

용어분류의 비교연구*

A Comparative Study on Lexical Classification

백지원, 이화여자대학교 대학원 문헌정보학과

jeoid@hanmail.net

Ji-Won Baek, Dept. of Lib. and Info. Science, The Graduate School, Ewha Womans University

최석두, 이화여자대학교 문헌정보학과

sdchoi@ewha.ac.kr

Suk-Doo Choi, Ewha Womans University

본 연구의 목적은 용어분류를 그 특성에 따라 유형 구분하고 그 각각의 장단점을 밝힘으로써 앞으로의 용어분류 연구를 위한 기반을 제공하는 것이다. 이를 위해 먼저 용어분류의 개념을 밝히고, 그 목적 및 활용분야를 논하였다. 다음으로 다양한 용어분류 관련 사례를 그 내·외형적 특성에 따라 크게 유의어 사전류와 분류형 시소러스류의 두 가지 유형으로 나눈 후, 이를 다시 세분하여 그 각각의 특징을 분석하였다. 이 분석을 바탕으로, 특히 색인과 검색 등 정보처리에 필요한 용어분류의 유형과 특성을 논하였다.

1 서론

원문정보에 대한 요구와 웹 환경에서의 정보활용이 증가됨에 따라 정보의 축적 및 검색에 사용하는 용어를 어떻게 효율적으로 제어하고 이용할 것인가에 대한 문제가 중요하게 인식되기 시작하였다. 이에 지식이나 문헌을 조직하듯이, 용어를 일정한 체계와 구조 내에 조직화할 필요가 생겼고, 이는 용어분류를 의미하는 것이라 할 수 있다.

문헌정보학계에서는 지식을 담고 있는 문헌을 분류하기 위한 다양한 분류론을 연구해 왔으며, 이에 따라 여러 가지 분류체계를 개발하여 사용해 왔다. 그러나 정보의 색인과 검색 등 정보처리를 위해 필수적인 용어분류에 대한 본격적인 논의는 그리 활발히 이루어지지 못했다. 앞으로 용어분류의 분류 방법과 체계 등

용어분류에 대한 심도 있는 연구가 이루어지기 위해서는 기존의 용어분류 작업은 물론, 직접적인 용어분류는 아니지만 다양한 용도로, 여러 방면에서 이루어져온 기존의 용어체계화 작업들이 어떠한 기준에 의해서, 어떠한 방식으로 이루어져왔는지 살펴볼 필요가 있다.

본 연구에서는 용어분류를 명시적으로 행하거나, 혹은 암묵적으로 포함하고 있는 각종 주제분석 도구들 중 특히 국내외적으로 널리 알려진 사전, 시소러스, 주제명표, 그리고 관련된 개별 연구 및 프로젝트들을 대상으로 용어분류의 유형과 원칙을 분석해 보고자 하였다. 이러한 분석 대상 자료를 그 형식과 특성에 따라 유형별로 그룹화하되, 외형적인 형태나 명칭이 그 자료에 내재되어 있는 분류론적 특성과 일치하지 않을 경우, 분류의 특성에 우선순위를 두어 그룹화 하였다. 각종 유형에 있어서 논의해야 할 다양한 측면이 있겠으나, 본 연구에서는 분류와 개념구조에만 초점을 맞추었다.

* 이 논문은 2002년도 두뇌한국21사업에 의하여 지원되었음.

2 용어분류 개요

2.1 개념

‘용어’는 ‘일정 전문분야에서 주로 사용되는 말’로, 그리고 ‘어휘’는 ‘어떤 일정한 범위 안에서 쓰이는 낱말의 수효 또는 낱말의 전체’로 정의된다. 그런데 언어를 구성하는 중요한 의미 단위는 어휘소이고, 실제 어휘집들이 어휘를 분류하기 위하여 채택한 기준은 모두가 개별 어휘소들의 의미이다(김광해 1993, 126). 그러므로, 엄밀한 의미에서의 용어분류는 ‘용어 또는 어휘를 어휘소의 뜻에 따라 분류하는 것’이라고 하는 것이 옳을 것이다. 여기서 ‘뜻’이라는 단어는 일상적인 수준에서는 개념, 의미 등과 거의 동일한 의미로 사용되고 있다.

용어분류와 관련된 단어들은 대상과 분야 및 연구자에 따라 조금씩 다르게 사용될 수 있다. 본 연구는 특정 대상을 가지고 실제 분류를 수행하고자 함이 아니라, 용어분류의 전반적인 특성과 접근방법을 밝히고자 하는 데 주된 목적이 있으므로, 용어, 어휘, 어휘소 등을 특별한 구분 없이, 어떤 뜻(개념; 의미)을 명시적으로 드러내는 말 혹은 기호를 뜻하는 동일한 것으로 간주한다.

