

## 국어와 독일어의 대조통사론과 기계 번역<sup>1)</sup> —격정보와 시제정보 표현의 대조성을 중심으로—

이 민 행 (제주대)

### 0. 머리말

국어와 독일어는 격정보와 시제정보의 부호화 *Enkodieren*에 있어 대조적인 방식을 취한다. 국어에서 독립적인 의미를 가진 시제형태소와 격형태소가 각기 동사와 명사의 어간에 부착됨으로써 시제정보와 격정보가 규칙성있게 표현된다. 독일어에서는 동사와 명사의 굴절 *Flexion*을 통해서 시제정보와 격정보가 표현되고 어간의 불규칙적인 변화들도 관찰된다. 때문에 독일어에서 어간과 시제형태소나 격형태소들을 분명하게 분할하는 것이 가능하지 않다. A.W.Schlegel (1818)이래 국어와 같은 언어유형은 부착어 *Agglutinierende Sprache*로서, 독일어와 같은 언어유형은 굴절어 *Flektierende Sprache*로서 규정되어 지고 있다 (Bußmann 1990:718). 이 글에서는 부착어의 하나인 국어와 굴절어에 속하는 독일어의 형태-통사론상의 대조적인 성격에도 불구하고 실험적인 국어-독일어 기계번역시스템의 구축이 어떻게 가능한 지를 보이려 한다. 이를 위해 두 언어의 대조적인 특성을 보완하는 의미에서 보편구조에 대한 연구가 뒷받침되어야 한다는 것을 주장하고자 한다.

논문의 구성은 다음과 같다. 제1절에서는 국어의 부착어적인 속성을 뚜렷하게 부각시켜 주는 등위접속구조를 독일어의 그것과 비교하여 논의한다. 제2절에서 인공지능어인 PROLOG 로 쓰여진 실험적인 국어-독일어

1) 이 논문은 대우재단 후원으로 개척된 '93 한국국어독문학 학술대회에서 발표한 것임.

기계번역시스템을 개관한 후 국어의 형태소 분석과정에 대해 설명한다. 제3절에서 국어와 독일어에 '공통적인 구조로서 시제구와 격구를 포함하는 보다 포괄적인 술어-논항 구조 Prädikat-Argument Struktur가 먼저 논의된다. 이어 기계번역의 한 단계인 원천언어 *Quellesprache* 국어 문장으로부터 술어-논항 구조가 얻어지는 국어의 통사분석과정에 대해서 설명한다. 제4절에서 공통의 술어-논항구조로부터 목표언어 *Zielsprache*인 독일어의 문장이 얻어지는 기계번역시스템의 생성단계에 대해 설명한다. 제5절에서는 맺는 말로서 이제까지 기술한 실험적인 국어-독일어 기계번역시스템의 한계와 개선방향에 대해서 논의한다. 논문말미의 부록에 몇가지의 PROLOG 술어들에 대한 정의와 번역시스템의 실행 예가 제시되어 있다.

### 1. 국어와 독일어의 등위접속구조에 대한 대조분석과 국어의 부착성 Agglutinität

부착어에서는 독립적인 통사적 기능을 지니는 형태소안에 다양한 문법적인 정보들이 들어있다. 국어의 등위접속구조는 국어의 격형태소와 시제형태소가 독립적인 통사적 기능을 한다는 사실을 명료하게 보여준다. 다음의 예 (1a)-(1f)들을 통해 우리는 국어의 격형태소들이 독립적인 통사적인 단위, 곧 하나의 구성성분이 된다는 것을 알 수 있다.

- (1) a. 수미와 미미가 독문학을 전공한다.  
 b. 수미는 방학중에 영어와 컴퓨터를 배우려 한다.  
 c. 수미가 미미와 영미에게 컴퓨터를 함께 배우자고 제안한다.  
 d. 수미는 미미와 영미의 컴퓨터에 관한 지식도 별 것이 아니라고 생각한다.  
 e. 그러나 미미는 오빠와 언니로부터 컴퓨터를 배워 벌써 컴퓨터를 사용할 줄 안다.  
 f. 영미는 컴퓨터와 자동차에 대해 두려움을 가지고 있다.

(1a)에서 고유명사 '수미'에 등위접속형태소인 '와'가 부착되어 있고

고유명사 '미미'에 격형태소 '가'가 부착되어 있다. 등위접속구문에 대한 연구자들의 일반적인 견해에 의하면, 통사적 범주가 동일한 언어표현들만이 등위접속된다. (정재현 1990:19). 이러한 입장을 따라 우리는 예문 (1a)에서, 명사구 '수미'와 격조사구 '미미가'가 등위접속사 '와'에 의해 접속된 것이 아니라, 명사구 '수미'와 명사구 '미미'가 등위접속된 것으로 이해한다.<sup>2)</sup>

이때 등위접속구 '수미와 미미'는 격조사 '가'의 보족어 Komplement로 기능하는 것으로 분석한다. 마찬가지로 (1b)에서는 격조사 '를'이 등위접속구 '영어와 컴퓨터'를, (1c)에서는 격조사 '에게'가 등위접속구 '미미와 영미'를, (1d)에서는 격조사 '의'가 등위접속구 '미미와 영미'를, (1e)에서는 격조사 '로부터'가 등위접속구 '오빠와 언니'를, (1f)에서는 격조사 '에 대해'가 '컴퓨터와 자동차'를 각각 보족어로 취한다고 분석할 수 있다. 이와 같이 국어의 격조사들은 형태론적으로 명사에 부착되어 있지만 통사적으로 독립적인 기능을 갖는다. 독립적인 기능을 갖는 점에서는 등위접속사도 격조사와 마찬가지로이다.

다음의 예문들은 (1a)-(1f)의 국어문장들과 같은 의미를 갖는 독일어문장들이다.

(2) a. Swumi und Mimi studieren Germanistik.

b. Während der Semesterferien will Swumi Englisch und Computer lernen.

c. Swumi schlägt Mimi und Yeongmi vor, Computer zusammen zu lernen.

d. Swumi glaubt, daß Mimis und Yeongmis Kenntnisse über Computer nicht besonders gut seien.

e. Aber Mimi kann schon Computer benutzen, weil sie es bei ihrem Bruder

2) 동일한 입장을 따르면서 접속구 '수미와 미미가'를 단순히 명사구와 명사구의 등위접속으로 간주하고 접속성분 '미미가'의 경우 격정보를 더 가지고 있는 것으로 분석할 수도 있다. 그러나 이러한 분석방법은, 예를 들어 '수미와 미미와 영미가'와 같은, 접속성분이 세 개 이상인 접속구의 구조에 대한 일관성 있는 설명을 제공하지 못한다.

und ihrer Schwester gelernt hat.

f. Yeongmi hat Angst vor Computern und Autos.

