

# 컴퓨터를 이용한 국어 자료 처리 방법

홍 윤 표

## I. 머리말

국어 연구는 다른 분야의 학문과 마찬가지로, 이론과 자료와 실험이라는 과정을 거쳐 이루어져 왔다. 학문 연구의 과정을 마부가 수레를 끄는 과정에 비유해 본다면, 마부는 연구자이고, 수레는 학문이다. 수레는 두 바퀴로 굴러 간다. 그리고 그것을 움직이게 하는 힘은 앞에서 끄는 말[馬]이다. 이론과 자료가 수레의 두 바퀴라고 한다면, 말은 실험도구이다. 두 수레바퀴의 크기가 서로 다르다면 수레는 앞으로 나아가지 못하고 정한 자리를 맴돌 것이며, 그 수레를 끄는 말이 없다면 그 수레는 바퀴가 있어도 무용지물일 것이다. 자료와 이론과 실험은 학문 연구에서 가장 중요한 요소인 셈이다. 양쪽의 두 수레바퀴에 해당하는 이론과 자료는 어느 것이 더 중요하다고 하기 어렵다. 뿐만 아니라 연구자는 실험도구를 선택하거나 새로 만들 수도 있다. 말[馬]을 엔진과 같은 기계로 대체하면 더욱 빠르게 앞으로 나아갈 수 있을 것이다.

디지털 시대는 국어 연구가 이론과 자료와 실험이라는 과정을 거치는 점에서는 아날로그 시대와 마찬가지로이지만, 이론과 자료와 실험에 대한 습득 및 전달 방법에서 아날로그 시대와 차이를 보인다. 자료와 이론은 이전 시대에는 주로 문헌을 통해서 습득할 수 있었고, 실험이라는 것도 고작해야 카드를 만들어 분류하고 이를 체계화하는 것이었다. 그러나 디지털 시대에는 정보를 생산하고 생산된 정보를 검색하여 필요한 자료를 추출하는 시대가 되었다. 따라서 자료나 정보의 양이 비교할 수 없을 만큼 많아졌고 또 다양하여졌다.

국어 연구도 연구실에서 카드를 만들어 분류하던 시대는 가고, 컴퓨터에서 자료를 검색하여 이를 정보화하고 이를 통해 새로운 이론을 창조하는 시대로 접어들었다. 뿐만 아니라 현실세계의 일을 가상세계가 대체하는 시대로 급변하여서, 학문의 세계에서 소위 전자적 연구 개발(Electronic Research & Development, 줄여서 e-R&D라고 한다)이 급속히 확산되어 갈 것이다. 실험이 가상의 공간에서 운영하는 실험실에서 진행되는 것이다. 일정한 언어 자료와 규칙만 제공해 주면 가상의 컴퓨터 시스템은 그 언어의 모음체계와 자음체계를 자동으로 구축해 주는 시대가 온 것이다. 그래서 '○○ 방언의 모음체계 연구' 등의 제목을 가진 논문이 사라질 날도 있을 것으로 예상된다. 이것은 지금 당장 가능한 일이다. 문제는 컴퓨터에 어떠한 자료를 제공해 주는가 하는 문제와 자음체계와 모음체계를 결정하는 명확한 기준을 제공해 주는 일이다. 그래서 연구목적과 연구방법이 달라지는 것이다.

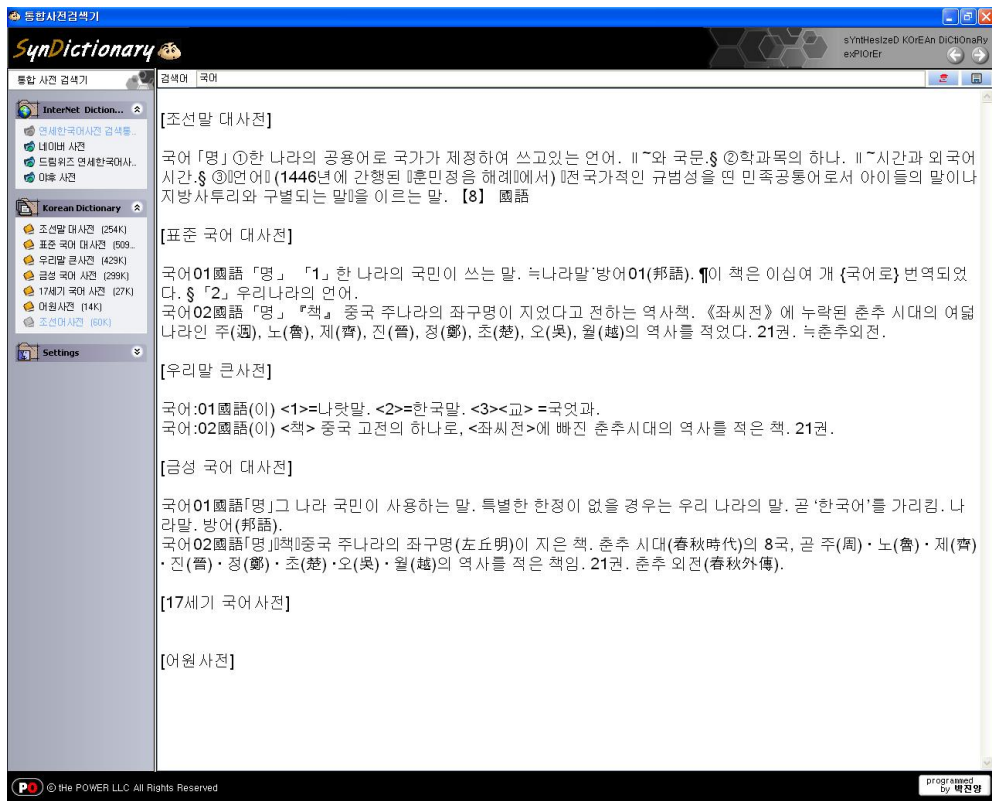
따라서 디지털 시대에 국어학자들의 컴퓨터 활용은 필연이라고 할 수 있다. 여기에서는 국어학자들이 컴퓨터를 이용하여 어떻게 자료를 처리할 수 있는가를 실증적으로 보이도록 한다. 그러나 시간의 제약으로 가장 기초적인 것들만 보이도록 한다.

## II. 컴퓨터의 자료 처리 기능의 활용

일반적으로 컴퓨터는 다음과 같은 아홉 가지 기능을 가지고 있다.

- ① 입력(input) 기능
- ② 제어(control) 기능
- ③ 연산(arithmetic) 기능
- ④ 기억(memory) 기능

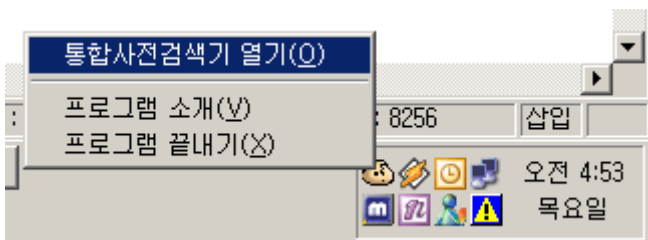




## (1) 프로그램 사용방법

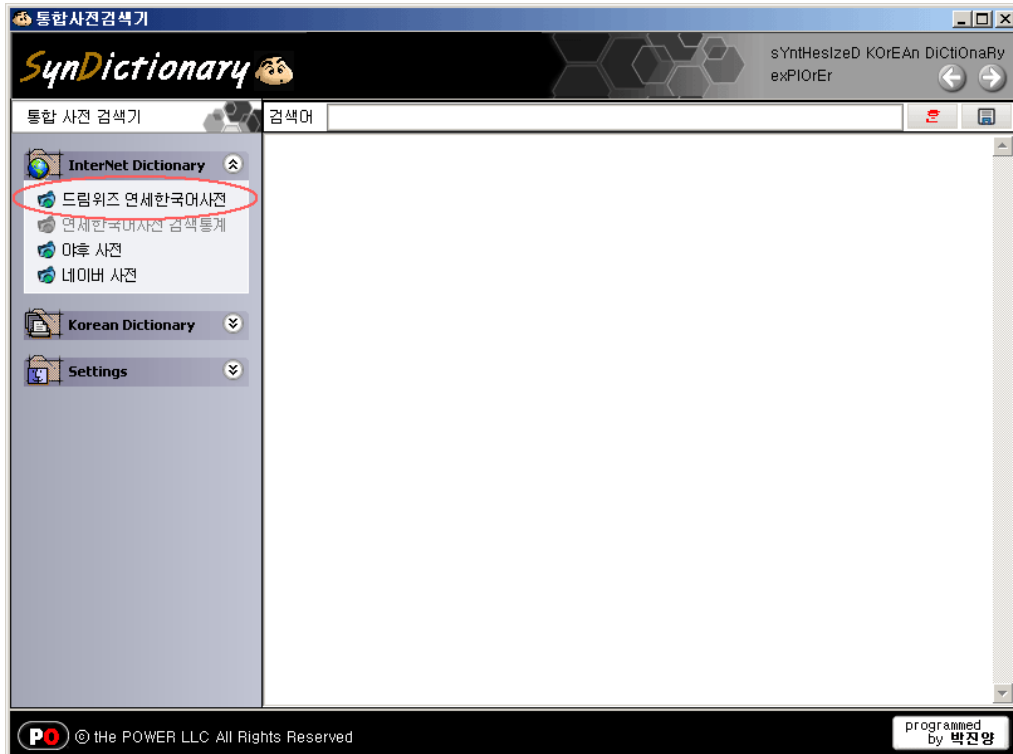
### ① 프로그램의 시작과 종료

컴퓨터를 부팅할 때 자동적으로 실행되도록 설정되어 있다. 따라서 사전 검색을 위해 프로그램을 호출하고자 하는 경우는 화면의 우측 하단에 있는 트레이 아이콘을 오른쪽 마우스 버튼으로 클릭하고 '통합사전 검색기 열기'라는 메뉴를 선택하면 된다. 프로그램을 완전히 종료하고자 하는 경우 역시 화면의 우측 하단에 있는 트레이 아이콘을 오른쪽 마우스 버튼으로 클릭하고 '프로그램 끝내기'를 선택하면 된다.



