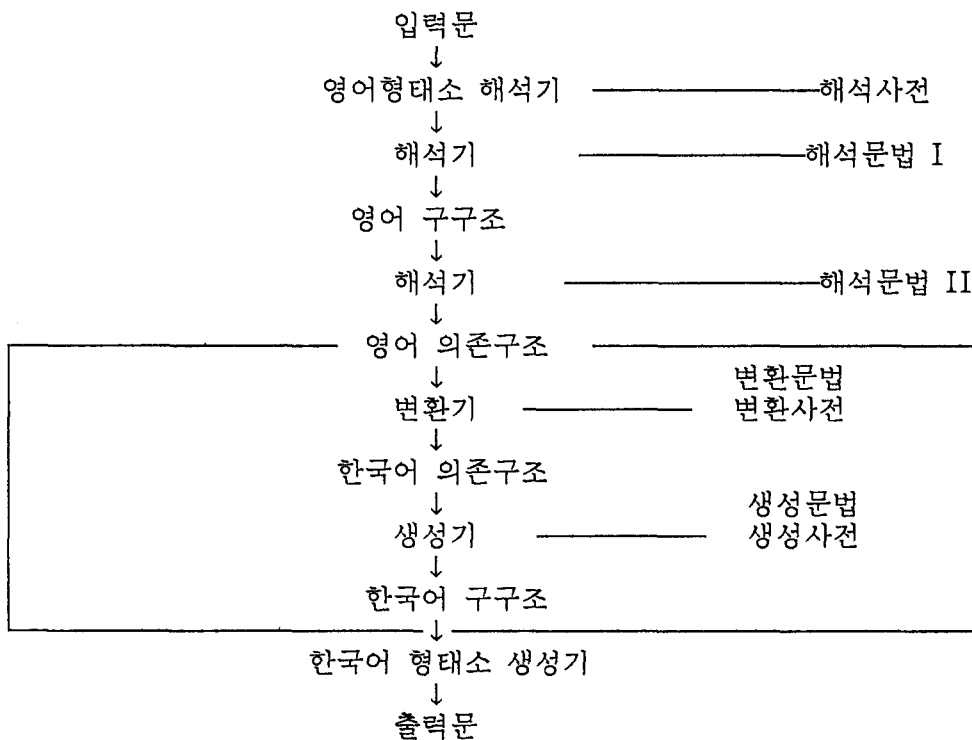


## 영-한 기계번역을 위한 영-한의존구조 대조분석에 관한 연구\*

신효필, 이호석, 장정일, 이정민  
(서울대학교 언어학과)

### 1. 서론

본 연구는 변환(Transfer)방식을 사용하는 시스템공학센터의 영한기계번역 시스템에 있어서 중간구조로 사용되는 의존구조를 체계화하고, 두 언어의 번역에 있어 효율적으로 사용될 수 있는 의존구조를 실제 예문을 중심으로 작성하는 것이 목표이다. 번역은 다음과 같은 과정을 통해 이루어 진다.



변환방식은 변환과정을 해석, 변환, 생성의 3단계로 구분한다. 해석의 결과가 변환단계에 입력이 되고, 변환의 결과가 생성단계의 입력이 된다. 여기서 각 단계 사이에 존재하는 것을 중간구조(intermediate structure)라고 하는데, 이 중간구조의 설정이 변환방식의 문법설계 및 작성에 중요한 열쇠가 된다. 본 연구에서 중점을 둔 부분은 다음과 같다;

\* 이 논문은 한국과학기술원 시스템공학센터의 동명의 위탁과제 중 일부이다.

- 영어의 특이성이 제거된 영어 의존구조의 작성
- 한국어로의 번역에 효율적인 정보를 제공할 수 있는 자질설정 (심층격, 표층격, 각 노드에 붙는 자질-값, 한국어 특유의 자질)
- 각 단계에 있어 효율적으로 사용될 수 있는 규칙설정.

