

# 한국어 초록 작성의 자동화에 관한 연구

— 미생물학분야 학술지의 논문을 대상으로 —

## A Study on the Automatic Abstracting System for Journal Articles in Korean in the Field of Microbiology

이 태 영 \*

### □ 목 차 □

- |                          |                     |
|--------------------------|---------------------|
| 1. 서 론                   | 3.3.1 연구목적문장의 모형    |
| 1.1 연구의 목적               | 3.3.2 방법·결과문장의 모형   |
| 1.2 연구 방법 및 제한점          | 4. 자동초록시스템의 설계 및 구현 |
| 2. 이론적 배경                | 4.1 자동초록시스템 개요      |
| 2.1 초록 후보문장의 발취          | 4.2 논문문장의 분석        |
| 2.2 초록의 구조분석             | 4.2.1 품사분석기         |
| 3. 초록문장의 모형:             | 4.2.2 구분석기          |
| 3.1 초록문장의 역할구조           | 4.2.3 절분석기          |
| 3.2 미생물학분야 표본논문의 초록문장 분석 | 4.3 초록문장 생성         |
| 3.2.1 첫문장의 분석            | 4.3.1 연구목적문장        |
| 3.2.2 중간문장의 분석           | 4.3.2 방법·결과문장       |
| 3.3 초록문장의 수리적 모형         | 5. 결 론              |

### 초 록

자연언어처리 문법 중 격문법, 개념의존문법, PATR-II, DCG를 응용하여 미생물학 분야에 대한 한국어 자동초록시스템의 구축을 연구하였다. 표본으로 선택된 초록들을 분석하여 초록의 구조가 연구목적문장, 방법·결과문장, 결론문장의 순으로 이루어지는 것을 알았고 이를 초록의 이상적인 구조형식으로 정하였다. 본문에서의 문장발취와 발취된 문장들이 초록의 문장으로 정제되도록 하는 과정에서 중요한 역할을 하는 명사역할자질과 그 자질을 부여하는 규칙을 개발하였다. 그리고 발취된 문장들을 재정렬시키기 위한 재정렬규칙, 초록문장을 생성하기 위한 문장 생성규칙을 두었다.

그리고 본 시스템의 방법론으로 작성된 자동초록들을 평가한 결과, 본 시스템의 방법론이 저자초록과 같은 수준에 이르려면 정밀한 역할자질의 확립과 문장 생성규칙에 대한 보다 깊은 연구가 필요하다고 사려된다.

### ABSTRACT

This study proposes a Korean automatic abstracting system in microbiology by applying Case Grammar, Concept Dependency Grammar, and Unification-Based Grammar(PATR-II, DCG). The sample abstracts are analyzed to clarify the ideal structure of abstract—a purpose sentence as first sentence, 2-3 method and result sentences as middle sentences, and a conclusion sentence as last sentence. To extract and refine the representative sentences constructing an automated abstract requires the rules giving the role features to nouns. And the rules rearranging the extracted sentences and the rules generating the abstract sentences are also required.

Evaluating the efficiency of this system, the method used in this automatic abstracting system needs the more precise role features and the rules of sentence generation to reach the level of the author abstracts.

## 1. 서 론

### 1.1 연구의 목적

처음으로 초록을 자동으로 작성하고자 하였던 이는 룬(H. P. Ruhn)이다.<sup>1)</sup> 그가 1958년도에 이 논제를 제시한 이후 여러 사람들이 그 뒤를 이어 연구하였고 이들이 발표한 자동초록 작성의 연구결과는 초록을 작성하는 데 여러가지 유익한 정보들을 제공하였다.<sup>2)</sup> 특히 핵심문장 및 논제를 자동적으로 추출하는 면에서는 괄목할 만한 지식을 초록자에게 주었으며 또 실제로 발췌문 형태의 초록작성도 가능하게 만들었다. 그러나 조사된 바에 의하면 현재까지는 수작업으로 작성되는 초록처럼 만족할 만한 수준의 초록을 생성하지는 못하고 있는데, 그 이유는 초록자가 초록을 작성하기 위하여 취하는 일련의 행위 중에서 논문의 문장을 분석하고 이로 부터 초록문장을 생성하는 언어처리 과정을 기존의 방법들로서는 해결할 수가 없었기 때문이다. 이점을 보완하기 위해서는 정보검색시스템의 하나인 질문응답시스템 등에서 응용하는 자연언어처리 기술을 접목시킬 필요가 있다.

이러한 문장의 분석 및 생성과 같은 언어학적인 문제와 더불어서 효과적인 자동초록시스템의 구축을 위해서는 명료한 초록을 구성하는 초록문장들의 형식적인 모형 정립과 이들 문장들의 서술순서에 대한 연구가 필요하다. 그동안 연구된 초록의 구성상 특징을 살펴보면 위에 열거한 종류들을 초월하여 공통분모적인 것

이 존재하고 있음을 시사한다.<sup>3)4)</sup> 즉 (1)연구의 목적과 관점, (2)방법론, (3)획득된 결과, (4)유추된 결론 등과 같은 사항들은 초록이 갖추어야 할 요소들로 규정된다. 그리고 하나의 초록을 구성하는 각 문장들의 서술순서에 준하고 있다는 사실을 알려 준다.

그러므로 자동초록시스템에서는 이러한 사항들을 보다 명확히 규명하고 일반화하여 시스템에 내장을 시켜야만 자동으로 생산된 초록이 비로소 그 기능을 다할 수 있다고 보는 것이다.

상기한 두 문제 다음으로 고려해야 할 요인은 주요 문장 및 논제를 추출하는 데 필요한 객관적인 방법론의 설정이다. 기존의 몇가지 기법-표제어 기법, 문헌내 소재 기법, 단서어(端緒語, cue word) 기법, 구문 의미적 기법, 통계적 기법-을 적절히 혼합하여 사용하는 발췌 방법론의 확립이 중요하며 이 중 특히 구문 의미적 기법은 통합기반문법(Unification-Based Grammar, 이후 UBG라 약칭함)에서<sup>5)</sup>

- 1) H. Borko, *Abstracting Concepts and Methods*, New York : Academic Press, 1975, pp. 160-161.
- 2) C. D. Paice, "Constructing Literature Abstract by Computer : Techniques and Prospects", *Information Processing & Management*, Vol. 26, No. 1, (1990), pp. 171-186.
- 3) J. E. Rowley, *Abstracting and Indexing*, London : Clive Bingley Ltd., 1982, pp. 9-27.
- 4) D. B. Cleveland & A. D. Cleveland, *Introduction to Indexing and Abstracting*, Littleton : Libraries Unlimited Inc., 1983, p. 108.
- 5) S. M. Shieber, *An Introduction to Unification-Based Approaches to Grammar*, Stanford : CSLI, 1986, p. 14

강조하는 자질(feature)을 응용할 수 있게 보완되어져야 한다.

따라서 본 연구는 초록작성의 자동화를 위하여 제기된 앞의 세 가지 문제점에 대하여 자연과학분야인 미생물학 논문의 본문과 초록을 대상으로 연구하여 자연과학 전 분야에 걸쳐 사용될 수 있는 자동초록시스템의 원형적인 모형을 구축하는 데 목적이 있다.

## 1.2 연구방법 및 제한점

미생물학분야의 연구논문을 실험대상으로 선택하여 초록의 구조를 분석하였으며, 초록과 본문의 분석과 문장생성을 위한 도구로서는 변형 규칙을 쓰지 않는 UBG중 PATR-II와 한정절문법(Definite Clause Grammar, 이후 DCG로 약칭함)을 혼용하였다. 특히 이들 문법의 필수요소인 자질구조(feature structure) - 품사, 격, 결합 등 - 는 문장을 이루는 요소인 단어들의 구문적 결합관계와 의미적 결합관계를 구분하여 파악하였다. 또한 시스템에서 설정된 규칙의 형식을 표시하고 초록의 문형을 정립하기 위해서 DCG에서 사용하고 있는 문법구조를 밝히는 술어해석적 표현식을 응용하고 형식언어에서 언어관계를 규명하는 수리적 언어표현을 도입하였다.

초록후보문장의 발췌는 기존의 방법외에 관용화된 중요 구나 절을 지식베이스 형태로 시스템에 기억시켜 중요 문장 식별에 사용하였다.

본 연구의 수행상 제한적인 사항들은 다음과 같다. 한국어 문법에서 품사구분은 7품사, 8품사, 9품사 등 여러가지로 나뉘는데 본 연구에

서는 기존의 문법에서 나누던 방식을 수정하여 본 시스템의 설계에 편리하게 형용사와 동사를 합하여 용언으로, 부사 중 전이어구 역할을 하는 '그리고'류는 연결부사, 나머지는 일반부사로 분류한다. 그리고 격과 자질들은 되도록 상세히 분류되 격은 조사의 기능을 중심으로 구분하고 자질들은 각 단어들이 갖는 품사적 속성, 각 품사 간의 결합관계, 그리고 쓰여진 문장 또는 문단에서 수행하고 있는 역할을 구별하여 단어 및 절과 문장의 자질구조를 파악하도록 하였다. 한편 초록 후보문장 발췌는 기존의 중요문장 발췌기법에 구문 의미적 방법을 첨가 응용하여도 완전하게 일반화된 규칙을 정립하지 못하는 한계가 있었다.

## 2. 이론적 배경

### 2.1 초록 후보문장의 발췌

초록문장의 발췌를 위해 그동안 여러가지 기법들이 대두되었는데 그것들을 정리하면 통계적 기법, 단서어 기법, 표제어 기법, 문헌내 소재기법, 구문·의미적 기법으로 요약된다.<sup>6)</sup>

본 연구에서는 초록 후보문장을 발췌하는 데 있어 통계적 기법에 속하는 룬<sup>7)</sup>과 ASCI-

6) C. Borkowski & J. S. Martin, "Structure, Effectiveness and Benefits of LEXtractor, an Operational Computer Program for Automatic Extraction...", JASIS, Vol. 26, No.2, (1975), p. 94.

7) H. P. Ruhn, "The Automatic Criterion of Literature Abstracts", in Pioneer of Information Science (ed. by C. K. Schultz), New York : Spartan Book, 1968, pp. 118 - 125.

matic 연구에서<sup>8)</sup> 사용한 주요어(대표어휘)와 비주요어(비대표어휘)를 대비하여 초록 후보 문장을 발췌하는 기법을 단순화시켜 주요어(후술하게 될 명사와 동사 또는 형용사의 역할 자질로 대치됨)만의 수로서 초록 후보문장을 판단하며, 표제어는 발췌된 초록 후보문장들이 불충분할 때에 후보문장에 연관되는 문장을 찾아 후보문장을 보완하기 위해 사용한다. 그리고 초록 후보문장을 발췌하는 데 결정적인 계기를 주는 단서어들을 별도로 모아 그 기능을 다하게 하였다. 또한 박센데일이 연구한 결과를<sup>9)</sup> 토대로 하여 한 문단에서 초록 후보문장이 다수 출현하였을 시에는 첫 문장과 마지막 문장을 선택하도록 하였다. 구문분석적 접근법에서는 일이 연구한 결과(각 단어에 품사만을 부여한 상태에서 문장의 유사패턴을 찾는 형식 이어서 실패하였음)를<sup>10)</sup> 본 연구에서는 단어에 의미적 자질(후술하게 될 역할자질)을 주어 이를 보완 사용하였다.

