

러시아어 텍스트의 기계번역을 위한 형태적 분석*

- ETAP 시스템을 모델로

강 덕 수(한국외대 교수)

최 기 선(과기원 전산과 교수)

김 연 수(한국외대 강사)

김 성 래(한국외대 강사)

1. 서론

Hutchins(1986:324)은 기계 번역을 "a marriage of practical needs and theoretical idealism"이라고 한 바 있다. 이것은 기계 번역에 대한 연구가 실용적 목적에 의해 시작되었지만 그 바탕은 이론적 이상주의임을 보여준다. 그 이상은 자연어의 구조를 완벽하게 언어학적으로 분석할 수 있다는 가능성을 전제로 한 것이라고 할 수 있다. 기계 번역을 위한 자연어에 대한 언어학적 접근은 형태론, 통사론, 의미론 영역을 중심으로 이루어지고 있다. 자연어 분석 방법으로서 멜추크(Мельчук)은 "Meaning \Leftrightarrow Text"모형을 제안하였다.

$$\begin{aligned} & \text{language proper} \\ & \{ \text{MEANING}_i \} \Leftrightarrow \{ \text{TEXT}_j \} \mid 0 < i, j \leq \infty \end{aligned}$$

LANGUAGE

(Мельчук 1988:44)

이 모델 이론은 자연어를 의미와 텍스트 사이의 대응 관계를 구축하는 논리적 장

* 본 논문은 학술진흥재단 연구비 지원(1997-1998)으로 연구된 것임.

치로 간주하고 있다. (“...a natural language is viewed as a logical device that establishes the correspondence between the infinite set of all possible meanings and the infinite set of all possible texts and vice versa.”)(ibid.)

이 모델에 의하면 언어는 7개의 영역으로 나누어진다.

1. 의미 영역(Semantic Representation), 즉 의미
 2. 심층 통사 영역(Deep-Syntactic Representation)
 3. 표층 통사 영역(Surface-Syntactic Representation)
 4. 심층 형태 영역(Deep-Morphological Representation)
 5. 표층 형태 영역(Surface-Morphological Representation)
 6. 심층 음운 영역(Deep-Phonological Representation)
 7. 표층 음성 영역(Surface-Phonetic Representation), 즉 텍스트
- (op. cit. 49)

기계 번역에서는 음운 영역과 음성 영역은 고려할 필요가 없으므로 기계 번역을 위한 자연어 처리는 입력어의 형태부에 대한 분석으로부터 연구가 시작된다.

1950년 대부터 시작된 러시아어의 기계 번역 연구는 현재 러.영 번역 시스템을 상당한 수준까지 개발하였다. 특히 아프레산(Апресян)과 보구슬랍스키(Богуславский), 욘딘(Июдин) 등을 주축으로 하는 ИИП(Institute for Information Transmission Problems)는 러시아어 기계 번역 시스템인 ЕТАР을 개발한 바 있다. 본 논문은 ЕТАР 시스템의 형태부 분석과 기술 방법을 고찰하고, 이것이 자연어로서의 러시아어 문법 형태부 기술과 어떻게 다른지 살펴 보고자 한다. 이를 통해 형태부와 관련하여 자연어 처리를 위한 효율적인 분석 개념과 방법을 제시하고자 한다. 더 나아가 이의 효용성을 검증하고 번역 프로그램을 개발하기 위한 기반을 구축하기 위해 순수한 형태부 분석의 예를 들고 형태소 사전과 형태 사전 구축을 위한 모델을 제시해 보고자 한다.

2. 형태부의 특성

러시아어는 형태부(morphological component)가 차지하는 비중이 크다. 따라서 기계번역에서도 형태부에 대한 분석은 정확하게 신속한 번역을 위해 중요한 과정으로 간주된다.

우선 러시아어에서는 성과 수와 격에 의한 구분이 복잡하다. 성은 남성, 여성, 중성으로 성에 따라 달라지는 굴절어미가 우선 인지되어야 한다. 수는 단수와 복수 둘 뿐이어서 단순히 보이자 단수형으로만 쓰이거나 복수형으로만 쓰이는 명사들의 경우에는 별도로 고려되어야 한다. 또한 단수 남성의 경우에는 대격형에서 “사람”의 개념이 독립적인 범주로 설정되어야 한다. 동시에 복수에서는 활동체와 불활동체 범주가 구분되어야 한다.

격은 학교문법에서 주격, 생격, 여격, 대격, 조격, 전치격의 6개격만 인정하고, 부분격과 처소격은 특수용법으로 간주한다. 그러나 기계번역에서는 нет сахара, в лесу와 같은 부분격과 처소격을 기존의 6격과 동등한 격으로 인정하여 모두 8격으로 다룬다.

이러한 성, 수, 격, 활동성의 개념은 또한 품사별로 적용되어야 한다.

명사의 변화형은 남성, 여성, 중성의 성에 대한 구분과 단수, 복수 그리고 주격, 생격, 여격, 대격, 조격, 전치격의 형태가 하나의 기본적인 패러다임으로 구성된다. 기계번역을 위한 분석에서는 이러한 패러다임이 먼저 고려되어야 한다.

명사만 예로 든다 해도 Zaliznjak(1980 39-55)에 의하면 남성 명사 7 개 유형(завод, житель, чайник, марш, месяц, герой, сценарий), 여성 명사 8 개 유형(карта, неделя, книга, туча, улица, статуя, линия, тетрадь), 중성 명사 8 개 유형(болото, горе, благо, жилище, солнце, здание, имя) 외에도 빈도수가 크지 않은 많은 예외들이 있다. ЕТАР에서는 이것들을 다시 압축하여 어간 끝 자음에 따라 모두 12 개의 유형으로 분류하였다. 남성형 4 개(завод, житель, герой, чайник), 여성형 5 개(карта, неделя, линия, книга, тетрадь), 중성형 3 개(болото, здание, пальто)로 표준화 되었다. 예외적인 것들은 은폐형에 의해서 실현되거나 사전 항목에 수록된다.

