

독한 전자사전을 위한 어휘항목의 미시구조에 대한 연구¹⁾

최병진(목포대)

1. 들어가는 말

정보통신기술의 발달로 인한 컴퓨터의 환경 변화는 교육 현장에 커다란 변화를 불러왔다. 이러한 변화 가운데 주목할 만한 것 중의 하나는 원격교육의 실현이다. 특히 멀티미디어 제작 기술의 발전과 WWW(World-Wide-Web)의 등장으로 교육환경은 급속도로 개선되어 가고 있다.

이러한 교육환경의 변화는 외국어 학습에도 커다란 전환점을 가져 왔다. 과거에는 외국어를 배우기 위한 학습 자료의 취득이나, 외국인과의 접촉, 대중 매체를 통한 대상 외국어와의 접촉이 쉽지 않았으나, 오늘날에는 이 모든 것들이 더 이상 문제가 되지 않는다.

또한 외국어를 배우는 과정에서도 다양한 매체를 통하여 보다 재미있고 편리하게 외국어를 학습하고 있다. 대부분의 외국어 학습시스템들에서는 학습효과를 높이기 위한 방법으로 멀티미디어 데이터, 학습자의 수준에 맞는 학습 내용, 그리고 언제든지 찾아볼 수 있는 온라인 전자사전 등을 제공하고 있다. 온라인 전자사전은 학습자에게 학습효과를 높일 수 있는 매우 중요한 요소 중의 하나이다. 더욱이 지금과 같이 신속하고 빠른 정보화 사회에서 대부분의 외국어 학습자들은 뜻을 모르는 단어의 의미를 알기 위하여 더 이상 인쇄된 사전을 뒤적이지기를 원하지 않는다.

이러한 이유에서인지 아니면 정보화의 흐름에 부응해서인지, 대부분의 출판사에서 는 기존의 사전을 전자화하여 전자사전의 형태로 출간해 나가고 있으며, 또한 인터넷 상에서도 다양한 종류의 온라인 전자사전을 쉽게 찾아 볼 수 있다.

이렇게 찾아 볼 수 있는 온라인 형태의 전자사전은 대부분 국어사전이나 영어사전 이 그 주를 이루고 있으며, 독일어 학습을 위한 독한사전이 전자사전의 형태로 제공 되고 있지는 않다. 요즘과 같이 독일어에 대한 관심이 점차로 줄어드는 시점에서 일반 출판사에서 독한사전을 전자사전으로 발행한다는 것은 사실상 기대할 수 없는 일이다. 하지만 학계에서도 이미 사전편찬과 관련하여 많은 관심을 가지고 있으며, 또 한 이 분야에 대한 연구인력도 점차로 늘어나고 있어서 독한사전을 보다 사용하기 편

1) 이 논문은 2001년도 목포대학교 학술연구비 지원에 의하여 연구되었음.

* 논문에 대해 세심한 조언을 해주신 익명의 심사자들에게 감사를 드립니다.

리한 전자사전의 형태로 발전시키는 것은 가능하다고 생각한다.

1949년 독한사전이 출간된 이래²⁾로 50여년의 세월이 지나 한국독어독문학회의 모델독한사전이 출간되기까지 우리의 독한사전은 나름대로 발전을 해 왔다고 할 수 있다.

이 논문에서는 독한사전이 더욱 발전된 모습으로 편찬되는데 조금이나마 기여하기 위하여 독한사전을 전자화하기 위한 방법론과 독한사전편찬의 근간이 될 수 있는 미시구조 Mikrostruktur를 검토함으로써, 앞으로의 사전편찬작업을 위해 고려되어야 할 사항³⁾을 제시하고자 한다.

2. 본론

2.1. 선행연구

우리는 컴퓨터로 작업되는 모든 사전을 전자사전으로 정의할 수 있다. 즉 컴퓨터가 읽을 수 있고, 컴퓨터가 가공·처리할 수 있고, 디스크나 CD-ROM에 저장되어 컴퓨터에서 사용할 수 있는 사전이라고 이해할 수 있다.

사전과 컴퓨터가 접목된 전자사전에 대한 관심은 90년대 들어서야 크게 증가했지만, 컴퓨터를 사전편찬작업에 이용하려는 생각은 이미 60년대부터 소그룹이지만 유능한 사전편찬학자들⁴⁾에 의해서 제기되었다.

독일의 Wahrig는 “사전편찬의 새로운 방향 Neue Wege in der Wörterbucharbeit (1967)”에서 컴퓨터 지향적인 사전편찬학을 제안하고 이러한 원칙에 따라 “Großes Deutsches Wörterbuch”을 편찬하였다. 그러나 Wahrig는 실제로 사전편찬작업을 자동화하기보다는 사전편찬작업의 전산처리를 위한 알고리즘을 제시하였고, 실제 작업은 수작업으로 이루어졌다.

사전편찬작업의 초기 자동화시대에는 많은 사전들이 컴퓨터의 도움으로 편찬되고 인쇄되었지만 대부분의 경우 어휘정보가 단순히 사람에 의해 입력되고 이것들이 자동

2) 이 시점은 구명철(2001)에 근거하고 있다.

3) 이 논문의 목적은 독한사전을 위한 완전한 미시구조를 확정하는 것이 아니라, 여기서 제안하는 미시구조를 바탕으로 완전한 미시구조를 정착시키기 위한 학문적 논쟁의 장을 마련하고자 함이다.

4) 불란서의 Bernard Quemada, 화란의 Felicien de Tollenaere, 독일의 Gerhard Wahrig를 들 수 있다.

적으로 정렬되어 처리되는 정도에 불과했다. 사전편찬을 위한 기초작업에서 편찬하는 데까지 컴퓨터가 읽어 처리할 수 있는 말뭉치 Korpus를 중심으로 사전을 편찬하는 프로젝트는 거의 없었다.

그러나 사전편찬작업에 컴퓨터를 이용하려는 노력이 점차로 증가하면서, 사전편찬 작업을 도와줄 수 있는 콩코던스 Konkordanz 프로그램이나 온라인 문서검색 시스템, 또는 사전편찬 워크벤치 등이 개발되었고, Longman Dictionary of Contemporary English(LDOCE 1978), Collins COBUILD English Language Dictionary(COBUILD 1987), Brockhaus-Wahrig(Wahrig 1981), Trésor de la Langue Française(Trésor 1971-1986)⁵⁾와 같은 사전이 전산처리방법의 도움으로 편찬되었다.

한편 일반 텍스트 형태의 사전을 기계가독 maschinenlesbar 형태의 전자사전으로 만든 최초의 작업은 1960년대 후반에 Systems Development Corporation(SDC)의 Olney에 의해 이루어졌다. 그가 일반 텍스트 형태의 사전을 종이 테이프위에 키펀치를 하여 기계가독형태로 만든 사전이 바로 “Webster’s Seventh New Collegiate Dictionary”이다⁶⁾. 이밖에 언어학적 코드와 분류 체계를 가지고 편찬된 Longman Dictionary of Contemporary English (LDOCE) 사전도 기계가독형태의 전자사전으로 만들어 졌다. 그리고 The Oxford English Dictionary (O.E.D)사전의 경우에는 CD-ROM의 형태로 사전을 저장하여 사용자로 하여금 손쉽게 사전을 이용할 수 있도록 하였다.

