문장 검색을 위한 색인시스템 구축
: 초·중등 학생의 한국어 및 영어 문장을 중심으로
A Construction of Indexing System for Sentence Retrieval

이 태 영(Tae-Young Lee)*

초 록

한국어 및 영어의 글쓰기를 도와주는 문장 및 문단 제공시스템을 구축하기 위하여 색인작성과 탐색시에 필요한 색인언어를 연구하였다. 색인언어로 명사어와 술어 및 부사어를 선정하였고 여러가지 보조 색인기호들도 추가하였다. 접근점으로 주제명과 키워드를 사용하였고 키워드 검색은 1절, 2절, 3절, 문맥첨가 탐색을 포함하였다. 검색의 만족도는 긍정적이었으며 데이터베이스의 양과 질을 충실히 보완한다면 문장이나 문단을 제공하여 주는 시스템은 효과적일 수 있다.

ABSTRACT

An indexing language were studied to construct the sentences and paragraphs providing system aided to write a Korean or English composition. The indexing language includes the index terms like noun, predicate, and adverb, and also various index symbols. The subject name and the keyword included the symbols, which indicate the connectives between clauses in a sentence, is used as the access point. The search results show this system will be effective with large database and developed retrieval methods.

키워드: 문장, 문단, 검색, 색인시스템, 문장검색, 한국어, 영어, Sentence, Paragraph, Retrieval, Indexing Language, Sentence Retrieval, Korean, English.

* 전북대학교 인문대학 문헌정보학과 교수(taehyun@moak.chonbuk.ac.kr)
■ 논문 접수일 : 2003. 2. 17
■ 게재 확정일 : 2003. 3. 10

1 서 론

1.1 연구목적

글짓기를 힘들어하는 학생들에게 구체적인 한국어 문장을 예시하여, 작문을 도와주는 참고도구는 매우 유용하게 사용될 것이다. 그리고 인터넷 이용이 생활화된 현 사회에서는 이러한 '글짓기 도우미'는 한국어 작문의 차원을 넘어 영어 작문의 영역까지 확대될 수 있다. 이를테면 국내의 이용자가 인터넷(특히 웹서비스)을 통하여 전 세계의 수많은 이용자(개인, 기관, 기업)들에게 메시지를 보낼 때, 영어로 작성된 데이터의 전송이 예측되는데, 영어 작문에 익숙하지 못하면 많은 고생을 하거나 통신을 포기하는 현상이 일어날 수 있다.

이와 같이 영어나 한국어 작문에 자신이 없을 때에 본인이 쓰려고 하는 내용과 유사한, 잘 작성된 기존의 문장들이 글짓기에 길잡이 역할을 할 수 있다. 다시 말하면 여러 장르(수필, 추천문, 상업문, 개인적 서신, 알림말 등)의 한국어와 영어 글들의 전체 또는 일부 문장들이 작문을 하는 데에 범례적 모형으로 작용할 수 있다는 것이다. 현재 한영번역시스템이 완벽하지 않은 상황에서는 이 중개시스템은 매우 적절하게 쓰일 수 있으며 또 미래에 우수한 한영번역시스템이 개발될지라도 하나의 한국어문장에 대해 여러 개의 유사한 영어문장의 수수 가능성이 있으므로 효율적으로 사용될 수 있다.

특히 자기 표현 기술을 연마해야 하는 십대의 학생들에게 열람식 문장제공시스템은 좋은 글쓰기 도구가 될 수 있다. 개인이 보고, 듣고, 체험한 내용 중에 꼭 적어놓고 싶은 부분들을 어려움 없이 글로 재현케 하고 또 필요한 영어문장을 즉시 참조케 하는 문장제공시스템은 바로 글쓰기 전문가를 대체하는 서비스가 될 수 있다. 또한 인간전문가와는 달리 이 시스템은 PC통신이나 웹을 통하여 항상 이용자 곁에 상주한다.

이러한 전문가적 시스템을 구축하려면 문헌검색에서 사용되는 색인언어의 속성인 망라성, 특정성, 정확성, 일관성(Lancaster, Warner 1993, 81; Moens 2000, 72)은 물론이려니와 그것에 덧붙여 문장내용을 대변할 수 있는 또 다른 특성을 갖는 문장검색용 색인언어들이 구상되어야 한다. 본고는 이러한 점들을 고려하며 한국어와 영어문장을 검색하는데 필요한 색인언어에 대한 연구를 하여 그 색인언어의 효과성을 검증하는데 목적이 있다.

1.2 연구 방법 및 제한점

실험용 데이터베이스에 포함되는 문장과 문단들을 발췌하기 위하여 초등학교 5, 6학년 국어, 사회 자연 교과서와 초등학교에서 가르치는 교실용 영어, 그리고 한국어와 영어 문장이 동시에 실려있는 동화 2편을 택하였다. 중등 한국어 및 영어자료는 중학생을 위한 국어와 영어 참

고서에서 발췌하였다.

한국어문장은 문학, 사회, 자연의 주제로 나누어 각각 100 문장씩으로 한정하였고 문단은 25개를 취택하였다. 영어문장은 교실영어, 동화, 중등영어도 각각 100 문장씩을 선정하였고 문단은 문학, 사회, 자연, 동화에서 총 50개를 포함시켰다.

문장과 문단의 색인을 위하여는 문헌검색에서 사용되는 주제명과 키워드색인 방식을 차용하였으며 색인언어(탐색언어)는 모두 한글형으로 통일하였고 문장의 색인대상단위는 단일문과 절을 택하였다. 단일문과 절의 색인언어는 명사어, 술어, 부사어를 중심으로 술어어미, 술어구(관용구, 선어말어미, 보조용언) 및 접속어구가 포함되었다. 그리고 불용어를 제외한 문장내의 전 단어를 색인어로 올렸다. 탐색시의 표목도 색인작성 방식과 동일한데 다만 질문식을 만드는 사람이, 보고자 하는 절이나 문장 내용의 단어를 중심으로 작성케 하였다.

검색 결과를 조사하기 위하여 초·중학교 선생 두 명과 초·중학교 학생 두 명에게 설문 평가를 하였다. 측정은 5항목의 만족도를 표시하는 리커드방식으로 행하였고 실험시스템 구현은 xml과 asp로 하였다

본고에서 사용한 추리기호를 아래의 예식1을 들어 설명하면 다음과 같다. '〈=〉'는 전후에 있는 내용의 비교, '/'는 좌측에 있는 Ti(색인어)를 포함하고 있는 Ci(절)들을 표현, '⊂'는 Ci가 Qi에 종속을 의미, '〈'는 매칭되어 나온 결과, '--〉'는 앞(위)의 결과에 이어서 또 추리, '-〉'는 결과 출력이다.

