비해 주관적인 다양한 의미를 갖기 때문이다.

1) 유일 의미의 원칙

인간은 특정 사물이나 개념을 인지하는 수준이 사람들이 갖고 있는 지식과 경험에 따라 크게 달라진다. 즉 하나의 용어는 사람들이 갖고 있는 인지도에 따라 의미적으로 각자 서로 달리 해석할 수 있다. 모든 용어는 대부분 다의성을 갖고 있기 때문에 시소러스가 검색어 사전의 역할을 수행하기 위해서는 하나의 디스크립터에 하나의 개념이 배정되어야 한다. 또한 디스크립터는 그 자체가 갖고 있는 심층적 의미보다는 용어에 대한 표층적 단일 의미로



4) 의소 : 통상적으로 의미 작용의 최소단위를 가리킨다.

해석되어야 한다. 예를 들면, 얼굴이라는 단어가 갖고 있는 단순한 표층적 의미는 인간의 머리에 있는 얼굴을 의미한다. 이러한 표층적 의미가 '얼굴을 들 수 없다'라는 표현에서의 개념은 단순한 얼굴을 지칭하는 것이 아니라 심층적 의미로 부끄럽다는 표층적 의미가 전이된 형태이다. '얼굴값을 한다'라는 표현에서 얼굴의 경우는 또 다른 해석이 가능하기 때문에 표층적 의미가 전이되는 것을 알 수 있다. 즉, 특정 단어를 단일한 의소로 해석을 하지 않을 경우에는 상황과 경우에 따라 원래 의미가 전이되어 전혀 다른 것으로 해석될 수 있다. 즉, 하나의 디스크립터가 하나의 개념을 갖기 위해서는 반드시 표층적 해석에 기준하여 사용되어야 한다. 또한 의미소의 표현 범주는 보편성에 기준하여 처리되어야 하는 것이다. 왜냐하면 디스크립터는 색인자와 이용자 모두에게 해당 개념에 대해 동일한 해석이 이루어져야 하기 때문이다.

2) 유일 계층의 원칙

하나의 개념어가 서로 다른 계층에 중복되어 출현하는 것은 해당 개념어가 다의성을 갖고 있는 것을 의미한다. 예를 들면, 서울대학교라는 개념어는 교육기관으로써 국립대학교라는 개념 하에 배정될 수 있다. 또한 서울대학교는 서울시라는 지역에 존재하는 지역단체로서의 역할도 있기 때문에 서울이라는 지역적 개념하에 배정될 수 있다. 이는 서울대학교라는 용어 자체가 함의하고 있는 내용을 단순하게 해당 용어를 표층적으로 해석하여도, 중의적으로 해석될 수 있는 경우이다. 이를 특정 계층에 중복하여 배정할 경우에는 대등관계를

비롯하여 연관관계, 계층관계로 구조화할 수 없게 된다. 서울대학교를 국립대학교라는 용어의 동의어로 처리할 경우에 서울지역의 지역단체로서의 의미가 배제되어 이를 수용할 경우에, 국립대학교와 지역단체와 대등관계가 설정되어 (비)디스크립터로서의 가치를 잃어버린다. 이러한 어려움 때문에 일반 상용 웹검색엔진에서는 @마크와 같은 기호를 사용하여 상황에 따라 해당 용어의 의미와 계층을 검색과정에서 차용을 한다. 이는 분류표(혹은 카테고리 검색)의 체계를 이용한 검색용으로는 활용이 가능하나 시소러스의 검색 기능에는 적합하지 않은 것으로 판단된다.

3) 명사 및 복합명사 우선 원칙

디스크립터의 형태가 구나 어절로 표현될 경우에는 명사(복합명사) 형태의 디스크립터의 경우보다 관계설정에 어려움이 상대적으로 많다. 왜냐하면 용어의 수가 증가할수록 용어에 수반되는 의미소가 증가하기 때문에 애매성이 증가되기 때문이다. 예를 들면, '일본에서 만든 만화영화'라는 표목은 일본이라는 단어가 갖고 있는 다의성의 수(n)과 만화영화가 갖고 있는 다의성의 수(n) 때문에 실제 해당 어절이 갖는 다의성에 대한 경우의 수는 $(n \times n)$ 개가 되기 때문에 어절의 용어수가 증가할수록 다의성의 수는 기하급수적으로 증가한다. 따라서 어절이나 어구로 이루어진 디스크립터는 계층 관계를 형성하고 보편성을 확보하기가 그렇지 않은 경우에 비해 상대적으로 높아진다.

