

(19)대한민국특허청(KR)

(12) 등록특허공보(B1)

(51) 。 Int. Cl.⁷
G06F 17/27

(45) 공고일자 2004년01월07일
(11) 등록번호 10-0413966
(24) 등록일자 2003년12월22일

(21) 출원번호	10-2000-0077945	(65) 공개번호	10-2002-0048715
(22) 출원일자	2000년12월18일	(43) 공개일자	2002년06월24일

(73) 특허권자 한국전자통신연구원
대전 유성구 가정동 161번지

(72) 발명자 정의석
대전광역시유성구신성동208-3은지빌라101호

차건희
대전광역시서구괴정동4-17

임수종
서울특별시성동구금호3동1143번지2층

윤보현
대전광역시유성구전민동462-4청구나래아파트110-801

왕지현
대전광역시유성구신성동148-4연구빌라301호

강현규
대전광역시유성구가정동236-2KIT교수아파트14-404

(74) 대리인 특허법인 신성

심사관 : 정재현

(54) 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치 및 그방법

요약

1. 청구범위에 기재된 발명이 속하는 기술분야

본 발명은 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것임.

2. 발명이 해결하려고 하는 기술적 과제

본 발명은 표준 문형과 같이 제한된 언어 표준을 따르는 응용 시스템 등에 있어, 사용자에게 실시간으로 표준 문형을 유도하기 위한 표준 문형 유도 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하고자 함.

3. 발명의 해결 방법의 요지

본 발명은, 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치에 적용되는 표준 문형 유도 방법에 있어서, 문장을 입력받아, 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고, 태깅 작업을 수행하는 제 1 단계; 형태소별로 분석된 문장의 어휘 정보를 추출하고, 형태소의 태깅 정보를 이용하여 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하되, 상기 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면 생략 추정 패턴을 검색하여, 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하는 제 2 단계; 및 문장 입력시 현재 시스템의 사용자

입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하여 점진적 구문분석의 피드백을 사용자에게 제공하는 제 3 단계를 포함함.

4. 발명의 중요한 용도

본 발명은 비정규적인 언어 표현을 표준화된 문장구조에 적합하도록 유도하는데 이용됨.

대표도

도 1

색인어

한국어, 표준 문형 규칙, 표준 문형 유도, 어절, 형태소, 언어 인식

명세서

도면의 간단한 설명

도 1 은 본 발명에 따른 표준 문형 규칙에 의한 표준 문형 유도 장치의 일 실시예 구성도.
 도 2 는 본 발명에 따른 표준 문형 규칙에 의한 표준 문형 유도 과정의 일 실시예 흐름도.
 도 3 은 본 발명에 따른 표준 문형 규칙에 의한 어휘 범주 목록의 일 예시 설명도.
 도 4a 및 4b 는 본 발명에 따른 표준 문형 규칙에 의한 문장 구성의 일 예시 설명도.
 도 5a 및 5b 는 본 발명에 따른 표준 문형 규칙을 이용한 점진적 구문 분석 과정의 일 예시 설명도.
 도 6 은 본 발명에 따른 논항 생략에 대한 추정 규칙의 일 예시 설명도. * 도면의 주요 부분에 대한 부호 설명 110 : 입력부 120 : 형태소 분석부 130 : 어휘 정보 추출부 140 : 표준 문형 규칙 결정부 150 : 생략 추정 패턴 검색부 160 : 표준 문형 검증 및 추정부 170 : 출력부

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은, 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치 및 그 방법에 관한 것으로, 더욱 상세하게는 표준 문형과 같이 제한된 언어 표준을 따르는 응용 시스템 등에 있어, 생략 추정 패턴과 바꾸어 쓰기 방식을 이용하여 사용자에게 실시간으로 자연스러운 표준 문형을 유도하기 위한 표준 문형 유도 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것이다.

제한된 문법을 이용한 표준 문형 유도 작업은, 영어권에서 지속적으로 진행되어왔다. 그 대표적인 것으로, 'N. E. Fuchs'와 'R. Schwitter'에 의해 제안된 "Attempto Controlled English(ACE), <CLAW 96, March 1996>" 등이 있는데, 여기서는 요구 사항 명세서를 작성하기 위해 상기 ACE라는 영어 표준안을 제안하고 있다.

표준안에 기반한 언어 처리 시스템의 가장 큰 문제점은, '어떻게 표준화된 유형으로 언어 생성을 유도할 것인가'이다. 상기 ACE의 경우에는, 표준안 유도 방법으로 '바꾸어 쓰기(Paraphrase)' 방식을 이용하였다. 그러나, 상기의 바꾸어 쓰기가 오직 문장 종료시에 한정되어 이루어지므로, 문장 입력자에게 실시간적인 피드백(feedback)을 줄 수 없어 문장 생성의 생산성을 감소시키는 문제점이 있었다.

발명이 이루고자 하는 기술적 과제

본 발명은 상기한 바와 같은 문제점을 해결하기 위하여 제안된 것으로, 표준 문형과 같이 제한된 언어 표준을 따르는 응용 시스템 등에 있어, 사용자에게 실시간으로 표준 문형을 유도하기 위한 표준 문형 유도 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는데 그 목적이 있다.