〈명사어미 목록〉

No	단 수					복 수					예		
	ИМ РОД	ДАТ	ВИН	ТВОР	ПРЕД	ИМ РОД	ДАТ	ВИН	ТВОР	ПРЕД			
1	А	Ы	Е	У	ОЙ	Е	Ы	#	АМ	Ы	АМИ	АХ	вод - А
2	Я	И	Е	Ю	ЕЙ	Е	И	Ь	ЯМ	И	ЯМИ	ЯХ	пул - Я
3	Я	И	И	Ю	ЕЙ	И	И	Й	ЯМ	И	ЯМИ	ЯХ	лини - Я
4	А	И	Е	У	ОЙ	Е	И	#	АМ	И	АМИ	АХ	книг-А, рук-А
5	Ь	И	И	Ь	ЬЮ	И	И	ЕЙ	ЯМ	И	ЯМИ	ЯХ	часть - Ъ
6	#	А	У	#	ОМ	Е	Ы	ОВ	АМ	Ы	АМИ	АХ	ротор - #
7	Ь	Я	Ю	Ь	ЕМ	Е	И	ЕЙ	ЯМ	И	ЯМИ	ЯХ	кул - Ъ
8	Й	Я	Ю	Й	ЕМ	И	И	ЕВ	ЯМ	И	ЯМИ	ЯХ	жреби - Й
9	#	А	У	#	ОМ	Е	И	ОВ	АМ	И	АМИ	АХ	сток - #
10	О	А	У	О	ОМ	Е	А	#	АМ	А	АМИ	АХ	лиц-О, войск-О
11	Е	Я	Ю	Е	ЕМ	И	Я	Й	ЯМ	Я	ЯМИ	ЯХ	мнени - Е
12	#	#	#	#	#	#	#	#	#	#	#	#	пальто - #

그러나 러시아어에는 이러한 패러다임을 벗어나는 많은 이형태들이 존재한다. 복수 주격에서 учителя, братья, друзья, сыновья와 같은 형태들, 복수 생격에서 учителей, братьев, +сестёр, студенток, товарищей와 같은 이형태들, 복수 조격에서 людьми, лошадьми, детьми와 같은 이형태들이 고려되어야 한다.

형용사의 경우, Zaliznjak은 9개의 패러다임(예 : новый, синий, редкий, свежий, куций, длинношей, живой, сухой, большой)을 기본형으로 제시하고 있다 (Zaliznjak 1980:56). 이것을 성에 따라 구분하면 27개의 패러다임이 만들어진 다. 복수에서는 성의 구분이 없으므로 단수의 남성, 중성, 여성 처럼 하나의 독립된 패러다임으로 인정하면 모두 36개로 늘어난다. 이것을 ЕТАР 시스템에서는 단수 남성형 5개(частый, хороший, тонкий, пустой, большой, лапласов), 여성형 4개(большая, хорошая, синяя, лапласова), 중성형 4개(чёрное, тонкое, синее, лапласово), 복수형 3개(простые, синие, лапласовы) 모두 16개의 패러다임으로 압축하였다. 이와 함께 술어형으로 쓰이는 단어미형을 위한 패러다임을 3개 더 추가 하였다.

〈형용사 및 형동사 어미목록〉

No.	단 수						예	
	ИМ	РОД	ДАТ	ВИН	ВИН	ТВОР		ПР
				неод	од			
	남 성							
13	ЫЙ	ОГО	ОМУ	ЫЙ	ОГО	ЫМ	ОМ	час-ЫЙ
14	ИЙ	ЕГО	ЕМУ	ИЙ	ЕГО	ИМ	ЕМ	хорош-ИЙ
15	ИЙ	ОГО	ОМУ	ИЙ	ОГО	ЫМ	ОМ	тонк-ИЙ
16	ОЙ	ОГО	ОМУ	ОЙ	ОГО	ЫМ	ОМ	пуст-ОЙ
17	ОЙ	ОГО	ОМУ	ОЙ	ОГО	ИМ	ОМ	больш-ОЙ
18	#	А	У	#	А	ЫМ	ОМ	лапласов-#
	여 성							
19	АЯ	ОЙ	ОЙ	УЮ	ОЙ	ОЙ		больш-АЯ
20	АЯ	ЕЙ	ЕЙ	УЮ	ЕЙ	ЕЙ		хорош-АЯ
21	ЯЯ	ЕЙ	ЕЙ	ЮЮ	ЕЙ	ЕЙ		син-ЯЯ
22	А	ОЙ	ОЙ	У	ОЙ	ОЙ		лапласов-А
	중 성							
23	ОЕ	ОГО	ОМУ	ОЕ		ЫМ	ОМ	чёрн-ОЕ
24	ОЕ	ОГО	ОМУ	ОЕ		ИМ	ОМ	тонк-ОЕ
25	ЕЕ	ЕГО	ЕМУ	ЕЕ		ИМ	ЕМ	син-ЕЕ
26	О	А	У	О		ЫМ	ОМ	лапласов-О

No.	복 수						예	
	ИМ	РОД	ДАТ	ВИН	ВИН	ТВОР		ПР
				неод	од			
27	ЫЕ	ЫХ	ЫМ	ЫЕ	ЫХ	ЫМИ	ЫХ	прост-ЫЕ
28	ИЕ	ИХ	ИМ	ИЕ	ИХ	ИМИ	ИХ	син-ИЕ
29	Ы	ЫХ	ЫМ	Ы	ЫХ	ЫМИ	ЫХ	лапласов-Ы

No	형용사 단어미형				예
	МУЖ	ЖЕН	СР	МН	
30	#	А	О	Ы	прост, прочитан
31	Ь	Я	Е	И	син-Ь
32	#	А	О	И	строг, хорош

형용사에서도 маленький, большой, хороший와 같은 어휘들을 고려해야 한다. 이와 함께 장어미형 외에 단어미형을 고려해야 한다. 단어미형에서는 короток, силен과 같이 출몰모음을 갖는 남성형을 위한 장치가 필요하다.

동사의 경우 Zaliznjak(1980:91-135)의 분류에 의하면 현재 시제 변화형은 16개 유형에 이른다. 이것을 다시 100여 개의 소 유형으로 나누었다. 그러나 일반적으로 러시아어 동사 활용 어미 유형은 제 1식(-ю, -ешь, -ет, -ем, -ете, -ют)과 제 2식(-ю, -ишь, -ит, -им, -ите, -ят)으로 구분된다. 이것을 ETAP 시스템에서는 어미의 첫 글자에 따라서 10개로 목록을 만들었다.