## 2.2 초록의 구조분석

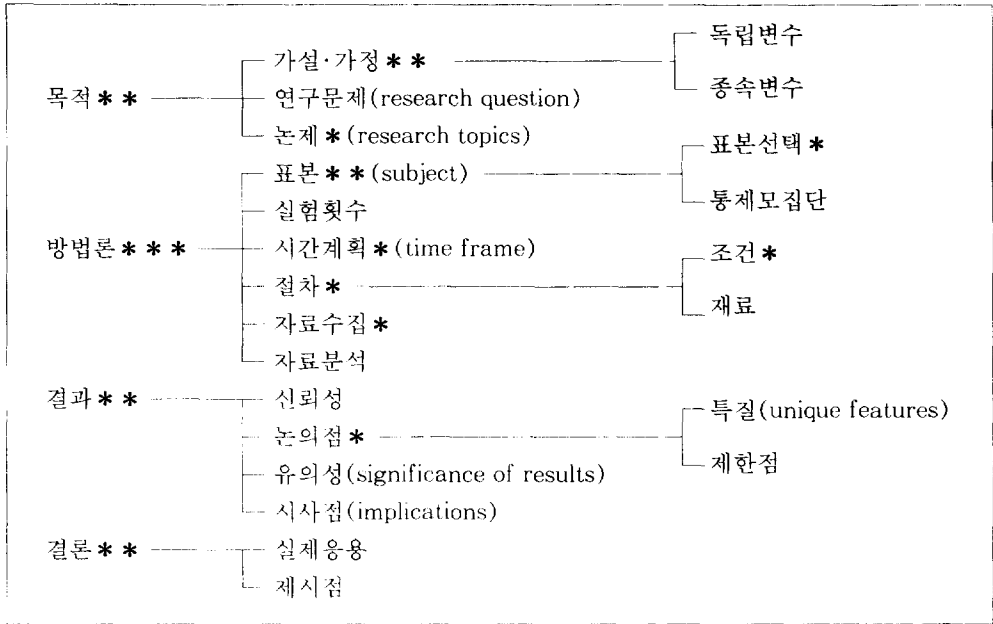
본고에서 연구대상으로 삼는 것은 지시 내지 통보초록의 형태이다. 초록문장은 서론에서도 언급하였듯이 초록의 성격상 일반적으로 쓰여진 문장들과 형식과 구조면에서 큰 차이를 보이며 또한 과학논문의 문장들과도 차이를 나타낸다. 과학논문에 쓰여지는 문장들은 일반문장보다 표현의 다양함에 있어 많은 제약을 받는다. 예를 들면 대화형의 문장—‘—하였지’, ‘—무엇을 하는거야?’ 등—과 존대법의 문장들은 전혀 사용되지 않고 ‘사려된다’, ‘제시하였

다’와 같은 설명문의 형식이 그 특색을 이룬다. 초록의 문장은 이러한 경향이 더욱 두드러지며 논문에서 내용을 전개하는 서론, 이론적 배경, 실험방법, 연구동기, 연구결과, 결론, 제시점 등 아주 핵심적인 내용을 축약하여 놓은 요약문이다. 따라서 이와 같은 초록문장은 구조상 초록을 이루는 각 문장들의 구성요소들이 핵심적인 내용을 표현할 수 있게 조직되어야 한다.

여기서 이러한 구성요소들을 조사하고 분석한 리디(E. D. Liddy)의 연구 결과를 살펴보면 다음과 같다. 그녀는 ERIC 데이터베이스에서 150개의 초록과 PsycINFO 데이터베이스에서 126개의 초록의 발췌하여 이를 두 분야의 초록자 각각 4인에게 20편씩을 의뢰하여 문장의 구성요소를 판정하게 하였다. 초록자들이 응답한 문장의 구성요소들을 정리한 것은 〈표 1〉과 같다.<sup>11)</sup>

- 
- 8) IBM Corp., Advanced Systems Development Division, ASCI-Matic Auto-Abstracting Project, Final Report, Vol. III, Yorktown Heights, N. Y. : IBM Corp., 1961. (H. Borko, op. cit., p. 170에서 재인용)
- 9) P. B. Baxandale, "Machine-made Index for technical Literature—an Experiment", IBM J. of Research and Development, Vol. 2, No. 4, (1958), pp. 354-361. (R. E. wylly, "Extracting and abstracting by computer", in Automated Language Processing (ed. by H. Borko), New York : John Wiley & Sons, 1967, p. 141에서 재인용)
- 10) L. L. Earl, "Experiments in Automatic Extracting and Indexing", Inf. Stor. Retr., Vol. 6, No. 6, (1970), pp. 313-334.

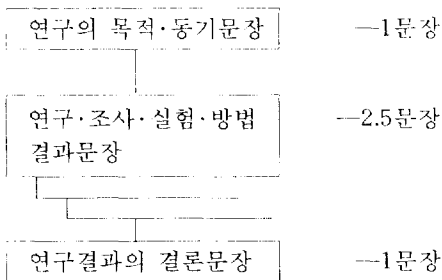
〈표 1〉 초록의 구성요소들



(여기서 \*\*를 친 요소들은 어느 초록에서나 흔히 발견되는 문장의 전형적 구성요소들이고 \*를 친 것들은 그 다음으로 보편적 구성요소들이며 아무 표시도 되어 있지 않은 것들은 세밀하게 분류할 때에 쓰이는 정밀한 구성요소로 작용한다.)

### 3. 초록문장의 모형

〈그림 1〉 초록의 역할구조



#### 3.1 초록문장의 역할구조

초록의 역할구조는 표본초록들을 분석한 결과 〈그림 1〉과 같이 파악할 수 있었다.

이렇게 세 유형으로 나누어지는 초록의 문장을 생성하기 위해서는 초록을 구성하는 각 문

11) E. D. Liddy, The Discourse-Level Structure of Natural Language Texts : An Exploratory Study of Empirical Abstracts. Syracuse, 1988. Syracuse University Ph. D. Dissertation, p. 116.

장의 구문구조와 그 구문들의 역할을 밝힐 필요가 있다. 우선 구문구조는 격(조사의 역할)<sup>12)</sup>, 명사와 용언의 결합자질<sup>13)</sup> 등을 포함하고 또 초록에서 가해지는 반복어구의 회피 등과 같은 제한점을 수용하면 그 구조를 밝힐 수 있다. 그런데 문장의 구문들의 역할을 분석하는 데는 문장의 역할인 “목적, 결과, 결론” 정도의 단순한 구분으로는 문장의 구문들의 역할 즉 문장의 역할구조를 정확히 파악할 수 없다고 본다. 그러므로 보다 구체적인 역할구조의 분석이 필요한데 이를 위해서 2장에서 소개한 리디가 분석하여 놓은 구성요소를 응용할 수 있다. 리디가 분석한 구성 요소는 영어 문장에서 술어와 결합할 때 <표 2>와 같은 양상을 보인다.<sup>14)</sup> <표 2>를 보면 명사구 등의 구성요소들이 술어를 사이에 놓고 결합되는 예가 나와 있는데 ‘결론’이란 구성요소는 술어의 종류에 따라 ‘발견점’, ‘논의점’, ‘가설’ 등과 결합한다. 이것으로 보아 술어가 문장구성 시 각 구성요소의 결합에 중요한 역할을 한다는 것을 알 수 있다.

이것은 생크(R. C. Schank)<sup>15) 16)</sup>의 개념의존 문법 이론에서 기본행위(동사)를 중심으로 문장의 형식구조를 파악하였던 것을 상기시키며 또한 기본행위가 문장의 역할구조 파악에 필수적임을 시사한다. 따라서 문장 분석 및 생성 시 문장의 역할구조를 파악하기 위하여 표본초록을 분석한 결과 다음의 <표 3>과 <표 4>와 같은 명사(구)와 용언들의 역할자질을 얻었다.

이상과 같이 초록문장을 구성하는 단어 및 구의 역할자질에 대한 개념을 기술하였는데, 문장구성 형식은 단어와 구로만 이루어지는 단문보다 절과 절이 복합되어 이루어지는 중문과

복문의 형식이 대중을 이룬다. 따라서 절과 절 사이의 연결형식에 대한 구문연결구조에 대한

- 12) 본 논문에서 설정하는 격관계는 여러 연구에서 흔히 공통적으로 출현한 격을 택하고 아울러 조사 다음에 오는 조사상당어구를 포함한 격도 적절히 사용하였다. 즉, “가, 이, 께서, 는, 은”→주격, ... “을, 를, -에 대하여, -에 대한”→대상격, “께, 더러, 보고, 한테, 에게, 에게로, 한테로”→처소 목표격, ... “-듯이, -과 같이, -과 비슷하게, -와는 다르게, -에서 알 수 있듯이”→유추격 과 같이 설정하였다.
- 13) 명사와 용언의 결합자질은 명사와 용언이 의미적으로 결합할 때 오류를 제거하기 위한 자질이다. 예를 들면 “산을 먹는다”는 문장은 통사적으로는 틀린 문장이 아니지만 의미적으로는 틀린 문장이다. 여기서 ‘산’을 ‘과자를’로 바꾸면 ‘과자를 먹는다’라고 되어 의미가 통하는 하나의 문장이 된다. 그리고 “산을 먹는다”에서 ‘먹는다’를 ‘추량한다’로 대체하면 “산을 추량한다”가 되어 의미가 성립된다. 따라서 우리는 ‘산’이나 ‘과자’란 명사와 ‘먹는다’, ‘추량한다’의 동사는 각기 서로 의미적으로 결합할 수 있는 상대가 한정되어 있음을 알 수 있다. 그러므로 명사들이 용언과 결합할 때 똑같은 의미적 결합 양상을 보이는 명사들을 한 집단으로 묶어서 문장의 생성 규칙에 적용시켜야 하는데 이 집단화시키는 수단으로 적절한 수준의 상위어를 찾게 되며 그 적절한 수준의 상위어가 바로 자질값이 되게 된다. 이것은 용언의 경우도 마찬가지여서 용언도 특정한 명사집단과 의미적으로 결합가능한 종류의 용언들을 묶을 필요성이 생긴다.
- 14) E. D. Liddy, op. cit., p. 186.
- 15) R. C. Schank & Mallory Selfridge, “How to Learn/What to Learn”, IJCAI-77, (1977), pp. 8-14.
- 16) R. C. Schank & R. P. Abelson, “Scripts, Plans and Knowledge”, IJCAI-75, (1975), pp. 151-157.

〈표 2〉 명사구 등의 구성요소와 술어의 결합관계

구성요소	관계의 어휘적 표현	구성요소
결론 (conclusion)	로부터 유추되었다(derived from) → ← 귀결되었다(used to reach)	발견점 (finding) 《 》

(여기서 ←, →는 “conclusion is derived from finding.” “finding used to reach conclusion”으로 해석하는 기호이다.)

〈표 3〉 명사어구의 역할자질 구분

기호	약칭	역할자질의 내용	리디와의 비교
x1	【연목】	연구목표 *	목적
x2	【가정】	가정 및 가설설정	가설
x3	【배경】	연구의 배경적 설정 *	배경
x4	【연체】	연구 및 실험대상체	
x5	【실물】	실험물	
x6	【실패】	실험용매, 실험매개체 《 》	

(이 역할자질은 초록의 문형정립에 필요하며 문장생성에 관여한다. ‘\*’한 자질은 문장, 절, 구, 단어의 구성소적 성격을 나타낸다. ‘\*’가 없는 자질은 구, 단어의 구성소적 성격을 나타낸다. 【연체】는 【실물】과 【실패】로 나누어 질 수 있다. 즉 【실물】과 【실패】가 없는 문에서는 【연체】하나만을 사용하며, 그렇지 않은 문에서는 세 개의 자질이 동시에 사용될 수 있다.)

〈표 4〉 용언역할자질 구분

기호	약칭	역할자질의 내용	용언 단어 종류
y1	【목적형】	연구목적 동기	조사하다, 분석하다, 측정하다, 시도하다, -고자 하였다
y2	【과정형】	연구 실험 조사 과정	도입하다, 넣다, 주사하다, 통하다, 따르다
y4	【결과형】	연구 실험 조사 결과 《 》	높았다, 낮았다, 있었다, 증가하다

분석이 이루어져야 한다. 보통 문장안에서 절과 절이 연결될 때에는 부절의 용언이 어미변화하여 서로 이어지며, 용언의 어미변화에 따라 “원인결과(-하여, -니, …)”, “관찰추측(-아, -볼 때, …)” 등을 나타내는 등 그 내포하고 있는 뜻이 있다. 본 연구에서는 이 용언의 어미변화가 내포하고 있는 연결성을 구문연결자질로 삼았다.

이 용언의 구문연결자질과 앞의 용언역할자질은 사실상 그것이 속한 절의 주요부자질을 결정하는 요소로 작용한다. 이것은 절의 주요부는 술어가 대변하도록 하였기 때문인데 실제로 하나의 절이 나타내고 있는 구문연결과 역할은 바로 문장에서 그 절이 수행하고 있는 의미를 나타내고 있다고 판단된다. 따라서 용언역할자질과 용언의 구문연결자질을 절의 주요부자질로 삼고 문장의 최종적 주요부자질은 주절의 용언역할자질과 구문연결자질로 대변한다.

