

## 결합설명 사전의 어휘 기술 방법론\*

이 병근·박진호  
(서울대학교 국어국문학과)

### 1. 서론

#### 1.1 다양한 형태의 사전이 필요하다

사전은 해당 언어에 대한 연구 성과가 집약되어 나타난 결과물인 동시에, 그 언어에 관한 정보를 필요로 하는 사람들(여기에는 일반인과 전문가, 특히 언어학자가 모두 포함된다)에게 유용한 정보를 제공해 주는 언어 정보의 원천이기도 하다. 그래서 어떤 언어에 대해 나와 있는 사전들을 보면 그 언어에 대한 연구의 깊이와 폭을 알 수 있고 그 언어를 사용하는 사회의 문화 수준도 짐작할 수 있다. 언어란 것이 여러 층위의 요소들이 서로 맞물려 있는 매우 복잡한 체계이기 때문에, 그리고 사람들이 하나의 언어에 대해 필요로 할 수 있는 정보의 종류가 매우 다양하기 때문에, 언어 정보의 집합체로서의 사전도 매우 다양한 모습, 다양한 용도를 가질 수 있다. 일반인들의 경우 책을 읽다가 만난 어떤 단어의 정확한 의미를 알기 위해 사전을 찾을 수도 있고, 글을 쓰다가 어떤 단어의 정확한 철자를 확인하기 위해 사전을 뒤적일 수도 있으며, (특히 외국어의 경우) 어떤 단어의 정확한 발음을 알기 위해 사전을 참조할 수도 있다. 전문 언어학자의 경우에도 사전을 필요로 하는 경우는 매우 다양하다. 역사언어학자들에게는 古語 사전, 語源 사전이 필요할 것이고, 어휘의미론 학자들에게는 단어들 사이의 의미 관계를

---

\* 이 연구는 1998년도 한국학술진흥재단 자유공모과제 공동연구지원을 받아 이루어졌다. 이 논문을 준비하는 데 김현 선생과 김진형 선생의 도움이 컸음을 밝혀 둔다.

알려 주는 유의어 사전, 반의어 사전, 그리고 단어들을 의미 영역에 따라 분류해 놓은 시소러스thesaurus 등이 유용할 것이고, 통사론 학자들은 어떤 동사가 어떤 문형으로 사용되는지 알아보기 위해, 또는 적절한 용례를 찾기 위해 사전을 이용할 수도 있다. 2개어 사전bilingual dictionary의 경우에도 독해에 도움을 주기 위한 사전과 작문에 도움을 주기 위한 사전은 매우 다른 모습을 지닐 것이다.

사전 이용자의 요구가 이처럼 다양하기 때문에 어느 하나의 사전에서 이 모든 요구를 만족시키기란 어려운 일이고, 각각의 용도에 맞게 특화된 다양한 사전이 존재하는 것이 바람직할 것이다. 실제로 영어의 경우 사전에 대해 전세계적으로 엄청나게 많은 사람들의 다양한 수요가 있기 때문에 매우 다양한 종류의 사전이 나와 있다. 그러나 한국어의 경우 아직 다양한 종류의 사전이 나와 있지 못하다. 사전 이용자들의 다양한 요구에 부응할 수 있게끔 특화된 다양한 종류의 사전을 개발하는 것은 우리 국민들이 좀더 수준 높은 언어 생활을 영위하기 위해서, 그리고 우리 문화와 학문이 한 단계 더 높은 수준으로 도약하기 위해서 필수적이고 시급한 일이다.

최근에는 일반인이나 전문 언어학자 외에 사전을 절실하게 필요로 하는 수요자가 하나 더 생겼다. 그것은 바로 컴퓨터이다. 최근 컴퓨터가 우리의 일상 생활 구석구석에 깊숙이 침투하고 있고 컴퓨터가 활용될 수 있는 일의 범위에 대해 무한한 가능성이 운위되고 있지만, 앞으로 컴퓨터를 사용하는 인구가 질적으로 성장하기 위해서는 사용자 인터페이스user interface의 문제가 근본적으로 해결되어야 할 것이다. 현재 컴퓨터는 인간의 말을 그대로 알아들을 수가 없기 때문에 인간이 컴퓨터를 사용하기 위해서는 컴퓨터가 알아들을 수 있는 형태의 언어를 배워서 컴퓨터에게 지시를 내려야 한다. 바로 이 점이 사람들로 하여금 컴퓨터를 사용하기 어려운 것으로 여기게끔 하는 근본 원인이다. 사람이 컴퓨터에게 맞출 것이 아니라 앞으로 만약 컴퓨터가 인간의 말을 알아들을 수 있게끔 할 수 있다면, 누구든지 컴퓨터를 손쉽게 이용할 수 있을 것이고 컴퓨터에 대한 수요와 이용 가능성은 폭발적으로 확대될 것이다. 그래서 컴퓨터가 인간의 말을 알아들을 수 있게 하기 위해 자연 언어 처리natural language processing에 대한 연구가 활발히 진행되고 있다. 그런데 인간의 언어라는 것이 엄청나게 많은 양의 정보를 담

고 있는 매우 복잡한 체계이기 때문에, 어떤 한 언어와 관련된 정보, 즉 컴퓨터가 그 언어를 알아듣고 유창하게 말할 수 있게 하기 위해 필요한 정보를 빠짐없이 체계적으로 구성하는 것, 그리고 그 지식을 컴퓨터가 저장하고 활용할 수 있는 형태로 만드는 것은 엄청나게 어려운 일이다. 자연 언어 처리 시스템을 구축하는데에 있어 현재 가장 부족한 것이 바로 컴퓨터가 읽을 수 있는 형태의 사전, 즉 기계 가독형 사전(machine-readable dictionary)이다. 한 언어의 통사 규칙이나 음운 규칙은 그 나름대로 복잡성을 가지고 있기는 하나 비교적 적은 수의 규칙들로 형식화할 수 있다. 반면에 어휘(vocabulary)라는 것은 數的으로 방대할 뿐 아니라 본질상 규칙으로 말끔히 설명될 수 없는 정보들로 이루어져 있기 때문에 하나하나 일일이 정보를 명시해 주어야 한다. 그래서, 컴퓨터가 자연 언어를 제대로 처리하기 위해 필수적으로 요구되는 사전은 어떤 모습이어야 하는가에 대해서 많은 연구가 이루어져야 할 것이고, 그렇게 만들어진 사전의 像에 맞추어 어휘에 대한 철두철미한 기술이 이루어져야 한다.