## 2. 의존구조

자연언어를 통사적으로 나타내는 데는 구성문법(constituency grammar)과 의존문법(dependency grammar)이라는 두가지 접근 방식이 있다. 전자가 문장을 이루는 구성성분과 그 문장간의 관계를 계층적, 선형적으로 나타내는데 비해 후자는 문장의 주서술어를 중심으로 구성성분간의 의존관계만을 나타낼 뿐 순서는 고려하지 않는다는 점에 그 특징이 있다. 따라서 의존문법을 근간으로 하는 의존구조는 보다 효율적인 기계번역을 위한 중간구조로서 바람직하다고 볼 수 있다.

의존구조는 기술하는 사람에 따라 여러가지 모양을 취할 수 있지만, 번역에 있어서 그 효율성을 높이기 위해서 우선적으로 고려되어야 할 사항은 의존구조를 이루는 기본노드를 가장 합리적인 방법으로 설정하는 일이라고 할 수 있다. 본 연구는 영한 두 언어의 체계적이고 효율적인 의존구조를 설정하여, 번역에 있어 야기되는 두 언어의 차이를 최대한도로 조정할 수 있도록 하였다.

### 2.1. 의존구조를 구성하는 노드

- (1) PRED(icate) 노드; 동사, 형용사
- (2) NOM(inal Phrase) 노드; 명사, 대명사...
- (3) MOD(ifier) 노드; 형용사, 한정사, 소유대명사...
- (4) AUX(iliary) 노드; 조동사
- (5) ADV(erb) 노드; 부사
- (6) 전환(Translator)노드; NOM, MOD, ADV전환노드.

$$\begin{array}{ccc} \text{PRED} & \text{PRED} & \text{PRED} \\ \parallel & \parallel & \parallel \\ \text{PRED} & \text{PRED} & \text{PRED} \end{array}$$
 --> 관계절, 동명사, to-부정사, 종속절...

### 2.2. 의존구조의 자질-값(feature-value)

- 가) 각 노드에 붙는 공통적인 자질-값  
(E\_LEX ;  
(E\_CON ;
- 나) 전환노드에만 붙는 자질  
(E\_TRANS TYPE )
- 다) PRED노드에 붙는 자질-값  
(E\_LEX);  
(E\_PRED TYPE ADJ/LIK(likely류의 서술어)/SEM(seem, appear류의 동사)/\*VER  
(E\_TENSE \*PRES/PAST/FUTURE/  
(E\_ASPECT PROG/PERFECT  
(E\_SENTENCE\_TYPE \*DEC(DECLARATIVE)/IMP(IMPERATIVE)/INT(INTERROGATIVE)/  
ECL(EXCLAMATIVE)/SJT(SUBJUNCTIVE)/PPT(PROPOSITIV)/  
OPT(OPTATIVE)  
(E\_VOICE PASSIVE/\*ACTIVE;  
(E\_NEG NOT/hardly/scarcely/few/little/never/  
(E\_COMPA -er/-est/-less/-lest
- 라) NOM노드에 붙는 자질-값  
(E\_LEX ;