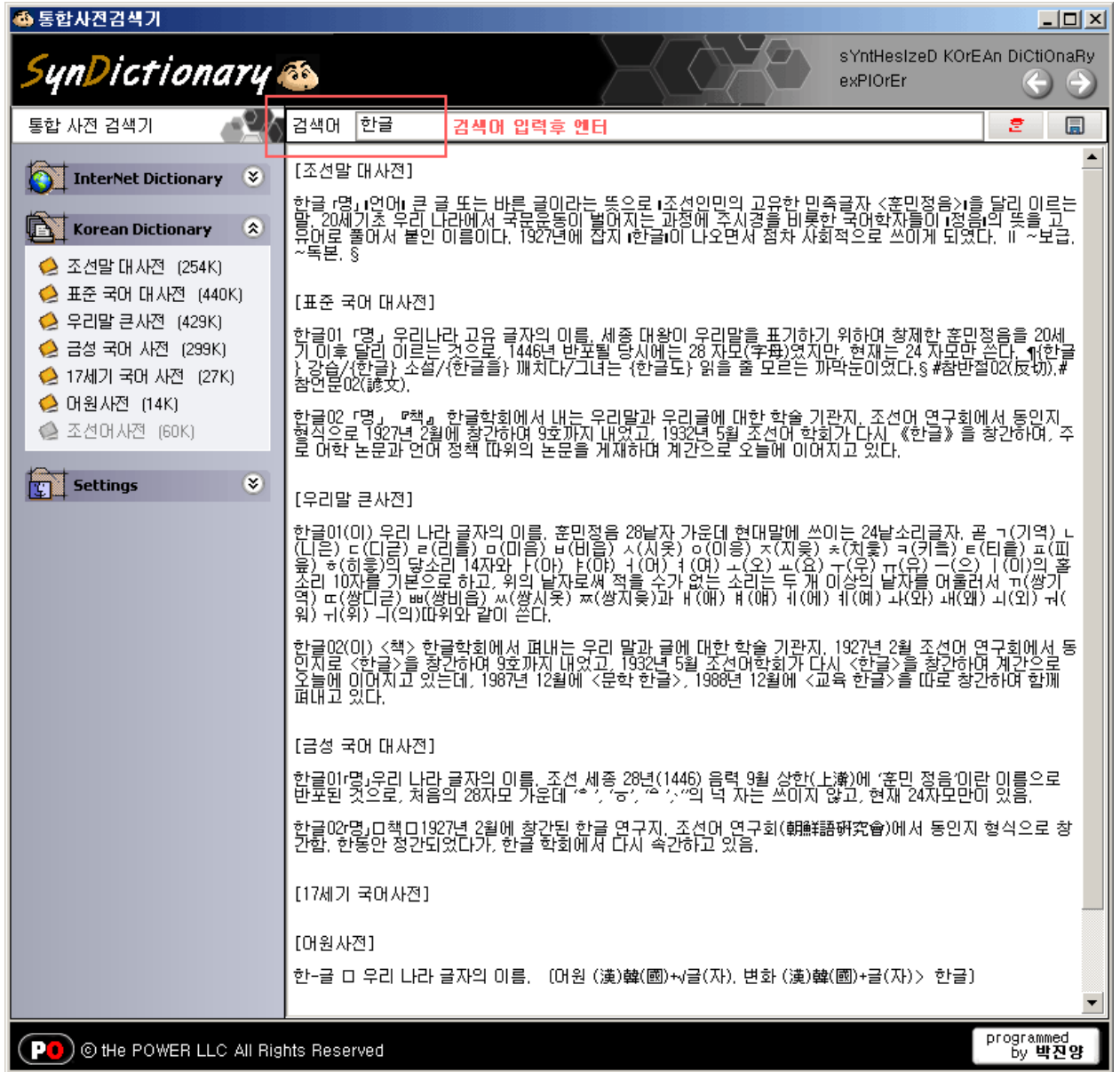
### (2) 인터넷 포털 사이트에서 제공하는 사전의 검색

아래 그림에서 표시된 부분을 클릭하면 연세 한국어 사전을 제공하는 드림위즈로 바로 접속된다. 야후 사전과 네이버 사전 역시 마찬가지이다.



(3) 사전 통합 검색

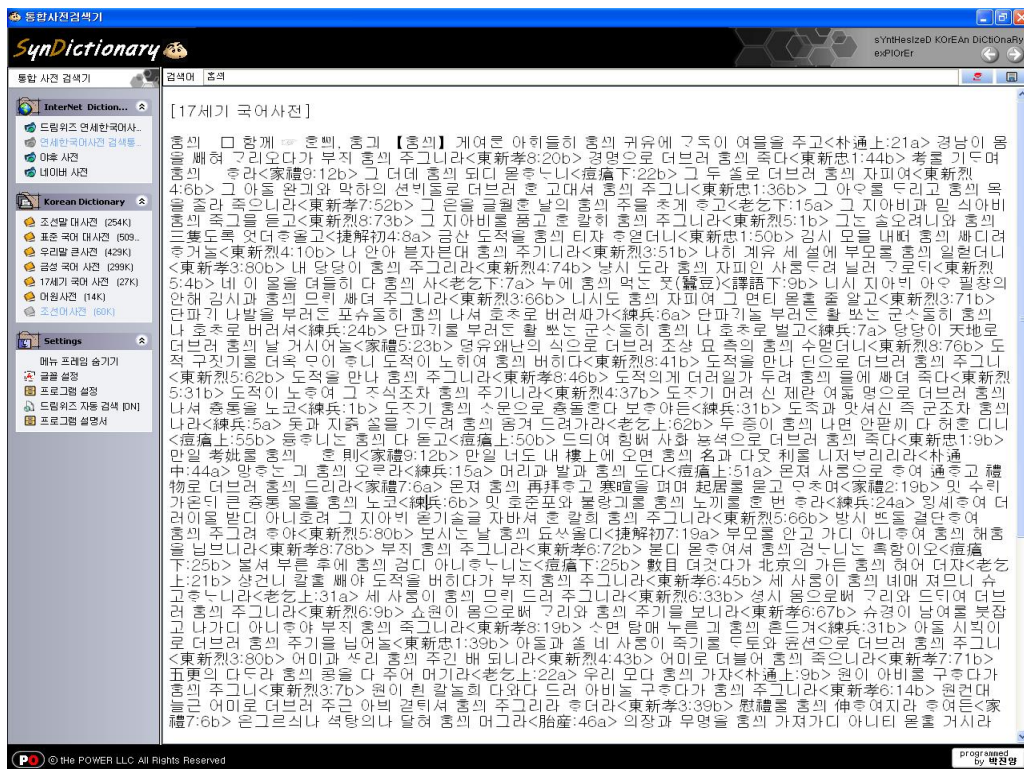
화면 상단 검색어 입력란에 찾고자 하는 검색어를 입력한 후 엔터를 치면 각각의 사전에서 검색된 내용을 모두 볼 수 있다. 여기서 검색되는 사전은 사용자가 지정할 수 있다. 기본값은 모든 사전을 검색하는 것으로 되어있다



만약 17세기 국어사전의 고어를 검색하려면 옛한글을 타자하여야 하는데, 옛한글 타자는 화면의 오른쪽 상단에 붉게 보이는 '훈'이란 아이콘을 클릭하여 옛한글입력기를 열어야 한다. 옛한글 입력기에서 옛한글을 타자하면 된다. 옛한글 타자 방법은 '훈글' 프로그램 타자 방법과 같다. 이렇게 옛한글 타자하는 방법은 다음 화면과 같다.



그리고 입력이란 아이콘을 클릭하면 검색란에 옛한글이 등장한다. 17세기 국어사전에서 ‘흥식’을 검색한 화면을 보이면 다음과 같다.



그런데 그렇게 해서 검색된 자료라고 해도 화면에 옛한글이 보이지 않을 때가 있다. 이때에는 화면 좌측 하단에 보이는 setting을 클릭해서 그중에 ‘글꼴’을 선택하고, 다시 수많은 글꼴 가운데 ‘새굴림’을 선택하면 옛한글을 볼 수 있다.

만약에 글씨가 너무 작아서 보기 힘들 때에는 setting에서 글꼴을 크게 하면 글자를 크게 볼 수 있다. 그 이외의 기능은 setting 아래에 있는 프로그램 설명서를 참조하면 된다.

이렇게 검색된 자료들은 마우스로 지정하여 원문 그대로 문서작성기에 그대로 불러 올 수 있다. 이러한 점에 ‘통합사전 검색기’의 장점이 있다.

이 통합사전 검색기는 입력되어 있는 다양한 사전, 예컨대 관용어 사전, 문세영의 조선어 사전, 심의린의 조선어사전, 고전소설독해사전, 신어사전, 의성의태어사전, 국어학사전, 세계신화사전, 역순사전 등과 각종 방언사전, 즉 강원도 방언사전, 경북방언사전, 제주어사전, 경남방언사전, 평안방언사전, 전남방언사전, 경주방언사전, 한국정신문화연구원 방언자료, 김병제의 방언사전 등등이 추가될 예정으로 보인다.

## 2. 어절별 색인 만드는 프로그램(halign.exe) [이전식 교수 개발]

어절별 색인 작업을 쉽게 해 주는 프로그램이다. 색인과 출전을 동시에 표시해 준다. 발표자는 저서의 색인 부분은 이 프로그램을 이용하여 만들고 있다. 아래에 그 어절별 색인을 만드는 프로그램의 입력의 예[표1-1]와 프로그램을 시행시킨 결과물[표1-2]을 표로 보인다.

[표1-1]

<1a>	보통 부부싸움을 할 때 여자가 바가지 긁는다는 말을 많이 하지요. 왜 바가지 긁는다는 말이 나왔을까요?
<2b>	여자가 남편에 대해 불만이 있으면, 안방과 붙어 있는 부엌으로 들어가서, 남편에게 들리도록 솔에 붙어 있는 누룽지를 떼기 위하여 솔에 물을 붓고서는
<중략>	
<3a>	바가지로 긁어 내는데, 덕이 있는 여자는 그 소리를 내지 않도록 되어 있지만, 남편에게 불만이 있는 여자는 남편이 듣도록 박박 소리를 내어 긁습니다.

[표1-2]

보통	<1a>
부부싸움을	<1a>
할	<1a>
때	<1a>
여자가	<1a>
<중략>	
소리를	<3a>
내어	<3a>
긁습니다.	<3a>
<이하략>	

### (1) 어절별 색인 만드는 프로그램 사용법

### (가) 자료 입력 방법

- ① < > 안에 문헌의 이름 및 페이지가 들어가야 한다.
- ② < > 안의 출전표시는 띄어 써서는 안된다.
- ③ < > 안의 출전 표시는 자료의 맨 위에 와야 한다.(매 페이지마다 출전을 밝혀 주어야 한다)

### (나) 자료 저장 방법

- ① '파일'을 선택하고 '다른 이름으로 저장하기'를 찾은 후
- ② '파일 형식'에서 2바이트 문서'를 선택한 후
- ③ 파일 이름을 '임의의 이름'(각자 선정) 타자한 후
- ④ '찾는 위치'에서 halign.exe가 들어가 있는 서브 디렉토리를 정한 후
- ⑤ '저장'을 한다.

### (다) 어절별 색인 작성 프로그램 활용방법

- ① 컴퓨터 왼쪽 하단의 '시작'을 누른 후
- ② '프로그램'을 누른 후
- ③ '보조 프로그램'을 선택한 후
- ④ 명령 프롬프트'를 선택하면 다음과 같이 DOS용 화면으로 바뀐다.