〈동사 어미 목록〉

* 1인칭 단수와 2인칭 복수 직설법

- | | | | |
|-----|---|-----|-----------|
| 33. | ю | ЕТЕ | (чита-ю) |
| 34. | ю | ИТЕ | (говор-ю) |
| 35. | у | ЕТЕ | (двин-у) |
| 36. | у | ИТЕ | (сообщ-у) |

* 2인칭 단수와 1인칭 복수 직설법

- | | | | |
|-----|-----|----|-------------|
| 37. | ЕШЬ | ЕМ | (чита-ЕШЬ) |
| 38. | ЕШЬ | ИМ | (говор-ИШЬ) |

* 3인칭 단수와 3인칭 복수 직설법

- | | | | |
|-----|----|----|------------|
| 39. | ЕТ | ЮТ | (чита-ЕТ) |
| 40. | ЕТ | УТ | (двин-ЕТ) |
| 41. | ИТ | ЯТ | (говор-ИТ) |
| 42. | ИТ | АТ | (сообщ-ИТ) |

ETAP 시스템에서 과거 시제는 단수 여성과 중성, 복수 어미형만 표시한다. 무형

어미를 갖는 단수 남성형은 따로 구분하지 않는다. 그것은 과거 시제 접미사 {-л}만 을 갖는 남성형을 관습형(трафареты)으로서 형태소 사전에 올리기 때문이다.

* 과거시제

43. А О И (читал-А)

반면에 명령법은 어미의 첫 글자에 따라 3개의 유형으로 구분된다.

* 명령법

44. й йТЕ (чита-й)

45. и иТЕ (ход-и)

46. ь ьТЕ (двин-ь)

동사에서도 читать와 같은 제 1식 변화형 동사와 говорить와 같은 제 2식 변화형 동사의 패러다임 외에 хотеть와 같은 혼합형, брать, взять, нести, везти 등과 같이 어간의 변형이 요구되는 어휘들에 대한 영역이 지정되어야 한다.

러시아어의 동사 체계에서는 시제 개념보다도 상의 개념이 중요하다. 거의 모든 동사 어휘들은 불완료상과 완료상이 하나의 쌍으로 이루어져 있다. 이것들은 형태적으로 구분될 뿐만 아니라 의미적으로 대립된다. 불완료상이 보통 반복, 경험, 미종료된 행위나 행위 사실 자체를 가리키는 반면, 완료상은 일회적 또는 결과적 행위를 가리킨다. 시제와의 관계에서도 불완료상, 완료상 형태가 각각 다른 시점을 표현한다. 불완료상 동사 читать의 1인칭 단수 현재형 читаю는 "(나는) 읽고 있다"라는 현재 진행적 의미를 가리키지만, 완료상에 의한 прочитаю는 "(나는) 읽을 것이다"라는 미래시제를 나타낸다. "(나는) 읽고 있을 것이다"라는 미래 진행적 의미는 я буду читать라는 복합 형태로 표현된다.

동사에서도 단수와 복수의 구분이 있으며, 각각의 수 개념은 인칭(1, 2, 3인칭) 개념과 연결되어 있다. 그리고 이 인칭과 수는 독립적인 어미 형태를 가지고 있다.

통사적으로도 러시아어에서 동사는 명사와 같이 주어, 목적어 기능을 수행할 수 있다.

Читать много - это очень важно. (많이 읽는 것 이것은 매우 중요하다.)

Знать значит понимать. (안다는 것은 이해한다는 것을 의미한다.)

위와 같은 예문들에서 동사 미정형들은 주어와 목적으로 기능하고 있다.

러시아어에서 수사는 기본적으로 남성 명사 путь와 같은 변화 패러다임을 갖는다. ETAP 시스템에서는 이것을 독립된 목록으로 설정한다.

<수사 어미 목록>

47. ь, и, и, ь, ь, ю, и (пять-ь)

러시아어에서는 수사 또한 복잡하다. "1"(один)과 "2"(два)는 성에 의해 구분되는 이형태를 가지고 있으며, 격에 따라 명사와 같이 변화하거나 대명사와 같이 변화한다. 또한 기수사와 서수사의 구분이 엄격하여 이것들은 서로 다른 어휘로 취급되어야 한다.

ETAP 시스템에서는 재귀 소사 -ся/-сь는 후접사로서 별도의 항목으로 취급되었다. 왜냐하면 러시아어의 재귀 동사는 타동사에 재귀 소사가 첨가됨으로써 파생되는 형태적 규칙성을 가지기 때문이다. 이를 위한 독립된 항목은 동사의 분석과 합성에 효율성을 부여한다.

<후접사 목록>

- 1. # сь (читала-сь)
- 2. # ся (читал-ся) 1인칭 단수와 2인칭 복수 직설법

러시아어에서 피동성 개념은 동사 어휘 끝에 붙는 재귀소사 -ся에 의해 나타내어질 수도 있지만, 형태적인 특성에 의해 수동태가 표현되어질 수 있다. 또한 동사의 모든 진술(representation)은 평서법, 명령법, 미정형, 형동사, 부동사로 구분된다.

3. 형태적 분석을 위한 분류

형태부에 자연어의 문장이 입력되면 제일 먼저 문장의 형태적 구조가 분석되고 분석된 어형에 형태적 특성이 수록된다. 이때 형태적 분석은 낱말 하나하나에 대해 가해진다. 이것은 입력 문장의 어형이 문장의 문맥과 상관없이 개별적으로 이루어진다는 것을 의미한다. 즉 각각의 어형에 관계 있는 모든 형태적 구조(morphological structure)가 열거된다. 어형이 동음이의어(homophony)를 갖는 경우에는 형태적 구조가 하나의 집합으로 나타내어진다. 예를 들어 *стекло*라는 어형은 3가지 동음이의어를 갖는다. 즉 명사 단수 주격과 대격, 그리고 동사 *стекать*의 완료상 *стечь*의 단수 과거 중성형. 형태 분석 과정에서 이 3가지 동음이의어가 모두 포착되어야 한다.

стекло : СТЕКЛО им, ед
 СТЕКЛО вин, ед
 СТЕКАТЬ сов, изъяв, прош, ср, ед

이것은 본 연구가 추구하는 형태론 모델의 목표가 완전성과 일치성이기 때문이다. 이것은 본 연구가 모델로 삼고 있는 ЕТАР 시스템의 기본 원칙이기도 한다 (강덕수, 박창호 1995 : 679).

본 연구는 언어학적 기술(linguistic description)과 알고리즘적 기술을 명확히 구분함으로써 노어의 형태적 모델이 실제형과 일치되도록 하고 모순되는 형태를 방지하고자 하는 것이다.

예를 들어 동사 어휘 *читать*의 능동형동사 현재 시제형의 여성 단수 대격형인 *читающуюся*는 다음과 같이 분석된다.

читающуюся ⇒ ЧИТАТЬ прич, непрош, страд, жен, вин, ед

이러한 동사 *читать*는 형태사전에 225개의 어형을 가지게 되며, 역시 225개의 개별적인 법칙이 수반된다.

