

확장된 동사형을 이용한 영어문장 검사기

차 의 영*, 김 영 태**

· 부산대학교 자연대학 전자계산학과

** 서울대학교 공과대학 컴퓨터공학과

English Sentence Critique Using Extended Verb Pattern

Eui Young Cha* and Young Taek Kim**

· Dept. of Computer Science, Pusan National University

** Dept. of Computer Engineering, Seoul National University

요 약

변환 방식의 기계 번역에서 가장 중요한 부분은 변환 단계이며 여기서 변환사 전이 매우 중요한 역할을 담당한다. 그러므로 인간이나 기계 번역기에 의해 생성 되는 영어 문장은 이들이 가지고 있는 동사 사전의 내용과 효율적인 생성 알고리 즘에 의해서 문장의 수준이나 정확성이 결정된다. 이렇게 생성된 문장을 검사하는 기존의 영어 문법 검사기들은 영어권의 사람들을 위주로 만들어졌기 때문에 문법 적인 중요한 규정들을 포함하지 않고 있어서 비영어권의 사용자가 이용하기에는 부적절하다. 본 논문에서는 인간이 번역하였거나 기계 번역기에 의해 생성된 문장 을 검사하고 교정할 수 있도록, 확장된 동사형을 기반으로한 동사 사전을 제안하 고 이를 이용한 영어 문장 검사기를 구현한다.

I. 서 론

외국어를 배우거나 말하려는 사람들에게 있어 가장 어려운 것은 그 언어의 문법을 정확 히 공부해야하는 것이다. 영어의 경우 영어를 말하는 영어권의 사람들이 관습적으로 이야 기하는 것들도 비영어권의 사람들에게는 어려운 문법으로 비춰질 수가 있다. 마찬가지로 영한 혹은 한영번역에서도 이러한 관습적인 문제들도 상세히 번역의 과정에 포함되어야 할

것이다.

기계번역은 사전(dictionary), 문법 규칙(grammar rules), 재구성 규칙(rewriting rules)과 같은 정확하고 복잡한 지식(knowledge)를 요구한다. 이러한 지식을 구축하기가 어렵기 때문에 단순히 Chomsky의 문법 등과 같이 언어학적인 이론만으로 실제적인 기계번역 시스템을 구현하기에는 많은 문제가 있다. 새로운 규칙을 추가하는 효과를 기대하기가 어려울뿐만 아니라 대규모의 규칙기반 시스템(rule-based system)을 사용하여 번역하는 것은 시간이 많이 소모되기때문에 번역에 있어 수행속도를 개선시키기가 어렵다[2, 4, 5, 6]. 그러므로 이러한 시스템의 단점을 개선하고 실제적인 기계번역 시스템을 만들기 위해서 collocation에 기반한 트랜스퍼 사전(transfer dictionary)을 이용한 기계번역 시스템과 예제(example)나 유사성(analogy)에 기반한 기계번역 시스템들이 제안되었다[4, 5].

이러한 기계번역 시스템중에서 한영 기계번역의 경우나 인간에 의한 직접 번역의 경우의 어떤 경우라도 단순한 사전적인 의미의 번역은 종종 영어 문장의 좀 더 정확한 어순이나 문법적인 구조를 도외시하는 경우가 많다. 그러므로 기계번역기나 인간의 번역에서 생성된 문장을 구조적이고 의미적 문법에 기반하여 검정하고 좀더 세련된 문어체의 문장으로 고쳐 줄 필요가 있다.

기존의 문법 검사기(grammar checker, critique)는 영어권의 사람들을 중심으로 되어있는 관계로 비영어권의 사용자에게는 심각하게 느껴지는 문법적인 문제를 사사롭게 넘어가고 있다. 지금 많이 사용되고 있는 소프트웨어로는 Grammatik IV, Express, Microsoft사의 Word 2.0 등이 있는데, 기계번역기나 인간에 의해 생성된 다음과 같은 영어 문장을 시험하면 이러한 문법 검사용 소프트웨어는 모두 올바른 문장으로 판정해준다.