## 1.2 결합설명 사전의 장점

사전은 한 언어를 구성하고 있는 어휘 요소들에 관한 갖가지 정보의 총체이므로, 우리는 자신의 모국어 또는 외국어에 관한 어떤 정보가 필요할 때 이 정보를 그 언어의 사전으로부터 얻을 수 있는 경우가 종종 있다. 그런데 지금까지 존재했던 사전들은 대개 언어 이해(compréhension)에 도움을 주는 역할을 해 왔을 뿐, 언어 산출(production)에 도움을 주기 위해 고안된 사전은 별로 없었다. 예를 들어, 책을 읽다가 의미를 잘 모르는 단어가 나오면 그 단어의 의미를 파악하기 위해 사전을 찾는 일은 흔히 있어도, 어떤 의미를 표현하기에 적합한 단어를 찾기 위해 사전을 참조하는 일은 별로 없다. 그것은 실제 언어 생활에서 후자의 필요성이 별로 없기 때문이라기보다는 그런 요구에 적합한 사전에 대해 사람들이 별로 생각해 보지 않았기 때문이다. 수요가 공급을 창출하기도 하지만, 공급이 수요를 창출하는 일도 종종 있다. 산출에 초점을 맞춘 사전이 나온다면, 이런 사전에 대한 수요와 이용 가능성은 엄청날 것이다.

어떤 의미를 표현하기에 적합한 언어 표현을 찾는 데에 사전이 도움을 줄 수

있는 상황은, 모국어 사용의 경우에도 일어날 수 있지만 외국어 사용의 경우에 더 극명하게 드러난다. 예를 들어, 미국인 학자가 한국에서 개최되는 학술대회에 와서 한국어로 논문을 발표한다고 하자. 그 학자는 한국어에 대한 초급적인 지식은 가지고 있지만 한국어를 유창하게 구사하지는 못한다. 그가 자신의 원고를 어렵사리 한국어로 작성할 수는 있겠지만, 그 원고는 제대로 된 한국어라고 할 수는 없을 것이다. 기존의 영한사전이나 한영사전도 큰 도움이 되지는 못할 것이다. 두 언어의 번역 대응어(translation equivalent)들을 제시해 놓았을 뿐, 비슷비슷한 의미를 갖는 여러 단어들 중에서 주어진 구체적인 문맥에 가장 어울리는 단어가 무엇인지 알려 주지는 않기 때문이다. 바로 이런 상황에서 이 학자가 자신의 거친 원고를 유려한 한국어로 가다듬는 데 도움을 주는 사전이 있다면 실용적 가치가 매우 높을 것이다.

기계 번역 프로그램을 개발하는 데에 있어서도 비슷한 문제가 발생한다. “썩은, 부패한”의 의미를 나타내는 영어 단어들로 *rot*, *rotten*, *spoilt*, *decayed*, *decomposed*, *putrefied*, *putrid*, *mouldered*, *rancid* 등이 있는데 이들은 아무렇게나 서로 대치되어 쓰일 수 있는 것이 아니라 피수식 명사가 무엇이냐에 따라 제약되어 쓰인다. 또한, 한국어에는 “死亡”의 의미를 나타내는 동사로 ‘운명하다’, ‘서거하다’, ‘작고하다’, ‘별세하다’, ‘돌아가(시)다’, ‘죽다’, ‘가다’, ‘(나가)돼지다’, ‘숨을 거두다’ 등 여럿이 있는데 주어에 해당하는 사람이 누구인가에 따라 선별적으로 사용된다. 이런 상황에서 주어진 문맥에 가장 적절한 단어가 무엇인지를 알려 주는 사전은 고품질의 기계 번역 시스템을 만드는 데에 필수적이다.

현대 자본주의 사회에서는 사전이 시장에서 팔려야 살아남을 수 있는 상품이기 때문에 위에서 말한 것과 같은 실용적인 목적을 염두에 두지 않을 수 없지만, 사전의 가치가 그런 실용적인 측면에만 있는 것은 아니다. Chomsky 학파의 대두 이래로 현대 언어학자들은 화자가 자신의 모국어에 대해 지니고 있는 지식(언어능력)을 명시적인 형태로 표상하는 것을 언어학의 주된 목표로 생각하고 있다. 그런데 화자의 모국어에 대한 지식 중에서 통사론적 지식, 음운론적 지식에 대해서는 많은 연구가 이루어져 왔고 이를 표상하기 위한 매우 구체적이고 정밀한 이론적 제안들이 많이 제시되어 왔지만, 사람들이 자신의 모국어 어휘에 대해 가

지고 있는 어휘적 지식, 즉 어휘부에 대해서는 상대적으로 관심이 적었고 최근에야 이에 대한 관심이 차츰 고조되고 있는 형편이다. 즉, 어휘부가 전체적으로 어떤 구조를 지니고 있으며 여기에 저장되어 있는 단어들에 관한 갖가지 정보가 어떤 방식으로 표상되어야 하는가 하는 것은 현대 언어학에서 매우 중요한 이론적 문제라 할 수 있다. 화자들이 자신의 모국어 어휘에 대해 지니고 있는 지식을 총체적으로 모형화(modelling)하는 것, 몇몇 소수의 어휘에 대해 단편적인 기술을 하는 데 그치는 것이 아니라 한 언어의 어휘 체계 전반에 대해 철저한 기술을 행하는 것, 그 기술의 結晶體인 사전을 만드는 것은 이론 언어학적 관점에서 볼 때에도 매우 가치 있고 절실한 과제이다.