- (E\_NUM PLU/\*SINGULAR;  
(E\_DCASE\_ ; AGT, OBJ, GOA...(총 40가지)  
(E\_SCASE\_ ; SUBJ, SUBJ1, ATRSUB, OBJ, OBJ2, ATROBJ  
(E\_PREP\_ ; in, for, at, by...  
\* NOM으로 전환된 노드에 붙는 자질  
(E\_TRANS\_ INF/GER/BSE/PRP/PSP/that/whether/if/IQU(간접의문문)  
/QUA(인용문)
- 마) MOD노드에 붙는 자질-값  
(E\_LEX\_ ;  
(MOD\_TYPE\_ DEF /INDEF/INDEF1/INDEF2/INDEF3/\*ADJ/ADN/POS/NUM/DEM/EMP/  
NOM1/NOM2/GRP  
\*MOD로 전환된 유형이 갖는 자질  
(E\_TRANS\_TYPE REL; 관계절
- 바) ADV구에 붙는 자질-값  
(E\_LEX\_ ;  
\* ADV로 전환된 노드가 갖는 자질-값  
(E\_TRANS\_CONC(양보)/INF(목적의 to-부정사)/COMPA(비교급, -할 수록)/RESLT  
(to-부정사의 결과적 용법)/COL(부대상황)
- 사) AUX노드에 붙는 자질-값  
(E\_LEX\_ ;  
(AUX\_TYPE\_ ABL(ity)/OBL(igation)/NEC(essity)1/NEC2/INE(vitability)/  
PRF(preference)/GUE(ss)1/GUE2/WIL(1)/PMS(permission)/  
OPT(ative)/JUF(justification)/REQ(uest)/HAB(it)/WIS(h)/  
IRO(ny)/DAR(ing)/STA(state)/PLN(plan)/DES(tiny)/ECL
- 사) 한국어 의존구조에만 붙는 자질-값.  
1) PRED노드에 붙는 자질-값  
(K\_END\_ \*DE/IR/IM/PR/EX/OP/SJT  
2) 전환된 노드에 붙는 자질-값  
(K\_CJT\_ SU(가정)/AR(나열)/NE(필요)/CO(양보)/EC(감탄)/EX(설명)/IC(증가)  
IN(의도)/PU(목적)/CH(선택)/CA(원인)  
<K\_EMB(EMBedded)\_K1; -다"고"/K2; -냐"고"/K3; -라"고"/K4; -자"고"/k5; -"라고"  
\_G1; -것/G2; -것으로/G3; -것에/G4; -데에  
\_N1; -음/ㅁ을/N2; -기/N3; -기로/N4; -기에  
\_I1; -는지/I2; -는가/I3; -(느)냐

### 3. 영어의존구조

본 논의에서 가장 중점을 둔 부분중의 하나는 한국어 번역이 고려된 영어의존구조의 작성이다. 영어와 한국어는 구조의 큰 차이를 보이는 언어이기 때문에 어느 한 언어만을 완벽하게 의존구조로 표시하다 보면 실제 변환과정에서 많은 문제점과 어려움이 야기된다. 따라서 본 연구과제에서는 영어의 많은 현상들을 자질로 처리하였고, 그 자질이 한국어에서도 관여적인 경우에만 각 노드에 명시하는 방법을 취했다. 본 연구에서 고려한 사항 중 일부는 다음과 같다;

1. 'be'동사 다음의 형용사를 전 문장의 서술어로 채택.  
'be'동사 다음의 명사(구)는 'be'동사를 서술어로 삼음.
2. 특수표현의 처리
  - a. 허사 it/there --> 생략
  - b. 동사 + particle --> 한 노드로 처리
  - c. 명령문의처리 --> 주어복원하지 않음
3. 불필요한 어구의 생략

- a. 해석되지 않는 재귀대명사의 생략
- b. 해석되지 않는 4형식 문장의 간접목적어 생략.
- 예) I envy you your fine garden.
- 4. 사역구문 --> 부사구로 전환  
예) Please let me go.
- 5. 'have'의 처리 --> 'be'로 전환하여 처리
- 6. 전치사구의 처리 --> 역어가 존재하는 것만 명시.
- 7. 복수표현 --> 한국어에서 필요한 경우만 명시.
- 8. 관사의 처리 --> 필요한 경우만 자질로 명시.
- 9. 관계대명사의 처리 --> 계속적 용법은 접속사와 대명사로 환원.
- 10. 부정의 처리 --> 자질로 명시.

