

# 어휘함수를 이용한 노-한 기계번역 시스템의 다의어 제어 변환규칙

김 성 래  
(한국외대 강사)

## 1. 도입

'정보'와 '네트워크'으로 표현되는 '지식,정보기반 사회'에서 정보를 구성하고 나타내는 자연어를 컴퓨터를 이용하여 처리하는 언어공학(language engineering) 분야는 급속하게 발전할 것이다. 특히 시스템 구성에서 전산언어학(computational linguistics)의 비중이 높은 기계번역(machine translation)에 대한 수요는 사이버 공간의 정보량이 증가할수록 높아질 전망이다.

현재 개발중인 노-한 기계번역(Russian-to-Korean Machine Translation : RKMT) 시스템은 한국외국어대학교, KAIST 그리고 러시아 과학원 산하 정보이체 문제 연구소(Institute for Information Transmission Problems: IITP)가 공동으로 연구를 진행하고 있다. 이 시스템은 과학-기술 텍스트의 번역을 목표로 삼고 있으며, IITP의 ETAP(ElectroTexničeskij Avtomatičeskij Perevod)을 이용하여 지금까지 약 100여 개의 문장을 실험적으로 번역하였다.

이 글은 기계번역 시스템에서 항상 난제로 여겨지고 있는 어휘 모호성(lexical ambiguity)의 처리, 특히 다의어(polysemy)의 번역문제를 다루고자 한다. 먼저 RKMT 시스템의 전반적인 구성과 다의어 번역의 문제점을 살펴보고, 이론 틀로 삼은 어휘함수(Lexical Function)를 관찰할 것이다. 그 다음 시스템에 실제로 입력한 문장에서 나타나는 다의어들을 효율적으로 제어할 수 있는 규칙과 사전구성을 설명할 것이다.

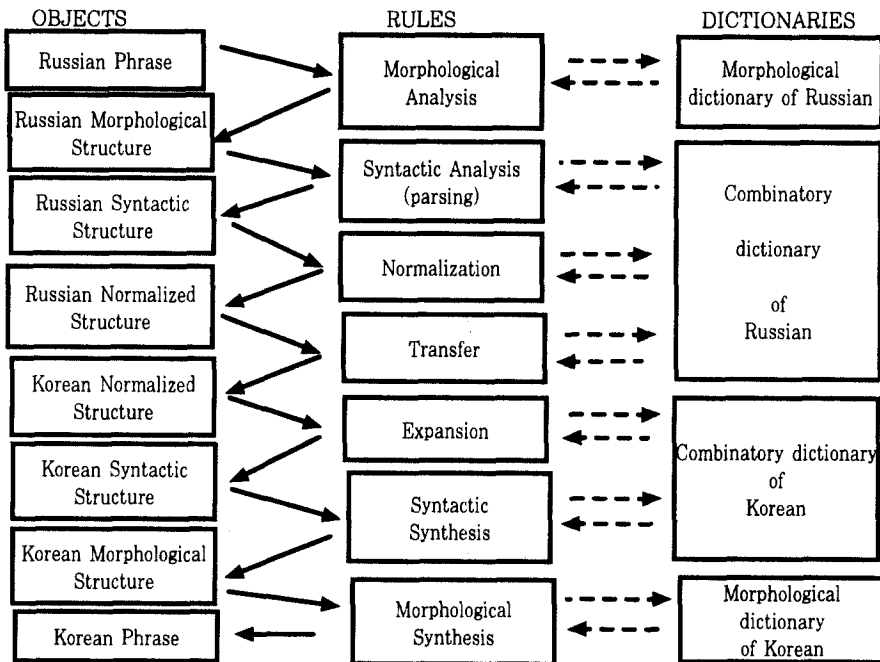
## 2. RKMT의 구성과 어휘 모호성

### 2.1 RKMT의 전체적인 구성

언어학적 규칙에 기반을 두고 있는 RKMT 시스템은 변환(transfer) 방식을 취하고 있으며, 시스템의 전체적인 구성은 노어분석, 노어문장 구조의 한국어로의 변환, 한국어 생성으로 구분할 수 있다. 노어 분석 과정은 노어의 형태구조, 통사구조, 표준 통사구조를 확립하는 것을 목표로 하고 있으며, 변환 과정은 노어의 표준 통사구조를 한국어 표준 통사구조로 바꾸는 역할을 한다. 번역의 최종단계인 한국어 생성과정은 한국어 통사구조와 형태구조를 확립하여 한국어 문장을 만들어 낸다. 이러한 번역 과정은 멜추(Mel'čuk)의 '의미-텍스트' 이론(Meaning-Text Theory)에 입각한 것이다.

다음 도표 [1]은 RKMT의 전체 구성을 보여 준다.

도표 [1] General Layout of RKMT Module



위의 전체적인 구성에서 이 글이 다루고자 하는 것은 변환단계이다. 변환단계에서 사용되는 규칙은 노어 표준 통사구조를 구성하는 모든 노어 어휘 및 자질 등을 한국어 어휘 및 자질로 번역해야만 한다. 즉 변환단계가 실행된 후에 확립된 한국어 표준 통사구조에는 단 하나의 노어 어휘나 형태자질도 남아 있지 않게 된다. 따라서 변환과정이 실행된 이후에는 입력어인 노어와는 상관없이 즉각 한국어 생성과정을 실행할 준비가 되는 것이다.