(2a)-(2c)에서 독일어의 1격과 4격, 3격 명사들이 등위접속되어 있다. 2격명사들이 등위접속되어 있는 (2d)에서 2격 표시의 격형태소인 's'가 'Mimi'와 'Yeongmi' 모두에 붙어 있는 것이 그 예문에 대응되는 (1d)에서의 국어의 경우와 다르다. (2e)에서는 3격 명사구들이 등위접속된 접속구 'ihrem Bruder und ihrer Schwester'가 전치사 'bei'의 보속어로 기능한다. 이 점에서 독일어의 전치사와 국어의 격조사가 유사성을 지닌다고 할 수 있다. (2f)에서도 전치사가 등위접속구를 보속어로 취하고 있다.

다음의 예들은 국어의 시제형태소들이 독자적인 통사기능을 지닌다는 것을 보여준다.

(3) a. 싸고 유익하-ㄴ 책

b. 동생이 책을 고르고 형이 값을 지불하-ㄴ다.

c. 동생이 책을 고르고 값도 지불하-였다.

d. 내일은 비가 오고 천둥이 치-겠다.

위의 (3a)-(3d)에서 등위접속된 두 개의 성분 중 두번째 성분에만 시제 형태소가 부착되어 있는 것을 볼 수 있다. 이 점에서 시제형태소와 격형태소의 유사성을 찾을 수 있다. 여기에서도 동일한 통사범주를 가진 언어표현들만이 등위접속된다는 일반론을 따라 (3a)의 경우처럼 형용사들이 등위접속되어 다시 형용사구를 이루거나, (3c)에서와 같이 시제중립적인 동사구들이 등위접속되어 다시 동사구를 만들거나, (3b)와 (3d)에서처럼 시제가 표현되지않은 문장들이 등위접속되어 다시 문장을 이루는 것으로 간주한다. 이때에 각 시제형태소들이 등위접속된 결과표현들을 보속어로 취하는 것으로 분석한다. 예를 들어 (3c)에서 과거시제형태소인 '였'이 등위접속구인 '책을 고르고 값도 지불하-'를, (3d)에서 미래시제형태소인 '겠'이 등위접속표현 '내일은 비가 오고 천둥도 치-'를

각각 보족어로 취한다.<sup>3)</sup>

(3b)-(3d)에서 시제형태소와 그 보족어가 함께 시제구 Tempusphrase를 형성하는데 이 시제구들은 다시 양상 Modus을 나타내는 형태소 '-다'의 보족어가 된다. 따라서 양상형태소도 시제형태소처럼 형태소분석시에 분리되어야 한다.

(4) a. das Buch, das billig und nützlich ist

b. Der jüngere Bruder wählt und der ältere bezahlt das Buch.

c. Der jüngere Bruder wählte das Buch und bezahlte auch.

d. Es wird morgen regnen und donnern.

(4a)-(4d)의 경우에 그것들에 대응되는 국어의 예들과 달리 등위접속된 두 성분 중 어느 한 쪽에만 시제형태소가 붙어 있는 경우는 없다. (4a)와 (4d)에서는 시제중립적인 성분들이 등위접속되어 있고, (4b)에서 3인칭/단수의 현재시제형태소 '-t'를 지닌 성분들이 등위접속되어 있으며 (4c)에서는 과거시제형태소 'te'를 지닌 동사구들이 접속되어 있다. (4b)와 (4c)의 경우에 시제형태소들을 동사의 어간으로부터 분리해 내는 것이 바람직하지 않다. 왜냐하면 불규칙 인칭변화를 하는 동사들의 경우에 어간과 시제형태소를 구분하는 것이 가능하지 않기 때문이다. 이 점에 독일어와 국어의 차이가 있다. 지금까지 논의한 바와 같이 격형태소와 시제형태소들이 형태론적으로 명사, 형용사와 동사의 어간에 부착되어 있다고 하더라도 통사론적으로 독립적인 기능을 갖기 때문에 국어의 경우 통사분석에 앞서 이 형태소들을 명사, 형용사와 동사의 어간들로부터 분리해 내

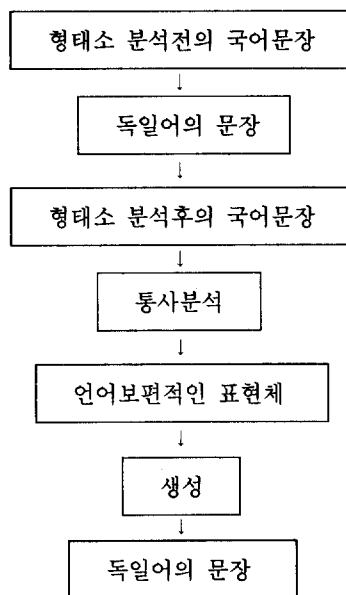
3) 국어에 미래시제가 존재하느냐의 문제를 놓고 의미론자들간에 논의가 분분하다. 그것의 쓰임이나 의미를 별도로 하고 형태론적 측면만을 고찰한다면 현재와 과거를 표현하는 시제 형태소들과 구분되는 또 하나의 시제형태소가 존재한다는 사실을 부인할 수 없다. 독일어의 경우에도 미래조동사 'werd.'가 화법조동사의 기능을 지니기 때문에 그것의 존재자체가 의문시된다.

는 형태소의 분석이 선행되어야 한다. 접속형태소와 양상형태소가 또한 통사론적으로 독자적인 기능을 가진다는 점은 이미 언급한 바 있다. 형태소분석의 과정에 대해 절을 달리해서 논의한다.

## 2. 국어의 형태소 분석기와 형태소 분석과정

여기에서 기술하는 실험적인 국어-독일어 기계번역 시스템은 언어보편적인 interlingual 모델에 속하는 것으로 그 구조는 다음과 같다.<sup>4)</sup> 머리말에서 언급했듯이 이 시스템은 PROLOG로 프로그램되어 있다.

(5)



4) 기계번역시스템의 모델에는 언어보편적인 모델 Interlinguales Modell외에 변환모델 Transfer-Modell과 직접모델 Direktes Modell이 있다.