예식1: {T1, T2, T3, T4}/(Ci⊂Qi) 〈=〉 {T7, T10, T4, T3}/(Cj⊂Sj⊂DB)

--〉 (Sj 〈 Ci⊂Qi) 〈=〉 (St 〈 Ck⊂Qk) -〉 S1

2 한국어와 영어의 문장 구성요소

한국어와 영어의 문장을 알려면 먼저 그 문법적인 특성을 고찰하여야 한다. 한국어 문장을 구성하는 단어들의 품사는 대체로 "명사, 대명사, 수사, 동사, 형용사, 관형사, 부사, 감탄사, 조사"와 같이 9품사로 분류하였다(임헌도 1987, 36-40). 여기에 하나 더한다면 지정사('이다'류)를 설정하기도 하였다(이길록 1981, 320). 영어의 경우는 사전과 학교문법을 보면 "명사, 대명사, 동사, 형용사, 관사, 부사, 전치사, 접속사, 감탄사"의 9종을 기준으로 삼을 수 있다(The Random House Dictionary of the English Language 1987, 1415).

한국어와 영어의 품사가 서로 다른 것은 종류에서 한국어의 "조사, 수사", 영어의 "전치사, 접속사"이고 쓰임새에서는 '형용사'이다. 이것들을 살펴보면 다음과 같다. 한국어의 '수사'는 영어의 명사, 형용사 등에 그대로 포함되어 있고, 한국어

의 '조사'와 영어의 '전치사'는 그 쓰임새가 비슷하다. 영어의 '접속사'는 한국어에서 용언이 어미변화를 일으켜 접속어구 구실을 하거나 명사 또는 조사에 속한 단어가 그 역할을 대신한다(영어의 관계대명사, 관계부사도 접속역할을 한다). 한국어 '형용사'와 영어 '형용사'는 어의는 같으나 용법상에는 현격한 차이가 있다. 한국어의 '형용사'는 동사와 같이 어미변화를 일으키므로 한국어에서는 '동사'와 '형용사'를 묶어 용언으로 통칭한다.

색인 시 구체적으로 고려해야 할 한국어와 영어의 문장 구성요소들을 분석하면 첫째, 한국어의 술어어미와 영어의 접속어 관계, 둘째, 한국어의 선어말어미 및 보조용언과 영어의 조동사 용법 관계, 셋째, 영어의 부정사와 동명사에 대한 한국어 용언구의 대응, 넷째, 한국어와 영어의 관용적 표현 및 형용사의 처리 문제가 있다(이태영 1998, 220-221). 다음의 (1)부터 (7)까지의 사항들은 위의 문제들을 처리한 결과이다. 고딕체 단어는 변수이고, 샘물체는 상수이며, '-'는 어간이나 단어를 나타냈다.

1) 접속(술어어미와 접속어구, 관계대명사와 관계부사 포함)
   (1) 대등접속 : ①-고;and, ②-(으)나;but, ③-나;or, ④보다;more-than, 등
   (2) 종속접속 : ①-면;if, ②전에;before, ③때문에;because,④-는 한;as long as,등
   (3) 관계대명사 : ①-ㄴ,-는(사람,인명);who,②-ㄴ,-는(사물);which, ③-는 것은(을);what, ③-(이)든(지) 간에; whatever, 등
   (4) 관계부사 : ①-ㄴ,-는(장소); where, ②-ㄴ,-는(방법);how, 등

2) 첨어1(선어말어미와 조동사 사용, &로 첨언함)
   (1) 존대 : -ㅂ니-,-습니-,-시-, -하여 주시-, -하여 주십니-,하고 있습니-;Would - please, 등
   (2) 과거 : -ㅆ-,-았-,-였-,-었-,-ㅆ,-왔, -■ㅓㅆ,-됐; -ed.
   (3) 현재 : -ㄴ-,-는-; -.
   (4) 미래 : -〈야〉겠-, -ㄹ것이다, -려고 한다; will, -be going to, try to, 등
   (5) 과완(과거완료) : -ㅆ었다,-하였었다, -했었다,-되었었다; had~-ed.
   (6) 과진(과거진행) : -고 있었다

3) 첨어2(보조용언과 조동사 사용, &로 첨언함)
   (1) 미래 : -ㄹ것이다,-려고 한다; will, -be going to.
   (2) 현진(현재진행) : -고 있다,-는 중이다,-가다; be -ing.
   (3) 현완(현재완료) : -〈해〉 버렸다, -적이 있다, -고 있다; have -ed.
   (4) 부정1 : -지〈도〉 않(못)하-, -지도 ; do not〈also, too〉-, 등
   (5) 가능(불가) : -ㄹ 줄 알다(모르

다), -ㄹ 수〈가〉 있다(없다); can, may〈not〉 -, 등

(6) 추측 : -는 듯 하다, -는 것 같다, -ㄹ 것이다; seem to -, would seem -, 등

(7) 허가 : -〈하여, 해〉도 좋다, -어도 된다; may -, 등

(8) 당위 : -야〈만〉 한다, -지 않으면 안된다; must -, have to-, 등

(9) 습관 : -곤 하였다, -기도 하였다;used to, 등

(10) 희망 : -고 싶다, -기를 원한다; like to -, 등

(11) 권유 : -는 것이 좋다, -ㄹ 필요가 있다; had better -, need to -, 등

(12) 가정 : -면 -ㄹ 것이다, -ㄹ 지도 모른다, -쓰을지도 모른다, -텐데; if-were~would-, would have-, 등

(13) 양태 : -게 되다, -게 하다, -게 시키다, -게 지다(곱게 지다), -해 지다(깨끗해 지다); let -, 등

4) 부정(부정사와 동명사 및 용언구, &로 첨언함, '형'은 형용사, '동'은 동사임)

(1) 부정2 : -기 위하여, -ㄴ 것, 등; -to(eat to live,is to make)

(2) 동명 : -는 것을, 등;-ing (like smoking)

(3) 형동 : ("동+동,형+형,동+형,형+동"일 때) -게~-다, -며~-다, -야만~-겠-

(4) 상태1 : -게 되다(하다, 시키다, 지다), -해 지다, -면 된다, -기도한다, -ㄴ다고 한다, -하기만 하다; let -, 등

(5) 상태2 : -단다, -잖아, -는 구나, 자구요, -는거(야,지 등), -는 건(가 등)

(6) 중첩2 : -고~-다(기다리고 기다리다), -다고~ 생각하다(이에 준하는 용언)

5) 관용(관용구, &로 첨언함)

(1) 것당 : -는 것이 당연하다; must, might, 등

(2) 것뿐 : -ㄴ 다는 것 뿐이-

(3) 록할 : -도록 할 수 있-; what if he could(somehow) get - to.