C:\WDocuments and Settings>

그러면 다음과 같은 명령을 준다. 즉 halign.exe가 들어 있는 서브디렉토리로 옮겨간다. 만약 그 서브디렉토리를 halign이라고 하면 다음과 같은 명령을 준다.

- ⑤ C:\W>cd halign
- ⑥ c:\Whalign>halign 파일 이름(자료화일).2b 만들고 싶은 파일 이름.hwp
- ⑦ 프로그램이 돌아간 후 컴퓨터가 멈춘다.
- ⑧ ×를 누르면 윈도 상태로 돌아간다.

### (라) 어절별 색인 파일 불러 오는 방법

- ① hwp를 실행시킨다.
- ② '불러오기'를 선택하고
- ③ '파일형식'에서 '2바이트파일'을 선택하면 만들어 놓은 파일 이름이 나온다.
- ④ 그것을 선택한 후 두 번 클릭하면 파일을 볼 수 있다.

이 프로그램은 연구자들의 활용목적에 따라 다양하게 활용될 수 있다. 예컨대 문학작품에서 사전에 등재되어 있지 않은 어휘를 뽑으려고 한다면, 물론 그 작품이 입력된 파일이 있어야 한다. 한 예를 들어 보이도록 한다.

우선 연변의 아동문학 작품을 하나 들어 보이도록 한다.

[표2-1]

<정문준, 정든백양나무, 1>

그 말에 분함을 깨우는 계집애의 입술이 파르르 떨리고 있었다. 그런 계집애를 더 골려줄 양으로 다시 제 이름을 새기려 들던 사내애의 손에서 칼이 떨어졌다. 계집애가 칼 쥔 그 애의 손을 때린 것이다. 얼른 칼을 주는 계집애는 내쳐 몸을 빼려 했다. 하지만 사내애의 빠른 달음박질을 뒤떨굴 수 없었다. 사내애가 앞을 막아섰다.

그 칼 내놓지 못하겠니?

다시 나무에 칼을 대지 않고야...

당돌하게 내뱉는 계집애의 고운 눈초리가 매섭게 치며졌다. 손칼은 계집애의 손아귀에서 꼬리를 감추고 있었다.

이 주먹맛 보고 싶어 그래?

제 힘을 톡대고 사내애는 주먹을 둘러메었다. 계집애는 뒤주춤했다. 겁질린 눈에서 이슬 같은 것이 구울다가 눈귀에 맺히고 있었다. 울음을 뱉으며 계집애가 성을 냈다.

넌 나와 함께 저 나무를 심었지 뭐야. 그런데도 넌...봐, 상처 입은 백양나무가 몸부림치는 걸.

대들어 손칼을 빼앗으려던 사내애의 눈길이 백양나무쪽으로 날아갔다. 드세진 듯한 바람 속에 태질하는 백양나무가 사내애의 눈을 꼭 채워주고 있었다. 저 백양나무는 그들이 이곳 가암소학교에 불던 날 심은 것이다. 달구지에 실어온 나무 묘목 가운데서 제일 키가 크고 굵직한 백양을 골라 계집애는 물초통을 기울이고 사내애는 흙을 복돋우면서 어서어서 자라기를 바라며 두 애가 한마음이 되었었다.

<정문준, 정든백양나무, 2>

눈바람이 세차던 어떤 날 새벽, 버짚더미를 안고 계집애가 교정에 나타났었다. 그 애가 한창 백양나무에 가마니를 둘러주고 있는데 새끼줄 통구리를 멘 사내애가 험금썩금 달려왔다.

왜 혼자 왔니? 나한테 알리지도 않고...

사내애가 나무람했다.

간밤 꿈에 나무가 감기에 걸려 쿵쿵 기침을 짓는게 아니겠니. 덜덜 떨고 있겠지 뭐야. 난 병든 나무를 안고 온밤 울었댔어...

스키모자를 올리붙이며 사내애는 이마의 땀을 뺉 문질렀다. 상기된 얼굴에 웃음이 병글거리고 있었다.

참, 이상하지. 왜 똑같은 꿈이 우릴 찾아왔을가, 응?

의문 실린 계집애의 눈에 정기가 반짝 빛났다.

우리가 함께 심은 나무이니깐 나무도 우리에게 똑같은 꿈을 준 게지 뭐.

호호호...우리 심은 꿈나무!

히히히...우리 심은 꿈나무!

그들이 웃고 떠들며 가마니를 두른 백양나무에 새끼줄을 촘촘히 감아주었다. 금시 잠퐁해진 듯한 눈바람을 자장가로 들으며 애어린 백양나무는 꿈속에 깃들고 있었다...

이 문학작품을 어절별 색인을 내면 [표2-2]와 같다. 일부만 보이도록 한다.

[표2-2]

그	<정문준, 정든백양나무, 1>
말에	<정문준, 정든백양나무, 1>
분함을	<정문준, 정든백양나무, 1>
깨무는	<정문준, 정든백양나무, 1>
계집애의	<정문준, 정든백양나무, 1>
입술이	<정문준, 정든백양나무, 1>
파르르	<정문준, 정든백양나무, 1>
떨리고	<정문준, 정든백양나무, 1>
있었다.	<정문준, 정든백양나무, 1>
	<중략>
금시	<정문준, 정든백양나무, 2>
잠풍해진	<정문준, 정든백양나무, 2>
듯한	<정문준, 정든백양나무, 2>
눈바람을	<정문준, 정든백양나무, 2>
자장가로	<정문준, 정든백양나무, 2>
들으면서	<정문준, 정든백양나무, 2>
애어린	<정문준, 정든백양나무, 2>
백양나무는	<정문준, 정든백양나무, 2>
꿈속에	<정문준, 정든백양나무, 2>
깃들고	<정문준, 정든백양나무, 2>
있었다	<정문준, 정든백양나무, 2>

위의 예에서 표준국어대사전에 등재되어 있지 않은 단어를 우선 눈으로 찾아보면 [표 2-3]과 같다.

[표2-3]

용례	어간	출전	확인
뒤떨굴	뒤떨구다	<정문준, 정든백양나무, 1>	있음
주먹맛	주먹맛	<정문준, 정든백양나무, 1>	있음
턱대고	턱대다	<정문준, 정든백양나무, 1>	없음
뒤주춤했다.	뒤주춤하다	<정문준, 정든백양나무, 1>	북한어로 등재
겹질린	겹질리다	<정문준, 정든백양나무, 1>	없음
물초롱을	물초롱	<정문준, 정든백양나무, 1>	없음
벼짚더미를	벼짚더미	<정문준, 정든백양나무, 2>	없음
통구리를	통구리	<정문준, 정든백양나무, 2>	있음
혈금씨금	혈금씨금	<정문준, 정든백양나무, 2>	있음
나무람했다.	나무람하다	<정문준, 정든백양나무, 2>	없음
깃는게	짚다	<정문준, 정든백양나무, 2>	북한어로 등재
스키모자를	스키모자	<정문준, 정든백양나무, 2>	없음
올리불이며	올리불이다	<정문준, 정든백양나무, 2>	없음
잠풍해진	잠풍하다	<정문준, 정든백양나무, 2>	없음
애어린	애어리다	<정문준, 정든백양나무, 2>	북한어로 등재

이 작업을 통해 우리는 ‘턱대다, 물초롱, 겹질리다, 나무람하다, 스키모자, 올리불이다, 잠풍하다’ 등의 새 어휘를 찾아낼 수 있다. 물론 이 중에는 ‘턱대다’가 ‘턱을 대다’, ‘겹질리다’는 ‘겹에 질리다’로, ‘스키모자’는 한 단어가 아닌 것으로, 그리고 ‘올리불이다’는 ‘올려붙이다’로, ‘잠풍하다’는 ‘잔풍(殘風)하다’로 인식할 수 있을지 모른다. 그러나 작품을 일일이 읽어서 새 어휘를 찾아내는 방법과 비교해 볼 때 시간면에서 엄청난 이득을 볼 수 있을 것이

다.

### 3. 용례 사전 만드는 프로그램(kwoc.exe) : 이진식 교수 개발

용례를 자동적으로 만들어 주는 프로그램이다. 다른 프로그램들이 대부분이 용례를 검색하면 핵심어가 가운데 정렬되고 나머지 요소들이 좌우로 배열되는 KWIC(Key Word In Context) 방식을 사용하고 있지만, 이러한 방식으로 검색한 자료들은 실제 이용에서 많은 애로가 있다. 특히 사전 편찬에서 용례를 뽑을 때, 앞뒤의 불필요한 어절들을 일일이 손으로 지우는 어려움을 겪게 된다.

이것을 극복하기 위해서 소위 KWOC(Key Word Out Context) 방식으로 만들어 줄 필요가 있는데, 이 프로그램이 곧 KWOC.EXE이다. 이 용례사전을 만드는 프로그램은 한국학 중앙연구원의 이진식 선생이 개발한 두 개의 간단한 프로그램을 이용하는 것이 좋다. 하나는 어절별 색인 작성을 해 주는 프로그램인 halign.exe이고, 또 하나는 이 프로그램을 이용하여 만든 파일을 다시 이용하여 용례사전을 만들어 주는 kwoc.exe 프로그램이다. 모두 DOS용 프로그램이라 약간 불편한 점은 있지만, 그리 복잡한 것이 아니기 때문에 쉽게 이용할 수 있다.

#### (1) 용례사전 만드는 프로그램 사용법

##### (가) 프로그램 복사 방법

우선 프로그램을 어느 디렉토리 안에 복사해 놓는다. 그 디렉토리의 이름을 '용례사전'이라고 하자. 그 디렉토리 안에 들어갈 파일은 다음과 같다.

㉠ halign.exe   ㉡ kwoc.exe   ㉢ 원문을 입력한 자료(2 바이트 문서)

##### (나) 자료 입력 방법

어절별 색인 작성 프로그램과 동일하지만, 반드시 지켜야 할 원칙이 있다. 그것은 문장 단위마다 반드시 마침표를 찍어 두어야 한다는 점이다.

##### (다) 자료 저장 방법

어절별 색인 작성 프로그램과 동일(확장자 .2b 의 파일로 저장)

##### (라) 용례사전 만드는 프로그램 활용법

① 윈도상에서 '시작'을 선택하고 다시 '모든 프로그램'을 선택, 그리고 다시 '보조프로그램'을 선택한 후 그 속에서 '명령 프롬프트'를 선택한다.

② 그러면 컴퓨터의 본 화면이 검은 바탕의 도스 환경으로 바뀔 것이다.

③ 디렉토리를 '용례사전'으로 바꾸어 놓고 다음과 같은 몇 가지 명령을 이용한다.

㉠ C:>cd 용례사전

㉡ C:\용례사전>halign 임의의 파일 이름.2b 만들 파일 이름.hwp

㉢ C:\용례사전> 로 끝난다.

㉣ C:\용례사전>kwoc 만든 파일 이름. 2b 다시 만들 파일 이름.hwp

##### (마) 용례사전 조사 파일 불러 오는 방법

어절별 색인 작성 프로그램과 동일

그러면 모든 작업이 끝나고 어절별 용례사전이 만들어질 것이다.