지금까지 말한 바와 같은 실용성과 이론적 타당성을 갖춘 새로운 형태의 한국어 사전을 만들기 위한 기초 연구로서, 본고는 I. A. Mel'čuk과 그의 동료들이 개발한 결합설명 사전Dictionnaire Explicatif et Combinatoire, DEC의 어휘 기술 방법론을 살펴보고자 한다. Mel'čuk은 오래 전에 러시아어의 결합설명 사전을 낸 바 있으며 1970년대에 Canada의 Université de Montréal에 정착한 뒤로는 프랑스어의 결합설명 사전 개발에 주력해 오고 있다. 지금까지 3권의 현대 프랑스어 결합설명 사전Dictionnaire Explicatif et Combinatoire du Français Contemporain, DECFC이 출간되었으며 DECFC 제4권이 곧 나올 예정이다.

결합설명 사전은 의미나 통사적 특성 등을 기술하는 데에 있어서도 매우 정밀한 방법론을 내놓고 있지만 가장 두드러진 특징은 단어들 사이의 어휘관계 lexical relation를 표상하는 데에 매우 많은 노력을 기울인다는 점이다. 기존의 사전들 또는 어휘부 모형은 개별 단어들의 특성을 원자적으로 기술하는 데에 그쳤다고 해도 과언이 아니다. 표제항으로 제시된 단어가 다른 단어들과 맺는 관계의 양상은 매우 소홀하게 취급되어 왔다. 결합설명 사전은 단어들이 서로 맺고 있는 어휘관계를 폭넓게 체계적으로, 그리고 명시적으로 기술하기 위한 방법론을 깊이 연구해 왔다. 그 결과 고안된 장치는 어휘함수fonction lexicale, FL라 불리는 데이에 대해서는 뒤에서 자세히 살펴볼 것이다. 어휘함수는 어휘관계 중에서도 특히 통합 관계, 즉 連語collocation 관계를 포착하는 데에 효과적인 것으로 정평이 나 있다. 그래서 어떤 수식어와 피수식어가 가장 잘 어울리는지, 어떤 술어 명사

와 어떤 기능동사가 공고한 결합 관계를 가지는지를 잘 드러내 준다. 그래서 앞서 말한 바와 같이 産出에 도움을 주고자 하는 목표에 잘 부합된다.

그리고 최근에는 결합설명 사전의 이론과 방법론에 대해 전산언어학자들의 관심이 높아지고 있다. 자연 언어 처리를 위한 전자 사전을 구축하는 데에 있어 결합설명 사전이 적절한 모형으로 이용될 수 있다는 생각에서이다. 컴퓨터 학자들과 결합설명 사전을 개발해 온 언어학자들 사이의 협동 연구가 진척되어, 최근에는 결합설명 사전의 지침에 따라 사전 항목을 기술하기 위한 사전 편집기가 개발되어 사용되고 있다고 한다. 이러한 사실은 한국어 사전 편찬의 방법에 대해서도 시사하는 바가 많다고 하겠다. 사전 편찬의 방법을 혁신하고 효율성을 제고하기 위해서도 컴퓨터를 적극적으로 활용해야 하며, 또한 언어 사전 편찬이 컴퓨터에 의한 자연 언어 처리에도 기여할 수 있어야 하는 것이다. 결합설명 사전은 그런 점에서 미래의 사전에 대한 像을 제시해 준다고 할 수 있다.

본고는 I. A. Melčuk과 그의 동료들이 개발한 결합설명 사전의 이론과 방법론을 살펴보고, 한국어 어휘 기술 및 사전 편찬에 활용될 수 있는 방안을 탐색해보려 한다. 이러한 연구를 바탕으로 하여 궁극적으로는 한국어의 구조와 특성에 적합한 사전 기술의 모형을 개발하는 것을 장기적인 목표로 생각하고 있다. 결합설명 사전은 기본적으로 이론적 어휘부에 대한 모형이지만 위에서 말한 바와 같은 실용성과 전산 응용성을 가지고 있기 때문에, 한국어에 이 방법론이 응용될 수 있다면 많은 유용한 결실을 맺을 수 있으리라 기대된다.

우선 2절에서는 결합설명 사전의 거시구조에 대해 살펴보고, 3절에서는 이 사전의 미시구조 및 여기에 제시되는 각종 정보, 이들 정보를 제시하는 방식과 관련된 원칙들을 살펴본다. 결합설명 사전에서 제시되는 여러 정보들 중 특히 어휘함수와 관련된 것은 4절에서 따로 자세히 논의한다. 5절에서는 결합설명 사전의 모형에 따라 한국어 단어 몇 개를 기술하여 제시함으로써, 이 사전의 형식과 특징에 대한 이해를 돕고자 한다.

## 2. 결합설명 사전의 거시구조

DEC는 거시구조(어휘요소의 배열), 미시구조(표제항에 대한 각각의 언어 정보의 구획), 언어 정보 제시 방식 등의 모든 면에 있어서 철저한 원칙에 입각해 있다. 우선 DEC 편찬의 지침이 되는 기본 원칙들을 살펴보겠다.

### 2.1 단어vocable와 어휘요소lexie

보통의 다른 사전들은 동형어들을 어깨번호로 구별하고 하나의 단어가 지니는 여러 의미들을 하나의 표제항 내에서 번호로 구별하는 방식을 대개 취한다. DEC에서도 하나의 단어가 여러 의미를 지닐 때(즉 다의어일 때) 이들을 한데 모아 배열한다는 점에서는 기존의 사전들과 비슷하다. 그러나 DEC에서 표제항이 되는 기본 단위는 여러 의미를 갖는 다의어가 아니라, 다의어가 갖는 여러 의미들 중 어느 하나의 특정 의미를 갖는 경우로 제한된 단어이다. 의미·텍스트 이론Meaning-Text Theory, MTT에서는 전자를 단어vocable, 후자를 어휘소lexème라고 부른다.<sup>1)</sup> 뒤에서 자세히 설명할 갖가지 정보도 어휘소를 단위로 하여 제시된다. 이것은 다른 사전들에서는 찾아보기 힘든 DEC만의 특징이라 할 수 있다.

예를 들어 설명하면, ‘다리1’(脚)과 ‘다리2’(橋)는 별개의 vocable이 되며, ‘다리1’의 의미 중 “사람의 신체부위로서의 다리”, “동물의 신체부위로서의 다리”, “탁자 등 가구의 다리” 등등은 vocable ‘다리1’에 속하는 어휘소들이 된다.