(4) 밖없 : -할 수밖에 없다

(5) 뿐아 : -ㄹ 뿐만 아니라

(6) 않것 : -지를 않는 것이-

(7) 않없 : -하지 않을 수 없다 ; cannot~but, cannot help -ing.

(8) 적뿐 : -ㄴ 적이 없(었)을 뿐만 아니라

(9) 지없 : 〈아무리〉 강조하여(해)도 지나침이 없다;cannot - too, cannot be to.

(10) 최것 : 최대한 발휘하는 것이다

(11) 할것1(될것1) : -게 할(될) 것(거)이-

(12) 할것2(될것2) : -야 할(될) 것(거)이-

(13) 척뿐 : -ㄴ(한) 척하는 것 뿐이-

(14) 하안 : -하여(해)서는 (하면, 하는 것은) 안된다(곤란하다)

(15) 할없 : -할 것 없-

6) 형용(형용사와 형용사 상당어구)

한국어에서 형용사는 '파랗다'에서 보는 바와 같이 동사처럼 어미가 있으며 어미 변화를 한다. 한편 "이다, 있다, 없다"는 어떤 명사의 뒤에 붙어 용언구실을 한다. (예; 산있다, 산이다, 맛있다, 맛이다, 맛없다, 등)

(1) 용조 : '이다'는 조사이나 문장이나 절 말미 단어에 왔을 때 용언상당어구로 취급함

① 파랗다; is blue,②산이다; is mountin, 등

7) 조사와 전치사

①의; of, ②으로; to, ③으로부터; from, ④에; on, 등

3 문장 색인언어의 형식

3.1 문장 색인의 구성소

문장이 나타내고 있는 의미(개념)을 색인한다는 것은 문헌의 주제색인과는 매우 다르다. 예를 들어 '가축사료'에 대하여 쓴 글은 '가축사료'라는 주제명으로 그 글이 '가축사료'에 대해서 섰다는 것을 쉽게 알 수 있지만, "우리는 가축의 사료를 영국에서 구입하였다"는 문장은 '가축사료'라는 복합어 하나만으로 이 문장을 다 대변한다고 할 수는 없는 것이다. 그러므로 전체적인 주제 개념보다는 개개 문장의 표현의미를 파악하여 "가축사료, 영국, 구입하다(문장은 술어가 있어야 완성된다)"와 같이 정립하게 되는데, 문장을 구성하는 단어 간과 절 간의 관계를 또 명확히 보여주어야 한다.

문헌색인의 색인어 간의 관계설정은 ① 문헌에서 뽑아낸 개념들의 정렬이나, ② '표제' 또는 '초록' 등에서 실린 그대로 나온 절 형태의 개념들을 분석하는 것이므로 "서양의 문헌에 나타난 한국의 위인들(한국의 문헌에 나타난 서양의 위인들)"이나 "물고기가 내는 독물(물고기에 해로운 독물)"과 같이 비교적 관계를 정립시키기 쉽고, 명사 중심적인 색인이 된다. (정영미 1993, 78) 그러나 한 문장에 속한 단어간의 관계는 기본적으로 '문장기본형식'에서 보는 바와 같이 'S(주어)', 'O(목적어), 'C(보어)', 'V(술어)와 같으며 또 9종류의 품사가 있어 문장 안에서 언어적 규칙에 따라 결합된다. 전통적으로 '어휘'는 '단어'와 문법요소(접사 또는 어미)로 구분하고, 단어는 명사, 동사, 형용사, 관형사, 부사, 조사 등 여러 가지 품사로 분류하였다. 이러한 분류법은 그 나름대로 필요하고, 또 여러 가지 장점이 있지만, 문장분석의 실용적 목적을 위하여, 보다 간소하게, 다음의 〈표 1〉과 같이 분류할 수 있을 듯하다. (정신문화연구원 1985, 69-70.)

이에 근거하여 한국어문장에서 각 품사별 특성과 영어문장에서 쓰일 용도를 (1)-(5)항으로 나누어 정리하였다.

〈표 1〉 단어의 분류

기본 분류	기능에 의한 분류	
근간어	명사어,	서술어
수식어	명사수식어(관형사어)	서술어수식어(부사어)
어 미	명사 어미	서술사 어미

(1) 명사와 술어

근간어는 문장의 근간을 이루는 말, 즉 문장이 표현하는 '의미'의 가장 중추적인 역할을 담당한다(정신문화연구원 1985, 69-70). 이런 관계로 보통 문헌들이 색인될 때에 명사(기본 5형식에서의 S,C,O)만이 색인어로 채택되지만 개념서술이 중요한 문장색인에서는 서술을 최종적으로 결정짓는 V도 색인어로 선택하지 않으면 않된다.

(2) 부사

언어의 제 성질에 따라 넓은 의미에서는 감탄사를 제외한 단어들이 개념이며, 협의로는 내용어 만이 개념이다. 이 내용어에는 체언과 용언이 포함되었다(김봉주 1988, 53). 대체로 용언의 의미를 수식한다고 하겠지만(김민수 1981, 163; Bolinger, Sears 1981, 84-5) 부사는 용언을 꾸미는 역할을 하여 개념을 형성하는 데에서는 비켜나 있다. 그러나 "아련히, 쓸쓸히, 외로이" 등과 같은 부사어는 글짓는 이가 충분히 찾아볼 수 있는 것이다.

(3) 조사

그러면 쓰여진 어휘의 격을 강하게 나타내는 조사는 어떠한가? 개념이 발화되어 분화될 때 절마다 명사와 동사가 구성되면 명사에 주어지는 조사는 저절로 결정된다. 예를 들면 아래의 〈표 2〉와 같다.

〈표 2〉 문장생성 예

용어 집합	문장 생성
(사람, 글, 쓰다)	사람이 글을 쓰다
(사람, 글, 쓴다)	사람이 글을 쓴다, 사람은 글을 쓴다.
(돼지, 글, 쓰다)	돼지를 글로 쓰다.

여기서 "돼지가 글을 쓰다"가 성립한다고 보는 것은 상식적으로 올바른 쓰임이 아니며 가정적 바람에 근거한 예술적 표현이다. 즉 언어적으로 보았을 때 진실이 아니라 허위를 얘기하는 것이 된다. 따라서 색인할 때 글의 의미를 손상치 않는 조사는 생략하도록 하며 영어에서의 전치사는 선별적으로 계상하도록 한다.