한편 국내에서는 1990년대에 들어오면서 사전편찬학에 대한 관심이 점차 활기를 띠면서, 국어사전편찬의 실제 작업과 관련하여 컴퓨터를 이용하려는 시도도 함께 이루어졌다. 국내에서는 크게 연세대학교 사전편찬실(현 언어정보개발연구원의 전신)과 국립국어연구원, 고려대학교, 한국과학기술원을 중심으로 사전편찬에 대한 연구가 이루어졌다. 사전 편찬작업을 수행하는 과정 속에서 개개의 사전편찬 연구팀은 각각 개별적으로 코퍼스를 수집하고 이를 가공함으로써, 사실상 시간과 비용에 있어서 중복된 작업을 수행하였다. 이러한 상황 속에서 한국과학기술원의 인공지능연구센터에서는 국내 전자사전의 표준화와 통합을 위한 노력을 시도하였으나, 그 연구가 지속적으로 이루어지지 못했다. 그 뒤 문화관광부의 지원을 얻어 세종계획이라는 국책 프로젝트가 시작되었고, 이 프로젝트를 통해 세종사전⁷⁾이라는 새로운 전자사전이 개발되고 있다.

5) B. Quemada에 의해 주도된 Trésor 사전편찬을 위해서는 상당히 많은 양의 불란서 문헌과 사전자료의 데이터가 상당히 일관성있게 모아졌다.

6) Olney, J. & Ramsay, D.(1972) 참조

7) 세종사전은 프랑스의 M. Gross의 어휘문법(lexicon-grammar)에 기초하여 어휘정보를 구축하고 있다.

국어사전의 전자사전화와 함께 이중언어사전에 대한 연구도 병행되었으나, 이는 주로 영어나 일본어를 중심으로 이루어졌으며, 독일어를 중심으로 하는 연구는 거의 이루어지지 않았다.

그러나 독어독문학계를 중심으로 독한사전의 거시구조(Makrostruktur)나 미시구조(Mikrostruktur)와 관련하여 이루어진 연구로는 다음과 같은 것이 있다.

우선 구명철(1997)에서는 외국인을 위한 독일어 사전 편찬을 위해 사전편찬과 관련된 주요 이론을 소개하고 이를 바탕으로 독한사전을 위한 사전학적인 모델을 제시하였다. 그의 모델에서는 독한사전의 거시구조를 위해서는 벽감 알파벳(nischenalphabetisch) 구조가, 미시구조를 위해서는 부분통합적(semiintegriert) 구조가 적합하다고 주장되었고, 그 외에도 거시구조와 미시구조에 있어서 고려해야 할 사항이 제시되었다.

한편 전경재(1998)는 표제어 유형, 정서법, 동어군과 동일규정어 합성어를 중심으로 독일어 사전(DU, BW)과 독한사전(모델, 엡센스, 동아, 교학사)을 비교하여 거시구조의 관점에서 독일어 사전의 표제어 배열과 관련된 문제점들을 논하였다.

이러한 연구와는 달리 양도원(1999)은 독일어 초보자를 위한 외국어 교육을 위하여 기존의 사전과는 다른, 가능한 한 사용하기 쉬운 학습자 중심의 사전개발을 시도하였다.

구명철(2001)에서는 1949년부터 지금까지 출간된 9개의 사전을 사전학적인 관점에서 거시구조와 미시구조를 비교하였고⁸⁾, 구명철(2002)에서는 외국어 교육을 위한 디지털 사전의 모델이 제안되었다.

이외에도 사전과 관련하여 이루어진 연구로 이해옥(2001)과 천미애(2001)를 들 수 있다. 이해옥은 독일 본 대학에서 이루어지고 있는 한독사전 편찬사업과 관련하여 주체별 한독사전의 편찬기술에 대한 문제점과 분류 체계를 소개하였고, 천미애는 미시구조의 한 부분이라고 할 수 있는 성구소에 대하여 성구소 사전정보의 문제점을 살펴보고 외국어 학습자를 위한 성구소의 새로운 미시구조를 제안하였다.

지금까지의 사전과 연관된 연구를 살펴보면, 구명철(2002)에서 디지털 사전의 모델을 제안한 것 이외에는 독일어 전자사전에 대한 연구가 이루어지지 않은 상황이며, 또한 사전에 대한 구체적이고 표준이 될 만한 논리구조 형태가 제안되어 있지 않아, 실제로 독한사전의 편찬을 위해서 표준이 될 수 있는 구체적인 지침이 마련되어 있지 않은 실정이다.

이 논문에서는 지금까지의 선행연구를 바탕으로 이상적인 독한사전의 편찬을 위해

8) 구명철(2001)과 유사한 연구로 김경(2002)이 있다.

고려되어야 할 사항을 검토하고 구명철(2002)에서 제시된 디지털 사전과 유사한 형태의 전자사전 구축을 위한 미시구조를 살펴본 후, 이를 컴퓨터로 구현하기 위한 기본 설계를 XML로 구현하고자 한다⁹⁾. 이러한 맥락에서 이 연구는 구명철(2002)과 일맥상 통하고 있으나, 구명철(2002)은 외국어 교육을 위한 디지털 사전의 특징과 모델을 제안한데 반해, 여기서는 외국어 교육뿐만 아니라 일반 사전 이용자나 전자 출판에도 유용한 보다 구체적인 전자사전의 원형을 제시하고자 한다.

3.2. 어휘항목의 미시구조 고찰

2.2.1. 사전과 사전편찬학

사전은 우리가 단어의 사용에 대하여 알아야 할 내용과 전통이나 관습에 근거한 단어의 의미, 그리고 그 단어의 의미와 관련되어 있는 약속을 기술함으로써 인간의 의사소통의 이해를 돕는데 기여한다.

따라서 우리가 언어와의 관계를 지속적으로 유지하는 한 사전은 존재할 수밖에 없고, 계속해서 발전되면서 출간될 것이다. 그것은 바로 언어가 끊임없이 변화하며, 사전이 언어에 대한 지식을 중재해 주기 때문일 것이다.

인간의 언어생활에서 사전은 사실상 유형 또는 무형의 형태로 존재하지만 우리가 사전이라고 머리 속에 떠올릴 때에는 유형의 사전을 생각한다. 이 논문에서도 인지과학이나 심리학적 관점에서의 무형의 사전은 고려하지 않고 유형의 사전을 그 연구 대상으로 한다.

우리가 알고 있는 유형의 사전에 대해 단순히 정의를 내린다면, 단어들이 알파벳 순서로 정리되어 그 의미가 설명되어 있는 책이라고 할 수 있을 것이다. 그러나 이러한 정의는 여러 가지 형태를 지니고 있는 사전의 다양성을 충족시키지는 못한다.

그 이유는 사전에 등재되는 표제어의 단위가 단어가 아닌 형태소가 되는 경우도 있다. 뿐만 아니라 거시구조의 관점에서 보면, 알파벳 순서가 아닌 주제별로 분류되어 있는 시소러스 Thesaurus와 같은 형태의 사전도 있으며, 그 매체도 책자가 아닌 디지털의 형식으로 되어 있는 경우도 있다. 또한 대부분의 사전에서 중요한 역할을 하고 있는 뜻풀이에 있어서도 어휘항목에 대한 설명이 다양한 방법을 통해서 이루어지고 있는 것이 일반적이지만, 경우에 따라서는 뜻풀이가 필요없이 단지 주어진 어휘항목

9) 현재 독일 만하임의 독일어 연구소에서도 사전의 논리적인 구조를 설정하여 XML 에디터를 이용하여 하이퍼텍스트의 형태로 전자사전을 구축하고 있다.

에 대하여 단순한 형태의 정보가 주어지는 것만으로 충분할 때도 있다¹⁰⁾.