### 3.2 미생물학분야 표본논문의 초록문장분석

먼저 전체적으로 보면 초록은 평균 4.5문장을 보유하고, 문장은 평균 20.8개의 단어(복합어는 한단어로 취급)였으며 한 문장의 절들의 복잡정도는 3.5인 특징을 나타내었다.

#### 3.2.1 첫 문장의 분석

첫 문장은 평균 15.7개 단어에 51.5개 음절이며 <표 5>와 같이 분석되었다. 본 모형의 기본구조는 문장의 구문구조를 밝히는 것이 그 목적이기 때문에 기술의 편의상 명사역할자질,

격조사, 용언결합자질, 구문연결자질만 모형의 자질구조에 포함시켰다. 그 표현식은 <(명사역할자질, 격조사), …(용언결합자질, 용언역할자질, 구문연결자질)>와 같다. <표 5>의 A에서 보면 주어, 목적어, 술어의 자질구조는 직접적으로 절의 표현식(<>안의 부분)에 들어가고 보어부분(【처소격】, 【소유격】, 【동격】, 등)은 (가), (나)로 인도되는 표현식으로 나타내어 (가), (나)의 표현식을 참조하게 되어 있는데 그것은 문장분석의 통계상 주어 또는 목적어와 술어가 결합되어 있는 절이 절대 다수로서 문장 구문구조의 중심이 되었기 때문이다.

<표 5>의 <모형 1>은 <예문 1>의 문장을 역할자질분석 1과 같이 분석하여 얻었다.

#### <예문 1>

Hemerocallis 속 식물 두 종 (<골잎원추리와 노랑원추리>)에 대해서, 분열증기의 핵형을 분석하였다.

#### <역할자질분석 1>

{【연체】[에] 대해[서], 【속성】[의] 【연목】[을] 【연목형】[였다]}

이상의 모형분석에서 결과를 정리하면 주절의 술어들은 “조사하였다, 연구하였다, 시도하였다, 측정하였다” 등이며, 부절의 술어들은 “-(에) 대한, -(에) 의한, -(에) 관한, -(에) 대해서, -(에) 관하여, -기 위해서, -한 뒤” 등이 출현한다. 그리고 명사역할자질은 <표 3>의 “【결과】, 【결론】, 【견점】”을 제외한



〈표 5〉 연구목적문장 모형 예

A. 자질구조 중심의 절단위 모형	
(1) < (가), (나), ([연목], [을, 를]), (vc1, [연목형], ed) >	
(가) ([부위], [에서])	— Cs
(나) ([속성], [의])	
(2) < (가)+(나)+([연체], [에]), (va2, [ ], zc) >	
(가) ([속성], [인])	
(나) ([연체], [와, 과])	— C1
<<	
>>	
ed : 종결형, mo : 수식형, ec : 대등연결형, zj : 양보서술형, zc : 원인결과형, zg : 제시선택형,	
B. 실제문장중심의 모형	
<<	
모형 1. —( (C1) Cs )	
>>	

전 자질값들이 첫 문장에 올 수 있음을 알았다.

3.2.2 중간 문장의 분석

중간 문장들은 평균 23.1개 단어에, 76.2개 음절이며 보통 2.5개의 문장이 나열되고 연구 조사 실험방법과 과정, 결과, 발견점, 해답, 비교 등이 기술된다. 아래의 〈예문 2〉와 〈역할자질분석 2〉는 중간문장들 중의 하나를 분석한 것이다.

〈예문 2〉

pKM101과 pSL4의 mutator 기능과 보호효과는 (repair 기능에 따라) 다른 나 전반적으로 (두 플라스미드의 UV

와 MNNG에 대한) 저항성과 돌연변이 율을 증가시켰고 pSL4는 pKM101보다 그 효과가 높았다.

〈역할자질분석 2〉

{ [실물][의] [연목][는] ([속성][에] 따라) [상태형][나] ([연체] | [에] 대하[는]) [연목][을] [과정형] [있고] [실물]1[는] [실물]2[보다] [결과][가] [결과형][있다] }

이를 정리하면 중간 문장은 주절의 술어로 “나타냈다, 높았다, 보였다, 낮았다, 증가시켰다, 감소시켰다” 등이 출현하며, 부절의 술어는 “-볼 때, -한 뒤” 등이며, 명사역할자질

로는 【연체】, 【실물】, 【결과】, 【실매】, 【표본】, 【절차】, 【수준】, 【시간】, 【속성】, 【부위】가 온다. 【연목】은 드물게 출현한다.

### 3.3 초록문장의 수리적 모형

언어에서 문장의 생성 가지수는 무한대로 무한집합의 성격을 띠운다고 말하고 있다. 문자열 집합  $\{a, b, c\}$ 가 각 문자열을 연결하여 문장을 이루는 가지수는  $aaa, bbb, aabbcc, \dots$ , 식으로 전개하면 무한대에 이르나 문장에서 단어들의 결합상의 특수성 때문에 집합  $\{a, b, c\}$ 가 주어지고 실제어휘가 주어지면 유한집합으로 전환된다. 여기서 집합의 가지수를 좌우하는 것은 단어들의 문장구성을 위한 결합속성-자질-들이다. 예를 들면 “생쥐를 실험한다”, “생쥐가 실험된다”는 문장을 두고 생각할 때 “((명사)를, 용언-능동형), ((명사)가, 용언-피동형)”으로 그 자질들이 분석되는데 이와 같은 문장의 자질구조의 집합은 문장의 생성을 위해 “생쥐(명사), 실험-(용언)”이란 실제 어휘가 주어졌을 때 역으로 문장생성을 위해 자질구조의 통합을 시도하면 당연히 “생쥐를 실험한다, 생쥐가 실험된다”란 문장만이 생성되는 유한집합으로 귀결된다.

이와 같은 맥락으로 명사역할자질과 용언역할자질이 문자열집합에 부여되면 마찬가지로 유한집합의 성격을 띤다. 집합  $\{a, b, c\}$ 에  $a$ 는 용언  $y_1$ -【연목형】,  $b$ 는 명사  $x_1$ -【연목】,  $c$ 는 명사  $x_{12}$ -【부위】로 한정되면 자연히  $x_1$ 은 【목적격】을,  $x_{12}$ 는 【처소격】,  $y_1$ 은 ‘조사하였다’와 같은 용언으로 범위가 줄어들기 때문에<sup>17)</sup>

그 집합은 {{【동기형】, 【연목】, 【부위】}}로 되고 문장의 생성 가지수는  $(bca), (cba)$ 와 같이 둘로 제한되는 유한적인 성격을 띤다.

그러므로 초록의 문장들도 우선 각 문장이 수용할 수 있는 모든 자질들을 포괄하는 전집합을 설정하면 그 집합내의 문자열들이 부분집합을 이루어 파생하는 문장들-초록의 문장모형들-을 유한적인 수치내에서 유도하여 낼 수 있다. 그리고 이들 문장의 모형에 특히 가해져야 할 제한적인 규칙을 적용시키면 규범적인 모형이 자동으로 설정된다.

#### 3.3.1 연구목적문장의 모형

앞 절의 첫 문장에서 출현한 연구목적문장 모형의 전 집합적 수리모형구조는 다음의 〈수리모형 1〉과 같다(표 5)에서 생략된 모형도 모두 포함하였음).

- 17) 명사역할자질인 【연목】과 【부위】를 초록문장에서 구분하는 근거는 용언역할자질과 명사의 격자질에 의존한다. “답자균 효모에서 속간 원형질체 융합을 시도하였다”란 문장의 용언 ‘시도하였다’는  $y_1$ -【연목형】 용언역할자질이고 그 앞의 ‘속간 원형질체 융합’이란 복합명사구는 대상격이면서 앞으로 진행될 연구의 목표적 역할을 한다. 따라서 용언역할자질의  $y_1$ 이 출현하고 그 앞에 【대상격】(대부분이 추상어)명사가 왔을 때 이 명사(구)는 【연목】이라고 할 수 있다. 또한 ‘답자균 효모에서’는 ‘에서’란 조사의 처소적 역할로 말미암아 【연목】이 일어나는 장소를 가르키고 있다. 그러므로 ‘에서’를 유발하는 【부위】와 【연목】이 어느 부분집합에 구문요소로 부여되면 그것들로 인하여 파생되는 문장은 매우 제한적인 유한집합으로 된다.

## 〈수리모형 1〉

S → C1..Cn

C → NP VP

NP → N2 Cas

VP → Vo Vi Vs

〈N2〉 = {x1, x4, x5, x6, x9, x12, x13, x14}

〈Cas〉 = {ob, a1, po, eq, s1, t1}

〈Vo〉 = 〈V2〉

〈Vi〉 = {ten, voi}

〈Vs〉 = {ed, ec, mo, zc, zk}

〈V2〉 = {y1, y2, y3}

〈S〉 = 〈C(1 = 〈n = 〈4〉〉〉

〈C〉 = 〈W(2 = 〈n = 〈16〉〉〉

〈Ten〉 = {r, a}

〈Voi〉 = {j}

〈Ed〉 = [-다]

〈Ec〉 = [-고]

〈Mo〉 = [-한, -는, -ㄴ]

〈Zc〉 = [-해서, -하여]

〈Zk〉 = [-한 뒤]

위에 사용된 기호 S는 문장, C, C1, Cn은 절, W는 단어, N2는 명사역할자질, Cas는 격, Vo는 용언의 어근, Vi는 선어말 어미, Vs는 어미, V2는 용언역할자질을 나타내고, Ten, Voi는 각각 시제, 양태, 능동 피동 사동의 정보들을 표현한다. 그리고 〈N2〉, 〈Cas〉, 〈Vs〉, 〈V2〉의 오른쪽 항에 있는 집합기호(집합내에 있는 자질값을 왼쪽의 자질들이 택할 수 있다는 뜻임)로 싸여져 있는 기호들은 부록의 '1. 기호일람표'를 보면 나와있다. 〈Voi〉는 능동·피동·사동을 말하는 자질로서 '-하다'는 'j', '-되다'는 'k',

'-시키다'는 '1'로 표기한다. 그리고 〈C(1 ≤ n ≤ 4)〉와 〈W(2 ≤ n ≤ 16)〉은 절이 1-4개, 단어가 2-16개 있다는 뜻이다.

이러한 전 집합적 모형에서 〈예문 3〉을 도출시킨 역할자질구조와 그 수리적 모형은 각각 〈역할자질분석 3〉과 〈수리모형 2〉와 같다.

## 〈예문 3〉

Mutator 기능을 가지는 플라스미드 pKM101과 이의 돌연변이체 pSL4를 DNA 수복능력이 다른 Esherichia coli B/r(phr-, recA-, uvrA-, uvrB-) 균주들에 도입하여 돌연변이원인 UV와 MNNG에 대한 보호효과와 mutagenicity를 조사하였다.

## 〈역할자질분석 3〉

〈 (〔속성〕1[을] [상태형][는]) [실물][를] (〔속성〕2[이] [상태형][ㄴ]) [실주][들에] [과정형][여] (〔속성〕1[인] [연체][에] 대하[ㄴ] [연목][를] [연목형][였다]) 〉

## 〈수리모형 2〉

〈N2〉 = {x1, x4, x9, x14}

〈V2〉 = {y1, y2, y3}

〈S〉 = 〈C(1 = 〈n = 〈4〉〉〉

〈C〉 = 〈Ws(2 = 〈n = 〈16〉〉〉

〈Ten〉 = {r, a}

〈Voi〉 = {j}

〈Ed〉 = { [연목형] }, [-다]

〈Mo〉 = { [상태형] }, [-는]

$$\langle Zc \rangle = \{ \text{【과정형】}, [-하여] \}$$

여기서  $x1=ob \in Cas$ , (명사역할자질의 【연목】( $x1$ )이 【대상격】( $ob$ )으로 되어 있다는 뜻임),  $x4=ob \in Cas$ ,  $x9=(ob, su) \in Cas$ ,  $x4=a1 \in Cas$ 와 같은 명사역할자질의 격성질과  $y1=VC1, y2=VC4, y3=VD6$ 와 같은 용언역할자질의 자질구조가 상기의 〈예문 3〉을 생성시켰다.

### 3.3.2 방법·결과문장의 모형

앞 절의 중간 문장에서 출현하는 방법·결과문장의 모형들을 근거로 하여 작성된 방법·결과문장의 전 집합적 수리모형구조는 다음의 〈수리모형 3〉과 같다.