## 2.2 변환단계의 어휘 모호성

변환단계는 노어 표준 통사구조를 한국어 표준 통사구조로 바꾸는 과정이다. 여기서 '표준 통사구조(Normalized Syntactic Structure)'란 의미-텍스트 이론의 총 7개 언어표현 단계에서 '심층통사표현(Deep-Syntactic Representation)' 단계에 해당하는 것으로 개별언어의 특징적인 문법현상은 제거하고 의미 표현에 중점을 둔 것이다. 변환단계에서 변환시켜야 할 대상은 일반적으로 형태자질(morphological feature), 통사관계(syntactic relation) 그리고 어휘소(lexeme)이다.

형태자질 변환은 노어 어휘의 형태자질을 상응하는 한국어 형태자질로 교체하는 것이다.

(1) ПРЕЗИДЕНТ им, ед, муж, од ⇒ 대통령 sg

(1)에서 муж(남성), од(활동체)가 삭제된 것은 한국어 문법에 성과 활동성 범주가 없기 때문이고, 격 범주 자질인 им(주격)은 한국어 생성과정에서 한국어 통사구조에 알맞게 바뀌어야 하기 때문에 삭제되었다. 결국 ед(단수)만이 한국어 형태자질 sg(단수)로 직접 교체된다.

통사관계 변환은 노어 통사관계를 한국어 통사관계로 바꾸는 것인데 때로는 구문 구조를 바꿔야 하는 경우도 있다.



위 예는 노어 통사관계인 수식관계(опредилительное отношение)가 한국어 통사관계인 한정관계(attributive relation)로 교체되는 것을 보여 준다.

마지막으로 어휘변환은 노어 어휘소를 노어 결합사전에 나타난 한국어 역어로 바꾸는 것이다. 이 어휘변환의 문제점은 다의어와 동형이의어(homography)의 변환 과정에서 발생하는 모호성이다.<sup>1)</sup> 동형이의어는 통사분석(parsing) 단계에서 결합사전에 들어있는 해당 어휘의 통사자질(syntactic feature), 의미자질(semantic feature) 그리고 지배유형(government pattern) 등의 정보를 이용하여 그 모호성이 해결되는 경우가 대부분이기 때문에 변환단계에서는 그 처리과정이 쉽다. 하지만 다의어의 경우에는 해당 어휘가 어떤 다른 어휘와 결합하느냐에 따라서 그 처리과정이 복잡해질 가능성이 높다. 따라서 다의어의 모호성 해결도 연어관계(collocation relation)에서 그 실마리를 찾는 경우가 대부분이다. (옥철영, 김영택 1993)

연어관계의 유형은 'V+N', 'V+ADV', 'A+N', 'ADV+A' 등으로 다양하며 (Wanner, L. (ed.), 1996:15-22), 각 언어마다 다양한 어휘로 나타난다. 다음은 노어에서 'V+N' 유형으로 표현되면서, 동사가 다양한 한국어 역어로 번역되는 경우를 보여준다.

- |                          |                    |
|--------------------------|--------------------|
| (3) пользоваться случаем | (기회를 <u>이용하다</u> ) |
| пользоваться правами     | (권리를 <u>갖다</u> )   |
| пользоваться уважением   | (존경을 <u>받다</u> )   |
| оказать помощь           | (도움을 <u>주다</u> )   |
| оказать сопротивление    | (저항을 <u>나타내다</u> ) |
| оказать давление         | (압력을 <u>가하다</u> )  |

1) 기계번역에서 발생하는 모호성은 어휘 모호성(lexical ambiguity), 통사 모호성(syntactic ambiguity), 공지시 모호성(coreferential ambiguity) 등이 있다.

'A+N' 유형에서도 영어의 'heavy rain', 'strong tea'와 같이 각 명사에 대한 연어가 다의어가 되는 경우가 많다. 다음 (4)는 노어의 'A+N' 구조 연어관계에서 노어 형용사를 기본적인 의미에 대응하는 한국어 형용사로 번역할 때 발생할 수 있는 모호성을 나타내는 예이다.

(4) жидкие аплодисменты	(* <u>뭉은</u> 박수)
круглый дурак	(* <u>둥근</u> 바보)
горький пьяница	(* <u>쓴</u> 술꾼)
тъма крошечная	(* <u>비참한</u> 어둠)

위와 같이 연어관계에서 나타나는 다의어 문제를 해결하기 위해 RKMT 시스템은 어휘함수(Lexical Function)를 변환규칙에 도입하였다. 다음 절에서는 어휘함수의 정의와 종류를 설명하고, 활용되는 분야를 보일 것이다.

### 3. 의미-텍스트 이론 과 어휘함수

#### 3.1 의미-텍스트 이론 개괄

1960년대 후반 구소련의 언어학자인 멜추크(Мельчук)은 의존 문법(Dependency grammar)에 기반을 둔 '의미-텍스트' 이론('Meaning-Text' Theory)을 발표하였다. 이 이론에서 '의미(meaning)'는 화자(speaker)에 의해 전달되는 정보 또는 내용(information or content)의 무한집합을 나타내며, 텍스트(text)는 수신자(addressee)에 의해 인지되는 언어학적 형태(forms)의 집합을 나타낸다. 멜추크는 언어를 의미 집합과 텍스트 집합 사이의 다대다 대응(correspondence) 혹은 사상(mapping)으로 이루어진 것으로 파악하고 다음과 같은 도식으로 나타내고 있다.

language proper

$$(5) \{MEANING_i\} \longleftrightarrow \{TEXT_j\} \mid 0 < i, j \leq \infty$$

위의 도식 (5)에 의하면 자연어는 모든 가능한 의미의 무한집합과 모든 가능한 텍스트의 무한집합의 대응(그리고 그 반대의 대응)을 성립시키는 논리적 장치라고 생각된다.