이 과정을 실제 『개수첩해신어』의 일례를 들어 설명해 보기로 한다.

[표3-1] 원문 입력 자료의 모습<sup>1)</sup>

<p>&lt;改捷1,01a&gt; 아모가히 이러 오라. 네 代官 中에 가 내 전갈로 &lt;改捷1,01b&gt; 니르기틀. 그적과 여기 느려와 어제라도 오을 거슬. 路次의 곧부매 &lt;改捷1,02a&gt; 계요 이제야 守門까지 완습닌. 안히 계시면 오려 헛더라 니르고 &lt;改捷1,02b&gt; 오라. 御念比흔 御使   읍도쇠. 어와 오옵시니 아람답스외. &lt;改捷1,03a&gt; 안히 일스오니 대되 同道하여 오쇼셔. 代官 中도 &lt;改捷1,03b&gt; 흔 끝에 일습닌. 아람답스외 여기 오로옵소. 아직 편히 안습소. &lt;改捷1,04a&gt; 나는 소임으로 완습거니와 처음이옵고. 그 외에 칭소흔 거시오니 &lt;改捷1,04b&gt; 各各 민망히 너기실가 근심하옵닌. 萬事   두루 쓰리심을 &lt;改捷1,05a&gt; 미털습닌. 인스장의 말씀이어니와 판당히 니르시는 양이로소이다.</p>
---

[표3-2] halign 프로그램을 실행시킨 후의 목적파일<sup>1)</sup>의 모습(이 파일은 곧 어절 색인이 될 것이다)

1) 프로그램에서 원문 입력 구조는 매우 중요하다. 그런데 이 프로그램이나 ‘깜짝새’ 프로그램에서는 원문 입력 구조를 동일하게 하였으므로, 이러한 형식으로 입력을 하면 거의 모든 프로그램을 이용할 수 있을 것으로 기대된다.

아모가히	<改捷1,01a>
이러	<改捷1,01a>
오라.	<改捷1,01a>
네	<改捷1,01a>
代官	<改捷1,01a>
中에	<改捷1,01a>
가	<改捷1,01a>
내	<改捷1,01a>
전갈로	<改捷1,01a>
니르기롤.	<改捷1,01b>
그적과	<改捷1,01b>
여기	<改捷1,01b>
느려와	<改捷1,01b>
어제라도	<改捷1,01b>
오을	<改捷1,01b>
거술.	<改捷1,01b>
<이하 생략>	

[표3-3] kwoc 프로그램을 실행시킨 후의 최종 목적파일의 모습

아모가히	아	히	이러	오라.	<改捷1,01a>		
이	아	히	이러	오라.	<改捷1,01a>		
오라.	아	히	이러	오라.	<改捷1,01a>		
네	내	代官	中에	내	전갈로	니르기롤	捷1 1
代	내	代官	中에	내	전갈로	니르기롤	捷1 1
中에	내	代官	中에	내	전갈로	니르기롤	捷1 1
가	내	代官	中에	내	전갈로	니르기롤	捷1 1
로	내	代官	中에	내	전갈로	니르기롤	捷1 1
니르기롤	내	代官	中에	내	전갈로	니르기롤	捷1 1b
그적과	그	여기	느려와	어제	도	을	거술 捷1 1b
	그	여기	느려와	어제	도	을	거술 捷1 1b
느	그	여기	느려와	어제	도	을	거술 捷1 1b
	그	여기	느려와	어제	도	을	거술 捷1 1b
오	그	여기	느려와	어제	도	을	거술 捷1 1b
거 .	그	여기	느려와	어제	도	을	거술 捷1 1b
<이하 생략>							

⑤ 앞의 어절 목록을 복사 기능을 이용하여 다음의 [표3-4]와 같이 만든다.

[표3-4]

아모가히	아 히	가 러 오	捷1 1	
이	이러	가 러 오	捷1 1	
오라.	오	가 러 오	捷1 1	
네	네	官中 가	로 니	를 <改 , a
代	代官	官中 가	로 니	를 <改 , a
中에	中	官中 가	로 니	를 <改 , a
가	가	官中 가	로 니	를 <改 , a
		官中 가	로 니	를 <改 , a
로	갈	官中 가	로 니	를 <改 , a
니르기롤	니 룰	官中 가	로 니	를 <改 ,
그적기	그	기 러	오	술 <改 ,
	기	기 러	오	술 <改 ,
느	느려와	기 러	오	술 <改 ,
	제라	기 러	오	술 <改 ,
오	오을	기 러	오	술 <改 ,
거 .	거술	기 러	오	술 <改 ,

<이하 생략>

⑥ [표3-4]의 왼쪽 어절에 어간형을 고쳐 타자하여 놓는다. 그러면 [표3-5]와 같은 모습을 보일 것이다.

[표3-5]

아모가히	아 가히	아 가히	이러 오 .	改捷1 01 >
이러	이러	아 가히	이러 오 .	改捷1 01 >
오다	오	아 가히	이러 오 .	改捷1 01 >
네	네	네 代官中 가	갈로 니 기롤.	<改捷 ,0 a>
代官	代官	네 代官中 가	갈로 니 기롤.	<改捷 ,0 a>
中	中	네 代官中 가	갈로 니 기롤.	<改捷 ,0 a>
가다	가	네 代官中 가	갈로 니 기롤.	<改捷 ,0 a>
		네 代官中 가	갈로 니 기롤.	<改捷 ,0 a>
갈	갈로	네 代官中 가	갈로 니 기롤.	<改捷 ,0 a>
니르	니 기롤	네 代官中 가	갈로 니 기롤.	<改捷 ,0 b>
그적기	그 기	그 기	기 느려	제 오을 거술. <改捷 ,0 b>
기	기	그 기	기 느려	제 오을 거술. <改捷 ,0 b>
느려오	느려와	그 기	기 느려	제 오을 거술. <改捷 ,0 b>
제	제라	그 기	기 느려	제 오을 거술. <改捷 ,0 b>
오다	오을	그 기	기 느려	제 오을 거술. <改捷 ,0 b>
것	거술	그 기	기 느려	제 오을 거술. <改捷 ,0 b>

<이하 생략>

⑦ 어간형을 가나다순으로 정렬한다  
그 결과 얻은 자료는 [표3-6]과 같다.

[표3-6]

가다	가	官中	가	로니	롤 <改 , a
것	거슬	기	려	오	솔 <改 ,
그적기	그	기	려	오	솔 <改 ,
		官中	가	로니	롤 <改 , a
네	네	官中	가	로니	롤 <改 , a
니르	니	官中	가	로니	롤 <改 ,
代	代官	官中	가	로니	롤 <改 , a
아모가히	아	가	러오	捷1	1
	제라	기	려	오	솔 <改 ,
	기	기	려	오	솔 <改 ,
오다	오	가	러오	捷1	1
오다	오을	기	려	오	솔 <改 ,
이	이러	가	러오	捷1	1
	갈	官中	가	로니	롤 <改 , a
中	中	官中	가	로니	롤 <改 , a
느 오	느려와	기	려	오	솔 <改 ,
				략	

⑧ 이 파일들을 이용하여 각종 형태의 용례사전을 만들 수 있다. 예상되는 형태는 매우 다양할 수 있다. 예를 하나만 든다면 [표3-7]과 같다.

[표3-7]

가다						
[가]	官中	가	로	니	를	<改 , a
것						
[거	의	려		오	술	<改 ,
그걸						
[그걸	의	려		오	술	<改 ,
[ ]	官中	가	로	니	를	<改 , a
네						
[네]	官中	가	로	니	를	<改 , a
니르						
[니르	官中	가	로	니	를	<改 ,
代						
[代	官中	가	로	니	를	<改 , a
아모가히						
[아모가히]	가	려	오			捷1 1
어제	도	그	여기	느려와	어제	도
					올	거
					술	捷1 1b
여기]	그	여기	느려와	어제	도	올
오다					거	술
[	아	히	이러	오라.		<改捷1, 01a>
오다						
을]	그	여기	느려와	어제	도	올
이					거	술
러]	아	히	이러	오라.		<改捷1, 01a>
전갈로]	내	代官	中	에	내	전갈로
中					니르	기
[	내	代官	中	에	내	전갈로
느					니르	기
오					를	捷1 1
려 ]	그	여기	느려와	어제	도	올
					거	술
						捷1 1b
						<이하 생략>

위의 형식은 올림말과 그 올림말의 곡용형과 활용형을 보여 주고 이에 대한 용례와 출전을 동시에 보여 준다.

⑨ 각 어휘에 문법 정보와 의미정보(뜻풀이 등)를 제공해 주면 용례사전의 작업은 모두 끝나는 것이다.

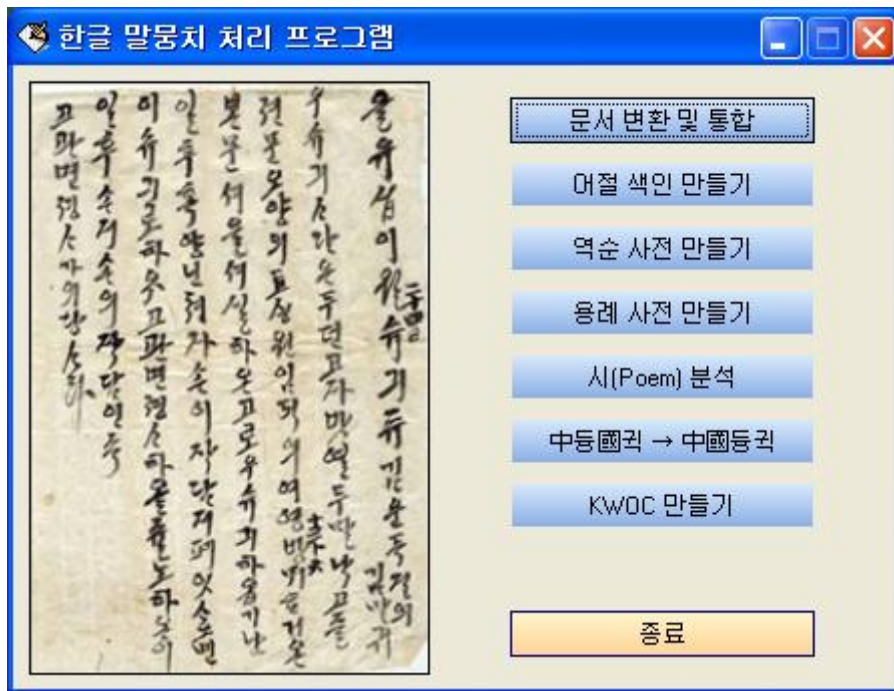
#### 4. 한국어 말뭉치 처리 프로그램 [박진양 개발]

한국어 말뭉치 처리 프로그램은 다음과 같은 7 가지 기능을 가진 프로그램이다.