〈수리모형 3〉

$$S \rightarrow C1..Cn$$

$$C \rightarrow NP VP$$

$$NP \rightarrow N2 Cas$$

$$VP \rightarrow Vo Vi Vs$$

$$\langle N2 \rangle = \{ x1, x4, x5, x6, x9, x12, x13, x14 \}$$

$$\langle Cas \rangle = \{ su, ob, cp, a1, po, eq \}$$

$$\langle Vo \rangle = \langle V2 \rangle$$

$$\langle Vi \rangle = \{ ten, voi \}$$

$$\langle Vs \rangle = \{ ed, ec, mo, zc, zd, zj \}$$

$$\langle V2 \rangle = \{ y1, y2, y3 \}$$

$$\langle S \rangle = \langle C (1 = \langle n = \langle 4 \rangle) \rangle$$

$$\langle C \rangle = \langle W (2 = \langle n = \langle 16 \rangle) \rangle$$

$$\langle Ten \rangle = \{ r, a \}$$

$$\langle Voi \rangle = \{ j \}$$

$$\langle Ed \rangle = [-다]$$

$$\langle Ec \rangle = [-고]$$

$$\langle Mo \rangle = [-한, -ㄴ]$$

$$\langle Zc \rangle = [-해]$$

$$\langle Zd \rangle = [-아]$$

$$\langle Zj \rangle = [-나]$$

이러한 전 집합적 모형에서 앞절의 〈예문 2〉를 도출시킨 역할자질구조와 그 수리적 모형은 각각 앞절의 〈역할자질분석 2〉와 아래의 〈수리모형 4〉와 같다.

〈수리모형 4〉

$$\langle N2 \rangle = \{ x1, x4, x5, x14, x15 \}$$

$$\langle V2 \rangle = \{ y1, y2, y3 \}$$

$$\langle S \rangle = \langle C (1 = \langle n = \langle 4 \rangle) \rangle$$

$$\langle C \rangle = \langle Ws (2 = \langle n = \langle 16 \rangle) \rangle$$

$$\langle Ten \rangle = \{ r, a \}$$

$$\langle Voi \rangle = \{ j \}$$

$$\langle Ed \rangle = \{ \text{【결과형】}, [-다] \}$$

$$\langle Ec \rangle = \{ \text{【과정형】}, [-고] \}$$

$$\langle Mo \rangle = \{ \text{【 】}, [-한] \}$$

$$\langle Zj \rangle = \{ \text{【상태형】}, [-나] \}$$

여기서  $x1=su \in Cas$ ,  $x4=a1 \in Cas$ ,  $x5=(po, su, cp) \in Cas$ ,  $x14=(ob, a1) \in Cas$ ,  $x15=su \in Cas$ 와 같은 명사역할자질의 격성질과  $y2=(VA2, VB9, VC2)$ ,  $y3=VD6$ ,  $y4=VD4$ 와 같은 용언역할자질의 자질구조가 앞의 〈예문 2〉를 생성시켰다.

#### 4. 자동초록시스템의 설계 및 구현

본 연구에서 목표로 하는 자동초록시스템은 2-3장에서 고찰된 사항들과 지식을 응용하여 설계하고 구현하였다. 특히 DCG를 토대로 하여 작성되는 문법규칙들은 본 시스템의 구현을 위하여 필요한 모든 자질들을 DCG의 용어구조(term structure) 즉 자질구조(feature structure)내에 포함시킨다. 그러나 본 장에서 문법규칙을 기술할 때에는 인식의 편리상 용어구조내에 포함되는 자질 중 프로그램 구현 상에서는 필요하나 규칙 설명에서는 그리 중요치 않은 자질들은 생략하였다.

##### 4.1 자동초록시스템 개요

본 자동초록시스템의 기본 개요는 <그림 2>와 같다.

자동초록작성의 대상이 되는 논문의 문장들이 입력되면 본 시스템은 그 문장들을 분석하고 절 또는 문장 별로 재정렬을 한 다음에 주요 문장을 발췌하고 발췌된 문장을 토대로 하여 초록의 분장을 생성한다.

논문 문장의 분석은 다음의 절차를 거쳐서 이루어진다.

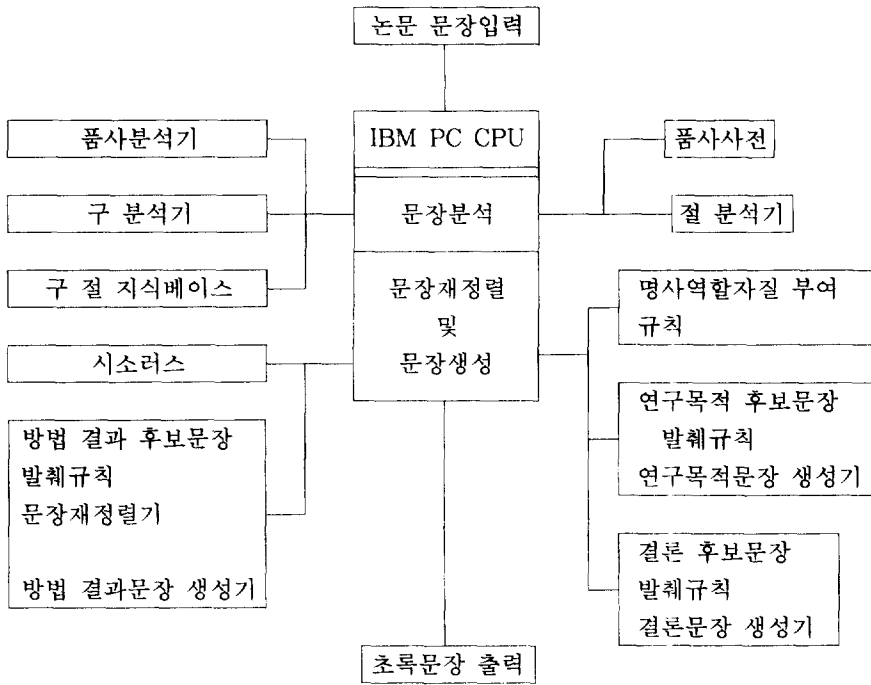
- (1) 입력된 문장에서 먼저 한 문장을 구성하고 있는 기본단위인 단어를 식별하고 그 단어에 품사를 부여한다. 단어에 품사를 부여하는 작업은 품사분석기를 통해 수행하며 이 때에 품사사전이 참조된다.

- (2) 품사를 부여받은 단어들은 구 분석기와 구 지식베이스에 의해 인접한 다른 단어들과 결합하여 구를 생성할 수 있을 때에는 구를 생성하게 된다.
- (3) 생성된 구와 구를 형성하지 못한 단어들은 절 분석기와 절 지식베이스를 통해 절 단위로 묶어져 문장의 구문구조를 이루게 된다.(본 연구에서는 이 낱개 단어들도 프로그램 작성의 편리상 광의의 구로 취급한다.)
- (4) 그리고 문장의 구문구조를 구성하는 구문요소(단어, 구)에 명사역할자질 부여 규칙으로부터 명사역할자질 값이 부여된다.

이렇게 분석되어진 문장으로부터 초록의 연구목적, 방법·결과, 결론 문장이 생성되는데 그 과정은 다음과 같다.

- (1) 연구목적문장은 연구목적 후보문장 발췌규칙에 의거해 해당 문장을 발췌한 후 연구목적문장 생성기에 의해 문장을 생성한다.
- (2) 방법·결과문장은 복합문일 때는 절 단위로, 단일문일 때는 문장 단위로 재정렬 규칙에 의해 입력된 수서와 문장내에서의 순서에 입각하여 정렬시키고 정렬된 문장(절)들에서 발췌규칙을 사용하여 방법·결과 후보문장을 발췌한 후 방법·결과문장 생성기에 의해 문장을 생성한다.
- (3) 결론문장은 결론 후보문장 발췌규칙으

〈그림 2〉 자동초록시스템의 기본구성도



로해당 문장을 발취한 후 결론문장 생성기로 문장을 생성한다.

이와 같이 자동초록이 생성되는데 본 기술에서는 초록문장의 생성 예로 연구목적문장과 방법·결과문장으로 한정하였다.

## 4.2 논문문장 분석

### 4.2.1 품사 분석기

단어의 품사 해석과정은 각 단어들을 각 품사사전들과 비교하여 해당하는 품사와 자질들을 부여받게 하는 과정을 거친다. 본고에서 구

분하는 품사의 종류는 명사(완전명사와 불완전명사로 구분), 대명사, 용언, 부사, 연결부사, 조사, 수사, 관형사로 나눈다. 다음의 〈그림 3〉은 품사 분석기의 개요도이다.

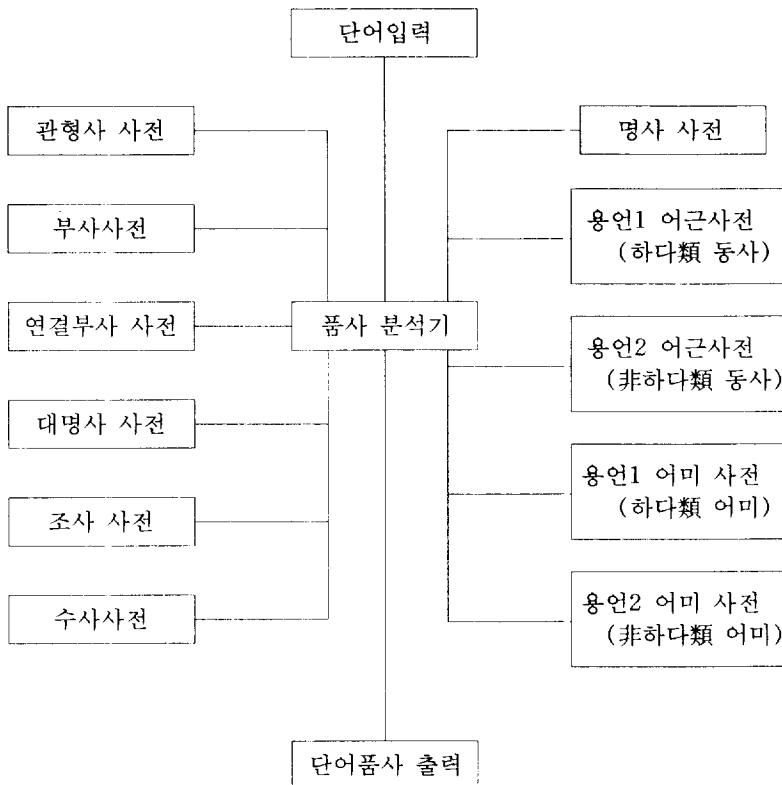
또한 용언을 하다類와 非하다類로 나누는 이유는 용언의 어미가 변화하여 어근과 결합할 때 용언 '측정하다'는 '측정하여', '측정한', '측정해'와 같이 '측정' 이후가 '하(여, 며, 고...)', '한(할)', '해(서, 도...)'의 3가지 유형의 형태변화를 하기 때문이다. 반면에 용언 '말기다'는 '말기고'와 같이 어근 '말기'가 어미 '고'와 결합하는 것이 있는 반면에 '하다'의 '해'처럼 어미 '어'가 어근의 '기'와 결합할 때 '겨'와 같이

축약형태를 이룬다. 그런데 문제는 非하다類의 용언들의 어미변화 축약형이 하다類의 ‘해’처럼 같은 형태소로 변화하지 않는다는 점이다. ‘떨어지다’는 ‘저’의 형태소를 가지므로 이것들을 어미사전에 일일이 열거하여야 하므로 비효율적이다. 한편 ‘어렵다’는 ‘ㅂ’이 탈락한 ‘(위)서’와 같은 어미변화를 하는데 이런 ‘ㅂ’탈락 어미변화 용언들은 프로그램상에서 특별한 조

치를 취해야 해결된다.

이러한 문제들을 해결하기 위해서 非하다類 용언들은 풀어쓰기형태로 사전이 만들어져야 한다. 즉 ‘마ㅌ겨’, ‘ㄸㄱㄹㅇㄱㅈㄱ’로 풀어쓰기식을 도입하면 통일적인 어미변화 형태소 ‘ㄱ’ 하나로 이들 용언의 어미변화를 나타낼 수 있는 것이다. 따라서 본고에서는 하다類와 非하다類의 용언사전을 분리하여 작성한다.

