(제 10회 한글 및 한국어 정보처리 학술대회)

YDK-Term : 한국어 용언의 다국어 통합정보사전

최용준+, 황도삼+, 최기선*

+영남대학교 컴퓨터공학과
경북 경산시 대동 214-1
우: 712-749,
yjchoi@nlp.yeungnam.ac.kr
dshwang@ynucc.yeungnam.ac.kr

*한국과학기술원 전산학과
대전시 유성구 구성동 373-1
우: 305-701,
kschoi@world.kaist.ac.kr

A Thesaurus for Korean Language

Yonjun Choi, Dosam Hwang, Keysun Choi
Department of Computer Engineering Department of Computer Science
Yeungnam University KAIST

요약

통합정보사전은 각종 자연언어처리 시스템에 있어서 고도의 언어처리 및 성능향상을 위한 필수 요소이며, 아무리 좋은 언어처리 도구와 처리 알고리즘이라도 계산언어학에 근거한 양질의 체계적인 전자사전이 없는 한 이의 실용화는 불가능하다. 기존에 출판되어 있는 사전은 자연언어처리 및 이해의 관점에서 개발된 사전이 아니며, 자연언어처리 도구 및 응용시스템에 사용되는 사전은 목적에 따라 각기 다른 체계에 의해 구축되어 있어 이용하는데 있어서 비효율적이다. 따라서, 고도의 언어처리 및 이해를 목적으로 한 체계적이며 과학적인 방법론을 이용하여 형태소, 구문, 의미정보 등 각종 정보가 통합된 통합정보사전의 개발이 반드시 필요하다.

본 논문에서는 다국어 통합정보사전 구축을 위한 한국어 용언의 통합정보사전을 설계한다. 이를 위해 사전구축 방법론을 정립하고, 정립된 방법론을 바탕으로 하여 통합정보사전의 개발을 위한 통합정보사전 개발 시스템을 설계하고 구현하였다.

◎ YDK-Term : Yongjun Dosam Keysun - Terminology Dictionary

◎ 본 연구는 1997.12.4부터 1998.10.3까지 과기부의 STEP2000과제인 "대용량 국어정보 심층처리 및 품질관리기술 개발"(KAIST)의 위탁과제로 수행하였다.

1 서론

전자사전은 출판되어 있는 사전과는 달리 자연언어처리 도구 및 응용시스템에 사용하기 위한 사전이며, 용도에 따라 각기 다른 체계에 의해 구축되어 있어 전자사전간의 호환성이 부족하다. 또한, 전자사전은 대부분 기계가독형 체계로 만들어져 있어 사람이 사용할 수 있는 형태로의 출판이 어렵다. 또한, 전자사전에 어떤 단어에 대해 기술되어 있지 않은 보다 더 상세한 정보를 얻기 위해서는 또다른 여러 사전들을 참조해야 한다는 불편함이 있을 뿐만 아니라, 이러한 사전들을 입수하는 것도 쉽지 않다. 따라서, 고도의 언어처리 및 이해를 목적으로 체계적이며 과학적인 새로운 방법론을 이용하여 형태소, 구문, 의미정보 등 각종 정보가 통합된 통합정보사전의 개발이 반드시 필요하다. 또한, 최근에는 전문용어에 대한 전문 사전의 구축 필요성이 제기되고 있으며, 일부 구축되고 있지만, 체언류에 대해서만 연구가 진행중이다. 따라서 특정분야에만 사용되는 전문용어 용언에 대한 연구가 필요하며, 아울러 국가간의 교류가 활발해지고 있으므로 다국어 정보를 포함하는 전문용어 용언의 다국어 통합정보사전의 개발이 필요하다.