이제 예를 통해서 국어의 형태소분석의 과정에 대해서 살펴보자.<sup>5)</sup>

(6) % 토니오가 연극을 좋아한다.

i. [tonioka,yeonkukul,coahanta]

ii. [tonio,ka,yeonkuk,ul,coaha,n,ta]

위의 (6i)은 국어문장 “토니오가 연극을 좋아한다”를 로마자로 표기하여 인공지능언어인 PROLOG의 자료구조인 리스트형식으로 나타낸 것이다. 이 입력 리스트를 형태소분석기가 분석하여 그 결과로서 (6ii)을 내놓는다. 출력된 리스트(6ii)에서 우리는 격형태소 “ka”, “ul”과 시제형태소 “n”과 양상형태소 “ta”가 분리되어 있는 것을 본다. 형태소분석은 순환적으로 적용되는 2항술어 morphsyn에 의하여 실행된다. 이 술어의 첫번째 논항은 리스트형식으로 쓰여진 분석전의 국어의 문장이고 두번째 논항은 형태소분석의 결과문장이다. 술어 morphsyn은 다른 2항술어 morphanal을 사용하여 첫 논항인 리스트의 첫번째 원소를 형태소별로 분할하고, 그 리스트의 나머지 원소들을 모아 만든 새로운 리스트에 순환적으로 적용된다. 이 과정은 리스트의 마지막 원소의 형태소분석이 끝날 때까지 계속된다. 그리고 내장술어인 append가 분석된 결과들을 모두 모아 하나의 리스트로 만들면 이것이 형태소분석의 최종결과가 된다. 위의 (6i)이 (6ii)로 바뀌는 형태소분석의 과정이 구체적으로 다음의 (8)에 보여지고 있다.

(7) ?- morphsyn([tonioka,yeonkukul,coahanta],Sm).

Sm = [tonio,ka,yeonkuk,ul,coaha,n,ta].

(8) ?- morphsyn([tonioka,yeonkukul,coahanta],Sm).

--> ?- morphanal(tonioka,W1).

W1 = [tonio,ka].

5) 국어표현의 로마자 표기를 위해 Yale Romanization 시스템이 사용된다.

--> ?- morphsyn([yeonkukul,coahanta],Sm2).  
 --> ?- morphanal(yeonkukul,W2).  
 W2 = [yeonkuk,ul].  
 --> ?- morphsyn(coahanta,Sm3).  
 --> ?- morphanal(coahanta,W3).  
 W3 = [coaha,n,ta].  
 --> ?- morphsyn([],Sm4).  
 Sm4 = [].  
 --> ?- append(W3,Sm4,Sm3).  
 Sm3 = [coaha,n,ta].  
 --> ?- append(W2,Sm3,Sm2).  
 Sm2 = [yeonkuk,ul,coaha,n,ta].  
 --> ?- append(W1,Sm2,Sm).  
 Sm = [tonio,ka,yeonkuk,ul,coaha,n,ta].

술어 morphanal은 여러 개의 형태소로 구성된 개별적인 단어를 시스템 내에 주어진 사전 Lexikon을 이용하여 최대한으로 형태소별로 분할하여 하나의 리스트로 만든다. 다음의 (9)에 이 시스템에서 사용되는 사전의 일부가 제시되어 있다. 위의 (8)에서 술어 morphanal이 사용된 경우들을 나열해보면 다음의 (10)과 같다.

- (9) a. word(tonio,npr(d(tonio))).  
 b. word(yeonkuk,cn(n(theater))).  
 c. word(ka,case(kp(ka),agent)).  
 d. word(ul,case(kp(ul),theme)).  
 e. word(coaha,v2(v(moeg\_))).  
 f. word(n,tense(pres(n),pres)).  
 g. word(ta,mood(decl)).



(10) a. ?- morphanal(tonioka, W1).

W1 = [tonio,ka].

b. ?- morphanal(yeonkukul, W2).

W2 = [yeonkuk,ul].

c. ?- morphanal(coahanta, W3).

W3 = [coaha,n,ta].

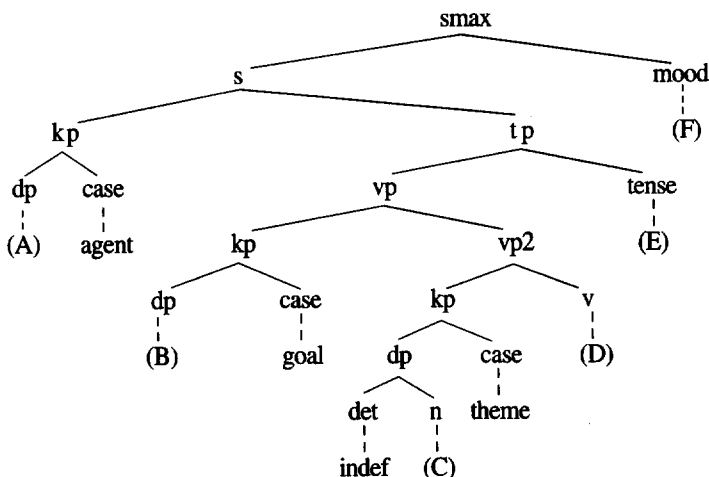
술어 morphanal이 어떻게 PROLOG로 정의되어 있는 지에 대해서는 여기서 설명하지 않겠다. 프로그램에 관심있으면 Lee(1992a:175ff.)을 참고하기 바란다.

### 3. 국어문장의 통사분석단계와 언어보편구조로서의 술어-논항 구조

이 절에서는 국어문장의 통사분석단계에 대해서 논의한다. 통사분석단계에서는 형태소분석이 끝난 국어문장의 통사구조를 분석하여 술어-논항 구조로 나타낸다. 이 술어-논항 구조로부터 후에 생성단계에서 목표언어의 문장이 생성된다. 따라서 우리는 통사분석의 결과로 나타날 술어-논항 구조를 언어보편적인 구조가 될 수 있도록 설정해야 한다. 이러한 맥락에서 보편문법을 지향하는 원리와 매개변향이론적인 틀내에서의 대조 연구의 결실인 Pollock(1989)의 시제구 Tempusphrase, TP분석과 Abney(1987), Shin(1991)의 격구 Kasusphrase, KP분석을 따라 시제구 TP와 격구 KP를 포함하는 다음의 (11)과 같은 구조를 단순문에 대한 하나의 언어보편적인 통사구조로 가정한다.<sup>6)</sup>

6) 여기서 국어와 독일어에 공통적인 통사구조로 가정되고 있는 수형도는 국어와 독일어가 형상적인 언어 Konfigurationale Sprache라는 전제에서 출발한 것이다. 국어가 형상적인 언어라는 언어학적인 논의가 Han(1987)과 Lee(1992b)에서 이루어지고 있다. 독일어의 형상성에 대해서는 Fanselow(1987)가 충분한 논거를 제시하고 있다.