〈그림 3〉 품사 분석기 개요도



4.2.2 구 분석기

4.2.2.1 접속부사구

문장의 서두에 출현하여 전 문장과의 흐름을 연결시켜 주는 “그렇게 하여”와 같은 접속어 구들을 접속부사구로 삼는다.

4.2.2.2 특수 용언구

용언+불완전명사, 용언+불완전명사+용언, 명사+용언과 같은 것을 특수 용언구로 설정하였다.

- (보고와 같-) → 한 용언구로 인식함
  - (보고와 다르-) → 한 용언구로 인식함
  - (-을 때 나타나-) → 한 용언구로 인식함
- 예 : 처리하였을 때 나타나는

4.2.3 절 분석기

절의 최소한의 필수조건은 명사와 동사, 형용사가 동시에 존재하여야 한다. 한 문장안에 절(부절)이 존재하려면 주절에 직접적으로 포함이 되지 않는 상기조건외의 구문요소가 있어야

한다. 그것을 충족시킨 절은 문장 안에서 쓰인 용법에 따라 중문의 경우 대등절, 복문의 경우 수식절, 명사절, 부사절로 나뉜다. 본 연구에서는 자연언어처리상 “..., 나타난 반면에”와 같은 문장형식에서 ‘나타난 반면에’를 한 용언구로 처리하여 문장분석 및 생성에 도움이 되도록 하였기 때문에 절 분석에서 이러한 사항이 참조가 된다. 그리고 절의 의미를 내포하는 절 주요부자질을 설정하는데 여기에서는 용언역할 자질과 용언의 어미변화에 따른 구문연결자질을 포함시킨다.

기본적으로 다른 구문연결자질을 포함시킨다.

한문장안에 “주어·목적어+용언, 주어·목적어+용언”의 형태가 반복이 될 때와 “주어·목적어, 주어·목적어+용언, 용언” 일때 콤마로 구분된 “주어·목적어+용언” 각각을 절로 인식한다. 아래의 <예문 4>는 그말의 절 식별 같이 수식절과 주절로 나뉘고 <예문 5>는 주절과 부사절로 나뉜다.

<예문 4>

Pleurotus ostreatus에서 분리된 laccase의 보결단을 추정하였다  
 명사구 + 용언구( ) + 명사구 + 용언구  
 | — 수식절 — | | — 주 절 — |

<예문 5>

숫쥐를 soybean oil과 beef tallow를 먹이고 관찰하였다  
 명사구 + 명사구 + 용언구(고) + 용언구  
 | — 부사절 — |  
 | ————— 주 절 ————— |



## 〈표본 1〉 서론의 ‘본고는’류 문장 예1

본연구에서는 *Sch. castellii* glucoamylase의 유전자의 구조를 밝히는 기초로서 glucoamylase를 분리정제하였으며 이 효소의 특성, 아미노산 조성 및 아미노말단 부위의 아미노산서열을 조사하였다.

## 〈표본 2〉 저자초록의 연구목적문장 예1

*Schwanniomyces castellii*로부터 glucoamylase를 정제하고 그 특성을 조사하였다.

## 4.3 초록문장 생성

## 4.3.1 연구목적문장

초록의 연구목적문장을 생성하려면 논문에서 연구목적 후보문장들을 발췌하고 그 문장들을 분석합성하여야 한다. 먼저 연구목적 후보문장을 발췌하여 문장의 구성요소-단어나 구-에 역할자질들을 부여하면, 어떤 논문의 연구목적 후보문장은 초록의 연구목적문장이 되는데 알맞은 구성요소를 갖고 있는 경우가 있고, 어떤 논문의 후보문장은 구성요소가 너무 많거나 또는 적은 경우가 있다. 여기서는 알맞은 경우를 기술하고자 한다.

## 4.3.1.1 연구목적 후보문장발췌

서론의 말미를 보면 “본고(에서)는.....—을 조사하였다”와 같이 그 논문에서 추구하는 목적, 방법 등을 기술한 부분이 있는데 바로 이 문장이 초록의 첫 문장인 연구목적 문장을 형성시켜주는 길잡이 역할을 한다. 논문 30편을 조사한 결과에 의하면 ‘본고는’류의 문장과 초록의 첫 문장은 그 내용상 매우 유사하였다.<sup>18)</sup>

실제로 〈표본 1〉의<sup>19)</sup> ‘본연구에서는’으로 연결되는 문장과 〈표본 2〉의 저자초록의 연구목적문장은 〈표본 1〉에서 ‘본연구에서는’, “유전자의 구조를 밝히는 기초로서”, “아미노산조성 및 아미노산말단부위의 아미노산서열”을 빼면 〈표본 2〉의 문장과 같아진다.

그러므로 문장분석을 거쳐 절의 용어구조가 완성된 문장이 입력되면 그 서두에서 다음의 〈표 6〉의 왼쪽 부분에 속하고 있는 ‘본고는’류의 단어와 문장의 말미에서 〈표 6〉의 오른쪽 부분에 나와 있는 용어의 유형이 있는가를 확인하여 이 두 종류의 단어가 존재하면 그 문장

18) 서론의 ‘본고는’류의 문장은 초록의 연구목적문장의 생성을 위하여 아주 적절한 문장이다. 그런데 문제는 서론에서 ‘본고는’류의 문장을 갖고 있지 않은 논문들이다. 그러나 이런 문장이 존재하지 않는 논문은 표본 30편을 조사한 결과 1편에 불과하여 통계적으로 무시하여도 괜찮은 수 이어서 그러한 논문의 연구목적문장 작성은 차후의 논제로 하기로 한다.

19) 배석, 외, “*Schwanniomyces castellii* Glucoamylase의 정제 및 성질”, *KOR. JOUR. MICORBIOL.*, Vol. 29, No. 2, (1991), pp. 104-110

〈표 6〉 연구목적 후보문장의 단서어

시작부분	끝 부분
본고는	— 하고자 한다(하였다).
본고에서는	조사하였다.
본 논문은	— 하는데 있다.
본 연구는	제시한다.
본 연구에서는	다루었다.

을 초록의 연구목적 후보문장으로 상정한다.

발체에 대한 규칙은 PART-II를<sup>20)</sup> 응용하여 서술하면 다음의 〈발체규칙 1, 2〉와 같다.

〈발체규칙 1〉

X → Y

Y → NP... VP

NP → N P

VP → Vo Vi Vs

〈NP 주요부〉=〈N 주요부〉

〈VP 주요부〉=〈Vo 주요부〉

〈VP [주격]주요부〉=〈NP 주요부〉

(‘본고는’류의 문장에서 본고類가 [주격]으로 사용될 때 적용된다.)

〈발체규칙 2〉

X → Y

Y → NP... VP

NP → N P

VP → Vo Vi Vs

〈NP 주요부〉=〈N 주요부〉

〈VP 주요부〉=〈Vo 주요부〉

〈VP [처소격]주요부〉=〈NP 주요부〉

(‘본고는’류의 문장에서 본고類가 ‘본고에

서는’과 같이 【처소격】으로 사용될 때 적용된다.)

#### 4.3.1.2 명사역할자질 부여

‘본고는’류의 문장이 발체되면 명사구와 절에 명사역할자질을 부여한다(용언과 수사구, 부사구는 문장분석시 이미 사전에서 역할자질을 부여받았다). 명사구나 절에 명사역할자질을 부여해야 하는 이유는 3장에서 설명되었다. 즉 문장이 갖고 있는 문맥의존적인 명사역할자질은 어떤 구성요소들의 집합이 각기 결합하여 문장을 생성하려고 할 때 문장생성의 가지수를 미리 알려주는 지표가 된다.

이와 같은 명사역할자질을 문장의 구문요소에 부여하는 규칙의 기술은 DGC<sup>21) 22)</sup> 표현방

20) S. M. Shieber, op. cit., pp. 24-25.

21) F. C. N. Pereira & D. H. D. Warren, “Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison With Augmented Transition Networks”, Artificial Intelligence, Vol. 13, pp. 231-278.

22) F. C. N. Pereira & S. M. Shieber, Prolog and Natural-Language Analysis, Stanford: CSLI, 1987.

식을 사용하되, 본 장에서는 인식이 용이하도록 명사구 용어구조(term structure)에 포함되는 자질들은 명사역할자질, 명사결합자질, 격, 그리고 실제 어구인 명사와 조사를 용어구조내에 포함시키고, 용언구의 용어구조내에 포함되는 자질들은 용언역할자질, 구문연결자질, 그리고 실제 어구로 제한하여 표현한다. 나머지 품사는 품사만을 표시한다. 각 단어나 구의 용어구조는 ( )로 표시되고 순서대로 나열되며 해를 얻기위한 부분은 X, Y 등 변수로 나타내고 불필요한 자질은 ‘\_’이나 ‘[ ]’로 넘어간다. 명사구에서 조사가 없을 경우는 ‘무’라고 해당 부위에 표기한다. 그리고 논항(자질)의 수를 꼭 명시할 필요가 없을 때에는 “...”로 표시한다. 예를 들면 np(【연체】, \_, 【동격】, 【식

물】, 무), np(【연체】...), vp(【연목형】..W)와 같은 표현식을 쓴다.

이러한 기술원칙으로 표본 1)의 ‘본고는’류 문장에 명사역할자질을 부여하면 다음과 같이 된다.

- 
- 21) F. C. N. Pereira & D. H. D Warren, “Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison With Augmented Transition Networks”, Artificial Intelligence, Vol. 13, pp. 231 - 278.
- 22) F. C. N Pereira & S. M Shieber, Prolog and Natural-Language Analysis, Stanford : CLSI, 1987.

### 명사역할자질 부여규칙 (A)

#### 〈부여규칙 1〉

〈 [ ], np(【속연】,\_, 【처소격】, 본연구, 에서는), np(X, 【화물】, 【소유격】, Sch. castellii glucoamylase, 의), np(Y, 【생부】, 【소유격】, 유전자, 의), np(Y, 【관념】, 【대상격】, 구조, 를), vp(【연목형】, 【조건】, 밝히는 기초로써) )

→ X=【연체】, Y=【연목】

( Y로 걸린 ‘유전자의’와 ‘구조’는 명사구 형성규칙에 의해 ‘유전자의 구조’로 재서됨)

( “밝히는 기초로써”는 용언구로 되어짐, 즉언구로 되어짐, 즉 〈 vp(..【수식】,), np(【속연】, )〉일 때는 한 용언절로 인식한다.)

#### 〈부여규칙 2〉

(【연체】, 【화물】, \_, X[Sch. castellii | Y[glucoamylase], - )

if X=【생물】

→ X=【실주】, Y=【연체】

#### 〈부여규칙 3〉

〈 d(이), np(X, 【성질】, 【동격】, 효소의 특성, \_), np(Y, \_, 【동격】, 아미노산조성, 및), np(Z, 【부분】, 【대상격】, 아미노산 말단 부위의 아미노산 서열, 을), vp(【연목형】, 【종결】, 조사하였다) )

→ X=【연목】, Y=【연목】, Z=【연목】

(생략된 조사 추정 후방향 규칙에 의거하여 【대상격】을 가져옴)

〈구문분석례 1〉

본연구에서는 Sch. castellii glucoamylase의 유전자의 구조를  
 【속연】 【실주】 【연체】 【연목】  
밝히는 기초로써 glucoamylase를 분리정제하였으며 이 효소  
 【연목형】(조사상당어구) 【연체】 【과정형】 【연  
 의 특성, 아미노산 조성 및 아미조말단 부위의 아미노산서열을  
 【연목】 【연목】 【연목】  
조사하였다.  
 【연목형】

명사역할자질 부여가 끝난 〈표본 1〉의 문장은 다음의 〈구문분석례 1〉과 같이 된다.

