

# 통합의존문법을 이용한 독일어 어휘목록 구성에 관한 연구

## - 자연어처리 시스템에 적용할 목적으로\*

김용경 · 김현수 (부산대)

### 1. 서론

먼저 생성문법 틀 내의 모든 연구에 대한 목표를 다음과 같이 설정할 수 있다.

1. 이상적인 생성문법은 인간언어에 있어서의 모든 문법적으로 가능한 변화를 생성시키는 규칙과 비문법적 발화를 생성시키지 않는 규칙을 구체적으로 설명할 수 있는 것이어야 한다.
2. 한 언어의 화자는 다른 일련의 발화를 생성시키거나 이해한다는 점에 이르기까지 나타나는 이 차이는 다른 집단의 화자를 위해서는 그 집단의 문법에 의하여 재고되어야 한다.
3. 완전한 문법성과 비문법성 사이의 중간영역에 발화가 놓여져 있다는 사실에서 생성문법은 그와 같은 발화를 생성할 수 있을 뿐만 아니라 그 발화에 그 언어에서의 최소한의 위상이 정확하게 할당될 수 있다.

본 연구에서 전개되는 생성문법 형식은 주로 1과 관련되어 있다. 이에 대해 본 연구가 살펴보고 넘어 가야할 사항들은 다음과 같다. 첫째, 이것은 생성문법 전통 이외의 것을 포함하고 있으면서도 몇몇 다른 초기연구에서 발견된다. 이것과 관련된 연구방향을 실현하는 출발점은 Richard Montague<sup>1)</sup>의 연구이다. 둘째, 통합(unification)개념이다. 이 개념은 1980년대 생성문법에서 처음으로 광범위하게 사용되었던, 문법단위의 상세한 기술을 만들어내기 위하여 다른 원천에서 나온 문법규칙을 조합하는 유연한 방법을 제공하고 있다. 통합은 Bresnan 1982, Gazdar, Klein, Pullard & Sag 1985<sup>2)</sup>(이하 GKPS1985)에서 이 개념을 사

\* 이 논문은 1998년도 한국학술진흥재단 자유공모과제의 연구비에 의하여 연구되었음.

1) Vgl. Montague, Richard(1977).

2) Gazdar, G., Klein, E., Pullum, G., Sag, I.(1985) : S.39.

용하는 몇몇 연구를 구분하는데 있어 중요하다. 셋째, 심층구조와 표층구조간의 구분보다는 단지 단순한 차원의 통사적 구조만을 생성하는 규칙을 사용하면서도 위에 언급된 목표를 달성하는 것이 가능할 것이라는 견해는 Gazdar 1981과 거기에 인용된 초기 연구들에서 그 기원을 찾을 수 있다. 이 견해를 실현하기 위한 필수적 전제조건은, 특정한 유형의 자질(자질) 통합의 발전에 따라서, 직접지배(immediate Dominance)/선형적 선행틀구조(Linear Precedence Framework) 그리고 수형도를 통해 정보를 제시하기 위한 방법으로서의 말단자질원칙(Foot Feature Principle) 등이었다. 넷째, 구구조규칙이 어휘목록과 투사에서 가장 잘 보여질 수 있다는 관점은 Stowell 1981에서 발전되었다. 그리고 어휘규칙이 다양한 종류의 변형을 대체할 수 있다는 생각은 일반적으로 Bresnan 1982에서 전개되었다. 복잡한 어휘목록을 만드는 일반적인 방법으로서 계층적으로 자질을 조합한다는 생각은 Flickinger 1987에서 유래한다. Pollard & Sag 1994는(이하 P&S94) 이러한 모든 생각들을 단순틀구조(single framework) 속에 총괄하여 광범위하고 다양한 영어구조를 중심으로 자세한 분석을 전개하고 있다.

본 연구의 중심적 생각은 Pollard & Sag 1987(이하 P&S87)에서 제공된 것이다. 이 연구에서 제시된 틀구조는 일반적으로 핵주도 구구조문법(Head-Driven Phrase Structure Grammar, 이하 HPSG)으로 알려져 있다.

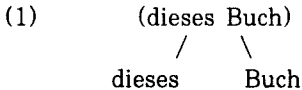
위에 언급된 모든 생성문법들은 생성문법에서의 다른 대부분의 연구들과 마찬가지로 구성(constituency)으로 알려져 있는 표현유형을 이용한다. 그러나 본 연구에서는 이것을 의존(dependency)으로 대체한다. 뒤에서 자세하게 논의되는 것처럼, 문법에 대한 구성으로서 의존개념의 전개는 생성문법 내에서는 비교적 관심이 적었다<sup>3)</sup>. 이러한 이유로 다음 장에서는 구성과 의존간의 관계에 대하여 비교적 상세하게 살펴본다.

3) 이것에 대한 주목할 만한 예외로는 Hudson, Richard(1984), Starosta, Stanley (1988) 참조.

## II. 이론적 전제

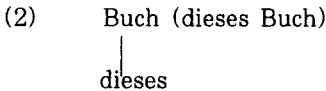
### 2.1. 구성과 의존 개념의 비교

구성은 일반적으로는 낱말들이 보다 더 큰 단위형식에 결합하는 것으로 정의될 수 있다. 아래 (1)의 수형도는 *dieses*와 *Buch*가 보다 더 큰 단위는 *dieses Buch*에 결합하고 있는 방식을 보여주고 있다. 이러한 결합의 결과는 독립된 교점 즉, 구교점에 제공된다.



이와 동일한 절차가 문장의 다른 부분에서 다른 낱말들의 결합에도 적용된다. 따라서 결합은 그 자체가 문장의 최상위 교점(S-교점)에 이를 때까지 각각 다른 낱말들을 더 높은 차원인 구교점으로 결합시킨다

생성문법이 시작되었던(구성<sup>4</sup>)개념에 완전히 의존했던 시기에는, 의존개념도 소수의 언어학자들에 의해서 발전되었다<sup>5</sup>). 아래 (2)의 수형도는 *dieses*가 *Buch*에 의존한다는 주장의 표현이다. 의존문법의 전통 내에는, 대개 조어(祖語, parent word)가 지배하거나(govern) 그것에 의존하는 낱말의 지배소(governor)라고 일컬어진다. 괄호 속의 구는 이 낱말들이 결합한 결과를 보여준다.



구성수형도와는 달리 의존수형도는 이러한 결합 결과를 보여줄 수 있는 독립적 교점을 만들지는 않는다. 그 대신에 의존문법에서의 작업은 이 낱말들이 어떻게 생

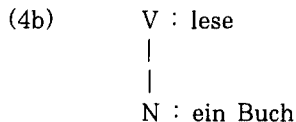
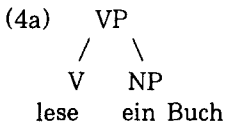
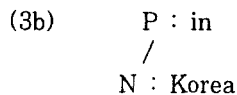
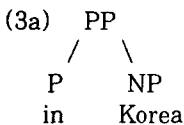
4) 구성개념의 발전은 실제로 생성문법보다 앞선다. Sgall은 구성은 미국의 구조주의 초기에 발전된 개념으로서 적용가능성에 대한 검토도 거치지 않은 채 Chomsky에 의해서 생성문법의 초기 연구에 편입되었다고 주장한다. Vgl. Sgall, Petr(1988).

5) Vgl. Tesnière, Lucien(1953).

성되는가와 이 낱말의 의미가 다른 보다 더 큰 단위의 의미를 생성시키기 위하여 어떻게 결합하는가가 아니라 문장 속에 있는 낱말들 사이의 관계를 표현하기 위하여 이 수형도를 어떻게 보아야하는가에 대한 문제에 초점을 맞추었다6).

구성소구조를 이용하는 연구는 분명히 언어학 연구에서 지배적인 것이긴 하나, X-Bar이론의 발전으로 생성문법을 토대로 한 연구는 의존의 전통에 의해 강한 영향을 받았다. Jackendoff의 연구7)덕분에 구성수형도도 의존을 표현하고 있다는 것은 명백하다. 즉 (1)에서 어휘적 명사인 Buch는 전체 NP의 핵이며, 의존소로서의 한정사 dieses를 가진다. 더 일반적으로 말하면, 구성수형도에 있는 어휘적 핵이 의존소로서의 자매어를 가지고 있다. 구성수형도와 의존수형도간의 차이는 후자에서 낱말이 조어 혹은 핵으로 특별히 선택된 것과 같은 범주를 가진 자매교점보다도 오히려 그 조어의 의존소라는 사실이다8).

위에서 언급된 것을 증명하기 위하여, 구성수형도와 의존수형도 모두를 표현하고 있는 아래 (3-4)를 살펴보자. 여기서는 그 교점에 대한 구성관계 해석은 생략하지만 통사적 범주는 덧붙인다. (3a)와 (4a)에 있는 수형도는 한 생성문법적 토대에서 나온 것이다. 의존수형도에 있는 낱말은 전체 수형도의 제일 아래 부분보다는 오히려 교점 그 자체에 보통 기술되는데, 왜냐하면 4.3장에서 고찰되는 바와 같이 범주와 의존관계는 이러한 낱말들의 선형적 순서(linear order) 그 자체를 독특하게 결정하지는 않기 때문이다.