- (1) A. I have always dreamed to visit Korea.
B. I have always dreamed of visit Korea.
C. I have always dreamed of visiting Korea.
- (2) A. I have been born in a town just outside Paris.
B. I was born in a town just outside Paris.
- (3) A. This causes that the children look for a cake.
B. This causes the children to look for a cake.

위의 예의 경우, 물론 영어권의 사용자에게는 별다른 문제가 되지 않을지 모르지만, 이들 소프트웨어가 철자교정이나 간단한 단어의 교환 등을 지시하는 수준인 만큼, 비영어권의 사용자들을 위한 문법 검사기로는 적절치 못하다.

본 연구는 이러한 기존의 문법 검사기가 가지고 있는 문제점을 해결하고 비영어권의 사

사람들이 영어문장을 작성하거나, 한영 기계번역시 영어문장의 생성에 사용되거나 생성된 영어문장의 검사에 필요한 영어의 동사형(verb pattern)에 기반한 동사사전 구조를 제안하고 이를 이용한 문법 검사기를 구현한다.

II. 동사 사전을 위한 구성요소

인간이나 기계에 의해 번역된 문장이 잘못되었을때, 잘못된 문장을 고쳐주기 위해서는 많은 지식을 필요로 한다. 잘못된 문장은 철자, 수동태, 도치, 동사형 등 고쳐야할 많은 문제점을 포함하고 있다.

- (4) A. It is not allowed to talk in the library.
B. People are not allowed to talk in the library.
C. Talking in the library is not allowed.
- (5) A. We are considering to visit Switzerland next year.
B. We are considering visiting Switzerland next year.
- (6) A. That these people commit crimes is not surprising me.
B. That these people commit crimes does not surprise me.

위의 영어문장의 경우 (4)B, C, (5)B, (6)B가 옳은 문장이다. 이 경우 (4)A, (5)A, (6)A로 생성되지 않도록 제한하기 위한 정보가 있어야 된다. (4)A의 경우 allow에서는 예비의 it가 사용될 수 없다는 제약때문이고, (5)A의 경우 consider는 동사의 -ing와 결합된다는 제약때문이며, (6)A는 동사 surprise는 진행형으로 쓰일 수 없다는 제약때문이다[1,9,10,11].

그러므로 올바른 문장을 생성해내기 위해서는 이러한 정보를 활용해야하는데, 이를 위해서 각 동사들의 제약조건들을 포함하는 효율적인 동사사전이 구성되어야 한다. 그리고 이 동사사전에 포함되는 제약조건들은 대부분 사전으로 부터 얻어지지만, 사전에 없는 정보들은 여러 소오스로 부터 모아져서 오랜시간에 걸쳐 입력되어야 한다.

이와 같은 많은 문법적인 제약요소들을 포함하는 효율적인 동사사전을 구성하기 위하여 필요한 용어를 정의하면 다음과 같다.

[정의 1] 확장된 동사형(Extended Verb Pattern) :

Hornby의 25가지 동사형을 중심으로 더욱 세분화한 50여가지의 동사형을 기본 동사형(basic verb pattern)이라하고, 여기서 파생변화된 형태의 동사형 30여가지를 변형 동사형(transformational verb pattern)이라 하며, 이를 통틀어 약 90여가지의 동사형을 확장된 동사형이라고 말한다.

(예 1) 기본 동사형(Basic Verb Pattern) :

- VP1 주어+be+보어+부사(구)
- VP2A 주어+자동사
- VP2B 주어+자동사+(for)+부사(구)
- VP2C 주어+자동사+부사(구)
- VP2D 주어+자동사+형용사/명사/대명사
- VP2E 주어+자동사+현재분사(구)
- ...

(예 2) 변형 동사형(Transformational Verb Pattern) :

- VP1-1 의문사+be+주어
- VP1-2 There+be+주어
- VP1-3 There+be+주어+부사(구)
- VP1-4 It+be+형용사/명사+to 부정사
- VP1-5 How/What+형용사/명사+(it be)+to 부정사(구)
- VP1-6 It+be+형용사/명사+동명사(구)
- VP1-7 It+be+형용사/명사+절
- VP1-8 It+be+형용사/명사+for+명사/대명사+to 부정사(구)
- VP2A-1 There+자동사+주어
- VP2A-2 It+seem/appear/happen/chance/follow+주어(that 절)
- VP2A-3 It+자동사+주어(절/부정사구)
- VP2C-1 There+자동사+부사(구)+주어
- ...