본 논문에서는 다국어 통합정보사전 구축을 위한 한국어 용언의 통합정보사전을

설계하고 구축한다. 이를 위해 사전구축 방법론을 정립하고, 정립된 방법론을 바탕으로 하여 통합정보사전의 개발을 위한 통합정보사전개발 시스템을 설계하고 구현한다. 개발하는 사전은 특정분야에 맞춰 전문화된 사전이 아니라 기존 사전의 정보를 통합하고 이미 개발되어 있는 여러 자연언어처리 도구를 이용항여 얻을 수 있는 각종 정보를 통합한 사전으로 국어정보처리 응용시스템 개발에 도움을 줄 수 있다. 또한, 통합정보사전 개발 시스템은 각종 전자사전 및 자연언어처리 시스템들의 정보를 손쉽게 통합할 수 있게 함으로써 고품질의 사전 개발을 가능하게 한다.

2 통합정보사전

현재까지 통합정보사전은 주로 미국, 유럽, 일본에서 많이 개발되어 있으며, 본 논문에서는 우리와 언어체계가 유사한 일본의 IPAL, EDR, 분류어휘표, 유어신사전, 일본어어휘대계와 부산대학교에서 구축한 용언의 하위범주화 사전[1]에 대하여 간략히 살펴본다.

2.1 IPAL

IPAL은 일본 정보처리 진흥사업 협회(IPA) 기술센터의 IPAL 그룹이 개발한 일본어의 전자화 사전(IPA Lexicon)으로, 다른 일반적인 일본어 사전에 비해 상세한 언어정보가 기술된 것이 특징이다. 현재 IPAL은 일본어 어휘 체계상 및 사용 빈도상 중요한 기본적인 동사(861 단어), 형용사(136 단어), 명사(1,081 단어)에 대해 의미 및 통사적 이름 특징에 근거하여 구분하고, 구분한 것을 하나의 단위로 형태, 의미, 통어, 관용 표현 등의 정보를 상세하게 기재하고 있다. 이 사전은 정보의 종류가 다양하고 상세하기 때문에 통합정보사전으로서의 가치가 매우 높지만 단어의 수가 적어 실용적이지 못하다는 단점이 있다.

2.2 EDR 전자사전

EDR 전자사전은 일상 문장에 사용된 모든 어휘를 커버할 정도의 대규모 사전이며, 특별한 응용 시스템이나 알고리즘을 위해 설계된 것이 아니고, 포괄적인 응용 목적을 가지고 개발되었다. EDR 전자사전의 정보량은 매우 방대하며 다양한 정보를 담고 있는 장점이 있지만 정보의 종류가 다양하지 못하다.

2.3 분류어휘표

분류어휘표는 일본 국립국어연구소에서 작성한 것으로 약 3만 현대어를 의미에 따라 분류 배열하였으며, 사용빈도가 높은 약 7천어를 기준으로 의미를 분류했다. 이 사전은 의미에 대해서만 다루었기 때문에 의미 정보는 가장 우수하지만 그 외의 정보가 전혀 없어 통합정보사전으로는 부족하다.

2.4 유의어사전

유의어사전은 관념을 언어화하기 위한 하나의 개념과 그 의미 내용을 대 표제어에 게재하고 이에 관련된 단어를 모아서 배열한 사전이다. 이 사전 역시 분류어휘표와 마찬가지로 의미에 대해서만 다루었기 때문에 의미 정보는 가장 우수하지만 그 외의 정보가 전혀 없어 통합정보사전으로는 부족하다는 단점이 있다.

2.5 일본어 어휘대계

일본어 어휘대계는 일본 NTT에서 개발한 사전으로, 의미체계, 단어체계, 구문체계 의미사전의 세 부분으로 구성되어 있다. 이 사전의 정보량은 매우 방대하며 다양한 정보를 담고 있는 장점이 있지만 EDR 전자사전과 마찬가지로 정보의 종류가 다양하지 못하다는 단점이 있다.

2.6 용언의 하위범주화 사전

부산대학교와 시스템공학연구소(현 전자통신연구소)의 국어공학센터가 컴퓨터에 의한 한국어 분석의 기본자료로 제공하는 것을 목적으로 1996년에 한국어 단어의 의미적 연어관계(하위범주와 정보)를 위하여 구축한 전자사전으로 사용빈도순에 따라 상위 3,500여개의 용언을 하위 범주화하였다. 이 사전은 그 정보의 종류가 다양하지 못하며, 사전개발과정이 수작업으로 진행되어 일관성이 부족한 것이 단점이다.