(4) 술어어미

술어어미는 복합문에서 절들 간이 어떤 의미나 경우로 이어져 있는가를 말할 때에 대두되는 요소이다. 복합문은 삽입확장, 연결확장 등이 있다. 삽입확장은 여덟

가지로 이루어지는데 술어어미가 포함된 것은 "...가느냐가, ...통일됨을, ...짓음에, ...돋는, ...돋을듯, ...돋기에, ...흐르는 것이" 등이 존재하였다. 연결확장은 "①중복적; ...고(며, 는 동시에, 거니와, 면(서), 자, 고서, 다(가) 등), ②선택적; ...거나(든지, 는가 등), ③대립적; ...으나(는데 불구하고, 지만, 거늘, 되 등), ④가정적; ...면(거든, 어도, 라도 등), ⑤원인적; ...으니까(기 때문에, 니, 어, 아, 므로, 기에 등), ⑥전제적; ...아야(어야, (는)다고 해야, 건만, ㄹ망정 등)"과 같이 6가지가 있었다. (김민수 1989, 241-254)

(5) 술어구

한국어나 영어에서는 두 개 이상의 술어가 합쳐져서 술어구 형태로 사용되는 경우가 흔히 발생한다. 그러므로 이런 술어구에 대처할 간단하면서 명료한 색인언어가 요구된다. 2장에서 살펴본 한국어와 영어의 문장 구성요소는 한국어에서 영어로 변환시키려 할 때나 또는 그 반대를 시도할 때 빈번히 부딪히는 사항들로서 바로 간단명료한 색인언어에 대한 해답을 제시하여 준다. 즉, 2장의 1)에서 6)항까지 거론된 것 등은 복잡한 술어구를 간단하게 색인할 수 있게 하며 또한 한국어로 영어문장을 찾을 때 이것들이 지표로 사용된다.

위에서 고찰한 결과 문장색인의 특성으로 인해 문장 색인언어에는 ①명사(주어, 목적어, 보어, 등), ②용언(어간만 사용), ③부사, ④접속(술어어미와 접속어구), ⑤첨어1(선어말어미와 조동사), ⑥첨어2(보조용언과 조동사), ⑦부정(부정사 및 동명사와 용언구), ⑧관용(관용구), ⑨형용(형용사 및 형용사 상당어구) 등이 포함될 수 있겠다.

3.2 주제명과 키워드 형식

3.2.1 주제명 형식

문장 색인에서의 주제명은 한 두개의 단어로 이루어져 관련 문장 전부를 검색해 내는 역할을 수행한다. 문장들에서 그 문장이 표현하는 바의 주제를 결정하는 일은 어떤 문장에서는 쉽게 파악이 될 때도 있고 또 어떤 경우에는 매우 난해하여 파악을 할 수 없을 때도 있다. 문장에 주제명을 부여할 때, 주제명이 쉽게 표출되지 않으면 문단이나 글 전체의 소주제 또는 주제에서 주제명을 도출한다. 글의 장르, 대상주제(여행, 바다여행, 산여행, 인물기행 등), 형식(사전, 편람 등) 등이 여기에 적용되었다. 본고의 실험을 위해 채택된 주제명은 아래의 〈주제명 색인식〉과 같이 보통 한 두 개의 단어로 이루어졌다. 영어 문장도 한국어 문장과 똑 같은 방법으로 한국어 색인어로 운영하였다.

〈주제명 색인식〉

① 주제어 : 한 품사로 이루어짐
② 주제어-명사어 : 주제어(글 또는 문단의 주제)+명사어(절 또는 단일문의

명사)

③ 주제어-술어 : 주제어(글 또는 문단의 주제)+술어(절 또는 단일문의 술어)

(*실제의 데이터베이스에서는 '명사어'와 '술어'는 '부주제명'이란 명칭으로, 주제어로 이끌어지는 '주제어파일' 내부에서 세분되어 진다. 먼저 주제어로 접근하여 파일을 찾은 연후에 다시 '명사어'와 '술어'로 구분하여 문장을 찾는다.)

3.2.2 키워드 형식

문장 색인에서 키워드는 이용자가 원하는 적합한 문장을 후조합 방식으로 검색하여 제공하는 역할을 수행한다. 먼저 문장을 색인하는 방법은 문장을 절단위로 나누어 절별 색인을 하고 절별 색인에 문장번호를 부여한다. 예를 들면 문장번호 100인 "가을이 오니 단풍이 든다"라는 문장은 "{가을, 오다 | 100}, {단풍, 들다 | 100}" 와 같이 색인한다. 이렇게 절별로 색인하는 이유는 문장 전체로 색인하면 키워드 숫자나 너무 많아지기 때문이며, 이용자가 문장의 개념을 추상할 때는 "가을, 단풍"과 같이 전체적이지만 실제로 작성단계에서는 절별로 "가을, 오다", "단풍, 들다"와 같이 절별로 접근한다고 판단되기 때문이었다.

둘째, 탐색을 할 때 탐색문도 절별로 작성하여 접근한다. 그러므로 한 절로만 탐색을 하면 엄밀한 의미에서 후조합 방식의 탐색이 이루어지지 않고 용어열색인과 같이 전조합 방식의 탐색이 이루어진다.

셋째, 복합문의 경우에는 절과 절이 이어지는 연결 형식을 고려한 문맥적 색인 기호도 부가하였다.

키워드는 명사, 술어, 부사로 한정하였으며 절과 절 사이의 관계표시를 위해서 한국어에서는 〈표 3〉의 술어어미를, 영어에서는 2장의 1)을 사용하였다. 그리고 한국어와 영어의 술어어미들의 변화 특성을 고려한 2장의 2)-6)도 색인언어에 포함시켰다. 의미어를 명사, 부사, 술어로 한정한 이유는 3.1에서 서술하였다.

한국어문장의 키워드형식은 ①키워드들의 단순나열, ②품사적 제한을 가한 키워드들의 나열(명사, 술어, 부사만으로), ③술어어미를 고려한 키워드들의 나열로 구성하였다. ③은 보다

정밀한 복합문장의 검색을 시도할 경우 사용된다. 복합문의 탐색문은 부절의 술어어미(영어문장은 접속사, 관계대명사, 관계부사)의 문맥(〈표 3〉 참조)을 이용하여 만들어진다. 본고의 실험적 검색을 위해 전체적인 〈키워드 색인식〉 모형을 다음과 같이 정리하였다.