하나의 동사 어형태는 하나의 어간과 5개의 분절형으로 구성된다.

예 : *читать*

접두사 : *про - чит - а - тъ¹⁾*

1) 접두사는 무형(#)¹⁾일 수도 있다.

- 매개모음 : чит - а - ть
 접미사 : чит - а - ющ - ую- ся
 어미 : чит - а- ющ - ую - ся
 후접사 : чит - а- ющ- ую - ся

이러한 분절형들은 형태 사전에 독립적으로 수록되어, 각각의 어형의 해당 형태들과 일치된다.

читающуюся :

- (접두사) # → несов (불완료)
 (어간) чит → читать : V(동사)
 (매개모음) а → ∅ (무의미 형태소)
 (접미사) ющ → прич, непрош (형동사, 비과격)
 (어미) ую → жен, вин, ед (단수, 여성, 대격)
 (후접사) ся → страд (수동태)

모든 어휘는 기본적인 구성 요소인 형태들의 특성을 집합적으로 갖는다. 이 특성들은 형태적 자질을 가리킨다. 예를 들어 형태소 -ами는 "복수"와 "조격"이라는 특성을 가지며, 복수는 제 1 요소로서 "수" 범주의 자질이고, 조격은 제 2 요소로서 "격" 범주의 자질이다. 이러한 문법적 자질은 품사, 활동성, 시제, 인칭, 피동성의 12개로 분류된다.

본 연구에서 품사는 대명사를 명사에 포함시켜 9개로 구분한다. 품사에는 고유 의미의 조건적 의미가 표시된다. 먼저 고유 의미가 영문 약자로 표시되고, 세미콜론 (;)에 의해 품사 어간이 갖는 자질들이 로마숫자로 표시된다. 이 가운데 항상 나타나는 자질이 아닌 선택적 자질들은 괄호로 주어진다. 수사는 격을 의무 자질로 가지지만, 활동성과 성은 일부 수사에서만 나타나는 선택 자질이다.²⁾

형태적 구조에 따라 해당 어휘의 문법적 특성을 제시하기 위해서는 특별한 기호

2) 활동성이 표시되는 수사 : "1"(один), "2"(два), "3"(три), "4"(четыре)
 성이 표시되는 수사 : "1"(один, одна, одно), "2"(два, две),
 "둘 다"(оба, обе), "1/2"(полтора, полторы)

를 지정해야 한다. 이러한 기호는 품사(part of speech), 활동성(animacy), 성(gender), 수(number), 격(case), 비교급(comparative), 단어미(short form), 진술(representation), 상(aspect), 시제(tense), 인칭(person), 태(voice)와 같은 문법적 특성을 나타내는데 사용된다.

I. 품사

1. 명사 - S ; II, (III), (IV), V
2. 형용사 - A ; (II, III, IV, V, VI, VII)
3. 동사 - V ; (II, III, IV, V, VII, VIII, IX, X, XI, XII)
4. 부사 - ADV ; (VI)
5. 수사 - NUM ; (II, III), V
6. 전치사 - PR
7. 접속사 - CONJ
8. 소사 - PART
9. 합성어 - COM

II. 활동성

1. 활동체 - од
2. 활동체 - неод

III. 성

1. 남성 - муж
2. 여성 - жен
3. 중성 - ∅

IV. 수

1. 단수 - ед
2. 복수 - мн

V. 격

1. 주격 - ми
2. 생격 - род
3. 부분격 - парт

- 4. 여격 - дат 5. 대격 - вни 6. 조격 - твор
- 7. 전치격 - пр 8. 처소격 - местн

VI. 비교급³⁾

- 1. 비교급 - срав 2. 최상급 - прев

VII. 단어미

- 1. 단어미 - кр

VIII. 진술

- 1. 평서법 - изъяв 2. 명령법 - пов 3. 미정형 - инф
- 4. 형동사 - прич 5. 부동사 - деепр

IX. 상

- 1. 불완료상 - несов 2. 완료상 - сов

X. 시제

- 1. 현재와 미래 - непрош 2. 과거 - прош
- 3. 현재(동사 есть의 경우) - наст

XI. 인칭

- 1. 1인칭 - 1-л 2. 2인칭 - 2-л 3. 3인칭 - 3-л

3) 형용사와 부사의 원급(positive degree)에는 특별한 개념을 설정하지 않는다. 비교급이나 최상급의 표시가 없으면 모든 형용사와 부사가 원급 외에 다른 급을 가지지 않는 것으로 간주한다.

X. 피동성⁴⁾

1. 수동태 - страд

4. 형태부의 분석 단계

мельчук에 의하면 형태부는 심층 영역(Deep-Morphological Representation)과 표층 영역(Surface-Morphological Representation)으로 구분된다. 심층 형태부는 어형의 내부 조직과 상관없이 문장을 구성하는 어형과 어구의 형태 그리고 선형적 어순을 명시한다. 그러나 이 단계에 들어가기 전에 먼저 입력된 문장의 어휘들의 초기 형태가 분석되어야 한다. 이 단계가 표층 형태부로서 여기에서는 어형의 내부 형태소 구조(the internal morphemic organization of word-forms)가 드러난다(Мельчук, 1988 : 71).

문장들의 형태소 분석이 이루어지는 예는 다음과 같다.

Были два друга в нашем полку. (우리 연대에는 두 명의 친구가 있었다)

БЫТЬ (прош, мн, несов, изъяв),

ВА (им, муж),

ДРУГ (ед, род),

В,

НАШ (муж, ед, пр),

ПОЛК (ед, местн)

Он стоял слева от меня. (그는 나의 왼쪽에 서 있었다)

ОН (им, ед, муж, од),

4) 피동상 자질이 표시되는 것은 후접사 -ся(또는 -сь)가 있는 동사와 없는 동사가 서로 “능동”(active)과 “피동”(passive)의 태적 대립(voice opposition)을 보이는 동사에 한한다. бриться, мыться와 같이 귀환성(reflexiveness)을 나타내거나 драться, видиться 처럼 상호성(reciprocity)을 의미하는 동사들은 -ся가 있는 형태를 개별적인 어휘소로 간주하여 사전에서 개별 항목으로 처리한다.