본 연구에서는 수작업으로 사전을 개발할 때의 문제점인 고비용/저효율을 문제를 해결하고 사전정보구축의 객관성을 높이기 위해 반자동으로 사전을 구축할 수 있는 사전 개발도구를 개발하며, 개발한 사전개발도구를 활용하여 다국어 정보를 포함한 다양한 정보를 포함하는 통합정보사전을 개발한다.

3. YDK의 설계 및 구현

2장까지 기존의 통합사전을 살펴보았다. 본장

에서는 사전구축과 사전개발 시스템의 설계과 구현에 관해 기술한다.

3.1 사전구축방법론

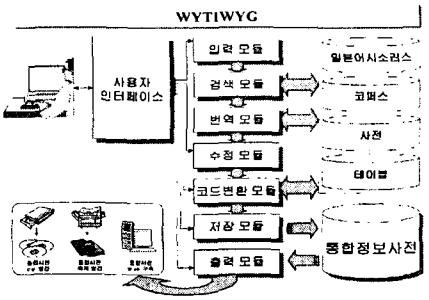
통합정보사전을 개발하기 위한 사전방법으로 수작업 방식과 반자동 방식이 있다. 수작업은 사람의 손으로 직접 모든 사전자료를 수집하고 입력하여 구축하는 방법으로 많은 노력과 시간 및 비용이 들어간다. 또한 사전정보를 입력하는 사람의 능력에 의해 사전의 품질이 결정되므로 객관성이 떨어진다는 단점이 있다. 이에 비해 반자동 방식은 사전구축 시스템을 이용하여 구축하는 방법으로 많은 양의 데이터를 빠른 시간에 입력할 수 있을 뿐 아니라 실질적이고 사전원시정보를 가지고 있는 코퍼스나 전자사전과 같은 정보원으로 부터 사전구축에 필요한 정보들을 자동적으로 추출할 수 있으므로 고품질의 사전을 개발할 수 있다. 본 연구에서는 반자동 방식으로 통합정보사전을 개발하기 위해, 통합정보사전 개발 시스템인 YDK(Yongjun Dosam Keysun) 시스템를 개발하였다[그림 1]. YDK는 WYTIWIG(What You Think Is What You Get)이라는 개발 개념을 가지고 있으며, 각종 사전정보의 참조 및 자연언어처리 도구들과의 통신을 통한 정보추출기능을 가지고 있어 효과적으로 사전을 개발할 수 있는 도구이다.

3.2 YDK의 설계

YDK는 다른 사전을 참조하기 위해 사전을 검색할 수 있어야 하며, 참조하는 사전에 일본어 사전이 있으므로 이를 번역시스템과 연계하여 번역결과를 얻어낼 수 있어야 한다. 또한 자연언어처리 도구들로부터 다양한 정보를 얻기 위해서는 하나의 인터페이스로 통합하기 위한 방법이 필요하다. YDK는 생각할 수 있는 모든 자연언어 자원들을 이용할 수 있도록 지원하기 위하여 전자사전과 자연언어처리 도구들을 하나의 인터페이스 내에 통합하는 방법이다. YDK는 분산되어 있는 자연언어처리 자원들이 각각의 자원에 적합한 모듈을 개발하고, 각 모듈들의 인터페이스를 표준화 시켜 통합하는 체계를 따른다. 이를 [그림 1]에 보인다.

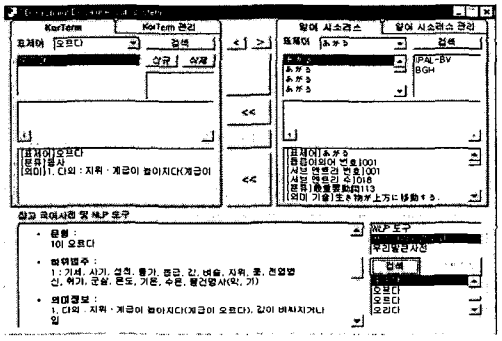
3.3 YDK의 구현

YDK는 Intel Pentium PC, MS-Windows95환경에서 MS-Visual C++ 5.0과 MS-Visual BASIC 5.0을 이용하여 개발하였으며,



[그림 1] YDK 시스템

서버에 설치되는 YDK지원 시스템은 SUN ULTRASparc, Solaris 2.5환경에서 gcc 2.7.2.3과 gdbm 1.7.3을 이용하여 개발하였다. 구현한 YDK Browser를 [그림 2]에 보인다.