3. 3. 2 새로운 대등관계 범주

선행연구와 함께 본 연구에서 서술한 바와

같이 대등관계는 특정 범주 내에 함께 존재하는 디스크립터에 대한 동의어간의 관계이다. 따라서 국내외 많은 시소러스에서 대등관계를 설정하는 범주는 크게 실제 대등관계와 유사 대등관계로 구분할 수 있다. 따라서 새로운 대등관계 기준을 제시하는 것은 실제 유사 대등관계의 범주를 설정하는 것과 이를 관리하는 방법을 제시하는 것으로 축약될 수 있다.

1) 실제 대등관계의 범주

동의어는 동일한 개념이나 물질에 대해 서로 다르게 표현하는 이형동의어라 할 수 있다. 동일한 개념을 다르게 표현하는 것은 크게 시간과 공간적 차이에 기인하는 것과 수준의 차이에 기인하는 것으로 구분할 수 있다.

- 시간의 차이 : 동일한 개념이나 물질이 시간의 경과에 따라 다르게 어휘소를 갖는 경우이다. 예를 들면, 전주대학교는 개교 초기에 영생대학교가 정식 명칭이었다. 이는 시간이 경과함에 따라 지금의 전주대학교로 개명이 되었으나 영생대학교와 전주대학교는 동일한 물질에 대해 다르게 표현되고 있다. 따라서 전주대학교와 영생대학교라는 용어는 시간의 차이에 기인하여 동일한 개념이 달리 표현됨으로 대등관계에 포함될 수 있다.
- 공간의 차이 : 동일한 개념이나 물질이 공간 및 지리적 차이에 따라 다른 어휘소를 갖는 경우이다. 표준어로 '거위'가 호남의 방언으로는 '떼가우'로 불리우고 있다. 이는 서울과 호남이라는 공간의 차이 때문에 발생하는 것으로써 대등관계로 이 용어들을 설정할 수 있다. 또한 거위를 영어

로는 goose 라는 용어를 사용하여 표현하고 있다. 이러한 차이도 공간적 차이에 근거하여 달리 표현되는 것으로써 대등관계로 이 용어들을 구조화할 수 있다. 공간의 차이에 의해 대등관계가 설정되는 것은 표준어와 방언, 외국어 대응어 등이 포함된다. 즉, 특정 개념에 대한 기준 용어와 함께 다르게 표현되는 모든 용어들을 대등관계로 설정하여 우선어와 비우선어로 구조화할 수 있다.

- 수준의 차이 : 동일한 개념이나 물질을 사용자의 경험이나 지식에 따라 다른 명명법을 갖고 있으며 이러한 것을 수준의 차이라 한다. 예를 들면, 일반인들이 펭귄이라고 용어를 사용하나 펭귄과 같은 극지방 동물학을 연구하는 분야에서는 sphenisciformes 라는 학명을 사용하고 있다. 또한 부모라는 개념에서 어머니라는 용어가 일반적으로 사용되고 있으나, 아이들의 수준에서는 엄마라는 용어를, 성인들에게는 어머님이라는 용어가 사용되고 있다. 이와 같이 표준어에 해당되지만 사회집단의 성향 수준에 따라 달리 표현되는 대등어 형태가 있다. 표준어나 방언의 형태는 갖고 있지 않으나 사회집단의 특정 계층이나 집단에서 사용되는 은어도 대등어로 처리될 수 있다. 예를 들면, 휴대용 전화기를 의미하는 용어가운데 '핸편'이라는 단어는 핸드폰의 축약형으로 표준어나 방언에 속하지 않고 지역이나 공간의 제약없이 특정 연령층에서 사용되는 것이기 때문에 수준의 차이에 의해 발생하는 대등어 형태라 할 수 있다.