6) Vgl. Schubert, Klaus(1987).

7) Vgl. Jackendoff, Ray(1977).

8) 이러한 표시법은 Hudson1984에서 나타나는데, 선형화를 보다 명확하게 하는 장점이 있는 반면, constituency와 dependency간의 관계를 불분명하게 하는 단점도 있다. Vgl. Hudson, Richard(1984).

이 수형도는 여러 경우에 있어서 의존수형도가 상응하는 구성수형도에 정확히 평행적이라는 것을 명확하게 보여주고 있다. 구성수형도에서는 독립된 교점이 어휘적 핵과 이것의 보충어와의 이러한 핵의 결합을 위해 도입되는 반면에, 의존수형도에서는 이 두 종류의 정보가 하나의 교점 내에 있다는 것이 바로 그 차이점이다. 달리 말하면 어휘적 핵은 구성관계의 교점에서는 부모교점으로의 구의 확장을 가지는 반면에 의존관계 수형도에서는 그렇지 않다는 것이다.

## 2.2. 선형적 순서

이제 여기서는 문장내 낱말의 선형적 순서에 대한 문제를 살펴본다. 전형적인 구성수형도는 문장의 낱말을 실제 발화에서 사용된 선형적 순서로 수형도의 아래 부분에 둔다. 그러나 이러한 선택은 의존문법에는 이용할 수 없는데, 왜냐하면 의존관계는 그 자체가 특정한 선형적 순서를 투사하지 않기 때문이다.

이 점을 증명하는 것으로 다음 (5)의 의존관계에 관한 진술을 살펴보자.

(5)  $N - - \rangle \text{Det}, A$

규칙 오른쪽에 있는 요소들의 순서가 수형도에서 요소들의 순서를 열거하고, 그에 따라 이러한 요소들의 순서가 구조의 단계에서 생성되는 구성틀구조에 초기의 모든 연구에서 이용된 규칙을 적용시키는 것은 생성의존문법학자들에게는 가능할 것이다. 그 다음에 규칙 오른쪽에 있는 요소들의 반대 순서를 상세히 설명하는 부가적인 의존관계 진술 즉, 아래 (6)를 기술하는 것도 가능할 것이다.

(6)  $N \rightarrow A, \text{Det}.$

이 예에서 규칙 오른쪽에는 단지 2개의 요소만이 있다. 그러나 얼마나 많은 요소가 있는가는 중요하지 않으며, 원칙적으로 존재하는 모든 요소의 순서를 상세하게 설명하는 일련의 규칙을 기술하는 것은 가능할 것이다.

통사적으로 표현된 것에서 선형적 순서로부터 계층적 관계를 구별하기 위한 또 다른 방법이 있다. 많은 언어에서 대부분 또는 모든 유형의 문장 낱말들은 거의 의미변화가 없거나 아예 없는 다양한 순서를 가질 수 있다. 달리 말하면, 낱말간의

문법적 관계는 그들의 선형적 순서의 변화에서 불구하고 불변으로 남아있다는 것이다. 언어와 관련된 이론이 해야할 임무 중 하나가 이러한 사실을 찾아내는 것일 것이다. 생성문법의 초기 연구에는 이 문제에 대한 단 한가지의 접근방식만을 제공했다. 즉, 여러 것 가운데 하나의 순서와 관련되는 구조를 생성시키기 위하여 구구조 규칙을 이용하며, 처음의 "심층구조" 순서로부터 다른 "표층구조" 순서를 파생시키기 위하여 "변형"을 이용하는 것이었다.

GPSG(GKPS1985)는 가능한 선택을 제시한다<sup>9)</sup>. 즉, 직접지배규칙에 의해 생성된 구조들의 연속이 결정되어 있지 않고, 구성소구조내에 있는 자매교점 중에서 선형적 순서를 결정하는 선형적 우선권(linear precedence, LP) 진술이라 명명되는 다른 일련의 규칙들을 그 어떤 순서도 가능하다는 것을 설명하기 위하여 이용한다는 것을 가정하는 것이다. 직접지배규칙과 진술의 결합은 ID/LP 형식이라고 명명된다. 더 일반적으로 이 체계가 계층적 관계와 선형적 관계 사이를 구별한다고 말할 수 있다.

구성틀구조 내에는 이 두 선택 모두가 아직도 이용되고 있다. 의존틀구조 내에는 단지 두 번째 선택 형식 즉, ID/LP 형식만이 가능한데, 왜냐하면 위에서 언급한 바와 같이 구구조의 의존 등가물도 의존수형도도 그 자체로는 독특한 선형적 순서를 보장하지 못하기 때문이다. ID/LP 형식(또는 몇몇 등가물)과 의존틀구조 사이의 상관관계는 ID/LP형식이 없는 하나의 구성성분으로서의 틀구조가 의존개념을 직접 즉, 어휘적 교점과 구의 교점 사이에 있는 구분을 없애고 직접 통합하는 것은 어렵다는 것을 암시하고 있다.

### 2.3. 의존기반의 초기 연구

2.1.장에서 언급된 바와 같이, 구성관계는 생성문법에서 통사적 표현의 지배적 형태였다.

언어학에의 컴퓨터 적용에 있어서 의존문법은 구문분석(parsing) 알고리즘과 기계번역을 위한 토대로 광범위하게 이용되어 왔다. 그 이유는 아주 명백하다. 구문분석기(parser)의 가장 전형적인 과제는 이미 존재하는 문장에 대한 통사적 표현을 만들어내는 것이다. 낱말 순서에 대한 정보는 이미 입력라인에 주어졌 있기

9) Gazdar, G., Klein, E., Pullum, G., Sag, I.(1985) : a.a.O., S.52f

때문에 이 정보를 다른 형식에서 이용할 수 있게 하는 수행도는 그리 중요하지 않다. 컴퓨터의 메모리를 절약하고 그럼으로써 표현을 만들어내는 알고리즘의 속도를 향상시킨다는 결과를 놓고 볼 때 구성수행도에 비해 의존수행도가 교점 수를 줄인다는 측면에서 더 중요하다. 이것은 수행도에 처음 어떻게 구현되는가가 보다 훨씬 더 중요한데, 왜냐하면 문장을 분석하는데 있어서 중요한 문제는 통사적 애매성을 어떻게 해결하느냐 하는 것이기 때문이다. 이 통사적 애매성 제거를 위한 한 접근 방식에서 분석의 첫 단계는 이 분석에서 선택하기 위한 담화요인(discourse factor)과 세상지식으로 방향을 돌리기 전에 먼저 그 문장의 가능한 모든 분석을 위한 표현을 만들어내는 것이다. 이러한 접근방식에서 보다 복잡한 구성소구조에 의해 요구되는 구문분석 알고리즘은 생성되는 많은 수행도 수에 의해 그 복잡성이 더 증가된다.

목표는 비슷하지만 생성문법과는 관련이 없는 의존문법에 대한 연구는 의미텍스트모델(Meaning-Text Model, MTM)이다. 이것은 Mel'uk에 의해 거의 완전하게 기술되고 적용된 것이다. Mel'uk에 따르면<sup>10)</sup>, "MTM은 전혀 생성문법을 이용한 시스템이 아니다. 이것은 완전히 등가적인 혹은 번역적인(translative) 장치다. 이것의 규칙은 모든 것을 상술하지도 않으며, 오직 문법적으로 정확하거나 의미있는 텍스트만을 생성시키지도 않는다. 즉, 이 규칙들은 단순하게 상용하는 텍스트에 의해 표현될 수 있는 모든 SemR(semantic relation, 의미관계)와 어떤 주어진 PhonR (phonological relation)을 연결시킨다."

비록 MTM이 문장을 생성시키는 목표를 가지지 않을지라도, 만약 일반적인 기술적 감각으로 "생성시킨다(generate)"는 말을 사용한다면, 위에 Mel'uk의 언급에서 MTM이 한 언어에 대해 인간의 세상지식을 모의실험하는 생성문법학자들의 목표를 공유하고 있다고 볼 수 있다. Mel'uk과 또 다른 언어학자들은 일곱 가지 서술단계를 구성하고 있는 정교한 시스템(ellaborate system)을 구축했는데, 의미를 위한 한 단계, 통사론을 위한 두 단계, 형태론을 위한 두 단계 그리고 음성학을 위한 두 단계가 그것이다. Mel'uk은 언어현상에 대한 많은 흥미있는 분석을 제공하고 있다. 그의 접근방식은 의미가 가장 근본적인 언어학적 단계로 받아들여지는 생성의미론에서의 연구<sup>11)</sup>와 유사한 것처럼 보인다. 비록 여기서 전개된 틀구조가 HPSG 보다 MTM에서 유래한 형식적 기제를 덜 언급한다 하여도 기본적인

10) Mel'uk, Igor A.(1987) : S. 45.

11) Vgl. Lakoff, George(1971).

방식에서는 전자 즉, MTM과 더 관련이 있다. 즉, 낱말은 그 일차적인 결합가로서 통사적 범주보다는 오히려 의미적 단위를 취한다고 생각을 받아들인다. 이에 대해서는 3.3장에서 보다 자세하게 언급한다.