(11)



위의 구조에서 첫째, 격(“case”)과 시제(“tense”)와 양상(“mood”)이 각각 독립적인 구성성분을 형성하고 있다는 것, 둘째, “agent”, “goal”와 “theme” 등 의미적인 기능들이 격으로 표현되어 있다는 것과 셋째, 이 의미기능들이 위계적인 구조를 통해 나타나 있다는 것이 특징적이다.<sup>7)</sup> (11)의 구

7) 의미적 기능에 대한 논의가 생성문법내에서 최근들어 아주 활발하다.

Parsons(1990:73-74)에 의미적 기능들과 통사적 기능들간의 전형적인 관계들이 아래의 (A7)와 같이 제시되어 있다. 이 글에서도 설명의 편의를 위해서 그의 견해를 수용하고 있다. 그러나 기본적으로 의미기능들은 술어에 따라 무한히 다양한 방식으로 문장내에 나타날 수 있다는 것이 필자의 입장이다. 특히 기계번역을 위해서는 대조연구에 바탕을 둔 의미기능들 사전이 크게 유익할 것이다.

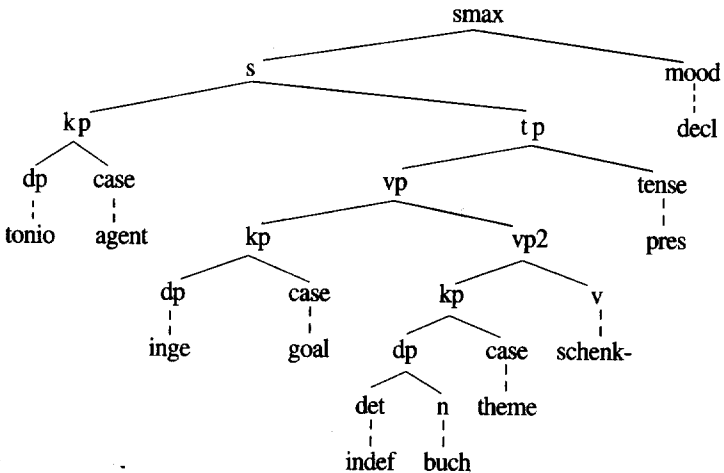
(A7)

Role	Typical Position in Active Sentence
Agent	Subject
Theme	Direct Object; subject of 'is'
Goal	Indirect object, or with 'to'
Benefactive	Indirect object, or with 'for'
Instrument	Object of with; subject
Experiencer	Subject

조는 언어보편적인 것이므로 국어와 독일어에도 공통으로 존재하는데, 다음의 (12a)와 (12b)은 이러한 통사구조를 갖는 국어와 독일어의 문장이다. 구조 (11)의 비어있는 자리들 (A),(B),(C),(D),(E),(F)에 각각 "tonio", "inge", "buch", "schenk-", "pres", "decl"을 채워넣은 구조 (12c)가 문장 (12a), (12b)가 공통으로 갖는 통사구조이다.

(12) a. 토니오가 잉에에게 책을 선물한다.

b. Tonio schenkt Inge ein Buch.



(12c)의 수형도는 아래의 (13)과 같은 PROLOG의 술어-논항 구조에 대응된다.

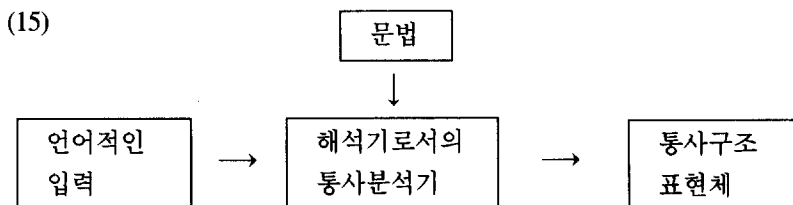
(13) smax(s(kp(dp(tonio),case(agent)),tp(vp(kp(dp(inge),case(goal)),  
vp2(kp(dp(det(indef),n(buch)),case(theme)),v(schenk\_))),tense(pres))),  
mood(decl))

국어문장의 통사분석과정은 바로 예문 (12a)에 대응하는 PROLOG의 리스트 (14a)가 형태소 분석과정을 거친 후에 얻어진 결과인 리스트 (14b)

의 통사구조로서 (13)과 같은 술어-논항 구조가 만들어지는 과정을 일컫는다.

- (14) a. [tonioka,ingeeykey,chaykul,senmwulhanta].  
 b. [tonio,ka,inge,cykey,chayk,ul,senmwulha,n,ta].

국어문장의 통사분석을 위해 Pereira/Shieber(1987)의 좌측코너통사분석기(“Left-corner parser”)를 채택하고 있는데, 이 분석기는 국어의 통사규칙에 빈번히 나타나는 좌측순환성(“left recursion”)에 의한 무한루프의 문제를 해결하는 데에 적합하다.<sup>8)</sup> 여기서 기술하고 있는 기계번역시스템에서 통사분석기는 원천언어의 문법을 해석하는 해석기 Interpreter로서의 기능만 할 뿐으로 보다 중요한 것은 문법이다. 다음의 도식은 문법과 통사분석기와의 관계를 잘 보여주고 있다.



여기에서 사용하고 있는 문법은 전통적인 구절구조문법을 확대/수정한 것인데 그 특성들을 살펴보자. 다음의 (16a)-(16g)에서 보듯이 격, 시제와 양상을 독자적인 통사범주로 취급한다.

- (16)a.  $smax(smax(S, mood(M))) \rightarrow [s(S), mood(M)]. /*\ smax \rightarrow s\ mood\ */$   
 b.  $s(s(KP, T)) \rightarrow [kp(KP, agent), tp(T, TNS)]. /*\ s \rightarrow kp\ tp\ */$

8) 좌측순환성이란 ‘수미의 언니의 친구의 동생’과 같이 어떤 구의 원편의 성분이 그것과 동일한 구조를 가지는 경우이다. 이때 어떤 규칙이 순환적으로 사용된다.

- c.  $tp(tp(VP, tense(T)), TNS) \rightarrow [vp(VP), t(T, TNS)]$ . /\*  $tp \rightarrow vp\ t\ */$
- d.  $kp(kp(DP, P), KAS) \rightarrow [dp(DP), case(P, KAS)]$ . /\*  $kp \rightarrow dp\ case\ */$
- e.  $word(ta, mood(decl(ta)))$ .
- f.  $word(ss, t(past(ss), past))$ .
- g.  $word(i, case(case(i), agent))$ .