2) 유사 대등관계의 범주

유사 대등관계로 설정되는지의 여부를 결정하는 기준은 개념의 허용범위이다. 즉, 동의어는 아니지만 시소러스와 같은 검색도구 구축과 같이 필요에 따라 동의어와 동일하게 취급할 수 있는지의 여부는 인간의 허용어부에 의해 결정된다. 유사 대등관계로 처리하는 것은 인간의 주관적 판단에 근거하는 것으로써 동의어에 비해 애매성과 다의성이 상대적으로 많이 발견될 수 있다. 시소러스 구축에 필요한 유사 대등관계의 범주에 속하는 것은 다음의 경우로 요약될 수 있다.

- 동일 수준에 있어 연관관계로도 설정이 가능한 경우 : 이 경우에 포함될 수 있는 용어간 관계는 두 개의 용어가 서로 개념적으로 일정부분을 공유하는 경우가 있다. 예를 들면, 디젤자동차와 LPG자동차의 경우, 자동차의 구조와 원리 등 개념의 상당부분을 공유하고 있으나 사용 연료가 다르기 때문에 이 두개의 용어를 개념적으로 구분하여 디스크립터로 채택될 수 있다. 또한 이 두 개의 용어 관계는 연관관계로 설정될 수 있다. 단, 시소러스에 등재될 디스크립터의 수가 제한되어 두 개의 용어를 모두 우선어로 선정할 수 없다면 이 두 개의 용어가운데 하나는 우선어로 처리하고 하나는 비우선어로 처리한다. 즉, 이 두개의 용어를 대등관계로 처리할 수 있다. 다른 경우로써, 두 개의 용어가 개념이나 물리적으로 일치하는 부분이 없으나 용어간의 공기정보 빈도수가 높아 개발자나 이용자의 주관적인 인식 때문에 대등관계로 설정할 수 있는 경우

가 있다. 예를 들면, FM과 음악방송이 이 경우에 해당한다. FM은 frequency modulation 의 약자로서 대응 번역어는 주파수변조이다. 이러한 의미가 사회 전반에 걸쳐 FM방송국이라고 인식되고 있으며, 특히 방송국보다 음악방송이라 인식되어 서로 동일한 개념을 갖는다고 인정한 경우이다. 원 의미와 개념은 서로 공유하는 부분이 전혀 없으나 사회 인식에 따라 대등관계가 형성된 경우이다.

- 상·하위 수준에 있어 계층관계로도 설정이 가능한 경우 : 이 경우는 앞에서 조사한 상위 개념어 선정(업워드 포스팅)이 대표적인 예가 될 수 있다. ‘셀룰러방식 핸드폰’이라는 단어와 ‘PCS방식 핸드폰’은 서로 다르게 운용되는 핸드폰이나 이를 특별히 구별할 필요가 없을 경우에 ‘핸드폰’이라는 우선어와 ‘셀룰러방식 핸드폰’과 ‘PCS방식 핸드폰’이라는 비우선어로 처리할 수 있다. 즉, ‘핸드폰’이라는 용어하에 전개될 용어이지만 시소러스의 효율성을 고려할 경우에 이를 대응관계로 처리할 수 있다.

3) 대등관계 설정의 필요성

위의 세 경우에서 대등관계나 혹은 다른 관계로 설정되는지의 여부는 전적으로 시소러스를 활용할 시스템의 환경과 이를 고려한 시소러스 개발자의 주관적 판단에 의존한다. 시스템의 환경은 소프트웨어적인 환경과 함께 시소러스의 효율을 고려하는 것이 우선한다. 일부 국내에서 개발된 시소러스에 등재된 디스크립터의 수가 2만개를 넘는 경우는 시소러스의 용