СТОЯТЬ (прош, ед, изъяв, муж, несов),
 СЛЕВА,
 ОТ,
 Я (род, ед, он / вин, ед, од)

기계 번역을 위한 자연어 처리 과정이 제일 먼저 착수되는 표층 형태 영역에 입력되는 문장은 자연어 상태 그대로이다. 따라서 형태적 분석을 시작하기 전에 입력된 문장의 어형태 가운데 이미 변형이 일어나 준비된 형태소 사전에 의해 처리될 수 없는 것을 미리 걸러내야 한다. 이러한 단계를 선 형태적 분석(pre-morphological analysis) 단계라고 할 수 있다. 그런데 노어에서는 영어의 isn't, can't, high-pressure, last-ditch, shirt-sleeve와 같이 hyphen이나 dash, apostrophe 같은 것에 의해 형태적으로 축약되거나 변형된 경우가 없기 때문에 노어 기계 번역에서는 이 단계를 고려하지 않아도 된다. 따라서 노어에서는 표층 형태부에서 이루어진 분석이 심층 형태부로 그대로 넘어갈 수 있다.

일반적으로 변형된 비 형태적 어형에 대한 처리가 끝난다 하더라도 형태소 사전에 의해 처리되지 못하는 부분이 남을 수 있다. 예를 들어, так как(왜냐하면), в соответствии с(...에 따라서), несмотря на(..에도 불구하고), кроме того(그 외에), друг другу(서로서로) 같은 표현들은 형태소 사전에 의해 올바르게 처리될 수 없다. 왜냐하면 형태적 분석은 낱말 하나 하나를 개별적으로 다루기 때문이다. 따라서 이러한 숙어적 표현들은 하나의 개별적 단위로 간주하여 사전 규칙(dictionary rules)으로 처리해야 한다. 이러한 사전 규칙들은 입력어의 모든 분석 과정에 관여하는 조합 사전(combinatory dictionary)에 수록된다. 따라서 위와 같은 숙어적 표현들은 러시아어 조합 사전에 포함된다. 이러한 표현들을 걸러내는 과정을 절대 어구(absolute phrases, безусловные обороты) 분석 단계라고 부를 수 있다.

이러한 과정을 거치면 언어학적으로 순수하게 형태(morphs)라고 부를 수 있는 어형(wordforms)만이 남게 된다. 이 마지막 과정을 순수 형태적 분석(purely morphological analysis) 단계라고 할 수 있다.

이와 같이 형태적 분석 단계는 번역 시스템 내에서 3개의 블록으로 구분된다.

- 1) 선 형태적 분석 블록(A bloc for pre-morphological analysis)
- 2) 절대 어구 분석 블록(A bloc for analysis of absolute phrases)
- 3) 순수 형태적 분석 블록(A bloc for purely morphological analysis)

본 연구는 제3 단계인 순수 형태적 분석을 위한 블록을 구축하는 방법을 제시해 보고자 한다.

5. 형태적 분석 방법

선 형태적 분석과 절대 어구의 분석 단계를 거치게 되면 언어학적으로 순수하게 형태라고 할 수 있는 어형만 남게 된다. 본 단계에서는 어형 마디 위에 제시된 문법적 특성이 수록되고 분류된다. 이 과정에서 중심적 역할을 하는 것은 형태소 사전이다. 이와 함께 굴절 언어(inflexional language)인 러시아어의 다양한 굴절 어미를 정교하게 분석하기 위해서는 어미 목록의 표준화가 중요하다. 형태소 사전과 표준 어미 목록을 구축하는 외에 ETAP 시스템은 형태부 분석을 효율적으로 수행할 수 있는 장치들을 고안하였다.

5.1 형태소 사전

순수한 형태적 분석의 중심 역할을 수행하는 형태소 사전은 사전 항목(dictionary items, словарная статья)으로 이루어져 있다. 이 사전 항목은 1개의 표제어와 몇 개의 부제어를 포함한다. 표제어는 어간(stems) 또는 어형(word-forms)으로 나타내어지며, 부제어는 5 자리의 숫자로 표시된다. 부제어를 가리키는 숫자는 조합 사전에 수록되는 해당 항목의 번호와 일치한다.

사전 항목에 포함되는 형태 정보 영역(a zone of morphological information)은 어간과 어미에 대한 정보를 제공한다.

동음이의어는 여러 항목으로 분산되어 수록되나, 그 밖의 비 동음이의어는 단일 항목으로 처리된다.

<형태소 사전 항목의 예>

예1 : СПИРТ

NAR : 03279

F : 1, NLY : 6, LY : ('Y'(ЕД, ПАТР), 'Y'(ЕД, МЕСТН))

예2 : ПОЛОЖИТЕЛЕН

NAR : 00558

F : 11

예3 : ПОЛОЖИТЕЛЬН

NAR : 00558

F : 11

NAR : 01858

F : 38

5.1.1 표제어

사전 항목의 표제어는 단어의 어간으로 나타내어진다. 철자상 구성이 같은 어간을 가진 단어들은 동일한 형태 항목에 기술된 다음 숫자로 표시되는 부제어에 의해 구분된다. 예를 들어 명사 вид와 동사 видеть는 동일한 항목에 의해 기술된다. 반면에 동일 단어의 교체 어간(alternative stems)은 положителен, положительн과 같이 다른 항목으로 나뉘어 기술된다. 예를 들어 동사 видеть와 명사 вид는 공통적으로 표제어를 항목 вид로 하고, 동사 비과거 시제형 виж-у, у-виж-у는 표제어를 항목 виж로 한다.

5.1.2 부제어

부제어는 5 자리의 숫자로서 조합 사전(combinatory dictionary)에 수록된 해당 항목의 번호와 일치한다. 이 부제어는 어간이 동일한 관계로 같은 항목에 기술되어야 하는 어휘들을 구분해 주는 기능을 한다.

5.1.3 열

열은 표제어로서 제공되는 어간과 결합하게 되는 접두사, 어간 매개소(theme vowel), 접미사, 어미, 후접사와 같은 분절 형태(segments)를 수록한다. 이 분절 형태들은 약자로 구성되는 키워드로 구분되며, 키워드의 내용은 다음과 같다.

LA : 접두사
 LY : 어미
 LW : 매개소
 LZ : 후접사
 LX : 접미사
 CTH : 어간

키워드에 의해 해당되는 형태는 철자형으로 주어진다. 이 철자형은 인용 부호에 의해 구분된다. 명사 남성 단수 주격에서 처럼 철자가 겹으로 나타나지 않는 형태는 #로 표시된다. 그와 함께 그 형태의 형태소적 특성이 괄호에 의해 제시된다.

