

노-한 기계번역에서 한국어 용언의 활용어미 합성 모델 연구*

강덕수 (한국외대), 최기선 (과기대)
김성래 · 김연수 (한국외대)

Kang, Ducksoo, Kisun Choi, Sungrae Kim, and Younsoo Kim. 1999. A Study of a Model for Composition of Conjugational Endings of Korean Predicates in Russian-Korean Machine Translation. *Korean Journal of Linguistics*, 24-3, 311-334. One of the problems to overcome in Russian-Korean machine translation is a variety of conjugational forms of Korean predicates. Russian, one of inflectional languages, has definite paradigms of endings in accordance with gender, number and case. In contrast, Korean, one of agglutinative languages, is considered not to have definite paradigms of inflectional and conjugational endings. But for machine translation it is necessary to systemize its endings and to make the definite paradigms. The paper tries to analyze problems of Korean predicates in Russian-Korean machine translation and to systemize Korean predicates, introducing the concept of conjugation to Korean, and finally to show the paradigm of effective generation of Korean predicates as an output in Russian-Korean machine translation. (Hankuk University of Foreign Studies)

1. 서론

노-한 기계 번역에서 극복해야 할 문제중 하나는 한국어 술어부의 다양한 형태이다. 굴절 언어로서 노어는 성, 수, 격에 따라 주요 품사의 모든 어휘들이 일정한 어미를 갖는데 반해, 한국어에서는 교착 언어로서 체계적인 활용어미 패러다임이 제시되지 않고 있다. 그러나 실제로 어휘 활용면에서는 한국어가 노어보다 더 복잡한 면을 보이고 있어 이에 대한 체계화 노력이 중요한 것으로 보인다.

Я буду читать книгу.

⇒ 나는 책을 읽고 있을 것이다.

Я не читал книгу.

⇒ 나는 책을 읽지 않았다.

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

위 예문에서 노어는 동사의 경우 미래 시제는 조동사가, 과거 부정문에서는 본동사가 활용 어미를 가진다. 그러나 한국어에서는 미래 시제에서 조동사 “것이다”앞에 진행 의미 조동사 “있다”의 미래 시제를 나타내는 “-르”을 갖는 “있을”과 본 동사 “읽다”의 활용형 “읽고”가 온다. 과거 부정문에서는 부정 조동사 “않다”의 과거형 앞에 “읽다”의 변화형 “읽지”가 온다.

본 논문은 노-한 기계 번역 체계에서 한국어 슬어부가 가지는 문제점을 분석하고, 노-한 기계 번역시 한국어 슬어부를 효율적으로 합성하기 위해 한국어에 활용 어미 개념을 도입하여 이를 체계화하고자 한다.

2. 용언의 품사적 개념

문장은 주어와 서술어라는 2대 필수 기능 성분으로 이루어지는데, 그 주 기능 성분의 하나인 서술어의 핵심요소가 용언이다. 용언(predicates) 또는 동사류어(verb class words)는 문장의 서술기능을 담당하는 낱말들의 범주이다. 서술 기능이란 그 주어가 가리키는 사람이나 사물의 행위나 상태를 풀이하는 것을 말한다. 즉, 주어가 가리키는 대상의 움직임, 상태, 성질 또는 속성 따위를 풀이하여 밝히는 것이다. 용언의 서술 기능은 그 뒤에 덧붙여진 서술 보조소(auxiliary)의 도움을 받게 된다. 용언은 서술 기능의 핵심 구실을 하지만 그 자체로는 시제, 상, 서법 따위의 문법적 형식을 갖추지 못하므로 그런 기능을 드러내는 서술 보조소의 도움을 받는 것이다. 서술 기능은 실질적인 풀이 작용과 문법적인 형식 면으로 나누어진다. 전자는 주로 용언의 어간이 지닌 의미 기능에 속하며, 후자는 그 어간에 덧붙여지는 문법 요소들이 드러내는 형식적 기능이다 (서정수 1994:503).

- (1) 그 학생이 책을 열심히 읽는다.
- (2) 그 학생은 조금 전에 그 책을 읽었다.

위의 예문들에서 ‘읽’은 어간이고, ‘는’이나 ‘었’은 형식적인 기능을 하는 서술 보조소들이다.

2.1. 한국어의 용언

한국어의 용언은 기본적으로 동사, 형용사, 존재사, 지정사 등의 네 가지로 나누어 볼 수 있다.¹ 이 네 가지 범주는 구문론적으로나 의미론적으로 서로 다른 특성을 보인다. 그 가운데에서 동사가 가장 대표적인 큰 범주이

며 다음으로 큰 범주는 형용사이다. 존재사와 지정사는 몇 개 낱말이나 의 존형태로 되어 있으나 독특한 문법적 기능을 드러내는 범주로서 앞의 동사나 형용사와는 다른 특성을 보인다. 이 네 가지 범주의 낱말들은 의미 또는 문법적인 특성으로 보아 약간의 차이가 있으나 문장에서 서술 기능을 담당한다는 점에서는 공통적이다.

2.1.1. 동사와 형용사

동사는 대표적인 용언이다. 동사는 명사와 함께 문장 성분의 가장 중요한 줄기를 이룬다. 명사구 또는 주어의 핵심이 명사라면 동사는 동사구 또는 서술어의 핵심이다. 서술어를 이루는 용언들을 동사류어라 하는 것도 바로 동사가 차지하는 비중이 크기 때문이다. 동사는 용언의 한 가지로서 형용사, 존재사 및 지정사와 구분되는 특성을 지닌다. 동사는 의미론적으로나 구문론적으로 다른 용언들과 확실히 구분되는 특성을 가진다. 동사는 전통적으로 '사물의 움직임을 나타내는 낱말'이라고 정의된다(서정수 1994: 591). 이것은 동사라는 범주가 의미적 관점에서 비롯된 것임을 말해준다. 그러나 우리가 말하는 동사는 그리 간단하게 뜻매김되지 않고 구문론적으로도 독특한 면을 보인다.