도 보다는 매크로적인 분류표의 상관색인이나 주제명 표목표의 목적으로 구축된 것이라 할 수 있다. 따라서 시소러스가 본연의 역할을 수행하기 위해서는 디스크립터를 적정 규모로 유지해야 한다. 왜냐하면 시소러스에 등재된 용어의 수가 적정한 수를 유지할 경우에 그 활용적인 가치를 갖기 때문이다. 이와 관련된 선행 연구(쓰리소프트 2002, 남영준 2001)에 따르면 국내의 주요 시소러스의 경우 등재된 디스크립터의 수가 4만 여개(NASA 시소러스 이상인 것도 존재하나 대부분 시소러스의 디스크립터는 15,000여개의 수준이 가장 바람직한 것으로 보고 있다. 왜냐하면 디스크립터의 수가 너무 적을 경우에 재현율이 높아지고, 용어가 너무 많아질 경우에 재현율과 정도율이 모두 저하되기 때문이다(쓰리소프트 2002, 남영준 2003, 남영준 2001). 이와 같이 디스크립터의 개수를 제한하는 것은 모든 개념이나 사물에 대해 개별적인 용어를 부여하는 것이 검색과 같은 정보관리 부분에서 비효율적인 행위가 될 수 있기 때문이다. 우리주위의 사물은 거의 무한에 가까운 반면에 언어자원은 유한하기 때문에 무한에 가까운 모든 사물에 독특한 이름이 주어진다면 우리는 낱말의 늪 속에서 헤어나지 못하게 된다. 따라서 비트젠슈타인은 인간이 특정한 사물이나 개념을 범주화하는 이유에 대해 '인간이 갖고 있는 유한한 용어로 무한한 사물을 다루기 위함'이라고 설명하며 적정한 언어자원이 인간생활에 더욱 효율적이라고 설명하고 있다(이기동 1982).

한편 새로운 개념이나 사물에 대한 정보가 증가할수록 이에 대한 검색적 배려가 필요하기 때문에 시소러스의 규모를 계속 제한하는 것도

현실적으로 불가능하다. 이러한 상반된 것을 극복하기 위해서는 디스크립터로 효용성이 상대적으로 떨어지는 것과 아직 효용성이 검증되지 않는 신규용어에 대해 일차적으로 비디스크립터화하는 방안을 채택한다. 국내의 지침에서 업워드 포스팅과 같은 계층이나 연관관계로 처리될 수 있는 용어들을 대등관계로 처리할 수 있는 지침을 갖고 있는 것도 이러한 상반된 요구를 해소하기 위한 것이다.

3. 4 대등관계 설정의 확장 지침

전통적으로 대등관계를 설정할 수 있는 경우는 복수 용어간에 개념적으로 유사관계에 속하는 경우에 한정하였다. 이런 원칙은 초기 시소러스 표준안에서 지켜졌으며, 용어관리의 필요성이 증대됨에 따라 개념적으로 유사관계에 속하는 것 이외에도 계층적으로 관련이 있는 용어가운데 필요에 따라 대등관계를 설정할 수 있도록 확대되고 있다. 또한 용어간 연관관계로 처리할 수 있는 용어들도 필요에 따라 대등관계로 처리하는 기준을 제안하고 있다. 본 장에서는 대등관계로 처리될 수 있는 기준을 확대하여 현행 기준을 그대로 수용하면서 향후 시소러스의 관계설정을 처리할 수 있는 확장기준을 다음과 같이 제안할 수 있다.

1) 동의어: 동의어는 대체어으로써 각 용어를 바꾸어도 항상 의미나 개념의 애매성이 거의 없는 경우이다. 이 관계에 속하는 것으로는 방언을 비롯하여 은어, 속어, 외국어 대응어가 이에 속한다. 이 경우는 항상 대등관계로 설정할 수 있다.

2) 유사동의어: 유사동의어에 속하는 것으로는 연관관계나 계층관계로 설정될 수 있는

모든 용어간 관계가 이에 속한다. ISO나 국내외 기준이나 지침에서 제안하는 연관관계의 예시나 계층관계의 예시가 모두 포함될 수 있다. 이는 사회적 환경이나 사용자의 집단의 인식수준을 기준으로 시소러스 개발자의 주관적 판단에 의존한다. 즉 유사동의어 그룹으로 군집화된 용어들은 외부 환경의 변화에 따라 필요시에 연관관계나 계층관계로 분할 혹은 확대 구조화될 수 있는 용어들이다.