[정의 2] 제약조건(Constraints) :

제약조건이란 문장을 변형시키는 기본 조건과 어떤 주어진 동사가 가지는 특수한 문장 구조나 특정한 동사형을 가지는데 필요한 구성요소들의 조건을 말한다. 제약조건은 문장 구성요소들의 길이나 도치, 수동태 불가조건 등과 같은 일반 제약조건(**general constraints**)과 동사의 동사형을 제한하는 특수 제약조건(**special constraints**)이 있다.

(예 3) 일반 제약조건 :

- A. 구성요소의 길이(length of elements)
 - (1) L(S) ... 주어를 구성하는 단어의 수.
일반적 기준은 $L(S) > 4$ 이면 도치의 조건, 예비의 it나 there의 사용조건이 된다.
 - (2) L(DO) ... 직접목적어를 구성하는 단어의 수.
일반적 기준은 $L(DO) > 4$ 이면 VP15B-1을 구성하는 조건(도치의 조건)이 된다.
 - (3) L(PREP+N/PRON) ... 전치사+명사/대명사를 구성하는 단어의 수

일반적 기준은 L(PREP+N/PRON) < L(DO)이면 VP14-1을 구성하는 조건(도치)이다.

...

B. 도치의 조건

(1) 대조, 강조 : VP10, VP13A-1, VP15A-1에서 도치의 조건이다.

(2) 감탄문, 의문문 : 일반적 도치의 조건.

C. 수동태 불가조건

(1) 타동사가 좋다, 싫다, help 등의 뜻이면 수동태 불가능

(2) 재귀동사, 동족 목적어 동사, have=떡다의 뜻이면 수동태 불가능

D. 기타 조건

(1) 주어=that 절, to 부정사, 동명사, 절 등의 조건이 문장의 구성에 영향을 준다.

(예 4) 특수 제약조건 :

예를 들면, dare라는 주어진 동사가 가지는 동사형은 VP5, VP7, VP13A, VP17A이며 dare가 이러한 동사형중 하나를 사용하는 조건은 의문문, 부정문 등과 같은 문장의 종류와 동사가 가지는 뜻에 따라 다르다.

[정의 3] 사용 강도(Use Strength) :

사용강도는 어떤 문장을 생성하는데 제약조건이 사용되는 중요도를 말하는 것으로 다음과 같은 종류가 있다.

SU(Strongly Used) : 주어진 조건을 반드시 만족하면서 사용되는 경우

GU(Generally Used) : 주어진 제약조건을 일반적으로 만족하는 경우

OU(Occasionally Used) : 가끔 사용될 수 있는 형태의 제약 조건

AI(Absolutely Impossible) : 주어진 제약조건이 절대 허용되지않는 경우

OI(Almost Impossible) : 대부분 주어진 제약조건을 만족하나 예외가 허용되는 경우

[정의 4] 시험 요소(Test Element) :

시험 요소란 동사형을 구분하는데 기본적인 표식이 될수있는 요소들로서 기본동사(Basic Verb), 변칙정형동사(Anomalous Finite), 부사적 불변화사(Adverbial Particles), 기타 요소(Other Elements) 등이 있다.

(1) 기본 동사(Basic Verb) :

A. 기본동사 1 : be, have

B. 기본동사 2 : seem, appear, happen, chance, follow, need, want, get 등이다.

(2) 변칙정형동사(Anomalous Finite) : will, would, shall, should, can, could, may, might, must, dare, need, do, does, did 등이다.

(3) 부사적 불변화사(Adverbial Particles) : on, off, in, out, up, down, back, away 등이다.

(4) 기타 요소(Other Elements) :

- A. 기타 요소 1 : as, like, as if, as though 등.
- B. 기타 요소 2 : ought, can't, won't, get 등.

[정의 5] 부호(Symbol) :

부호는 논리 부호로 AND(&), OR(!), NOT(-)이 있고, 그밖에 의미를 나타내는 M (Meaning), 길이를 나타내는 L(Length), S3(3인칭 단수때 s의 사용 여부), 공백을 나타내는 NUL 등이 있다.