따라서 사전이란 '특정한 매체를 통해 특정사용자에게 필요한 정보를 쉽게 찾아볼 수 있도록 정돈된 어휘요소(주로 단어)의 목록'이라고 정의하는 것이 타당하다¹¹⁾.

지금까지 사전은 문법이론과 독자적인 것으로 간주되었고, 사전을 편찬하는 작업 역시 별개의 것으로 여겨졌으며 큰 관심을 끌지 않았다. 그러나 90년대로 접어들면서 사전은 언어학의 세부분야에서 집중적으로 조명을 받았으며, 자연언어를 처리하는 언어학이나 전산학의 응용분야에서도 그 중요성이 부각되었다. 이러한 변화와 함께 드디어는 사전편찬학 Lexikographie이라는 영역의 확립에까지 이르게 되었다.

사전편찬학이란 용어는 어휘학 Lexikologie이라는 용어와 대립되어 자주 사용되고 있는데, 언어학의 한 분야로서 어휘학 Lexikologie이란 명칭은 그리스어의 "단어에 관한, 단어와 관련된"이란 의미의 'lexikós'와 "이론 내지는 학문"의 의미를 지닌 'logos'에서 유래되었다. 한마디로 어휘학은 단어에 관한 학문으로, 어휘 Lexem가 무엇인가에 대한 질문에 대답하기 위해 노력한다. 어휘를 그 연구대상으로 삼는 어휘학은 어휘정보를 서술하기 위한 언어학적 이론이나 방법론과 관련된 기술언어학의 한 분야로, 어휘적 언어, 관용어, 어휘의미론¹²⁾, 어장 Wortfeld의 구조, 의미성분이나 의미 관계와 밀접한 관련이 있다.

반면에 사전편찬학 Lexikographie이란 과거에는 사전을 편찬하는 과정이나 행위¹³⁾를 의미하였지만, 최근에 와서는 그 과정뿐만 아니라, 응용언어학의 한 분야로 실질적인 사용을 위한 사전의 구축 및 디자인과 관련된 이론으로까지 그 개념이 확장되었다. 사전편찬학의 대상은 사전으로, 사전은 미리 주어진 것이 아니며, 의도적으로 창조되어진 것이다. 사전은 그 사용자에게 단어나 어휘에 대한 질문에 상담자로서의 역할을 수행해야 한다. 따라서 사전편찬학자들은 사용자들의 요구에 부합하기 위해서 어떻게 사전을 구성해야 할 지에 대해서 연구를 해야 한다.

사전편찬학의 원칙적인 방법은 기호이론적인 연구에 기반을 두고 있다. 전통적인 사전편찬학의 기호모델은 Saussure의 이원적 모델로서 단어의 형태 Form와 개념으로 해석되는 의미를 구별하고 있다. 이 단순한 기호모델이 어휘를 검색하는 절차적인 기준에 따라 사전의 유형을 구별하는데, 전통적인 유형의 사전에서와 같이 단어의 형태

10) 이런 경우에는 정서법사전이나 역순사전, 또는 형태소 사전과 같이 사전에 어휘항목이 존재하는 것으로 충분히 그 기능을 하고 있는 경우이다.

11) Hausmann(1985), S. 369.

12) 최근까지도 어휘의미론은 단어의 통사적, 형태론적, 음운론적 특질과 별도로 다루어졌으나, 최근의 언어학 이론에서는 이러한 여러 층위의 내용들이 점차로 통합하고 있다.

13) Wiegand(1983, 38)

를 중심으로 사전을 검색하고, 검색된 단어의 정보가 의미가 되는 사전이 어의론 사전 semasiologisches Wörterbuch이며, 검색의 기준이 개념이고 그 정보가 단어형태인 사전이 명칭론 사전 onomasiologisches Wörterbuch이다.

따라서 사전이 거시구조에 있어서 어의론적 구조를 지니면 알파벳 순서의 사전의 모습을 지니고, 명칭론적 구조를 지니면 시소러스나 개념사전의 형태를 지닌다.

한편 개별어휘항목의 정의부분을 어떻게 구성해야 할지에 대한 문제는 바로 미시구조를 어떻게 설정하느냐와 연관되어 있다.

사전의 분류요소는 사전어휘항목 Wörterbuchartikel이며, 각각의 어휘항목은 표제어 Lemma와 그에 대한 설명부분으로 나눌 수 있다. 미시구조에 대한 연구는 바로 이 어휘항목의 설명부분에 대한 연구로 설명부분에 대하여 세부 구성성분요소를 설정하고 어떠한 구조를 지닐 것인가에 대한 윤곽을 정하는 것이다. 이것은 바로 어휘의 구조화와 관련이 있다고 할 수 있다. 다시 말하면, 어휘항목들 사이의 관계의 관점에서 전체적인 거시구조를 정하고, 개별 어휘항목에 대하여 어떠한 어휘정보를 어떻게 설정해야 할 지의 미시구조를 정함으로써 어휘를 구조화하는데 기여할 수 있다¹⁴⁾.

2.2.2. 메타언어의 설정: XML

컴퓨터로 하여금 문서 데이터를 인식하여 처리할 수 있고, 더 나아가서는 다른 목적에 응용할 수 있게 하기 위해서는, 이 문서 데이터를 기호 표시의 체계로 바꾸어 정형화하여야 한다. 컴퓨터가 처리할 수 있는 기호는 문서 데이터에 비해 적은 수이므로 특별한 마크업(mark-up) 체계의 개발이 필요하다. 마크업이란 문서의 형태 정보를 유지하기 위해 부가적으로 필요로 하는 정보 처리 명령의 표시이다. 즉 복잡한 형태 정보를 표현하기 위해 약속하는 일종의 기호이다. 이러한 마크업 체계의 예로는 SGML (Standardized Generalized Markup Language), HTML (HyperText Markup Language), LaTeX 등을 들 수 있다.

그런데 이러한 표시 체계에서 항상 문제가 되는 것은 바로 표준화의 문제이다. 표준화된 코드 체계를 만들기 위한 노력은 80년대 초 SGML 개발을 시작으로, TEI (Text Encoding Initiative)의 작성, 지금의 XML의 개발을 바탕으로 어떠한 형식의 문서나 데이터도 호환의 어려움없이 컴퓨터가 읽을 수 있도록 코드화 Kodierung하려는 연구 등이 그동안 지속적으로 추진되고 있다.

SGML은 마크업 언어가 아니라 마크업 언어 구문을 정의하는 메타언어로, 1980년

14) 최병진(2002) 34ff 참조.

대 초부터 IBM 표준인 GML을 바탕으로, Charles F. Goldfarb와 ISO (International Organization for Standardization)의 한 작업 그룹에 의해 개발되었다. SGML을 이용하여 문서를 작성하기 위해서는 일반적인 마크업 언어를 정의하고 이 언어에 맞춰서 실제 문서를 만드는 작업이 필요하다. SGML의 원칙은 내용상 다른 여러 텍스트 부분들이 지시자 (혹은 태그)를 통하여 표시된다는 것이다. 내용상 다른 텍스트 부분들이란 예를 들면, 제목, 단락, 언어학적 설명, 인용, 주, 도표, 수식, 특수 문자 등과 같은 것을 말한다. 그밖에도 텍스트 외부에 놓여있는 대상을 지시하는 지시자도 정의할 수 있다.

지시자는 기타 텍스트와 구분되도록 '<' 과 '>'에 의해 끝이 표시된 임의의 문자열을 말한다. 표시의 끝은 '</>'으로 나타낸다. 텍스트의 표시를 위해 사용된 지시자는 문서의 특별한 부분에 문서유형정의 Document Type Definition(DTD)¹⁵⁾라는 이름으로 정의된다. 문서유형정의는 문서에 사용된 모든 지시자들과 경우에 따라서는 그들의 계층구조적인 관계를 포함한다.