- ① 문서 변환 및 통합
- ② 어절 색인 만들기
- ③ 역순 사전 만들기
- ④ 용례사전 만들기

- ⑤ 시(poem) 분석
- ⑥ 中등國권 → 中國등권
- ⑦ KWOC 만들기

이 프로그램은 한 장의 CD로 되어 있는데, 이것을 장착시키면 화면에 ‘한국어 말뭉치 처리 프로그램’이라는 아이콘이 생길 것이다. 그 화면은 매우 작아서 잘못 생각하면 프로그램이 잘못 실행된 것인 줄로 알겠지만, 실제로는 사용에 큰 지장이 없다. 그 프로그램을 실행시키면 다음과 같은 화면이 보일 것이다.

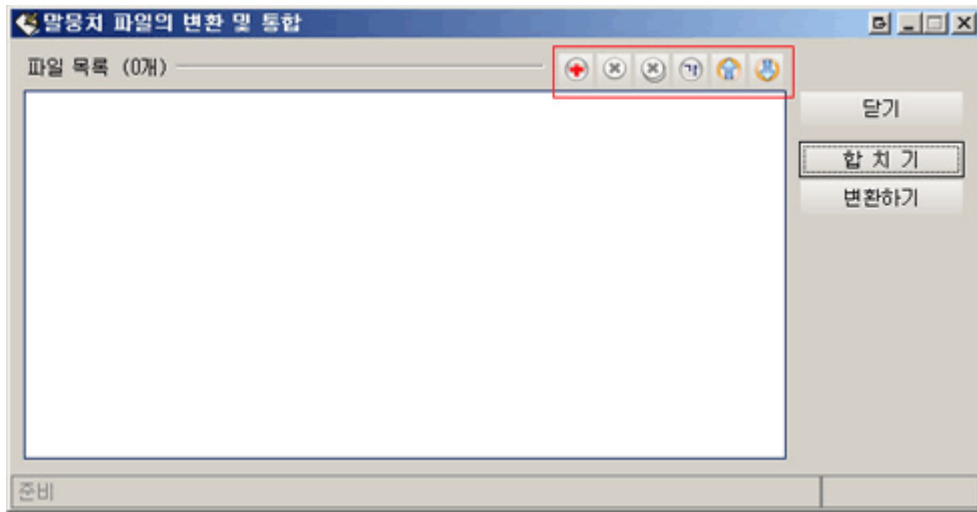


### (1) 문서 변환 및 통합

이 프로그램은 말뭉치 파일의 형식을 변환하거나 여러 개의 말뭉치 파일을 하나로 통합할 수 있는 프로그램이다. 가능한 파일의 형식은 HWP, 2B, 유니코드 텍스트, 완성형 텍스트, 조합형 텍스트 등이다. 그러나 문제가 없이 프로그램이 작동하는 파일은 유니코드 텍스트 파일이다.

#### (가) 프로그램 사용 방법

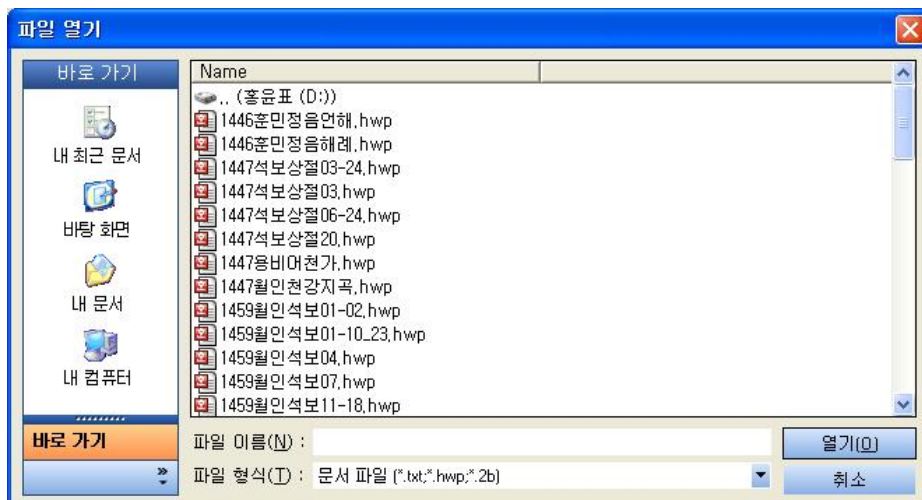
- ① 프로그램에서 '문서 변환 및 통합'을 클릭한다.
- ② 아래의 화면에서 아이콘을 클릭하여 작업할 파일의 목록을 선택한다. 각 아이콘들의 기능은 아래와 같다.



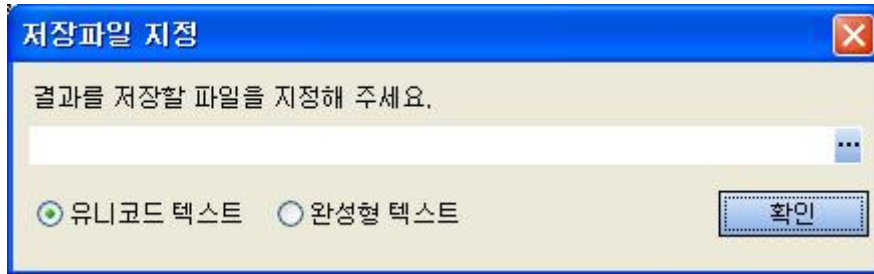
아이콘	기능
+	파일들을 파일목록에 추가한다.
⊗	파일 목록에서 선택한 파일을 파일목록에서 삭제한다.
⊘	파일 목록에서 모든 파일을 삭제한다.
⊎	파일 목록을 정렬한다.
↑	선택된 파일을 한 칸 위로 올린다.
↓	선택된 파일을 한 칸 아래로 내린다.

이중에서 가장 요긴하게 사용할 수 있는 기능은 ‘파일 합치기’이다. 왜냐 하면 검색을 위해서는 모든 말뭉치를 합쳐야 하기 때문이다. ‘한글’ 프로그램에서는 이어붙이기를 하여야 하는데, 이 일은 매우 번거로운 일이며, 시간도 꽤나 소비되는 작업이다. 그렇지만 이 프로그램을 이용하면 순식간에 많은 파일을 한 파일로 만들 수 있다.

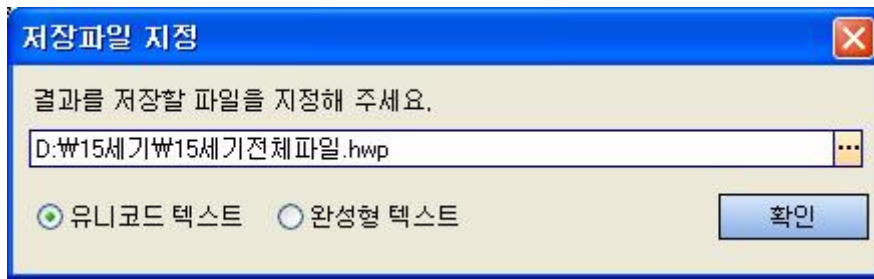
말뭉치 파일들을 통합할 경우에는 먼저 + 아이콘을 클릭하여 합칠 파일이 저장되어 있는 디렉토리를 정하고, 그 중에서 합칠 파일들을 마우스로 클릭하여 선정한다. 다음 화면이 그러한 경우이다.



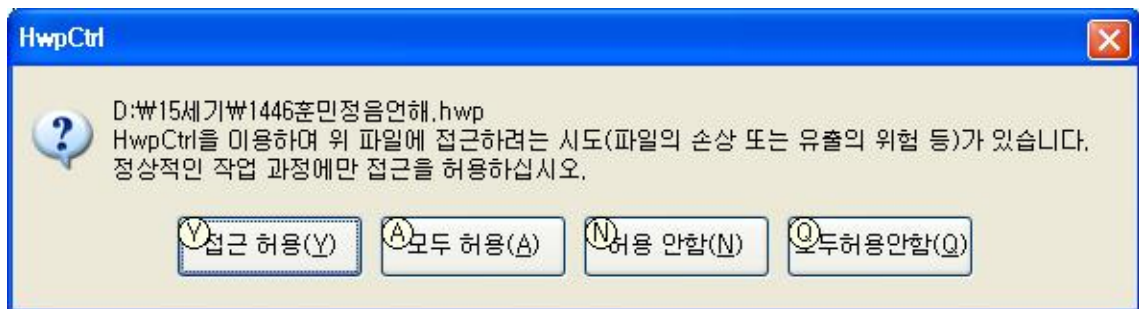
그 후에 '합치기' 버튼을 클릭한다. '합치기'를 클릭하면 다음 그림과 같이 '결과를 저장할 파일을 지정해 주세요'란 안내문이 나온다.



오른쪽의 '...' 기호를 클릭한 후에 서브디렉토리를 선택한 후에 파일이름을 타자한 후 '저장'을 클릭하면 다음과 같은 화면이 나올 것이다.



'확인'을 클릭하면 다음과 같은 화면이 될 것이다.



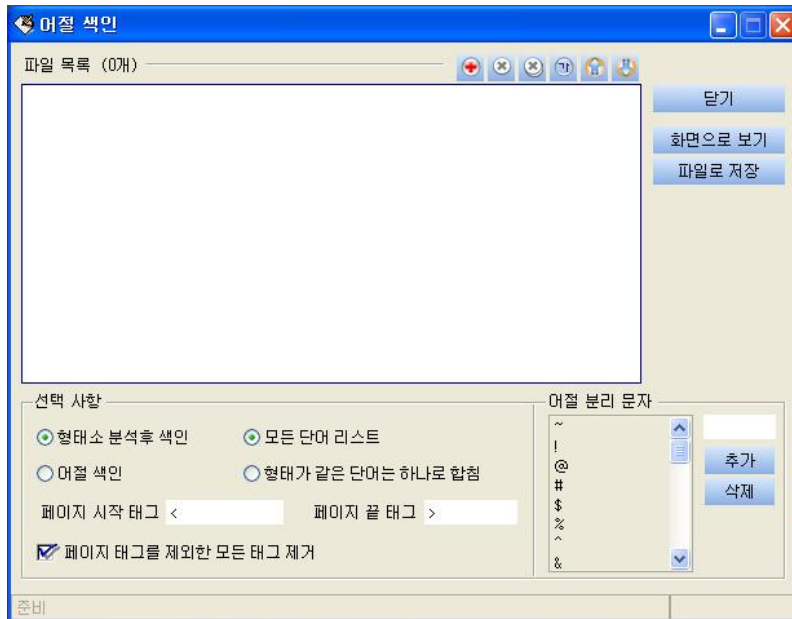
여기에서 '모두 허용'을 클릭하면 오른쪽 하단에 파일이 합쳐지는 과정이 보일 것이다. 그러면 모든 작업은 끝난다.

## (2) 어절 색인 만들기

이 프로그램은 말뭉치 파일에서 어절 색인을 만들기 위한 프로그램이다. 그 사용법은 다음과 같다.