이상의 구성요소들은 동사사전에서 사용되어 기계번역기나 인간에 의해 생성되는 문장의 검정과 수정에 사용되기위한 제약조건들을 표현하거나 검사하는데 사용된다.

III. 동사사전의 구조 및 이용

동사형을 포함하는 동사사전의 구조를 살펴보자. 동사사전은 동사, 동사형, 제약조건, 의미의 4부분으로 나누어 지는데 그 구조는 다음과 같다.

```

(Verb   (Verb Pattern 1 (Constraints 1) (Meaning 1))
         (Verb Pattern 2 (Constraints 2) (Meaning 2))
         ( ...           ...           ...           )
         (Verb Pattern i (Constraints i) (Meaning i))
         ( ...           ...           ...           )
         (Verb Pattern n (Constraints n) (Meaning n)))

```

그림 1. 동사사전의 구조.

동사 Verb를 이용하여 만들어지는 문장은 그 동사가 의미 Meaning i로 사용되기위해서 동사형 Verb Pattern i의 형태로 구성되어야 하며, 이때 사용되는 제약조건은 Constraints i를 만족하여야 한다. 이와 같은 구조로 만들어진 cause, dare, seem과 같은 동사의 동사사전은 다음과 같다.

```

(cause (6A,6B,12,14 (NUL)      (..의 원인이 되다. 야기하다))
       (17A         (NUL)      (.. 하게 하다. 시키다))
       (NUL         (AI(VP9))  (NUL)))

```

```

(dare (5 (SU(부정문|의문문|감탄문) & AI(7 & -S3) & SU(조동사))(감히..하다))

```

(7A (SU(긍정문) & AI(5) & SU(-조동사)) (감히 ..하다))
 (13A, 17A (SU(긍정문) & AI(7)) (..에게 도전하다))

(seem (2A-2 (SU(S=that 절) (..인 듯이 생각되다))
 (3A (NUL) (..으로 보인다))
 (4D (SU(S=-(to 부정사|동명사|절)))(..인 듯이 생각되다))
 (4D-1 (SU(S=to 부정사|동명사|절)) (..인 듯이 생각되다))
 (4E (NUL) (..인 것 같다.)))

그림 2. 동사사전의 예.

동사 cause는 ‘..의 원인이 되다’라는 의미로 사용될때 동사형 6A, 6B(주어+동사+명/대), 동사형 12(주어+타동사+명사/대명사(간목)+명사/대명사(직목)), 동사형 14(주어+타동사+명사/대명사(직목)+전치사+명사/대명사(구))로 구성되고 특별한 제약조건은 없다. 그리고 ‘..하게 하다. 시키다’의 뜻으로 사용될때는 동사형 17A(주어+타동사+명/대+(not)+to 부정사(구))로 사용되며 특별한 제약조건은 없다. 그리고 동사형 9(주어+타동사+that 절)로는 절대 사용될 수 없음을 나타내고 있다.

동사 dare의 경우 조동사로 쓰일때에 부정문, 의문문, 감탄문에서 동사형 5(주어+변칙정형동사+부정사(구))로 쓰이며 3인칭 단수때 dare에 s를 붙이지 않으며 동사형 7은 허용되지 않는다. 조동사가 아닐때는 긍정문으로 쓰여서 동사형 7A(주어+타동사+(not)+to 부정사(구))로 사용된다. 위의 두 경우 그때 dare의 의미는 똑같이 ‘감히 ..하다’이므로 두가지의 차이는 긍정문인가 아닌가에 따라 동사형 5와 7중 어느 것을 선택할 것인가가 결정된다. 그밖에 ‘..에게 도전하다’의 뜻일 때는 동사형 13A(주어+타동사+명사/대명사(직목)+to+명사/대명사(구))나 17A(주어+타동사+명사/대명사+to 부정사(구))로 사용된다.

- (7) A. She dares speak about it.
- B. How dare you to come in without knocking!
- C. She dares not to speak about it.
- D. She dares not speak about it.