또한 Abney(1987), Olsen(1989)의 관사구 분석을 받아들여, 전통적인 명사구를 관사구 Determinator Phrase 로 간주하는데, 고유명사는 0가 관사로서 그리고 보통명사는 1가 관사로서 범주화한다. 다음을 보자.

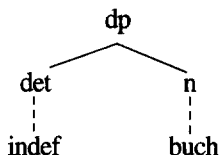
- (17) a.  $dp(dp(D)) \rightarrow [npr(D)]$ .
- b.  $dp(dp(Det, N1)) \rightarrow [det(Det), n\_bar(N1)]$ .
- c.  $dp(dp(det(indef), N)) \rightarrow [cn(N)]$ .
- d.  $word(tonio, npr(d(tonio)))$ .
- e.  $word(ku, det(demon))$ . /\*  $d\_ \ */$
- f.  $word(chayk, cn(n(buch)))$ . /\*  $buch\ \ */$

위의 (17c)는 보통명사가 관사없이 나타나서 관사가 채워져야 할 자리가 비어 있을 때에, 비한정성("indefiniteness")을 나타내는 통사정보인 "indef"를 관사 det의 논항의 자리에 자동으로 삽입하기 위한 하나의 일반규칙("Default Rule")이다.<sup>9)</sup> 이 규칙이 리스트 (14b)의 통사분석식에 적용됨으로 해서 통사분석의 결과인 술어-논항구조 (13)안에  $dp(det(indef), n(buch))$ 라는 관사구가 생긴다. 이 관사구의 구조를 수형도로 나타낸 것이 다음의 (18)이다.

---

9) 명사가 관사나 지시대명사가 없이 홀로 나타나는 경우가 국어에 아주 빈번한데, 그 때에 독일어의 부정관사나 정관사 중 어느 쪽이 더 적합한 지는 문맥이 결정한다. 그래서 언어보편적인 것으로 간주되는 관사를 위해 마련된 관사자리를 특화하지 않은 채(unterspezifiziert) 놓아두는 것도 하나의 방법이 될 수 있겠다.

(18)



통사분석은 3항 술어 parse에 의해 실행된다. 이 술어가 형태소분석이 끝난 리스트형식의 국어문장을 입력으로 하여 그것에 대한 통사구조를 술어-논항 구조로 내보낸다. 앞서 논의한 (14b)의 리스트로 표현된 문장으로부터 (13)의 술어-논항 구조가 얻어지는 과정을 다음과 같이 예로 들 수 있겠다.

(19) a. parse([tonio,ka,inge,eykey,chayk,ul,senmwulha,n,ta],[ ],SR1), SR1 =. [\_ ,SR].

b. SR =

smax(s(kp(dp(tonio),case(agent))),tp(vp(kp(dp(inge),case(goal))),  
vp2(kp(dp(det(indef),n(buch)),case(theme)),v(schenk\_))),tense(pres))),  
mood(decl))

문장 “tonioka ingeeykey chaykul senmwulhanta”에 대한 통사구조가 (19b)의 술어-논항 구조인 SR이다. 이 술어-논항 구조를 별도로 Lehner (1990:168-173)에 의해 정의된 1항 술어 drucke\_baum을 통해서 수형도로 나타낸 것이 위 (12c)이다. 언어보편구조인 술어-논항 구조로부터 독일어의 문장이 얻어지는 과정이 다음 절에서 논의된다.

#### 4. 언어보편구조로부터 독일어 문장이 생성되는 단계

독일어 문장의 생성단계에서는 자질구조 Merkmalsstruktur를 보강한 한정절문법 (“Definite-Clause-Grammar”)이 생성기 Generator의 기능을 한다. 문장의 생성은 문장의 통사구조분석과 역방향으로 실행된다.<sup>10)</sup> 어떤 통사

10) 이 시스템에서 통사구조로부터 독일어 문장의 생성을 위해 채택되고 있는 하향

구조가 주어졌을 때에 그 구조를 갖는 문장을 찾아내는 과정이 바로 생성과정이라고 할 수 있다. 예를 들어 앞 절에서 다룬 통사 구조 (18b)로부터 독일어 문장을 생성하는 것은 그 구조를 갖는 독일어 문장을 찾아내는 것이다. 바꾸어 말하자면 어떤 독일어 문장을 통사분석했을 때 그러한 통사구조를 얻게 되는 지를 알아내는 것이 생성의 과정이다. 앞 절에서 논의했듯이 통사구조 (19b)를 갖는 독일어 문장은 다음의 (20a)이고 이를 PROLOG의 리스트로 표현한 것이 (20b)이다.

- (20)a. Tonio schenkt Inge ein Buch.  
 b. [tonio,schenkt,inge,ein,buch,.]

이 생성단계에서 통사구조 (19)로부터 (20b)의 리스트가 얻어진다. 생성하는 과제를 떠맡는 술어는 4항술어 *gsmax*이다. 이 술어의 첫 논항이 생성되어야 할 독일어 문장을 위한 자리이고 두번째 논항은 빈 리스트이며 세번째 논항이 통사구조이고 네번째 논항은 독일어 문장의 자질구조를 나타내기 위한 자리이다. 다음의 (21a)-(21c)는 리스트 (20b)의 생성과정을 보여준다.

- (21) a. ?- *gsmax*(DS,[ ],  
*smax*(s(*kp*(*dp*(tonio),*case*(agent)),*tp*(*vp*(*kp*(*dp*(inge),*case*(goal)),  
*vp2*(*kp*(*dp*(*det*(indef),*n*(buch)),*case*(theme)),*v*(schenk\_))),*tense*(pres))),  
*mood*(decl)),FS).  
 b. >>> K-D-Uebersetzung:

DS = [tonio,schenkt,inge,ein,buch,.]