#### 2.4. 두 연구의 통합

이 장은 본 연구에서 전개되는 틀구조의 명칭으로 어떤 것을 사용할 것인가에 대한 문제를 간략하게 다룬다. 이것은 HPSG와 많은 부분이 관련되어 있기 때문에 “의존기반 HPSG (Dependency-based HPSG)”라고 명명될 수 있다. 이 명칭은 의존개념이 문법의 중심이 된 것으로 더 이상 전통적인 의미에서의 구구조문법(Phrase Structure Grammar, PSG)가 아니라는 것을 의미한다. 그리고 그 다음에 선택할 수 있는 것은 핵주도 의존문법(Head-Driven Dependency Grammar, HDG)이다. 그러나 이미 언급한 바와 같이, Hudson과 Starosa의 연구가 이들이 비록 이 용어를 사용하지 않지만 이 역시 “핵주도적(head-driven)” 특성을 가지고 있다. 그러면 Hudson과 Starosta의 연구와 본 연구간에 어떤 차이가 있는가? 이를 간략하게 기술하면, Hudson과 Starosta에 의해 전개된 생성문법의 한 연구와 이와는 일반적인 차이는 자질의 본질 그리고 자질간의 상호관계에 있다. 이 둘은 주어진 어휘목록 내에 서로간의 구조적 관계를 가지지 않는 이원적 자질을 이용하고 있다. 그리고 의미적 역할과 때로는 화용적 정보가 어휘목록에 통합된다는 것도 그 차이 중 하나다. 이러한 차이 가운데 첫 번째와 두 번째는 여기서 제안된 생성적 틀구조를 위해 통합의존문법(Unification Dependency Grammar, UDG)이라는 명칭을 사용하도록 유도한다. 이미 앞에서 언급된 바와 같이, 통합개념은 HPSG에서 뿐만 아니라 어휘기능문법(Lexical Functional Grammar, LFG)와 GPSG에서도 중심적인 개념이다. 덧붙여서, 자연어처리에 대한 컴퓨터 활용에 이 개념을 사용하게 하며, 따라서 생성문법 내에서뿐만 아니라 전산언어학에서도 이러한 경향을 따르고 있다. UDG는 이러한 방식을 이룬 통사론 내에서 많은 주의를 끌지 못했던 의존문법의 오랜 전통과도 통합한다.



### III. 어휘목록

#### 3.1. 자질표

어떤 한 종류의 모델에 대한 정보표현방법으로 구구조 표지(marker) 또는 수형도를 이용하는 것은 생성문법에서의 많은 사용으로 비교적 낯설지 않다. P&S94에 의해 사용된 것처럼<sup>12)</sup>, 수형도는 단순히 속성값 자질표(Attribute-value Matrices, AVM)을 표현하는 빠른 방법이다. 이 때 이러한 자질표의 많은 항목들은 그것과 관계가 없다. 그러나, 비록 모든 항목들이 포함되었을지라도, AVM은 그 수형도와 등가적인 것은 아니다. 왜냐하면 후자 즉, 수형도는 구조할당을 이용하고 있기 때문이다.

AVM의 수형도는 (i) 속성 표식과 (ii) 종류기호로 이루어져 있다. 본 연구에 이용된 자질구조는 오히려 P&S의 AVM에 사용하였던 두 가지보다는 한 종류의 기본적인 언어적 대상을 가지고 있다. 이것은 자질이라 명명되며, 일반적으로 AVM's에 있는 종류기호(sort symbol)와 상응하는 것이다.

대문자로 쓰여진 속성 표식들은 CASE(격), NUMBER(수) 그리고 CATEGORY(범주)와 같은 다양한 언어적 자질을 지정하는데 사용된다. 소문자로 쓰여진 종류기호들은 주어진 속성과 관련된 특정한 값을 나타내는데 사용된다. 예를 들면 acc(대격)는 언제나 격 속성과 관련되어 있다. 이것은 CASE acc로 쓰여진다. 종류기호의 다른 곳에서 사용하는 것은 그에 관련되는 하부구조유형을 가리키는 것이다.

##### (1) CATnoun|CASEacc

이것은 속성값 구조 CASEacc가 속성값 구조 CATnoun의 하부구조라는 것을 말하고 있다. 반대로 CATnoun은 CASEacc의 상부구조이다. 그와 동시에 종류기호 noun은 이에 관련된 구조 유형을 지시하고 있다.

종류기호는 보통 계층적 망에 있는 다른 종류기호에 연결되어 있다. 만약 그러한 기호 A와 B가 A가 보다 높은 곳에 위치해 있는 방식으로 연결되어 있다면, A는 B

12) Pollard, Carl & Ivan Sag(1994) : S.15-17

의 상위종류(supersort)라고 말해지며, A가 핵이 된(headed) 구조는 B가 핵이 된 구조의 상위유형이다. 반대로, 기호 B는 기호 A의 하위종류(subsort)이며, 구조 B는 구조 A의 하위유형(subtype)이다. 위의 예에서 nom과 acc 모두는 종류기호 case의 하위종류들이다. 한 상위종류는 그 자체로 어떠한 종류기호가 계층의 상위로 올라가는 상위종류를 가진다. 마찬가지로, 한 하위종류 자체가 계층의 아래로 내려가는 하나 혹은 그 이상의 하위종류를 가진다.

위의 상부/하부구조 CASE와 종류기호 case는 P&S94에서 제시된 바와 같이<sup>13)</sup> 전체적으로 기술된 AVM에 있는 몇몇 정보가 잉여적이라는 것을 분명하게 한다. 즉, 종류기호 case나 속성표식 CASE 중 하나를 제거하는 것이 바람직하다. 사실상 계층적으로 정돈된 종류기호 개념은 속성표식과 종류기호 사이의 구분을 불필요한 것으로 한다는 것은 분명하다. 속성은 단순히 연합된 종류기호의 상위종류이며, 따라서 생략될 수 있다. 예를 들면, 만약 CASE가 acc의 상위종류이면, 전자를 간단히 생략할 수 있다. 왜냐하면 어떤 종류기호는 그 상위종류를 함축하고 있기 때문이다. 따라서 UDG의 자질구조에서는 단지 하나의 기본적 대상 즉, 자질만을 이용하는 것이다. 이러한 변경은 보통 P&S와 그 밖의 HPSG 지지자들에 의해 사용되는 AVM을 기술하는 단순형태를 공식적으로 변형시키며, 더 나아가 음운적 낱말을 소문자로 자질을 대문자로 기술하는 것이 가능하다. 여기서는 더 나가서 이제부터 자질구조에 대한 논의에서 실제적인 문법분석 이후 하위/상위종류보다는 하위/상위유형이라는 용어로 기술하고, 부분을 구성하는 특정 기호보다는 전체구조에 관심을 둔다.

하위유형은 그것의 상위 유형과의 "is-a"관계에 있으며, 그 역은 아니다. 따라서 CASE와 관련된 위의 예에서 볼 때 ACC는 CASE이지만, CASE는 ACC가 아니다. 이것은 어휘목록 속에 주어진 상위유형이 하부구조로서 이러한 어휘적 표지를 사용하는 특정한 구 표지에서 그의 어떤 하위유형에 의해 실현될 수 있다는 결론을 도출한다. 예를 들면 PPRO(Personal Pronoun, 인칭대명사)는 ANOUN(argument noun, 논항명사)의 하위유형이며, 구성소들구조에서 자주 나타나는 NP의 가장 밀접한 UDG의 등가물이다. 이것은 PPRO가 구 표지에서도 ANOUN이 나타나는 어떤 슬롯(slot) 즉, 그 어떤 곳에서도 나타날 수 있다는 것을 의미하고 있다. 이것은 인칭대명사가 ANOUN의 다른 유형과 같은 분포를 가진다는 사

13) ebd.



여 있다면, 앞에 있는 자질 WORD는 불필요한 것이되고, 그렇다면 이것을 UDG에서는 생략한다. HPSG에 있는 WORD의 세 하부구조(그리고 UDG의 첫 번째 자질)은 다음과 같다. 즉,

PHON(음운) : 모든 낱말이 음성적 구성요소를 가지고 있다. 이것은 아마 낱말의 각각의 분절에 대한 음운론적 자질뿐만 아니라, 그 어떤 음운론적 규칙도 이러한 분절의 결합에 적용된다는 사실을 포함하고 있다.

CAT(범주) : 이것은 동일한 낱말의 이(異)형태에 대한 주어진 낱말과의 관계, 이 낱말이 결합할 수 있는 표지종류, 이 낱말의 해석 그리고 VAL에서 언급된 표시유형과 이 낱말의 결합에 대한 해석 등의 정보를 포함하고 있다. 이런 류의 정보는 본 연구의 일차적 주제이다.

QUANT(양화사) : 이것은 P&S94<sup>15)</sup>에 의해서 도입된 마지막 요소다. 양화사의 특별한 논리적 자질은 자신의 하부구조를 실제로 주어진 낱말에 제공하는 것이지만, 그러나 UDG에서는 단지 그의 통사적 분할에만 다루고 하부구조는 이용하지 않는다.