#### 4. 변환

앞장에서 제시된 영어의존구조를 한국어 의존구조로 번역하는 과정을 변환이라 한다. 변환은 크게 어휘형태를 한국어 어휘로 번역하는 어휘규칙과 각 노드에 붙는 자질-값을 번역하는 자질규칙으로 구성되어 있다. 본 연구에서는 이 둘을 변환규칙(Transfer rule)이라고 칭한다.

가) 어휘변환규칙; 영어어휘를 한국어 어휘로 번역.

나) 자질변환규칙; 영어의존구조의 자질을 한국어 자질로 변환

ex) (E\_DCASE\_ ) --> (K\_DCASE\_ )

1) 술어변환규칙(Pred Trans\_Rule); 7개의 PTR

2) 전환노드변환규칙(Trans\_Node\_Trans-Rule); 115개의 전환규칙.

TNTRG16;	NOM(E_TRANS_TYPE GER)	-->	NOM(K_TRANS_EMB I3)
	PRED(E_LEX_ )		PRED(K_LEX_ )
TNTRI1;	ADV(E_TRANS_TYPE INF)	-->	ADV(K_TRANS_CJT AR4)
	PRED(E_LEX_ )		PRED(K_LEX_ )
TNTRP2;	ADV(E_TRANS_TYPE PRP)	-->	ADV(K_TRANS_CJT AR1)
	PRED(E_LEX_ )		PRED(K_LEX_ )
	PRED(E_LEX_ )		PRED(K_LEX_ )
TNTRB2;	NOM(E_TRANS_TYPE BSE)	-->	NOM(K_TRANS_EMB K1)
	PRED(E_LEX_ )		PRED(K_LEX_ )
TNTRS2 ;	NOM(E_TRANS_TYPE PSP)	-->	NOM(K_TRANS_END AR4)
	PRED(E_LEX_ )		PRED(K_LEX_ )
	(E_TENSE PAST)		(K_TENSE PAST)
TNTRT1;	NOM(E_TRANS_TYPE THAT)	-->	NOM(K_TRANS_EMB K1)
	PRED(E_LEX_ )		PRED(K_LEX_ )
TNTRU1;	NOM(E_TRANS_TYPE QUA)	-->	NOM(K_TRANS-EMB K5)
	PRED(E_LEX_ )		PRED(K_LEX_ )
TNTRW1;	NOM(E_TRANS_TYPE WHETHER)	-->	NOM(K_TRANS-EMB I1)
	PRED(E_LEX_ )		PRED(K_LEX_ )

```

TNTRF1;  ADV(E_TRANS_TYPE  if)          ADV(K_TRANS_CJT  SU1)
          |
          | PRED(E_LEX_ )                --> PRED(K_LEX_ )
          |
TNTRQ1;  NOM(E_TRANS_TYPE  IQU)         NOM(K_TRANS_EMB  I1)
          |
          | PRED(E_LEX_ )                --> PRED(K_LEX_ )
          |
TNTRR1;  NOM(E_TRANS_TYPE  REL)         NOM(K_TRANS-EMB  G1)
          ||
          || PRED(E_LEX_ )                --> PRED(K_LEX_ )
          ||
3) 부정변환규칙(Negation_Pred_Trans_Rule); 10개의 NPTR
   NPTR1; PRED(E_LEX_ ) --> PRED(K_LEX_ -)
          (E_NEG not)      (K_NEG1a)
4) 변환규칙을 적용받지 않는 자질-값
   AUX(E_LEX_ ), (AUX_TYPE_ ), (E_PREP_ ), (MOD_TYPE_)

```

변환시 영어의존구조의 자질-값은 위와 같은 변환규칙에 의해 해당 한국어 자질-값으로 변환된다. 한국어 의존구조는 영어의 특이성이 제거된 영어의존구조를 그대로 복사하여 사용하고, 그 자질-값도 변환규칙에 의해 한국어 해당 자질-값을 갖는 것 그리고 영어의존구조의 자질-값이 그대로 내재되어 있는 것등이 섞여 있는 혼합구조라고 할 수 있다.