---

식("top-down")의 한정절문법은 등위접속구문의 생성에 적합치 못하다. 이에 대한 보완이 후에 이루어져야 할 것이다.

c. FS = [phon:schenkt,cat:v,  
 subc:[first:[agr:[num:sg,per:3 : \_],  
 case:nom,phon:tonio,cat:det : \_],  
 second:[case:dat,phon:inge,cat:det : \_],  
 third:[case:acc,phon:ein,cat:det,  
 comp:[gen:neut,phon:buch,cat:n  
 :\_:\_:],  
 mood:[phon:.,cat:mood:\_:\_:]

위의 독일어 문장의 생성과정에서 흥미로운 점이 몇 가지 있다. 그들 중의 하나는 언어보편적인 구조인 (21a)의 세번째 논항자리에 관사구의 비한정성(“indefiniteness”)에 대한 통사적인 정보가 들어 있고 그 언어보편적인 구조에 대응되는 생성문 (20b)에는 부정관사 “ein”이 나타나 있다는 사실이다. 이 부정관사의 생성에는 생성단계에 적용되는 다음의 (22a),(22b)와 같은 독일어의 통사규칙과 어휘규칙이 관여한다.

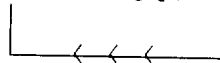
- (22)a.  $gdp(dp(Det,N),KAS,Fdp) \rightarrow gdet(Det,KAS,Fdet), gcn(N,Fcn),$   
 {Fdp == Fdet,Fdp:comp == Fcn}  
 b.  $gdet(det(indef),theme,Fdet) \rightarrow [ein], \{Fdet:phon == ein,$   
 Fdet:cat == det, Fdet:comp:gen == neut}.

독일어의 경우, 굴절어로서의 특성상 격형태소와 시제형태소를 명사와 동사의 어간으로 분리시키는 것이 가능하지 않다. 독일어 문장을 생성할 때에 어떻게 보편적인 구조에 표현되어 있는 격정보와 시제정보를 지니는 명사와 동사의 표현을 찾아낼 수 있는가도 흥미있는 문제이다. 이 문제의 해결을 위해 독일어에도 시제(T)와 격(K)의 두가지 추상적인 통사범주와 그것들을 위한 자리가 통사규칙내에 존재한다고 가정하겠다. 일종의 통사규칙내에 마련된 전승기재 *Vererbungsmechanismus*에 의해서 그 자리들은 동사구와 관사구가 지니고 있는 해당정보들에 의해 채워진다.



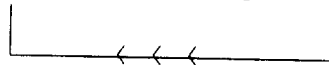
다음이 규칙의 예들이다.

(23) a.  $gtp(tp(VP, tense(T), Ftp) \rightarrow gvp(VP, T, Fvp), \{ \dots \})$ .



시제정보의 전송

b.  $gkp(kp(DP, case(KAS)), KAS, Fkp) \rightarrow gdp(DP, KAS, Fdp), \{ \dots \}$ .



격정보의 전송

독일어의 경우에 주어와 정동사간에 인칭과 수에 관한 일치 *Kongruenz* 가, 관사와 명사간에는 성, 수와 격에 관하여 일치가 이루어져야 한다. 이러한 일치현상은 영어와 독일어 등의 일부 언어에만 존재하므로 언어보편적인 관점에서 다룰 것이 아니고 해당언어의 분석이나 생성과정에서 별도로 다루어야 한다. 이 시스템에서는 이를 위해 독일어의 생성기에 자질구조를 보강했다. 예를 들어 다음 (24a)-(24d)의 어휘기재항들은 모두 *v(schenk\_)*라는 통사정보를 지니고 있다. 그럼에도 불구하고 보편구조 (19b)로부터 독일어 문장의 리스트 표현인 (21b)가 생성될 때에 (24a)만이 선택되는 것은 독일어의 생성을 위한 한정절문법의 한정절 (25a)에 자질구조로 표현된 주어와 정동사간의 일치에 대한 제약때문이다.

(24) a.  $gv3(v(schenk_), pres, Fv) \rightarrow [schenkt], \{ Fv:phon = schenkt,$

$Fv:cat = v,$

$Fv:subc:first:agr:num = sg,$

$Fv:subc:first:agr:per = 3,$

$Fv:subc:first:case = nom,$

$Fv:subc:second:case = dat,$

$Fv:subc:third:case = acc \}$ .

b. gv3(v(schenk\_),pres,Fv) --> [schenkst], {Fv:phon == schenkst,  
 Fv:cat == v,  
 Fv:subc:first:agr:num == sg,  
 Fv:subc:first:agr:per == 2,  
 Fv:subc:first:case == nom,  
 Fv:subc:second:case == dat,  
 Fv:subc:third:case == acc}.

c. gv3(v(schenk\_),pres,Fv) --> [schenken], {Fv:phon == schenken,  
 Fv:cat == v,  
 Fv:subc:first:agr:num == pl,  
 Fv:subc:first:agr:per == 3,  
 Fv:subc:first:case == nom,  
 Fv:subc:second:case == dat,  
 Fv:subc:third:case == acc}.

d. gv3(v(schenk\_),pres,Fv) --> [schenke], {Fv:phon == schenke,  
 Fv:cat == v,  
 Fv:subc:first:agr:num == sg,  
 Fv:subc:first:agr:per == 1,  
 Fv:subc:first:case == nom,  
 Fv:subc:second:case == dat,  
 Fv:subc:third:case == acc}.

(25) a. gs(s(KP,T),Fs) --> gkp(KP,agent,Fkp), gtp(T,Ftp),

{Fs == Ftp, Fs:subc:first == Fkp}.

b. gtp(tp(VP,tense(T)),Ftp) --> gvp(VP,T,Fvp),

{Ftp == Fvp}.

- c. gvp(vp(DP1,DP2,V),T,Fvp) -->  
 gv3(V,T,Fv),gkp(DP1,goal,Fdp1), gkp(DP2,theme,Fdp2),  
 {Fvp == Fv, Fvp.subc:second == Fdp1,  
 Fvp.subc:third == Fdp2}.
- d. gd0(d(tonio),\_,Fd) --> [tonio],  
 {Fd.phon == tonio, Fd.cat == det,  
 Fd.agr:per == 3, Fd.agr:num == sg,  
 Fd.case == nom}.

주어와 시제구와의 결합에 관여하는 한정절 (25a)는 주어인 격구 *gkp*가 지니는 자질구조 *Fkp*와 시제구가 가진 자질구조 *Ftp*가 통합될 수 있어야 한다는, 곧 두 자질구조가 전달하는 정보간에 상충이 일어나지 않아야 한다는 제약을 표현하고 있다. PROLOG의 2항 술어 “==”은 두 자질구조의 통합여부를 체크한다. 예를 들어 어휘기재항 (25d)에 제시된 주어인 “tonio”의 자질구조에 따르면 주어의 인칭(*per*)은 3인칭이고 수(*num*)는 단수(*sg*)이다. 반면 어휘기재항 (24b)-(24d)가 지닌 자질구조에는 이와 다른 정보들이 들어 있기 때문에 정보간의 상충이 일어나 그것들의 선택이 이루어지지 않은 것이다. 이상이 보편구조로부터 독일어가 생성되는 과정에 대한 설명이다. 설명의 편의를 위해 너무 전문적인 프로그래밍에 관한 해설은 생략했다.