용어학은 개념을 언어로 표현하는 것에 관한 학문으로, 본질상 분류와 깊은 관련을 맺고 있다(김태수 2000, 8). 또한 분류는 유사한 사물들을 함께 그룹화하는 행위이고(정필모, 오동근 1998, 17), 유용성을 전제로 질서를 부여하는 일종의 사회적 기술이다(김태수 2000, 144). 결국, 용어분류란 일정한 언어 체계에 속하는 단어의 집합을 일정한 원리에 따라 동질적인 집합으로 나누는 것(田鳥 1993; 국립국어연구원 1993, 2에서 재인용)이라 할 수 있다.

한편, 용어가 나타내는 것이 지식이고, 지식의 집성체가 학문 또는 문헌이라고 할 때, 결국 용어와 지식, 그리고 문헌은 모두 다른 것이기는 하지만 서로 깊이 연관되어 있다고 할 수 있다. 따라서 지식분류와 문헌분류는 그 대상과 성격 등에서 용어분류와 근본적인 차이를 보이는 하나, 이들 분류 역시 특정한 분류표를 통해 개념을 구체적으로 조직하고 있다는 점에서, 어느 정도 용어분류의 성격을 지닌다고도 할 수 있다.

2.2 목적 및 활용분야

용어분류의 목적은 다양하며, 이는 주로 학문 분야별로 특색 있게 활용되고 있다. 첫째, 어휘 조사 및 정리를 통해 한 언어의 어휘를 집대성하고 그 체계를 세우고자 하는 목적이다. 주로 언어학적 측면에서의 연구에서, 기초어휘, 방언, 특정 분야 어휘 등 어휘 조사 및 기술, 작문 및 표현적 요구를 위한 어휘 검색, 어휘 및 언어교육, 그리고 사전 편찬 및 집필 등을 위해 어휘를 분류하게 된다.

둘째, 기계번역, 자연언어처리, 전자사전 편찬, 그리고 인공지능 연구 등의 기반으로 용어분류를 연구할 수 있으며, 주로 전산학 분야에서 이루어진다. 하나의 언어를 의미별로 분류하여 계층화시킨 의미계층 구조는 문장 분석시 구조적 모호성과 단어 의미 모호성을 해소하는 중요한 단서를 제공하고, 지식베이스로서의 역할을 하는 것이 보통이다.

셋째, 색인·검색을 비롯한 정보처리의 목적으로 용어분류 하는 경우로서 주로 문헌정보학 분야의 연구와 관련을 갖고 있다. 색인기법은 분류의 응용인 한편, 분류표는 통제되고 구조화된 색인어휘의 한 예라 할 수 있는 것이므로(정필모, 오동근 1998, 22), 용어분류체계는 색인의 근거가 되고 용어간의 관계 표시 및 검색이 체계적이고 효율적으로 이루어질 수 있도록 도와준다. 이 외에도 문헌과 웹 문서 자동분류, 의미에 의한 정보검색, 질의어 확장, 질의와 문헌사이의 유사도 측정을 통한 문헌순위 부여, 그리고 시소러스 통합 등을 위해 용어분류의 결과를 활용할 수 있다.

한편, 이러한 세 가지 분야에서 용어분류는 모두 상호관련성이 있으므로 제각기 따로 떼어 생각할 수 없다. 즉 전산학적 응용을 위해서는 언어학적인 어휘 수집, 조사, 분석이 선행되어야 하는 것이고, 문헌정보학의 분류 이론 및 경험, 그리고 색인·검색에서의 다양한 연구 성과는 전산학적 응용에 있어서 반드시 수반되어야 하기 때문이다. 또한 언어학과 문헌정보학에서의 연구는 전산학적 구현에 의해 실체화되고 활용될 수 있다. 따라서 용어분류는 학제적인 성격이 강하다고 할 수 있다.

3 용어분류의 유형과 특성

분류 체계를 가진 어휘 목록의 수록집은 분류 결과가 어떤 형태로 제시되느냐에 따라 그 이름이 달라진다. 일반적으로 사전의 형식을 가진 것은 '분류사전'이나 '갈래사전'이라 칭하고, 여기에 사전의 형식이 아닌 것을 포함시킬 때는 '분류어휘집'이라 부른다. 또한 이를 구미나 전산 분야에서는 '시소러스'라 칭하고, 일본에서는 '類語辭典'이라는 이름을 쓴다(국립국어연구원 1993, 8). 이러한 모든 유형은 분류 체계를 가진 어휘집이라는 점에서 그 근본 성격이 모두 같은 것이지만, 주로 결과의 외형적 형태에 따라 특별한 기준 없이 관행적으로 구분되어 온 측면이 강하다.