### 3.3. 범주 하위유형과 형태적 특성

CAT는 두 개의 하위유형 즉, HEAD와 LINK를 가지고 있다. 여기서는 HEAD 구조를 보다 상세하게 다룬다. HEAD는 네 가지 하위유형을 가지고 있다. 즉, VERB(동사), NOUN(명사), ADJECTIVE(형용사) 그리고 ADVERB(부사)이다. 이들 각각은 하위유형 CONTENT(내용)와 CONTEXT(문맥)를 가지고 있으며, ADVERB를 제외한 나머지는 하부구조 FORM(형태)을 가지고 있다. 이 하부구조 자체는 그것의 상부구조인 CAT의 하위유형에 따라 하위유형을 가진다.

그리고 NOUN의 하위유형은 NAME(명칭), COMMON(보통명사), PPRO(인칭대명사), RELPRO(관계대명사) 그리고 가끔씩 나타나는 EXPL(expletive, 허사). P&S94도 또한 "대용어(anaphor)"(재귀대명사) 자질결속에

---

15) ebd., S.47

대한 분석을 제공하고 있다<sup>16)</sup>. 이 분류는 고유명사, 보통명사 그리고 다양한 유형의 대명사들 간의 전통적인 분류에 상응하고 있다.

여기에 이미 언급된 명사유형을 위하여 CAT|FORM 구조에 대한 목록을 제시한다.

(3a) NAME

(3b) COMMON|NUM {SG:PL}

(3c)

PPRO FORM	NUM {SG:PL}
	PERS {1:2:3}
	GEN {MASC:FEM:NEUTER}
	CASE {ACC:NOM}

(3d) RELPRO|CASE {ACC:NOM:POSS}

(3e) EXPL

범주 하위유형 VERB를 위하여 FORM의 하위유형은 본질적으로 그 동사의 형태에 일치한다. 독일어에서 이것들은 다음과 같다. 즉, FIN(정형, finite), BSE(기본어, base), PRP(현재분사, present participle), PSP(완료분사, perfect participle)이다. FORM의 부가적 하부구조 즉, TMA(시제-화법-양상, tense-mode-aspect)에 따라 구별되는 두 개의 정형 형태가 있다.

### 3.4. 내용과 결합

CAT의 하위유형 NOUN 자체는 유형의 복합적 하부구조를 가진다. 이것의 직접 하위유형은 REFN(지시적 명사, referring noun)와 EXPL(허사)이다. REFN 자체는 하위유형 ANOUN과 RELPRO를 가진다. 만약 그 동사들을 위한 어휘목록 VAL(valency, 결합)이 그렇게 요구한다면, 비록 허사도 논항명사와는 다른 하위유형이며, 허사 또한 기상관계를 나타내는 동사의 주어가 되는 예에서처럼 논항위치에 나타날 수 있다. 그러나 논항명사는 그 하위유형과 하부구조에 있어서 허사와는 많은 차이를 나타내고 있다.

ANOUN은 하위유형으로 NAME(고유명사), POSS(소유격), PPRO(인칭 대

---

16) ebd., S.267-272.

명사), DETN(한정명사) 그리고 CMON(보통명사)를 가지고 있다. 이 모든 것이 가지고 있는 자질 CTNT(내용)는 유형 RESTR(제약)의 여러 하부구조와 관련되어 있다. 이것들은 이 세계의 몇몇 실재(實在) 혹은 실재의 부류에 대한 자질에 관한 것을 말해주고 있다.

이제 여기서는 결합개념에 대해 살펴본다. HPSG와는 달리 UDG에서의 결합은 단순히 독립적 하부구조로 다루어지지 않는다. 그 대신에 결합가(valents)는 CONTENT 구조내의 ROLE의 다양한 유형에 대한 하부구조로 다뤄진다. 이런 관점에서 본 연구는 의존틀구조 내에 있는 Mel'uk을 비롯한 언어학자에 의해 발전된 의미-텍스트 모델의 전통을 따르고 있다. 통합문법의 이것에 대한 보다 상세한 동기는 한 특정한 역할 혹은 제한이 여러 다른 방법을 통해 통사적으로 실현될 수 있다는 것이다.

이러한 하부구조는 어휘목록의 음운론적 낱말과 결합할 수 있는 낱말의 범주와 그 밖의 정보에 제한을 결정한다. 결합이라는 용어는 생성문법 내에 있는 대부분의 연구에서의 "하위범주화"라는 낱말과 일치한다. 결합개념은 Tesnière<sup>17)</sup>에 의해서 발전되었다.

VAL의 다양한 하부구조들의 모든 값이 관련 낱말의 의존소로 반드시 실현될 필요는 없다. 많은 경우 실현은 선택적인 것이며, 그 밖의 경우에는 어떤 이유로 실현이 억제된다. 이러한 가변성을 다루기 위하여 하위유형 '+', '-' 그리고 '+-'를 가지는 자질유형 REALIZABILITY가 이용된다. '+'는 관련된 결합가가 실현되어야 한다는 것을 의미하며, '-'는 그것이 실현되지 않을 것임을 의미한다. 이 자질은 형식적으로는 VAL의 하부구조 혹은 그 하위유형 가운데 하나이지만, 일반적으로는 관련 범주에 대한 접미사로 단축될 것이다.

## IV. 어휘목록 구성

### 4.1. HPSG에서의 문제와 해결방안

3장에서는 어휘목록이 표지의 통사적, 의미적, 화용적 측면을 위한 필수적 정보를 기호화하는 것이 매우 복잡하다는 것을 증명하고 있다. 사실 문법에 있어서의

17) Vgl., Tesnière, Lucien(1953).

모든 복잡성은 거의 어휘사전에 있다고 할 수 있다. 2장의 어휘목록 구성을 위한 논의에서 아직 해결되지 않은 다음과 같은 문제가 남아 있다. 즉, 하나 하나 상세하게 기술되어야 하는 특정한 낱말에 대한 특유의 정보의 어휘목록 묶음이 어디까지 확장될 수 있는가 그리고 이러한 복잡함을 극복하고 어휘목록을 구성하는 체계적인 방법을 도출하는 보다 더 큰 규칙은 어디까지 확장될 수 있는가에 대한 문제이다.

낱말을 전통적인 방법으로 다루는데서 나타나는 일반성은 몇몇 언어 또는 대부분의 언어 또는 모든 언어에서 발견된다고 가정한다.

(a) 낱말은 예를 들면 발화 부분에 따라 분류될 수 있다.

(b) 많은 언어에서 낱말은 어형변화로 알려져 있는 접사를 가지는데, 이것은 그것들의 정확한 음운적 형태를 위해 적어도 같은 문장에 있는 다른 낱말의 자질에 의존하고 있다.

(c) 대부분의 낱말들은 단지 다른 낱말의 특정한 유형과만 결합할 수 있다. 이와 같은 제약은 종종 결합가에 대한 요구로 알려져 있다.

앞장에서의 어휘목록에 대한 논의는 UDG가 CAT, CASE, FORM 그리고 VAL과 같은 '자질' 을 가짐으로써 이러한 어휘목록의 양상을 다룬다는 사실을 명확하게 하는데, 이 양상들은 이 종류의 정보를 기호로 바꾸는 것을 허용하고 있다. 더욱이 이 자질들은 임의적인 것이 아니며, 매우 제한된 패턴을 따르고 있다. 즉,

(a') 보통 다섯 가지 이상의 주 문법적 범주는 고려되지 않는다. 즉, 명사, 동사, 형용사, 부사 그리고 전치사가 그것이다.

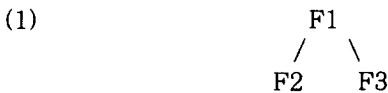
(b') 대부분의 언어는 격 구분, 성 혹은 부류, 명사와 형용사에 있어서의 수와 같은 그 이상의 것을 가지지는 않으며, 동사에 있어서는 시제, 양상, 태 등이 있다. 이러한 차이를 명명하기 위해 사용된 항목은 많은 언어에서 상당한 유사성을 보여주고 있다.

(c') 또한 가장 중요한 결합가를 기술하기 위해 사용된 항목은 다른 언어들 과 유사하다. 즉, 주어, 목적어, 전치사적 목적어, 비교 절, 시간적, 공간적, 장소적 수식어가 그것이다.

HPSG는 체계적인 방법으로 어휘목록을 구성하고자 시도한 최초의 틀구조이다. 이 어휘목록구성 목적으로 P&S는 Flickinger1987에서 처음으로 발전된 이른바 어휘계층(lexical hierarchy, LH)과 Bresnan에 의해 제안<sup>18)</sup>되었던 것과 같은 어휘규칙을 이용한다<sup>19)</sup>.

어휘계층은, 이 진술 단위가 낱말대신에 오는 자질이라는 사실은 제외하고, 연결된 의존구조 진술집합과 매우 유사하게 작용한다. 따라서 F1 → F2와 같은 진술은 만약 어떤 어휘목록이 F1을 가지며, F2도 가지고 있다는 것을 의미한다. F1, F2 또는 그 둘 모두는 또한 F3, F4 등등의 부가적인 자질의 원천이 될 수 있다 (F1 → F3, F2 → F3 등 등). 만약 규칙의 오른쪽 부분이 하나 이상의 자질을 가진다면, 이것은 언급된 자질 가운데 어떤 것이 규칙의 왼쪽에 있는 자질과 함께 나타날 수 있다는 의미로 해석된다.