## 5. 맺는말

이 글은 격정보와 시제정보를 표현하는 방식이 국어와 독일어간에 차이가 난다는 관찰로부터 출발했다. 두 언어에서 관찰되는 이러한 대조적인 성격에도 불구하고 우리가 보편문법을 지향하는 원리와 매개변항의 연구결과를 수용하여 언어보편적인 구조를 상정함으로써 국어에서 독일어로의 기계번역이 가능하다는 것을 보였다. 여기서 소개한 기계번역시스템은 극히 제한된 통사구조의 국어와 독일어문장을 대상으로 하는 실

협적인 연구목적의 시스템이었다. 언어학적으로 문제가 덜한 문장구조들만을 다룬 것이 하나의 한계로 지적될 수 있겠다. 이 시스템을 위해 설정된 몇가지 언어학적인 가정들, 예컨대 “agent”, “goal”과 “theme” 등의 의미기능들간에 일정한 위계성이 성립한다는 가정과 국어에서 관사가 나타나지 않을 때에 일반규칙을 통해서 부정관사를 삽입할 수 있다는 가정 등은 좀더 구체적인 검토가 필요하다. 끝으로 대조연구의 한 응용분야이며 정보화시대의 총아인 기계번역시스템의 실용화가 성공하기 위해서는 음성/음운론, 형태/통사론, 의미/화용론과 사전학 등 여러방면에서의 국어와 독일어간의 심도있는 대조연구가 뒷받침되어야 한다는 점을 강조하고자 한다. 국어와 독일어에 관한 모든 자료중심의 대조연구가 초기단계에 있는 한-독 기계번역분야의 발전에 크게 기여할 것이다.

### 참고문헌

- Abney, St. 1987. The English Noun Phrase in its Sentential Aspect. Ph.D. Diss. MIT.
- Bußmann, H. 1990. Lexikon der Sprachwissenschaft. 2. Auflage. Stuttgart.
- Chomsky, N. 1986. Barriers. Cambridge(Mass.).
- Fanselow, G. 1986. Konfigurationslität-Untersuchungen zur Universalgrammatik am Beispiel des Deutschen. Tübingen.
- Fillmore, C. F. 1968. The Case for Case. In: E. Bach und R.T. Harms (Hrsg.). Universals in linguistic Theory. New York. 1-88.
- Gazdar, G. und C. Mellish. 1989. Natural Language Processing in PROLOG. Wokingham.
- Han, H.-S. 1987. The Configurational Structure of the Korean Language. Seoul.
- Kohrt, M. 1976. Koordinationsreduktion und Verbstellung in einer generativen Grammatik des Deutschen. Tübingen.
- Lee, M. 1992a. Kontrastive Syntax und maschinelle Sprachanalyse im Rahmen einer Unifikationsgrammatik -- Untersuchungen zum Deutschen und

Koreanischen.

Frankfurt a. M.

- Lee, M. 1992b. Configurationality Parameter in Korean and its Computational Implication—An HPSG Approach. In: C. Lee und B.-M. Kang (Hrsg.). Language, Information and Computation, Proceedings of Asian Conference, Seoul. 206-217.
- Lehner, C. 1990. Prolog und Linguistik. München.
- Olsen, S. 1989. Das Possessivum: Pronomen, Determinans oder Adjektiv?. In: Linguistische Berichte 120. 120-153.
- Parsons, T. 1990. Events in the semantics of English: A Study in subatomic semantics. Cambridge, M.A.
- Pereira, F.-C.N. und S.-M. Shieber. 1987. Prolog and Natural-language Analysis. CSLI Lecture Notes Nr. 10. Stanford University.
- Pollock, J.Y. 1989. Verb Movement, Universal Grammar, and the Structure of IP. In: Linguistic Inquiry 20. 364-424.
- Shieber, S.-M. 1987. Unification and Grammatical Theory. CSLI Report. Stanford University.
- Shin, H.-S. 1991. Kasus als Funktionale Kategorie. Dissertation. Universität Regensburg.
- Sommers, H.L. 1987. Valency and Case in Computational Linguistics. Edinburgh.
- Stechow, A.v. 1990. Morphologie und Syntax. Arbeitspapier Nr.16 der Fachgruppe Sprachwissenschaft. Universität Konstanz.
- 신수송. 1989. 독일어 통사구조의 기계적 분석을 위한 기초연구. 어학연구 Vol. 25.1.
- 정재현. 1990. 현대독일어 UND-등위접속구문에 대한 연구. 서울대 박사학위논문.

## Zusammenfassung

Kontrastive Syntax und Maschinelle Übersetzung Koreanisch-Deutsch unter besonderer Berücksichtigung der unterschiedlichen Ausdrucksweisen von Kasus- und Tempusinformationen

Min haeng Lee (Cheju Nt'l Univ.)

Koreanisch und Deutsch gehen andere Wege beim Enkodieren der Informationen bei Kasus und Tempus. Im Koreanischen werden die Informationen durch Hinzufügen der einschlägigen Morpheme an den Nomina und Verbstämmen ausgedrückt. Im Deutschen hingegen werden sie über Flexion zum Ausdruck gebracht. Damit trotz dieser Unterschiede eine maschinelle Übersetzung zwischen beiden Sprachen möglich ist, muß man die Forschungsergebnisse der Universalgrammatiker der Chomsky-Schule nutzbar machen. In dieser Abhandlung wird ein experimentelles Maschinelles Übersetzungssystem Koreanisch-Deutsch dargestellt, das als ein interlinguales Modell konzipiert ist und über eine sprachuniverselle syntaktische Struktur verfügt. Das Übersetzungssystem ist lediglich in der Lage, ein Fragment, nämlich die Grundkonstruktion des Koreanischen, ins Deutsche zu übersetzen und die Koordinationskonstruktion des Koreanischen in die interlinguale Repräsentation zu transferieren.