예1 : 동사 обстреливать 항목의 분절 형태:

LW : 'ива' несов,

LY : 'я'(депр, непрш, 'ю'(изъяв, несов, ед, 1-л), 'ете'(изъяв, непрш, мн, 2-л),

LZ : ('#', 'сь', страд)

위와 같은 방법에 의해 불완료상 능동태와 수동태의 부동사 обстреливая, обстреливаясь, 비과거 시제 단수1인칭 обстреливаю, обстреливаюсь, 복수 2인칭 обстреливаете, обстреливаетесь가 실현된다.

위 방법에서는 불완료상 동사 능동태와 수동태의 부동사, 비과거 시제 단수 2, 3인칭형, 복수 1, 3인칭형은 주어지지 않는다. 이 형태들은 수동태를 나타낼 때 후접사 ся를 가지므로 독립적으로 주어진 열에 의해 실현된다.

예2 : нашёл

СТН : (сов, изъяв, прош, ед, муж), NAR : 06006

동사 найти와 같이 어형(wordform)이 특정한 어간이나 형태소를 가지고 있다면, 그 어형을 분절 형태로 나누는 것은 기계 번역에서 효율성을 떨어뜨릴 수 있다. 이런 경우 형태소 사전에는 어형과 함께 그 어형의 특성을 전부 수록하는 것이 효율적이다. 그렇지 않으면 예를 들어 동사 найти의 과거 시제 남성 어형 нашёл은 대명사 наш의 중성형 наше와 혼동될 수 있다. 그러므로 이 경우에는 어형 нашёл이 독립적인 열에 의해 제시된다.

три

СТН : (NUM, твор), NAR:02317

또 다른 예로서 три '3'의 조격형 триа는 특정 어미 -емя를 가지므로 역시 독립적으로 사전에 수록된다.

6. 형태부의 구성

형태(morphs)와 문법적 특성은 형태소 사전에 수록된 해당 항목에서 충분히 제시될 수 있다. 그러나 이러한 정보를 사전에만 의존한다면 사전이 대단히 커질 것이다. 이와 함께 형태부도 방만하게 될 것이다. 이러한 단점을 보완하기 위해 ЕТАР 시스템에서는 4 가지 유형의 규칙을 고안하고 있다. 그 첫 번째 것은 수많은 활용 어미들을 표준화시키기 위한 표준형(standard paradigms of endings), 두 번째 것은 반복적으로 사용되는 형태적 정보의 양을 축소하여 효율성을 높이기 위한 규격형(formats), 세 번째 것은 여러 규격형에 반복해서 포함되어 있는 형태적 정보를 통합해서 관리하기 위한 관습형(trafarets), 네 번째 것은 어미 유형의 예외적 활용을 규격화 하기 위한 은폐형(masks)이다.

6.1 표준형

러시아어는 명사, 대명사, 형용사, 동사 등 주요 품사들이 성, 수, 격에 따라 변화하는 굴절형 언어이다. 굴절어미들은 기본적인 패러다임으로 형식화될 수 있다. 형식화된 어미 형태는 표준형으로 형태부의 기반을 구성한다. 제 2 장 “형태부의 특성”에 수록된 목록이 바로 표준형이 된다.

6.2 규격형(форматы)

표준 어미 목록은 성, 수, 격에 따라 분류되었기 때문에 어휘의 합성 과정에서 비효율적일 수 있다. 예를 들어, *хороший*와 같은 형용사는 단수 남성, 여성, 중성과 복수의 4가지 패러다임을 가진다. 이러한 패러다임의 중복 현상은 거의 모든 형용사에서 나타난다. 이러한 형태적 정보의 목록을 효율적으로 이용할 수 있도록 축소시키는 방법으로 규격형을 만들 수 있다. 형태소 사전의 사전 항목에는 그 항목의 어간과 결합하게 되는 분절형(접두사, 매개소, 접미사, 어미, 후접사 등)을 일일이 열거하는 대신 규칙적으로 반복되는 분절형들의 집합인 규격형의 번호가 주어지게 된다.

규격형의 예 :

예1. F:7

NLY:(14,20,25,28)

규격형 F:7은 형용사 남성(표준형 14번), 여성(표준형 20번), 중성(표준형 25번) 단수 뿐만 아니라 복수형(표준형 28번)을 포함한다. 따라서 *минувший*라는 형용사는 형태소 사전에 규격형 번호 F:7을 수록하고 있기 때문에 다음과 같이 실현될 수 있다.

예1. *минувш-ий, минувш-его, ...*

минувш-ая, минувш-ей, ...

минувш-ее, ...

минувш-ие, ...

예2. F:16

LY:'я' (ДЕЕПР, НЕПРОШ)

〈예〉 вид-я зна-я

F:16은 부동사 비과거 시제형을 규칙적으로 실현한다.

예3. F:31

LX:'юш'(ПРИЧ, НЕПРОШ), NLY:(14,20,25,28)

규격형 F:13은 접미사 'юш'를 가진 형동사 비과거 시제형의 모든 형태를 제공한다.

〈예〉 кол-юш-ИЙ, кол-юш-ЕГО, ..., кол-юш-АЯ, ..., кол-юш-УЮ, ..., кол-юш-ИЕ.

예4. F:40

LX:'вш'(ПРИЧ, ПРОШ), NLY:(14,20,25,28)

규격형 F:40은 접미사 'вш'를 가진 형동사 과거 시제형의 모든 형태를 제공한다.

〈예〉 чита-вш-ИЙ, ..., чита-вш-АЯ, ..., чита-вш-ЕЕ, ..., чита-вш-ИЕ

예5. F:42

LX:'л'(ИЗЪЯВ, ПРОШ), NLY:43, LY:'#'(ЕД, МУЖ)

규격형 F:42에 의해 평서문 과거 시제형이 제공되며, 남성 단수형은 무형 어미를 갖는다.

〈예〉 кол-о-л, кол-о-л-а, кол-о-л-о, кол-о-л-и

6.3 관습형(трафареты)

규격형에 해당하는 형태소 분절형들이 여러 어휘들에 공통적으로 이용될 때 이것을 또 다른 집합으로 묶어 하나의 항목으로 만들면 어휘의 합성에 효율성을 더욱 높

일 수 있다.

예 1 : T:77

LX : 'е' срав

LX : 'айш' прев, F:7

위 관습형 77번은 2개의 열을 포함하고 있다. 첫번째 열은 비교급 형성에 관한 정보를 제공해 준다. 이 정보에 의해 형용사들은 крепкий-крепче와 같은 비교급을 만들어낼 수 있다. 두번째 열은 최상급에 관한 정보를 담고 있다. 이 정보에 의해 접미사 {-айш-}를 갖는 형용사 최상급이 만들어진다. 최상급 형태는 형용사와 마찬가지로 성, 수, 격에 따라 활용 어미를 가진다. 이에 대한 정보는 같은 열에 포함된 F:7에 의해 주어진다. 왜냐하면 규격형 F:7은 성, 수, 격에 따라 필요한 모든 형태 정보를 압축시킨 NLY:14,20,25,28을 포함하는 집합이기 때문이다. 이에 따라 형용사 крепкий는 다음과 같이 비교급과 최상급 변화형을 갖게 된다.