## 5. 한국어 생성

입력된 영어문장이 형태소 분석을 거쳐 영어의존구조로 표현되고 다시 변환을 거쳐 한국어 의존구조로 표시된 후 이 의존구조는 한국어 표면문장으로 전환 되는데 이 과정을 생성(generation)이라고 한다. 한국어 생성에는 한국어 의존 구조를 토대로 조사를 결합하고, 부정표현, 조동사를 서술어와 결합시키는등 적합한 표면문장을 생성하는 일련의 규칙들로 이루어 진다. 생성에 관여하는 규칙을 생성규칙(Generation Rule)이라고 한다.

### 5.1. 생성규칙

#### 가) 술어화규칙(Predicate-Generation\_Rule)

```

PGR1;
  PRED (K_LEX_ 이다)
  |
  | NOM (K_LEX A1) --> 'A1 + 이다'
  | (K_SCASE ATRSUB)

```

```

PGR2;
  PRED (K_LEX_ 이다)
  |
  | NOM (K_LEX A1) --> 'A1 + 이다'
  | || (K_SCASE ATRSUB)
  | PRED(K_LEX_ )

```

#### 나) 문장종결어미 생성규칙(Sentence\_Ending\_Generation\_Rule); 총 7개

```

SEGR2; (K_SENT TYPE IR1/2/3) --> '-(는)가/-냐/-(으)가/-(습)니까'

```

#### 다) 조동사 결합규칙(Auxiliary\_Combining\_Generation\_Rule); 총 23개

```

ACGR1; PRED (K_LEX_ A1)
        |
        | AUX (E_LEX_ ) --> 'A1 + -리 수 있다'
        | (AUX_TYPE ABL)

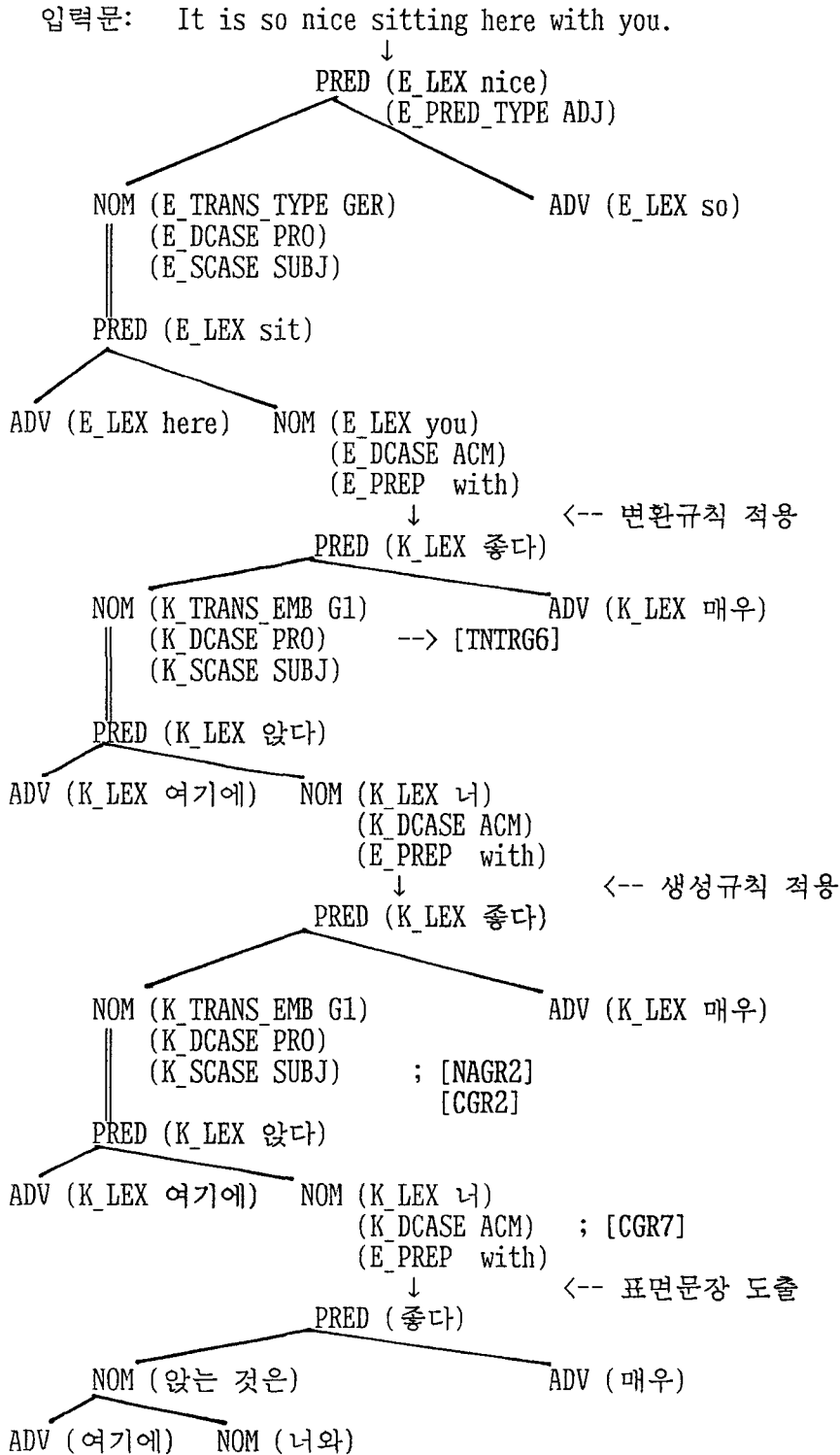
```