〈 키워드 색인식〉

① 명사어들 -- 일반적 색인 방식

② 명사어들(3개 이내) + 술어

③ 명사어들(2개 이내) + 술어 + 부사어

④ 명사어들(3개 이내) + 술어 + 문맥

⑤ 명사어들(2개 이내) + 술어 + 부사어 + 문맥

〈표 3〉 문맥(연결어미)의 종류

종 '류	연결 어미	종 류	연결 어미
+가정	-면, -거든, -(ㅏ)아도, -(ㅓ)어도, -라도 / if	+과상	들고 있었다(과진)/들려 있었다(과상)
+과정1	-여, -여서, -해, -해서, 〈동사의 경우〉: -ㅓ(서), -ㅏ(서) -어(서) -아(서)	+과정2	-ㄴ(는)데, -다가, -(ㄴ)는 중에 / while
+과정3	-게	+나열1	-고 / and
+나열2	'고'로 대체될 수 있는 것: -며, -면서, -고서	+대립	-지만, -거늘, -되, -(으)나, -는데 불구하고
+목적	-려고, -ㄹ, -러, -도록, -라, -도록, -려면	+반반	-지~지, -던(든)지~-던(든)지
+불문	-(이)든(지) 간에 / whatever, whoever, whichever, 등	+비교	-보다, -ㄴ(는) 것 보다 / more~than
+삽입	-ㅁ(이,을), -음(이,을), -기(가,는,를,로), -느냐가, -는 것(이,은 등), -(을)듯, -기 만을, -(하,라)고(는), -라는	+상태	-ㄴ 채(로, 로는 등), -는 데(에, 에 있어 등), -ㄴ(는) 거(것이)-, -ㄴ 결과, -ㅣ(이,리)
+선택	-거나, -든지, -는가 / or	+수식	-ㄴ, -는 / -ing, -ed, who, that, which
+시간	-기 전, -ㄴ(는) 후, -ㄹ 때(까지)/ before, after, until	+염려	-까봐
+원인1	-(으)니까, -기 때문에, -므로, -기에 / because	+원인2	-(다)니, -자, 〈형용사의 경우〉: -어(서) (는), -아(서)(는), -ㅓ(서)(는), -ㅏ(서)(는),
+의문	-면서도, -고서도,	+전제	-아야, -어야, -(는)다고, -해야(만), -건만, -ㄹ 망정, -는 한 / as~as, so~as
+종결	-다, -까, -오, -세, -게나	+현상	들고 있다(현진)/들려 있다(현상)

(문맥의 종류는 김민수의 연구를 중심으로 삼고 여기에다 영어와 관련된 항목을 부가하였다.(3.1의 (4) 참조)

영어문장의 키워드형식도 기본적으로 한국어와 같은 방법으로 찾는다. 그런데 영어에서 자주 부닥치는 문제는 술어와 전치사, 관사들의 용법이다. 예를 들면 "학교에 다닌다"와 "학교에 간다"는 "I go to school", "I go to the school"로, " 관한 연구"는 "study on the library, study on learning"와 같이 두가지로, "서울에 간적 있다"와 "서울에 산적 있다"는 "I have ever been to(in, at) Seoul"와 "have been living in Seoul"로 되며, "서울에 살고 있다"는 "I live in Seoul"과 "I am living in Seoul"(일시적인 삶)로, 그리고 "장사를 하고 있다"는 "I have been in business"와 같이 표현된다. 한편, "월급으로살다"는 "live on my salary", "서울에 살다"는 "live in Seoul", "붓으로 살다"는 "live by my pen"으로 작성된다.

위의 예에서 보았듯이 술어(조동사 포함)와 전치사 및 관사들이 여러 가지 쓰임새로 사용되고 있는 것을 볼 수 있다. 따라서 지금까지 언급한 색인언어에 전치사(영어 관용구와 문단의 검색에서 적용)도 포함시켰다.

4 색인과 탐색 식의 실제

4.1 색인어 작성규칙

색인언어는 자연언어 상태에서 발생하는 재현율의 저조를 방어하기 위하여 명사는 이음동의어, 용언은 파생어들을 통제하였다. 예를 들면 명사인 '달'은 "월, 개월"을 대신하여 색인어로 쓰이고 술어인 '걷다'는 "걸어가다, 걸어오다, 거닐다"를, '생각하다'는 "생각되다, 생각시키다, 생각나다 등"을 대표하여 색인어로 상정된다. 즉, 파생 용언들의 기본형을 선택어로 삼았다.

한국어나 영어 문장들을 색인작성하거나 탐색할 때에 수행되는 색인어 추출을 위하여 공통으로 적용한 사항들은 다음과 같다. 색인하여야 할 명사어가 "N의 N, NN, NNN" 등과 같이 복합적인 의미를 담은 단어일 때에는 하기의 규칙1을 따른다. (여기서 ↑는 색인어로 채택되는 주요단어를 뜻한다)

↑ ↑
규칙1 : ① IF 'N1의 N2' THEN 'N1N2'
↑ ↑
② IF 'N1N2...' THEN "N1, N2, .., .., ..., N1N2..."

그런데 색인 및 탐색 식 작성의 실제에서 중심 주제어라고 판단되는 단어와 함께 쓰이는 명사어와 술어가 두 개 이상 출현할 수가 있다. 첨가되어지는 단어가 2개 이상 존재할 때에는 하기의 '규칙2, 3'을 따른다(여기서 ↑는 강조된 단어를 뜻한다).

↑ ↑
규칙2 : ① IF 'N1 N2...' THEN 'N1N2' OR "N1, N2,..."
↑
규칙3 : ① IF 'N1에 N2를' THEN 'N1'
↑
② IF 'V1고 V2다' THEN 'V2'

그리고 주제명이 확실하게 파악되지 않는 문장은 아래의 규칙4와 같이 처리한다.

규칙4 : ① IF 'N1 N2, N3, V' THEN 'S' (여기서 S는 글이나 문단의 주제임)

그리고 키워드 형식의 색인작성을 할 경우에 복합문장은 규칙5, 6과 같이 처리한다.

↑ ↑
규칙5 : ① IF "N1 V1고, N2 V2" THEN "N1, V1 (전절)"
↑ ↑ ↑ ↑
규칙6 : ① IF "N1 V1고, N2 V2" THEN "N1, V1 (전절), N2, V2(후절)"

4.2 한국어문장의 실제

3.2.1과 3.2.2에서 서술한 바를 토대로

여기에서 실험하고자 하는 한국어의 색인 및 탐색 식의 종류는 주제명 형식의 "① 주제어, ②주제어+명사, ③주제어+술어", 와 키워드 형식의 "①명사어들, ②명사어들+술어, ③명사어들+술어+부사어, ④명사어들+술어+문맥, ⑤명사어들+술어+부사어+문맥," 등 기본적으로 8개이며 실제는 아래와 같다.

1) 주제명 형식 표현

주제명을 표현하는 방법은 이재철 주제명표목표에 의하면 "위생, 위생-목적, 위생/치아, 보건·위생", "브라질 이민의 역사" 등과 같은 형식이 있다. 아래에 작성된 〈색인 예1〉의 (1), (2), (3)은 규칙1,2에 의거하여 〈예문1〉을 재료로 해서 색인을 작성한 경우이다.