의미와 텍스트의 상징적 표현은 의미표현(Semantic Representation : SemR)과 음성표현(Phonetic Representation : PhonR)으로 나타내며 두 표현 사이에 통사표현(Syntactic Representation : SyntR), 형태표현(Morphological Representation : MorphR)의 중간 단계를 포함시켜 다음과 같이 기술할 수 있다.

$$(6) \{SemR\} \leftrightarrow \{SyntR\} \leftrightarrow \{MorphR\} \leftrightarrow \{PhonR\}$$

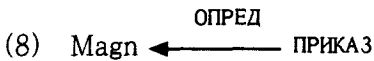
위에서 의미표현(SemR)은 제멋대로인 그물망(network)으로, 통사표현(SyntR)은 나무(tree) 구조로, 형태표현(MorphR)과 음성표현(PhonR)은 선형적인 사슬(chain)로 표현된다. 멜축은 이러한 4가지 표현 중에서 의미표현을 제외한 나머지 표현을 의미에 중점을 둔 것들은 심층적인 것(deep, D~)으로, 형태에 중점을 둔 것은 표층적인 것(surface, S~)으로 구분하여 다음과 같이 총 7단계의 성층적인 모델(stratificational model)을 제시한다.

(7) 의미-텍스트 이론의 표현 단계

1. 의미표현 (Semantic Representation : SemR), 혹은 의미(meaning)
2. 심층통사표현 (Deep-Syntactic Representation : DSyntR)
3. 표층통사표현 (Surface-Syntactic Representation : SSyntR)
4. 심층형태표현 (Deep-Morphological Representation : DMorphR)
5. 표층형태표현 (Surface-Morphological Representation : SMorphR)
6. 심층음성표현 (Deep-Phonetic Representation : DphonR ; 즉, 음소표현)

### 7. 표층음성표현 (Surface-Phonetic Representation : SphonR ; 음성표현 혹은 텍스트)

어휘함수는 위 7단계의 표현 중 심층통사표현을 이루는 심층통사구조 (Deep-Syntactic Structure)의 구성 성분에 속한다. '정도의 심함' 혹은 '매우 집중력이 있는 정도'를 나타내는 어휘함수 Magn은 심층통사구조에서 다음과 같이 나타날 수 있다.



위의 (8)에서 어휘소(lexeme) 'ПРИКАЗ'와 어휘함수 'Magn'은 각각 노드를 구성하며 그 노드를 방향성 있는 호(arc), 즉 화살표가 연결하고 있다. 화살표 방향은 지배와 종속관계를 표시하며 화살표 위의 'ОПРЕД'은 수식관계라는 통사관계를 나타낸다.

발화자의 입장에서 러시아어 어휘 ПРИКАЗ(명령)의 '정도의 심함'을 나타내려면 'СТРОГИЙ'로 혹은 'КАТЕГОРИЧЕСКИЙ'와 같은 구체적인 어휘들로 어휘함수 Magn 노드를 대치할 것이다.

### 3.2 어휘함수의 정의 및 종류

어휘함수는 특정 어휘 표현이 해당 어휘에 특별히 의존적일 때 어휘들 간의 규칙적인 관계를 기술하기 위해 고안된 것이다. '의미-텍스트' 이론에서 어휘함수는 문장의 심층통사구조(Deep-Syntactic Structure)와 해석결합사전(Explanatory-Combinatorial Dictionary)에서 특정 어휘의 어휘관계(lexical relations)를 나타내고 있다. 어휘함수의 쓰임새는 다음 (9)와 같이 수학적 사용과 동일하다.

$$(9) f(L) = \{Li\}$$

위 식에서  $f$ 는 어휘함수를 나타내고  $L$ 은 해당 어휘를 나타내는 변수(argument 혹은 keyword)이다. 집합인  $\{L_i\}$ 는 어휘 함수  $f$ 의 값(value)에 해당하는 어휘 표현을 의미한다.(Mel'čuk, 1996)

구체적으로 어휘함수 **Bon**을 예로 들어보자. 어휘함수 **Bon**은 라틴어 *bonus*(‘좋은’)로부터 그 이름을 따온 것으로 해당 어휘의 ‘좋은’ 혹은 ‘긍정적인’ 평가를 표현한다. 다음은 어휘함수 **Bon**에 실제 어휘들을 대입한 예이다.

- (10) **Bon** (aid) = valuable  
**Bon** (behavior) = excellent  
**Bon** (атмосфера) = приятная  
**Bon** (влияние) = благотворное

어휘함수는 동의어, 반의어, 파생어 등과 같이 해당 어휘를 대치하는 계열적(paradigmatic) 혹은 대치(substitute) 어휘함수와 해당 어휘와 다른 어휘간의 관계를 나타내는 통합적(syntagmatic) 혹은 매개적(parametric) 어휘함수의 두 가지 종류로 구분할 수 있다.<sup>2)</sup> 계열적 어휘함수와 통합적 어휘함수의 일부를 실제 어휘들을 대입하여 예시한다. 여기서  $L$ 은 해당 어휘(lexeme)를 나타낸다.

(11) 계열적 어휘함수의 예

- 1) **Syn** [Lat. *synonymum*] - 동의어  
**Syn** (to telephone) = to phone  
**Syn** (бросать) = кидать  
**Syn** (лингвистика) = языкознание
- 2) **Anti** [Lat. *antonymum*] - 반의어  
**Anti** (victory) = defeat  
**Anti** (друг) = враг  
**Anti** (первый) = последний

2) 어휘함수의 종류와 분류에 관해서는 Мельчук(1974), Mel'čuk(1996), Ramos & Tutin (1996), Steel & Meyer(1990)을 참고하시오.