즉, 본 발명은 표준 문형 규칙에 따른 텍스트를 생성함에 있어 사용자에게 표준 문형 규칙에 대한 특별한 지식을 요구하지 않기 위하여, 사용자의 비정규적이고 일상적인 문장 표현을, 생략 추정 패턴과 바꾸어 쓰기 방식에 의해 실시간적으로, 자연스럽게 표준 문형으로 유도하는데 그 목적이 있다. 즉, 비정규적인 언어 표현을 표준화된 문장구조에 적합하도록 유도하고자 하는 것이다. 아울러, 표준 문형안을 정규적 표현 방식으로 기술하여 표준 문형 해석기 모듈에

실시간으로 로딩될 수 있게 형식화하여 제시하고자 한다.

발명의 구성 및 작용

상기 목적을 달성하기 위한 본 발명은, 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치에 있어서, 사용자에게 의해 문장을 입력받기 위한 입력수단; 상기 입력수단을 통해 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고, 태깅 작업을 수행하기 위한 형태소 분석수단; 상기 형태소 분석수단에 의해 형태소별로 분석된 문장의 어휘 정보를 추출하기 위한 어휘정보 추출수단; 형태소의 태깅 정보를 이용하여, 상기 어휘정보 추출수단을 통해 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하여 적용하기 위한 표준문형규칙 결정수단; 및 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과(점진적 구문 분석 결과)를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하는 구문분석 피드백수단을 포함하되, 상기 표준문형규칙 결정수단에서 결정된 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면, 생략 추정 패턴을 검색하기 위한 생략추정패턴 검색수단; 및 상기 생략추정패턴 검색수단을 통해 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하기 위한 표준문형 검증/추정수단을 포함하여 이루어진 것을 특징으로 한다.

그리고, 본 발명은 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치에 적용되는 표준 문형 유도 방법에 있어서, 문장을 입력받아, 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고, 태깅 작업을 수행하는 제 1 단계; 형태소별로 분석된 문장의 어휘 정보를 추출하고, 형태소의 태깅 정보를 이용하여 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하되, 상기 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면 생략 추정 패턴을 검색하여, 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하는 제 2 단계; 및 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하여 점진적 구문분석의 피드백을 사용자에게 제공하는 제 3 단계를 포함하여 이루어진 것을 특징으로 한다.

한편, 본 발명은 한국어 표준 문형 규칙에 의한 표준 문형 유도를 위하여, 프로세서를 구비한 표준 문형 유도 시스템에, 문장을 입력받아, 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고 태깅 작업을 수행하는 제 1 기능; 형태소별로 분석된 문장의 어휘 정보를 추출하고, 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하되, 상기 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면 생략 추정 패턴을 검색하여, 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하는 제 2 기능; 및 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하여 점진적 구문분석의 피드백을 사용자에게 제공하는 제 3 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다. 본 발명의 기술적인 과제를 살펴보면, 표준 문형을 이용한 점진적 구문 분석 기법으로서, 본 발명에서는 실시간 표준 문형 검증기의 기본 기술로써 문장 입력자에게 표준 문형에 적합한지 아닌지에 대한 실시간 반응이 요구된다. 이를 위해, 기존 언어학에서 언급되어 온 문형이라는 개념을 확장한 구조화된 표준 문형 집합이 필요하며, 그 표준 문형으로 사용자의 생성 문장을 유도하는 언어 처리 기술이 필요하다. 또한, 한국어의 특수 현상으로 간주되는 빈번한 중요성분 생략 현상과 어순 자유 현상에 어느 정도 대처할 수 있어야 한다. 본 발명은 사용자의 비정규적인 언어 표현을 표준 문형으로 유도할 수 있는 것으로, 이때 사용자에게 표준 문형안에 대한 특별한 지식을 요구하지 않으며, 일상적 문장 표현과의 차이점을 줄일 수 있게 표준 문형 유도를 제시하며, 표준 문형안을 정규적 표현 방식으로 기술하여 표준 문형 해석기 모듈에 실시간으로 로딩될 수 있게 형식화하여 제시한다. 이를 위해, 본 발명에서는 표준 문형안을 정규화된 형식으로 제시하고, 이를 이용한 점진적 구문 분석 장치를 기술한다. 그리고, 표준 문형의 구성에 대하여 정리하고, 이를 이용해 점진적 구문 분석 기법 및 바꾸어 쓰기에 대하여 기술한다. 또한, 표준 문형 유도를 위한 논항 추정 방식을 정리하여 표준 문형 유도에 대한 방식을 제시한다. 본 발명에 따르면, 표준 문형은 한국어 문장 패턴을 구조적으로 기술하여 언어처리의 응용에 타당하도록 기술되었으며, 이를 이용한 점진적 구문 분석 기법은 어절 단위의 동적 분석을 가능하게 하였다. 또한, 한국어에 빈번한 언어 현상인 자유어순과 생략현상은 표준 문형 규칙과 논항 추정 방식으로 해결하여 사용자의 표준 문장 생성에 자연스러움을 유도하였다. 그리고, 문장 입력시 동적으로 분석되는 결과를 사용자에게 제시함으로써, 문장 종료후 수행되는 바꾸어 쓰기(Paraphrase) 과정을 한단계 발전시켰다.

상술한 목적, 특징들 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해 질 것이다. 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 일실시예를 상세히 설명한다.

도 1은 본 발명에 따른 표준 문형 규칙에 의한 표준 문형 유도 장치의 일실시예 구성도이다.