그러므로 dare의 (7) A는 문장이 긍정문이므로 7A, 13A, 17A로 쓸 수 있으며 문장에서 동사가 ‘감히 .. 하다’로 사용되었으므로 7A로 바뀌어야 한다. 물론 여기서 주어진 문장의 동사가 ‘감히 ..하다’로 사용되었는가는 따로 정보가 주어지지 않으면 판단하기 어려우므로 현재의 문장 구성 성분의 형태로 판단할 수 밖에 없을 것이다. 이문장은 (7) A’로 되어야 한다. (7) B는 감탄문이므로 동사형 7A로 될 수 없으므로 (7) B’처럼 동사형 5로 구성되어야 한다. (7) C는 부정문이므로 (7) C’와 같이 동사형 5로 구성되어야 한다. (7) D는 부정문일때는 dare가 조동사로 사용되므로 dare에 s를 붙이지 않는다.

- (7) A'. She dares to speak about it.
 B'. How dare you come in without knocking!
 C'. She dare not speak about it.
 D'. She dare not speak about it.

동사 seem의 경우 동사적 의미는 모두 비슷하게 사용되어 구분이 힘들고, 문장의 구문적 형태에 따라 사용할 수 밖에 없다. 이러한 구문적 제약을 동사형을 이용하여 동사사전에 나타내고 있다. 동사 seem을 이용한 예는 다음과 같다.

(예 5) 비타민 C는 여러가지 면에서 인체에 도움을 주고 있는 것으로 보인다.

S = Vitamin C

V = seem

여러가지면에서 인체에 도움을 주다 = help the body in many ways.

이 문장에서 필요한 역어를 적절한 방법에 의해 선정하였다고 하면, 동사 seem이 사용 가능한 동사형은 seem의 동사사전에 의해서 다음과 같다.

(VP2A-2) it+seem+주어 (제약조건=주어가 that 절)

(VP3A) 주어+자동사+전치사+명사/대명사/동명사

(VP4D) 주어+seem+(to be)+형용사/명사 (제약조건=주어가 4D-1이외의 것)

(VP4D-1) it+seem+형용사/명사+주어 (제약조건=주어가 to 부정사/동명사/절)

(VP4E) 주어+seem+to 부정사(구)

주어가 Vitamin C로 단독 명사이므로 여기서 가능한 동사형은 4D와 4E이다. 여기서 도움을 주다(help)라는 동사에 의해 동사형 4D가 불가능하므로 4E가 가장 적절하다. 그래서 생성되는 문장은 다음과 같다.

(8) Vitamin C seems to help the body in many ways.

이와 같이 동사사전에 특정 동사의 동사형, 제약조건, 의미를 상세하게 묘사함으로써 기계나 인간이 번역한 문장을 적절한 구문적, 의미적 형태를 고려하여 교정할 수 있다. 그러므로 기계번역기나 인간이 생성한 문장을 검사하고 고치는 문장 검사기의 수준의 정도는 특히 동사사전에 포함된 동사형, 특수 제약조건 등이 얼마나 정확하고 광범위하게 조사되었는가에 좌우된다.

IV. 영어 문장 검사기의 구현

영어 문장 검사기에서 가장 중요한 부분과 동사형 검사와 문법 검사 모듈에서 사용되는 동사사전과 일반 제약조건이다. 일반 제약조건은 우리가 사용하는 사전에는 없는 영어 문법의 기본적인 규칙들이므로 이를 어떻게 조사해서 저장하는가가 중요하다. 또한 동사사전에서는 얼마나 정확하고 상세한 제약조건들이 조사되고 규정되어 있는가가 정확한 문장을 생성해낼 수 있는지 여부를 결정하여 준다. 그림 3은 일반 제약조건이 사용된 문법 검사 알고리즘이다.

```
[Algorithm] GrammarCheck(VP, S, DO, PREP+N|PRON, ...)  
{  
  /* Grammar Check Module Using General Constraints */  
  If L(S) > 4 then invert the sentence using preparatory it or there.  
  If L(DO) > 4 & VP = 15B then change this sentence into 15B-1.  
  If L(PREP+N|PRON) < L(DO) & VP = 14 then use 14-1.  
  If L(to+N|PRON) < L(DO) & VP = 13A then use 13A-2.  
  If L(Present Particle) < L(N) & VP = 19A then use 19A-1.  
  ...  
}
```

그림 3. 일반 제약조건을 사용한 문법 검사 알고리즘.