- 3) **Gener** [Lat. genus] - L의 총괄 명칭  
**Gener** (carrots) = vegetables  
**Gener** (радость) = чувство / эмоция  
**Gener** (республика) = государство
- 4) **S0** - 동일 의미의 명사  
**S0** (to analyze) = analysis  
**S0** (защищать) = защита  
**S0** (лгать) = ложь
- 5) **S1** - L을 행하는 행위자(Agent) 혹은 대상  
**S1** (to teach) = teacher  
**S1** (бороться) = борец  
**S1** (лекция) = лектор

(12) 통합적 어휘함수의 예

- 1) **Magn** [Lat. magnus] - '매우', '정도의 심함'  
**Magn** (to laugh) = heartily  
**Magn** (кричать) = громко, во всё горло  
**Magn** (дурак) = круглый
- 2) **Bon** [Lat. bonus] - '좋은'  
**Bon** (analysis) = fruitful  
**Bon** (случайность) = счастливая  
**Bon** (отель) = первоклассный / пятизвездочный
- 3) **Ver** [Lat. verus 'real, genuine'] - '그래야만 되는, 의도되는 요구에 부합하는'  
**Ver** (punishment) = well-deserved  
**Ver** (приказ) = законный  
**Ver** (документ) = подлинный
- 4) **Oper1** [Lat. operari 'to do, carry out'] - L을 행하는 어휘  
**Oper1** (investigation) = conduct  
**Oper1** (приказ) = дать

**Oper1** (помощь) = оказать

5) **Func0** [Lat. \*functionare 'to function'] - L의 기능이 수행됨

**Func0** (snow) = falls

**Func0** (война) = идти

**Func0** (ночь) = стоять

멜축은 위와 같은 표준 어휘함수 약 60 여개를 제시하였다. 표준 어휘함수가 다른 어휘함수 등과 결합하여 다음과 같은 복합(complex) 어휘함수로 사용될 수 있다.<sup>3)</sup>

(13) **AntiMagn** (temperature) = low

**AntiMagn** (приказ) = мягкий

**AntiBon** (атмосфера) = зловещий

어휘함수는 80년대 중반이후 전산언어학과 자연어처리 분야에서 광범위하게 응용되고 있다. Апресян, Ю.Д., Цинман, Л.Л. (1998)은 어휘함수를 이용한 35개의 구문전환 공식을 작성하여 구문 바꿔쓰기(paraphrasing) 시스템을 도입하였고,<sup>4)</sup> MS ACCESS를 이용하여 노어 어휘 학습 프로그램을 개발하였다. Lee, W., Evens, M.(1996)는 응급실에서 사용되는 의료용 전문가 시스템(Medical Expert System)의 텍스트 생성기에 어휘함수를 도입하였다.<sup>5)</sup> 또한 Iordanskaja도 텍스트 생성과 구문 바꿔쓰기에 어휘함수를 이용하였다.

3) 아프레산(Ю. Д. Апресян)은 복합 어휘함수를 포함한 약 100 여개의 어휘함수를 사전구성에 사용하고 있다.(Апресян, Ю.Д., Цинман, Л.Л. 1998)

4) 다음 공식과 예를 참고하십시오.

$$\text{Oper1}(S0(X)) + S0(X) \Leftrightarrow X$$

дать приказ  $\Leftrightarrow$  приказать

$$\text{Oper2}(S0(X)) + S0(X) \Leftrightarrow X$$

Город подвергается обстрелу.  $\Leftrightarrow$  Город обстреливают.

5) 이 시스템에서는 Gener(Aspirin)=drug, S0AntiAble1(to talk)=aphasia와 같은 일반적인 어휘함수와 함께 avg\_oral\_dose(Aspirin)=325mg 과 같은 의료용 어휘함수를 도입하였다.

위에서 예시된 어휘함수에서 다의어 번역에 주로 이용되는 것은 통합적 어휘함수이다. 4절에서는 변환규칙과 사전 그리고 RKMT에 입력한 문장들을 살펴보겠다.

## 4. 어휘함수를 이용한 다의어 번역

### 4.1 변환규칙과 해석결합사전

변환단계에서 사용되는 규칙들은 형태자질 번역규칙, 통사관계 번역규칙, 어휘 번역규칙이다. 다의어 번역을 위해서 지금까지 사용된 대부분의 어휘 번역규칙은 다의어에 해당하는 어휘의 사전에 기입되는 사전규칙(dictionary rule)이었다.<sup>6)</sup> 예를 들어 ГОРЬКИЙ란 어휘는 일반적으로 ‘쓰다’라고 번역하지만 ПЬЯНИЦА를 수식할 때는 ‘지독하다’로 번역하라는 규칙을 ГОРЬКИЙ의 사전항목에 기입하는 것이다. 그러나 다의어에 해당하는 모든 어휘의 사전에 규칙을 작성해야 하는 이러한 방법은 비효율적이다. 따라서 다의어 처리를 위한 어휘변환 규칙은 일반규칙이어야 하며, 이 규칙이 실행되기 위해서는 어휘함수가 각 어휘의 결합사전 항목에 기입되어야한다.

다의어의 번역을 위한 규칙은 어휘함수 번역규칙으로 존재한다. 어휘함수 번역규칙은 먼저 노어 어휘와 한국어 어휘가 동일한 어휘함수를 갖는 지를 확인하여 두 어휘가 동일한 어휘함수를 갖는다면 노어 어휘의 어휘함수 값(value)을 한국어 어휘함수의 값으로 교체하는 것이다. 다음은 어휘함수 번역규칙을 보여준다.