[그림 2] YDK의 실행예

4. YDK-Term : 한국어 용언의 다국어 통합정보사전

YDK-Term(YongjunDosamKeysun-Terminology Dictionary)은 현재 동사와 형용사에 대해 연구가 진행되어 있으나, 본 논문에서는 동사에 대해서만 언급한다. 동사는 <표 1>과 같이 6개의 항목과 25개의 부항목으로 이루어져 있다.

4.1.1 표제어 정보

YDK-Term 동사사전은 한국어의 동사(용언)를 대상으로 하였으며, KAIST 코퍼스 1996년판에서 사용 빈도순으로 추출한 300단어를 대상으로 하였다.

(1) 표제어 : 표제어를 정할 때에는 동사의 원형으로 하였으며, 오름차순으로 배열하였다.

예) 가다, 가다듬다

(2) 동음이의번호 : 예를 들면 「쓰다」에 있어서 「쓰다(冠),「쓰다(書)」「쓰다(用)」.....은 하나의 표제어로 하지 않고 별도의 표제어로서 수록하고 있다.

예) 001, 002, ...

<표 1> YDK-Term 동사 항목

항목	부항목			
1. 표제어정보	<1> 표제어			
	<2> 동음이의번호			
	<3> 부엔트리 번호			
	<4> 대역정보	일어		
		영어		
		중국어		
		독일어		
		불어		
		:		
	<5> 출처			
	<6> 분류			
	<7> 음운정보			
2. 의미정보	<8> 의미 기술	다의		
		부사 (BC)		
	<9> 관련어	상위어		
		유의어		
		반의어		
	<10> 시소러스			
	<11> 의미분류			
3. 형태정보	<12> 전성형			
	<13> 어간			
	<14> 자타			
	<15> 활용			
	<16> 파생			
	<17> 높임			
4. 통어정보	<18> 문형			
	<19> 문례			
	<20> 술어소			
	<21> 의미소	N1		
		N2		
		N3		
		:		
5. 문법정보	<22> 사동	장 (BC)		
		단 (BC)		
		대역	영어	
			일어	
	<23> 피동	장 (BC)		
		단 (BC)		
		대역	영어	
			일어	
6. 기타정보	<24> 관용표현			
	<25> 비고			

(3) 부엔트리 : 표제어를 의미 및 통사적 특징에 기초하여 하위구분한 것을 부엔트리라고 하였다. 여기서 의미에 기초한 하위구분은 개개의 동사가 가지고 있는 여러가지 어휘적 의미에 의해서 그 동사를 분류하는 것을 가리킨다. 따라서, 어휘적 의미가 다르면 하나의 표제어에 대해 여러가지의 부엔트리가 존재한다.

예) 001, 002 ...

(4) 대역정보 : 다국어통합정보사전에 대비한 필드로 한-일간, 한-영간, 한-중간 대역 표제어를 기재한다.

(예) 표제어 영 중 일
쓰다 write

(5) 출처 : 각 항복 정보 개재에 이용된 사전 자료의 출처를 기재한다.

(예) 「국어대사전」 민중서림, IPAL-BV, KAIST 코퍼스(97) .....

(6) 분류 : 표제어의 품사를 기재한다.

(예) 동사/형용사.

(7) 음운 정보 : 표제어의 한국어 발음에 대한 정보로 소리나는 그대로 표기했다.