따라서 본 연구에서는 기존의 유형 구분과 상관 없이, 다양한 용어분류의 유형을 크게 유의어 사전류와 분류형 시소러스류의 두 가지로 나누었다. 사실상 용어간의 의미 관계를 파악하고 체계화한다는 의미에서는 시소러스 또한 유의어 사전류에 속하는 것이라고 할 수 있다. 그러나 독특한 편성형식과 체제를 갖추고, 정보의 색인과 검색에 주로 사용되는 현대적 의미에서의 시소러스, 그 중 특히 분류형 시소러스는 그 외형적 체제나 특성에서 볼 때 유의어 사전류와는 차이가 있으므로 이 둘을 구분하였다.

3.1 유의어 사전

'유의어 사전'으로 통칭되는 어휘집을 구분하는데 있어 김광해(1987)의 유형 구분을 적용하여, 어휘를 중심으로 분류하는 것은 (협의의)유의어 사전, 개념의 범주를 중심으로 분류하는 것은 분류 어휘집(혹은 類語辭典), 유의어군을 이루고 있는 어휘들의 의미 변별과 용법의 제시에 중점을 두고 있는 것은 의미변별사전이라고 구분한다. 이 중 의미 변별 사전은 본 연구의 범위를 넘는다고 보아 이번 논의에서는 제외하고 나머지 두 개, 즉 유의어 사전과 분류 어휘집, 그리고 별도로 개념사전이라는 명칭으로 범주화 될 수 있는 유형 등 총 세 가지로 유의어 사전의 유형을 세분하였다.

3.1.1 유의어 사전

유의어 사전은 의미가 유사한 어휘들이 군집을 이루며, 군집 중에서 가급적 쉬운 말, 즉 보편적인 언중이 공유하는 일상어를 표제 어휘로 채택하여 가나다 순(알파벳 순)으로 배열하는 것이 보통이다. 표제 어휘 및 각 유의어들의 의미나 용법 등에 관한 정보는 생략되는 것이 보통이나, 표제 어휘를 명확히 해 주고 표제어의 사용을 돕는다는 의미에서 반의어들을 함께 수용하는 경우가 많다.

『類義語·反意語 辭典』(1987)은 국내 유의어 사전 유형으로서 거의 유일한 것으로 볼 수 있으나, 외국의 경우 *Webster's New World Thesaurus*(1985), *The Longman Synonym Dictionary*(1986), *Roget's Superthesaurus*(1995), 그리고 *Collins Thesaurus*(1995) 등 이 유형에 속하는 많은 사전이 있다.

유의어 사전의 용어분류는 무엇보다도 유의어에 대한 정의를 어떻게 할 것인지에서부터 시작된다고 할 수 있다. 어떠한 관계에 있는 어휘들을 유의어로 볼 것인가라는 유의성의 정의에 대한 문제와 함께, 이러한 유형의 어휘집의 주된 용도가 무엇이냐에 따라 유의관계의 폭이 달라지게 된다.

유의어 사전과 같은 유형은 특정한 범위의 어휘를 대상으로 분류를 시작하게 되는 것이므로, 범위가 한정되어 있지 않은, 한 언어의 전체 용어를 분류 대상으로 할 때에는 사용하기 힘든 방법이다. 또한 이는 단순히 유의성을 기준으로 범주화되어 있을 뿐이므로, 하나의 범주 내의 유의어간의 의미 관계에 대해서는 알 수 없는 평면적인 체계를 가졌다고 할 수 있다.

3.1.2 분류 어휘집

분류 어휘집은 유어사전, 분류사전 혹은 갈래사전 등으로 불리기도 한다. 이 유형의 어휘집에서는 한 언어의 어휘 전체와 같이 큰 어휘 목록 전체를 대상으로 하여, 개념을 중심으로 범주화하여 분류하고, 그에 따라 배열하는 방식을 취한다. 따라서 이는 어휘의 의미를 중심으로 유의어군을 집결시키는 것이라기 보다는 언중이 세계를 범주화하는 양상을 중심으로 그에

관련되는 어휘들의 집합을 체계적으로 분류한 것이라 할 수 있다(김광해 1987, vi-vii). 이는 한 언어의 개념정리라는 측면, 그리고 어휘 의미에 대한 성분 분석이나 범주화 작업에 있어서 지침을 제공한다는 측면에서 중요성을 갖는다. 여기에는 유의어가 아닌 하위어들도 수용되며, 또한 사용의 편의를 위하여 반의어가 포함되는 경우도 많다(남영신 1989, 5).

Roget's Thesaurus(1992)는 그 개념에 따라 용어를 범주화시켰다는 점에서 현대적인 의미의 시소러스라기 보다는 분류 어휘집으로 유형화 할 수 있다. 이는 분류 어휘집(또는 시소러스)의 시초로서 획기적인 것이긴 하지만, 인간 사고 및 사물의 범주화에 있어서 객관성 있는 분류 기준을 찾기 힘들다.