이와 같은 규칙결합은 P&S에 의해서 이것이 공간상의 제한 때문에 왼쪽에서 오른쪽 대신에 위에서 아래로(top-downward) 표현된다. 위의 예는 다음과 같이 나타난다. 즉,



3장에서 제시된 자질구조의 항목에서는 이것은 두 개의 어휘목록 즉, F1|F2 그리고 F1|F3를 만들어 낸다.

보다 구체적인 예는 다음과 같다.



이것은 하나의 명사가 주격 또는 목적격을 가질 수 있고, 이 두 가지 가능성을 위한 어휘목록을 생성시킨다는 것을 의미하고 있다. 이 단순한 예를 기초로 해서

18) Bresnan, Joan(1982) : S.168-170.

19) ebd., S.36-37.



어휘계층은 문장구조의 표현에서 나오는 수형도구조를 가지고 있는 것처럼 보인다. 그러나 보다 자세한 예는 기대되는 모든 자질관계를 허용하기 위하여 각각을 교차시키는 수형도 가치를 허용할 필요가 있는 것처럼 보인다.

#### 4.2. 어휘지도 : FORM 하부구조

P&S87에서 제시된 어휘계층의 단순화는 하위유형 개념을 통해 가능하다. 조동사는 단순히 동사의 하위유형 즉, AUX이다. 따라서 이것을 GPSG와 HPSG에서 처럼 +AUX와 -AUX간의 구별을 할 필요가 없다.

조동사 자체는 다음과 같은 하위유형을 가지고 있다. 즉, MOD(modals, 화법), COP (copulas, 계사) 그리고 그 이외의 것들이다. 이 마지막 집단은 haben을 포함하고 있다.

VERB|FORM 하부구조가운데서 자질 PAS(passive)는 나타나지 않는다. 비록 다른 서유럽어와 마찬가지로 독일어가 gesprochen과 같이 PAS (passive, 수동)와 PSP (past participle, 과거분사)사이에서 형태적 구별을 하지 않을 지라도, GKPS85와 P&S87 둘 다는 그러한 형태의 통사적 습성(behavior)에 대한 두 개의 다른 유형 사이를 구별하기 위하여 VERB구조내에 있는 자질 FORM의 다른 하위유형을 사용한다. 과거분사가 haben이나 sein으로 지배되고 현재분사와 정동사형태를 일치시키는 것으로서 보충어와 유사한 역할 구성을 가지는 반면에, 이른바 수동형 분사들은 타동사의 X-ER(agent, 동작주)을 주어에서 von구로 밀어 넣고, X-ED(patient, 피동자)를 주어위치로 올린다. 이것은 결합가에서의 한 차이이다. 그러나 결합가를 매우 정확하게 지시하고 있는 문법체계에서 FORM 구조내에서의 그러한 구분은 불필요한 것처럼 보이며, 이 장에서 논의된 시스템에서는 이러한 구분을 하지 않는다.

#### 4.3. CONTENT 하부구조 : 역할과 결합가

의미적 역할은 결합패턴에 대한 다음의 논의로 직접 이끈다. Gruber1965 이후로 "동작주(agent)", "피동자(patient) 그리고 그 밖의 다른 것들과 같은 명사 보충어의 의미적 역할을 논의하는 것은 언어학자들에게 있어서는 일상적인 것이다. 비교적 최근에 Jackendoff 1987은 의미적 역할에 대한 보다 정교한 이론이 필요

하다는 주장을 했는데<sup>20</sup>), 왜냐하면 이것은 부분적으로 보충어인 동사에 의해 결정되기 때문이다. P&S는 일반적으로 명사적 보충어 대한 의미적 할당에서 Jackendoff를 따랐다. 여기서는 이 관계의 동작주와 피동자보다는 오히려 관계 X를 위해 "X-ER"과 "X-ED"를 참조함으로써 그렇게 할 것이다. 그러나 세 번째의 명사적 논항을 위해서는 수여동사의 "RECIPIENT(수령인)"을 참조하는 Fillmore의 견해를 따른다. 이러한 선택은 RELATION이 처음 두 역할을 결정하는 것보다 세 번째 역할을 결정하는 데는 거의 영향력이 없다는 것과 따라서 전체 낱말부류를 위해서 사용된 보다 일반적인 항목이 있을 수 있다는 것은 그럴듯한 것으로 여겨진다.

P&S는 LH에 대한 논의에서 의미적 역할을 다루지 않는다. 이들은 특정한 낱말의 결합에서 특정한 통사적 유형으로 될 수 있는 범주의 다양한 결합 각각을 고려함으로써 결합패턴을 다룬다. 따라서 정확하게 두 개의 결합가 즉, 명사 주어와 명사 목적어를 가지는 동사들은 "엄격한 타동사적(strict transitive)" 유형이며, 세 번째의 명사적 논항을 가지는 동사들은 "수여적" 유형이다.

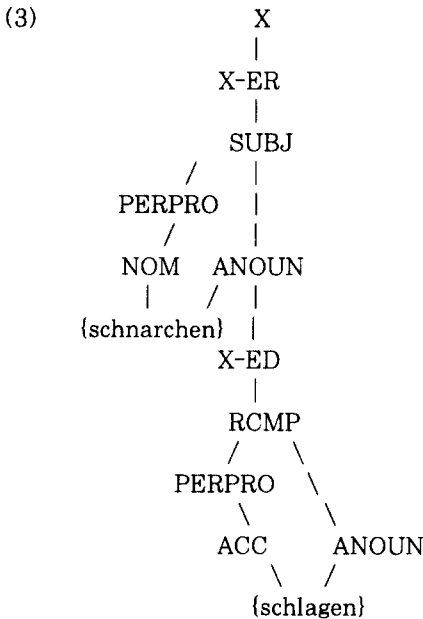
결합을 위해 LRM(Lexical Road Map, 어휘지도)을 발전시키는데 있어 여기서는 분할할 수 없는 것이 아니라 결합으로서의 의미적 역할과 이러한 다양한 결합 패턴을 다루는 다른 접근방식이 도입된다. 이때 위의 결합은 "X-ER"(LRM의 같은 경로에서 더 먼저 지시된 RELATION유형에 의해 이미 결정된 값에 의해 대체되는 X를 가진), "SUBJECT|ANOUN", "COMP|ANOUN" 등과 같이 개별 부분속으로 쪼개질 수 있는 것을 말한다. 이것은 의미적 역할에 있어 주어진 결합을 위해 다른 결합패턴을 만들어 내는 것을 가능케 하고, 따라서 능동과 수동, 문장의 주어와 그것의 다른 형태를 취하고 있는 변이형들 사이의 관계를 변형, 상위규칙<sup>21</sup>(metarule) 또는 어휘적 규칙없이 획득하는 것을 가능하게 한다. VERB|FORM의 하부구조인 자질과 더불어 이 LRM의 해석은 다른 것이 아니라, 현재의 자질의 상위자질이 각각의 경우에서 다른 방법으로 지시되지 않는 한 선택하는 자질의 하부구조로서의 그 자질구조의 부분들이다. 그와 같은 선택적 해석 가운데 한가지 경우는 다시 여기서는 { }에 의해 지시되는 PHON 하부구조인데, 왜냐하면 아마도 한 특정 언어를 위한 RELATION 구조의 아주 완전한 LRM은 그 대부분에서 많은 낱말을 가지기 때문이다. 덧붙여서 이 LRM에 있는 몇몇 동사들

20) Vgl., Jackendoff, Ray(1977) : a.a.O.

21) Gazdar, G., Klein, E., Pullum, G., Sag, I.(1985) : a.a.O., S.65f

이 두 개 또는 세 개의 결합가를 가지기 때문에, 그 경로에 있는 선행하는 자질보다는 RELATION의 하부구조로서 각각의 새로운 의미적 ROLE을 도입하는 것이 필요하다.

결합패턴의 가장 복잡하고 다양한 유형은 동사에서 발견되며, 유형 X의 RELATION을 표현하고 있다. 동사의 첫 번째 보충어는 결합유형 SUBJ(주어)를 가진다. 동사는 주어에 대해 하위범주화되어 있으며, 대부분 의미적 역할을 할당하고 있는데, 이것은 X-ER로 나타낸다. 이 논항은 ANOUN으로 표시된다. 단순한 타동사들도 또한 역할 X-ED와 범주 ANOUN을 가진 논항을 가지고 있다. 결합가의 이러한 결합가운데 하나에서, 만약 실제로 ANOUN이 하위유형 PERPRO(인칭대명사)를 가지고 있다면, 그것은 격형태 즉, 격 하위유형 NOM(주격)을 가지고 있는 SUBJECTS와 격유형 ACC(대격)을 가지는 다른 결합가 유형을 구분한다. 이것은 다음과 같이 LRM에서 표현된다.



ANOUN과 PERPRO사이의 분할은 위의 LRM에서는 두 번 나타난다. 이것은

각 ROLE를 위한 그리고 각 ROLE내에서 이 역할을 실현할 수 있는 각 결합유형을 위한 이러한 분할을 가능하게 한다. 그러나 이것은 매우 중요한 두 가지 일반화를 잃어버리는 것처럼 보이는데, 이 두 가지는 일반화는 (i) 인칭대명사의 격이 ROLE이 아니라 단지 결합유형에만 의존한다는 것, (ii) 격에 대한 이러한 사실은 SUBJ와 이에 대한 모든 경우를 위해 각각 반복될 수 있다는 것이다. 적절한 결합 유형의 서로 다른 경우들을 위해 LRM의 적절한 부분에 연결되는 몇 가지 방법으로 다음의 두 가지 경로를 가지는 것은 타당하다.