4.3.1.3 연구목적문장 생성

서론의 ‘본고’류의 문장은 ‘연구(조사)하였다’와 같이 ‘연구형’인 것과 ‘—됨을 밝혔다’처럼 ‘공표형’인 것이 있다. 초록의 연구목적문장으로는 ‘연구형’을 택하여 서론의 문장에서 ‘공표형’인 것이 있다. 초록의 연구목적문장으로는 구목적문장 생성시 ‘연구형’으로 전환시킨다. 연구목적문장은 그 구성이 [수식절]\*

{[도입절]+[과정절]+[마감절]}과 같이 되며 필수적으로 갖추어야 하는 명사역할자질은 【연목】, 【연체】, 【실물】, 【실매】, 【실주】이고 용언역할자질은 【연목형】이다. 여기서 도입절에 흔히 나타나는 절구조는 아래의 (1)이며, (4)는 수식으로 절의 명사 앞에 첨가될 수 있다. 과정절은 (2)의 형식으로 많이 출현하며 역시 (4)가 의미확장 구실을 하여 준다. 마감절은 (3)이 절 형식이며 (4)가 수식역할을 한다. 여기서 N2는 명사역할자질, V2는 용언역할자질, V3는 구문연결자질이다.

- (1) N2 = { 【실주】(x9), 【연목】(x1), 【연체】(x4) }  
 V2 = { 【과정형】(y2), 【연목형】(y1) }  
 V3 = { 【원결】(zc) → 「을(기) 위해(하여), -에 관해(하여)-에 대해(하여)」 }  
 파생되는 절 형식 예  
 (가) { 【실주】|의| 【연목】|을| 【연목형】 }

〈예문 6〉 S. castelli의 돌연변이율을 조사하기 위하여  
 【실주】 【연목】 【연목형】

·  
·  
·

- (2) N2 = { 【실주】, 【연체】, 【실매】(x6), 【방법】(x7), 【절차】(x8) }  
 V2 = { 【과정형】 }  
 V3 = { 【원결】, 【관추】(zd) → [... 도입하여, -했을 때] }
- (3) N2 = { 【연목】 }  
 V2 = { 【연목형】 }  
 V3 = { 【종결】(ed) }
- (4) N2 = { (1), (2)의 n2에 준하는 것이 출현함 }  
 V2 = { 【과정형】, 【상태형】(y3) }  
 V3 = { 【수식】(mo) }

이와 같은 구성요소를 갖는 절들이 도입, 과정, 마감절의 순서를 이행하여 결합되면 연구 목적문장이 되는 것이다. 단 도입절과 과정절은 ‘본고는’류의 문장에서 둘 다 존재하거나 또는 둘 중에 하나만 존재하는 경우가 있다.

후자의 경우는 존재하는 해당절만 연구목적문장에서 생성된다. 이러한 범용적인 모형과 원칙을 준수하여 <표본 1>에서 초록의 연구목적문을 만드는 생성규칙의 예는 다음과 같다.

**연구목적문장 생성규칙**

<문장 생성규칙 1>

〈 np(【속연】..본고, ), np(【연목】...)...vp(【연목형】, 【조건】, 밝히는 기초로써) 〉

<예문 7>

본연구에서는 Sch. castellii glucoamylase의 유전자의  
 【속연】            【실주】            【연체】            【연목】  
구조를 밝히는 기초로써  
 【연목형】

<문장 생성규칙 2>

〈 dp(이), np(【연목】, 【성질】, 【동격】, 효소의 특성, ..), np(【연목】, 【성질】, 【동격】, -, -), np(【연목】, 【추상】, 【대상격】, , ), vp(【연목형】..) 〉  
 → 〈 dp(그), np(【연목】, 【성질】, 【대상격】, 특성, 을), vp(【연목형】..) 〉

이러한 규칙으로 인하여 생성되는 연구목적문장은 다음과 같다.  
 “Sch. castellii glucoamylase의 유전자의 구조를 밝히기 위하여 glucoamylase를 분리정제하

였으며 그 특성을 조사하였다.”  
 위의 자동생성문장은 <표본 2>와 비교하여 보면 문장이 긴 흠이 있으나 전달하고자 하는 의미는 일치하고 있다.

### 4.3.2 방법·결과문장

방법·결과문장 생성은 방법·결과 후보문장 발췌, 발췌된 문장들의 재정렬, 문장생성의 순으로 진행된다.

#### 4.3.2.1 방법·결과 후보문장 발췌

방법·결과 후보문장을 발췌하는 방법은 연구의 결과를 서술하고 있는 논문의 '결과 및 고찰'장에서 명사역할자질이 【연체】, 【연목】, 【실매】 등으로 나타나는 명사 단서어를 포함하고 있고 용언역할자질이 【결과형】이며 동시에 구문연결자질이 【종결형】인 단서어를 갖는 문장을 방법·결과 후보 문장으로 일차 상정한

다. 특히 연구목적 후보문장('보고는'류의 문장)에서 주절의 주요부인 【연목형】 용언(조사, 측정, 실험, 연구)이 종속절인 부사절에서도 쓰였을 경우 그 종속절을 따로 한 의미항으로 취급하여 초록의 방법·결과문장 생성에 이용한다.

표본으로 선정된 논문의<sup>23)</sup> 결과 및 고찰부분의, 문장들에 아래와 같은 명사역할자질 부여규칙을 적용하여 자질값을 부여한다.

---

23) 배석, 외, op. cit., pp. 104 - 110.

#### 명사역할자질 부여규칙 (B)

<부여규칙 4>

< np(【실주】, -, 【시처격】, -,에서), np(X, 【관념】, 【도처격】, 절단회복, 예), vp(【상태형】, -,관계하는) >, np(【주속】, -, -, -, 는)  
→ X=【연목】

<예문 8>

Escherichia coli에서 excision repair에 관계하는  
    【실주】  X  【상태형】  
uvrA와 uvrB 유전자는 uvrC와 더불어  
  【주속】

<부여규칙 5>

np(X, 【관념】, -, DNA 절단, -) use np(【연목】, 【관념】, -, DNA회복, -)  
→ X=【연목】

<부여규칙 6>

< np(【주속】, -, 【주격】, -, -)...c(np(【연목】, 【관념】, 【도처격】, -, DNA 절단, -), vp (【상태형】, -, 필요한) >, < np(X, 【유기】, 【목적격】, endonuclease, -), vp(【과정형】, -, 활성화시킴으로써) >  
→ X=【실매】



### 질 생성규칙

〈질 생성규칙 1〉

〈 np(..【대상격】, 1.), vp(..【수식】.), np(..【소유격】, 2.), np(..【대상격】, 3.) 〉

→

〈 np(..【소유격】, 2.), np(..【주격】, 3.), np(..【대상격】, 1.), vp(...) 〉

〈예문 9〉

<u>repressor LexA 단백질</u>	질을	절단하는	<u>protease의 기능</u>	을
1			2	3

〈질 생성규칙 2〉

〈 np(..【주격】, 1.), np(..【도처격】, 2.), vp(..【수식】), np(..【대상격】, 3.) 〉

→

〈 np(..【주격】, 1.), np(..【도처격】, 2.), vp(...) 〉

〈질 생성규칙 3〉

〈 np(..【시처격】, 1.), np(..【도처격】, 2.), vp(..【수식】), np(..【동격】, 3, 와),  
np(...【주격】, 4, 는), np(..【동격】, 5, 와 더불어) 〉

→

〈 np(..【동격】, 3, 와), np(..【동격】, 4, 및), np(..【주격】, 5, 는), np(..  
【시처격】, 1), np(..【도처격】, 2), vp(...) 〉

·  
·  
·

위에 예시된 질 생성규칙 등을 가지고 부록의 (2)와 (3)문장을 재정렬시키면 다음의 재정렬문 (A)와 같다. (재정렬문에서 np와 rp는 논항 중 뒷 부분에서 두 논항이 실제 어휘이며, vp, qp는 끝 논항이 실제어휘이고, 다른 품사구의 용어구조는 괄호 안의 논항이 실제어휘이다.)

이 중에서 방법·결과문장 생성에 작용하는

절들은 용언의 【결과형】을 가지며 명사역할자 질 중 【연목】을 갖고 【실주】, 【실매】, 【연체】 중 2개 이상의 명사역할자질을 포함하여야 한다. 이러한 절들을 방법·결과문장 생성의 대상 절로 1차 선정하면 (8), (9), (15)절이 된다.

이러한 방법으로 부록의 (4)에서 (16)문장들을 중요한 절들만 간단히 정리하면 다음의 재정렬문 (B)와 같이 된다.



재정렬문 (A)

- (1) < np(【부매】, 【유기】, -, protease, 의), np(【속연】, 【관념】, -, 기, 능, 은), np(【부매】, -, -, repressor LexA 단백질, 을), vp(【과정형】, 【종결】, 절단하) >

<예문 10>

protease의 기능은 repressor LexA 단백질을 절단하는  
 【부매】    【속연】                    【부매】                    【과정형】  
 .  
 .

- (8) < < (7)의 도치문장 >, np(【실매】, -, -, UV, 와), np(【실매】, -, -, MNNG, 에 대하여), np(【실주】, -, 【구치격】, wild type\*(H/r30R)\*수준, 으로), np(【연목】, -, -, 보호효과, 를), vp(【결과형】, -, 나타내었고) >

- (9) < np(【실매】, -, -UV, 와는 다르게), np(【실매】, -, -, MNNG, 에 대하여), np(【연체】, -, -, pSL4, 는), np(【연체】, -, -, pKM101, 보다), qp(높은), np(【연목】, -, -, 보호효과, 를), vp(【결과형】, -, 나타내었다.) >

- 15) < vp(【과정형】, 【수식】, 도입된), np(【연체】, -, -, pKM101, 과), np(【연체】, -, 【주격】, pSL4, 는), np(【실매】, -, 【도치격】, UV, 에), np(【연목】, -, -, 보호효과, 를), vp(【결과형】, -, 나타내었다.) >

재정렬문 (B)

- (16) < np(【연체】..pKM101, 과), np(【연체】..pSL4, 는), np(【실매】..UV, 에), np(【연목】.. 보호효과, 를), vp(【결과형】..나타내었고) >

<예문 11>

pKM101과 pSL4는 UV에 보호효과를 나타내었고  
 【연체】    【연체】 【실매】 【연목】    【결과형】

- (17) < ap(특히), np(【실주】, -, 【시치격】, uvrA -균주, 에서), np(【연체】, -, 【주격】, pSL4, 는), np(【연체】, -, 【비교격】, pKM101, 보다), np(【실매】, -, 【대상격】, UV, 에 대하여), qp(높은), np(【연목】, -, 【대상격】, 보호효과, 를), vp(【결과형】.. 나타내었다.) >

- (19) < np(【실주】..uvrB-균주, 에서는), np(【실매】..MNNG, 에 대하여), np(【연체】..

pKM101, 과), np(【연체】..pSL4, 는), np(【연목】.. 저항성, 을), vp(【결과형】.. 보였고) )

(24) < np(【실주】.. HS30R \* (uvrA-) \* 균주, 와), np(【실주】.. HS30R \* (uvrB-) \* 균주, 에서), np(【연체】.. 플라스미드 pKM101, 과), np(【연체】.. pSL4, 는), np(【연목】.. 돌연변이 유발능, 을), vp(【결과형】, -, 억제시켰으며) >

(25) < np(【실주】.. HS30R \* (uvrA-) \* 균주, 와), np(【실주】.. HS30R \* (uvrB-) \* 균주, 에서), np(【연목】.. npc(【연체】.. pKM101, 보다), np(【연목】.. 돌연변이 유발 억제 효과, 가), vp(【결과형】, -, 낮았다) >

(29) < < np(【연체】.. pSL4, 는), np(【연체】.. pKM101, 보다), ap(전방적으로), np(【속연】.. 기능, 이), vp(【결과형】, -, 강화되었음을), vp(【결과형】, -, 알 수 있었다.) >

여기서 용언역할자질의 【결과형】과 명사역할자질의 【연목】을 중심으로 절들을 정리한다. 방법·결과문장의 절 형식은 “단순 방법·결과절”, “비교 방법·결과절”로 나눌 수 있는데 그 구분은 방법·결과 후보절들이 재정렬을 하였을 때 “-보다”라는 구가 출현하였으면 비교 방법·결과문으로 상정을 하고 그렇지 않으면 단순 방법·결과문으로 한다.

비교 방법·결과문 :

9보, 17보, 18보, 20보, 25보, 28보, 27돌, 29전

(여기서 숫자는 절번호, 전은 전반적인 결과의 언급이고, 돌은 돌연변이율, 보는 보호효과에 대한 언급을 한 것임)

단순 방법·결과문 :

8보, 15보, 16보, 19보, 21돌, 22돌, 23돌,

24돌, 26돌

이러한 방법·결과문장을 문형→【연목】→【연체】→【실매】→【실주】→【결과형】 순서로 계층을 만들어 이원행렬화 하면 다음의 <표 7>과 같이 된다.