DEC의 거시구조가 보이는 또 하나의 두드러진 특징은, 둘 이상의 단어로 이루어진 숙어도 표제항으로 등재한다는 것이다. 이러한 요소를 MTT에서는 숙어소 phrasème라고 부른다. 다른 사전들에서는 이런 숙어를, 그 숙어 속에 포함된 핵심 단어의 부표제항 정도로 취급하는 것이 보통이다. 그러나 MTT에서는 숙어소도 둘 이상의 단어로 이루어져 있다 뿐이지, 의미나 결합상의 특성 등의 여러 측면에서 보면 어휘소와 다를 바가 없다고 보며, 그렇기 때문에 이 두 가지 단위

1) vocable의 적당한 번역어를 찾기 어려워 일단 단어라고 해 둔다. 어휘소의 개념도 보통 다른 언어학자들이 사용하는 개념과 차이가 있다.

모두 사전에서 표제항으로 등재하는 것이 타당하다고 주장한다. 그리고 어휘소와 숙어소를 아울러서 어휘요소lexie라고 부른다. 요컨대, DEC에서 언어 정보 제시의 단위가 되는 표제항은 어휘요소라고 할 수 있다.

어떤 두 개의 어휘소가 서로 다른 의미를 지니고 있다고 할 때, 이 둘을 하나의 *vocable*에 속하는 것으로 묶을 것인가, 아니면 별개의 *vocable*로 처리할 것인가 하는 것은 간단한 문제가 아니다. 이것은 바로 오래 전부터 많은 학자들 사이에 논란이 되어 온 동형어와 다의어의 구별 문제인 것이다. DEC에서 이 문제를 처리하는 기본 원칙은, 형태상으로 일치하는 두 어휘소가 정의*définition*에서 공통된 부분이 있으면 하나의 *vocable*로 묶고, 정의에 공통된 부분이 없으면 동형어로 간주하여 별개의 *vocable*에 소속시킨다는 것이다.

이 문제에 대한 명확한 기준을 제시하기 위해 MTT에서는 다음과 같은 개념들을 도입하고 있다.

- (1) 의미 가교*pont sémantique*:  $L_1$ 과  $L_2$ 라는 두 어휘요소가 있을 때 이 둘의 의미 사이에 존재하는 특징적인(자명하지 않은) 공통 성분을 둘 사이의 의미 가교라고 부른다.

여기서 ‘특징적’이라는 말이 무슨 뜻인지에 대해 보충 설명이 필요하다.

- (2) 특징적인 공통 성분: 두 어휘요소의 의미의 공통 성분은 다음의 3 가지 조건을 만족시킬 때 특징적인 공통 성분이라 할 수 있다.
- ① 문법적인 의미 또는 매우 일반적인 의미(“and”, “or”, “존재”, “사람” 등)이어서는 안 된다.
  - ② 두 의미에서 각각 차지하는 중요성이 비슷해야 한다.
  - ③ 두 의미에서 각각 비슷한 계층상의 기능을 해야 한다.

의미 가교의 개념을 이용하여 다음과 같이 어휘요소들 사이의 관련성, *vocable*, 기저 어휘요소 등을 정의할 수 있다.



- (3) 직접적/간접적 관련성 *lien direct/indirect*:  $\langle L_1 \rangle$ 과  $\langle L_2 \rangle$ 라는 두 의미 사이에 의미가 교차 존재하면, 이 둘은 직접적으로 관련되어 있다고 한다.  $\langle L_1 \rangle$ 과  $\langle L_2 \rangle$  사이에 직접적인 관련은 존재하지 않지만,  $\langle L_1 \rangle$ ,  $\langle L' \rangle$ ,  $\langle L'' \rangle$ ,  $\langle L''' \rangle$ , ...,  $\langle L_2 \rangle$ 과 같은 연쇄에서 각각의 인접한 둘 사이에 직접적인 관련이 존재하는 경우  $\langle L_1 \rangle$ 과  $\langle L_2 \rangle$ 는 간접적으로 관련되어 있다고 한다.
- (4) *vocable*:  $n$ 개의 어휘요소  $L_1, L_2, \dots, L_n$ 이 모두 記標 *signifiant*가 같고, 이들 중 어느 둘 사이에도 직접 또는 간접적인 관련이 존재할 때, 이  $n$ 개의 어휘요소의 총체는 하나의 *vocable*이 된다.
- (5) 기저 어휘요소 *lexie de base*: 어떤 *vocable*의 어휘요소들 중  $L$ 이라는 어휘요소가 있다고 할 때, 다른 모든 어휘요소들은 직접적/간접적으로  $L$ 을 참조하고  $L$ 은 다른 어휘요소를 전혀 참조하지 않을 때,  $L$ 을 이 *vocable*의 기저 어휘요소라고 한다.

## 2.2 어휘요소들의 배열 순서

하나의 *vocable* 내에서 어휘요소들을 어떠한 순서로 배열할 것인가 하는 문제에 대해서도 DEC는 명시적인 원칙을 제시하고 있다. 이 원칙들은 각 어휘요소의 의미 정의에 입각해 있으므로, 각 어휘요소의 의미 정의를 어떻게 내릴 것인지의 문제가 선결되어야 한다. 이 문제는 3.2.2에서 다룬다. 원칙들을 소개하고, 각각에 대해 예를 들어 설명해 보겠다.

원칙 7.2) 의미 포함의 원칙:  $L_1$ 의 정의가  $L_2$ 의 정의에 포함되면,  $L_1$ 이  $L_2$ 보다 앞에 온다.

자동사 ‘움직이-’와 타동사 ‘움직이-’를 비교해 보면, 전자의 정의가 후자의 정

---

2) 본고에서 소개하는 원칙이나 기준의 번호는 Mel'čuk et al.(1995)를 따른다.

의에 포함된다는 것이 분명하다. 따라서 전자를 후자보다 앞에 배열한다.