<예> 먹다 먹따

4.1.2 의미정보

(1) 의미기술

1) 다의 : 이 정보는 컴퓨터가 사용하기 위한 것이 아니라 사람이 보기 위한 정보로서, 부엔트리의 의미를 알기 쉽게 설명한 것이다. 표제어 중의 부엔트리 각각의 의미를 구별하기 위하여 기술한 것이며, 동사의 의미는 같지만 각 구조에 대한 의미가 달라서 부엔트리가 나누어져 있는 경우와 구별하기 위해서도 사용된다.

예) 가다

[다의] 움직이다(시계가 가다), 전달되다(기별이 가다), 변하다(맛이 가다), 생기다(주름이 가다), 꺼지다(전기가 가다), 쏠리다(마음이 가다, 관심이 가다, 호감이 가다), 죽다(선생님께서 졸지에 가시다)

2) 부사 : 사용할 수 있는 부사의 예를 든다. 부사를 알 경우 동사의 의미를 보다 명확하게 할 수 있기 때문이다.

(예) 가다 [부사] 일찍, 자주 ...

(2) 관련어 : 부엔트리와 의미적 관련을 가진 단어로서 상위어와 유의어 및 반의어에 대한 정보를 갖는다. 관련어를 생각하기 어려운 경우에는 무리하게 기재하지 않았다.

1) 상위어 : 두 개의 단어가 상대적으로 포함의 관계에 있을 경우에 그 두 단어는 상하 관계를 구성하는데, 보다 일반적이고 넓은 의미를 가진 단어를 상위어라고 부른다.

예) 가다 [상위어]움직이다

2) 유의어 : 표제어와 같거나 아주 닮은 개념을 나타내는 단어를 유의어라고 부른다. 따라서, 동의어도 유의어에 포함된다

예) 가다 [유위어]움직이다.

3) 반의어 : 부엔트리로 나타내는 개념이 많은 부분을 공유하고 있으면서, 일부 대립하는 관계에 있는 단어를 반의어라고 부른다.

예) 가다 [반의어]오다

(3) 시소러스 : 표제어를 참고할 수 있는 다른 시소러스의 이름을 기재했다. 부산대 시소러스, 연세대 시소러스, IPAL-BV 등 참고할 수 있는 시소러스명을 기재했다.

예) IPAL-BV

(4) 동사의 의미적 분류

동사를 의미에 따라서 다음과 같이 분류하였으며, 각 동사가 어느 그룹에 속하는가를 <표 2>에 나타낸다.

<표 2> YDK-Term의 의미분류체계

대분류	세분류
상태	존재.소유
	관계 인정
	단순상태
동작	추상적관계
	시간
	상태변화
	설치
	이탈
	접촉
	생리.심리
	지각.사고
	발견
	경제활동
	사회활동
	언어활동
	자연현상

4.1.3 형태정보

(1) 전성형

전성은 품사가 바뀌는 것을 말한다. 여러 가지의 전성정보가 있을 경우에는 각 부엔트리가 다른 품사에서 바뀌어 동사가 된 경우를 먼저 기재하고, 그 다음에 부엔트리의 단어가 다른 품사로 바뀔 수 있는 경우를 기재했다.

예) 쓰다 [전성] 쓰기

(2) 어간 : 용언에서 활용한 때에 변화하지 않는 부분을 「어간」이라고 부른다. 여기에서는 그 어간을 기재한다.

예) 오르다 [어간] 오ㄹ

(3) 자타 : 해당하는 부엔트리가 「자동사」인가 「타동사」인가를 기재하고 있다. 자동사인 경우에는 "자", 타동사인 경우 "타" 라고 쓴다.

예) 가다 [자타] 자

(4) 활용형 : 불규칙활용이 있는 경우에만 그 활용형을 기재한다.

예) 가다 [활용형] 거라 불규칙

(5) 파생형 : 합성어 가운데 구성요소의 어느 하나가 접사인 합성어를 파생어라 하고 그러한 조어를 파생어이라 한다.

예) (형용사) 빨갛다 [파생형] 새빨갛다

(6) 높임형

높임형으로 변환할 경우 어간 자체가 변화되는 경우에만 기재한다.

예) 먹다 [높임형]잡수시다.