한편, 우리나라에서의 대표적인 사례는 남영신의 『우리말분류(대)사전』(이름씨편 1987, 1989, 1994; 풀이말편 1988; 꾸밈씨 기타편 1992)과 박용수의 『우리말 갈래사전』(1989), 『겨레말 갈래 큰사전』(1992) 등이다. 남영신은 「이름씨(명사)편」에서는 학술적·상식적인 분류기준을 사용했다고 밝히고 있고, 「꾸밈씨(관형사, 부사, 조사, 어미, 접두어, 접미어, 선어말어미, 감탄사, 대명사, 준말 등) 기타편」에서는 품사별로 분류항목을 별도로 설정하였으며, 「풀이말편」에서는 쓰임새에 따라 서술어를 분류하였다. 박용수(1989)의 경우 쓰임새와 동류성(유의성)에 따라 대항목을 나누고, 일부는 품사별로, 그 나머지는 의미별로 분류하였다. 한편 박용수의 다른 분류사전(1992)에서는 표제어를 임의의 기준으로 대분류하고, 이를 품사--뜻--쓰임새를 차례대로 분류의 기준으로 삼아 세분하였다.

『分類語彙表』(國立國語研究所 1964) 역시 두 단계로 나누어 복수의 분류 기준을 적용하고 있다. 즉, 전체 체계를 품사별로 4류로 나눈 후, 이를 의미별로 분류하여 각 항목에 그 의미에 해당하는 어휘들을 배당하고 있다.

한편, 국립국어연구원의 『어휘분류에 관한 연구』(1993)의 경우, 의미 범주에 따른 주제별 분류법을 사용하고 있다. 이는 동일 주제와 관련되는 주체와 환경 및 대상, 도구, 방식, 활동 등을 나타내는 단어들을 같은 표제항에 포함시키는 방법이다. 교차분류가 발생하

지 않도록 문제의 대상과 상황을 대표할 수 있는 전형적인 예를 중심으로 동일 유형의 대상을 추출하고, 이를 다른 대상이나 상황에서는 제거하였다(국립국어연구원 1993, 1). 이 경우 어떤 용어 항목이 어떤 대상이나 상황에서 전형성을 가지는지를 판단하는 데 객관적인 판단의 근거가 있을 수 있는지의 여부가 분류표의 신뢰성을 좌우하게 될 것이다.

이러한 분류 어휘집에서는 한 언어의 총체적 어휘 목록을 효과적으로 범주화하기 위한 기준을 수립하는 것이 가장 중요한 문제이다. 하지만 사례들을 보면 편의에 따라 혹은 편자의 주관에 따라 품사, 의미, 쓰임새 등의 분류기준을 체계와 일관성 없이 적용하거나, 아예 명확한 분류 기준이 없는 경우도 있다. 이에 어휘 항목의 소진성(消盡性)을 위해 무리하게 하나의 항목으로 분류된 어휘들도 있고, 같은 의미라도 품사에 따라 분류 기준이 달라지는 등의 문제가 있다. 이처럼 유형은 분류의 관점이 불명확한 경우가 있고, 분명한 분류의 원칙을 파악하기 어렵다는 점에서 문제점이 있다.

3.1.3 개념사전

문헌정보학, 용어학, 심리학, 전산학 등의 분야에서는 다양한 목적으로 인간의 사고 및 개념을 구조화하고 분석하려는 많은 연구들이 진행되어 왔다. 이러한 개념의 구조화는 본질적으로 용어를 그 개념으로 구분 또는 범주화하며 체계를 세운다는 것, 즉 용어를 분류한다는 것과 일맥상통한다. 이처럼 개념 분석을 통해 어휘의 개념체계를 세운 어휘데이터베이스를 개념사전, 혹은 지식베이스라 할 수 있다.

개념사전의 대표적인 것들로는 EDR(1993), WordNet(1998), 그리고 CYC(2002)를 들 수 있다. EDR은 「개념 분류 사전」에서 개념을 5개 범주로 나누어 하위에서 세분하고 있으며, 「개념 기술 사전」에서 명사에 대한 동사의 격관계, 즉 자연언어 처리를 위한 구문적 관계를 위주로 하여 다양한 관계를 설정하고 있다. WordNet은 개념분류를 하는데 있어 동의 관계를 기본으로 하되, 명사, 동사, 형용사의 품사별로 각각 계층관계, 반의관계, 양식관계를 위주로 하고 있

다. CYC는 인간 세계의 전 분야에 대한 광범위한 어휘를 총 43개 범주로 나누어 수천 개에 이르는 다양한 의미관계를 설정할 수 있도록 되어 있다.