#### (14) 어휘함수 번역규칙

국지변항 선언:RALFA

RALFA: \_ANTIREAL1/\_ANTIREAL1-M/..... (어휘함수 목록)

CHECK

X가 노어 어휘이고, X의 사전항목에 RALFA에 있는 어휘함수가 존재한다면

- 
- 6) 규칙은 그것이 적용되는 범위에 따라서 다음과 같이 세가지로 구분된다.
1. 일반규칙(general rule): 번역하는 모든 문장에 적용되는 규칙
  2. 특수규칙(specific rule): 일부 구문에만 적용되는 규칙
  3. 사전규칙(dictionary rule): 특정 단어에만 적용되는 규칙

N:01

CHECK

1.1 X와 Z이 S,V,A,ADV중 하나이고, Z에 대응하는 X의 어휘함수가 RALFA에 있고, Z의 어휘함수가 RALFA에 존재한다면

1.2 Z가 한국어 어휘 중에 하나라면

DO

1 X의 어휘함수 값을 Z의 어휘함수 값으로 교체하라

N:02

CHECK

1.1 X와 Z이 S, V, A, ADV 중 하나이고, Z에 대응하는 X의 어휘함수가 RALFA에 있고,

1.2 Z의 어휘함수가 RALFA에 존재하지 않는다면

DO

1 어휘함수에 의한 번역을 포기하라

이러한 규칙이 시스템에 적용될 때는 형식어(formal language)를 사용하는 형식기법(formalism)이 사용된다.<sup>7)</sup> 규칙 (14)은 시스템에서 다음과 같은 형식어로 표시된다.

(15) REG:RA-TRADUCT.56

LOC:RALFA

RALFA: ANTIREAL1/\_ANTIREAL1-M/\_ANTIREAL2-M/\_CAUSFACT0/\_CAUSFACT0-M/ \*  
 \_CAUSFUNC0/\_FINOPER1/\_FINOPER2/\_FINREAL1/\_FINREAL1-M/\_INCEPFUNC1/ \*  
 \_INCEPOPER1/\_INCEPOPER2/\_INCEPREAL1/\_INCEPREAL2/ \*  
 \_LIQUFACT0/\_LIQUFACT0-M/\_LIQUFUNC0/\_MANIF/\_OPER1/\_OPER2/\_PREPAR/ \*  
 \_REAL1/\_REAL1-M/\_REAL2/\_REAL2-M/\_CAUSFACT1/\_CAUSFACT1-M/\_CAUSFACT2/ \*  
 \_CAUSFACT2-M/\_CAUSFUNC1/\_CAUSFUNC2/\_LIQUFACT1/\_LIQUFACT1-M/ \*  
 \_LIQUFUNC1/\_LIQUFUNC2/ \*  
 \_CAUSOPER1/\_CAUSOPER2/\_CAUSREAL1/\_CAUSREAL2/\_INCEPLABOR1-2/ \*  
 \_LABOR1-2/\_LIQUOPER1/\_LIQUOPER2/\_LIQUIREAL1/\_LIQUIREAL1-M/\_LABREAL1-2/ \*

7) 형식어는 용어(terms)와 조건 검토에 사용되는 술어(predication) 그리고 명령실행에 사용되는 명령어(instruction)로 이루어지며, 언어학적 분석을 자유롭게 표현할 수 있도록 체계적으로 구성되어 있다. 규칙의 구성방법과 규칙에 사용된 형식어에 대해서는 Апресян и др. (1989,1992)를 참고하라.

```

_DEGRAD/_FACT0/_FACT0-M/_FACT1/_FACT1-M/_FACT2/      *
_FINFACT0/_FINFUNC0/_FINFUNC1/_FUNC0/_FUNC1/_FUNC2/_INCEPFAC0/  *
_INCEPFUNCO/_INCEPFUNC1/_SON/_ANTIFACT0-M/              *
_MAGN/_ANTIMAGN/_BON/_ANTIBON/_VER/_ANTIVER/           *
_CAP/_EQUIP/_MULT/_SING/_GENER
CHECK
1.1 LEXR(X)&COLF(X,* .RALFA,*)
N:01
CHECK
1.1 M-EQUN(X,Z,* .S,V,A,ADV)&COLF(X,Z,RALFA,*)&LF(Z,RALFA)
1.2 LEXA(Z)
DO
1 ZAMAUZLF:X(Z,RALFA)
N:02
CHECK
1.1 M-EQUN(X,Z,* .S,V,A,ADV)&COLF(X,Z,RALFA,*)
2.2 LF(Z,RALFA)
DO
1 STERALT:
    
```

위의 규칙에 사용되는 해석결합사전은 품사영역(POR), 통사자질영역(SYNT), 의미자질영역(DES), 지배모델영역(Di,j), 번역어 영역(TRANS)으로 구성되어 있다. 다음은 노어 'ПРИКАЗ'와 한국어 '명령'의 사전내용을 나타내고 있다.

(16) 'ПРИКАЗ'의 사전항목

```

POR:S
SYNT:МУЖСК,ИСЧИСЛ
DES:'ДЕЙСТВИЕ','ФАКТ','ИНФОРМАЦИЯ','АБСТРАКТ'
D1.1:РОД,'ЛИЦО'
D2.1:ИНФ
D2.2:02
D3.1:ДАТ,'ЛИЦО'
V0:ПРИКАЗЫВАТЬ
SYN:РАСПОРЯЖЕНИЕ
SYN1:ПОВЕЛЕНИЕ
    
```

MAGN:СТРОГИЙ/КАТЕГОРИЧЕСКИЙ  
ANTIMAGN:МЯГКИЙ  
VER:ЗАКОННЫЙ  
ANTIVER:НЕЗАКОННЫЙ  
OPER1:ДАВАТЬ/ОТДАТЬ  
OPER3:ПОЛУЧАТЬ  
FUNC0:ПОСТУПАТЬ1  
FUNC1:ПОСТУПАТЬ1<ОТ>  
LIQUFUNC0:ОТМЕНЯТЬ  
REAL2-М:ВЫПОЛНЯТЬ/ПОДЧИНЯЙТЕСЯ/ПОВИНОВАТЬСЯ  
ZONE:K  
TRANS: 명령