〈색인 예1〉

(1) 주제어 : ①탱자나무, ②거울
(2) 주제어+명사어 : ①탱자나무-가을, ②거울-나
(3) 주제어+술어 : ①탱자나무-자랑하다, ②거울-들다

〈예문1〉

① 아, 이제는 환한 가을 햇살 아래 노오란 금빛 단추를 달고 자랑스럽기만 한 탱자나무야.
② 내 호주머니 조그만 거울 속엔 내가 들어 있고 장난감도 있다.

2) 키워드 형식 표현 : 기존의 데이터베이스에서 키워드 검색시스템은 키워드들만 표현하는 방식과 키워드와 문법적 요소를 같이 합쳐서 나타내는 방법이 병존해 왔다. 본고에서도 그와 유사한 방법을 택하였다. 실험하려고 하는 키워드 표현방식은 "①명사어들, ②명사어들+술어, ③명사어들+술어+부사어, ④명사어들+술어+문맥, ⑤명사어들+술어+부사어+문맥이다. 다음의 〈색인 예2〉는 아래의 〈예문2〉를 소재로 하여 색인을 작성한 예이다.

〈색인 예2〉

(1) 명사어들 : ①{과일, 망신, 모과}, ②{시골, 흙냄새}
(2) 명사어들+술어 : ①{과일, 망신, 모과, 시키다}, ②{시골, 흙냄새, 나다}
(3) 명사어들+부사어+술어 : ①{과일, 망신, 모과, 시키다}, ②{시골, 흙냄새, 물씬, 나다}
(4) 명사어들+부사어+술어 : ③{아랫니, 뾰족하게, 올라오다}, ④{눈, 사뿐사뿐, 내리다}
(5) 명사어들+부사어+술어 : ⑤{새소리, 싣다}, ⑥{이야기, 나누다, &습관}
(6) 명사어들+부사어+술어 : ⑥{바닷가, 나오다}, ⑦{초원, 위, 번지다}
(7) 명사어들+부사어+술어+문맥 : ①{과일, 망신, 모과, 시키다, +중복}, ⑦{초원, 위, 번지다, +수식}
(8) 명사어들+부사어+술어+문맥 : ④{눈, 사뿐사뿐, 내리다, +종결}, ⑥{별

빛, 내리다, +수식, 저녁이다, +가정}
(9) 명사어들+부사어+술어+문맥 : ⑤{새소리, 싣다, +중복}, ⑥{이야기, 나누다, &습관, +종결}
(10) 명사어들+부사어+술어+문맥 : ⑥{바닷가, 나오다, +원인}, ⑦{진홍빛, 보랏빛, 타오르다, +수식}

〈예문 2〉

① 과일 망신은 모과가 시키고, 어물전 망신은 꼴뚜기가 시킨다.
② 흙냄새가 물씬 나는 시골에서 귀여운 동물과 함께 살면 얼마나 좋을까?
③ 아랫니 두 개가 새싹처럼 뾰족하게 올라와 있었다.
④ 날이 어두워지면서 흰 눈이 사뿐사뿐 내렸습니다.
⑤ 상쾌한 아침 공기가 지저귀는 새 소리를 싣고 방안으로 밀려 들어왔습니다.
⑥ 별빛이 내리는 저녁이면 고니는 바닷가로 나와 훈이와 많은 이야기를 나누곤 했습니다.
⑦ 저녁이면 함께 앉아 초원 위로 번지는 붉고 진한 노을을, 진홍빛과 보랏빛으로 타오르는 구름들을 바라보았답니다.

4.3 영어문장의 실제

영어문장을 검색할 때의 색인언어는 한국어문장을 검색할 때와 마찬가지로 한국어 색인어를 사용하였고 색인방식도 한국어문장 색인방식과 동일하였다. 다만 경우에 따라서 2장에서 기술하였던 조사(조사와 전치사)에서 전치사 역할을 하는 조사를 문맥 색인언어로 추가시켰다. 다음의 〈'색인 예3〉은 〈예문3〉으로 작성된 문맥을 포함한 색인의 예이다.

〈색인 예3〉

(1) 명사어들+술어 :
① {우물, 동, 서, 미터이다},
④ {친구, 선택, 주의, 지나치다, &당위}
(2) 명사어들+술어+문맥 :
① {우물, 동, 서, 미터이다, +중복},
② {당신, 말하다, +삽입},
③ {회의, 메모, 하다, &권유},
④ {청년, 일, 능력, &최것}, {그, 부여하다, &현완, +수식}

〈예문3〉

① 우물은 동에서 서로 22미터이고, 남에서 북으로 12미터이다; The well is 22m from east to west, 12m from north to south
② 당신이 말하고자 하는 것을 안다; I see what you mean
③ 회의에서는 메모를 하는 것이 좋다; It has better for you to take notes at the meeting.
④ 친구 선택은 아무리 주의해도 지나침이 없다; We cannot be too careful

in the choice of our friends

⑤ 젊은이가 세상에서 아마 할 수 있는 가장 큰 일은 그에게 부여된 능력을 최대한으로 발휘하는 것이다; The great thing a youth can possibly do in this world is to make the most possible out of the ability that has been given him

4.4 문단의 실제

한국어나 영어 문단의 색인어는 주제명 형식이다. 주제명은 일반적인 웹사이트의 카테고리와 같이 계층적으로 전개되기 때문에 문단의 내용을 잘 표현한다고 생각되는 짧막한 구로 자유롭게 표현하였다. 그 예를 들어보면 다음과 같다. '봄'과 같이 단일어로 나타내거나 '탱자나무'와 같은 복합어 또는 '부엌-저녁', '슬픔-느낄 때', '영화의 주제곡' '고양이-고르다' '문화의 차이와 공통점', '지난 학습내용 확인', '호랑이-미끌어지다', '해-누이와 달-오빠'와 같이 자유롭게 표현하였다.

영어문단과는 별도로 영어의 어떤 특정 술어구와 전치사구를 포함하는 여러 문장예들의 탐색을 원하는 경우를 대비하여 "used to, on the wall"과 같이 영어 색인어구로 이끌어지는 문장들의 집합을 '영어관용구'로 묶어 영어문단의 한 종류로 포함시켰다.

5 데이터베이스 운영

5.1 html과 xml화일

html과 xml화일로 작성된 내용들은 첫째, 유사한 문장들의 집합(〈그림 1〉 참조), 둘째, 문단의 전체(〈그림 2〉 참조), 셋째, 영어 술어구나 전치사구(영어관용구)에 따른 문장들의 집합(그림 3 참조)이다. 첫째와 둘째의 경우에는 주제명으로 인도되는 계층적 범주 단계로 해당 문장이나 문단을 링크를 통해 찾아갈 수 있게 설계되었다. 셋째는 데이터베이스로 작성되어 있는 술어구와 전치사구의 목록을 살펴보고 역시 링크로 참조케 하였다.