이러한 개념사전은 어휘와 그 의미에 대한 지식이 컴퓨터가 인간언어를 처리하는데 필수적이며, 개념은 어휘관계로 구조화 될 수 있다는 것을 전제로 한다는 점에서 공통점을 갖는다(Lenat, Miller, and Yokoi 1995). 반면 이들은 지식의 수준과 개념 관계에서 차이를 보인다. WordNet은 품사별로 어휘를 구분하여 각각 다른 분류를 하고 있는데, 이는 특히 한국어와 같이 품사간의 의미 개념 차이가 뚜렷하게 구분되지 않는 언어에는 적용하기에 어려움이 있다. 이러한 경우 품사에 상관없이 개념적으로 관련 있는 용어들을 링크할 수 있는 방법이 필수적으로 마련되어야 한다(김수정 1998, 22). 한편, 용어분류에 있어서 분류체계 설정 및 개념 표현이 우선시 되기 때문에 EDR이 가지고 있는 구문적 정보는 작성하고자 하는 분류의 용도에 따라 추가될 필요가 있을 것이다. 또한 CYC와 같은 경우는 수많은 개념과 관계가 설정되어 지식베이스로 역할을 하는 것이기 때문에 상대적으로 작성이 복잡하고 어렵다는 문제가 있다.

3.2 분류형 시소러스

분류형 시소러스는 크게 분류법을 적용한 것과 패시기법을 적용한 것으로 나눌 수 있고, 분류법을 적용한 경우는 다시 문헌분류표를 적용한 것, 자체적인 분류표를 적용한 것, 그리고 구체적으로 분류법을 적용하지는 않았으나 분류의 장점을 도입하고 있는 것 등으로 구분할 수 있다.

3.2.1 분류법 적용

첫째, 시소러스에서 용어분류 및 체계 구축을 위해 문헌분류법을 적용하는 사례로는 미국국립의학도서관 분류법(NLMC 1995)을 적용하고 있는 MeSH와 미국 의회도서관분류법(LCC)을 적용하고 있는 LCSH가 있다. COSATI 분류표를 적용한 예로는 *Thesaurus of Engineering and Scientific Terms*(1969)를 들 수 있다. 또한 도서십진분류법을 적용한 사례로는 용어를

먼저 그 의미로 분류한 후, 하위 분류에 있어서 도서십진분류법을 적용하고 있는 일본의 『角川類語新辭典』(1985; 송영빈 2002, 98)과 KDC를 적용한 우리나라의 『주제명 검색요어집』(1994)이 있다.

다음으로, 시소러스 자체적인 분류법을 적용하고 있는 경우도 있다. *INSPEC Thesaurus*(1999)는 INSPEC 분류를 사용하고 있는데, 디스크립터는 'CC'를 이용하여 해당 분류를 연결하고 있다(IEE 2001). 국내의 경우 『경제신문 시소러스』(1993)가 기사자료 표준분류표(1992)를 적용하고 있는데, 이는 원칙과 검색 관행을 함께 고려하여 작성된 분류표이다.

한편, 어떤 시소러스는 경우에 따라 의견상 명확히 드러나지는 않더라도 분류구조를 가지고 있으며, 이러한 체계를 통해 용어의 의미와 용어들의 연관성을 더 명확히 이해할 수 있게 해준다. 이러한 예로는 *Thesaurus of ERIC Descriptors*(1995)와 *Unesco Thesaurus*(1995)가 있는데, *Unesco Thesaurus*의 경우 디스크립터에 이 시소러스를 구성하는 87개 마이크로시소러스의 용어 번호를 부기하는 등의 방법으로 분류체계의 장점을 도입하여 색인과 검색시에 용어의 개념체계를 쉽게 파악할 수 있도록 하였다(UNESCO 1995, ix-xvii).

최석두(2002, 83-84)는 한글 시소러스의 분류체계로서 문헌분류법인 한국십진분류법(KDC)의 적용을 주장한바 있다. 그 이유 중 용어분류와 관련지을 수 있는 것은 다음과 같이 요약할 수 있다; 1) 새로운 분류표의 작성 없이 기존의 분류표를 적용하는 것이므로 경제적이다. 2) 문헌분류와 문헌의 주제를 표현하는 용어는 밀접한 관계를 가지고 있다. 3) 범용문헌분류표는 의미적 분류체계로서, 문헌의 색인과 검색에 많이 이용되어 잘 알려지고 검증된 분류체계이다. 4) 용어분류의 소진성 조건을 만족시키는 분류체계이다. 5) 근본적인 체계 변환 없이 다른 분류체계와 호환 혹은 변환이 가능하다. 6) 분류표를 갱신하면 용어분류체계가 갱신되며, 분류번호의 조기성, 분류표의 색인, 필요 분야의 추가, 확장 등 분류의 여러 기능을 용어분류에 이용할 수 있다.