(17) '명령'의 사전항목

POR:S  
DES:'ДЕЙСТВИЕ', 'ФАКТ', 'ИНФОРМАЦИЯ', 'АБСТРАКТ'  
D1.1:GEN1,'лицо'  
D2.1:raneun-form  
D2.2:대한  
SYN:지시  
MAGN:엄격하다/단호하다  
ANTIMAGN:느슨하다  
VER:당연하다  
ANTIVER:부당하다  
OPER1:내리다  
OPER3:받다  
FUNC0:시행되다  
LIQUFUNC0:취소하다  
REAL2-М:실행하다/수행하다  
ZONE:R  
TRANS:ПРИКАЗ

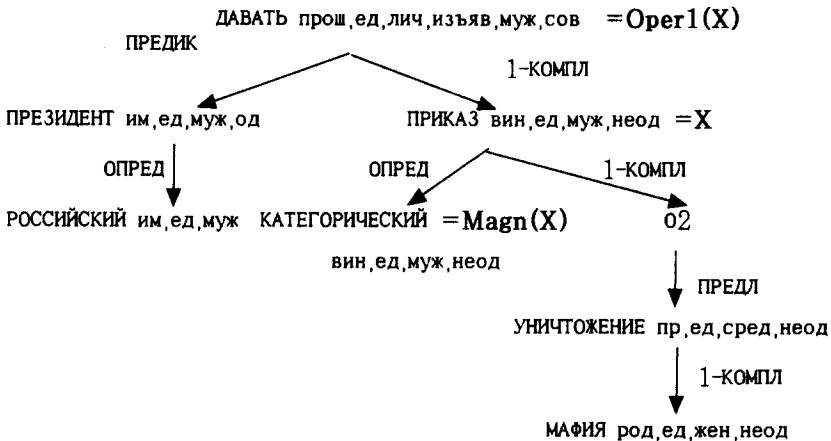
## 4.2 다의어 번역 예문8)

실험 예문은 이전에 RKMT에서 해결하지 못한 다의어들이 포함된 것으로 선정하였다. 다음 (18)은 어휘함수 Oper1과 Magn을 이용하여 번역한 예문으로 번역전의 노어 통사구조와 번역후의 한국어 통사구조를 함께 보여주고 있다.

(18) 1) Российский президент дал категорический приказ об уничтожении мафии.

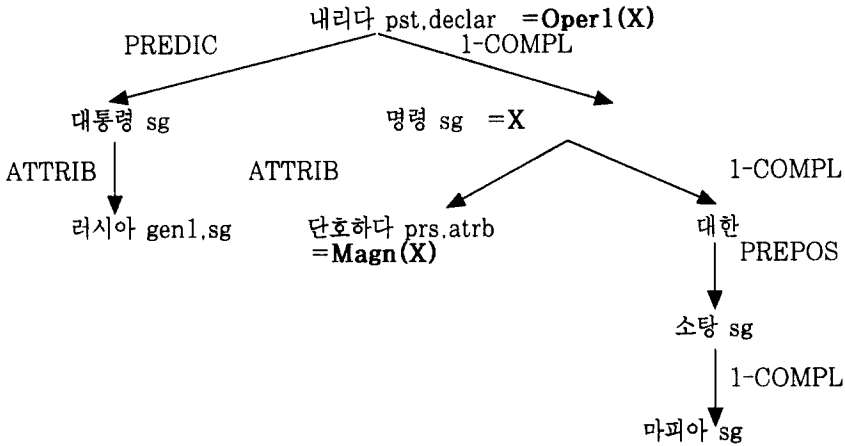
(러시아의 대통령은 마피아의 소탕에 대한 단호한 명령을 내렸다.)

### 2) 노어의 통사구조



8) 번역 실험은 IITP의 ETAP 시스템으로 실행되었고, 하드웨어는 MicroVax 3100을 OS는 Unix를 사용하였다.

3) 한국어 통사구조



이 예문에서 어휘함수 Magn을 이용하여 'КАТЕГОРИЧЕСКИЙ'를 '범주적이다' 대신 '단호하다'로 번역하였고 Oper1을 사용하여 'ДАВАТЬ'를 '주다'가 아닌 '내리다'로 번역할 수 있었다.

다음 (19)의 예문들은 번역에 성공한 예문 중 일부를 사용된 어휘함수와 함께 나타낸 것이다.

(19) 번역예문

- 1) Профессор прочитал студентам лекцию о машинном переводе.  
(교수는 대학생들에게 기계번역에 대한 강의를 하였다.)  
Oper1(лекция)=читать, Oper1(강의)=하다
- 2) Наш профессор пользуется глубоким уважением студентов.  
(우리들의 교수는 대학생들의 깊은 존경을 받고 있다.)  
Oper2(уважение)=пользоваться, Magn(уважение)=глубокий  
Oper2(존경)=받다, Magn(존경)=깊다
- 3) Он бросил университет.  
(그는 대학교를 중퇴하였다.)