3) 전거어 : 전거어 파일에 포함되는 것은 서명(총서명 포함)에 한한다. 인명이나 기관·단체명의 경우는 모두 동의어 수준이기 때문에 별도로 처리한다. 주제명 전거의 경우는 유사동의어 수준이기 때문에 유사동의어 처리기준에 준거한다.

4) 대립어 : 전통적인 대등관계와 반대되는 관계를 의미하며, 상위 개념어 선정에 해당하는 용어를 포함, 실제 반대에 해당하는 용어도 포함한다. 이 때 대등관계로 처리할 수 있는 용어는 두 용어가 필요에 따라 연관관계로 설정될 수 있는 용어만을 대상으로 한다. 이 기준에 따라 대등관계로 설정된 용어들은 관련정보의 양이 증가하여 이를 구별하는 것이 검색효율성에 유리하다고 판단될 경우에 대등관계에서 연관관계로 설정할 수 있다. 또한 상위 개념어 선정은 외부 환경에 따라 언제든지 계층관계로 전환하여 처리할 수 있는 디스크립터 풀의 역할도 수행할 수 있다.

5) 새로운 대등관계 기호 : 현재 대등관계를 표기하는 기호는 우선어의 경우에는 USE라는 기호를 비우선어의 경우에는 UF(used for)라는 기호를 사용하여 대등관계로 설정된 경우는 모두 구별없이 이 기호를 사용하고 있다. 한편,

계층관계 설정에 있어서는 속관계를 비롯하여 사례관계, 부분관계로 구분하여 구조화를 하고 있다. 대등관계도 계층관계에 있어 국면정보를 활용하는 것과 같이 동의어와 유사동의어, 전거어로 그 국면정보 쓰임에 따라 구분하는 것이 바람직하다. 이 때 국면정보를 제한하는 대등관계 기호를 선행연구(남영준 2003)에서 다음과 같이 제시하였다.

- UF(used for) : 동의어
- UFQ(used for quasi synonyms) : 유사동의어
- UFQ/H(used for quasi synonyms/hierarchical) : 유사동의어/계층
- UFQ/R(used for quasi synonyms/relational) : 유사동의어/연관
- UFA(used for authority file) : 전거어

이와 같은 기존의 대등관계 기준에 디스크립터의 확장과 관계구조의 확장성을 고려하여 반의어나 대비어의 관계도 대등관계로 설정할 수 있도록 다음과 같은 대등관계 설정기준을 추가 확대할 필요가 있다.

- UFC(used for coordinate words) : 대립어
- UFC/A(used for antonym) : 반의어
- UFC/C(used for comparison words) : 대비어

이 가운데 유사동의어(UFQ) 파일과 대립어(UFC) 파일은 그 자체만으로 시소러스에 등재될 디스크립터 관리용 용어풀⁸⁾의 기능을 수용할 수 있다. 즉, 디스크립터로 채택되기 전에 관련된 용어의 비우선어로 활용되고, 관련 자

료가 증가하여 검색어로서 가치가 발생할 때 디스크립터로 채택하는 방법을 의미한다. 향후 관리성을 고려할 경우에 유사동의어를 계층관계로 확대할 수 있는 대등어(UFQ/H)와 연관관계로 구조화할 수 있는 대등어로 구분할 수 있다. 또한 대립어 파일에 수록된 용어는 시스템의 조건에 따라 대립어군을 계층과 연관관계로 설정과정을 상대적으로 용이하게 수행할 수 있다.