SGML은 일반적인 틀로서 여러 가지로 사용될 수 있어서, 현재는 수많은 양의 데이터를 표준화하여 코드화하는데 사용되고 있다. 예를 들면, Oxford Advanced Learners Dictionary of Current English와 CD-ROM에 기계가독형태로 저장된 Oxford English Dictionary가 바로 이러한 SGML형태로 구조를 이루고 있으며, British National Corpus와 같은 대량의 코퍼스도 SGML을 바탕으로 하는 코드 체계를 이루고 있다.

그러나 SGML은 문서를 표준화된 형태로 표기하는 데에는 유용하게 쓰일 수 있으나, 이를 배우고 이용하기에는 매우 복잡하고 불편하다. 또한 현재 널리 유용하게 사용되고 있는 인터넷의 WWW환경에서 바로 이용할 수 없다. 이러한 이유로 SGML과 WWW의 HTML의 장점을 살린 XML(eXtensible Markup Language)이라는 언어가 개발되었다.

이 논문에서는 바로 이 XML¹⁶⁾이라는 언어를 기반으로 전자사전의 논리구조를 설정하려고 한다.

2.2.3. 사전 어휘항목의 구성요소

일반적으로 사전을 편찬할 때에 기본적으로 고려되는 사항은 표제어를 배열하는 거

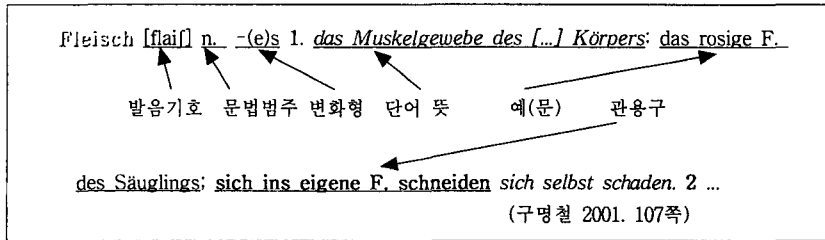
15) DTD에 대해서는 최병진(2001)을 참조.

16) XML의 장점이나 구체적인 사항에 대해서는 최병진(2001)을 참조.

시구조와 표제어에 대한 정의가 어떠한 형태로 구조를 이루어야 할지의 미시구조에 관한 문제이다.

이 논문에서는 우선 인쇄된 사전을 거시구조보다는 미시구조적인 관점에서 사전 어휘항목의 정의를 위하여 어떠한 구성요소가 나타나야 할지를 고찰한다. 국내에서 출간된 주요 독한사전의 미시구조와 거시구조에 대한 사전학적 분석은 이미 기존의 연구(17)에서 이루어졌으므로 여기서는 자세히 다루지는 않고, 기존 사전의 구조를 비교, 분석한 결과를 바탕으로 독한 전자사전 편찬을 위한 어휘항목의 미시구조 설정에 초점을 맞추고자 한다.

우선 미시구조의 설정에 앞서 사전에서 미시구조를 구성하는 요소는 어떠한 것이 있는지를 살펴 보아야 할 것이다.



<표 1> 사전어휘항목의 실례

<표 1>은 구명철(2001)에서 사전의 미시구조에 대한 구체적인 예를 든 것이다. 이 예에서처럼 사전 어휘항목을 구성하고 있는 일반적인 구성요소를 먼저 간단히 살펴본다.

사전에서 가장 기본적인 요소라고 할 수 있는 것은 표제어이다. 일반 사전의 경우 표제어는 알파벳순으로 배열되어 있다. 영어의 경우에는 동일한 표제어에 대하여 품사에 따라 발음이 다르게 실현되는 경우가 자주 나타나지만 독일어의 경우에는 이러한 경우가 매우 드물다. 단지 동음이의어 Homonym나 동철자어 Homograph일 경우에는 표제어와 함께 어께번호(1, 2, 3 ...)가 쓰이고 있다. 한편 표제어와 함께 나란히 표시되는 것 가운데 하나가 단어의 중요도이다. 이러한 단어의 중요도는 일반적으로 '*' 부호를 통해 그 중요성을 표시하고 있다. 일반적으로 독일어 외말사전에는 이러한 중

17) 전경재(1998), 구명철(2001) 참조

요도의 부호가 나타나지 않지만 대부분의 독한사전에서는 이러한 중요도의 표시를 쉽게 찾아 볼 수 있다. 이것은 두말사전에서의 이러한 표시가 외국어를 습득하는 과정에서 어떤 단어가 먼저 배워야 할 필수 단어인지를 아는데 도움을 주기 때문이다.

사전의 표제어 다음으로 제시되어야 할 것은 그 표제어의 발음이다. 독일어의 경우에는 표제어의 발음이 일반적으로 필요하지 않은 경우가 대부분이지만, 분절을 잘 구분하지 못하는 초보자의 경우에는 매우 유익하게 쓰일 수 있다¹⁸⁾. 이 외에도 전자사전의 특성을 살려서 발음정보와 관련하여 발음을 음성으로 변환하거나 또는 음성파일을 저장함으로써 해당 단어의 발음을 듣는 것이 가능하다.

표제어와 발음 뒤에는 일반적으로 문법범주에 대한 정보가 나타나야 한다. 문법범주와 관련된 정보로는 우선 단어의 품사를 생각할 수 있다. 일반적으로 인쇄사전에서는 명사인 경우에는 ‘명사’라는 품사정보가 생략된 채, 직접 그 명사의 성이 표시되고 있다. 그리고 이와 나란히 명사의 단수 2격과 복수 1격의 어미가 해당 단어의 활용정보를 위하여 제공되고 있다. 반면에 동사나 형용사의 경우에는 품사정보 이외에 활용정보는 거의 제시되어 있지 않다. 단지 동사에서는 그 동사가 ‘haben’동사를 필요로 하는 지, 아니면 ‘sein’동사를 필요로 하는 지에 대한 정보만 언급되어 있다(표 2).

<p>*Licht [lɪçt] (∇ leuchten; engl. light) <i>n.</i> -(e)s, -er, ① 빛, 밝기, 밝음, 광명(光明); 광휘(光輝), 광채(光彩); (Beleuchtung) 등불;</p>	<p>gēh[e]n* [gɛ:(ə)n] (ging, gegangen) [=engl. go] (I) <i>i.</i> (s.) ① 건다, 걸어가다; breitbeinig ~, 성큼성큼 걸다; gegangen kommen, 걸</p>
<p>*gēgen [gɛ:gən] <i>prp.</i> mit <i>acc.</i> († <i>dat.</i>) ① [방향] ...에 대하여, 면하여, 향하여, 쪽으로; ~ den Himmel, 하늘을 향하여; ~ den Wald</p>	<p>leicht <i>a.</i> [方] (add., alem.; leicht) 가벼운, 쉬운. lichter <i>adv.</i> (vielleicht) 아마.</p>

<표 2> 인쇄사전의 실례

한편 전치사는 일반적으로 격변화한 명사를 목적어로 취하기 때문에 전치사와 함께 나타나는 명사의 격변화가 표시되어 있다.

앞에서 언급한 인쇄사전에서 찾아볼 수 있는 문법정보와 관련된 사항은 독한 전자사전을 위한 미시구조를 고려할 때에도 필요한 구성요소이다.