- FinReall(университет) = бросать, FinReall(대학교) = 중퇴하다
- 4) Московские студенты наградили нового профессора жидкими аплодисментами.  
 (모스크바의 대학생들은 새로운 교수를 가벼운 박수로 맞이하였다.)  
 Labor1(аплодисменты) = награждать,  
 AntiMagn(аплодисменты) = жидкий  
 Labor1(박수) = 맞이하다, AntiMagn(박수) = 가볍다
- 5) Туристы нагуляли страшный аппетит.  
 (관광객들은 왕성한 식욕이 났다.)  
 IncepOper1(аппетит) = нагуливать, Magn(аппетит) = страшный  
 IncepOper1(식욕) = 나다, Magn(식욕) = 왕성하다
- 6) Идёт спор о жизни между философами.  
 (삶에 대한 논쟁이 철학자들 사이에 진행되고 있다.)  
 Func0(спор) = идти, Func0(논쟁) = 진행된다.

## 5. 결론

어휘합수는 멜추크의 이론에서 가장 세련되게 다듬어진 분야이기 때문에 최근 전산 언어학에서 연구가 활발히 진행되고 있다. 이 글에서는 다의어의 모호성을 해결하기 위해 RKMT에 도입하여 부분적으로나마 중간언어(interlingua)방식을 시스템에 도입할 수 있었다. 그러나 광범위한 해석결합사전 구축을 위해서는 많은 시간이 필요로 하고, 다양한 어휘합수를 능숙하게 다루기 위해서는 사용자들의 인내심 있는 노력이 요구된다는 점은 여전히 숙제로 남아있다. 그러나 Mel'čuk, I.A.(1996:96)이 준숙어적 표현(Semi-phasemes)이라고 부른 제약된 어휘관계를 어휘합수만큼 체계적으로 나타낼 수 있는 다른 대안은 아직 제시되지 않고 있다.

어휘합수를 적용시킬 수 분야는 아직도 남았다. 정보검색(information retrieval)용 사전에서 어휘합수를 이용한다면 보다 정교한 검색결과를 기대할 수 있을 것이고, 기계번역의 통사분석과정에서도 어휘합수가 효과적으로 작용할 가능성이 있다.

### 참고문헌

- 강덕수, 김성래, 김연수, 박창호 (1995), "러한 기계번역 기술개발", 전자도우미 시스템 개발, 과학기술처, pp.607-735.
- \_\_\_\_\_ (1996), 러시아어-한국어 기계번역 시스템(2), 과학기술처.
- 강덕수, 김성래, 김연수, 최기선 (1999), "노-한 기계번역에서 한국어 용언의 활용어미 합성모델 연구", 언어 24권 3호, 한국언어학회, pp.311-334.
- 강덕수, 최기선, 김연수, 김성래 (1998), "러시아어 텍스트의 기계번역을 위한 형태적 분석", 노어노문학 10권2호, 한국노어노문학회, pp.3-30.
- 옥철영, 김영택 (1993), "연어에 기반한 최상의 번역 선택", 한국정보과학회 논문집 20권4호, PP.561-571.
- Apresjan, J.D., Mel'čuk, I.A., Žolkovskij, A.K. (1973), "Materials for Explanatory Combinatory Dictionary of Modern Russian", in: Kiefer, F (ed.), *Trends in Soviet Theoretical Linguistics*, Foundations of Language Supplementary Series, Vol.18, Dordrecht: Reidel, pp.411-438.
- Apresjan, J.D., Boguslavskij, I.M., Iomdin, L.L., Lazurskij, A.V., Sannikov, V.Z., Tsinman, L.L. (1992), "ETAP-2: The Linguistics of A Machine Translation System", *META*, Vol.37, No1, Montréal: Les Press de l'Université de Montréal, pp.99-112.
- Boguslavskij, I.M. (1995), "A Bi-directional Russian-to-English Machine Translation System (ETAP-3)", *MT summit V*, July 10-13, Luxembourg.
- Boyer, M., Lapalme, G. (1985), "Generating Sentences from Semantic Networks", in: Dahl V. and Saint-Dizier P. (eds.), *Natural*

- Language Understanding and Logic Programming: First International Conference on Natural Language Understanding and Logic Programming, Rennes, France, Amsterdam: North-Holland, pp.181-189.*
- Evens, M.W. (ed.) (1988), *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*, Cambridge: Cambridge University Press.
- Grimes, J.E. (1996), "Lexical functions Across Language", in: Wanner, L. (ed.) (1996), pp.103-114.
- Heid, U. (1996), "Using Lexical Functions for the Extraction of Collocations from dictionaries and Corpora", in: Wanner, L. (ed.) (1996), pp.115-146.
- Heylen, D. (1995), "Lexical Functions, Generative Lexicons and the World", in: *Computational Lexical Semantics*, Saint-Disier, P., E. Viegas (eds), Cambridge: Cambridge University Press, pp.125-140.
- Iordanskaja, L., Arbatchewsky-Jumarie, N. (1992), "Lexicographic applications of lexical functions: two sample lexical entries from an explanatory-combinatorial dictionary", in: *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Department of Slavic Languages and Literatures, University of California, pp.364-372.
- Iordanskaja, L., Kim, M., Polguère, A. (1996), "Some Procedural Problems in the Implementation of Lexical Functions for Text Generation", in: Wanner, L. (ed.) (1996), pp.279-298.
- Iordanskaja, L., R. Kittredge, A. Polguère (1991), "Selection and Paraphrase in a Meaning-Text Generation Model", *Natural Language Generation in Artificial Intelligence and*