4. 결 론

시소러스가 적절한 검색효율을 유지하기 위해서는 등재된 디스크립터의 질과 구조에 대한 지속적인 유지관리가 필요하다. 특히 대등관계로 처리하는 용어들은 전통적인 시소러스 개발 지침에서 제안했던 동의어만을 처리했던 수준에서 계층관계와 연관관계로 처리될 수 있도록 그 범위가 확대되고 있다. 그 이유는 시소러스에 등재된 디스크립터의 수에 대한 제한 때문에 기인한다. 모든 용어를 디스크립터로 처리할 경우에 시소러스가 갖는 적절한 검색효율 확보라는 원칙이 손상될 수 있기 때문이다. 따라서 기존의 디스크립터와 새로운 개념어를 시소러스에 등재하기 위해서는 유사어를 비롯하여 하나의 개념을 공유하는 단어들을 대등관계로 처리해야 한다. 이를 위해 본 연구에서는 대등관계를 처리할 수 있는 개념을 기존의 동

의어관계를 포함하여 대립어까지 포함할 수 있도록 확장하였다. 이 연구를 진행하면서 도출된 결론을 요약하면 다음과 같다.

1) 대등관계의 사례는 전통적인 동의어관계를 비롯하여 유사동의어 관계, 전거어 관계 대립어 관계 등 네가지 경우로 확대할 수 있다.

2) 동의어 관계에는 특정 시각이나 상황에서 서로간의 용어를 대치하여도 의미적 잡음이 최소화하는 관계이다. 이에 속하는 예로써, 방언을 비롯하여 은어, 속어, 외국어 대응어가 포함되며, 이 관계는 항상 대등관계로 설정할 수 있는 관계이다.

3) 유사동의어 관계는 특정한 개념을 서로 공유하고 있는 관계이다. 공유의 형태가 상황에 따라서는 연관형태나 계층형태로 중복될 수 있다. 유사동의어로 처리하는 것은 사회적 환경이나 사용자의 집단의 인식수준을 기준으로 시소러스 개발자의 주관적 판단에 의존한다.

4) 업워드 포스팅 관계는 시소러스 개발자의 주관에 따라 계층관계로도 처리할 수 있는 동일한 특성을 공유한 원어와 파생어의 관계이다. 즉 상위 개념어 선정은 동의어나 유사동의어 개념보다는 계층관계에 속하는 용어에 포함되는 용어를 시소러스의 규모나 수준에 따라 대등 관계로 처리하는 경우에 해당한다. 이는 개념적이거나 논리적인 방법보다는 시소러스의 용어의 총수를 적정 규모로 유지하기 위해 검색결과와 건수가 적거나 혹은 시소러스 디스

5) 용어풀은 유사어 사전의 pool을 의미하는 것으로써 시소러스의 디스크립터로 채택되어 완전한 표목으로 활용되기 이전 수준의 단어로써 해당 디스크립터에 비디스크립터를 모아 둔 저장 공간이다. 또한 해당 단어가 독립어로서 특정 용어와 관계설정이 되기 이전 단계로써 표목으로 등재되기 이전 수준의 키워드를 모아 둔 저장 공간을 의미한다.

크립터로 너무 세분화된 계층 용어군에서 하위 개념어들을 상위개념어의 동일어 처리를 하는 방법을 의미한다.

5) 대립어의 관계는 반대의미를 갖는 반의어와 대비어로 구분할 수 있으며, 이의 선정은 시소러스를 활용하는 시스템이나 환경을 고려하여 구축할 수 있다. 특히 대비어는 연관관계로의 설정을 기본으로 할 수 있다.

6) 향후 시소러스 관리를 위해 대등관계 기호를 유형별로 구분하면 다음과 같이 4가지 유형과 6가지 종류로 구분할 수 있다.

- UF(used for) : 동의어
- UFQ(used for quasi synonyms) : 유사동의어
- UFQ/H(used for quasi synonyms/hierarchical) : 유사동의어/계층
- UFQ/R(used for quasi synonyms/relational) : 유사동의어/연관
- UFA(used for authority file) : 전거어
- UFC(used for coordinate words) : 대립어