이밖에도 앞의 문법정보와 함께 제시되어야 할 것은 어원에 관한 정보이다. 기존의

18) 'abästen'이 'a-bästen'으로 또는 'ab-ästen'으로 실현되는지 초보자의 경우에는 구분하기가 쉽지 않다.

인쇄사전에서는 표제어에 따라 어원정보가 나타나지 않는 경우도 종종 있다.

지금까지 언급한 표제어, 발음, 문법정보, 어원정보의 영역은 사전의 구조에 있어서 큰 영향을 미치고 있지 않으나, 의미정보는 사전의 거시구조와 미시구조의 설정에 있어서 매우 중요한 역할을 하고 있다.

즉 표제어의 배열과 관련이 있는 거시구조에서는 표제어가 의미정보와 관련하여 어떻게 배열되느냐에 따라 ‘단순 알파벳 구조 glattalphabetische Makrostruktur’, ‘등지 알파벳 구조 nestalphabetische Makrostruktur’, ‘벽감 알파벳 구조 nischenalphabetische Makrostruktur’로 나뉜다¹⁹⁾.

단순 알파벳 구조의 사전에서는 표제어를 단순히 알파벳 순서로 나열하기 때문에 표제어 검색이 용이하다는 이점이 있지만, 동일한 의미의 부류에 속하는 표제어의 관계를 파악하기가 쉽지 않다.

이와는 반대로 등지 알파벳 구조의 사전은 의미상 서로 연관이 있는 표제어를 하나의 등지 Nest로 묶어 나열을 한다. 한 표제어의 등지에 나열되는 하위 표제어들 사이에서는 알파벳 순서로 나열이 되어 있지만, 의미적 관련이 없는 표제어는 이 등지에 포함되어 있지 않아 전체적으로 보면 알파벳 순서가 뒤섞여 있다. 이 사전은 표제어 들간의 의미관계를 쉽게 파악할 수 있어 어휘 학습에 용이하지만, 초보자의 경우에는 표제어를 찾는데 약간의 어려움이 뒤따른다.

벽감 알파벳 구조란 앞의 두 가지의 장단점을 모두 고려한 구조라고 할 수 있는데, 벽감 알파벳 구조의 사전에서는 알파벳 순서를 엄격하게 지키면서 동일한 부류의 의미를 지닌 단어를 제한적으로 구별하여 나열한다. 따라서 표제어를 찾거나 표제어간의 의미관계를 파악하는 것이 모두 용이하다.

반면 한 어휘항목의 미시구조는 앞에서 언급한 발음, 문법정보, 어원정보, 그리고 곧 언급하게 될 의미정보 등 다양한 어휘항목의 구성요소가 어떻게 그 구조를 정할 것인가와 관련이 있다. 앞에서도 언급했듯이 이러한 다양한 구성요소 가운데 의미정보가 미시구조의 설정에도 중요한 역할을 하고 있다. 즉, 의미정보와 의미정보에 의존적인 용례나 성구의 배열모습에 따라 미시구조는 크게 ‘통합적 미시구조 integrierte Mikrostruktur’, ‘비통합적 미시구조 nichtintegrierte Mikrostruktur’, ‘부분통합적 미시구조 semiintegrierte Mikrostruktur’로 구분할 수 있다²⁰⁾.

통합적 미시구조에서는 용례나 성구가 표제어의 의미정보의 구분에 따라 해당 의미군에 통합적으로 배열된다. 반면에 비통합적 미시구조에서는 용례와 성구에 관한 내

19) 이 세 가지의 거시구조에 관해서는 구명철(1997), 전경재(1998), 김경(2002) 참조.

20) 미시구조의 자세한 설명은 구명철(1997), 김경(2002) 참조.

용이 표제어의 의미와는 상관없이 해당 표제어의 용례와 성구가 각각 개별적으로 구분되어 나타난다.

이 두 가지를 절충한 부분통합적 미시구조에서는 일정한 의미를 지닌 표제어의 용례가 해당 의미정보의 바로 뒤에 나타난다. 그러나 성구의 정보는 한꺼번에 따로 나타난다.

여기에서는 의미정보에 해당하는 내용을 논하기 위하여 앞에서 열거한 어느 특정한 구조나 특정 사전에 의존하지 않고, 일반적으로 의미내용에 포함될 수 있는 사항들만을 나열하고자 한다²¹⁾.

먼저 표제어의 의미가 여러 개인 경우에는 의미를 구분하여 정의할 수 있는 구분자가 있어야 하는데 이를 위하여 의미의 구분번호가 필요하다. 표제어가 다의어가 아닌 경우에는 의미의 설명을 위한 구분번호 없이 의미정보가 나타날 수 있다.

의미의 정의에 있어서는 우선 해당 표제어의 전문적인 사용 영역에 대한 정보가 표시되어야 한다. 이에 해당하는 것은 표제어에 대한 문체적 평가, 표제어가 사용되는 지역과 시대의 구별에 대한 설명, 그리고 그 표제어가 사용되는 전문영역에 대한 내용이 포함된다. 표제어의 사용영역에 대한 정보는 나타나지 않을 수도 있다.

다음으로 의미의 정의를 위해서는 실제로 표제어의 의미를 서술하여 풀이하는 의미풀이가 있다. 그러나 의미의 정의를 위하여 표제어의 의미를 기술적으로 풀어서 설명하지 않고, 동의어나 다른 표제어를 참조하게 함으로써 의미를 정의할 수도 있다.

이러한 경우에는 의미풀이 다음에 관련어라는 항목을 설정하여, 표제어와 의미적 상관관계(동의어, 반의어, 상위어, 하위어)에 있는 단어를 통하여 표제어의 의미를 정의할 수 있다.

의미에 대한 정의가 이루어지면, 실제로 그 표제어가 사용된 용례가 소개되어야 한다. 이때 용례에 대한 출처도 가능하면 표시하는 것이 좋다.

한편 용례와 성격이 유사하지만, 특정한 단어와 결합하여 자주 사용되는 성구에 대한 내용은 해당 언어를 모국어로 사용하지 않는 외국인 학습자에게는 필수적으로 설명되어야 할 필요가 있기 때문에 이를 간과해서는 안 된다.

마지막으로 사전 어휘항목의 미시구조를 구성하고 있는 요소로는 파생어에 대한 정보를 들 수 있다. 파생어에 대한 정보를 통하여 조어관계를 파악할 수 있으며 해당 파생어의 정보는 하이퍼텍스트의 형식으로 구체적인 의미를 파악할 수 있어서 표제어와 관련된 조어의 의미를 쉽게 파악하는데 도움을 줄 수 있다. 이것이 또한 전자사전의 강점이라고 할 수 있다.

21) 의미정보 설명을 위하여 어떠한 구조를 취할 지는 나중에 다시 거론한다.

이외에 표제어의 의미를 그림이나 동화상의 멀티미디어 요소를 통하여 설명할 수 있는 것도 가능하다. 이러한 사항은 구성성분요소를 위한 자리를 하나 추가하여 줌으로써 간단히 구현할 수 있지만, 이 논문에서는 생략하기로 한다.

지금까지 미시구조를 구성하고 있는 구성 요소를 간략하게 살펴보았는데, 이를 요약하여 정리 해보면 다음의 표와 같다.