도 1에 도시된 바와 같이, 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치는, 사용자에게 의해 문장을 입력받기 위한 입력부(110)와, 입력부(110)를 통해 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고, 태깅 작업을 수행하기 위한 형태소 분석부(120)와, 형태소 분석부(120)에 의해 형태소별로 분석된 문장의 어휘 정보를 추출하기 위한 어휘 정보 추출부(130)와, 형태소의 태깅 정보를 이용하여, 어휘 정보 추출부(130)를 통해 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하여 적용하기 위한 표준 문형 규칙 결정부(140)와, 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과(점진적 구문 분석 결과)를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하는 출력부(170)를 포함한다.

여기서, 상기 표준 문형 규칙 결정부(140)는, 입력되는 문장에 따라, 본 발명에 따른 표준 문형 유도 장치가 적용되는 시스템의 내부에 미리 저장되어 로딩(loading)된 한국어 표준 문형 규칙 중 적합한 것을 선택하여 적용시키는 기능을

수행한다.

또한, 상기 표준 문형 규칙 결정부(140)에서 결정된 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않을 경우, 생략 추정 패턴을 검색하기 위한 생략 추정 패턴 검색부(150) 및 상기 생략 추정 패턴 검색부(150)를 통해 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하기 위한 표준 문형 검증 및 추정부(160)를 더 포함한다.

상기 각 구성 요소들의 상세 기능 및 동작 과정은, 이하에서 설명하기로 한다.

도 2는 본 발명에 따른 표준 문형 규칙에 의한 표준 문형 유도 과정의 일실시에 흐름도이다.

도 2에 도시된 바와 같이, 본 발명에 따른 표준 문형 유도 과정은, 우선 표준 문형 규칙이 로딩됨으로써(210) 초기화된 시스템에, 사용자에게 의해 문장이 입력되면서(220) 시작된다.

이때, 상기 로딩된(210) 표준 문형 규칙은, 메모리에 저장되어 표준 문형 검증을 위한 구문 분석에 적용된다.

상기 사용자에게 의해 입력된(220) 문장은 어절 단위로 분석이 진행된다(230). 즉, 하나의 어절이 입력되고, 이에 대한 확인인 공백이 입력되면, 즉시 형태소 분석 단계로 진행되어 형태소 분석 및 태깅 작업을 수행한다(231).

이후, 필요할 경우 어휘 정보 추출을 수행(232)하게 되는데, 이러한 상기 어휘 정보 추출 과정(232)은, 서술어의 경우 어휘별 논항 구성 정보를 어휘 사전으로부터 얻기 위함이다. 즉, "가다"라는 서술어가 입력되었을 때, "--가--로가다"라는 서술어의 하위 범주화 정보를 얻어, 표준 문형 검증에 적용하기 위한 것이다.

상기 어휘 정보 추출 과정(232)을 통해, 분석된 어절에 대한 적합한 표준 문형 규칙이 결정된다(233).

좀더 자세한 설명을 위하여, 하기의 도 5a에 예시한 "나는 연구소에 간다"라는 문장을 보면, 첫 어절 "나는"이 "나/nc는/jc"으로 분석된 후(512), 분석된 각각에 대하여 적합한 표준 문형 규칙들이 할당되는 것을 볼 수 있는데(513), 이러한 과정들이 어휘 정보 추출을 통해(232) 표준 문형 규칙을 결정하는(233) 과정이다.

상기 표준 문형 규칙 결정(233)은 형태소의 태깅(tagging) 정보를 이용하여 수행된다.

이후, 부분적 구문 분석 과정인 표준 문형 규칙 매칭(Matching) 작업이 수행되는데(240), 여기서 분석된(231) 형태소가 표준 문형 규칙에 적합하지 않을 경우, 생략 추정 패턴 검색을 수행하게 된다(242). 이러한 생략 추정 패턴 검색(242)은 내포문이나 이어진 문장의 경우에, 내포문이나 선행절에서 생략된 논항을 안은 문장이나 후행절의 논항으로 추정해서(243) 표준 문형 매칭 작업을 계속적으로 진행할 수 있게 하는 과정이다.

상기 분석된(231) 형태소가 상기의 표준 문형 규칙에 적합하거나(240), 또는 생략 추정 패턴 검색(242)을 통해 표준 문형 추정을 완료하게 되면(243), 바꾸어 쓰기 출력을 통해(241) 현재 진행중인 분석 결과를 출력하게 된다. 따라서, 문장 입력자는 자신이 현재 생성중인 문장이 표준화 문형에 적합한지 아닌지를 실시간으로 알 수 있으며, 지속적인 문장 입력으로 특정 표준 문형 유형을 학습하게 되는 것이다.

상기 표준 문형 규칙을 구성하는 용어 정보들은, 하기의 도 3에 예를 들어 기술하였는데, 여기서 지시관형사 dt, 수관형사 nu 및 성상관형사 ad는 모두 관형사 mm의 세부 어휘 범주를 나타낸다.

도 3은 본 발명에 따른 표준 문형 규칙에 의한 어휘 범주 목록의 일예시 설명도이다.

도 3에 도시된 바와 같이, 본 발명에서 사용되는 어휘 범주의 목록은 크게 31 가지가 있으나, 이는 어디까지나 자의적인 일례일 뿐이며, 경우에 따라 얼마든지 범주의 종류와 명칭에 있어 다른 목록을 사용하여도 무방하다.

단, 본 발명에서는, 도 3에 도시된 어휘 범주의 목록과 명칭에 한정하여 설명하기로 한다.

도 4a 및 4b는 본 발명에 따른 표준 문형 규칙에 의한 문장 구성의 일예시 설명도이다.