- Computational Intelligence*, Paris, C.L., W.R. Swartout, W.C. Mann (eds), Dordrecht: Kluwer Academic Publishers, pp.293-312.
- Iordanskaja, L., Paperno, S. (1996), *A Russian-English Collocational Dictionary of the Human Body*, Columbus: Slavica Publishers.
- Lee, W., Evens, M. (1996), "Generating Cohesive Text Using Lexical Functions", in: Wanner, L. (ed.) (1996), pp.299-306.
- Mel'čuk, I.A. (1979), *Studies in Dependency Syntax*, Ann Arbor: Karoma.
- \_\_\_\_\_ (1981), "Meaning-Text models: a recent trend in Soviet Linguistics", *Annual Review of Anthropology* 10, pp.27-62.
- \_\_\_\_\_ (1982), "Lexical functions in lexicographic description", in: *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Department of Slavic Languages and Literatures, University of California, pp.427-444.
- \_\_\_\_\_ (1988a), *Dependency Syntax: Theory and Practics*, Albany, N.Y.: SUNY press.
- \_\_\_\_\_ (1988b), "The Explanatory Combinatorial Dictionary", in: Evens, M.W. (ed.) (1988), pp.41-71.
- \_\_\_\_\_ (1996), "Lexical function: A tool for Description of Lexical Relations in a Lexicon", in: Wanner, L. (ed.) (1996), pp.37-102.
- Mel'čuk, I.A., A. Polguère. (1988), "A formal lexicon in Meaning-Text Theory (or How to do lexica with words)", *Computational Linguistics* 13(3-4), pp.261-275.
- \_\_\_\_\_ (1988), "Semantic description of lexical units in an Explanatory Combinatorial Dictionary: Basic principles and heuristic criteria", *International Journal of Lexicography*,

Vol.1, 3, pp.165-188.

Mel'čuk, I.A., Žolkovskij, A.K. (1988), "The Explanatory Combinatorial Dictionary", in: Evens, M.W. (ed.) (1988), pp.41-74.

Nakhimovsky, A.D. (1990), "Word Meaning and Syntactic Structure: Some Comparative Notes", in: Steele, J. (ed.) (1990), Ottawa: University of Ottawa Press, pp.3-17.

Ramos, M.A., Tutin, A. (1996), "A Classification and Description of Lexical Functions for the Analysis of their Combination", in: Wanner, L. (ed.) (1996), pp.147-168.

Ramos, M.A., Tutin, A., Lapalme, G. (1995), "Lexical functions of the *Explanatory Combinatorial Dictionary* for lexicalization in text generations", Saint-Dizier, P. & E. Viegas, (eds.) (1995), *Computational Lexical Semantics*, Cambridge : Cambridge University Press, pp.351-366.

Reuther, T. (1996), "On Dictionary Entries of Support Verbs: The Cases of Russian VESTI, PROVODIT' and PROIZVODIT'", in: Wanner, L. (ed.) (1996), pp.181-208.

Steele, J. (ed.) (1990), *Meaning-Text Theory: Linguistics, Lexicography, and Implications*, Ottawa: University of Ottawa Press.

Steele, J., Meyer, I. (1990), "Lexical Functions in an Explanatory Combinatorial Dictionary: Kinds, Descriptions, and English Examples", Steele, J. (ed.) (1990), Ottawa: University of Ottawa Press, pp.41-61.

Wanner, L. (ed.) (1996), *Lexical functions in lexicography and Natural language processing*, Amsterdam/Philadelphia: John Benjamins Publishing Company.

Апресян, Ю.Д. (1991), "Словарная статья глагола *рисовать*", *Семиотика и*

- информатика, Вып.32, Москва: ВИНТИ, с.34-41.
- \_\_\_\_\_ (1995), *Избранные трубы. Т.2, Интегральное описание языка и системная лексикография*, Москва: Школа «Языки русской культуры».
- Апресян, Ю.Д., Богуславский, И.М., Иомдин, Л.Л., Лазурский, А.В., Перцов, Н.В., Санников, В.З., Цинман, Л.Л. (1989), *Лингвистическое обеспечение системы ЭТАП-2*, Москва: Наука.
- Апресян, Ю.Д., Богуславский, И.М., Иомдин, Л.Л., Лазурский, Митюшин, Л.Г., Санников, В.З., Цинман, Л.Л. (1992), *Лингвистический процессор для сложных информационных систем*, Москва: Наука.
- Апресян, Ю.Д., Цинман, Л.Л. (1998), "Перифразирование на компьютере", *Семиотика и информатика*, Вып.36, с.177-202.
- Жолковский, А.К., Мельчук, И.А. (1967), "О семантическом синтезе", *Проблемы кибернатики*, Вып.19, с.177-238.
- Иомдин, Л.Л. (1990), *Автоматическая обработка текста на естественном языке : модель согласования*, Москва: Наука.
- Мельчук, И.А. (1974), *Опыт теории лингвистических моделей 'Смысл ⇔ Текст', Семантика, синтаксис*, Москва: Наука.
- \_\_\_\_\_ (1995), *Русский язык в модели «Смысл⇔Текст»*, Wien: Wiener Slawistischer Almanach.
- Мельчук, И.А., Жолковский, А.К. (1984), *Толково-комбинаторный словарь, Опыты семантико-синтаксического описания русской лексики*, Wien: Wiener Slawistischer Almanach.

## **A Transfer Rule based on Lexical Function in R-K MT System**

The purpose of this study is to make a transfer rule for the disambiguation of Russian polysemy in Russian-Korean Machine Translation System. This paper introduces the Lexical Function which was proposed by I.A. Mel'čuk and explains definitely the restricted lexical collocations to determine lexical ambiguity.

This study begins with a brief review of general layout of R-K MT Module and shows lexical ambiguity in the transfer level. Second, this paper classifies the lexical functions and gives some examples. Third, the paper observes the construction of the transfer rule based on lexical function and explains the structure of Russian and Korean Explanatory Combinatory Dictionaries which contain the lexical functions.

Finally I observed the result of experiments. The experiments were carried out with ETAP system in IITP. There are some sentences which were successfully translated with the help of lexical function in R-K MT system.