표제어	어깨번호, 중요도
발음	분절구분에 유용, 음성파일 저장
문법 정보 (품사 표시)	명사(품사정보 생략, 명사의 성 표시, 명사의 단수 2격과 복수 1격의 어미가 활용 정보, 동사· 형용사의 경우에는 활용형에 대한 정보 없음), 동사('haben'이나 'sein' 선택 표시), 전치사(목적어로 나타나는 명사의 격변화 표시)
어원정보	표제어에 따라 어원정보가 나타나지 않는 경우도 있다.
의미정보	구분번호(생략 가능) 표제어의 사용 영역에 대한 정보(생략 가능) 표제어에 대한 문체적 평가(생략 가능) 표제어가 사용되는 지역과 시대의 구별에 대한 설명(생략 가능) 표제어가 사용되는 전문영역에 대한 내용 의미풀이: 풀이 없이 관련어(동의어)로 설명하기도 함 관련어: 의미적 상관관계(동의어, 반의어, 상위어, 하위어)어
용례	용례에 대한 출처도 가능하면 표시
성구	외국어 학습자에게는 필수적인 요소
파생어 정보	하이퍼텍스트를 통한 표제어와 조어간의 의미 파악 가능

<표 3> 사전 어휘항목의 구성요소

2.2.4. 독한 전자사전의 미시구조 설정

여기에서는 II.2.3.에서 소개된 사전 어휘항목의 구성요소를 바탕으로 독한 전자사전 편찬을 위한 미시구조 설정에 대하여 살펴보도록 한다.

미시구조의 설정에 앞서 사전의 편찬을 위해서는 표제어의 배열과 관련된 거시구조를 언급하지 않을 수 없다. 그 이유는 인쇄되어 있는 독한사전의 경우에도 이미 1차적으로 거시구조가 알파벳 순서로 결정이 되어 있지만, 알파벳 순서의 배열에 따라서도 앞에서 언급한 바와 같이 3가지로 구분이 되기 때문이다. 인쇄사전의 경우에는 표제어의 검색에 있어서 사전을 일일이 찾아보아야 하는 번거로움이 있기 때문에, 일반

사용자에게는 벽감 알파벳 구조가 편이성을 제공하고 있다고 볼 수 있다. 그러나 전자사전의 경우에는 표제어를 검색하기 위해서 표제어를 입력하는 것으로도 검색이 충분하기 때문에 단순 알파벳 구조로 거시구조를 설정하는 것이 별로 문제가 되지 않는다. 따라서 인쇄사전에서는 등지 알파벳 구조나 벽감 알파벳 구조에서 표제어의 하위 범주에서 검색이 가능했던 표제어가 전자사전에서는 더 이상 어떠한 표제어의 그룹에 속하지 않고 독자적인 표제어로 분류되는 것이 가능하다.

또한 단순 알파벳 구조로 거시구조를 설정하는 것은 사전편찬에 있어서 의미정보를 고려해야 하는 번거로움을 덜어줌으로써 사전편찬을 더 용이하게 한다는 이점도 가지고 있다.

다음으로 단순 알파벳 구조의 틀 안에서 미시구조를 생각해 보도록 한다. 이 논문에서는 미시구조의 설정에 있어서 표제어를 포함하여 다루고 있다. 그것은 한 어휘항목을 구성하고 있는 요소에 표제어도 포함되기 때문이다.

이 표제어를 중심으로 의미정보를 구성하는 미시구조는 전자사전의 이점을 살려 부분통합적 미시구조의 형태를 선택한다²²⁾. 부분 통합적 미시구조를 선택하는 이유는 우선 용례의 사용을 해당 표제어의 의미가 설명되어 있는 영역에서 실제로 사용된 예문을 참조함으로써 표제어의 의미를 정확히 알 수 있기 때문이다. 반면에 성구의 경우에는 성구의 실제 사용이 표제어의 세분화된 의미에서 어느 영역에 해당하는지를 정확히 구분하기가 쉽지 않다.

정형화된 독한 전자사전의 편찬을 위하여 앞에서 언급된 어휘항목의 구성요소의 논리적인 구조가 정해져야 한다. 이러한 구조를 정하는 이유는 사전편찬작업을 일관되게 수행함으로써 사전 어휘항목이 통일된 모습을 지닐 수 있도록 하기 위함이다. 실제로 사전에서의 어휘항목에 대한 구조는 사전마다 매우 다양하게 나타난다. 다른 사전이 다른 구조를 지니는 것은 극히 자연스러운 일이지만, 같은 사전에서조차도 동일한 품사에 대하여 다양한 형태의 구조가 종종 나타난다. 이처럼 각 어휘 항목들의 구조가 사전마다 다르고, 또한 한 사전에서도 매우 다양하게 나타나기 때문에, 일반 사전을 정형화하기 위해서는 모든 사전에 응용될 수 있는 공통적 구조를 찾아내어 코드화하는데 중점을 두어야 한다. 여기서 주목해야 할 것은 우리가 사전의 필요한 정보를 효과적으로 다시 추출하기 위해서는 사전의 어휘 정보를 컴퓨터에 어떠한 방식으로 코드화하느냐는 것이다.

22) 그러나 이 논문에서 사전을 구조화하여 어휘정보를 입력할 경우, 필요에 따라서는 비통합적 미시구조로 쉽게 구조를 바꿀 수 있다. 전자사전의 구축을 위해서는 비통합적 미시구조가 더 효율적일 수도 있으며, 개발하는 과정도 쉬울 수 있으나, 이 논문에서는 부분통합적 미시구조를 선택한다.

어휘정보의 코드화를 위해서 이 논문에서는 전 세계적으로 문서호환의 표준화를 지향하고 있는 XML을 이용하며, 정형화된 사전의 논리적인 구조를 XML의 문서유형정의로 나타낼 수 있다.

<code><!DOCTYPE woerterbuch SYSTEM "dkw.dtd" [</code>	
<code><!ELEMENT wortartikel -- (lemma, definition)* ></code>	어휘항목
<code><!ELEMENT lemma -O (#PCDATA) ></code>	표제어
<code><!ELEMENT definition -- (orth, gram, sem*, phras*, wbnng*) ></code>	정의부분
<code><!ELEMENT orth -- (lem_nr*, lem_w?) ></code>	형태정보
<code><!ELEMENT lem_nr -O (#PCDATA) ></code>	표제어 어깨번호
<code><!ELEMENT lem_w -O (#PCDATA) ></code>	표제어 중요도
<code><!ELEMENT gram -- (ausspr, silbetr, syn, etrn?) ></code>	문법정보
<code><!ELEMENT ausspr -O (#PCDATA) ></code>	발음
<code><!ELEMENT silbetr -O (#PCDATA) ></code>	분절표시
<code><!ELEMENT syn -- (pos_n pos_v pos_a pos_p pos_etc) ></code>	품사정보
<code><!ELEMENT etm - O (#PCDATA) ></code>	어원정보
<code><!ELEMENT pos_n -- (pos, genus, gen_2, pl_1) ></code>	명사정보
<code><!ELEMENT pos -O (#PCDATA) ></code>	품사
<code><!ELEMENT genus -O (#PCDATA) ></code>	명사의 성
<code><!ELEMENT gen_2 -O (#PCDATA) ></code>	명사의 단수 2격
<code><!ELEMENT pl_1 -O (#PCDATA) ></code>	명사의 복수 1격
<code><!ELEMENT pos_v -- (pos, infl, ha_se)</code>	동사정보
<code><!ELEMENT infl -O (#PCDATA) ></code>	활용정보
<code><!ELEMENT ha_se -O (#PCDATA) ></code>	haben, sein의 선택정보
<code><!ELEMENT pos_a -- (pos, komp) ></code>	형용사 정보
<code><!ELEMENT komp -O (#PCDATA) ></code>	형용사 비교급
<code><!ELEMENT pos_pe -- (pos, obj_gen) ></code>	전치사 정보
<code><!ELEMENT obj_gen -O (#PCDATA) ></code>	목적어의 격
<code><!ELEMENT pos_etc -O (#PCDATA) ></code>	기타 품사정보
<code><!ELEMENT sem -- (sem_nr, fach*, bed, sem_rel*, zb)></code>	의미정보
<code><!ELEMENT sem_nr -O (#PCDATA) ></code>	의미구분번호
<code><!ELEMENT fach -O (#PCDATA) ></code>	표제어 사용영역
<code><!ELEMENT bed -O (#PCDATA) ></code>	의미풀이
<code><!ELEMENT sem_rel -- (synon anton hypern hypon) ></code>	관계어 정보
<code><!ELEMENT synon -O (#PCDATA) ></code>	동의어
<code><!ELEMENT anton -O (#PCDATA) ></code>	반의어
<code><!ELEMENT hyperon -O (#PCDATA) ></code>	상위어
<code><!ELEMENT hypon -O (#PCDATA) ></code>	하위어
<code><!ELEMENT zb -- (zb_text, zb_ql?) ></code>	용례 정보
<code><!ELEMENT zb_text -O (#PCDATA) ></code>	예문
<code><!ELEMENT zb_ql -O (#PCDATA) ></code>	예문 출처
<code><!ELEMENT phras -- (phrs, phrs_bed?) ></code>	성구정보
<code><!ELEMENT phrs -O (#PCDATA) ></code>	성구
<code><!ELEMENT phrs_bed -O (#PCDATA) ></code>	성구풀이
<code><!ELEMENT wbnng -- (wb_lemma, wb_art) ></code>	조어정보
<code><!ELEMENT wb_lemma -O (#PCDATA) ></code>	조어표제어
<code><!ELEMENT wb_art -O (#PCDATA) ></code>	파생의 종류
<code>] ></code>	