도 4a 및 4b에 도시된 바와 같이, 기본문(410,411)은, 1항 술어, 2항 술어, 3항 술어로 구성된다. 각 기본문을 구성하는 NP1은, 도 4b의 논항(421)에 도시했듯이, 명사구 _NP와 주격조사 jc1이 결합되어 구성된다. 동일한 방법으로 목적격 조사 jc2, 필수 부사격 조사 jc3에 의하여 NP2, NP3가 각각 구성된다.

도 4a의 논항(420)은 상기 NP1, NP2, NP3를 말하는 것이다.

도 4a의 논항 확대(450)는, 관형사의 논항 확대(451), 접속조사, 접속부사의 논항 확대(452), 속격조사의 논항 확대(453)에 의해 구성된다. 여기서, 논항 확대란, 단일 NP에 의해 구성되는 논항이 아닌 수식어를 포함한 논항을 의미한다. 즉, 관형사의 논항 확대(451)에서 "이 사람이"의 경우, 관형사 "이"에 의해 "사람"이 수식되어 "((이/mm 사람/nc)/_NP+이/jc1)/NP1"으로 구조 분석이 될 수 있음을 의미한다.

여기서, "NP"와 "_NP"와의 차이점을 설명하면, NP(471)는 복합명사나 단일명사를 말하며, _NP는 "(이/mm (사람/nc)/_NP)/_NP"와 같이 확장된 NP를 말한다.

도 4a 및 4b에서 기술되고 있는 표준 문형 규칙들은 CFG(Context Free Grammar)형식을 따르며, 다중 트리 방식을 취하고 있다.

표준 문형 규칙의 표기법을 살펴보면, A:(B+[C] D{ }...)에서 A는 (B+[C] D{ }...)의 상위 개념이고 "[]"는 필수 요구 항목이 아니라는 표시이다. 또한, "{ }"는 구성 성분의 제약 정보를 나타낸다.

상기한 도 2의 표준 문형 규칙의 로딩 단계(210)에서 도 4a 및 4b의 규칙 항목들을 각각 메모리에 저장하여, 표준 문형 규칙으로써 구문 분석에 적용하게 되는 것이다.

도 4b의 관형사 논항 확대(351)에서 mm{!.[dt]} 중 "{ }"는, 각 규칙 구성 항목의 제약 정보를 기술한다. 표준 문형 규칙의 제약 정보는 모두 세가지로 구성되어 있다.

"!"은 세부 어휘 범주 정보, 즉, 관형사 논항 확대(451)의 관형사 mm{!.[dt]}은 반드시 지시관형사 dt에 한하여 허락한다는 말이며, 접속조사, 접속부사의 논항 확대(452)의 "!"@는 "NP+''"가 재귀적으로 쓰일 수 있다는 것을 말한다. 그리고, 속격조사의 논항 확대(453)의 "!"^은, 속격조사 jm인 "- 의"에 의해 수식되는 명사구 _NP는 내부에 단일 NP만을 가질 수 있다는 것을 말한다. 예를 들면, "철수의 엄마와 동생"은 "((철수의 엄마)와 동생)"으로 항상 "철수의 엄마"와 결합되며 "(철수의 동생)"으로는 결합될 수 없다는 말이다.

도 4a의 서술어 확대(460)는, 도 4b의 부사의 서술어 확대(465)와 보조용언의 서술어 확대(463)로 구성되어 있다. 그리고, 절(430)은 관형절 ETMS(454), 명사절(472), 인용절 JQTS(464), 부사절(465), 서술절(466)로 구성되며, 절에

다른 논항 확대 또한 도 4b에 각각 기술되어 있다.

관형절의 논항 확대(454)는, 문장 _S와 관형사형 전성어미 etm이 결합하여 관형어의 역할을 하고, 명사절(472)은 문장 _S와 명사형 전성어미 etn 혹은 "관형사형 전성어미 etm + 것"에 의해 명사절을 구성하게 된다. 그리고, 인용절(464)은 문장 _S와 "라고, 하고, 고"의 인용격 조사가 결합되어 구성된다. 부사절(465)은 서술성 조사 "같이, 없이, 달리"와 "게, 듯이, 도록"에 의해 생성된다. 마지막으로, 서술절(466)은 이중 주어문을 표현하는 역할로 "코끼리가/NP1 코가/NP1 길다."와 같이 한 문장에 두개의 NP1이 발생하는 구문을 처리하기 위해 필요하다.

도 5a 및 5b는 본 발명에 따른 표준 문형 규칙을 이용한 점진적 구문 분석 과정의 일예시 설명도이다.

즉, 도 5a 및 5b는 본 발명에 따른 표준 문형 규칙을 이용한 점진적 구문 분석 과정의 일예시로서, 문장 "나는 연구소에 간다"에 대한 점진적 구문 분석 기법 및 바꾸어 쓰기 방식을 보여 준다.

본 발명에서 말하는 점진적 구문 분석 기법이란, 구문 분석이 문장 입력 과정에 병행해서 진행되는 것을 의미한다.

먼저, 도 5a의 점진적 구문 분석 과정은, 전술한 바와 같이, 첫 단계의 "나/nc 는/jc"(511)에는 각각 해당 표준 문형 규칙(513)이 할당되며, "NP1:(_NP(나/nc)+jc1(는/jc))"(514)과 같이 병행 분석이 진행된다.

도 5b의 첫 단계(530)는, 상기의 이러한 과정을 트리 형식으로 표현하고 있다.