4.1.4 통사정보

(1) 문형 : 문형은 다음과 같이 표기한다.

1. 명사구는 N1, N2, ... 로 나타낸다.

2. ( )로 묶여져 있는 것은 임의적(optional)인 명사구이다.

3. 명사구의 순서는 가장 표준적으로 보여지는 어순에 따른다.

4. 동사는 기재하지 않는 것을 원칙으로 한다.

예) 가다 [문형] N1이

(2) 문예 : 문형의 형식으로 실제 사용된 예를 기재한다.

예) 가다 [문예] 철수가 가다

(3) 의미소 : 예를 들면, 「살다」라고 하는 동사는 존재를 나타내는 것이지만, 그 대상이 되는 명사구는 인간과 「유정물」에 한정되어 있고, 생명이 없는 물체(책, 상자등)은 「살다」라고는 말하지 않고, 「있다」를 사용해야만 한다. 이와 같은, 명사구에 대한 제한을 「의미소」이라고 부른다. 본 사전에서 채택한 의미소를 <표 3>에 나타낸다. N1, N2등의 하나의 항목에 대응하는 의미소는 네 개를 한도로 해서, 약어로 기재하고 있다. 어떤 명사구가 하나의 의미소에 대응한다는 것은 아니다. 다음과 같은 통사적 환경에 있어서 동일의 명사가 복수 의미 속성을 가질 수 있다.

(4) 술어소

술어소는 격형식을 하나의 수식처럼 정의하여 컴퓨터가 단어를 처리할 때 보다 정확한 결과를 얻기 위한 부분이다. 이 부분은 지면 관계상 생략한다.

4.1.5 문법정보

문법정보는 용언의 하위범주화사전의 정보를 참조하며, 2개의 부항목으로 구성된다.

(1) 사동 : 예) 오르게 하다, 올리다

(2) 피동 : 예) 올려지다

4.1.6 기타정보

(1) 관용표현 : 일상 자주 사용되는 것을 기재했다. 구체적으로는 의미가 가장 가깝다고 여겨지는 단어 뜻의 「관용 표현」난에 기재했다. 해당「관용 표현」에 포함되는 형용사의 의미가 어느 단어 뜻에 가장 가까운가, 그 판정이 어려운 경우도, 그때마다 검토해서 기입했다. 따라서, 하나의 관용 표현이 두가지의 단어 뜻에 걸쳐서 등록되는 것은 없다.

예) 속담

<표 3> 의미소

약호	전체이름	속성명	예
CON	concrete	구체명사	
ANI	animal	동물명	개. 말. 새. 원숭이. 코알라
HUM	human	인간	누이. 선생. 남성. 학생
ORG	organization	조직,기관	국가. 기업. 경찰. 연구소
PLA	plant	식물	꽃. 벚꽃 , 소나무. 장미
PAR	parts	생물의 부분	머리. 다리. 팔. 허리. 뿌리. 날개
NAT	natural	자연물	산. 하늘. 돌. 강. 언덕
PRO	products	생산물,도구	종이. 차. 빵. 천. 칼
PHE	phenomenon	현상명사(자연/생리)	빛. 소리. 불. 바람. 비. 눈물
ABS	abstract	추상명사	목소리. 냄새. 병. 주름
ACT	action	동작,작용	공부. 연습. 견학. 산책
MEN	mental	정신	마음. 의식. 추억. 고민
LIN	linguistic products	언어작품	명사. 뉴스. 설교
CHA	character	성질	아름다움. 결점. 외관. 관용
REL	relation	관계	인연. 원인. 조건. 근거
LOC	location	공간,방향	바깥. 공원. 동쪽. 오른쪽
TIM	time	시간	어제. 일요일. 저녁
QUA	quanty	수량	삼일. 세명. 5kg. 3미터, 전부. 한 사람씩
DIV	diverse		

(2)기타 : 지금까지 기술한 항목에 포함되지 않는 정보를 기재하는 항목이다.