만약 지배하는 동사가 타동사라면 FORM 하부구조 자체는 하부구조 PSP를 가지며, 그 속에서 X-ER 역할을 가진 격 NPs가 von에 의해 도입되는 전치사적 보충어로 표현된다는 사실은 잘 알려진 바다. 이것은 X-ER에서 비롯한 부가적 가치를 요구함으로써 LRM에 표현된다. 이러한 방법으로 동작주를 표현하는 것은 동사 형태의 변화에 의해 수행될 것이다. 이것은 하나의 연결이 동사변화에 대한 논의에서 도입된 자질 PSP(과거분사)와 von-구로 이끄는 COMP 사이를 묘사할 것이라는 것을 의미한다. X-ER에서 나온 두 가지는 두 번째 논항의 역할인 X-ED에서 재결합되며, 이러한 논항의 통사적 역할의 표현을 위해 다시 분기되는데, 이 논항은 주어로서 X-ER을 가진 경로에 있는 보충어적 목적어 또는 von-구로서 X-ER을 가진 그 경로에 있는 주어를 말한다.

#### 4.4. 독일어의 부가어적 형용사

어형변화하는 형용사는 전통적으로 수, 성, 격과 같은 자질을 가지고 있다. 논리적으로 가능한 형태의 수는 그 언어의 이러한 각 자질이 만들어지는 구분의 산물이다. 독일어는 두 개의 수 하위유형(단수, 복수), 세 개의 성 하위유형(남성, 중성, 여성) 그리고 네 개의 격 하위유형(주격, 소유격, 여격, 목적격)을 가지고 있다.

이것은 논리적으로 전부  $2 \times 3 \times 4 = 24$ 개의 가능한 형태를 제공하고 있다.

독일어 형용사를 여기서는 “강변화”와 “약변화” 형태로 구분한다. 이들 사이의 선택은 (4a-b)에서 보여지는 것처럼 형용사가 나타나는 정확한 통사적 환경에 의존하고 있다.

(4a) ein gutes Buch ( stark nom/acc neuter singular )

(4b) dieses gute Buch ( schwach nom/acc neuter singular )

이러한 보다 상세한 구분은 논리적으로 전부 48개의 가능한 형태에서 나온 결과이다. 하지만 이 분석은 아주 불필요한 것처럼 보이는데, 왜냐하면 실제로는 단지 다섯 가지의 서로 다른 형태만이 존재하기 때문이다. 즉, gute, gutem, guten, guter, gutes. 이 모두는 각각 어형변화표에 하나 이상의 슬롯을 채우고 있다. guten은 실제로 전부 25 개의 슬롯을 채우고 있다.

P&S94는 다음의 두 규칙을 이용하여 48개의 출현가능 형태를 5개로 줄이는 방향으로의 중요한 단계를 취하고 있다<sup>22)</sup>.

(i) 자질 F를 위한 보충어인 x의 값을 의미하는 F,x 표시법을 사용하는 것

그리고

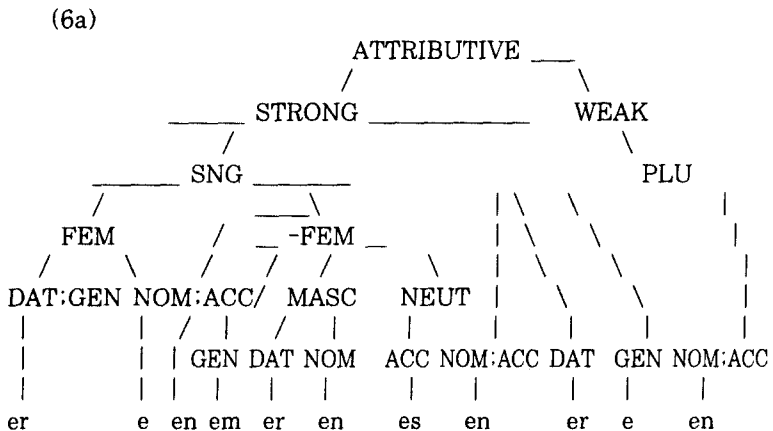
(ii) 몇몇 형태는 네 가지 자질 즉, 수, 성, 격, 강도(strength) 가운데 하나 혹은 그 이상의 자질을 특별히 지정되지 않는다는 것을 가정하는 것. 이것은 HPSG 틀구조의 구조분할에 관한 일반적인 장치 덕분에 가능하다. 구조분할은 일치 자질이 그것이 수정하는 ANOUNs의 값에 일치하는 값을 가질 필요가 없다는 것을 의미한다. 관련된 구성요소들에 있는 일치자질은 서로 양립될 수 있다. 이것은 예를 들면 만약 하나의 ANOUN이 GEN-MASC로 상세하게 열거된다면 변화하는 형용사는 GEN-FEM 또는 GEN-NEUTER로 상세히 열거되지 않는다는 것을 의미한다. 그러나 만약 그 형용사가 GEN-FEM 혹은 GEN-NEUTER로 상세하게 열거되거나 성에 대해서 상세히 기술되지 않고 남아 있다면, 이것은 여전히 남성 ANOUN과 양립할 수 있는 것이다. 이 전략은 P&S가 48개의 어형변화 형

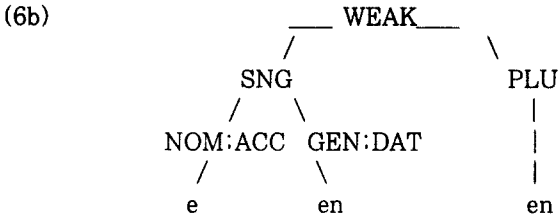
22) Pollard, Carl & Ivan Sag(1994) : a.a.O., S.35f.

태를 14개로 줄이는 것을 가능하게 한다. 즉,

(5)	TYPE	NUM	GND	CASE
guten	1 WEAK	SNG	—	GEN : DAT
	2 WEAK	PLU	—	—
	3 —	SNG	MASC	ACC
	4 STR	SNG	-FEM	GEN
	5 STR	PLU	—	DAT gute
	6 —	SNG	FEM	NOM : ACC
	7 WEAK	SNG	MASC	NOM
	8 WEAK	SNG	NEUT	NOM : ACC
	9 STR	PLU	—	NOM : ACC gutes
	10 STR	SNG	NEUT	NOM : ACC guter
	11 STR	SNG	FEM	GEN : DAT
	12 STR	SNG	—	GEN
	13 STR	SNG	—	NOM gutem
	14 STR	SNG	-FEM	DAT

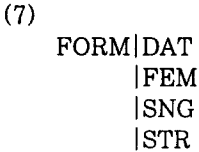
여기서 LH의 적용에서 나오는 결과와 LRM의 적용에서 나오는 결과에 대한 분석을 비교하는 것은 효과적이며, 아래에서 LH를 이용하는 첫 번째 분석을 제시한다.



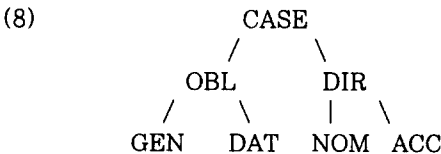


표시법 a:b는 관련된 형태가 값 a와 b사이를 구별하지 않는다는 것을 의미하며, 이것들 중 하나는 연합된 음운론적 형태와 양립할 수 있다.

이 수형도 아래에 있는 모든 형용사는 주어진 경로에 있는 모든 자질을 받아들인다. 따라서 예를 들면 제일 왼쪽에 있는 형용사 *guter*는 다음의 FORM 구조를 받아들인다. 즉,



만약 우리가 자질 OBL(oblique, 斜格) 즉, CASE의 하위유형으로서 그리고 DAT와 GEN의 상위유형으로서의 OBL을 도입한다면 이 도식은 더 간단해진다. 마찬가지로 DIR(direct, 직접격)은 NOM과 ACC의 상위유형이다. 이러한 항목은 전통문법에서 나온 것이다. CASE 유형의 계층관계는 따라서 다음과 같다. 즉,



두 번째, 독일어의 전통적인 어형변화에 대한 일반화는 자질명세 수를 효과적인 방법으로 줄이기 위하여 유형 계층을 사용하는 것을 가능하게 한다. 그 일반화는 복수형태들 즉, 유형 NUM의 하위유형 PLU가 성을 구분하지 않는다는 것이다.

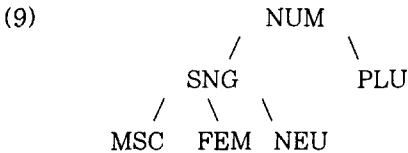
또 한편으로 이것은 만약 NUM의 하위유형이 PLU라면 하부구조 GND(gender, 성)를 간단하게 생략할 수 있게 한다. 그러나 만약 NUM의 하위유형이 SNG라면 또한 GND 하부구조도 생략하고자 할 것이다. 이것은 원칙적으로 가능하다. 왜냐하면 통합의 사용은 정해진 것보다 적은 즉, 어떤 낱말의 범주를 위한 문법의 한 부분인 자질을 단지 몇 가지만 정해놓은 그런 어휘목록을 가지는 것을 가능하게 하기 때문이다. 즉, 통합은 낱말의 자질구조 간의 동일성(identity)보다는 단지 양립성만이 일치하기를 요구한다.