<표 4> 사전 어휘항목을 위한 문서유형정의

독한 전자사전의 문서유형정의를 설정하기 위해서는 사전 어휘항목을 구성하는 항목을 최소한의 요소²³⁾로 나누어 이것을 사전의 미시구조 요소로 확정한다.

II.2.3.의 내용을 바탕으로 독한 전자사전에 포함되어야 할 정보를 <표 4>와 같은 문서유형정의로 그 구조를 설정할 수 있다.

이 문서유형정의에서 ‘,’ 기호는 스트링이 다음에 연결되어 나타날 수 있음을 의미하며, ‘|’ 기호는 여러 개 중에서 선택할 수 있는 것을 나열할 때 쓰인다. 또 ‘*’ 기호는 해당요소가 한 번도 나타나지 않거나, 무한히 나타날 수 있음을 표시하는 반면, ‘+’ 기호는 해당 요소가 최소한 한 번 나타나야 하며, ‘?’ 기호는 해당 요소가 최대한 한 번 선택적으로 나타날 수 있음을 표시한다. 이러한 규약을 바탕으로 <표 4>의 문서유형 정의가 의미하는 바는 대략 다음과 같다.

우선 최상위 요소로 어휘항목은 표제어 Lemma와 정의 Definition로 이루어진 것을 알 수 있으며, 표제어와 정의는 최소한 한 번 나타나야 한다는 것을 알 수 있다. 즉 표제어와 정의를 포함하는 어휘항목이 사전의 최소한의 기본 구성요소라고 할 수 있다.

표제어는 사전의 검색어로 이용되며, 정의부분은 형태정보, 문법정보, 의미정보, 성구정보, 그리고 조어정보로 나누어질 수 있다.

형태정보에는 동철자어를 위한 어깨번호 표시와 단어의 중요도를 나타낼 수 있다.

문법정보에는 발음, 분절표시, 품사와 그에 따른 부수적인 정보, 그리고 어원정보도 포함된다.

문법범주의 하위정보에 속해있는 품사정보에는 명사, 동사, 형용사, 전치사, 기타 품사에 해당하는 정보가 선택적으로 나타날 수 있음을 보여 준다.

의미정보에는 의미를 구분하는 의미 구분번호, 표제어의 사용 영역, 의미풀이, 관계어, 용례를 들 수 있다.

그리고 성구정보에는 해당 표제어가 들어있는 성구의 예와 그 풀이가 속해 있으며, 조어정보에는 표제어를 기본으로 하는 관련조어의 정보가 들어 있다.

이 문서유형정의에 따라 어휘항목의 한 예 ‘gesund’를 다음과 같이 나타낼 수 있다.

```
<wortartikel>
  <lemma> gesund </lemma>
  <definition>
    <orth>
      <lem_w> * </lem_w>
    </orth>
  <gram>
```

23) 사전의 어휘 정보를 세분화하면 할수록 우리는 세분화된 정보를 더 잘 사용할 수가 있다.


```

<ausspr>gazúnt</ausspr>
<silbetr>ge-sund</silbetr>
<syn>
  <pos>a</pos>
  <komp>gesúnder; gesunder</komp> </syn>
  <etm> ahd. gisunt; engl. sound </etm>
</gram>
<sem>
  <sem_nr>1</sem_nr>
  <bed> 건강한, 튼튼한, 기운 좋은; (정신이) 건전한; (목재 따위가) 벌레먹지
  않은, 썩지 않은 </bed>
  <zb>
    <zb_text> -e Farbe, 건강한 혈색; -er Schlaf, 숙면; -e Ansichten, 온당한
    견해; der -e Menschenverstand, 상식 </zb_text> </zb>
  <sem_nr>2</sem_nr>
  <bed> 건강에 좋은, 위생적인; 유익한, 효험 있는, 자양이 되는</bed>
  <zb> <zb_text> -es Klima, 건강에 좋은 풍토; -</zb_text> </zb> </sem>
</phras>
  <phrs> - wie ein Fisch im Wasser </phrs>
  <phrs_bed> 원기 발랄한 </phrs_bed>
  <phrs> ein -er Geist in e-m -en Körper </phrs>
  <phrs_bed> 건전한 정신은 건전한 신체에 있다. </phrs_bed>
  <phrs> das ist ihm ganz - </phrs>
  <phrs_bed> a) 그것은 그의 건강에 도움이 된다. b) [俗]그것은 그에겐 좋은
  교훈이다. </phrs_bed>
  <phrs> sich - machen </phrs>
  <phrs_bed> a) 건강해지다. b) [俗] 지금 능력이 생기다 금전 형편이
  좋아지다. </phrs_bed> </phras>
</definition>
</wortartikel>

```

3. 결론

지금까지 독한 전자사전을 위한 미시구조의 설정에 대하여 간단히 살펴 보았다. 여기서 제시한 미시구조는 단지 전자사전만을 위한 구조가 아니라 일반 사용자들을 위한 인쇄사전의 편찬을 위하여서도 필요하다고 생각한다.

이 논문에서는 전자사전의 편찬을 위하여 거시구조의 관점에서는 단순 알파벳구조를 중심으로 표제어를 배열하고, 개별 어휘항목의 서술에 있어서는 부분통합적 미시구조를 바탕으로 어휘정보를 기술하고자 하였다. 이것은 전자사전의 이점을 최대한

살리며, 사전편찬작업을 좀 더 용이하게 하고자 하는 입장에서이다. 거시구조를 단순 알파벳 구조로 설정한 것에는 전자사전의 편찬이라는 입장에서 앞으로도 큰 이점이 없으리라 여겨지지만, 미시구조의 설정에 있어서는 다양한 주장이 제시되리라 기대한다.