원칙 8. 의미 근접성의 원칙: 의미상  $L_1$ 이  $L_2$ 보다  $L$ (기저 어휘요소)에 가까울 경우,  $L$ ,  $L_1$ ,  $L_2$ 의 순서로 제시한다.

동사 ‘타다’의 경우 ‘종이가 타다’ 같은 예는  $L$ , ‘밥이 타다’ 같은 예는  $L_1$ , ‘얼굴이 붉벌에 타다’ 같은 예는  $L_2$ 라고 할 수 있다.<sup>3)</sup> 여기서  $L_1$ 이  $L_2$ 보다  $L$ 에 가까우므로  $L_1$ 을  $L_2$  앞에 위치시킨다.

형용사 ‘좋다’의 경우 ‘품질이 좋다(양호)’ 같은 예를 기저 어휘요소  $L$ 이라 할 수 있는데, ‘감기에는 이 약이 좋다(적합)’ 같은 예를  $L_1$ , ‘나는 철수가 좋다(호감)’ 같은 예를  $L_2$ 라 하면,  $L_1$ 이  $L_2$ 보다  $L$ 에 가까우므로  $L_1$ 을  $L_2$  앞에 위치시킨다는 것이다.

원칙 9. 환유 우선의 원칙:  $L$ 에 대해  $L_1$ 의 의미는 환유métonymie에 의한 확장이고  $L_2$ 의 의미는 은유métaphore에 의한 확장일 때,  $L$ ,  $L_1$ ,  $L_2$ 의 순서로 제시한다.

‘머리’라는 *vocable*의 경우 “사람의 신체부위로서의 머리”가 기저 어휘요소가 될텐데, 여기서 확장된 의미를 갖는 어휘소들 중에 “머리털/머리카락”과 “우두머리”가 있을 수 있다. 이 둘 중 전자는 기저 어휘요소의 환유적 확장이고 후자는 은유적 확장이므로 전자를 후자 앞에 배열한다는 것이다.

동사 ‘밟다’를 가지고 예를 들어 보면, “밭로 바닥을 누르다”를  $L$ , ‘고국 땅을 밟다(방문)’ 같은 예를  $L_1$ , ‘그는 내 의견을 완전히 밟아 버렸다(무시)’ 같은 예를  $L_2$ 라 하면,  $L_1$ 은  $L$ 의 환유적 확장이고  $L_2$ 는  $L$ 의 은유적 확장이므로  $L_1$ 을  $L_2$  앞에 배열한다.

3) 연세한국어사전 타다1 ㉠의 ①~③.

원칙 10. 의미 성분 위계의 원칙: L의 정의에 출현하는 의미 성분들 중 L<sub>1</sub>의 은유적 확장의 기반이 되는 의미 성분이 L<sub>2</sub>의 은유적 확장의 기반이 되는 의미 성분보다 앞에 오면, L<sub>1</sub>이 L<sub>2</sub>보다 앞에 온다.

‘날개’의 기저 어휘요소 L을 “새나 곤충의 몸통 양쪽에 붙은, 날 때에 펄럭거리는 부분”이라고 정의한다고 할 때, L<sub>1</sub>은 “비행기의 몸통 양쪽에 붙은, 날 때에 공기의 떠받치는 힘을 받는 부분”이라고 정의할 수 있고, L<sub>2</sub>는 “기계나 기구에 붙은, 바람개비 모양의 펄럭거리는 부분”이라고 정의할 수 있을 것이다(‘선풍기의 날개’ 같은 예). 여기서 L<sub>1</sub>과 L<sub>2</sub> 모두 L의 은유적 확장이지만, L<sub>1</sub>의 은유적 확장의 기반이 되는 “...의 몸통 양쪽에 붙은, 날 때에”라는 의미 성분이 L<sub>2</sub>의 은유적 확장의 기반이 되는 “펄럭거리는 부분”이라는 의미 성분보다 L의 정의 내에서 앞에 나오므로, L<sub>1</sub>을 L<sub>2</sub> 앞에 배열한다.

원칙 11. 특징적 성격의 원칙: 해당 어휘장에 대해 L<sub>1</sub>이 L<sub>2</sub>보다 특징적이라면 L<sub>1</sub>이 L<sub>2</sub>보다 앞에 온다.

‘주먹’이라는 *vocable*에 속하는 어휘소들 중에는 수량 단위 의존명사로서 “주먹I(신체 부위)에 담을 수 있는 양”이라는 의미를 갖는 것이 있다(‘강냉이 한 주먹’ 같은 예). 그런데 신체 부위의 어휘장에서 이러한 어휘소를 갖는 *vocable*은 별로 없다. 따라서 이 어휘소는 신체 부위 어휘장에서 매우 예외적이므로 *vocable* ‘주먹’의 다른 어휘소들보다 뒤에 배열된다.

### 3. 결합설명 사전의 미시구조

#### 3.1 어휘요소 분할의 기준

앞에서 말했듯이 DEC는 하나의 *vocable*을 우선 의미에 따라 어휘요소들로 구분하여 배열하고 있다. 따라서 DEC의 미시구조 구성에 있어서 가장 먼저 해결해야 할 문제는 하나의 *vocable*을 어휘요소들로 어떻게 분할할 것인가 하는 것이

다. DEC에서는 이 문제를 해결하기 위해 명시적인 기준들을 제시하고 있다.

기준 1. 다중 해석의 기준: 어휘요소 후보 L'(...σ'...σ"...)'을 포함하는 문장을 만들었을 때, 이 문장이 L로 인해 두 가지 해석을 가지며 이 중 한 해석은 (...σ'...)'을, 다른 한 해석은 (...σ"...)'을 포함할 경우, 이 L은 L<sub>1</sub>(...σ'...)'과 L<sub>2</sub>(...σ"...)'로 분리한다.

‘그는 어제 산 두 물건을 저울질해 보았다’라는 문장은 실제로 저울로 두 물건의 무게를 측정해 보았다는 뜻일 수도 있고 두 물건을 비교해 보았다는 뜻일 수도 있다. 그리고 이 중의성은 오로지 ‘저울질하다’라는 단어 때문에 생기는 것이다. 따라서 다중 해석의 기준에 의해 ‘저울질하다’는 두 개의 어휘소로 구분되어야 한다.