### (가) 프로그램 사용법

프로그램에서 '어절 색인 만들기'를 클릭한다. 그러면 다음과 같은 화면이 나타날 것이다.



이 화면에 나타나는 몇 가지 선택사항은 다음의 의미를 지닌다.

표지	기능
형태소 분석 후 색인	형태소 분석을 수행하여 형태소 단위로 색인을 만든다. 단, 대상 말뭉치가 현대 한국어 말뭉치일 때만 사용한다.
어절 색인	어절 단위로 색인을 만든다.
모든 단어 리스트	중복을 고려하지 않고 어절이나, 형태소를 배열한다.
형태가 같은 단어는 하나로 합침	같은 어절이나 형태소가 중복되어 나타나는 경우 어휘는 하나로 표시하고 뒤에 출전 표시를 추가한다.
페이지 시작 태그, 페이지 끝 태그	페이지를 표시하는 태그의 시작과 끝 태그를 기록한다. 일반적인 경우 페이지 시작, 끝 태그는 '<', '>'이고, 세종 말뭉치인 경우는 '<pb', '>'이다.
페이지 태그를 제외한 모든 태그 제거	말뭉치에 여러 가지 태그가 나타나는 경우(예, 세종 말뭉치) 위에서 명시한 페이지 태그를 제외한 모든 태그를 무시한다.

이용자는 위의 표지 중에서 선택할 수 있다. 어절 색인을 만들 때에는 두 가지 방식이 있다. 하나는 모든 말뭉치를 형태소 분석을 한 후에 어절 색인을 해 주는 방식이고, 또 하나는 단순히 어절 색인만 해 주는 방식이다. 우선 형태소를 분석한 후에 색인을 내고 싶으면 이것을 선택하도록 한다. 그리고 대부분은 페이지 태그를 제외한 모든 태그를 제거하고 어절 색인을 내는 것이 좋을 것이다.

오른쪽 상단에 있는 아이콘의 의미는 앞의 '문서 변환 및 통합'에서 보인 것과 마찬가지로이다. 우선 어절색인을 만들 파일을 열도록 한다. 이 파일은 '유니코드 txt'로 저장된 파일인 것이 좋다. hwp 파일은 실행결과가 좋지 않은 것으로 보인다.

예를 들기 위해, 앞에서 어절 색인을 위해 만들었던 연변 아동문학 작품을 대상으로 하기로 한다.

이를 실행하되, '화면으로 보기'를 선택하면 다음과 같은 화면을 얻을 것이다.

어휘	출현 페이지
하지만 / MAJ	<정문준 정든백양나무.1>
논 / ETM	<정문준 정든백양나무.1>
논 / ETM	<정문준 정든백양나무.1>
논 / ETM	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.1>
미 / VCP	<정문준 정든백양나무.2>
미 / VCP	<정문준 정든백양나무.2>
미 / VCP	<정문준 정든백양나무.2>
미 / VCP	<정문준 정든백양나무.2>
미 / VCP	<정문준 정든백양나무.2>
미 / VCP	<정문준 정든백양나무.2>
미 / VCP	<정문준 정든백양나무.2>
논바람 / NNG	<정문준 정든백양나무.2>
논바람 / NNG	<정문준 정든백양나무.2>
매 / NNG	<정문준 정든백양나무.1>
매 / NNG	<정문준 정든백양나무.1>
매 / NNG	<정문준 정든백양나무.2>
으로 / JKB	<정문준 정든백양나무.1>
으로 / JKB	<정문준 정든백양나무.1>
손칼 / NNG	<정문준 정든백양나무.1>
손칼 / NNG	<정문준 정든백양나무.1>
모퉁 / NNG	<정문준 정든백양나무.1>
주은 / NNG	<정문준 정든백양나무.1>
때리 / VV	<정문준 정든백양나무.1>
머서머서 / MAG	<정문준 정든백양나무.1>
미 / MM	<정문준 정든백양나무.1>
하 / VX	<정문준 정든백양나무.1>
마 / EF	<정문준 정든백양나무.1>
마 / EF	<정문준 정든백양나무.2>
목욕 / NNG	<정문준 정든백양나무.2>

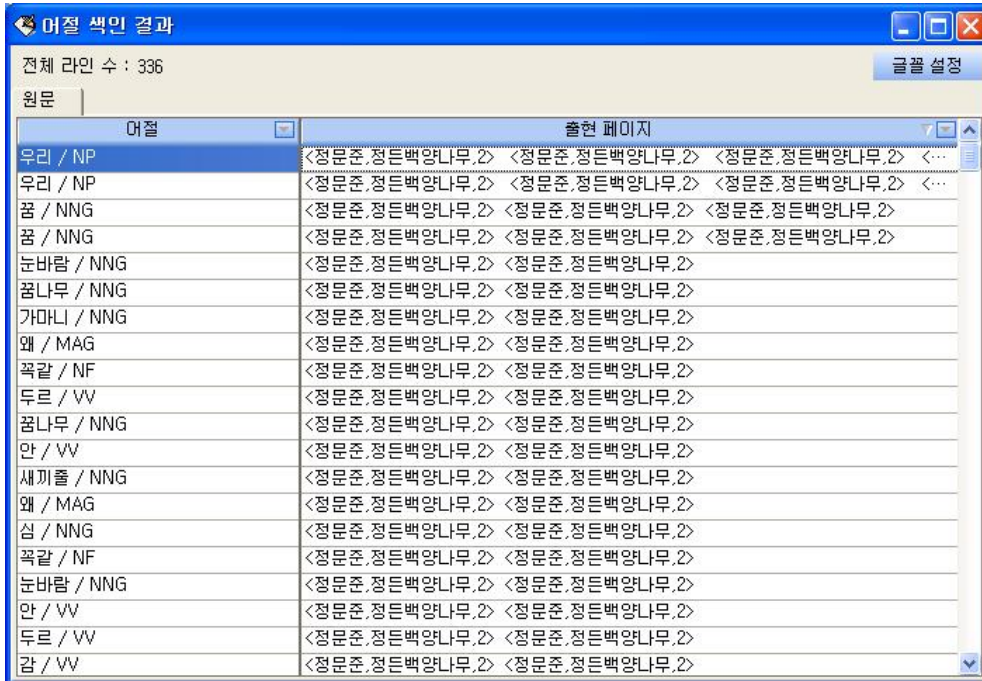
이것을 '파일로 저장'을 선택하면 결과 파일이 만들어진다. 그런데 이때 조심하여야 할 일 이 있다. 그것은 결과 파일이 원래의 원문 파일과 동일한 이름으로 저장된다는 점이다. 그러므로 서브디렉토리를 달리 하여 저장을 하는 편이 좋다.

만들어진 파일을 불러 오면 [표4-1]과 같이 될 것이다. 이중 일부를 보이면 다음과 같다.

[표4-1]

하지만 / MAJ	<정문준,정든백양나무,1>
는 / ETM	<정문준,정든백양나무,1>
는 / ETM	<정문준,정든백양나무,1>
는 / ETM	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
이 / VCP	<정문준,정든백양나무,2>
눈바람 / NNG	<정문준,정든백양나무,2>
눈바람 / NNG	<정문준,정든백양나무,2>
<중략>	
던 / ETM	<정문준,정든백양나무,1>
던 / ETM	<정문준,정든백양나무,1>
던 / ETM	<정문준,정든백양나무,2>
떨리 / VV	<정문준,정든백양나무,1>
울리붙이 / NF	<정문준,정든백양나무,2>
ㄴ / JX <정문준,정든백양나무,1>	
ㄴ / JX <정문준,정든백양나무,1>	
팍 / MAG	<정문준,정든백양나무,1>

그러나 선택에서 ‘형태소 분석후 색인’과 ‘형태가 같은 단어는 하나로 합침’을 지정하고, ‘화면으로 보기’를 선택하면 다음과 같은 화면이 될 것이다.



이것을 파일로 저장해서 열어 보면 [표4-2]와 같을 것이다. 그 일부를 보인다. 이 부분은 탭을 이용하여 필드를 지정할 수 있다.

[표4-2]

하지만 / MAJ	<정문준,정든백양나무,1>
는 / ETM	<정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1>
이 / VCP	<정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,2> <정문준,정든백양나무,2> <정문준,정든백양나무,2> <정문준,정든백양나무,2> <정문준,정든백양나무,2>
눈바람 / NNG	<정문준,정든백양나무,2> <정문준,정든백양나무,2>
애 / NNG	<정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,2>
으로 / JKB	<정문준,정든백양나무,1> <정문준,정든백양나무,1>
손칼 / NNG	<정문준,정든백양나무,1> <정문준,정든백양나무,1>
묘목 / NNG	<정문준,정든백양나무,1>
주은 / NNG	<정문준,정든백양나무,1>
때리 / VV	<정문준,정든백양나무,1>
어서어서 / MAG	<정문준,정든백양나무,1>
이 / MM	<정문준,정든백양나무,1>
하 / VX	<정문준,정든백양나무,1>
야 / EF	<정문준,정든백양나무,1> <정문준,정든백양나무,2>
웃음 / NNG	<정문준,정든백양나무,2>
는 / JX	<정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,2> <정문준,정든백양나무,2>
가 / XPN	<정문준,정든백양나무,1>
바람 / NNG	<정문준,정든백양나무,1>
금 / XSN	<정문준,정든백양나무,2>
있 / EP	<정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,2> <정문준,정든백양나무,2> <정문준,정든백양나무,2> <정문준,정든백양나무,2>
계 / EC	<정문준,정든백양나무,1> <정문준,정든백양나무,1> <정문준,정든백양나무,2>
려 / EC	<정문준,정든백양나무,1> <정문준,정든백양나무,1>
ㄴ데도 / EF	<정문준,정든백양나무,1>

만약에 ‘형태소 분석후 색인’을 택하지 않고 단순히 ‘어절색인’을 선택하고, 역시 ‘모든 단어 리스트’를 선택하여 ‘화면으로 보기’를 클릭하면 다음과 같은 화면을 볼 것이다.

머질 색인 결과

전체 라인 수 : 163

글꼴 설정

머질	출현 페이지
그	<정문준,정든백양나무,1>
말에	<정문준,정든백양나무,1>
분함을	<정문준,정든백양나무,1>
깨무는	<정문준,정든백양나무,1>
계집애의	<정문준,정든백양나무,1>
입술이	<정문준,정든백양나무,1>
파르르	<정문준,정든백양나무,1>
떨리고	<정문준,정든백양나무,1>
있었다	<정문준,정든백양나무,1>
그런	<정문준,정든백양나무,1>
계집애를	<정문준,정든백양나무,1>
더	<정문준,정든백양나무,1>
골려줄	<정문준,정든백양나무,1>
양으로	<정문준,정든백양나무,1>
다시	<정문준,정든백양나무,1>
제	<정문준,정든백양나무,1>
이름을	<정문준,정든백양나무,1>
새기려	<정문준,정든백양나무,1>
들던	<정문준,정든백양나무,1>
사내애의	<정문준,정든백양나무,1>

이것을 파일로 저장하여 열어 보면 [표4-3]과 같은 모습일 것이다. 일부만 보이도록 한다.