#### 4.3.2.3 방법·결과문장 생성

앞 절에서 재정렬된 절들은 방법·결과문장을 생성하기 위한 재료 역할을 하게 되는데 이들 단순 또는 비교문장들은 발생 숫자에 있어 최소일 때는 【실주】, 【연체】, 【실매】, 【연목】, 【결과형】이 모두 모여 하나의 문장을 이루는 전 집합적인 경우와 최대일 때는 경우의 수 만큼 문장이 발생하는 부분 집합적인 경우가 있다. 그래서 연구목적문장과 마찬가지로 방법·결과문장의 생성에 지침이 되는 범용적인 모형들을 정립하여 문장생성에 참조가 되도록 한

다. 방법·결과문장에 흔히 출현한 명사역할자 질은 【실주】, 【연체】, 【실매】, 【수준】, 【시간】, 【결과】, 【환경】, 【속성】, 【속연】이었으며, 용언 역할자질은 【과정형】, 【상태형】, 【결과형】이었

다. 그리고 문장은 ‘과정절+마감절’의 형식을 띄거나 ‘마감절’ 홀로 구성되어지며 각각 수식 절에 의해 수식된다. 이 두 경우에 빈번히 출현하는 절형식은 다음과 같다.

〈표 7〉 방법·결과 후보질의 이원행렬의 예

(명사는 명사역할자질과 실제 어휘만 기재하고 용언은 용언역할자질(【결과형】)만 기재 하는데 0은 ‘나타냈다’, +는 ‘높았다’, -는 ‘낮았다’를 표시한다.)

문 번호	【연목】		【연체】		【실매】		【실주】				【결과형】
	보호 효과	돌연 변이율	pKM 101	pSL 4	MN NG	UV	rec A-	uvr A-	uvr B-	phr-	
8	를		과	는	《 에 대하여		에서				0
16	를		과	는		에					0
17	를		보다	는		에		에서			0
19	를		과	는	에	《 대하여			에서		0
25	가		보다	는		《		와	에서		-
29	이	이	보다	는		《 》					+

(1) 마감절 :

를 나타내었다.

(가) { 【연체】[는, 은] 【실매】[에 (대  
한)] 【연목】[을] 【결과형】 }

【결과형】

(나) { (【연체】1, 【실매】1)[는] (【연  
체】, 【실매】, 【환경】, 【방법】)[에  
서] (【연체】2, 【실매】2)[보다]  
【연목】[이, 가] 【결과형】 }

예문 12 : pLS4는 MNNG에 보호효과  
【연체】 【실매】 【연목】

(다) {【실매】[에 의하여, 에 의한]【연  
목】[은] ((부사), 또는 【환경】[에  
의하여])【결과형】}

(라) {【연체】[의]【연목】[은]은【결  
과】[이다]}

(2) ‘과정절+마감절’

(가) {〈(【방법】, 【속성】 또는 【연목】)  
[을]【과정형】[비교할 때, ...]〉〈  
마감절〉 ((1)에 기술된 유형중  
하나가 음)}

이와 같은 방법·결과문장의 범용적인 문장  
형식을 준수하여 방법·결과문장을 생성하게  
된다. 생성시키는 방법·결과문장의 수는 보통  
과학논문의 초록문장이 4-5개인 점(동기문장,  
결론문장 포함)을 들러 2-3개를 만든다. 그런  
데 대상으로 올라 오는 절들이 5개 이상 발생  
한 경우에는 아래의 원칙에 의하여 4-5개를  
선택한다.

1. 자질2의 구성소를 많이 소유하고 있는 절  
순으로 선택한다.
2. 동일한 의미를 갖는 절이 반복하여 출현  
할 경우 그 절은 선택한다.
3. 선택된 방법·결과절이 원래 제삼의 절을  
이끌고 있었을 때 제삼의 절들이 부매를  
갖고 있을 경우 선택된 절과 결합시킨다.

이러한 원칙에 따라 재정렬된 방법·결과 후  
보절들에서 방법·결과문장을 생성하기 위한  
후보문장이나 절을 <표 7>에서 선정하면 우선  
전 집합적인 (29)가 대상으로 떠오른다. 선정  
된 전 집합적 문장이 비교문 하나이므로 ‘돌연

변이율’과 ‘보호효과’를 대표할 수 있는 후보문  
장을 위 원칙에 근거하여 발췌하면 <표 7>에  
서는 (16)이 선택된다. 또 하나는 <표 7>에는  
나와 있지 않지만 (24)로 정할 수 있다.

이렇게 후보문장이 발췌되면 결론문장과의  
연관관계를 고려하면서 논문의 순서대로 나열  
한다. 발췌된 후보문장으로부터 방법·결과문장  
을 생성하면 다음과 같다.

“uvrA-(HS30R)와 uvrB-(HS30R)균주  
들에서 pKM101과 pSL4는 UV에 보호효  
과를 나타내었고 돌연변이 유발능을 억제  
시켰다. pSL4는 pKM101보다 전반적으로  
기능이 강화되었다.”

이때 용언의 시제는 과거형을 사용하고 문장  
의 길이가 허용하는 한 대등절로 이어 기술하  
며 주어, 목적어 용언이 중복되면 재정렬규칙  
4, 5의 역순으로 접순하여 중복된 어구를 생략  
하여 간단한 복합문으로 만든다.

5. 결 론

이와 같은 방법과 절차에 의해 운영되는 본  
자동초록시스템을 가동하여 10편의 자동초록  
을 작성하고 그것들을 3인의 미생물학분야 전  
문가에게 평가를 의뢰한 결과는 다음과 같았  
다.

- (1) 저자초록에 대한 자동초록이 갖는 내용  
유사도는 약 70% 수준이었다.
- (2) 유사문장의 수도 약 70%이었다.
- (3) 자동초록의 문장의 의미흐름의 자연성  
정도는 60%의 초록의 의미흐름에 무리

가 없었다.

또한 미생물학분야의 자동초록시스템이 자연과학분야의 타 영역에서 적용이 될 수 있는지의 여부를 검증하기 위하여 미생물학분야에서 설정한 자질구조를 지질학과 전기공학분야에 실험하여 보았다. 자질구조의 적용성을 조사하기 위하여 지질학과 전기공학분야의 논문 각 3편씩을 분석한 결과 자동초록시스템을 구축하는 데 있어 근간을 이루는 명사결합자질, 명사역할자질, 용언역할자질의 자질구조상의 체계를 그대로 유지하되 다만 몇몇의 자질들만 추가시키면 타 분야들에서도 적용할 수 있는 것으로 판단할 수 있었다.

앞으로의 과제는 초록문장의 수리적 모형을 이용한 문법규칙을 강화하여 재정렬규칙과 문장생성 규칙중에서 구절의 변환 과정을 일일이 정리하는 체계를 개선하여 규칙들을 간소화하는 방법에 대한 연구가 있어야 한다. 그러기 위해 본 연구에서 설정한 역할자질의 수정, 개선 및 새로운 자질에 대한 연구와 본 연구에서 제시한 용언의 구문연결자질의 활용에 대한 연구가 필요하다고 생각된다.

### 참고문헌

강승식, 『자연언어 처리를 위한 통사분석기의 설계 및 구현』, 서울, 1988. 서울대학교 석사학위논문.  
권혁철, 『자연언어 처리를 위한 인식문장의 형식논리 표현』, 서울, 1987. 서울대학교 박사학위논문.  
금장철, 『중심어 주도형 구절구조 문법을 이용

한 한국어 구문구조 해석기의 구현』, 서울, 1987. 한국과학기술원 석사학위논문.

김명철, 『자연스러운 표층문 생성을 위한 한국어 표현 특성에 관한 연구-PIVOT방식 기계번역에서-』, 서울, 1987. 한국과학기술원 석사학위논문.

김영진, 『자질구조로부터의 자연언어 생성에 관한 연구』, 서울, 1990. 서울대학교 석사학위논문.

송민선, 『자연언어 이해 시스템을 위한 한글의 의미구조 및 추론에 관한 연구』, 서울, 1986. 서울대학교 대학원 석사학위논문.

윤덕호, 김영택, “세계의 기계번역 그 추세와 전망”, 정보과학회지, Vol. 7, No. 6, (1989), pp. 82-90.

이기용, “영한 기계번역체계 구축을 위한 소고”, 정보과학회지, Vol. 7, No. 6, (1989), pp. 10-24.

이인배, 『격시스템을 응용한 한글 자연어 질의어 처리 시스템에 관한 연구』, 서울, 1985. 한국과학기술원 석사학위논문.

이정현, 『한국어처리를 위한 구절문법과 질문응답시스템』, 인천, 1988. 인하대학교 박사학위논문.

정영미, “우리말 정보자료를 처리하는 지능형 정보검색시스템의 설계”, 정보관리학회지, Vol. 8, No. 2, (1991), pp. 3-31.

최원태, 『격문법을 이용한 자동 색인 및 탐색 확장에 관한 연구』, 서울, 1986. 연세대학교 대학원 석사학위논문.

- 최인숙, 『자동초록을 위한 지식기반시스템 설계에 관한 연구』, 서울, 1987. 연세대학교 대학원 석사학위논문.
- 최증민, 『자연언어 처리를 위한 한글 Conceptual Analyzer에 관한 연구』, 서울, 1986. 서울대학교 석사학위논문.
- 한성국, 『한·영 기계번역 시스템의 구현』, 인천, 1988. 인하대학교 대학원 박사학위논문.
- 허미숙, 『지식베이스를 이용한 자동색인시스템에 관한 연구』, 서울, 1991. 연세대학교 대학원 석사학위논문.
- Barr, Avron & Feigenbaum, E.A., The Handbook of Artificial Intelligence vol. 1, Reading : Addison - Wesley Pub. Co., 1981.
- Borko, Harold, Abstracting Concepts and Methods, New York : Academic Press, 1975.
- Borkowski, C. & Martin, J.S., "Structure, Effectiveness and Benefits of LEXtractor, an Operational Computer Program for Automatic Extraction...", JASIS, Vol. 26, No. 2, (1975), pp. 94 - 102.
- Earl, L.L., "Experiments in Automatic Extracting and Indexing", Inf. Stor. Retr., vol. 6, no. 6, (1970), pp. 131 - 334.
- Edmundson, H.P., "New Methods in Automatic Abstraction", J. of ACM, vol. 16, no. 2, (1958), pp. 264 - 285. Encyclopedia of Artificial Intelligence, New York : John Wiley & Sons, 1987.
- Encyclopedia of Library and Information Science, New York : Marcel Dekker Inc., Vol. 1, 1968.
- Fillmore, C.J., "The Case for Case", in Universals Linguistic Theory, (ed. by Bach and Harms), New York : Holt, Rinehart and Winston, 1968. 국역판 : 『격문법이란 무엇인가』, 남선우 외역, 서울 : 을유문화사, 1986.
- Gazdar, G. & Mellish, C., Natural Language Processing in Prolog, Workingham, England : Addison - Wesley Pub. Co., 1989.
- Harris, M.D., Introduction to Natural Language Processing, Reston, Virginia : Reston Pub. Co., 1985.
- Liddy, E.D., The Discourse - Level Structure of Natural Language Texts : An Exploratory Study of Empirical Abstracts, Syracuse, 1988. Syracuse University Ph.D. Dissertation.
- Luhn, H.P., "The Automatic Criterion of Literature Abstracts", in Pioneer of Information Science (ed. by C. K. Schultz), New York : Spartan Books, 1968.
- Morris, A.H., Supporting Environmental Scanning and Organizational Communication with the Processing of Text : The Use of Computer - Generated Abstracts, Lubbock, 1988. Texas Tech.