5.2 asp화일

키워드 형식으로 운영되는 asp화일은 두 개의 화일로 구성되었다. 하나는 색인화일 역할을 하는 도치화일(Manolopoulos et al. 2000, 56)로서 검색 기준이 되는 색인 표목어는 통제된 개개의 단어들이며 표목어에 부여되는 리스트 항목들은 문장 내지 문단의 번호들이다. 문장과 문단화일은 문장번호와 실제의 문장으로 이루어졌다.

5.3 키워드 탐색 유형

키워드로 운영되는 한국어 문장 탐색순서는 다음과 같다.

```
<?xml version="1.0" encoding="euc-kr"?>
<?xml-stylesheet type="text/xsl" href="class001-01.xsl"?>
<document>
<산>
산 길은 계곡을 끼고 이리구불 저리구불 산을 기어오르고 있었다.
산 아래로 눈을 돌리면 푸른 골짜기가 한 폭의 그림처럼 눈안에 들어옵니다.
산골짜기는 온통 검푸른 빛으로 물들어 있었다.
산꼭대기를 바라보면서 한가롭게 앉아 담배를 피우고 있었습니다.
산나물로 끓인 국이라 향긋한 내음이 무척 좋았습니다.
산바람은 밤이 깊어갈수록 더욱 더 세차게 전나무 가지를 흔들었습니다.
산바람은 밤이 깊어갈수록 더욱 더 세차게 전나무 가지를 흔들었습니다.
산에는 단풍이 빨갛게 물들어 사람들을 손짓하고 있습니다
산촌의 풍치가 아름답고 그윽한 정서가 넘치는 마을에 이르렀습니다.
서산마루를 향해 막 지려는 햇살이 붉은 빛을 띠면서 쏟아져 들어오고 있다.
</산>
<document>
```

<그림 1> 연관 문장의 집합 예

```
<?xml version="1.0" encoding="euc-kr"?>
<?xml-stylesheet type="text/xsl" href="class001-01.xsl"?>
<째document>
<겨울>
한 해가 저물어 가는 섣달 그믐날. 잿빛 하늘이 무겁게 내려앉고 목화송이 같은 하얀 눈이 내리기 시작하고 있었다. 궁궐 뜰 연못가의 잣나무 가지에서 혼자 앉아 있는 외쪽눈 방울새는 그대로 하얀 꽃송이었다 외쪽눈 방울새의 모습은 더욱 아름다워 보이기 시작했다. 외쪽눈 방울새는 지난 몇 달 동안이 그대로 몸서리쳐지는 한바탕의 꿈이었다고 생각을 해 본다.
</겨울>
</document>
```

<그림 2> 문단의 예

"①질문내역 형성 -> ②절별 정리 -> ③절의 명사, 술어발췌 -> ④질문입력 -> ⑤통제언어로 변환 -> ⑥색인화일 탐색 -> ⑦부합된 문장(문단) 번호 추출 -> ⑧문장(문단) 파일에서 개별 레코드들 출력" 순으로 전개한다(영어문장 검색도 이와 같다). 실제로 검색하는 요령을 예들 들어 살펴본다. 질문내역이 "우측의 그림을 보자"라고 형성되었고 질문내역 안의 개념 간 우선순위가 첫째, '보자', 둘째,

화일명 : used to
〈html〉
〈body〉
He used to get up early　　He is used to getting up early
He is used to living in Seoul　　I am not used to this weather
He'll get used to speaking English　　I am used to go to London
It used to be believed that the sun moved round the earth. /옛날에는 태양이 지구를 돈다고 믿고 있었다.
He came earlier than he used (to). /그는 여느 때보다 일찍 왔다.
There used to be a storehouse. /원래는 여기에 창고가 있었다.
He didn't used to play golf, did he? /그는 전에는 골프를 치지 않았지. _
She is very different than she used to be. /그녀는 그전과는 상당히 다르다.
〈/body〉
〈/html〉

〈그림 3〉 영어 관용구에 따른 문장 집합

'우측의', 셋째, '그림을'이라고 판명되었다면 그 탐색식은 1차는 (보자) 2차는 (우측, 보자) 3차는 (우측, 그림, 보자)로 작성하여 입력시킨다.

탐색의 종류는 다음과 같이 3가지로 운영하였다. 복합문의 경우에 있어서 질문은 절별로 하여 공통으로 나타나는 문장을 검색하였다.

① 보통탐색 : {T1, T2, T3, T4}/(Ci⊂Qi) 〈=〉 {T3, T4, T7, T10}/(Cj⊂Sj⊂DB) --〉 (Sj 〈 Ci⊂Qi) -〉 S1
(한 절 및 단일문 위주로 문장화일을 검색하여 매칭되는 문장들을 모두 출력한다.)

② 정밀탐색 : {T1, T2, T3, T4}/(Ci1⊂Qi) 〈=〉 {T3, T4, T7, T10}/(Cj1⊂Sj⊂DB) {T5, T6, T7, T8}/(Ci2⊂Qi) 〈=〉 {T7, T8, T11, T13}/(Cj2⊂Sj⊂DB) --〉 (Sj 〈 Ci1⊂Qi) 〈=〉 (Sj 〈 Ci2 ⊂ Qi) -〉 Sj
(둘 내지 세 절 위주로 문장화일을 검색하여 매칭되는 모든 문장을 출력한다.)

③ 문맥탐색 : 주절에 얽힌 부절들의, 혹은 부절에 얽힌 부절들의 술어어미의 속성들을 가미하여 검색한다. ②에 문맥 색인기호를 하나 또는 둘을 더 추가하는 형식이다.

6 문장 검색 및 평가

6.1 만족도 평가

앞장에서 논의한 주제명(문장, 문단, 영어관용구)과 키워드(한 절, 두 절, 세 절, 문맥 포함)로 이루어지는 여러 종류의 탐색식의 결과를 살펴보기 위하여 설문 대

상자에게 원하는 문장내역을 색인언어로 정리하여 문장 및 문단데이터베이스에 질문토록 하였다. (http://taehyun.chonbuk.ac.kr/의 글짓기도우미2를 클릭하면 실험시스템에 접속된다) 대상자들에게 입력되어 있는 문장과 문단 그리고 영어 관용구들이 소수이고 색인어 작성 상의 오류가 있을 수 있고 문맥 관련 기호들은 실험용이란 것을 주지시켰다. 문장을 검색하는 질문횟수는 10번이었다. 이렇게 10번을 탐색한 후에 다음의 설문을 하였다. 질문에 따라 출력된 문장들이 주는 만족도는 리커드척도 방식(아주 긍정, 긍정, 보통, 부정, 아주 부정의 5항목)으로 측정하였다.