바꾸어 쓰기는 분석 결과의 가시화(Visualization)를 말한다.

본 발명에서 제안하는 가시화 방법에는, 도 5a의 분석 결과 제시와 도 5b의 트리 분석 결과 제시의 두가지가 있다. 상기 두가지 분석 결과는 문장의 어절별 구문 분석 결과를 보여줌으로써, 점진적 구문 분석의 피드백을 사용자에게 제공하여 현재 시스템의 사용자 입력 이해 정도를 사용자 스스로가 알 수 있게 한다.

도 5a의 (522)는 "나는 연구소에 간다"의 모든 논항과 서술어가 찾아진 결과를 표현하고 있다. 이는 도 5b의 (550)으로 표현할 수도 있다.

이 경우, 도 5b의 (560)과 같이 각 논항 NP1, NP3와 서술어 VP 및 "가/pv"에 의해 기본문이 선택된다. 여기서는 S:(NP1 NP2 NP3 VP)의 3항 기본문이 선택되겠다. 도 5b의 (570)은 최종 분석된 결과를 보여준다.

또한, "연구소에 나는 간다"와 같이 주어와 부사어와의 위치가 바뀌었을 경우에도 본 발명에서 제안하는 표준 문형 규칙의 적용에는 문제가 없다.

NP3와 NP1 자체의 결정은 순서와는 상관없다는 것이다. 즉, 단지 각 논항과 서술어가 서로 일치하는 항목에 결합되기 때문에, 최종 서술어 VP에 의한 문형 S:(NP1 NP2 NP3 VP)의 선택에 있어 논항의 순서는 그다지 문제가 되지 않는다.

따라서, 한국어 문장의 논항간 자유 어순 문제는, 본 표준 문형 규칙의 적용에 있어 어려움 없이 해결할 수 있다.

도 6은 본 발명에 따른 논항 생략에 대한 추정 규칙의 일예시 설명도이다.

도 6에 도시된 논항 생략에 대한 추정 규칙 테이블은, 한국어에 빈번한 언어 생략 현상에 대한 대처 방안의 일례로서 제시된 것이다.

즉, 도 6에 도시된 논항 생략에 대한 추정 규칙은, 상기한 도 2의 생략 추정 패턴 검색(242)과 표준 문형 검증 및 추정(243)에 이용된다.

이러한 생략 현상이 문제가 되는 이유는, 표준 문형 자체가 서술어의 논항이 충족되어야만 한데 있다. 다시 말해, "철수가 먹는다"에서 서술어 "먹는다"의 목적격이 생략되어 있으므로 위 문장은 정문으로 볼 수 없다는데 문제가 있는 것이다.

본 발명에서 생략을 다루는 방식은 모두 두가지로, 그 첫째는, 단문이 상기 예문과 같은 비결정 논항을 가질 때, 바꾸어 쓰기에서 "철수가/NP1 /NP2 먹는다/VP"로 분석하여, 그대로 입력자에게 보여주는 것이다. 즉, 분석 결과에 대한 진위 판정을 사용자의 확인에 의존하도록 하는 것이다. 이러한 방법은, 사용자 입력의 자연스러움과 점진적 분석을 편리하도록 하는데 그 의의가 있다.

둘째, 복문일 경우 일반적 논항 생략은 관계 문장에서 논항을 추정하는 것이다. 여기서, 관계 문장이란 내포문일 경우 안은 문장이 되고, 이어진 문장의 경우 선행절 또는 후행절이 된다.

도 6의 (610),(620),(630)의 경우, 각각 NP1, NP2, NP3에 대한 추정 예문을 보여주고 있으며, (640)은 NP1과 NP2에 대한 추정 예문을 보이고 있다.

(670)과 (680)에는, 이어진 문장의 논항 추정 예문이 기술되어 있다.

이상에서 설명한 본 발명은, 문자(character)의 입력과 출력에 한정하여 설명하였으나, 현재의 음성 인식 기술이 음성 신호를 입력받아 문자로 변환하여 인식함으로써 내부 동작을 수행하는 경우가 대부분이므로, 음성에 의한 입출력 시스템에 적용하여 구현하는 데에 있어서도 아무런 문제가 없음은 명백하다.

상술한 바와 같은 본 발명의 방법은 프로그램으로 구현되어 컴퓨터로 읽을 수 있는 형태로 기록매체(씨디롬, 램, 롬, 플로피 디스크, 하드 디스크, 광자기 디스크 등)에 저장될 수 있다.

이상에서 설명한 본 발명은, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 있어 본 발명의 기술적 사상을 벗어나지 않는 범위내에서 여러 가지 치환, 변형 및 변경이 가능하므로 전술한 실시예 및 첨부된 도면에 한정되는 것이 아니다.

발명의 효과

상기한 바와 같은 본 발명은, 표준 문형 규칙에 따른 점진적 구문 분석 기법을 이용하여 어절 단위의 동적 분석을 가능하게 함으로써 실시간적인 응답을 통한 표준 문형 유도할 수 있는 효과가 있다.

또한, 본 발명은 한국어에 있어 빈번한 자유 어순과 생략 현상을 표준 문형 규칙과 논항 추정 방식으로 해결함으로써, 사용자의 입력 문장에 대한 표준 문형 유도 과정을 보다 더 자연스럽게 할 수 있는 효과가 있다.

(57) 청구의 범위

청구항 1.

삭제

청구항 2.