[표4-3]

그	<정문준,정든백양나무,1>
말에	<정문준,정든백양나무,1>
분함을	<정문준,정든백양나무,1>
깨무는	<정문준,정든백양나무,1>
계집애의	<정문준,정든백양나무,1>
입술이	<정문준,정든백양나무,1>
파르르	<정문준,정든백양나무,1>
떨리고	<정문준,정든백양나무,1>
있었다	<정문준,정든백양나무,1>
그런	<정문준,정든백양나무,1>
계집애를	<정문준,정든백양나무,1>
더	<정문준,정든백양나무,1>
골려줄	<정문준,정든백양나무,1>
양으로	<정문준,정든백양나무,1>
다시	<정문준,정든백양나무,1>
제	<정문준,정든백양나무,1>
이름을	<정문준,정든백양나무,1>
새기려	<정문준,정든백양나무,1>
들던	<정문준,정든백양나무,1>
사내애의	<정문준,정든백양나무,1>
손에서	<정문준,정든백양나무,1>
칼이	<정문준,정든백양나무,1>
떨어졌다	<정문준,정든백양나무,1>
계집애가	<정문준,정든백양나무,1>
칼	<정문준,정든백양나무,1>
권	<정문준,정든백양나무,1>

이렇게 하지 않고 ‘어절색인’을 택하고 역시 ‘형태가 같은 단어는 하나로 합침’을 선택한 후에 화면으로 보기를 클릭하면 다음과 같은 화면이 보일 것이다.

어절	출현 페이지
그	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1>
말에	<정문준,정든백양나무,1>
분합을	<정문준,정든백양나무,1>
깨무는	<정문준,정든백양나무,1>
계집애의	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1>
입술이	<정문준,정든백양나무,1>
파르르	<정문준,정든백양나무,1>
떨리고	<정문준,정든백양나무,1>
있었다	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1><정...
그런	<정문준,정든백양나무,1>
계집애를	<정문준,정든백양나무,1>
더	<정문준,정든백양나무,1>
골려줄	<정문준,정든백양나무,1>
양으로	<정문준,정든백양나무,1>
다시	<정문준,정든백양나무,1><정문준,정든백양나무,1>
제	<정문준,정든백양나무,1><정문준,정든백양나무,1>
이름을	<정문준,정든백양나무,1>
새기려	<정문준,정든백양나무,1>
들던	<정문준,정든백양나무,1>
사내애의	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1><정...

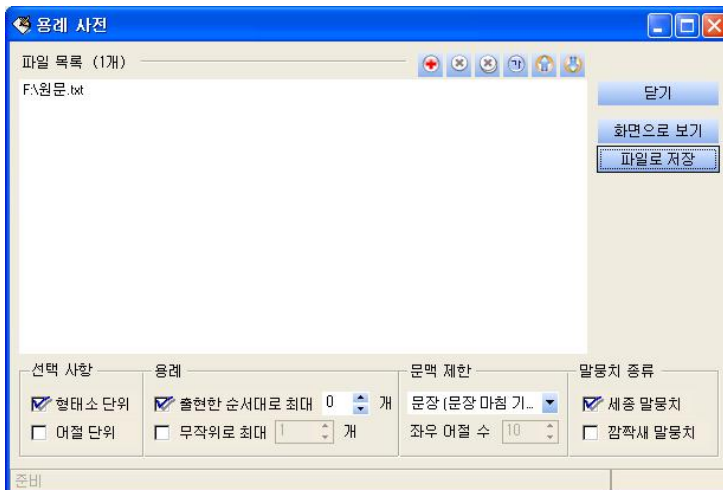
이것을 ‘파일로 저장’하여 그 파일을 보면 [표4-4]와 같은 모습을 보일 것이다. 그 일부를 보이면 다음과 같다.

[표4-4]

그	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1>
말에	<정문준,정든백양나무,1>
분함을	<정문준,정든백양나무,1>
깨무는	<정문준,정든백양나무,1>
계집애의	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1>
입술이	<정문준,정든백양나무,1>
파르르	<정문준,정든백양나무,1>
떨리고	<정문준,정든백양나무,1>
있었다	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1>
그런	<정문준,정든백양나무,1>
계집애를	<정문준,정든백양나무,1>
더	<정문준,정든백양나무,1>
골려줄	<정문준,정든백양나무,1>
양으로	<정문준,정든백양나무,1>
다시	<정문준,정든백양나무,1><정문준,정든백양나무,1>
제	<정문준,정든백양나무,1><정문준,정든백양나무,1>
이름을	<정문준,정든백양나무,1>
새기려	<정문준,정든백양나무,1>
들던	<정문준,정든백양나무,1>
사내의	<정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1><정문준,정든백양나무,1>
손에서	<정문준,정든백양나무,1>
같이	<정문준,정든백양나무,1>
떨어졌다	<정문준,정든백양나무,1>

### (3) 용례사전 만들기

‘역순사전 만들기’는 생략하고 ‘용례사전 만들기’를 설명하도록 한다. ‘용례사전 만들기’의 화면은 다음과 같다.



우선 +를 클릭하여 용례사전을 만들 파일을 선택한다. 이 파일도 역시 유니코드의 txt 파일이어야 한다. 두 가지 방식 중에서 한 방식을 선택할 수 있다. 하나는 형태소를 분석해 주고 그 형태소에 따라 용례를 만들어 주는 것이다. 이것은 동음이의어의 용례를 한 형태소

로 처리할 위험을 덜기 위해서 취해진 방식이다. 또 하나는 어절단위 그대로 용례를 만들어 주는 것이다. 그리고 용례 출현한 모든 용례를 다 보기 위해서는 ‘무작위로 최대’를 선택하는 것이 좋다. ‘문맥제한’은 ‘사용자 입력’이란 내용을 클릭하면 ‘문장’ 단위를 화면의 ‘리턴 키’로 인식하는지, 아니면 ‘마침표’로 인식하는지를 결정하는 것이다. 현대국어 자료는 가능한 한 마침표로 인식하도록 정해 줄 필요가 있다. 왜냐 하면 현대국어 글들은 대개 문장부호를 사용하고 있기 때문이다. 그러나 옛문헌을 입력한 자료는 리턴키를 이용하여 문장 단위로 인식할 수 있도록 할 수 있을 것이다.

화면으로 보기를 선택하면 화면의 왼쪽에 표제항이 나오고 그 표제항을 클릭하면 하나하나 예문을 볼 수 있도록 하였다. 그러나 번잡스러우므로 파일로 저장하는 것이 좋다. 그 파일을 열어 보면 다음의 [표5-1]과 같이 되어 있을 것이다.

[표5-1] 형태소 분석을 하지 않고 어절별로 용례를 만든 것

가마니를 :	그 애가 한창 백양나무에 ___가마니를___ 돌려주고 있는데 새끼줄 통구리를 맨 사내애가 혈금씨금 달려왔다. 그들이 웃고 떠들며 ___가마니를___ 두른 백양나무에 새끼줄을 촌촌히 감아주었다.
가암소학교에 :	저 백양나무는 그들이 이곳 ___가암소학교에___ 붙던 날 심은 것이다.
가운데서 :	달구지에 실어온 나무 묘목 ___가운데서___ 제일 키가 크고 굽직한 백양을 골라 계집애는 물초롱을 기울이고 사내애는 흙을 북돋우면서 어서어서 자라기를 바라며 두 애가 한마음이 되었었다.
간밤 :	___간밤___ 꿈에 나무가 감기에 걸려 퐁퐁 기침을 짓는게 아니겠니.
감기에 :	간밤 꿈에 나무가 ___감기에___ 걸려 퐁퐁 기침을 짓는게 아니겠니.
감아주었다. :	그들이 웃고 떠들며 가마니를 두른 백양나무에 새끼줄을 촌촌히 ___감아주었다.___ 금시 잠퐁해진 듯한 눈바람을 자랑가로 들으면서 에어런 백양나무는 꿈속에 깃들고 있었다...
감추고 :	손칼은 계집애의 손아귀에서 꼬리를 ___감추고___ 있었다.
같은 :	겹질린 눈에서 이슬 ___같은___ 것이 구울다가 눈귀에 맺히고 있었다.
걸. :	그런데도 넌...봐, 상처 입은 백양나무가 몸부림치는 ___걸.___ 대들어 손칼을 빼앗으려던 사내애의 눈길에 백양나무쪽으로 날아갔다.
걸려 :	간밤 꿈에 나무가 감기에 ___걸려___ 퐁퐁 기침을 짓는게 아니겠니.
겹질린 :	___겹질린___ 눈에서 이슬 같은 것이 구울다가 눈귀에 맺히고 있었다.
것이 :	겹질린 눈에서 이슬 같은 ___것이___ 구울다가 눈귀에 맺히고 있었다.
것이다. :	계집애가 칼 쾅 그 애의 손을 때린 ___것이다.___ 얼른 칼을 주은 계집애는 내쳐 몸을 빼려 했다. 저 백양나무는 그들이 이곳 가암소학교에 붙던 날 심은 ___것이다.___ 달구지에 실어온 나무 묘목 가운데서 제일 키가 크고 굽직한 백양을 골라 계집애는 물초롱을 기울이고 사내애는 흙을 북돋우면서 어서어서 자라기를 바라며 두 애가 한마음이 되었었다.

만약에 ‘형태소 분석을 선택한 후에 보인 파일의 모습은 [표5-2]와 같을 것이다.

[표5-2]

하지만 / MAJ :  
 \_\_\_하지만\_\_\_ 사내애의 빠른 달음박질을 뒤떨굴 수 없었다.

는 / ETM :  
 <정문준,정든백양나무,1> 그 말에 분함을 \_\_\_깨무는\_\_\_ 계집애의 입술이 파르르 떨리고 있었다.  
 다시 나무에 칼을 대지 않고야... 당돌하게 \_\_\_내뱉는\_\_\_ 계집애의 고운 눈초리가 매섭게 치며졌다.  
 드세진 듯한 바람 속에 \_\_\_태질하는\_\_\_ 백양나무가 사내애의 눈을 팍 채워주고 있었다.

이 / VCP :  
 계집애가 칼 쥔 그 애의 손을 때린 \_\_\_것이다.\_\_\_ 얼른 칼을 주는 계집애는 내쳐 몸을 빼려 했다.  
 제 힘을 \_\_\_턱대고\_\_\_ 사내애는 주먹을 돌려매었다.  
 제 힘을 턱대고 사내애는 주먹을 \_\_\_돌려매었다.\_\_\_ 계집애는 뒤주춤했다.  
 넌 나와 함께 저 나무를 심었지 \_\_\_뉘야.\_\_\_ 그런데도 넌...봐, 상처 입은 백양나무가 몸부림치는 걸.  
 저 백양나무는 그들이 이곳 가암소학교에 불던 날 심은 \_\_\_것이다.\_\_\_ 달구지에 실어온 나무 묘목 가운데서 제일 키가 크고 굵직한 백양을 골라 계집애는 물초롱을 기울이고 사내애는 흙을 복돋우면서 어서어서 자라기를 바라며 두 애가 한마음이 되었었다.