그리고 실제로 이것이 아주 훌륭한 단순화를 이룰 것처럼 보인다. 그렇게 된다면 예를 들어 guter와 같이 MASC, NOM으로 상세화된 형용사는 Männer와 같이 NOM, PLU로 상세화된 ANOUN과 양립할 수 있을 것인데, 이것은 정확한 것이 아니다.

이 문제를 둘러싸고 있는 한가지 방법은 기본규칙이 있다는 것을 가정하는 것인데, 이 규칙 가운데 하나가 LRM에서 NUM을 상세하게 정하지 않은 ANOUNs의 경우 이 하부구조를 SNG로 명기하는 것이다. 이것은 기본규칙이 할 수 있는 일반화된 개념 내에서는 주의를 끈다. 예를 들면 이것은 음운적 실현이 없는 형태소의 경우 즉, 적어도 ANOUNs의 경우에 자질을 하나만 가지고 있는 것으로 명기하는 그런 경우이다.

그럼에도 불구하고 여기서 유형계층을 사용하는 다른 시도를 행하는데, 이것은 GND가 SNG의 하부구조라고 보는 것이다. 이것은 자질 규정에 대한 일반적인 생각을 뒤집는 것이기 때문에 인간의 직관에 반하는 것으로 보일지도 모른다. 하지만 성이 명기된 형용사가 암묵적으로 단수로 명기되는 것을 보장하는 효과를 가진다.

이 제안에 따른 NUM의 유형계층은 다음과 같다. 즉,



마지막으로, STRONG과 WEAK 형용사들 간의 차이를 보다 더 자세히 검토한다면 그 이상의 자질 축소가 가능한 것처럼 보인다.



이것에 대해서는 다음의 두 가지 중요한 사항을 고려해야 한다.

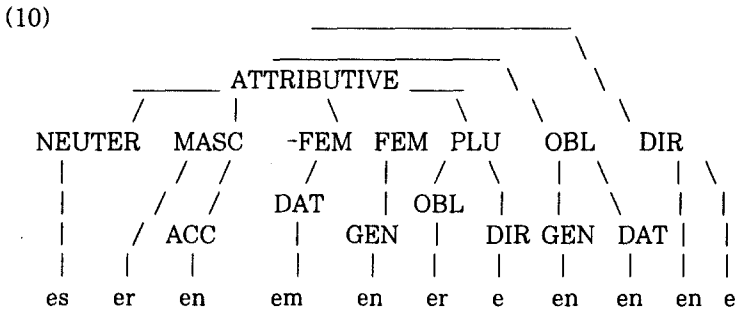
(a) 약변화 형용사들은 많은 슬롯에서 강변화 형용사와 동일하다.

(b) 다섯 개의 강변화 형태에 대립되는 것은 단지 두 개의 약변화 밖에 없다. 바꾸어 말하면 몇몇 차이는 강변화의 환경에서보다는 약변화 환경에서 일어난다.

이것은 다음 가설을 암시하고 있다. 즉, 약변화 형태가 강변화 형태와 동일하지 않는다는 점에 있어서, 약변화 형태는 강변화 형태의 정해진 것보다는 다양성이 적다는 것이다.

이러한 접근은 두 개의 양립할 수 있는 형용사 가운데 어떤 것이 보다 더 상세화되는가를 결정하는 명확한 절단(cut)방법이 있다는 것을 전제한다. 즉, 이것은 하나의 형태를 다른 것과 구분할 것을 요구하는 자질의 수이다. 보다 많은 자질을 요구하는 형태는 적은 자질에 의해 기술된 형태들보다 더 상세화된다.

이제 위에서 다루어진 어형변화를 처리하기 위하여 특정한 LH를 제시하는 다음과 같은 도식을 제시한다.

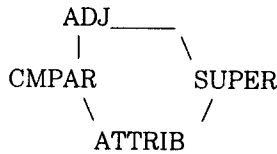


지금까지 본 연구는 P&S에 의해 사용된 종류의 어휘계층으로 독일어 형용사에 대한 두 가지 분석을 제시했다. 이것은 그 계층이 상당히 일반적인 규칙을 만들기에 충분한 것처럼 보인다. 그러나 여기서 비교급과 최상급 형태를 LH에 삽입할 수는 없다. 왜냐하면 그렇게 한다면 엄청난 양의 불필요한 계층이 나타나게 되며, 분기한 교점이 다시 하나로 모이는 것을 허용하지 않기 때문에 각 단계마다 세 번씩

반복적으로 다시 분기되어야 하기 때문이다. 이러한 반복은 대부분의 형용사 원급-비교급-최상급 변화가 동일한 방법으로 어형변화한다는 것 때문에 필요하다. 단지 소수의 형용사만 비교급과 최상급의 형태에서 어간모음이 변음된다.

그러나 만약 분기된 교점이 다시 한 교점에 모이는 것을 허용한다면 다음 단계를 만드는 규칙적인 방법에 대한 분석은 간단해진다. 처음에는 다음 차원으로 세 개의 길을 따라 분기된다. 이것은 다시 한 교점으로 수렴되고, 이 점에서 위에서 제시한 어형변화 자질 분석을 다시 시작한다. 이것은 자질구조만 고려한다면 다음 (21)과 같은 LRM의 형태를 가질 것이다.

(11)

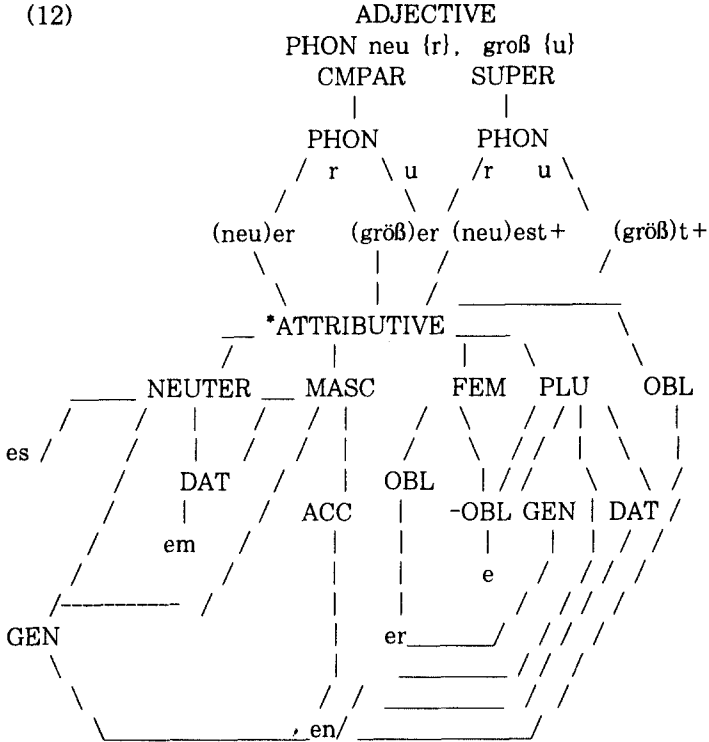


여기에 음운형식을 추가하는 것은 LRM을 보다 복잡하게 만드는 것이다. 그 이유는 비교급과 최상급을 만드는 접미사와 더불어 어간모음이 변하는 어근도 허용해야 하기 때문이다. 따라서 비교급과 최상급에서의 두 개의 읽기 분절인 r(regular)과 u(umlaut)를 제시한다.

최상급 어근을 가진 '+' 표시는 형태소 경계가 된다. 이것은 이러한 형태들이 분리된 낱말로는 존재하지 않지만, LRM에서 더 아래로 내려가는 어형변화 어미 중 하나에 의해 완전하게 될 필요가 있다는 것을 나타내고 있다.

어형변화되지 않은 술어형태를 만들기 위해 발전된 다섯 가지 경로 모두는 ADJECTIVE의 하위유형 ATTRIBUTIVE가 LRM에서 ADJECTIVE에 선행하는 경우에만 자질 FORM으로 모인다. 이 집적화 경향은 그들의 다양한 자질구조들로 다섯 가지의 어형변화된 형태를 만드는데 필요한 다중분기(multiple divergence)을 예비하는 것이다.

(12)



이 분석은 지금까지 제시되었던 것 가운데서 비교급과 최상급 접미사를 다루고 있는 것이기 때문에 가장 완전한 것이다. 어형변화 접미사에 대한 처리와 이전에 제공된 것들과 비교해 보면, 이것은 다섯 가지 형태를 기술하기 위해 11개의 자질을 이용하는 가장 간단한 것이라는 사실을 알 수 있다. 이러한 단순화를 얻기 위해 LH대신에 LRM을 재분류할 필요는 없었지만, 만약 관심을 어형변화 어미를 기술하기 위해 이용되는 LRM의 부분에 한정한다면, 4.3.장에서 제시된 것 보다 단순한 LH 확장이라는 사실을 알 수 있다. 왜냐하면 그 어떤 경로도 다른 경로와 한번 이상 교차하지 않고, LRM에 경로표지가 필요 없으며, 더구나 이렇게 한 점으로 수렴된 후에는 새로운 분기가 일어나지 않기 때문이다.