이 논문에서는 어휘항목을 위한 완전한 미시구조를 제시하기보다는 미시구조의 설정이 사전편찬의 일관된 작업을 위하여 얼마나 중요한 지를 부각시키면서, 앞으로의 사전편찬작업을 위하여 미시구조의 설정에 대한 토론의 장을 마련하고자 함이다.

따라서 이 논문에서 제시한 미시구조의 문제점과 다른 관점에서 보는 다양한 이견과 반박이 제시됨으로써 보다 보편적이고 객관적인 미시구조가 완성될 수 있다고 확신한다.

참고문헌

- 구명철 (1997): 독한사전의 사전학적 모델. 외국어로서의 독일어 2집272-294쪽.
 구명철 (2001): 우리나라 독한사전의 역사적, 사전학적 조명. in 차봉희(위음), 『한국의 독일문학 수용 100년, 1』. 한신대학교 출판부, pp 97-116.
 구명철 (2002): 디지털사전 구현을 위한 방향 설정 및 모델링. - 독일어 어휘사전을 예로 들어. in 『외국어교육(Foreign Languages Education)』 제9권, 제2호, 한국외국어교육학회, 2002년(pp. 343-361).
 김 경 (2002): 독-한 두말사전의 미시구조 조직. 독어학 제5집. 183-202쪽.
 양도원 (1999): DaF 독한사전 모델 개발 연구. 외국어로서의 독일어 5집. 30-54쪽
 이해옥 (2001): 주제별 분류사전 편찬기술방법 연구. 독어학 제4집. 395-411쪽.
 전경재 (1998): 알파벳에 의한 독일어 사전에 있어서 표제어의 설정 및 배열. 독일언어문학. 제10집. 143-162쪽.
 천미애 (2001): 사전의 미시구조. - 성구소 사전정보의 문제점과 개선안 - 독어교육 22집. 1-17쪽.
 최병진 (2001): 독어독문학 학술자료 정보화를 위한 방법론적 고찰. 독일문학 제80집. 396-415쪽.
 최병진 (2002): 어휘정보구축을 위한 사전텍스트의 구조분석 및 변환. 언어와 정보 6권 2호. 33-56쪽.
 Bierwisch, M. (1983): Semantische und konzeptuelle Repräsentationen lexikalischer Einheiten, Studia Grammatica XXII, Berlin.
 Chomsky, N. (1965): Aspects of the Theory of Syntax, Cambridge, Massachusetts
 Christoph Schwarz/Dieter Wunderlich(Hrsg.) (1985): Handbuch der Lexikologie, Königstein-Ts: Athenäum.
 Fraas, C./ Haß-Zumkehr, U. (1998): Vom Wörterbuch zum lexikalischen Infor-

- mationssystem: LEXXIS - ein neues Projekt des Instituts für deutsche Sprache. in Deutsche Sprache Jg. 26, H4. Zeitschrift für Theorie, Praxis, Dokumentation. S. 289 - 303.
- Gibbon, D. (2000): Computational Lexicography. in: Frank Van Eynde & Dafydd Gibbon eds., *Lexicon Development for Speech and Language Processing*. Kluwer Academic Publishers.
- Hallig, R., von Wartburg, W. (1963): *Begriffssystem als Grundlage für die Lexicographie. Versuch eines Ordnungsschemas*, Berlin.
- Hausmann, F. J. (1985): *Lexikographie*, in: Christoph Schwarz/Dieter Wunderlich(Hrsg.): *Handbuch der Lexikologie*, Königstein- Ts: Athenäum.
- Katz, J., and J.A. Fodor (1963): The Structure of a Semantic Theory, in *Language* 39: 170-210.
- Lutzeier, P. R. (1995): *Lexikologie: ein Arbeitsbuch*, Tübingen.
- Olney, J. & Ramsay, D. (1972): From machine-readable dictionaries to a lexicon tester: progress, plans, and an offer, in: *Computer Studies in the Humanities and Verbal Behavior* 3,4, 213 -220.
- Quillian, M. R. (1968): Semantic Memory, in: M. Minsky (ed.), *Semantic Information Processing*, Cambridge, Mass: MIT Press, 227-270.
- Roget, P. (1959): *Thesaurus of English Words and Phrases*. London: Dent & Sons.
- Schaeder, B. (1981): *Lexikographie als Praxis und Theorie*
- Schank, R. C. (1972): Conceptual Dependency: A Theory of natural language understanding, in: *Cognitive Psychology* 3.
- Schippan, T. (1992): *Lexikologie der deutschen Gegenwartssprache*, Niemeyer.
- Wiegand, H. E. (1998): *Wörterbuchforschung*, Walter de Gruyter.
- Wilks, Y. (1972): *Grammar, Meaning and the Machine Analysis of Language*, London.
- Winograd, T. (1972): *Understanding Natural Language*, Academic Press, New York.

Zusammenfassung

Eine Untersuchung zur Mikrostruktur der lexikalischen Einträge für ein elektronisches Deutsch-Koreanisches Wörterbuch

Choi, Byung-Jin(Mokpo national Univ.)

In diesem Aufsatz wird versucht, eine Mikrostruktur für ein elektronisches Deutsch-Koreanisches Wörterbuch zu entwerfen.

Bei gedruckten Wörterbüchern sind die lexikographischen Daten und die Wörterbuchstrukturen mehr oder weniger standardisiert, dagegen bei elektronischen Wörterbüchern strikt formalisiert.

Im Laufe der letzten 50 Jahre sind zwar viele gedruckte, jedoch kaum auf Elektronik basierte zweisprachige Wörterbücher für Deutsch-Koreanisch erschienen. Elektronische Wörterbücher haben mehr Vorteile und sind in diesem Informationszeitalter als Hilfsmittel für den Fremdspracherwerb unabdingbar.

Die Festlegung der logischen Struktur des Wörterbuchs ist für die Wörterbucharbeit unerlässlich und für die Herstellung elektronischer Wörterbücher von Bedeutung, wird jedoch in der koreanischen Praxis oft vernachlässigt. Daher weisen die meisten koranischen Wörterbücher keine konsistente Form in der Struktur von Wörterbuchdefinitionen auf.

Aus diesen Gründen schlage ich eine formale logische Struktur für ein elektronisches Deutsch-Koreanisches Wörterbuch vor.

Zunächst werden grundlegende lexikalische Begriffe eingeführt, anschließend wird auf die Analyse der lexikalischen Einträge für ein Deutsch-Koreanisches Wörterbuch eingegangen.

In diesem Beitrag wird keine fertige Mikrostruktur präsentiert, sondern eine Anregung zur weiteren Diskussion zu einer Mikrostruktur gegeben.

Als logische Struktur für das elektronische Deutsch-Koreanische Wörterbuch wird eine glattalphabetische Makrostruktur aufgenommen, die Mikrostruktur als semiintegriert.

Die lexikographische Arbeit ist zeitaufwändig und braucht viele Fachkräfte. Deshalb ist die Festlegung einer konsistenten und objektiven Mikrostruktur eine Voraussetzung der Wörterbuchherstellung. Sie regt auch dazu an, die lexikalischen Informationen im Artikel genauer zu analysieren. Erst dadurch wird ein effektives elektronisches Wörterbuch möglich, das eine Weiterbenutzung der klassifizierten Informationen unterstützt.

[검색어] 사전편찬학, 독한사전, 전자사전, 미시구조
Lexikographie, Deutsch-Koreanisches Wörterbuch, elektronisches Wörterbuch,
Mikrostruktur

최병진

534-729

전남 무안군 청계면 도림리 61

목포대학교 어문학부 독일언어문화학 전공

bjchoi@apollo.mokpo.ac.kr