형용사는 용언의 한 가지로서 동사와 마찬가지로 주어에 대한 서술 기능을 한다. 그런데 형용사는 주어의 동작이나 변화를 나타내는 동사와는 달리, 그 성질이나 상태를 나타내는 의미적 특성을 드러낸다. 형용사는 서술 기능 면에서는 동사와 마찬가지로 그 의미적인 면에서 상태성 서술을 드러내는 특성이 있는 것이다. 즉, 사물의 성질이나 상태를 나타내는 것이 형용사의 의미적 특성이다. 이런 점에서 형용사를 상태동사(stative verb)라고 부르기도 한다. 이렇게 형용사를 상태성 동사라 할 때에 동사는 동작 동사(action verb)라고 하여 대립적으로 기술된다. 한국어에서 형용사는 서술적(predicative) 기능이 우선하고 한정적(attributive) 기능은 부차적이다. 이것은 형용사가 본디 용언 또는 상태 동사라는 것을 뒷받침한다.

(3) 그 여자가 예쁘다.

(4) 예쁜 그 여자.

¹ 용언의 하위부류에 대해서는 다양한 의견이 있다. 많게는 동사, 형용사, 존재사, 지정사의 네 가지에서 적게는 동사와 형용사만을 용언의 범주로 인정하기도 한다. 본 논문에서는 의미론적 특성과 구문론적 특성을 모두 고려해 네 가지로 분류한 서정수 교수의 안을 바탕으로 이를 기계번역을 위해 동사, 형용사, 연사의 세 가지로 재분류할 것이다. 용언의 하위부류에 대해서는 서정수(1994:504-505)를 참조하라.

그런데 한국어에서는 일부 형용사가 동사와 형태나 의미가 비슷하여 서로 넘나들며 쓰이는 일이 있다. 즉 동일 형태의 어근을 지닌 형용사와 동사가 있는 것이다.

- (5) 달이 밝다. (형용사)
- (6) 날이 밝는다. (동사)
- (7) 그 사람은 너무 늙다. (형용사)
- (8) 너도 얼마 안 있으면 늙는다. (동사)

이들이 동사로 쓰일 때에는 문말 형태 “는/다”나 “고 있”과 어울리게 된다.

- (9) 이제 날이 밝는다.
- (10) 바야흐로 새 날이 밝고 있다. (서정수 1994:635)

이 무리에 드는 형용사는 늦다, 낫다, 크다, 맞다, 틀리다, 굳다, 못하다, 얇다, 흐리다 등과 같은 것들이 있다.

문장에서 서술 기능을 담당하는 낱말 범주를 용언이라고 정의할 때, 노어에서는 동사와 형용사 뿐 만 아니라 명사, 부사, 수사 등도 서술 기능을 담당할 수 있다. 그러나 노어는 형태부가 매우 발달된 언어로서 각 품사가 형태적으로 명확하게 구분되며, 동사와 형용사, 명사는 각기 그 어미변화 유형을 따르므로 품사 분류상에 어려움이 없다. 즉 노어에서의 동사와 형용사는 형태적으로 명확히 구분된다. 그러나 한국어에서 용언의 분류는 위에서 예로 든 문장들을 볼 때 그리 명확히 구분되지 않는다. 유목상 교수는 동사와 형용사를 정의함에 있어서 동사나 형용사란 이름 자체가 의미 중심적인 것이기 때문에 의미성에 따라 정의하는 것이 마땅하다고 주장하며, 동사를 ‘움직임을 나타내는 단어로 활용이 있는 품사’로, 형용사를 ‘성질이나 상태를 나타내는 단어로 활용이 있는 품사’로 정의한다. 예를 들어, ‘크다’라는 단어에는 ‘대’(大)의 뜻과 ‘성장’이라는 뜻이 있는데, 전자의 경우는 형용사, 후자의 경우는 동사이다 (유목상 1993:73). 이러한 경우에 노-한 기계 번역에서는 동음 이의어로 처리하고 있다. ‘대’의 의미를 갖는 ‘크다’는 이에 해당되는 노어 단어 *большой*와의 대응을 고려하여 형용사로 분류하고 형태 사전에 ‘크다 1’로 기입한다. 반면 ‘성장’의 의미를 갖는 동사 ‘크다’는 노어 동사 *вырасти*와의 대응을 고려하여 ‘크다 2’로 기입된다.

한국어에서 동사와 형용사를 변별하는 기준은 의미성이다. 어떤 어미는 수용하고 어떤 어미는 수용하지 않는다는 것은 어간의 의미자질에 의한 제

약일 뿐이다. 예컨대 동사 ‘달다’와 형용사 ‘달다’의 어미 선택이 다르고, 이에 따라 활용상에 차이가 나는 것은 어핵인 어간이 의미를 지배하기 때문이다. 따라서 동사와 형용사를 어미 변화의 특징에 의해 구분하는 것은 식별상의 하나의 편법일 뿐이다 (유목상 1993:73).

러-한 기계번역 시스템에서 동사와 형용사의 품사 분류는 기본적으로 의미적 기준에 따른다. 즉 상태성을 지닌 용언은 형용사로 분류한다. 그리고 형태부에서 효율성과 경제성을 고려해 동사와 형용사를 위한 표준 어미 목록을 따로 설정한다. 이 때 형용사와 동사를 위한 표준 어미 목록의 기준이 되는 것은 어미변화 상의 특성이다. 즉 형용사는 명령형과 청유형, 그리고 목적을 나타내는 서법, 그리고 진행상 등이 없다는 것이다.

2.1.2. 존재사와 지정사

존재사는 사물의 “있음”과 “없음”을 나타내는 낱말이다. 그 낱말 형태는 일반 형태인 “있다”, 존재 형태인 “계시다”와 부정 형태인 “없다”의 세 가지로 되어있다. 존재사의 일반 형태인 “있다”는 이 낱말 범주가 지니는 모든 문법적, 의미적 특성을 지니는 대표적인 형태이다.