하지만 다음의 관점에서 보자면 LRM은 4.2.와 4.3.의 LRM보다 간단한 것은 아니다. 즉, 첫째, PHON 하부구조가 각 경로를 위해 다른 부분들로 쪼개진다는

것이다. 어근을 가진 부분(이것은 또한 기본형태와 비교급 형태를 가진 낱말이지만, 최상급 형태는 가지지 않은 것이다)은 어형변화 자질 이전에 나타나고, 어형변화 어미를 가진 부분은 이러한 자질 이후에 나타난다. 둘째, 자질 ATTRIBUTE는 들어오는 경로와 나가는 경로를 교차하는 교차점으로서 기호\*에 의해 상세화되는 것이다. 이 표지는 그 두 경로를 연결한다는 측면에서 받아들여지는 것이다.

## V. 결론

이상으로 통합의존문법을 이용하여 독일어에 있어서의 어휘목록 구성에 관해서 살펴보았다. 본 연구에서는 생성문법에서 출발한 HPSG와 GPSG같은 문법이론을 함께 분석하면서 의존개념을 바탕으로 한 문법이론에 초점을 맞추으로써, 자연어 처리를 위한 보다 효율적인 한 방안을 제시하고자 했다. 자연어처리를 위한 많은 관심은 연구자들로 하여금 다양한 접근방식으로서의 연구를 접하게 하고, 이것은 또한 이전보다 더 나은 결과를 제시할 수 있는 계기가 된다는 측면에서 볼 때, 본 연구는 긍정적인 측면을 내포하고 있다고 생각한다.

본 연구의 전개과정은 다음과 같다. 2장에서는 HPSG와 GPSG의 기본개념인 구성과 통합의존문법의 기본 개념인 의존을 비교함으로써 본 연구의 이론적 기반을 세웠다. 그와 더불어 낱말에서 구로의 확장, 구의 선형적 순서, 의존에 입각한 초기연구 등을 고찰함으로써 다음 단계로의 연구에 대비했다. 3장에서는 어휘목록에 관한 사항들을 중점적으로 고찰했다. 먼저 각 낱말의 자질표현에 대해 살펴보고, 자질의 분할, 범주 하위유형과 형식적 특성 그리고 내용과 결합에 대해 고찰했다. 그리고 4장은 본 연구의 핵심적인 장으로서, 어휘목록구성을 다루었다. 여기서는 먼저 자연어처리에 있어서 생성문법의 대표적인 HPSG에서의 문제점과 그 해결책을 간단하게 살펴보았으며, 이어서 어휘지도(Lexical Road Map)와 CONTENT의 하부구조인 역할과 결합에 대해서 그리고 마지막으로 독일어에서의 부가어적 형용사를 실제적인 분석과 더불어 어휘계층을 이용하여 독일어 형용사의 어휘지도를 제시했다.

본 연구는 비록 형용사 분석에 한정되어 있지만, 앞으로 좀 더 많은 기회가 주어진다면, 독일어의 다른 품사들에까지 그 연구범위를 확장시켜 고찰하는 것이 필요하리라 본다. 그런 다음 각 연구를 총괄하여 보다 효율적인 자연어처리를 위한 하나의 모델을 제시할 수 있도록 하는 것이 본 연구의 궁극적인 목표이다. 따라서

필자를 비롯한 많은 연구자의 관심과 노력이 필요하리라 본다.

## 참고문헌

- Bresnan, Joan. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
- Bresnan, Joan, Kaplan, Ronald, Peters, Stanley, Zaenen, Annie. 1982. Cross-Serial Dependencies in Dutch. *Linguistic Inquiry* 4. 613-635.
- Carpenter, Robert. 1991. The Generative Power of Categorical Grammars and Head Driven Phrase Structure Grammars with lexical rules. *Computational Linguistics* 17.3:300-313.
- Chomsky, Noam. 1982. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht:Foris.
- Fillmore, Charles. 1968. The Case for Case. in Bach & Harms 1968.
- Flickinger, Dan. 1987. *Lexical Rules in the Hierarchical Lexicon*. Phd. Dissertation. Stanford University.
- Gazdar, Gerald. 1981. Unbounded dependencies and Coordinate Structure. *Linguistic Inquiry*. 12:155-184.
- Gazdar, G., Klein, E., Pullum, G., Sag, I. 1985. *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.
- Gazdar, G., Pullum, G. 1982. *Generalized Phrase Structure Grammar. A theoretical synopsis*. Bloomington: Indiana University Linguistics Club.
- Gruber, Jeffrey. 1965. *Studies in Lexical relations*. Phd. Dissertation. Massachusetts Institute of technology.
- Hudson, Richard. 1984. *Word Grammar*. Oxford: Basil Blackwell
- \_\_\_\_\_. 1993. review of *Perspectives on Phrase structure: Heads and Licensing*, ed. by Susan D. Rothenstein. *Journal of Linguistics*: 29.1: 179-183.

- Jackendoff, Ray. 1977. X-Bar Syntax. A study of phrase structure. 2nd Printing. 1981. Cambridge: MIT Press.
- \_\_\_\_\_. 1987. The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry* 18.3:369-411.
- Lenerz, Jürgen. 1977. Zur Abfolge nominaler Satzglieder im Deutschen. Vol. 5, Studien zur deutscher Gramatik. Tübingen: Günter Narr.
- Lyons, John. 1968. Introduction to Theoretical Linguistics. Cambridge: University Press.
- Maxwell, Daniel. 1984. A typologically based principle of word order. *Language* 59.
- \_\_\_\_\_. 1992. Comparing Constituency and Dependency Grammar. in *Selections from the second annual Computational Linguistics in the Netherlands Meeting*.
- Maxwell, Dan, Klaus Schubert, & Toon Witkam(editors). 1988. *New Directions in Machine Translation*. Dordrecht: Foris.
- Mel' uk, Igor A. 1987. *dependency syntax: theory and practice*. Albany: State University of New York Press.
- Montague, Richard. 1977. *Formal Philosophy: Selected Papers of Richard Montague*. ed. by Richard Thomason. 3d Printing: 1981. New Haven: Yale University Press.
- Pollard, Carl & Ivan Sag. 1987. *InformationBased Syntax and Semantics: Volume 1: Fundamentals*. Stanford: CSLI Lecture Notes, #13.
- \_\_\_\_\_. 1991. An Integrated theory of Complement Control. *Language* 67: 63-113.
- \_\_\_\_\_. 1994. *HeadDriven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Pullum, Geoffrey. 1982. Syncategorematicity and the Analysis of Auxiliaries. *Glossa* 16:181-215.
- Schubert, Klaus. 1987. Metataxis: contrastive dependency syntax for

- machine translation. Dordrecht: Foris.
- Sgall, Petr. 1988. On some results of the conference. 243-250 in Maxwell, Schubert, & Witkam 1988.
- Starosta, Stanley. 1988. The Case for Lexicase: An outline of lexicase grammatical theory. London: Pinter. Stowell, Timothy. Origins of Phrase-Structure. Unpublished MIT Dissertation.
- Tesnière, Lucien. 1953. Esquisse d'une Syntax Structurale. Paris: Klincksieck.
- \_\_\_\_\_. 1959. Éléments de Syntaxe Structurale. Paris: Klincksieck [2. ed. 4 print. 1982].
- Uszkoreit, Hans. 1987. Word Order and Constituency Structure in German. Stanford: CSLI Lecture Notes #8.

## Zusammenfassung

### Eine Studie über die deutsche lexikalische Eintragung mit Unifikation-Dependenz Grammatik

Yongkyung Kim · Hyunshu Kim (Pusan Univ.)

In dieser Arbeit wird versucht, deutsche lexikalische Eintragung in Natural Language Processing(NLP) mit der Unifikation- Dependenz Grammatik zu untersuchen, weil der lexikalische Eintrag für die maschinellen Übersetzungen sehr wichtig ist.

Der Begriff der Unifikation, zuerst allgemein verwendet in der generativen Grammatik in den achtziger Jahren, stellt eine flexible Weise des Kombinierens der grammatischen Richtlinien von unterschiedlichen Quellen zu Einzelspezifikationen des Erzeugnisses einer Maßeinheit der Grammatik zur Verfügung. Unifikation ist in Bresnan 1982, Gazdar,

Klein, Pullum & Sag 1985 wichtig, um einige Arbeiten zu benennen, die es gebrauchen.

Im zweiten Abschnitt werden zuerst der Vergleich zwischen "Constituency" und "Dependency", Lineare Ordnung, die vorhergehende Arbeit über die Dependenz Grammatik und die Vereinigung erwähnt.

Im dritten Abschnitt wird das Lexikon gezeigt, das die Eigenschaft Matrix, die grundlegenden Abteilungen, Kategorie Subtypen und formale Eigenschaften und Inhalt und Wertigkeit umfasst.

Im vierten Abschnitt wird der Aufbau der lexikalischen Einträge ausgedrückt, der das Problem und seine Lösung in HPSG, lexikalische Road Map, Roles and Valents und Attributive adjektivische Flexionen auf Deutsch enthalten.