- University Ph.D. Dissertation.
- Paice, C.D., "Constructing Literature Abstract by Computer : Techniques and Prospects", Information Processing & Management, vol. 26, no. 1, (1990), pp. 171 - 186.
- Partee, B.H., et al., Mathematical Methods in Linguistics, Dordrecht : Kluwer Academic Publishers, 1990.
- Pereira, F.C.N. & Warren, D.H.D., "Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks", Artificial Intelligence, Vol. 13, pp. 231 - 278.
- Radford, Andrew, Transformational Grammar : a First Course, Cambridge, Cambridge University Press, 1988. 국역판 : 『변형문법』, 서정목 외 역, 서울 : 을유문화사, 1990.
- Rich, E., Artificial Intelligence, New York : McGraw - Hill, 1983.
- Rich, M.K & Robinson, P.R., Using Turbo Prolog, 2ed., Berkeley : McGraw - Hill, 1988.
- Rowley, J.E., Abstracting and Indexing, London : Clive Bingley Ltd., 1982.
- Schank, R.C & Selfridge, M., "How to Learn/What to Learn", IJCAI-77, (1977), pp. 8 - 14.
- Schank, R.C. & Abelson, R.P., "Scripts, Plans and Knowledge", IJCAI-75, (1975), pp. 151 - 157.
- Shieber, S.M., An Introduction to Unification-Based Approaches to Grammar, Stanford : CSLI, 1986.
- Tennant, Harry, National Language Processing, New York : PBI, 1981.
- Wall, R., Introduction to the Mathematical Linguistic. 국역판 : 『수리 언어학』, 이영현 역, 서울 : 한신문화사, 1987.

## 부 록

- (1) { < 플라스미드 pKM101과 pSL4를, [연체]  
[DNA 회복기능이, 다른], Escherichia coli B/r  
[연목] [상태형] [실주]  
균주에, 도입하였고, < 두 플라스미드들은,  
[과정형] [연체]  
도입된, 균주들에서 안정하였다. > }  
[결과형] [실주] [상태형]
- (2) { < recA 유전자는, [SOS repair 기작  
[주속] [연목]  
에서, (repressor LexA 단백질을 절단하는),  
[부매] [과정형]  
protease의 기능을, 가지며], [(DNA 회복계  
[부매] [속연] [상태형] [연목]  
에, 중요한), 역할을, 한다고], 알려져 있는데  
[상태형] [속연] [과정형] [결과형]  
> \* (Roberts 등, 1970) \* } [이 유전자의 기능  
[속성]

을, 상실한], 균주 \* (NG 30)\* 는, MNNG에,  
 【상태형】 【실주】 【실매】  
 매우, 민감하였으며), < [recA - 균주에, 도입  
 【상태형】 【실주】 【과정형】  
 된], pKM101과 pSL4는, UV와 MNNG에, 대  
 【연체】 【실매】  
 하여, wild type(H/r30R) 수준으로, 보호효과  
 【실주】 【연목】  
 를, 나타내었고 >, < UV와는, 다르게, MNNG  
 【결과형】 【실매】 【상태형】 【실매】  
 에, 대하여, pSL4는, pKM101보다, 높은, 보호  
 【연체】 【연체】  
 효과를, 나타내었다. > >  
 【연목】 【결과형】

(3) { < [Escherichia coli에서, excision re-  
 【실주】 【연목】  
 pair에, 관계하는], uvrA와 uvrB 유전자는,  
 【상태형】 【주속】  
 uvrC와 더불어, [(UV에, 손상받은), DNA의  
 【실매】 【상태형】 【연목】  
 절단에, 필요한], endonuclease를, 활성화시킴  
 【상태형】 【실매】 【과정형】  
 으로서, [error-freepair에, 중요한], 역할을,  
 【연목】 【상태형】 【속연】  
 하는데 >, \* (Seeberg, 1978)\* < [이들의 기능  
 【과정형】 【속연】  
 이, 결핍된], uvrA-(HS30R)와 uvrB-(HS30R)  
 【상태형】 【실주】  
 균주들은, UV에, 매우, 민감하였으며 >,  
 【실매】 【상태형】

< 도입된, pKM101과 pSL4는, UV에,  
 【과정형】 【연체】 【실매】  
 보호효과를, 나타내었다. > } \* (Fig. 16, c) \*  
 【연목】 【결과형】

(4) { < [R46과 pKM101이, uvrA -와  
 【연체】 【주속】  
 uvrB-의 E. coli K12 균주에서, UV에 대하여,  
 【실주】 【실매】  
 저항성을, 보인, 것처럼 > } \* (Whleh 등,  
 【연목】 【결과형】 【불명】  
 1979) \* < pKM101과 pSL4는, [Fig.1b와 c  
 【연체】  
 에서, 알 수 있듯이], UV에, 보호효과를, 나타  
 【결과형】 【실매】 【연목】  
 내었고 > < 특히, uvrA -균주에서, pKM101는,  
 【결과형】 【실주】 【연체】  
 UV에 대하여, 낮은, 저항성을, 보였지만 >, <  
 【실매】 【연목】 【결과형】  
 pSL4는, 높은, 보호효과를, 나타내었다. > }  
 【연체】 【연목】 【결과형】  
 \* (Fig.1b와 c) \*

(5) { < 돌연변이원 MNNG는, uvrB-균주  
 【실매】 【실주】  
 에 대하여, 치사효과를, 나타낸 반면에 >,  
 【연목】 【결과형】  
 < uvrA - \* (HS30R) \* 균주는, Wild type \*  
 【실주】 【실주】  
 (H/r30R) \* 과 비슷하게, MNNG에, 높은,  
 【실매】



저항성을, 보였다.\* (Fig. 2a b) \* > }

【연목】 【결과형】

(6) { < uvrB - 균주에서는, MNNG에 대하

【실주】 【실매】

여, pKM101과 pSL4는, 저항성을, 보였고 > ,

【연체】 【연목】 【결과형】

< pSL4는, pKM101보다, 높은, 보호효과를,

【연체】 【연체】 【연목】

나타내었다.\* (Fig. 2a) \* > }

【결과형】

(7) { Fig. 3, 4는, < [UV 조사시, 유발되는]

【실매】 【과정형】

돌연변이율에, 미치는 > , 플라스미드 pKM101

【연목】 【상태형】 【연체】

과 pSL4의 영향을, 나타낸 것이다. }

【속연】 【결과형】

(8) { < [Photoactivating 효소가, 결핍된],

【주속】 【상태형】

phr - (H/r30)에, 도입된 > , 두 플라스미드는,

【주속】 【과정형】 【연체】

UV에 대하여, 돌연변이율을, 증가시켰다. }

【실매】 【연목】 【결과형】

\* (Fig. 3) \*

(9) { < recA - (NG30)균주는, UV에 대한,

【실주】 【실매】

돌연변이능이, 완전히, 폐지되었고 > , < 이 결과

【연목】 【결과형】

는, Kondo 등의 보고와, 일치하였다. > }

(10) { 그리고, 플라스미드 pKM101과

【연체】

pSL4는, recA - 균주에서, UV의 돌연변이

【실주】 【실매】 【연목】

유발능을, 회복시키지 못하였다 } \* (Fig. 3) \*

【결과형】

(11) { < R46과 pKM101는, recA - 유전자

【실주】 【주속】

의존성이기 때문에 > , \* (Venturini 1978 :

【상태형】

Walker, 1978b) \* < pSL4도, recA 유전자에,

【실주】 【주속】

의존함을, 나타내었다. > }

【상태형】 【결과형】

(12) { < HS30R \* (uvrA -) \* 균주와

【실주】

HS30R \* (uvrB -) \* 균주는, UV에 대하여,

【실매】

돌연변이율이, 급격히, 증가하였고 > , < [두

【연목】 【결과형】

균주에 도입된] 플라스미드 pKM101과 pSL4

【실주】 【과정형】 【연체】

는, 돌연변이 유발능을, 억제시켰으며 > ,

【연목】 【결과형】

< pSL4는, pKM101보다, 돌연변이 유발억제

【연체】 【연체】 【연목】

효과가, 낮았다 > } \* (Fig. 4) \*

【결과형】

•

•

(15) { < [recA - 균주에, 도입된], pKM101  
 【실주】 【과정형】  
 과 pSL4는, 약간의 증가를 보였지만 >, < 큰,  
 【연체】 【결과】 【결과형】  
 영향을, 미치지 못하였고 >, \* (Fig. 5a) \*  
 【속연】 【상태형】  
 < [recA - 균주에, 도입된], 두, 플라스미드는,  
 【실주】 【과정형】 【연체】  
 돌연변이 유발능에, 별다른, 영향을, 못미쳤지  
 【연목】 【속연】 【상태형】  
 만 > \* (Fig. 5b) \*, < [recB - 균주에, 도입된],  
 【실주】 【과정형】  
 두, 플라스미드는, 돌연변이율을, 증가시켰고  
 【연체】 【연목】 【결과형】  
 >, < 플라스미드 pSL4는, pKM101보다, 높은  
 【연체】 【연체】  
 효과를, 나타내었다. > \* (Fig. 5c) \* }  
 【연목】 【결과형】

(16) { < [DNA 회복능력이 다른], Esche-  
 【연목】 【상태형】  
 richia coli B/r 균주들에서, [(돌연변이원 UV  
 【실주】 【실매】  
 와 MNNG에 대한), 치사효과와 돌연변이 유  
 【연목】

발에 대한], 실험결과는, Kondo 등 \* (1970)  
 【속연】  
 \*의 보고와 일치하였으며 >, < 도입된,  
 【속연】 【상태형】 【과정형】  
 pKM101v과 pSL4의 영향은, DNA 회복 기능  
 【연원】 【속연】 【연목】  
 에 따라, 상호관련성이, 달랐고 >, < pSL4는,  
 【연목】 【상태형】 【연체】  
 pKM101보다, 전반적으로, [기능이, 강화되었  
 【연체】 【속연】 【결과형】  
 음을] 알 수 있었다. > }  
 【결과형】

(17) { < 이와같은, 차이점은, [두, 플라스미  
 【속연】 【연체】  
 드에, 존재하는], mutator유전자의 조절부위  
 【상태형】 【주속】 【부위】  
 또는 구조부위의 변화로, 보여지기 때문에 >, <  
 【부위】 【결론】 【결론형】  
 이들, [mutator유전자에 대한], 분자 유전학적  
 【주속】 【제시】  
 연구가, 계속되어야 >, < 그, 상호관련성이, 밝  
 【상태형】 【연목】  
 혀지리라 생각된다. > }  
 【결론형】



반각 전각	ï í, ì î ï í ì î
반각 전각	ø œ
반각 전각	cm <sup>2</sup> m <sup>2</sup> km <sup>2</sup> cm <sup>3</sup> m <sup>3</sup>
반각 전각	ñ ñ <sup>n</sup>
반각 전각	ö ó ò ô ö ó ò ô o o ø 3
반각 전각	ü ú ü ú û û
반각 전각	ÿ ŷ
반각 전각	A B Γ Δ E Z H Θ I K Λ Μ Ν Ξ Ο Π Ρ Σ T Υ Φ Χ Ψ Ω
반각 전각	ε ι μ ν π ω ρ σ υ ζ α β γ δ ζ η θ κ λ ε φ χ ψ
0	반각 ' , x 전각 × ⑩
1	반각 ' , i 전각 I ① ½ ¼
2	반각 ' , ii 전각 II ② ¾
3	반각 ' , 전각 III ③ ¾
4	반각 ' , 전각 IV IV ④ ¾
5	반각 ' , v 전각 V ⑤
6	반각 ' , 전각 VI VI ⑥
7	반각 ' , 전각 VII VII ⑦
8	반각 ' , 전각 VIII VIII ⑧
9	반각 ' , 전각 IX ⑨

Alt + = 문서처리목록호출	Alt + Z 화면전체를 한줄위로 이동
Alt + F1 문서유평 Key	Alt + S 모아쓰기모드
Alt + F2 Page 지정	Alt + Q 풀어쓰기모드
Alt + F3 중심잡기	Ctrl + A →
Alt + F4 Tap 지정	Ctrl + D →
Alt + F5 특수기호 입력	Ctrl + S →
Alt + F6 도표, 양식 Drawing	Ctrl + O →
Alt + F7 배각	Ctrl + Z →
Alt + F8 밑줄	Ctrl + C →
Alt + F9 한자	Ctrl + E →
Alt + F10 영문 반/전각	Ctrl + F →
Alt + 1 범위설정(삭제, 이동 복사할때)	Ctrl + V →
Alt + 3 범위내 문서 삭제	Ctrl + X →
Alt + 4 범위내 문서를 현 Cursor	Ctrl + W →
Alt + 5 범위내 문서를 현 Cursor위치에 복사	작업취소 ESC
Alt + 7 대치	지움 Del
Alt + 8 문서폭 조정	삽입·정정 Ins
Alt + 9 어구 등록	지정취소 Space bar
Alt + 0 표계산	