- (11) 그 사람이 늘 집에 있다.
- (12) 그는 책이 많이 있습니다.
- (13) 저기 있는 사람이 내 아내입니다.
- (14) 너는 여기 좀 있어라. (서정수 1994:640)

(11)에서는 “있다”가 공간적으로 존재하는 상태를 나타내고, (12)에서는 ‘소유’ 상태를 나타낸다. (13)에서는 “있다”가 관형사형으로, (14)에서는 명령법으로 쓰인다. 이렇게 “있다”는 독특한 기능을 보인다.

존재사는 존재 상태를 나타내는 점에서 보면 형용사로 여겨진다. 의미적으로 기준 할 때 형용사 개념은 ‘성질이나 상태 또는 존재등을 나타내는 말’로 정의되며, 이 개념은 다시 성질 형용사, 상태 형용사, 존재 형용사 등으로 나눌 수 있다 (유목상 1993:94). 그러나 존재사는 어느 형용사와는 다른 구문론적 특성, 즉 동사성을 보이기도 한다. 실제로, 이 존재사는 형용사와 동사의 양면성을 보여 어느 범주에도 소속시키기 어려운 특성이 있다. 구문론적으로도 존재사는 동사와 동일한 특성이 있다. 그런 동사적 특성은 존재사를 형용사라고 여기는 처리 방식으로는 쉽사리 설명하기 어렵다. 무엇보다도 존재사는 동사가 나타나는 환경에서 흔히 쓰이기 때문이다. 존재사는 동작 부사와 자연스럽게 어울릴 수 있고, 명령법, 청유법, 약속법

형태로도 쓰일 수 있으며, 의도를 표현하는 형태들과 어울려 쓰이기도 한다는 점에서 동사적인 특성을 보인다. 반면 존재사는 동사성과는 전혀 다른 형용사성을 보이기도 한다. 존재사는 과거 시제나 완료상 및 진행상과 관련될 때는 형용사적인 특성을 강하게 드러낸다. 또한 존재사는 평서술형 형태와 어울려서 형용사적 특성을 드러낸다. 존재사가 동사와 마찬가지로 “는/ㄴ다”의 형태와 어울려 동사성을 드러낸다면, 다른 한편으로는 일반 형용사와 마찬가지로 술어적 “다”형태하고만 어울린다.

(15) 여기에 책이 많이 있다/ (?)있다/ *없는다.

(16) 할아버지께서는 지금 방에 계시다/ 계신다. (서정수 1994:647)

(15)에서 보는 바와 같이 “있다”와 “없다”는 “는/ㄴ다”와 결합하는 경우 일반적으로 어색하거나 전혀 불가능하므로 형용사성을 드러낸다고 할 수 있다. 이럴 때에는 주어진 존재 상태를 나타내므로 상태성을 강하게 보이게 때문이다.

반면에 지정사관 앞선 명사구와 결합하여 서술 기능을 드러내는 용언이다. 그 명사구는 서술 내용을 나타내는 보어(complement)이고 거기에 결합하는 지정사는 형식적 서술어 구실을 드러내는 것이다.

(17) 저 집이 우리 회관이다.

위의 문장에서 “이다”가 지정사이고 “우리 회관”은 보어이다. 지정사 “이다”는 보어에 덧붙여 쓰이는 의존 형태이며 서술 보조사와 결합하여 여러 가지 문법적 기능을 드러낸다. 그러나 이 지정사 자체는 실질적 의미가 없으므로 그 앞의 보어와 반드시 어울려야 한다. 그런데 실질적인 서술기능을 하는 것은 지정사 앞에 쓰인 보어 “우리 회관”이다. 즉, “저집”이라는 주어가 “우리 회관”에 해당함을 나타내는 것이다. 지정사 “이다”는 이런 실질적 서술에는 관여하지 않고, 그 보어에 서술어로서의 형식을 갖추게 하는 구실을 함과 동시에 뒤따르는 시제, 상, 서법 따위의 문법 요소를 연결시키는 일을 맡는다.

지정사는 서술 보조소와 결합하는 관계에서는 형용사와 거의 같다. 지정사는 시제/상 형태 “었”과 결합하여 단순 과거 시제를 나타내고, 영(零) 형태(∅)와 결합하여서는 현재 시제를 나타낸다. 진행상이나 완료상을 나타내는 문법 형태와 결합되지 않는 점에서도 지정사는 형용사와 같다. 평서술형 “다”와는 결합하지만 “ㄴ다”형태와는 결합하지 못한다.

- (18) 그 사람은 선생이었다/ 이다/ *인다.
 (19) *그 친구는 지금 막 선생이었다.
 (20) *그 친구는 한 동안 선생이고 있다.

또 “이다”는 의도법의 “겠”과 어울리지 못하고 약속법, 명령법, 청유법 및 의도형 연결 형태와 함께 쓰일 수 없는 점에서도 형용사와 같다.

2.1.3. 기계번역을 위한 용언의 재분류

앞에서 한국어의 용언은 동사, 형용사, 존재사, 지정사의 네 가지로 분류된다고 하였다. 그러나 노-한 기계번역 시스템에서는 학교문법이나 기존의 국문법과는 다른 품사분류 체계를 도입하고 있다. 노어와의 대응관계를 고려하여 한국어의 품사를 명사, 수사, 동사, 형용사, 연사, 조동사, 부사, 접속사, 감탄사, 후치사의 10개로 분류한다.² 이러한 분류에서 연사(copula)는 한국어 문법에서 지정사 혹은 계사라고 일컬어지는 ‘이다’와 ‘아니다’ 등의 낱말을 가르키는 것으로 노어의 *быть, являть* 또는 *-(тире)*에 대응된다.