기준 2. 局所的/全局的 의미 차이의 기준: 어휘요소 L의 두 가지 용법이 보이는 의미 차이가 국소적일 때에는(즉 정의 내의 다른 의미 성분에 영향을 미치지 않으면) 하나의 어휘요소로 처리하고, 의미 차이가 전국적일 때에는(즉 정의 내의 다른 의미 성분에 영향을 미치면) 두 개의 어휘요소로 분리한다.

‘팔다’는 “돈을 받고 어떤 물건의 소유권을 영구적으로 이전하다”라는 의미(‘집을 팔다’ 같은 예)와 “돈을 받고 자기 몸의 性的인 소유권을 일시적으로 이전하다”라는 의미(‘몸을 팔다’ 같은 예)를 나타낼 수 있다. 그런데 이 둘의 의미 차이는 국소적이고 통사적 특성이나 어휘적 공기 관계에 있어서도 차이를 보이지 않으므로 이 기준에 따르면 이 둘을 별개의 어휘소로 분리하지 않는다.

기준 3. 양립가능한 共起의 기준: 어휘요소 후보 L'(...σ'...σ"...)'을 사용하여 양립 가능한 공기를 가진 문장(즉 두 의미 성분 (...σ'...)'과 (...σ"...)'에 각각 대응하는 별도의 어휘요소를 포함하는 문장)을 만들 수 있다면, L을 분리

하지 않고 하나의 어휘요소로 처리하며 L의 정의 속에 (...σ' 또는 σ"...)'  
과 같이 이접disjunction을 사용한다.

‘동생’은 “한 부모의 자식으로서 자기보다 나이가 적은 남자나 여자”로 정의할 수 있다.<sup>4)</sup> 그런데 다른 사전에서는 이것을 두 개의 어휘소로 분할하여 하나는 “…… 자기보다 나이가 적은 남자”로, 다른 하나는 “…… 자기보다 나이가 적은 여자”로 정의할 수도 있을 것이다. 이 두 방법 중에 어느 쪽을 선택할지를 알려 주는 것이 바로 양립가능한 공기의 원칙이다. ‘철수와 영희는 모두 자기 동생과 사이좋게 지낸다’라는 문장은 철수의 동생이 남동생이고 영희의 동생이 여동생인 경우에도 아무런 문제 없이 자연스럽게 성립된다. 즉 ‘동생’의 두 의미 성분이 하나의 문장 속에서 각각 활성화될 수 있는 것이다. 따라서 ‘동생’을 두 개의 어휘소로 분리하지 않고 하나의 어휘소로 처리한다. 양립가능한 공기의 원칙은 이 원칙을 처음 제안한 학자들의 이름을 따서 Green-Appresjan의 기준이라고도 불린다.

기준 4. 차별적 공기의 기준: 어휘요소 후보 L(...σ'...σ"...)'에 대해 서로 완전히 구별되는 (형태적, 통사적, 어휘적) 공기 요소들의 집합 둘을 뽑아낼 수 있고 이 두 집합이 각각 (σ')과 (σ'')에 대응할 경우, L을 두 개의 어휘소로 분리해야 한다.

‘알다’의 여러 어휘소들 중에 “어떤 사실을 인식하다”(L<sub>1</sub>), “사실이 어떠어떠하다고 생각하다”(L<sub>2</sub>), “어떤 기술이나 능력을 지니고 있다”(L<sub>3</sub>) 등이 있다. 그런데 ‘-니 줄로’ 형태의 보문은 L<sub>2</sub>의 경우에만 가능하고, L<sub>3</sub>는 ‘-리 줄’ 형태의 보문만 취할 수 있다. 따라서 이 셋은 별도의 어휘소로 분리되어야 한다.

기준 5. 차별적 파생의 기준: 어휘요소 후보 L(...σ'...σ"...)'에 대해 서로 완전히 구별되는 파생 요소들의 집합 둘을 뽑아낼 수 있고 이 두 집합이 각각 (σ')

---

4) 연세한국어사전 동생 ①.

과 (σ<sup>m</sup>)에 대응할 경우, L을 두 개의 어휘요소로 분리해야 한다.

‘식다’는 “더운 기가 없어지다”(‘밥이 식다’ 같은 예)와 “감정이나 기운의 강도가 약해지다”(‘벤처 열기가 식기 전에’ 같은 예)의 두 가지 의미를 가질 수 있는데, 사동사 ‘식히다’는 “더운 기를 없애다”의 의미만 갖는다. 따라서 차별적 파생의 기준에 따라 ‘식다’를 두 개의 어휘소로 분리해야 한다.

### 3.2 결합설명 사전의 미시구조 구획

결합설명 사전은 각 표제항에 대해 미시구조를 다음과 같이 구획을 나누어 정보를 제시하고 있다.

1. 음운 구획: 발음, 운율적 특징.
2. 의미 구획: 정의, 함축적 의미.
3. 통사적 결합 구획: 지배도식, 논항들의 共起 제약.
4. 어휘적 결합 구획: 표제항의 의미나 통사적 속성에 의해 결정되지 않는 특이한 결합상의 특징.
5. 용례 구획
6. 관용표현 구획

다음에서는 각 표제항lexie vedette에 대해 DEC에서 구분하고 있는 미시구조의 각각의 구획에 대해 살펴본다.

#### 3.2.1 음운 구획

표제항의 記標signifiant의 음운 형식을 명시하는 부분이다. 철자법 관련 정보도 이 곳에 제시할 수 있다. 그러나 기존의 DEC에서 실제로는 아무런 정보도 제시하지 않았다. 운율적 정보도 기존 사전들에서는 대개 무시되었지만 매우 중요하다는 것을 Mel'čuk이 지적하고 있다.

### 3.2.2 의미 구획

DEC의 의미 구획은 사전적 정의 *définition lexicographique* 부분과 내포적 의미 *connotation* 부분으로 나뉜다. DEC가 엄격한 언어학적 원칙에 입각하여 편찬된 사전이라는 사실을 가장 극명하게 보여 주는 부분이 바로 의미 구획, 그 중에서도 특히 정의 부분이다. DEC의 이름에서 ‘*explicatif*’라는 말은 정의를 일정한 원칙에 입각하여 명시적으로 완벽하게 제시한다는 의미를 갖는다. 그만큼 정의는 DEC에서 중요시되는 부분이다. DEC에서 사전 정의 작성의 원칙으로 제시하고 있는 것들을 하나하나 살펴보기로 하겠다.