- ACGR3: PRED (K\_LEX A1) --> 'A1 + -(어)야만 하다'  
 |  
 AUX (E\_LEX )  
 (AUX\_TYPE OBL)
- ACGR9: PRED (K\_LEX A1) --> 'A1 + -임에 틀림없다'  
 |  
 AUX (E\_LEX )  
 (AUX\_TYPE GUE1)
- 라) 시제-상 결합규칙(Tense\_Aspect\_Combining\_Generation\_Rule); 총 13개
- TACGR3: (K\_LEX A1) (K\_TENSE FUTURE) --> 'A1 + -리 것이다'  
 (K\_ASPECT NIL)
- TACGR4: (K\_LEX A1) (K\_TENSE NIL) --> '-고 있다'  
 (K\_ASPECT PROG)
- TACGR5: (K\_LEX A1) (K\_TENSE NIL) --> '-었다'  
 (K\_ASPECT PERF1)
- 마) 비교급 결합규칙(Comparative\_Combining\_Generation\_Rule); 총 4개
- CCGR1: PRED (K\_LEX A1) --> '더 A1 -다'  
 (K\_COMPA -er)
- 바) 조사 생성 규칙(Casemarker\_Generation\_Rule); 총 26개
- CGR1: NOM (K\_LEX A1) ---> 'A1 + 이/가/은/는'  
 (K\_SCASE SUBJ)
- CGR3: NOM (K\_LEX A1) ---> 'A1 + 을/를'  
 (K\_SCASE OBJ)
- CGR8: NOM (K\_LEX A1) ---> 'A1 + 에/에서'  
 (K\_DCASE LOC)  
 (E\_PREP in/on/at)
- 사) MOD생성규칙(Modify\_Generation\_Rule); 총 6개
- MGR1: MOD (K\_LEX A1) --> 'A1 + -의'  
 (MOD\_TYPE POS)
- 아) NOM노드 관형화 규칙(NOM-node\_Adnominal\_Generation\_Rule); 총 4개
- NAGR2: NOM (K\_TRANS\_TYPE G1/G2/G3/G4) --> 'A1 + (으)ㄴ 것'  
 ||  
 PRED (K\_LEX A1)  
 (K\_TENSE PAST)
- 자) 접속사 결합규칙(Conjunct\_Combining\_Generation\_Rule); 총 9개
- CCGR3: PRED (K\_CON and) --> 'A1 + 그러면 + A2'  
 PRED (K\_LEX A1) PRED (K\_LEX A2)  
 (K\_SENT\_TYPE IM)
- 차) 복수표지 결합규칙(Plural-marker\_Combining\_Generation\_Rule)
- PCGR1: NOM (K\_LEX A1) --> 'A1 + 들'  
 (K\_NUMBER PLU)