존재를 의미하는 어휘 ‘있다’³와 그 부정 형태인 ‘없다’는 동사로 분류되었다. 이것은 이 단어 자체가 갖는 의미에 따른 것이 아니라, ‘있다’의 활용상의 형태적 특성 때문이다. 즉 존재사 ‘있다’가 진행상을 갖지 않는다는 점을 제외하고는 명령형과 청유형, 약속법등의 서법 형태를 가지는 등 다른 동사와 동일한 용법을 보인다. 부정 형태인 ‘없다’는 쓰임새에 있어서 ‘있다’와는 다르지만 긍정 형태 ‘있다’와의 관계를 고려해 동사로 분류한다. 그러나 ‘없다’는 독특한 활용 패턴을 따르므로 한국어 형태부에는 ‘없다’를 위한 독자적인 표준 어미 목록을 설정한다.

이와 같이 러-한 기계번역 시스템에서 한국어의 용언은 동사와 형용사 그리고 연사로 분류될 수 있으며, 이들 품사들은 각각의 표준 어미 목록을 갖게 된다.

²이에 대해서는 강덕수, 김성래의 “노-한 기계 번역에서 한국어 합성을 위한 품사 분류 체계 연구”를 참조하라.

³노-한 기계 번역을 위한 한국어 사전에서 ‘있다’는 동음 이의어로 두 개의 사전 항목에 기입된다. 진행상을 나타내는 보조 용언 ‘있다 1’은 조동사, 존재를 의미하는 ‘있다 2’는 동사로 기입되어 있다.

3. 한국어 용언의 활용어미

문장 내에서 용언은 서술 기능을 담당하는데 서술 기능은 실질적인 풀이 작용과 문법적인 형식 면으로 나누어진다. 즉 용언의 서술 기능은 그 뒤에 덧붙여진 서술 보조소의 도움을 받는데, 서술 보조소는 시제(tense)와 상(aspect) 그리고 서법(mood)을 가리키는 기능을 한다. 한국어에서는 서술 보조소가 문장의 끝에 나타나며 그 중에서도 서법 형태는 가장 끝에 위치하고 시제나 상 형태가 그 앞에 온다. 서술 보조소는 단순히 용언의 굴절어미라기 보다는 문장 전체의 시제 관계나 양상 그리고 문장 전체의 성격이나 종류를 결정짓는 요소이다. 동사나 형용사 등의 어휘 범주는 문장의 주어에 관한 내용 서술을 담당하고, 서술 보조소는 주어와 서술 용언이 연결되어 이루어지는 문장 전체의 문법적인 양상을 마지막으로 결정짓는 구실을 한다. 이 서술 보조소에 따라 문장은 현재, 과거, 진행, 완결 등의 시제/상적 기능과 함께, 서술문, 의문문, 명령문 등의 갖가지 서법 양태를 갖추게 된다(서정수 1994:189). 용언의 문법적 기능을 나타내기 위하여 어미 변화하는 것을 활용(conjugation)이라 한다. 용언의 굴절이 어미의 교체로 이루어질 때, 그 교체되는 어미를 활용어미라 한다. 여기는 단어의 뜻을 나타내고 활용어미는 주로 문법적 관계를 나타낸다(이철수 1994:108).⁴ 용언의 활용어미는 문법적 기능에 따라 종지법 어미와 연결법 어미, 그리고 전성법 어미로 나누어진다.⁵ 종지법 어미는 문장을 끝맺는 역할을 하는 것으로 서술, 의문, 명령, 청유, 응낙, 감탄 등을 나타낸다. 연결법 어미는 문장을 끝맺지 못하고 다음 말에 연결되는 것으로 나열, 선택, 비교, 목적 등 다양한 의미를 표현한다.

(21) 고기를 잡으려 냇가에 간다. <목적>

전성법 어미는 용언이 다른 품사처럼 쓰이게 한다.

(22) 책을 보는 사람이 많다.

(23) 가기도 잘도 간다 서쪽 나라로.

(22)의 어미 '는'은 용언이 체언을 꾸미는 관형어의 구실을 하도록 하는

⁴앞으로 편의상 서술 보조소를 통칭하는 말로 활용어미라는 용어를 사용할 것이다.

⁵활용어미의 유형에 관해서는 이철수(1994:109-112)를 참조하라.

관형사형 어미이고, (23)의 ‘기’는 용언이 명사와 같은 구실을 하게 하는 명사형 어미이다.

3.1. 시제, 상, 서법

시제란 사건이나 상태의 시간적 위치를 문법 형태로 구분하여 나타내는 문법 범주이며 일반적으로 현재, 과거 및 미래의 세 갈래로 구분되나 언어에 따라 달리 나뉘이기도 한다. 한국어에서는 과거와 현재 시제의 문법 형태는 갖추어졌다고 할 수 있으나 미래 시제는 일정한 문법 형태가 대응되지 않는다. 미래 시간의 표시는 부사어나 그 밖의 어휘적 도움으로 드러나므로 실제로는 이루어질 수 있으나 시제 범주로 인정하기는 어렵다. 즉, 한국어에는 “과거 시제”와 “현재 시제”에만 그 나름의 독자적인 시제 범주가 존재한다 (서정수 1994:196). 이러한 사실에 근거하여 노-한 기계 번역을 위한 형태부에도 현재와 과거 두 시제만 설정된다.

상은 한 사태가 일정한 시역(時域) 안에서 시간적으로 변화하는 양상을 나타내는 문법 범주이다. 일정한 시역이란 한 사태의 시간적 위치 곧 시제적으로 지정된 시간 영역을 가리킨다. 시간적 변화 양상이란 문제의 사태가 정해진 시간 안에서 어떤 모습으로 바뀌는지를 가리킨다. 이것은 기본적으로 상황의 완결 또는 미완결의 양상으로 파악되는데, 관점에 따라서는 그 밖의 여러 양상으로 파악되기도 한다. 상이 문법 범주라는 것은 그것이 주로 문법 형태로 표시된다는 말이다. 곧 상이란 현재, 과거 또는 미래의 한 시제 영역에서의 상황에 대하여 완결상(perfective)과 미완결상(imperfective) 또는 그 밖의 여러 양상으로 파악되는 바를 문법 형태로 나타내는 것이라고 할 수 있다 (서정수 1994:204).