한편 이처럼 분류표를 용어분류에 도입하는 경우,

문헌분류표의 고정적인 체계가 용어의 다양하고 복잡한 의미 관계와 변화를 다각적으로 나타낼 수 있을지를 고려해야 한다. 일반적으로 문헌분류표와 같은 계층 구조는 잘 정의 내려지고, 분명한 류(class)의 경계를 가지고 있는 실체를 위해 유용하지만(Kwasnik, 1999), 용어는 지식이나 학문에 비해 끊임없이 생성, 소멸되고 계속적으로 변화하기 때문이다. 또한 문헌분류표가 상당 부분 검증된 것이기는 하지만, 해당 주제 의 출판물의 양과 연구에서의 중요도에 따라 분류의 상세함이 좌우되는 MeSH 나무 구조의 사례))에서와 같이, 기존 분류표에서 계층의 비약이나 불균형, 작성자의 주관성이나 편견 등 여러 가지 문제점들이 노출되고 있다. 따라서, 용어분류를 위해서는 분류표의 분류체계로서의 신뢰성에 대한 근본적인 검토가 다각적으로 이루어져야 할 것이다. 한편 NLM 분류와 COSATI 분류, 그리고 시소러스 자체 분류법 등과 같이 특정한 주제에 한정되어 있는 분류표의 경우 분류 대상 용어를 모두 포괄할 수 있는지의 문제가 기본적으로 고려되어야 한다.

3.2.2 패싯기법 적용

패싯기법은 용어학에서 한 주제분야를 다수의 분류 방식으로 표현할 수 있는 가능성을 제공한다(Bowker and Timothy 1994).

시소러스와 패싯의 결합을 가장 잘 보여주는 대표적인 사례는 *Thesaurfacet*(1969)으로, 이는 시소러스를 포함한 분류 체계 혹은 패싯을 이용한 시소러스라 할 수 있다. 이는 직선적인 구조를 통해 하나의 계층 관계밖에 제시하지 못하는 분류체계의 한계를 극복하기 위하여 패싯화된 시소러스를 통해 개념간의 다중 계층관계를 보여주고자 한 것이다(Aitchison 1970).

시소러스에 패싯 분류를 적용한 다른 사례들로는 *Art & Architecture Thesaurus*(1994), *ROOT Thesaurus*(1988), 그리고 *Metathesaurus*(1998)가 대표적이다. AAT는 다른 사례에 비해 비교적 심층적인 계층수준을 가진 패싯분류를 적용한 시소러스로서, 패

싯의 조합에 있어서 상당한 정도의 자유를 허용하는 자유패싯모형의 하나이다(*Art & Architecture Thesaurus* 1994, 28; 손대형 1998, 20). 한편, ROOT에서 각각의 주제는 패싯 또는 기본적인 범주에 따라서 분야별로 분류되었고, 여기에 해당기호를 표시하였다(1988, i). 여기서 패싯 분석은 엄격하게 적용된 것은 아니지만, 주제분야의 논리적인 배열을 지원하기 위한 유용한 도구로 사용되고 있다. 또한 *Metathesaurus*에서 각각의 용어 의미는 정의나 주석에 의해, 계층관계에서 상·하위의 관계로, 동의어와 용어 사이의 다른 관계어로, 또는 기술이나 분류로 색인에서의 사용법에 의해서 명확하게 정의된다(정삼희 1999, 23-24).

이와 같이 패싯분류기법을 적용한 시소러스의 경우, 문헌분류에서와 마찬가지로 유연성과 표현력이 좋고, 다양한 관점에서 대상을 볼 수 있다는 장점이 있다. 그러나 용어 전체를 균형있게 분류할 수 있는 명확하고 망라적인 패싯의 설정 자체가 어렵고, 패싯들간의 관계가 부족하며, 패싯 조합이 까다롭다는 등의 단점이 있다. 더 나아가서는 일관성 있는 색인이 어렵고, 전체적인 분류 체계, 즉 용어들간의 의미체계를 조망할 수 있는 방법이 없다는 등의 문제를 갖게 된다. 또한 패싯 기법은 제한된 주제분야에는 효과적일 수 있으나 전 분야에 적용할 경우 그 규모가 너무 방대해질 수 밖에 없는 속성을 가지고 있으므로, 궁극적으로 전 지식분야 및 주제를 대상으로 하는 용어분류에 적용하기에는 어려움이 있다.

4 정보처리를 위한 용어분류

용어분류는 그 목적이나 대상 영역에 따라 성격과 유형이 상당히 달라질 수 있다. 실제로 각종 용어분류 관련 자료들에 나타나는 계층구조 또는 범주화의 차이는 구축 목적의 차이에서 기인하는 바가 크다. 따라서 용어분류는 기본적으로 보편성을 갖추되, 그 구체적인 대상 용어와 분류의 목적에 따라 그 원칙, 구조, 계층의 깊이, 그리고 유형 등 모든 것이 합목적적이어야 할 것이다. 본 연구에서는 다양한 활용분야 중 특

1) MeSH Tree Structure. <<http://www.hsl.creighton.edu/hsl/Searching/MeSH-Tree-TOC.html>>.