한국어 표준 문형 규칙에 의한 표준 문형 유도 장치에 있어서,
 사용자에게 문장을 입력받기 위한 입력수단;
 상기 입력수단을 통해 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고, 태깅 작업을 수행하기 위한 형태소 분석수단;
 상기 형태소 분석수단에 의해 형태소별로 분석된 문장의 어휘 정보를 추출하기 위한 어휘정보 추출수단;
 형태소의 태깅 정보를 이용하여, 상기 어휘정보 추출수단을 통해 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하여 적용하기 위한 표준문형규칙 결정수단; 및
 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과(점진적 구문 분석 결과)를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하는 구문분석 피드백수단을 포함하되,
 상기 표준문형규칙 결정수단에서 결정된 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면, 생략 추정 패턴을 검색하기 위한 생략추정패턴 검색수단; 및
 상기 생략추정패턴 검색수단을 통해 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하기 위한 표준문형 검증/추정수단을 포함하는 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치.

청구항 3.

제 2 항에 있어서,
 상기 문장의 입력과 출력은,
 문자에 의한 입력과 출력 및/또는 음성에 의한 입력과 출력을 포함하는 것을 특징으로 하는 한국어 표준 문형 규칙에 의한 표준 문형 유도 장치.

청구항 4.

삭제

청구항 5.

한국어 표준 문형 규칙에 의한 표준 문형 유도 장치에 적용되는 표준 문형 유도 방법에 있어서,
 문장을 입력받아, 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고, 태깅 작업을 수행하는 제 1 단계;
 형태소별로 분석된 문장의 어휘 정보를 추출하고, 형태소의 태깅 정보를 이용하여 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하되, 상기 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면 생략 추정 패턴을 검색하여, 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하는 제 2 단계; 및
 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하여 점진적 구문분석의 피드백을 사용자에게 제공하는 제 3 단계를 포함하는 한국어 표준 문형 규칙에 의한 표준 문형 유도 방법.

청구항 6.

제 5 항에 있어서,
 상기 문장의 입력과 출력은,
 문자에 의한 입력과 출력 및/또는 음성에 의한 입력과 출력을 포함하는 것을 특징으로 하는 한국어 표준 문형 규칙에 의한 표준 문형 유도 방법.

청구항 7.

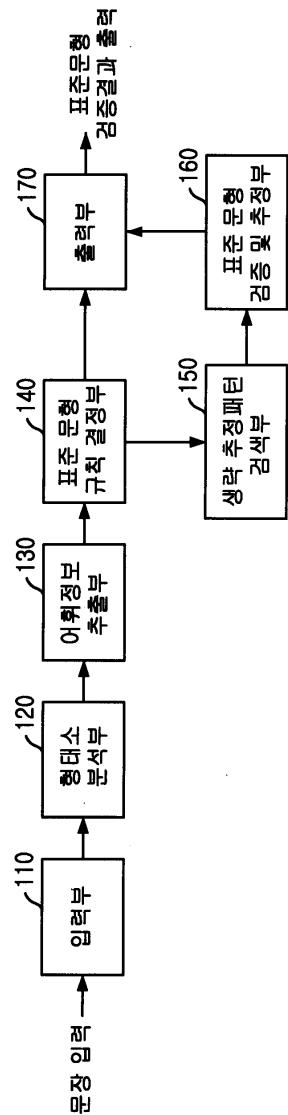
삭제

청구항 8.

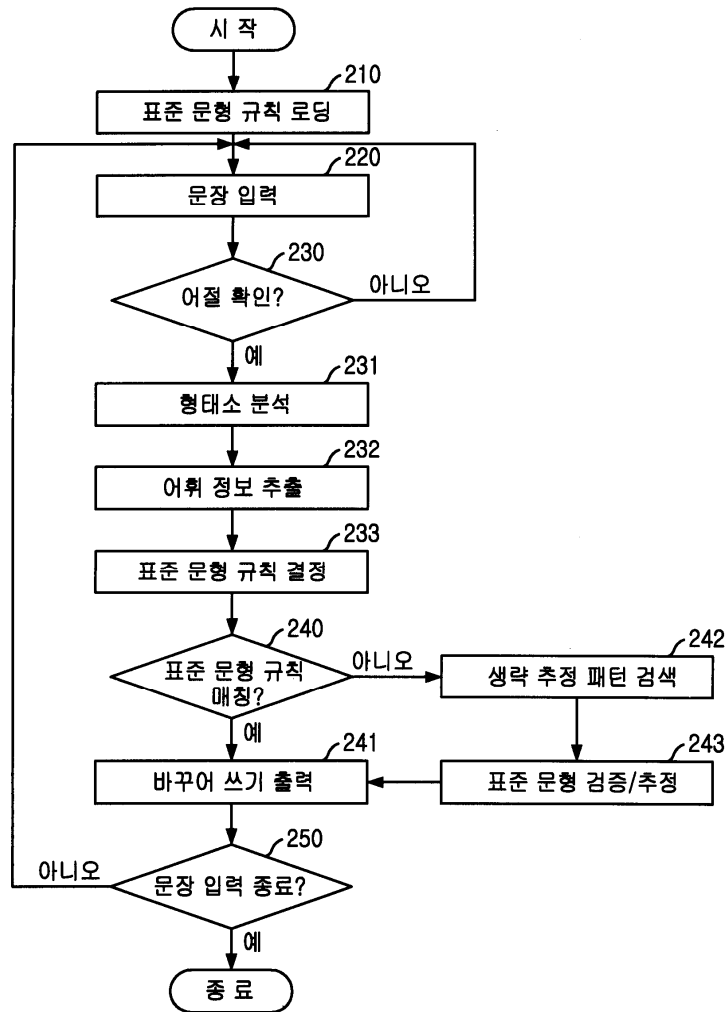
한국어 표준 문형 규칙에 의한 표준 문형 유도를 위하여, 프로세서를 구비한 표준 문형 유도 시스템에,
 문장을 입력받아, 입력받은 문장을 어절별로 구분하여, 구분된 어절을 구성하는 형태소를 분석하고 태깅 작업을 수행하는 제 1 기능;
 형태소별로 분석된 문장의 어휘 정보를 추출하고, 추출된 어휘 정보에 적합한 표준 문형 규칙을 결정하되, 상기 표준 문형 규칙이 상기 입력된 문장에 직접 적용되지 않으면 생략 추정 패턴을 검색하여, 검색된 생략 추정 패턴에 따라 표준 문형을 검증 및 추정하는 제 2 기능; 및
 문장 입력시 현재 시스템의 사용자 입력 이해 정도를 사용자가 실시간으로 알 수 있도록, 상기 표준 문형 규칙이 적용된 입력 문장의 어절별 구문 분석 결과를 바꾸어 쓰기(paraphrase) 방식으로 가시적으로(visualization) 출력하여 점진적 구문분석의 피드백을 사용자에게 제공하는 제 3 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