서법이란 문장의 내용에 대하여 말하는 이가 지니는 정신적 태도를 나타내는 문법 범주이다. 일반적으로 말을 할 때에는 낱말 따위 구성 요소들을 구문 규칙에 따라 배열하여 문장의 내용을 만들면서 여기에 말하는 사람 자신의 심리적 태도를 결들여지게 된다. 예를 들면, “비가 오겠다”라는 말에서 “비가 오”까지는 문장의 내용이고, “겠다”라는 형태소는 말하는 이 자신이 그 내용을 미루어 짐작하고 있다는 것을 나타낸다. 이렇게 우리가 말을 할 때에는 반드시 그 내용에 대한 자신의 태도가 결들여지는 데, 이런 태도를 드러내는 문법 범주를 서법이라 한다. 이 문법 범주는 용언의 굴절 형태나 어미 또는 문말 형태 따위로서 서법적 의미(modality)를 나타내는 형태들을 가르킨다.

한국어에서는 대개 문장의 끝부분에 나타나는 문법 형태들로 서법이 표시된다.

- (24) 비가 온다. [-니다]
- (25) 비가 오겠다. [-겠다]
- (26) 비가 오지? [-지?]
- (27) 비야 비야 오너라. [-너라]
- (28) 비가 오는군. [-는군]

(서정수 1994:245)

서법적 문법 범주는 문말 형태와 비문말 서법 형태를 다 포함한다. 즉, 서술법, 의문법 등의 종결법이나, “겠”, “더” 따위의 선어말 어미의 비문말 형태를 통틀어 서법이라 한다.

3.2. 용언과 활용어미

용언은 뒤따르는 활용어미(=서술 보조소)와 결합하여 서술 기능을 온전하게 드러낸다. 용언은 그 자체로서 서술적 의미를 드러냄과 아울러 뒤따르는 활용어미와 결합하여 시제, 서법 등의 문법적 서술 기능을 갖추게 된다. 한국어의 활용어미 형태들의 대부분은 각 범주의 용언들과 두루 어울리지만 일부는 특정한 범주의 용언하고만 결합할 수 있다. 용언과 활용어미와의 결합관계를 표로 나타내면 다음과 같다.

(29) 서술 보조소의 종류와 형태

범 주	시제/상/서법 형태	결 합 용 언		
		동 사	존재사	형용사/ 지정사
1) 시제/상		○	○	○
	고 있, {어} 있	○		
2) 추정법	겠, 르것어, 리	○	○	○
3) 의도법	겠, {르}껴	○	○	
4) 알림법	{더}	○	○	○
5) 서술법	다, 네, {어}{요}, 지{요}, 오소, (스)비니다	○	○	○
	는/니다, 는구나, 는군{요}	○	{○}	
	구나, 군{요}		○	○

6) 의문법	나, 느냐, {어}(요), 지(요), 르까(요), 요, 뵈니까	○	○	○
	느냐, 논가(요), 나가(요)	○	○	
7) 명령법	{어라}, 라, 게, {어}(요), 지(요), (시)오, 십시오, 럼, 려무나, 구려	○	○	
8) 청유법	자, 세, 시오, {어}(요), 지(요), 뵈시다, 십시다	○	○	
9) 약속법	마, 으세, 리다	○	○	

(서정수 1994:511)

용언은 비종결 문법 기능소(연결법 어미와 전성법 어미)와 결합하여 뒤따르는 문장 성분과 이어진다. 비종결 문법 기능소란 관형화소(관형사형 어미), 명사화소(명사형 어미), 접속 기능소(연결 어미) 등을 일컫으며, 용언에 덧붙여서 뒤따르는 문장 성분과 문법적 결합관계를 보인다. 비종결 문법 기능소들은 각 범주의 용언들과 두루 결합하는 것도 있고, 특정한 용언하고만 어울리는 것들도 있다. 비종결 문법 기능소와 용언과의 결합관계를 표로 나타내면 다음과 같다.

(30) 비종결 문법 기능소의 갈래와 형태

범 주		형 태	선 행 용 언		
			동 사	존재사	형용사/지정사
1) 관형화소	현재	는	○	○	
		ㄴ			○(ㄴ)
	과거	ㄴ	○	(○)	
	추정	르	○	○	○
2) 명사화소		ㅁ, 기	○	○	○
3) 접속화소	병렬	며, 면서, 고	○	○	○
	대립	나, 지만	○	○	○
	선택	거나, 든지	○	○	○

3) 접속화소	상황	니, 되	0	0	0
		는데	0	0	
		ㄴ데		(0)	0
	계기	{어}{서}	0	0	0
	까닭	니(까), 므로, {어}서, 거든, 기에, 길래	0	0	0
		느라고	0	(0)	
	조건	면, 거든, {어}야	0	0	0
	양보	{어}도, 더라도, 르망정, 르지언정, 르지라도, ㄴ들	0	0	0
	목적	려고, 고 자, 러	0	0	
	도달	도록, 게	0	0	(0)
	전환	다가	0	0	0
	첨가	르뿐더러 /아니라	0	0	0
	점중	르수록	0	0	0
	비교	거든, 듯 (이)	0	0	0
연발	자(마자)	0	(0)	(0)	

4. 기계번역을 위한 한국어 용언의 활용어미 합성 모델

4.1. 한국어 용언 활용형 목록

4.1.1. 동사

앞장에서 용언과 종지법 어미, 연결법 어미, 및 전성법 어미의 결합관계를 살펴보았다 (표 29, 30). 그러나 노-한 기계번역을 위한 한국어 형태부에는 위에서 제시된 모든 시제와 서법 활용형들이 다 사용되고 있는 것은 아니다. 본 작업은 러시아어 과학 기술 텍스트의 번역을 목표로 하므로 존대법 형태는 모두 제외시키고, 서법의 여러 변이형들 중 대표적인 형태 하나만을 기입하고 있다. 한국어 동사 활용형을 기술하기 위해 현재 사용하고 있는 자질들은 28개 형태이지만, 앞으로 이 목록은 더 정확하고 섬세한 번역을 위해 추가 또는 수정될 수 있다. 시제와 상에 있어서는 {-있었다} 등과 같은 형태는 {-었다}로 사용하기에 무리가 없으므로 현재와 과거 두 시제 범주만 사용하고 있다. 다음의 목록은 노-한 기계번역 시스템에 구축된 한국어 동사 활용 표준 목록이다.