히 정보의 색인과 검색 등 정보처리를 위해 필요한 용어분류의 내용과 특성을 다음과 같이 분석하였다.

첫째, 용어의 범위는 일반용어와 전문용어를 모두 포함해야 한다. 또한 주제범위에 있어서는 특정한 영역에 한정됨 없이 전 영역을 망라해야 한다. 따라서 궁극적으로는 기초어휘를 주로 다루는 유의어 사전류와 전문용어를 주로 다루는 분류형 시소러스류가 다루는 모든 용어가 대상이 될 것이다. 이처럼 용어분류가 이론적으로는 모든 용어를 대상으로 하지만, 실제로 모든 용어를 소진하는 분류를 하기 위해서는 사실상 분류 이전에 분류 대상 용어의 목록이 확정되어 있어야 한다. 그러나 모든 영역에 걸친 기초 어휘 및 전문용어 전체의 목록을 작성하는 것은 간단한 일이 아니므로, 분류대상 용어의 수집과 선정에 대한 원칙과 방법에 대한 논의가 분류에 앞서 이루어져야 한다.

둘째, 분류의 관점 혹은 기준의 측면에서 볼 때 용어분류는 용어의 의미에 의해 이루어져야 한다. 현실적으로 모든 분류 계층에 일관되게 동일한 의미 기준이 적용되기는 어렵기 때문에, 여러 사례들에서 보면 부분적으로는 문법 정보에 의해 분류하고 자모순으로 배열하는 방법을 병행하는 것이 보통이다. 그러나 정보검색에 있어 일차적으로 필요한 것은 용어가 가진 음성/음운의 특성이나 품사 등의 통사적 특성이 아닌, 용어가 가진 의미라 할 수 있다. 따라서 의미에 의한 분류를 기본으로 하되, 목적과 용도에 따라 다른 기준을 부가적으로 적용할 수 있을 것이다.

셋째, 분류의 유형에 있어서는 문헌분류표와 같은 열거식 분류가 가지는 망라성 및 체계성의 장점과, 패킷기법과 같은 분석·합성식 분류가 가지는 다차원적 분류 가능성의 장점을 통합할 수 있는 체계여야 한다. 또한 용어는 그 수요가 방대하고 끊임없이 생성·변화·소멸하는 속성을 지니고 있으므로, 분류에 있어서 수정의 용이성과 새로운 용어의 삽입에 대한 신속성이 있어야 한다.

넷째, 분류의 구분 원리에 있어서는 일관성이 있어야 하고, 모든 대상 용어를 포괄(또는 소진)할 수 있도록 망라적이어야 한다. 실제로 동일한 어휘(소)가 여러 가지 기준에 동시에 소속되는 경우가 많기 때문

에 일관되고 명확한 분류 기준을 찾아내는 일은 용어분류에 있어서 가장 근본적이고도 중요한 일 중 하나이다. 또한 항목 구분에 있어서 분류된 항목은 상호배타적이어야 한다. 상호배타성은 분류의 구분 원칙으로서 준수되어야 하나, 실제 용어분류 작업에서는 항목 간 교차분류를 해야 할 필요성도 있을 수 있다.

다섯째, 정보처리를 위해서는 다차원적이고 상황의 존적인 용어분류를 구성해야 할 필요가 있다. 인지적 정보검색에서는 메시지의 대상이나 기호가 아닌 주제나 수신자가 중요하다고 간주하므로, 정보의 의미를 문맥과 상황에 따라 해석하게 된다. 따라서 각 개인에게는 여러 가지 분류가 각기 정당할 수 있으며, 이에 따라 용어관계의 세분화만으로는 지식의 다양한 계층관계와 이용자의 다양한 정보요구를 반영하기에 어쩔 수 없는 한계가 있다. 결국 앞으로의 용어분류 연구는 용어의 의미를 중심으로 개별적인 요구와 상황에 맞게 사용할 수 있는 다차원분류 혹은 상황분류 등의 가능성을 고려해야 할 것이다.

5 결론

용어분류를 위해서는 아직까지 해결해야 할 여러 가지 어려움이 있다. 하지만 반대로 생각하면 바로 이러한 이유에서 용어분류론의 확립이 요구된다고 볼 수 있다. 이를 위해서는 용어분류에 대해 어느 정도 보편성을 가진 기준이 마련될 수 있는 가능성을 찾아야 한다. 본 연구에서는 그 한 가지 방법으로 용어분류에 있어 관련된 선행 작업들을 분류의 특성에 따라 유형화하여 그 각각의 특성을 분석하는 방법을 사용하였다. 그리고 이를 바탕으로 특히 정보처리를 위한 용어분류의 내용과 방향을 제시하였다. 앞으로는 용어분류의 대상과 목적에 적합한, 보다 구체적인 분류의 원칙과 방법에 대한 연구가 진행되어야 할 것이다.