## 5.2. 생성

5.1의 생성규칙이 영어의존구조에서 변환된 한국어 의존구조에 적용되어 한국어 문장을 도출하게 된다. 생성시에는 보통 의존구조를 구구조로 전환하는 방식을 사용한다. 그러나 본 연구에서는 생성시 한국어 의존구조를 구구조로

전환하는 방식을 채택하지 않고 의존구조에서 곧장 한국어 표면문장을 생성하는 방식을 취한다. 지금까지 논의한 사항을 번역과정을 통해 개략적으로 살펴보면 다음과 같다;



## 6. 결론

지금까지 변환방식을 채택하고 있는 기계번역시스템에 있어서 중간구조로 사용되는 영,한 의존구조에 대하여 논의하였다. 본 과제에서는 변환의 출발점이 되는 영어의존구조를 한국어 번역에 적합한 구조로 만드는데 중점을 두고 영어의 특이현상들을 한국어 표현에 맞도록 중화시키는 입장을 취하였다. 이러한 방식으로 체계화된 영어의존구조는 변환규칙을 거쳐 한국어 의존구조로 변환되는데, 본 논의에서는 구조는 그대로 복사하고 의존구조 각 노드에 사용되었던 자질-값들은 번역, 변환하는 방식을 취했다. 한국어 의존구조는 다시 생성규칙에 의해 한국어 표면문장으로 전환되는데 본 논의에서는 한국어 구구조를 생성하지 않고 의존구조에서 곧장 표면구조를 도출하는 방식을 채택하였다.

## 참고문헌

1. 김정수(1990) <영한 기계번역을 위한 영어 구문해석시스템의 설계 및 구현>, 한국과학기술원 전산학과 석사학위 논문.
2. 남기심, 고영근(1985), <표준국어 문법론>, 탑출판사.
3. 송성문(1990), <성문종합영어>, 성문출판사.
4. 신기철, 신용철(1975), <새우리말 큰사전>, 삼성출판사.
5. 신호필(1990), <HPSG를 기초로 한 한국어 동사의 하위범주화>, 서울대학교 언어학과 석사학위 논문.
6. 이정민(1975), "국어의 보문화에 대하여", <어학연구> 11-2.
7. 임홍빈(1988), "국어의 명사구 의미역에 대하여", 원고.
8. 장석진(1989), <한영동사의 하위범주화와 대응에 관한 연구>, 한국전자통신연구소.
9. 장석진(1989), <자연언어처리의 기초연구>, 한국과학재단.
10. 정희성(1989), <한영 기계번역 시스템(I)에 관한 연구>, 과학기술처.
11. 최현배(1977), <우리말본(제6판)>, 정음사.
12. 한국과학기술원 시스템공학센터(1989), <기계번역을 위한 언어 모델링 및 골격 시스템 설계> (연구보고서), 과학기술처.
13. Hornby, A.S.(1975), Guide to Pattern and Usage in English, Oxford University Press.
14. Pollard, C. and I. Sag (1987), Information-based Syntax and Semantics vol. 1, Lecture notes No.12, CSLI.
15. Quirk, Randolph and Sidney Greenbaum and et al.(1985), A Comprehensive Grammar of the English Language, Longman.