## 부록 Appendix

## %% Grammatik für das Koreanische %%

sc(S) ---> [smax(S), conj(C)]. /\* smax == Satz der  
Maximalprojektion \*/

smax(smax(S,mood(M))) ---> [s(S), mood(M)].

s(s(KP,T)) --->

[kp(KP,agent), tp(T,TNS)].

s(s(S1,C,S2)) ---> [s(S1),coord(C), s(S2)].

kp(kp(DP,case(KAS)),KAS) --->

[dp(DP),case(P,KAS)].

pp(pp(DP,P)) --->

[dp(DP),p(P)].

dp(dp(D)) ---> [npr(D)].

tp(tp(VP,tense(TNS)),TNS) ---> [vp(VP),i(T,TNS)].

tp(tp(TP,C,TP1),TNS) ---> [tp(TP,TNS),coord(C),tp(TP1,TNS)].

vp(vp(V1)) --->

[v1(V1)].

vp(vp(VP,C,VP1)) --->

[vp(VP),coord(C),vp(VP1)].

## %% Lexikon %%

word(senmwulha,v3(v(schenk\_))).

word(ca,v1(v(schlaf\_))).

word(po,v2(v(seh\_))).

word(coaha,v2(v(moeg\_))).

word(top,v2(v(helf\_))).

%% "Definite-Clause"-Grammatik für das Deutsche %%

%% -- mit eingebauten Merkmalsstrukturen %%

gsmax(smax(S,mood(M)),Fsmax) --> gs(S,Fs), gmood(M,Fmood),  
 {Fsmax == Fs, Fsmax:mood == Fmood}.

gs(s(KP,T),Fs) --> gkp(KP,agent,Fkp), gtp(T,Ftp),  
 {Fs == Ftp, Fs:subc:first == Fkp}.

gkp(kp(DP,case(KAS)),KAS,Fkp) --> gdp(DP,KAS,Fdp),  
 {Fkp == Fdp}.

gdp(dp(D),KAS,Fdp) --> gd0(D,KAS,Fd), {Fdp == Fd}.

gdp(dp(Det,N),KAS,Fdp) --> gdet(Det,KAS,Fdet), gcn(N,Fcn),  
 {Fdp == Fdet, Fdp:comp == Fcn}.

gdp(dp(Det,Nbar),KAS,Fdp) --> gdet(Det,KAS,Fdet), gnbar(Nbar,Fnbar),  
 {Fdp == Fdet, Fdp:comp == Fnbar}.

%% Lexikon %%

gmood(decl,Fmood) --> ['.'], {Fmood:phon == '.',  
 Fmod:cat == mood}.

gconj(conj(dass),Fc) --> [dass], {Fc:phon == dass,  
 Fc:cat == conj}.

gv1(v(schlaf\_),past,Fv) --> [schlafen], {Fv:phon == schlafen,  
 Fv:cat == v,  
 Fv:subc:first:agr:per == 3,  
 Fv:subc:first:agr:num == pl,  
 Fv:subc:second == end}.

gvc(v(glaub\_),pres,Fv) --> [glaubt], {Fv:phon == glaubt,  
 Fv:cat == v,



```

    Fv:subc:first:agr:num == sg,
    Fv:subc:first:agr:per == 3}.
X == Y :-
    denotes(X,A),
    denotes(Y,B),
    unify(A,B).
denotes(Var,Var):-
    var(Var),!.
denotes(Atom,Atom):-
    atomic(Atom),!.
denotes([HIR],[HIR):- !.
denotes(Dag:Path, Value):-
    pathval(Dag,Path,Value,_).

%% Testen %%

/* Test1: [tonioka,ingeeykey,chaykul,senmwulhanta] */

>>> K-D-Uebersetzung:

[tonio,schenkt,inge,ein,buch,.]

/* Test2: [tonioka,canta] */

>>> K-D-Uebersetzung:

[tonio,schlaeft,.]

```

/\* Test3: [tonioka,ingelul,coahanta] \*/

>>> K-D-Uebersetzung:

[tonio,mag,inge,.]

/\* Test4: [tonioka, ku, yeonkukul,coahanta] \*/

>>> K-D-Uebersetzung:

[tonio,mag,das,theater,.]

/\* Test5: [tonioka,heikeeykey,kkochul,senmwulhanta] \*/

>>> K-D-Uebersetzung:

[tonio,schenkt,heike,eine,blume,.]

/\* Test6: [tonioka,ingelul,topnunta] \*/

>>> K-D-Uebersetzung:

[tonio,hilft,inge,.]

/\* Test7: [tonionun,ingeka,ku,yeonkukul,coahantako,mitnunta] \*/

>>> K-D-Uebersetzung:

[tonio,glaubt,inge,mag,das,theater,...]

/\* Test8: [ingeka,toniolul,ttenako,censanhakul,kongpwuhanta] \*/

>>> Transfer zur interlingualen Struktur

=>>> Interlingua in P-A-Struktur

```

smax(s(kp(dp(d(inge)),case(agent)),tp(vp(vp(kp(dp(d(tonio)),case(theme)),v
(verlass_))
,coord(und),vp(kp(dp(det(ein_),n(informatik)),case(theme)),v(studier_))),tense
(pres))
),mood(decl))

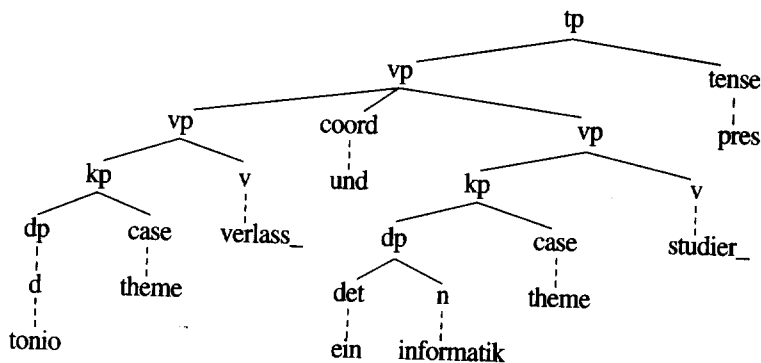
```

=>> Interlingua in C-Struktur

%% ein Teil davon

/\* Test9: [tonionun,ingeka,ku,yeonkukul,coahantako,mitnunta] \*/

>>> Transfer zur interlingualen Struktur



=>> Interlingua in P-A-Struktur

```

smax(s(kp(dp(d(tonio)),case(ag)),tp(vp(smax(s(kp(dp(d(inge)),case(agent)),tp
(vp(kp(dp(det(d_),n(theater)),case(theme)),v(moeg_)),tense(pres))),mood(decl)),v
(glaub_)),tense(pres))),mood(decl))

```

=>> Interlingua in C-Struktur

%% ein Teil davon