참고문헌

- 國立國語研究所. 1964. 『分類語彙表』. 東京: 秀英出版.
국립국어연구원. 1993. 국어 어휘의 분류 목록에 대

- 한 연구. [서울] : 동연구원.
- 국립중앙도서관. 1994a. 주제명 검색요어집 개발에 관한 연구. 서울: 국립중앙도서관.
- 김광해. 1987. 『類義語·反意語 辭典』. 서울: 한샘.
- . 1993. 『국어 어휘론 개설』. 서울: 집문당.
- 김수정. 1998. 한국어 개념 사전의 구축에 관한 연구. 석사학위논문, 연세대학교, 문헌정보학과.
- 김태수. 2000. 『분류의 이해』. 서울: 문헌정보처리연구회.
- 남영신. 1987-1992. 『우리말분류사전』. 서울: 한강문화사.
- . 1994. 『우리말 분류대사전』. 서울: 성안당.
- 박영자, 송만석. 1992. 자연언어 처리를 위한 한국어 동사·명사의 개념 분류. 『제4회 한글 및 한국어 정보처리 학술대회 논문집』. 141-149.
- 박용수. 1989. 『새우리말갈래사전』. 서울: 한길사.
- . 1992. 『겨레말갈래큰사전』. 서울: 한글문화연구회.
- 손대형. 1998. 패시분류체계를 이용한 시소러스 작성에 관한 연구. 석사학위논문, 연세대학교, 문헌정보학과.
- 송영빈. 2002. 전자사전과 어휘 기술. 『국어정보학 세미나 강의자료집』. 제3회 국어정보화 아카데미, 83-106.
- 田鳥いこう. 1993. “語彙分類のえ方,” 『日本語學』, 12. 재인용: 국립국어연구원. 1993. 국어 어휘의 분류 목록에 대한 연구. [서울] : 동연구원.
- 정삼희. 1999. 의미 네트워크를 이용한 시소러스 개발에 관한 연구. 석사학위논문, 전남대학교, 문헌정보학과.
- 조평옥, 옥철영. 1999. 의미속성에 기반한 한국어 명사 의미 체계. 『정보과학회논문지(B)』, 26(4): 584-594.
- 최석두. 2002. 『한글 시소러스의 구축 기준에 관한 연구』. 박사학위논문, 연세대학교, 문헌정보학과.
- Aitchison, Jean. 1970. "Thesaurofacet: a multipurpose retrieval language tool." *Journal of Documentation*, 26(3): 187-203.
- Bowker, Lynne and Timothy C. Lethbridge. 1994. "Terminology and Faceted Classification: Applications Using CODE4". In: *Knowledge Organization and Quality Management. Proceedings of the Third International ISKO Conference, 20-24 June 1994. Frankfurt/Main* : Indeks Verlag. 200-207.
- Buchanan, Brian. 1998. 『문헌분류이론』. 정필모, 오동근 역. 서울: 구미무역. 원전: *Theory of library classification*. London: Clive Bingley, 1979.
- CYCorp. Inc. 2002. Available at <<http://www.cyc.com>>. [cited 2002. 6. 25].
- Fellbaum, Christiane ed. 1998. *WordNet: An Electronic Lexical Database*. Cambridge: The MIT Press.
- Japan Electronic Dictionary Research Institute. 1993. *EDR Electronic Dictionary--Technical Guide*. 3rd Draft. EDR Technical Report TR-042.
- Kwasnik, Barbara H. 1999. "The Role of Classification in Knowledge Representation and Discovery." *Library Trends*, 48(1): 22-47.
- Lenat, Doug, George Miller, and Toshio Yokoi. 1995. "CYC, WordNet, and EDR: Critiques and Responses." *Communications of the ACM*, 38(11): 45-48.
- National Library of Medicine. 1999. *The NLM Classification*. 5th ed., rev. NLM.
- . *Introduction to the Classification*. [cited 2002. 3. 2]. <<http://www.nlm.nih.gov/class/nlmclassintro.html>>.
- The Institution of Electrical Engineers. 2001. *Outline of INSPEC Classification 1999*. [cited 2002. 2. 12]. <<http://www.iee.org/publish/support/inspec/document/class/classif.cfm>>.
- . 2001. *About INSPEC* [cited 2002. 2. 12]. <<http://www.iee.org/publish/inspec/about/>>.
- UNESCO. 1995. *Introduction to the 1995 edition*. [cited 2002. 2. 20.]. <<http://www.ulcc.ac.uk/unesco/intro/introeng.pdf>>.