(31) 한국어 동사 활용 표준목록

No.	자질 세트	활용 어미
1	mf	다
2	declar, prs	ㄴ다
3	declar, pst	였다
4	interr, prs	는가
5	interr, pst	였는가
6	condit, prs	ㄴ다면
7	condit, pst	였다면
8	sbj, pst	였을
9	expl, prs	니
10	expl, pst	였으니
11	altern, prs	거나
12	altern, pst	였거나
13	imper	여라
14	atrb, prs	는

No	자질세트	활용어미
15	atrb, pst	었던
16	gernd, prs	며
17	za-form	자
18	reo-form	러
19	zi-form	지
20	l-form	르
21	ya-form	여야
22	ryeogo-form	려고
23	ge-form	게
24	go-form	고
25	conj, prs	ㄴ다는
26	conj, pst	였다는
27	contr, prs	지만
28	contr, pst	였지만

위의 목록에서 17번 항목의 za-form은 한국어에서 '-자'로 실현되는 것인데, 이는 동사의 청유형이나 연발의 의미를 갖는 접속화소를 실현시키는 형태이다. 18번 항목의 reo-form(러-형)은 목적을 나타내는 형태이고, 19번 항목의 zi-form(지-형)은 부정문에서 '-지 않다'를 위해 사용된다. 그리고 20번 항목의 l-form(르-형)은 노어의 미래 시제를 한국어로 번역할 때 사용된다.

(32) 쓰다 fut =====> 쓰다 l-form 것이다 mf =====> 쓸 것이다

(33) 하다 neg, pst, progr =====> 하다 ko-form 있다 zi-form 않다 mf
=====> 하고 있지 않다.

한국어 용언의 명사화소는 “口”과 “기”의 두 가지인데, 전자에는 “음”이라는 변이형이 있다. 이들은 모든 용언에 공통적으로 쓰인다.

(34) 입음, 좋음, 있음, (책)입

(35) 먹기, 예쁘기, 있기, (책)이기

명사화(nominalization)란 서술 용언을 명사형으로 바꾸어 명사의 기능을

하도록 하는 동명사(gerund), 명사형 또는 동사적 명사(verbal noun)를 이루는 절차이다. 명사형 “(으)ㄴ”과 “기”사이의 의미적인 차이가 있지만, 여기에서는 그 형태적 활용형을 논하므로, 의미적 특징에 대해서는 언급하지 않겠다. 현재 기계 번역을 위한 한국어 형태부에는 용언의 명사화소 ‘-기’형만 기입하고 있다. 명사형 ‘-기’형은 그 자체가 명사처럼 변화하므로 이를 위한 명사 변화 표준 목록을 별도로 설정하고 있다.

(36) 용언의 명사형 변화형 목록(Z:022)

자 질 세 트	활 용 어 미
genr	#
nom1	가
nom2	는
nom3	도
gen1	의
gen2	부터
gen3	까지
ess1	로 부터
acc	를
comit	와

따라서 한국어 형태부에서 동사는 보통 28개의 어미변화형과 명사형 ‘-기’를 포함한 10개의 ‘-기’ 변화형을 가지므로 (38)에서와 같이 총 38개의 활용형을 갖는다.

4.1.2. 형용사와 연사

형용사 활용은 동사활용 목록에서 목적을 나타내는 ‘-러’형, ‘-려고’형, ‘-자’형과 명령형을 제외한 활용형을 갖는다. 그리고 형용사는 진행상을 갖지 않는다. 그러나 활용 어미 목록의 ‘-고’형은 진행상 뿐 아니라 병렬 접속의 기능도 하므로 ‘-고’형은 그대로 기입된다. 따라서 하나의 형용사 어휘는 모두 34개의 활용형을 갖는다.

연사 ‘이다’와 ‘아니다’를 위한 별도의 활용 표준 목록은 없다. 연사가 갖는 문법 범주는 형용사와 동일하기 때문에 형용사의 활용 표준 목록을 그

대로 이용한다.

4.2. 표준 어미 목록

4.2.1. 동사

(31)에서 동사의 활용 표준목록을 열거하였다. 그러나 모든 한국어 동사가 이 목록에 기입된 활용형에 따라 변화하는 것은 아니다. 이것은 단지 활용형의 목록을 정하기 위한 기준일 뿐이다. 따라서 한국어의 동사들을 그 활용 유형에 따라 몇 그룹으로 분류할 필요가 있다. 각 그룹을 대표하는 하나의 동사들을 선정해 표준어미목록을 작성한다. 즉, 표준어미목록은 활용 유형이 동일한 동사군의 어형변화를 제시하는 것이다. 표준어미목록은 Z로 표시하며 이 철자 뒤에 어형 변화 유형이 숫자로 표시된다. 현재 동사의 표준어미목록은 6개가 작성되어있다. 이 목록은 앞으로 계속 보충될 수 있다.

(37) 한국어 동사의 표준어미목록

- Z:001 ['일하다' 유형의 동사]
- Z:002 ['제안하다' 유형의 동사]
- Z:003 ['남다' 유형의 동사]
- Z:008 ['떠나다' 유형의 동사]
- Z:009 ['알다' 유형의 동사]
- Z:025 ['돕다' 유형의 동사]

표준어미목록 Z:001은 '-하다'로 끝나는 동사들을 위한 것으로 예를 들면 다음과 같다.