원칙 1. 명제 형식의 원칙: 술어적 의미를 갖는 어휘요소는 피정의항 자체가 의미 논항<sup>5)</sup>에 해당하는 변항들을 포함한 명제 형식으로 되어 있어야 한다.

이것은 피정의항의 형식에 대한 규정이다. 동사나 형용사같이 논항을 필요로 하는 단어는 피정의항이 논항들을 포함하는 온전한 명제의 형식을 갖추고 있어야만, 정의를 제대로 줄 수 있다.<sup>6)</sup> 그래서 ‘약속하다’라는 동사의 정의를 제시할 때 피정의항을 ‘X가 Z에게 Y를 약속하다’와 같은 명제 형식으로 제시하는 것이다. 이렇게 해야만 정의항에서 X, Y, Z 등의 논항이 갖는 의미론적 제약 등을 제대로 진술할 수 있다. 생성문법에서 말하는 선택 제약 정보가 이런 식으로 정의에서 자연스럽게 포착될 수 있다.

원칙 2. 해체의 원칙: 정의항에 등장하는 어휘요소들은 피정의항이 되는 어휘요소 L보다 의미가 더 단순해야 한다.

기존 사전들에서 흔히 찾아볼 수 있는 문제점 중 하나는, 피정의항보다 어려운

5) 의미 논항과 통사 논항에 대해서는 다음 절에서 설명한다.

6) *Collins Cobuild English Language Dictionary*에서도 정의를 제시할 때 이러한 점을 고려하고 있다. 예를 들어 ‘hear’라는 단어를 정의할 때 ‘when you hear sounds’와 같은 식으로 시작하는 것이다.

단어가 정의항에 사용되는 일이 많다는 것이다. 이것은 사전에서 정의(뜻풀이)를 제시하는 근본 취지에 어긋나는 일이다. 그리고 이와 밀접하게 연관된 문제로 순환적 정의의 문제를 지적할 수 있다.  $A=B+C$ 로,  $B=D+E$ 로,  $D=A+F$ 와 같은 식으로 정의를 하다 보니, 결국 A를 정의하는 데 A를 이용하는 셈이 된다. 이러한 문제들을 피할 수 있는 가장 좋은 방법은, 정의항에는 피정의항보다 의미가 더 단순한 단어들만 사용하는 것이다. ‘해체의 원칙’이 말하고자 하는 바가 바로 이것이다. 이런 식으로 의미가 복잡한 단어들을 좀더 단순한 단어들로 정의(해체)해 나가다 보면, 중국에는 해당 언어의 다른 단어들로 정의할 수 없는 가장 단순한 의미를 지닌 단어들에 도달할 것이다. 이것이 바로 의미 원소 *primitif sémantique* 이다.

원칙 3. 최대 블록 원칙: 정의를 할 때 의미 해체는 가능한 한 얇은 수준에서 이루어져야 한다. 즉 어휘요소 L을 정의하기 위해 사용되는 n개의 어휘요소  $L_1, L_2, \dots, L_n$  중에서 어떤 둘 이상의 어휘요소들의 결합이 하나의 어휘요소로 대체될 수 있어서는 안 된다.

Wierzbicka 같은 학자들은 정의항에 의미 원소들만 사용하자는 입장을 취한다. 그렇게 함으로써 정의 작성에 있어서 작성자의 주관성을 배제할 수 있고, 앞에서 말한 순환적 정의의 문제, 피정의항보다 더 어려운 단어를 정의항에 사용하는 문제 등을 피할 수 있다는 것이다. 그러나 MTT에서는 의미 원소의 수준까지 단번에 깊은 의미 해체를 하기보다는, 가능한 한 얇은 수준에서 의미 해체를 하는 것이 바람직하다고 본다. 항상 의미 원소들만 사용해서 정의를 하려 하면 정의가 너무 길어져서 피정의항의 의미를 파악하기가 오히려 어렵게 되며, 또한 서로 의미상 관련이 있는 단어들 사이의 관계가 오히려 흐려지는 문제가 있기 때문이다. Simon Dik의 *Functional Grammar*에서도 *Stepwise Semantic Decomposition*이라는 원칙을 제시하고 있는데, MTT의 최대 블록 원칙과 일맥상통한다.

원칙 4. 표준화의 원칙: 정의항에서 중의적인 용어나 동의적인 용어들을 사용하



는 것은 피해야 한다. 즉, 정의항에 사용되는 용어는 항상 하나의 의미만 지녀야 하고 하나의 의미는 항상 하나의 용어로 표현되어야 한다.

기존 사전의 문제점 중 하나는 정의에서 사용되는 단어가 다의적인 경우가 많다는 것이다. 해당 언어를 잘 알고 있는 모국어 화자는 문맥 등을 고려하여 뜻을 이를 이해할 수 있겠지만, 외국인 학습자를 위한 사전이나 컴퓨터를 위한 전자사전에서는 정의에 사용되는 단어의 의미가 정확하게 구분되어 있어야 한다. 이를 위해서는 정의항에 다의어가 사용될 때 이 *vocable*의 어떤 어휘요소가 사용되고 있는지를 구분해서 표시해 주어야 한다. 동형어는 위첨자 숫자로, 의미 차이가 크고 불규칙적인 어휘요소들은 로마숫자로, 의미 차이가 작고 규칙적일 때는 아라비아숫자로, 의미 차이가 매우 작고 의미·통사적으로 규칙적인 관계일 때는 로마자 소문자로 구분한다. ‘다리’를 예로 들어 이 규약을 보이면 다음과 같다.