도면

도면1



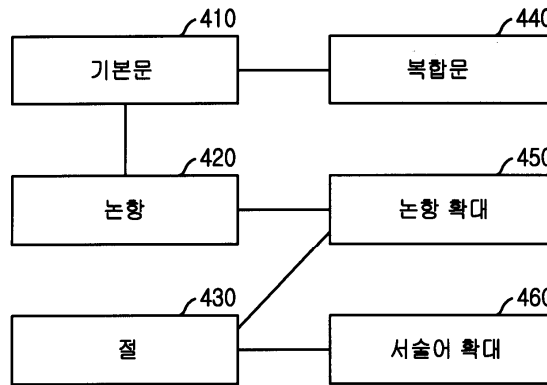
도면2



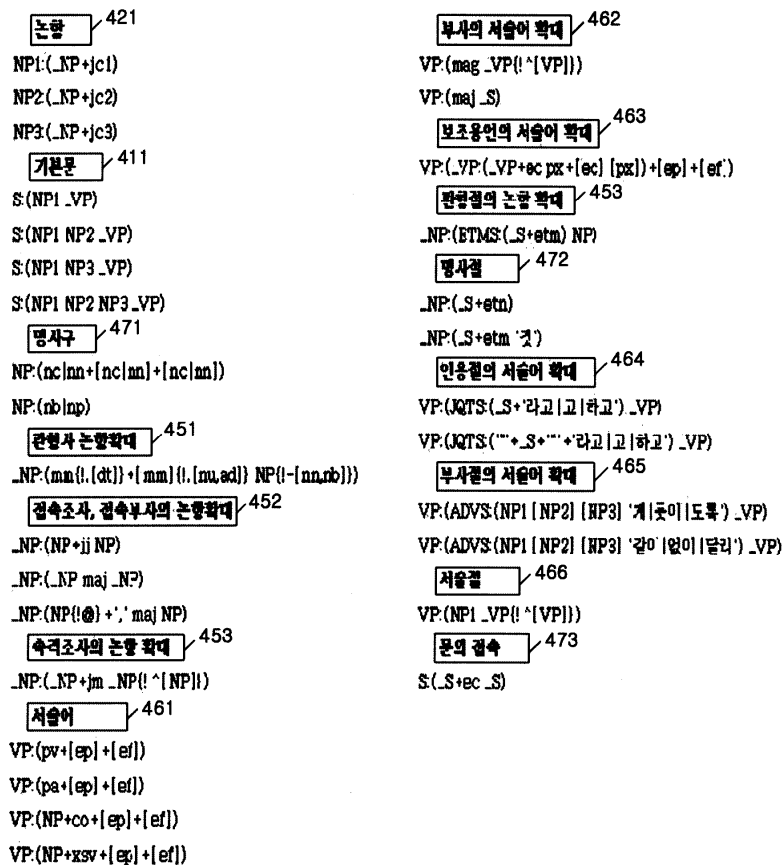
도면3

f	외국어	jc	적조사
nc	자립명사	jx	보조사
nb	의존명사	jj	접속조사
np	대명사	jm	속격조사
nn	수사	ep	선어말어미
pv	동사	ef	종결어미
pa	형용사	ec	연결어미
px	보조용언	etn	명사형어미
co	지정사	etm	관형형어미
mag	일반부사	dt	지시관형사
maj	접속부사	nu	수관형사
mm	관형사	ad	성상관형사
li	감탄사	jc1	주격조사
xsn	명사 파생 접미사	jc2	목적격조사
xsv	동사 파생 접미사	jc3	필수부사격조사
xsm	형용사 파생 접미사		