설문 내용 및 응답 평균치

질문 1 : 이러한 시스템이 필요한가? 필요함

질문 2 : 필요하다면 한국어와 영어 중 어느 것이 더 필요한가 ? 영어

질문 3 : 필요하다면 개별문장과 문단 중에 어느 것이 더 효과적인가 ? 문단

질문 4 : 검색한 결과가 어느 정도의 만족을 주었는가 ? 긍정

질문 5 : 검색 방법에서의 차이점은 있었는가 ?

1) 문단 검색에서 '단어와 기호조합' 방식과 '어구 및 어절' 방식의 만족도는 ?
'단어와 기호조합' -> 보통, '어구 및 어절' 방식 -> 긍정
2) 범주별 검색에서 부주제명의 효과는 ? 아주 긍정
3) 키워드 검색에서 '1절 단독', '2절 조합', '3절 조합'의 만족도는 ?
'1절 단독' -> 긍정, '2절 조합' -> 보통, '3절 조합' -> 부정
4) 키워드 검색에서 '문맥삽입 탐색'과 '문맥배제 탐색'의 만족도는 ?
'문맥삽입 탐색' -> 보통 '문맥배제 탐색' -> 보통
5) 키워드 검색에 있어 술어, 부사어의 효과는 ? 긍정
6) 영어 검색 시 한국어 색인어 사용은 ? 아주 긍정

6.2 결과 분석

질문 1, 2, 3, 4에서 확인된 바와 같이 개인이 글을 쓸 때 참고할 수 있는 문장이나 문단을 제공하여 주는 시스템의 존재는 긍정적인 가치를 갖고 있었다. 영어의 경우는 무척이나 필요한 존재였다. 문장들이 상당한 수준까지 축적되어 있다면 영어의 경우는 중요한 역할을 하는 데이터베이스가 될 수 있을 것이다. 문단은 전체적이 기술을 할 때 참조용으로 유용하게 작용할 수 있었고, 문장은 작문하려고 할 때의 막막함을 해소시켜주는 것으로 인정되었다. 그리고 한 문장을 제공하는 것보다는 묶음 문장이나 두세 개의 연결된 문장(문단)이 제공되는 것을 선호하였다.

질문 5의 1)과 2)의 결과는 문장이나 문단에 주제명을 부여할 때, 두 개 이상

의 개념을 복합시키는 것이 효과적인 한편, '-, /' 등의 기호로 주제명과 부주제명을 연결하는 것보다 말로 이어가는 형식이, 더 쉽게 개념화될 수 있다는 것을 보여준다. 정리하면 주제명 하나를 사용하는 것은 문헌검색에서는 유효할지 모르나 문장이나 문단검색에서는 좀더 내용을 구체적으로 묘사하는 것이 필요하다.

키워드 검색에서 '1절 단독' -> '2절 조합' -> '3절 조합' 의 순으로 만족도가 낮아진 원인은 절의 조합이 많아질수록 검색되어 나오는 문장의 숫자가 적어지는 현상에서 비롯된 것으로 추정하였다. 같은 맥락에서 '문맥삽입 탐색'과 '문맥배제 탐색'의 만족도가 비슷한 것은 출현 문장수가 비슷하기 때문이라고 사려되었다. 끝으로 영어문장이나 문단의 탐색에 우리말 색인어 사용을 원하는 현상은 영어 단어 사용에 익숙하지 않았기 때문이라 생각되었다.

설문 결과의 분석을 통해 전체 응답의 만족도는 평균적으로 '보통' 이상이었다.

7 결 론

한국어 및 영어의 글쓰기를 도와주는 문장 및 문단 제공시스템을 구축하기 위하여 색인작성과 탐색시에 필요한 색인언어를 연구하였다. 색인언어로 명사어와 술어 및 부사어를 선정하였고 한국어문법 교과서와 영어의 학교문법 참고서를 고찰하여 (1) 접속(삽입,중복,선택,대립,가정, 원인 등), (2) 첨어1(존대,과거,현재 등), (3) 첨어2(미래,현재진행,현재완료,부정,가능,추측,허가 등), (4) 부정(부정사,동명사), (5) 관용(관용적 표현들), (6) 형용(형용적 표현들)과 같은 보조 색인기호들도 추가하였다. 검색방법을 다양화하는 취지에서 주제명으로 이끌어지는 전조합색인 방식의 범주별 검색과 키워드로 이끌어지는 후조합방식의 키워드 검색으로 나누고 키워드 검색에서는 1절 탐색, 2절 탐색, 3절 탐색, 문맥부가 탐색 등 여러 가지를 두었다. 데이터베이스에 포함되는 데이터의 종류는 문장(한국어와 영어), 문단(한국어와 영어). 관용구(영어), 문장모음(한국어와 영어)으로 크게 나누었다.

본 시스템을 실험 검색한 결과, 설문상에서 이러한 문단 및 문장제공시스템은 대체로 그 필요성에 있어 긍정적이었으며 영어의 경우는 더 높은 긍정을 이끌어 내었다. 따라서 데이터베이스의 양과 질을 충실히 보완한다면 문장이나 문단을 제공하여 주는 시스템은 효과적인 것이 될 수 있다.

참 고 문 헌

김민수. 1981.『국어의미론』. 서울: 일조각.
김민수. 1989.『국어문법론』. 서울: 일조각.
김봉주. 1988.『개념학』. 서울: 한신문화사.
리재철. 1961.『주제명표목표』. 서울: 연세

대학교.

이길록. 1981. 『국어문법연구』. 서울: 일신사.

이태영. 1998. 영어문장 제공시스템에서 한국어문장의 색인방법. 『제5회 한국정보관리학회 학술대회 논문집』, 219-223.

임헌도. 1987. 『문장원론』. 서울 : 학문사.

정신문화연구원. 1985. 『현대 국어문장의 실태분석』. 서울 : 시사문화사.

정영미. 1993. 『정보검색론』. 서울 : 구미무역 출판부.

Bolinger, D. and Sears, D. A. 1981. *Aspect of Language*. New York : Harcourt Brace Jovanovich, Inc.

Lancaster, F.W. 1998. *Indexing and Abstracting in Theory and Practice*. 2nd ed. London : Library Association Publishing.

Lancaster, F. W. and Warner, A. J. 1993. *Information Retrieval Today*. Arlington : Information Resources Press.

Manolopoulos, Y., Thedoridis, Y. and Tsotras, V.J. 2000. *Advanced Database Indexing*. Norwell : Kluwer Academic Publisher.

Moens, Marie-Francine. 2000. *Automatic Indexing and Abstracting of Document Texts*. Boston : Kluwer Academic Publishers.

The Random House Dictionary of the English Language. 1987. New York : Random House.