(38) '일하다' 유형의 어미 유형 모델(Z:001)

자질세트	활용형태
mf declar, prs declar, pst interr, prs interr, pst condit, prs condit, pst sbj, pst expl, prs expl, pst altern, prs altern, pst imper	일하다 일한다 일하였다 일하는가 일하였는가 일한다면 일하였다면 일하였을 일하니 일하였으니 일하거나 일하였거나 일하여라
atrb, prs atrb, pst gernd, prs za-form reo-form zi-form l-form ya-form ryeogo-form ge-form go-form conj, prs conj, pst contr, prs contr, pst	일하는 일하였던 일하며 일하자 일하러 일하지 일할 일하여야 일하려고 일하게 일하고 일한다는 일하였다는 일하지만 일하였지만

'subs'	
' genr	일하기 #
nom1	일하기가
nom2	일하기는
nom3	일하기도
gen1	일하기의
gen2	일하기부터
gen3	일하기까지
ess1	일하기로부터
acc	일하기를
comit	일하기와

4.2.2. 형용사

동사를 위한 표준 어미 목록이 6개 작성되어 있는 반면, 형용사를 위한 표준 어미 목록은 5개가 있다.

(39) 형용사 표준어미목록

- Z:006 ['새롭다' 유형의 형용사]
- Z:007 ['중요하다' 유형의 형용사]
- Z:023 ['예쁘다' 유형의 형용사]
- Z:024 ['작다' 유형의 형용사]
- Z:026 ['굵다' 유형의 형용사]

'새롭다' 유형의 동사들이 갖게 되는 어미 목록은 다음과 같다.

No	자질 세트	활용 어미
1	mf	새롭다
2	declar, prs	새롭다

No	자질세트	활용어미
3	declar, pst	새로웠다
4	interr, prs	새로운가
5	interr, pst	새로웠는가
6	condit, prs	새롭다면
7	condit, pst	새로웠다면
8	sbj, pst	새로웠을
9	expl, prs	새로우니
10	expl, pst	새로웠으니
11	altern, prs	새롭거나
12	altern, pst	새로웠거나
13	atrb, prs	새로운
14	atrb, pst	새로웠던
15	gernd, prs	새로우며
16	zi-form	새롭지
17	l-form	새로울
18	ya-form	새로워야
19	ge-form	새롭게
20	go-form	새롭고
21	conj, prs	새롭다는
22	conj, pst	새로웠다는
23	contr, prs	새롭지만
24	contr, pst	새로웠지만

4.3. 은폐형

은폐형은 표준 어미 목록에 속하지 않는 불규칙한 어형변화를 처리하기 위한 개념으로, M(masks)로 표시된다. 이것은 어떤 어휘 항목이 예외적인 어형들을 가져 하나의 표준 어미 목록만을 선택할 수 없을 때, 목록들 중 가장 근접한 패러다임을 가진 표준 어미 목록을 선택하여 이 가운데 일치하는 어형들은 그대로 실현되게 하고, 일치하지 않는 어형들은 실현되지 못하도록 은폐시키는 역할을 한다.

은폐형에서 숫자 1에 해당하는 어미 항목은 표준목록과 일치하여 그대로 실현되고, 0은 표준목록에서 불필요한 형태들을 은폐시킨다. 이 숫자들의 배열 순서는 표준 어미 목록에 열거된 자질들의 순서와 일치해야 한다. 예를 들면, 형용사 '새롭다'를 위한 표준 어미 목록 Z:006이 작성되어 있지만, 이 형용사는 어형 변화할 때 '새롭-'과 '새로-'라는 두 개의 어간이 필

요하다. 즉 위의 표준 어미 목록은 어미 형태를 완벽하게 기술하고 있지만 어간의 변화에 대해서는 언급할 수 없다. 따라서 형용사 '새롭다'는 완전 어간인 '새롭-'과 불완전 어간인 '새로-'의 두 경우를 조합할 수 있는 은폐 목록(mask:M)을 필요로 한다. 그 목록은 다음과 같다.

(41) '새롭다' 유형 형용사를 위한 은폐 목록

- M:008 ['새롭/다' 유형의 형용사 완전 어간: 새로-]
 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0
 M:009 ['새롭/다' 유형의 형용사 불완전 어간: 새로-]
 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1

은폐목록이 필요한 이유는 형태분석을 효율적으로 처리하기 위해서이다. 은폐목록을 이용하면 불규칙한 어미 형태를 갖는 단어의 경우 불규칙한 어미 형태를 사전에 일일이 반복적으로 나열해야 하는 비경제성을 배제할 수 있다. 표준 어미 목록과 은폐형 목록을 통하여 '새롭다'의 활용형이 어떻게 실현되는지를 다음의 예가 잘 보여 줄 것이다.

(42) 형용사 '새롭다' 활용형 합성 모델

	표준어미목록 (Z:006)	은폐형 M:008	'새롭-' 어간 실현형	은폐형 M:009	'새로-' 어간 실현형
--	-------------------	--------------	-----------------	--------------	-----------------

1	다	1	새롭다	0	---
2	다	1	새롭다	0	---
3	웠다	0	---	1	새로웠다
4	운가	0	---	1	새로운가
5	웠는가	0	---	1	새로웠는가
6	다면	1	새롭다면	0	---
7	웠다면	0	---	1	새로웠다면
8	웠을	0	---	1	새로웠을
9	우니	0	---	1	새로우니
10	웠으니	0	---	1	새로웠으니
11	거나	1	새롭거나	0	---
12	웠거나	0	---	1	새로웠거나
13	운	0	---	1	새로운
14	웠던	0	---	1	새로웠던
15	우며	0	---	1	새로우며
16	지	1	새롭지	0	---
17	울	0	---	1	새로울
18	워야	0	---	1	새로워야
19	게	1	새롭게	0	---
20	고	1	새롭고	0	---
21	다는	1	새롭다는	0	---
22	웠다는	0	---	1	새로웠다는
23	지만	1	새롭지만	0	---
24	웠지만	0	---	1	새로웠지만

4.4. 관습형

몇 가지 어미 형태들이 여러 단어들에 공통적으로 나타나는 경우 이 어미 형태들을 하나의 묶음으로 조합하여 집합적으로 표시해 주면 어미 활용 체계를 단순화하는 한편 기계 번역상 형태적 분석과 합성 과정에서 효율성을 높일 수 있다. 이를 위해 고안된 규칙이 관습형이다. 이 관습형은 T(trafarets)로 표시된다. 현재 한국어 형태부에서는 28개의 관습형이 만들어져 사용되고 있다.