- UFC/A(used for antonym) : 반의어
- UFC/C(used for comparison words) 대비어

시소러스 개발은 실제적인 것과 이론적인 작업으로 구분이 가능하다. 특히 실제 정보검색시스템에 사용되는 시소러스는 시스템요구에 따라 유사동의어나 전거 파일이외에 용어도 대등관계로 설정하여 사용하여야 한다. 즉, 연관관계나 계층관계로 설정될 수 있는 용어도 필요에 따라 대등관계로 설정될 수 있다. 따라서 대등관계를 설정하는 작업은 매우 주관적이며 고도의 판단을 요구하는 작업이기 때문에 이에 대한 기준이 절대적으로 필요하였다. 본 연구는 이러한 고려사항을 모두 수용하고 실제 시소러스 개발시에 대등관계를 설정할 수 있는 범위와 그 기준을 제안하였다. 이러한 작업은 주관적으로 이루어질 수 있는 시소러스의 대등관계 구축작업을 체계적으로 운영할 수 있는 표준적인 자료로 활용될 수 있을 것이다.

참 고 문 헌

김기찬. 2000. 영어와 한국어의 기본 색채어의 비교분석. 『언어과학연구』, 언어과학회, 18: 21-48.

김성도. 2002. 『구조에서 감성으로』 서울 고려대학교 출판부.

남영준. 2001. 지식관리시스템을 위한 의미형 한영 시소러스 구축에 관한 연구. 『도서관정보학회지』, 32(4): 77-98

남영준. 2003. 시소러스의 대등관계에 관한 연구. 『문헌정보학보』, 중앙대학교 문헌정보학회, 6: 143-175.

남영준. 안동연. 2000. 다국어 시소러스의 개념 동일화 작업에 관한 실험적 연구. 『지식처리연구』, 문헌정보처리연구회, 1(2):

- 57-78.
- 류주희. 2004. 한국역사용어 시소러스의 구축 방법론. 인문학과 문화콘텐츠 인문콘텐츠. 학술세미나 자료집. 중앙대학교 문화콘텐츠기술연구원. 59-68
- 쓰리소프트. 2002. 『고속철도건설공단 시소러스 개발 및 구축』. [서울] : 쓰리소프트
- 오름정보. 1997. 『디지털도서관용 시소러스 및 전거데이터 구축에 대한 연구』. 오름정보. 연구보고서. 대전
- 이기동. 1982. 언어와 인지. 언어. 언어과학회, 7(2): 284-298.
- 이종열. 2000. 동의어의 인지구조와 의미적 비대칭성. 『언어과학연구』, 언어과학회. 18: 257-280.
- 한국데이터베이스진흥센터. 2000. 『시소러스 개발지침』, 이화여자대학교. 최종연구보고서. 서울: 97~200
- 한국정보관리학회. 1988. 『법률분야 관련어집』. 최종보고서. 서울: 법원도서관.
- American National Standards Institute. 1980. *Guidelines for Thesaurus Structure, Construction, and Use*. New York: ANSI.
- D. Soergel. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*(L.A.: Merville, 1974) p.30.
- Galinski, Christian, Goebel Jurgen W. 1996. *Guide to Terminology Agreements*. Infoterm. Termnet.
- ISO. 1985. *Documentation - Guidelines for the Establishment and Development of multilingual Thesauri*. 2nd edition(ISO 5964-1985(E))
- ISO. 1986. *Documentation - Guidelines for the Establishment and Development of Monolingual Thesauri*. 2nd edition (ISO 2788-1986(E)) 정동열 역. 단일언어 시소러스 제정 및 개발 지침. 시소러스 개발지침. 문헌정보처리연구회. 1994.
- Jean Aitchison, Alan Gilchrist, David Bawden. 2000. *Thesaurus Construction and Use: a Practical Manual*. 4th. ed. Aslib.
- NISO. 2003. 「Guidelines for the Construction, Format, and Management of Monolingual Thesauri : ANSI/NISO Z39.19 - 2003」the American National Standards Institute.
- Seetharama, S. 1975. Term-Concept Relationship in an Information Retrieval Thesaurus. Seminar on thesaurus in Information Systems. DRTC and INSDOC. Banglore. :A80-A87.
- UNESCO. 1981. 「Guidelines for the Establishment and Development of Monolingual Thesauri-2nd revised edition」UNESCO.

к с і