문장 검사기는 그림 4과 같이 크게 5부분으로 구성된다. 즉 파서, 동사형 분석기, 동사형 검사기, 문법 검사기, 문장 교정기이다.

먼저 파서는 한영 번역기나 인간에 의해 생성된 영어문장을 파싱하게 되는데, 여기서는 기존의 파서와는 달리 문장의 구문적 구조외에도 확장된 동사형에서 사용하고 있는 여러가지 문법적 요소들과 같은 상세한 정보를 제공한다. 즉 주어나 목적어 등의 길이, 문장이 의문문, 부정문, 감탄문인가와 같은 정보, 시험요소의 존재여부 등이다. 이러한 정보는 동사형 분석기의 입력정보로 사용된다.

동사형 분석기와 동사형 검사기는 파서로부터 출력된 정보, 확장된 동사형 표, 동사형 구성요소 등을 이용하여 영어 문장의 동사형을 알아내고, 동사사전과 주어진 한국어 동사를 이용하여 동사형이 일치하는 동사의 의미와 함께 적절히 사용되었는가를 판정한다.

문법 검사기는 일반 제약조건, 동사사전 등을 이용하여 적절한 문장으로 고치도록 권하게 된다. 이러한 권유는 문장 교정기의 입력이 되어 더욱 정확하고 올바른 새로운 문장이 출력되게 된다.