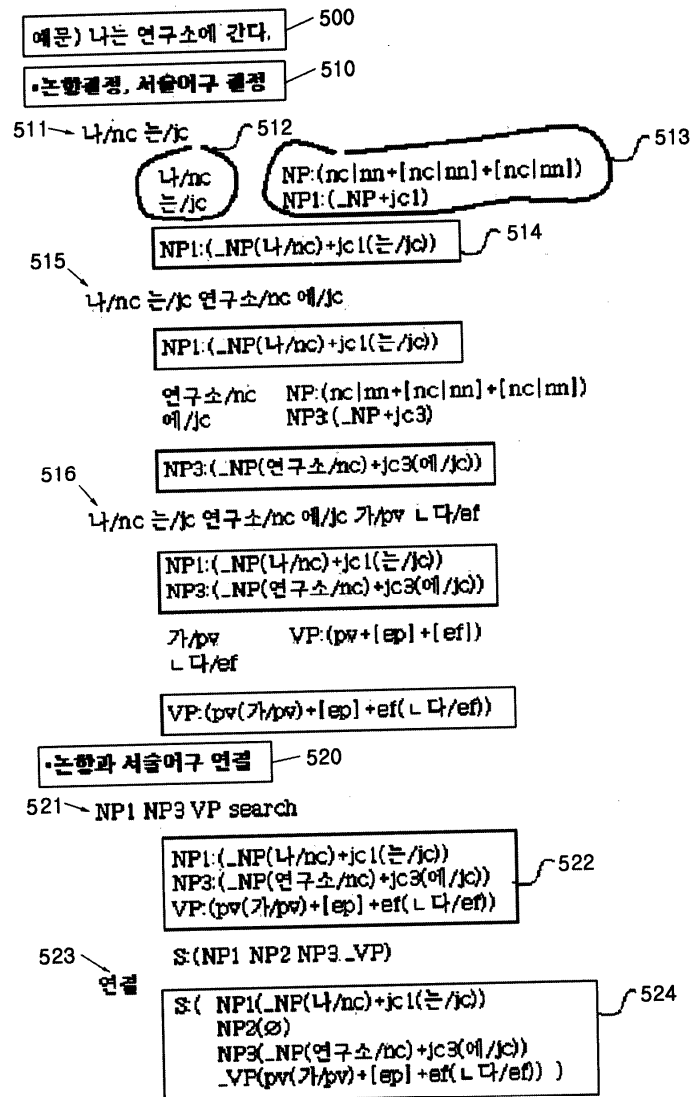
도면4a



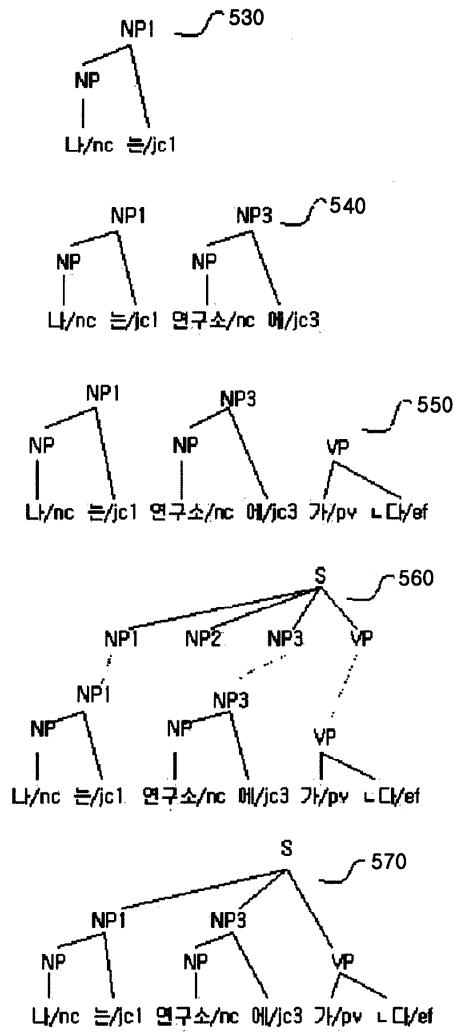
도면4b



도면5a



도면5b



도면6

폐수식어⇒NP1		610
(1) 불편한 OS는 없어진다.	: [(OS가) 불편한] OS는 없어진다.	
폐수식어⇒NP2		620
(2) 바이러스가 침투한 시스템은 포맷이 낮다.	: [바이러스가 (시스템을) 침투한] 시스템은 포맷이 낮다.	
폐수식어⇒NP3		630
(3) 사용자 중에는 인터넷에 필요한 통신장비를 사지 않는 사람이 많다.	: 사용자 중에는 [인터넷에 (통신장비가) 필수적인] 통신장비를 사지 않는 사람이 많다.	
안은문 주어⇒ NP1 ~ 폐수식어⇒NP2		640
(4) 맥킨토시가 인식하는 디바이스가 따로 있다.	: 맥킨토시가 [(맥킨토시가 디바이스를) 인식하는] 디바이스가 따로 있다.	
안은문 주어⇒ NP1		650
(5) 서버가 해킹을 당한 사실을 스스로 인식한다.	: 서버가 [(서버가) 해킹을 당한] 사실을 스스로 인식한다.	
NP+in ETNS의 NP⇒ ETNS의 NP		660
(7) 프로그래머는 보여준 사실을 몰랐다.	: 프로그래머는 [(프로그래머가) 보여준] 사실을 몰랐다.	
후행결의 논항 추정		670
(1) 1번 프로세스가 프로그램을 설치하면 관리를 하여준다.	: 1번 프로세스가 프로그램을 설치하면 [1번 프로세스가] 관리를 하여준다.	
선행결의 논항 추정		680
(2) 버그가 존재하여 프로그램을 수정하였다.	: 버그가 [프로그램에] 존재하여 프로그램을 수정하였다.	