4.2 고찰

본 연구에서 개발한 사전개발시스템인 YDK는 사전개발과정에서의 작업절차와 작업시간을 줄여준다. 이 시스템을 사용한 사전의 구축은 평균적으로 한시간에 6단어의 사전정보 입력이 가능하며 일괄처리 기능을 이용하면 시간당 100단어에 대한 사전자료 구축이 가능하여 매우 빠른시간에 사전자료의 구축이 가능하며 사전자료 입력절차를 최소 40% , 최대 85%를 줄여 사용하기 편리하다는 장점이 있다. 그 절차 비교를 <표 4>에 보인다. 또한 다른 사전과의 비교 평가를 <표 5>에 보인다.

5. 결론

본 논문에서는 다국어 통합정보사전 구축을 위한 한국어 용언의 통합정보사전을 설계하였다. 이를 위해 사전구축 방법론을 정립하고, 정립된 방법론을 바탕으로 하여 통합정보사전의 개발을 위한 통합정보사전 개발 시스템을 설계하고 구현하였다. 개발한 사전은 특정 분야에 맞춰 전문화된 사전이 아니라 기존 사전의 정보를 통합하고 여러 자연언어처리 시스템을 이용하여 얻을 수 있는 정보를 통합한 사전으로 국어정보처리 응용시스템 개발에 도움을 줄 수 있다. 또한, 통합정보사전개발 시스템은 각종 전자사전 및 자연언어처리 시스템들의 정보를 손쉽게 통합할 수 있어 고품질의 사전을 개발할 수 있다. 향후 연구과제로는 개발한 사전체계에 맞추어 단어를 입력해 실제사전을 구축하고 평가를 통해 검증하는 것등이 있으며, 형용사에 대한 항목설계도 필요하다.

<표 4> 사전자료 입력 절차 비교

		단어별 직접입력	일괄입력기능활용
YK의 사전입력	코퍼스로부터 추출한 단어입력	표제어입력 -> 표제어 번역 -> 일어시소러스 검색 -> 항목선택 -> 항목정보 번역 -> 항목정보 입력 -> 자료저장 ::: 7단계	없음
	일어 시소러스 번역	표제어선택 -> 표제어 번역 -> 일어항목선택 -> 한국어항목선택 -> 일어항목의 번역 -> 항목정보 입력 -> 자료저장 :::: 7단계	없음
	코퍼스로부터 추출한 단어입력	표제어입력 -> 일어시소러스 검색 -> 일어 항목선택 -> 한국어항목선택 -> 항목정보의 import -> 자료저장 :::: 6단계	표제어입력 -> 일어시소러스 검색 -> 항목정보의 import -> 자료저장 :::: 4단계
	일어 시소러스 번역	일어표제어 선택 -> 표제어 import -> 일어 항목선택 -> 한국어항목선택 -> 항목정보의 import -> 자료저장 :::: 6단계	입력개시 ==> 입력완료 :: 1단계

<표 5> 각 사전의 비교평가

	항목	IPAL	용언의하위범주화	YDK
사전	단어수	A+	A-	B
	정보수	A	B	A+
	품질	A+	A	A-
사전개발도구	사전개발절차	B+	F	A-
	자료입력속도	A	F	A+
	입력절차	A	F	A+

참고문헌

[1] 시스템공학연구소, 부산대학교 "한국어 문장 분석을 위한 용언의 하위범주화에 관한 연구-최종연구보고서", 시스템공학연구소, 1997

[2] 이재성 외3, "텍스트 및 전자사전 관리 시스템의 설계", 한국정보과학회&한국인지과학회, 제8회 한국어 정보처리 학술대회 논문집, pp.408-414, 1996

[3] 최병진 외3, "표준화를 위한 일반사전의 논리 구조", 한국정보과학회&한국인지과학회, 제8회 한국어 정보처리 학술대회 논문집, pp.415-423, 1996

[4] 한국과학기술원, "텍스트코퍼스 및 전자사전 관리시스템(TDMS)", 과학기술처, 통합 국어정보베이스 최종보고서, pp.17-150, 1996

[5] 황도삼 외2, "자연언어처리", 홍릉과학출판사, 1998