4.4.1. 동사를 위한 관습형의 예

동사의 관습형은 보통 품사와 표준 어미 목록 번호, 그리고 동사에서 파생된 명사형의 격변화를 위한 별도의 항목을 갖는다. 예를 들어, 규칙 동사 '일하다' 유형은 다음과 같은 관습형 T:001을 갖는다.

(43) T:001 ['일하다' 유형의 규칙 동사]

cth:V, nlz:001

cth:V, ly: 'gi'subs, nlz:22

위의 관습형은 '일하다'와 같은 활용 유형에 속하는 동사들을 위한 관습형이다. Cth는 품사를 표시하는 것으로 V는 동사를 의미한다. 이 동사는 표준 어미 목록(nlz)001을 따른다. 그 다음 행의 ly는 어미를 표시하는 것으로 subs(명사형)의 어미가 '기'임을 명기해 주고, 이 '기'형은 표준 어미 목록(nlz)022에 따라 변화한다는 것을 나타낸다. 따라서 '일하다' 유형의 동사들의 사전 항목에는 품사, 어미 목록, 명사형 변화가 일일이 열거될 필요가 없이 T:001만 기입되면 필요한 모든 항목을 불러서 쓸 수 있다.

4.4.2. 형용사를 위한 관습형의 예

예를 들어, 형용사 '새롭다'를 위한 관습형에는 '새롭-' 어간을 위한 것(T:022)과 불완전 어간 '새로-'를 위한 것(T:023)의 두 가지가 있다.

(44) '새롭다'의 관습형

T:022 [형용사 '새롭다'의 완전 어간 '새롭-']

cth:A, nlz:6/8

cth:A, ly: 'gi'subs, nlz:22

T:022 [형용사 '새롭다'의 불완전 어간 '새로-']

cth:A, nlz:6/9

T:022의 nlz:6/8은 표준어미목록Z:006에 은폐목록 M:008을 씌우라는 것이다. 그래서 표준어미 목록에 수록된 형태들 중에서 은폐목록 008에서 숫자 0으로 표시된 항목의 어형은 어간이 '새롭-'일 경우에는 실현되지 않는다.

5. 결론

본 논문은 노-한 기계 번역에서 노어와 한국어의 동사와 형용사 체계의 차이를 극복하여 기계 번역에 알맞도록 한국어의 용언을 체계화하는 것이다. 노어에서 동사와 형용사는 우선 형태적으로 확실히 구분되지만, 한국어에서는 애매한 점이 많다. 예를 들어 노어의 형용사 *большой*와 동사 *вырасти*는 한국어에서 둘 다 '크다'이다. 이런 점은 기계 번역 상에 어려운 문제 중 하나로 지적 될 수 있다. 이 경우 한국어에서 의미적으로 '크다 1'과 '크다 2'로 구분함으로써 해결의 실마리를 찾을 수 있다. 또한 기존 한국어 문법에서 논란이 되어 온 존재사와 지정사도 연사와 조동사로 개념을 재정립함으로써 동사와 형용사의 구문을 분석하기가 용이하게 하였다.

이와 함께 동사와 형용사의 다양한 활용형을 시제, 상, 서법 기능에 따라 활용 어미로 분류함으로써 한국어 용언 체계를 기계 번역에 알맞게 분석하였다. 이 결과 용언의 변화형을 위한 표준 어미 목록 작성이 가능해졌다. 활용 어미는 모두 38개로서 이것들이 조합되는 표준 어미 목록은 동사가 6개, 형용사가 5개이다. 이러한 표준 어미 목록을 약간씩 벗어나는 불규칙형들도 은폐형 장치로 규칙화할 수 있음을 보여 주었다. 마지막으로 여러 규칙들이 공통적으로 나타나는 현상을 하나의 집합으로 묶어 관습형을 만들어 내었다. 이러한 장치들에 의해 노-한 기계 번역은 신속하고도 정확한 번역 결과를 얻을 수 있게 된다.

본 논문은 노-한 기계 번역에서 최종 단계에 나타나는 한국어의 형태적 합성에 초점을 맞추었다. 그 결과 한-노 번역을 위한 한국어의 형태적 분석에는 다소 한계가 있을 수 있다. 그러나 본 논문에서 도입한 개념들을 활용하면 복잡해 보이는 한국어의 형태소들을 체계적으로 분석하는 것이 그렇게 어려운 일은 아닐 것이다.

참고문헌

- 강덕수, 김성래, 김연수, 박창호. 1996. 러시아어-한국어 기계번역 시스템 (2), 과학기술처.
- 강덕수, 김성래, 박창호. 1994. 러시아어-한국어 기계번역 기반기술에 관한 연구, 과학기술처.
- 서정수. 1994. 국어 문법, 서울:뿌리깊은 나무.
- 유목상. 1993. 한국어 문법의 이해, 서울:일조각.
- 이철수. 1994. 국어 형태학, 인하대학교 출판부.
- 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," МЕТА, Vol. 37, No. 1, pp. 99-112.

Апресян, Ю. Д. и др. 1989. Лингвистическое обеспечение системы ЭТАП-2, Москва:Наука.

서울특별시 동대문구 이문동 270

한국의국어대학교 노어과

130-791

E-mail: kangds@maincc.hufs.ac.kr

Fax: +82-2-968-0330

접수일자: 1998. 10. 13.

게재일자: 1999. 2. 27.