〈예〉 비교급 : крепче,

최상급 : крепчайший, крепчайшего,...

крепчайшая, крепчайшей,...

крепчайшее,...

крепчайшие,...

예 2 : T:61

LA:(A1, '#' несов), NLY:(33,37,39,44), LY : 'ть' инф

LA : '#' несов, F:31

LA : '#' несов, F:16

LA:(A1, '#' несов), F:42

LA:(A1, '#' несов), F:40

LA : A1, LY : 'в'(депр, прош)

관습형은 해당 그룹의 어휘가 필요로 하는 모든 형태적 정보를 제공한다. 위 예 T:61은 동사에 관한 관습형이다. 이 관습형에 의해 (по)страдать 유형의 동사들은

모든 활용 어미 형태를 가질 수 있다. 이 유형의 동사들의 불완료상은 무형 접두사 (#)에 의해 형성된다. 완료상은 다양한 접두사에 의해 형성되는데, 이에 대한 구체적인 정보는 해당 동사들의 형태소 항목에 별도로 주어진다. 접두사는 관습형의 1, 4, 5, 6 열의 A1에 의해 변항(parameter)으로 주어진다. 관습형 T:61은 부동사 비과거 시제형, 형동사 비과거 시제형, 형동사 과거 시제형, 평서문 과거 시제형, 부동사 과거 시제형을 모두 제공한다. 따라서 (по)страдать 동사는 다음과 같은 어형을 갖게 된다.

- 〈예〉 현재 시제형 (по)страда-ю, (по)страда-ешь,.....
- 형동사 #-страда-ющ-ий,.....
- 부동사 #-страда-я,
- 과거 시제형 (по)страда-л-#, (по)страда-л-я,.....

6.4 은폐형(маски)

은폐형 규칙은 예외적이거나 표준화 시키기 어려운 활용 어미 형태나 표준 어형 변화들(standard paradigm)에 나타나지 않는 형태를 가진 어휘의 활용형을 합성할 때 효율적이다. 예를 들어 명사 брат ‘형제’의 경우 복수형은 братья, братьев, братьям, братьев, братьями, братьях이다. 이것은 예외적으로 나타나는 것이므로 규칙화하기가 어렵다. 이러한 경우 기존의 어미 목록을 활용하는 대신 예외적인 부분만 대체할 수 있다면 어미 목록의 수도 줄이고 규칙도 간편하게 할 수 있다. 은폐형 규칙을 활용하면 예외적인 형태를 일일이 열거하게 되는 수고를 덜어줄 수 있다는 점에서 그 의미가 있다.

은폐형은 숫자 0과 1로 구성된다. 어휘의 해당 어미 유형에 은폐형을 대응시켜 숫자 0에 해당되면 생성 과정에서 그 어미형은 삭제된다. 반면에 숫자 1에 해당되면 활용된다.

- 예1 : M:007
- 1.1.1.1.1.1.0.0.0.0.0

예를 들어 명사 брат는 형태소 사전 항목에 M:007이라는 은폐형 규칙을 갖는다. 따라서 명사 брат는 어미 목록 6번에 의해 어미형을 갖게 되지만 은폐형 7번에 의해 복수형에서는 표준형의 출현이 봉쇄되고, 대신 위와 같은 예외형이 실현된다.

예2: T:018

Φ :2, сок:5/7

관습형 규칙 18번은 규격형 2번 가운데 어미 활용은 표준 어미 목록 5번을 따르되, 은폐형 규칙 7번에 의해 복수 어미형 6개를 표면으로 생성시키지 않는다는 것을 의미한다.

7. 형태소 사전 항목의 구성

형태소 사전의 개별 항목이 실제로 어떻게 구성되는가는 구체적인 예를 분석하면서 예시해 보도록 한다.

예 1: 00060 библиотек|а ОСН|= T:4

библиотека라는 단어는 여성형으로서 어간이 연구개 자음으로 끝나므로 книга, рука의 유형과 같이 표준어미 목록(сок)의 4번에 해당한다. 따라서 표준 어미 목록 [сок:4]를 갖게 된다. 그런데 이 단어는 품사가 명사이고, 성 범주는 여성, 활동성 범주는 불활동성이기 때문에 규격형 2번으로 통합된다. 즉 (Φ :2, хар:S, жен, неол)로 표시된다. 그런데 표준어미 목록 сок:4와 규격형 Φ :2는 활용범위가 일치한다. 따라서 이 두 번호를 묶어 T:4라는 하나의 관습형으로 만들 수 있다. 즉 T:4 \in Φ :2, сок:4. 이 단어에는 최종적으로 형태소 사전에서 다음과 같이 표시된다.

위의 사전구성에서 제일 왼쪽의 5자리 숫자는 이 단어의 고유번호이고, 그 다음에 표제어가 나타난다. 표제어에 나타나는 ' | ' 표시는 그 표시의 왼쪽이 어간이고 오른쪽은 어미라는 것을 구분해준다. 다음 블록의 'осн:'는 어간을 나타내주는 것으로

이 단어에서는 표제어에서 'l'로 구분된 어간이 아무런 변화형 없이 그대로 활용되기 때문에 동일하다는 의미로 '=' 표시를 하였다. 마지막 블록은 이 단어의 모든 어형변화형을 나타내주는 항목으로 관습형의 번호만이 주어졌다.