다리<sup>1</sup> 橋

- I. 물이나 골짜기를 건너갈 수 있도록 설치해 놓은 시설. ‘강에 다리를 놓다’
- II.1. 물건이나 정보를 전달할 때 거치게 되는 중간 단계. ‘몇 다리 건너서 소식을 전해 듣다’
- 2. 사람이나 단체 사이의 관계를 매개해 주는 수단. ‘김 사장님과과의 사이에 다리를 놓아 주세요’

다리<sup>2</sup> 脚

- I.1a. 사람의 몸의 아래 부분. ‘철수가 다리를 다쳤다’
- 1b. 동물의 몸의 아래 부분. ‘우리집 개가 다리를 짚뚝거린다’
- 2. 오징어, 문어, 히드라 등의 촉수. ‘문어는 다리가 8개이다’
- II.1. 옷에서 다리 I.1a에 해당하는 부분. ‘이 바지는 한 쪽 다리가 짧다’
- 2. 사물의 떠받치는 부분. ‘이 책상은 다리가 하나 부러졌다’
- 3. 안경의 테에 붙어서 귀에 걸게 되어 있는 부분. ‘다리가 부러진 안경’

‘다리’가 다의어이기 때문에, 위의 다리<sup>2</sup> II.1.의 정의항에서 그냥 ‘다리’라고 하지

않고 '다리 I.1a'와 같이 구분 번호를 붙여 주어야 하는 것이다. DEC에서 이미 기술되어 있는 단어의 경우에는 DEC의 해당 구분 번호를 붙여 주면 되겠지만, DEC에서 아직 기술되지 않은 단어의 경우에는 임시로 다른 사전을 기준으로 삼아 구분 번호를 붙여 줄 수밖에 없다.

원칙 5. 대치 가능성의 원칙: 피정의항과 정의항은 모든 발화에서 의미 차이 없이 서로 대치될 수 있어야 한다.

정의를 내린다는 것은 피정의항이 사용될 수 있는 필요충분조건을 제시하는 것이라고도 할 수 있다. 정의항이 피정의항에 대한 필요충분조건을 나타낸다면 이 둘은 서로 대치되어 사용될 수 있어야 할 것이다. 물론, 정의항이 피정의항보다 길고 장황한 것이 일반적이기 때문에, 피정의항을 정의항으로 대치하면 자연스럽게 못한 표현이 될 수도 있다. 그렇다 하더라도 인지적 의미의 측면에서 둘 사이에 차이가 있어서는 안 될 것이다.

사전 정의에서는 해당 단어와 관련된 백과사전적 지식은 제외하고 언어적 정보만 제시한다. 언어적 정보란 그 단어에 대해 화자들이 가지고 있는 지식을 말하고, 백과사전적 정보란 그 단어가 가리키는 대상에 대해 화자들이 가지고 있는 지식을 말한다. 어떤 정보가 언어적 정보인지 백과사전적 정보인지 판정하는 것은 쉬운 일이 아니다. MTT에서는 다음의 원리를 제안하여 이 문제에 대한 해답을 제시하고자 한다.

기준 6. 언어적 관여성의 기준: 어휘요소 L의 정의에 포함시킬 만한 의미 성분의 후보( $\sigma$ )가 있는데, 이 의미 성분이 L의 지시 대상의 내재적 속성을 나타내기는 하나 이 속성이 변별적인 속성은 아니라고 할 때, 해당 언어에 L과 형식적으로 연결되어 있고 그 정의에 ( $\sigma$ )가 포함되어 있는 어휘요소  $L_1$ 이 있을 때 그리고 그럴 때에만 ( $\sigma$ )를 L의 정의에 포함시킨다.

L과 L<sub>1</sub> 사이의 관계는 다음의 세 가지가 가능하다.

- (a) L<sub>1</sub>이 L과 같은 vocable에 속하는 어휘요소인 경우
- (b) L<sub>1</sub>이 L에서 파생된 단어 또는 L을 포함한 합성어인 경우
- (c) L<sub>1</sub>이 L을 포함하는 숙어소인 경우

이 세 경우의 예를 들어 보면 다음과 같다.

- (a) 신체부위를 나타내는 ‘머리’라는 어휘소의 정의에 “지능의 중추 기관”이라는 의미 성분을 포함시켜야 할지, 아니면 그런 것은 백과사전적 정보로 간주하여 정의에서 제외해야 할지가 우리의 관심사이다. 그런데 ‘머리가 좋다/나쁘다’, ‘머리를 쓰다’와 같은 표현을 보면 ‘머리’라는 vocable 내에 “지능의 중추 기관”이라는 의미를 갖는 어휘소도 포함되어야 함을 알 수 있다. 그렇다면 위의 (a)에 의해 “지능의 중추 기관”이라는 의미 성분이 신체부위를 나타내는 어휘소 ‘머리’의 정의에 포함되어야 한다고 결론짓게 된다. 사람의 신체부위를 가리키는 어휘소 ‘다리’의 경우에도 탁자 등의 사물을 떠받치는 부분을 가리키는 어휘소가 존재하므로 “支持”라는 의미 성분이 신체부위 어휘소 ‘다리’의 정의에도 포함되어야 한다.
- (b) ‘일손’ 같은 합성어가 존재하므로 신체부위를 나타내는 어휘소 ‘손’의 정의에 “노동의 주요 수단”이라는 의미 성분이 포함되어야 한다.
- (c) “빠르게”의 의미를 갖는 ‘총알처럼’이라는 숙어소가 존재하므로 ‘총알’의 정의에 “빠르다”라는 의미 성분이 포함되어야 한다.

DEC는 어휘요소의 共起 관계cooccurrence relation에 관한 정보에 세심한 주의를 기울이기로 유명한 사전이다. 어휘요소의 공기 관계에는 의미적 공기 관계와 어휘적 공기 관계의 두 가지가 있다. 전자는 해당 어휘요소의 의미 정의로부터 예측할 수 있는 공기 관계이고, 후자는 의미 정의로부터 예측할 수 없는, 어휘적으로 제약되어 있는 공기 관계이다. 예를 들어 동사 ‘기다리다’가 ‘애타게’ 같은 부사의 수식을 받을 수 있는 것은 ‘기다리다’의 의미 정의로부터 예측할 수 있는 사실이므로 의미적 공기 관계의 예이다. 그러나 ‘애타게’와 비슷한 의미를 갖는