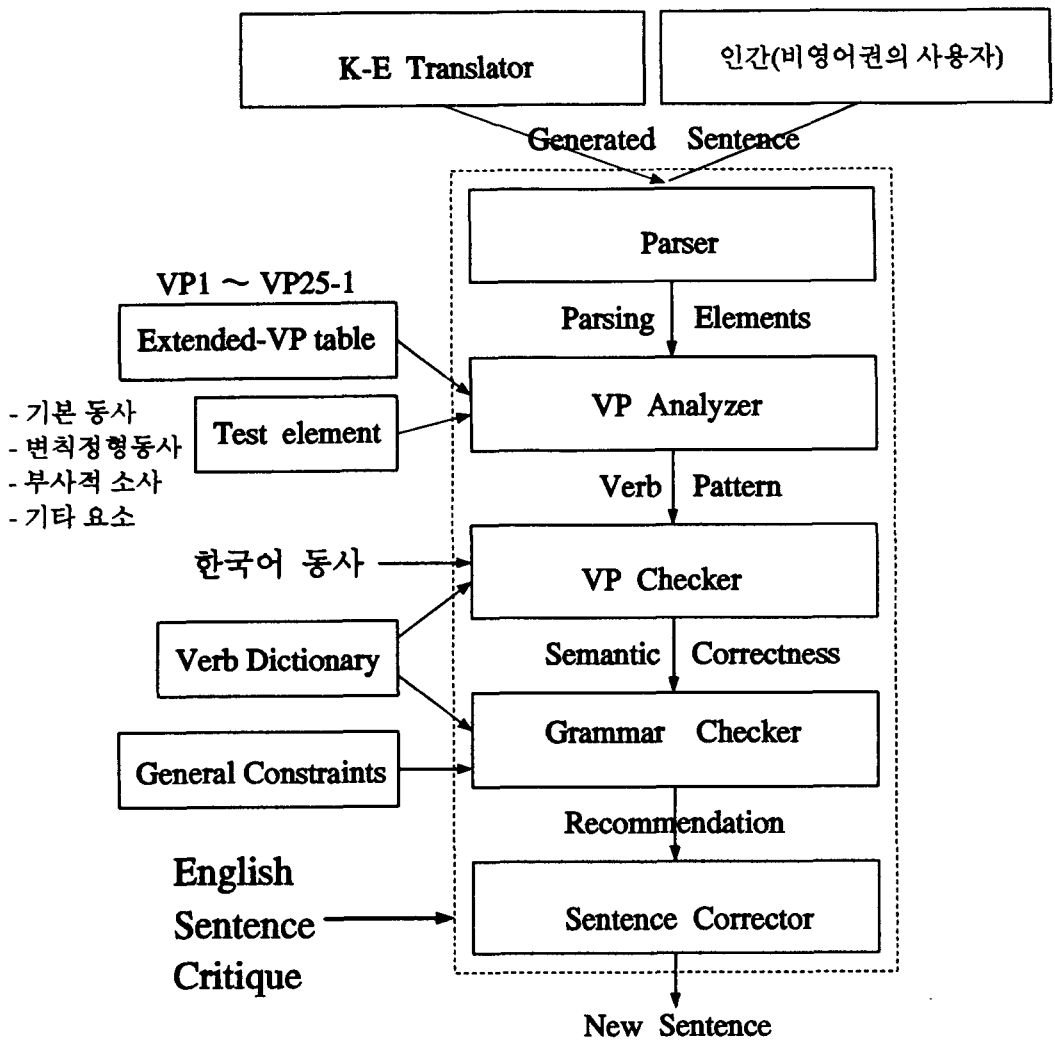


그림 4. 영어 문장 검사기의 구조.

V. 결론

확장된 동사형을 포함하는 동사 사전은 기존의 변환 사전이 고려하지 않는 동사형을 기본으로 하고 있다. 그리고 우리가 사용하는 사전에 없어서, 동사사전에 포함할 수 없는 영어 문법적인 요소들은 따로 일반 제약조건으로 표로 저장하고 있다. 이렇게 함으로서 기계 번역기나 비영어권의 사용자가 생성한 영어 문장이 문법적인 규정을 잘 따르고 있는가를 판정하고, 좀 더 문어체적이고 올바른 문장으로 고쳐준다. 영어 문법 검사기의 성능은 이러한 더욱 세밀하고 정확하게 조사된 제약조건들을 포함하는 동사사전과 영어의 기본적인 문법적 요소를 더욱 보강한 일반 제약조건에 의해 좌우된다.

이 논문에서의 영어 문장 검사기는 우선 비영어권의 사용자에 의해 생성된 문장을 검사

하기위한 것이지만, 기존의 collocation에 기반한 한영 트랜스퍼 사전[5]과 잘 결합시킨다면 기계번역기에서 좀 더 정확한 영어 문장이 생성되게 할 수 있다. 그리고 많은 동사에 대한 동사사전을 구성하는데 있어서 제약조건을 규정하고 입력하는 것은 큰 작업이므로 이를 위한 더욱 효율적인 사전 기술 언어[6]가 개발되어야 할 것이다.

참 고 문 헌

[1] J. B. Heaton and N. D. Turton, Longman Dictionary of Common Errors, Longman Group UK Limited, 1990.

[2] SNU, Natural Language Processing Status Report, Natural Language Processing Lab., Seoul National University, December 1991.

[3] B. Boguraev and T. Briscoe, "Large Lexicons For Natural Language Processing: Utilizing The Grammar Coding System of LDOCE", Computational Linguistics, Volume 13, Numbers 3-4, July-December 1987.

[4] E. Sumita, H. Iida and H. Kohyama, "Translating with Examples : A New Approach to Machine Translation", The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language, 1990.

[5] 김 나리, "Collocation에 기반한 한영 트랜스퍼 사전의 구성에 관한 연구", 석사학위 논문, 서울대학교 컴퓨터공학과, 1992.

[6] 심 광섭, 김 영택, "변환 사전 기술 언어", 정보과학회 논문지, 제19권 제 1호, 한국정보과학회, 1992.

[7] 양 승현, "숙어번역 방식을 이용한 영한 기계번역 시스템의 한국어 생성기에 관한 연구", 석사학위 논문, 서울대학교 컴퓨터공학과, 1992.

[8] 박 성재, "영한 번역기에서의 부사구 처리에 관한 연구", 석사학위 논문, 서울대학교 컴퓨터공학과, 1992.

[9] 시사영어사 편집국, 시사 엘리트 영한 사전, 시사영어사, 1987.

[10] 김 주현, 최 립, 스탠더드 영한 사전, 성문각, 1976.