예2: 02674 станок осн|стан(о)к р:2 т:9

станок이라는 단어를 예로 들면, 이 단어는 명사 сток 유형에 해당하는 표준어미 목록 9번에 속하므로 сок:9로 표시될 수 있다. 또한 이것은 남성, 불활동성의 자질을 포함하는 규격형 1번에 해당한다. 즉, $(\phi:1 \text{ хар:S, муж, неод})$. 그런데 сок:9와 $\phi:1$ 은 활용범위가 일치하므로 т:9라는 하나의 관습형으로 묶을 수 있다. 즉, $\tau:9 \in \phi:1, \text{сок:9}$. 이 단어에서 한가지 유의해야 할 사항은 교체어간이 있다는 점이다. 예를 들어 단수 생격 станка에서는 모음 o가 빠져 있다. 따라서 사전항목의 어간 표시 'осн:'에는 'стан(о)к'와 같이 교체어간이 표시되어야 한다. 이러한 교체어간을 표시하는 방법으로 제한자(ограничитель ; r로 표시)라는 블록이 사용된다. 이 제한자는 기본 어간 외에 또다른 이형태를 갖는 경우 이형태를 불러내기 위한 장치이다. 단어 станок의 경우에는 단수 주격과 대격에서만 станок이라는 어간을 갖고, 나머지 격형태에서는 o가 빠진 станк-를 어간으로 가진다. 이것은 제한자 제 2번으로 р:2 ед. им/ ед. вин으로 표시된다. 이 장치에 의해 станок의 완전 형태의 어간을 갖는 것은 단수 주격과 대격으로 제한된다. 단어 станок에 대한 형태사전은 다음과 같다.

04104 кислород осн|= $\phi:1$ сок:6/7

кислород라는 단어는 남성, 불활동성의 자질을 갖기 때문에 규격형 1번에 해당한다. 즉, $\phi:1 \text{ хар:S, муж, неод}$. 단수에서 변화유형은 ротор와 같이 표준어미 목록 6번에 해당하지만, 이 단어는 복수 변화형을 갖지 않기 때문에 은폐형을 사용하여야 한다. 즉 표준어미목록 6번에서 단수형은 나타나도록 하고, 복수형은 은폐형 7번을 사용하여 표시되는 것을 봉쇄한다.

сок : 006

'#ед,им, 'а'ед,род, 'у'ед,дат, '#ед,вин, 'ам'ед,твор, 'е'ед, пр *
'ы'мн,им, 'ов'мн,род, 'ам'ми,дат, 'ы'мн,вин, 'ами'мн,твор, 'ах'мн,пр

м: 007

1.1.1.1.1.1.0.0.0.0.0.0

위의 예에서 6번 표준어미목록(сок:6)에 7번 은폐형(м:7)을 사용하여 은폐형에서 '1'에 해당되는 단수형 변화 6개는 그대로 두고, '0'에 해당하는 복수형 변화 6개는 출현을 봉쇄할 수 있다. 그러나 이러한 변화유형은 많은 단어에 해당하는 것이 아니기 때문에 관습형을 별도로 작성하지 않고 어미목록 번호에 은폐형 번호를 덧붙이는 방식에 의해 сок : 6/7과 같이 표시한다.

8. 결론

본 논문에서는 ETAP 시스템에서 구축된 형태소 사전을 모델로 하여 러시아어의 형태소 분석 방법과 사전 구성 체제를 분석해 보았다.

이 시스템에서는 형태부를 선형태적 분석 블록, 절대 어구 분석 블록 그리고 순수 형태적 분석 블록으로 구분하였다. 이 중 주요한 과정은 순수 형태적 분석 블록에서 일어난다. 이 과정의 분석을 위해서 제일 먼저 활용 어미를 정형화한 표준형의 확립이 이루어졌다. 그리고 러시아어의 중요 문법 개념인 성, 수, 격에 따라 활용어미들이 중복되는 현상을 집합 개념에 의해 축약시켜 운용의 효율성을 극대화하고자 하였다. 이것이 규칙적으로 반복되는 분절 어미형들을 분류하여 집합적으로 지정된 규격형이다. 이에 의해 개별 어휘마다 표준형 어미를 하나 하나 열거하는 번거러움을 제거할 수 있다. 또한 동일 그룹의 어휘들이 별개의 규격형 어미들과 표준형 어미를 동시에 요구하는 경우들이 있다. 이런 경우 규격형 어미와 표준형 어미를 하나의 집합으로 묶어 관습형을 구성하여 어미 활용의 효율성을 더욱 높이고자 하였다. 따라서 형태부 규칙 구성에서 표준형, 규격형, 관습형 개념은 ETAP 시스템의 요체라

할 수 있다.

이와 함께 표준화시키기 어려운 활용어미 형태나 표준어형 변화들에 나타나지 않는 이형태를 가진 어휘들이 있다. 이런 어휘들은 대개 부분적으로만 표준형 어미들을 채택한다. 이 경우 은폐형 규칙에 의해 부분적 봉쇄와 부분적 실현을 가능하게 하였다. 이에 의해 이형태를 어휘 항목으로 지정해야 하는 체제상의 문제를 극복하고 자연어의 심층부를 기계번역 체계 안에 끌어들이 수 있게 되었다.

한국어의 분석과 합성에서도 이러한 집합적 개념을 응용한다면 한국어의 분석과 합성에서 더욱 더 큰 효율성을 기할 수 있게 될 것이다.

<참 고 문 헌>

- 강덕수, 김상국, 김성래, 박창호 (1993), 러시아어-한국어 기계번역 시스템의 기반 기술개발에 관한 연구(1), 한국과학기술원, 시스템공학연구소.
- 강덕수, 박창호 (1995), 노어-한국어 기계번역 기술개발 - 언어적 비교연구 및 노-한 계산 언어 모델화 부문, 과학기술처.
- 강덕수, 김성래, 김연수, 박창호 (1996), 러시아어-한국어 기계번역 시스템(2), 과학기술처, 시스템공학연구소.
- Hutchins (1986), *Machine Translation : Past, Present, Future*, Chichester : Ellis, Horwood Limited.
- Mel'čuk (1988), *Dependency Syntax : Theory and Practice*, Albany : SUNY Press.
- Апресян, Ю.Д. и др.(1989), *Лингвистическое обеспечение системы ЭТАП-2*, Институт проблем передачи информации, Москва : <<Наука>>.
- _____ (1992), *Лингвистический процессор для сложных информационных систем*, Институт проблем передачи информации, Москва : <<Наука>>.
- Зализняк, А.А.(1980), *Грамматический словарь русского языка*, Москва : <<Русский язык>>.

〈Abstract〉

Machine translation is said "a marriage of practical needs and theoretical idealism." It means that a study of machine translation has started by practical purpose, but its goal is theoretical idealism. This idealism assumes the possibility of linguistically perfect analysis of the structure of natural languages. Linguistic approach to natural languages for machine translation is focused on morphology, syntax and semantics.

A study of machine translation of Russian since 1950s has reached to significant level for Russian-English translation system. Especially, IPPI, led by Apresjan, Boguslavsky and Iomdin, has developed the ETAP system.

This paper analyzes the morphological component of this system and its difference from the natural structure of Russian. And the paper provides the effective system for the natural language processing, related to morphology, further, presenting a model for the construction of a morphological dictionary.