달구지에 실어온 나무 묘목 가운데서 제일 키가 크고 굵직한 백양을 골라 계집애는 물초롱을 기울이고 사내애는 흙을 복돋우면서 어서어서 자라기를 바라며 두 애가 한마음이 \_\_\_되었었다.\_\_\_ <정문준,정든백양나무,2> 눈바람이 세차던 어떤 날 새벽, 버짚더미를 안고 계집애가 교정에 나타났었다.

간밤 꿈에 나무가 감기에 걸려 킁킁 기침을 \_\_\_짓는게\_\_\_ 아니겠니.  
 덜덜 떨고 있겠지 \_\_\_뉘야.\_\_\_ 난 병든 나무를 안고 온밤 울었댔어... 스키모자를 올리붙이며 사내애는 이마의 땀을 뺀 문질렀다.

난 병든 나무를 안고 온밤 울었댔어... 스키모자를 \_\_\_올리붙이며\_\_\_ 사내애는 이마의 땀을 뺀 문질렀다.  
 우리가 함께 심은 \_\_\_나무이니깐\_\_\_ 나무도 우리에게 똑같은 꿈을 준 게지 뭐.  
 우리가 함께 심은 나무이니깐 나무도 우리에게 똑같은 꿈을 준 \_\_\_게지\_\_\_ 뭐.  
 금시 \_\_\_잠퐁해진\_\_\_ 듯한 눈바람을 자장가로 들으면서 애어린 백양나무는 꿈속에 깃들고 있었다...

눈바람 / NNG :  
 <정문준,정든백양나무,2> \_\_\_눈바람이\_\_\_ 세차던 어떤 날 새벽, 버짚더미를 안고 계집애가 교정에 나타났었다.

금시 잠퐁해진 듯한 \_\_\_눈바람을\_\_\_ 자장가로 들으면서 애어린 백양나무는 꿈속에 깃들고 있었다...

애 / NNG :  
 계집애가 칼 쥔 그 \_\_\_애의\_\_\_ 손을 때린 것이다.  
 달구지에 실어온 나무 묘목 가운데서 제일 키가 크고 굵직한 백양을 골라 계집애는 물초롱을 기울이고 사내애는 흙을 복돋우면서 어서어서 자라기를 바라며 두 \_\_\_애가\_\_\_ 한마음이 되었었다.

그 \_\_\_애가\_\_\_ 한창 백양나무에 가마니를 돌려주고 있는데 새끼줄 통구리를 맨 사내애가 혈금씨금 달려왔다.

으로 / JKB :  
 그런 계집애를 더 골려줄 \_\_\_양으로\_\_\_ 다시 제 이름을 새기려 들던 사내애의 손에서 칼이 떨어졌다.  
 대들어 손칼을 빼앗으려던 사내애의 눈길이 \_\_\_백양나무쪽으로\_\_\_ 날아갔다.

이 용례 만들기 프로그램은 옛문헌을 입력한 자료에도 가능하다. 앞에서 제시했던 개수첩 해신어를 대상으로 하여 이 프로그램을 실행시킨 결과 파일을 [표5-3]으로 보이도록 한다.

[표5-3]

가 : 네 代官 中에 \_\_\_가\_\_\_ 내 전갈로 <改捷1,01b> 니르기를.

各各 : 그 외에 칭소흔 거시오니 <改捷1,04b> \_\_\_各各\_\_\_ 민망히 너기실가 근심흐옵닉.

거시오니 : 그 외에 칭소흔 \_\_\_거시오니\_\_\_ <改捷1,04b> 各各 민망히 너기실가 근심흐옵닉.

거술 : 그적기 여기 느려와 어제라도 오울 \_\_\_거술\_\_\_ 路次의 곧부매 <改捷1,02a> 계요 이제야 守門까지 완습닉.

계시면 : 안히 \_\_\_계시면\_\_\_ 오려 흐더라 니르고 <改捷1,02b> 오라.

계요 : 路次의 곧부매 <改捷1,02a> \_\_\_계요\_\_\_ 이제야 守門까지 완습닉.

끝에 : 代官 中도 <改捷1,03b> 흐 \_\_\_끝에\_\_\_ 일습닉.

그 : \_\_\_그\_\_\_ 외에 칭소흔 거시오니 <改捷1,04b> 各各 민망히 너기실가 근심흐옵닉.

그적기 : \_\_\_그적기\_\_\_ 여기 느려와 어제라도 오울 거술.

근심흐옵닉 : 그 외에 칭소흔 거시오니 <改捷1,04b> 各各 민망히 너기실가 \_\_\_근심흐옵닉\_\_\_ 萬事 | 두루 쓰리심을 <改捷1,05a> 미턴습닉.

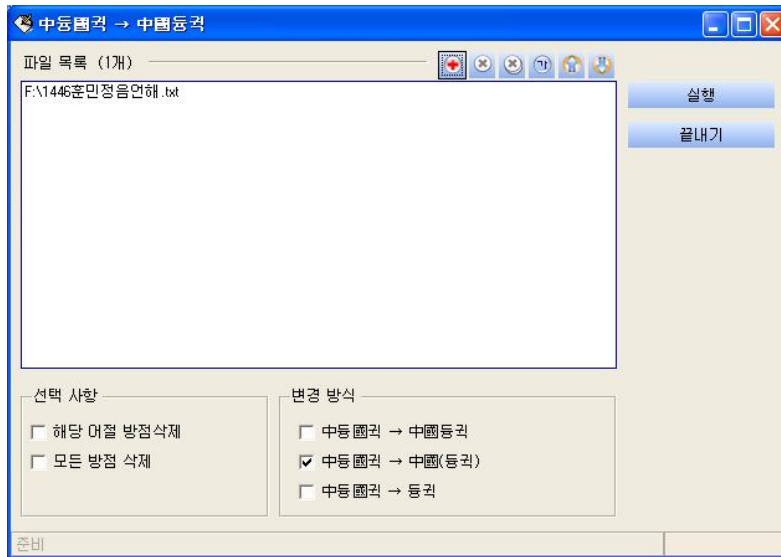
나논 : <改捷1,04a> \_\_\_나논\_\_\_ 소임으로 완습거니와 처음이옵고.

내 : 네 代官 中에 가 \_\_\_내\_\_\_ 전갈로 <改捷1,01b> 니르기를.

너기실가 : 그 외에 칭소흔 거시오니 <改捷1,04b> 各各 민망히 \_\_\_너기실가\_\_\_ 근심흐옵닉.

(3) ‘中등國帛 → 中國등帛’ 프로그램

이 프로그램은 옛문헌을 입력한 파일을 효율적으로 검색하기 위하여 만든 프로그램이다. 예컨대 한자음이 입력되어 있는 파일 중에서 한자음을 한자 바로 뒤에 입력한 것이 있으면, 자료 검색이 어려워진다. 예컨대 ‘世宗御製訓民正音’에서 ‘世宗’은 한자로 검색이 가능하지만 이것의 한자음을 기록한 ‘世宗宗宗御製訓民正音’처럼 입력한 자료는 ‘世宗’이 검색될 수 없다. 그래서 ‘世宗宗宗御製訓民正音’이란 문자열을 ‘世宗(성종)御製(영정)訓民(훈민)正音(정음)’이란 문자열로 바꾸어 주는 프로그램이다. 단, 한자 하나에 한자음이 있는 예들은 해당되지 않는다. 왜냐하면 한자 하나로 된 어휘는 검색이 가능하기 때문이다. ‘中등國帛 → 中國등帛’ 프로그램을 실현시켰을 대의 첫 화면은 다음과 같다.



이것을 실행시키기 위해 원문을 준비해 보도록 한다. 그 원문은 [표6-1]로 보인다.

[표6-1]

```
[head] 진언권공 [/head]
<1496진언권,1a> 진원권공
法界계를 조계 하는 眞진언언이라 세 닐굽 번 넓고 저스오라
<1496진언권,1b> 供공養양 받습는 眞진언언이라 세 번 넓고 저스오라
威위덕덕이 그지 업스시며 光광明明이 自自在在히시고 妙묘力력이 어디르샤 供공養양하습게 하는 眞진언언이
라 세 닐굽 번 넓고 저스오라 [gap reason='진언생략']
<1496진언권,2a> 供공養양을 내며 내는 眞진언언이라/ 세 닐굽 번 넓고 저스오라
<1496진언권,2b> 供공養양을 조계하는 眞진언언이라/ 세 닐굽 번 넓고 저스오라
너비 供공養양하습는 眞진언언이라/ 세 번 넓고 저스오라
<1496진언권,3a> 和화云운/ 모다 니르라
香향 휘워 供공養양하습노니 慈즈悲비를 브리디 가르샤 供공養양을 바드쇼셔
<1496진언권,3b> 흥 번 저스오라
불 혀 供공養양하습노니 慈즈悲비를 브리디 마르샤 供공養양을 바드쇼셔 [note] 흥 번 저스오라 [/note]
<1496진언권,4a> 仙선間간스 고즈로 供공養양하습노니 慈즈悲비를 브리디 마르샤 供공養양을 바드쇼셔
[note] 흥 번 저스오라 [/note]
<1496진언권,4b> 仙선間간스 菓과로 供공養양하습노니 慈즈悲비를 브리디 마르샤 供공養양을 바드쇼셔
[note] 흥 번 저스오라 [/note]
<1496진언권,5a> 仙선間간스 다로 供공養양하습노니 慈즈悲비를 브리디 마르샤 供공養양을 바드쇼셔 [note]
흥 번 저스오라 [/note]
<1496진언권,5b> 香香味미供공養양하습노니 慈즈悲비를 브리디 마르샤 供공養양을 바드쇼셔 [note] 흥 번
저스오라 [/note]
```

프로그램을 이용한 후의 결과 파일의 모습은 [표6-2]와 같다.

[표6-2]

<p>[head] 진언권공 [/head]          &lt;1496진언권,1a&gt; 진원권공          法界(법계)를 조케 하는 眞言(진언)이라 세 닐굽 번 넓고 저스오라          &lt;1496진언권,1b&gt; 供養(공양) 받좁는 眞言(진언)이라 세 번 넓고 저스오라          威德(위덕)이 그지 업스시며 光明(광명)이 自在(즈직)히시고 妙力(묘력)이 어디르샤 供養(공양)하습게 하는 眞言(진언)이라 세 닐굽 번 넓고 저스오라          &lt;1496진언권,2a&gt; 供養(공양)을 내며 내는 眞言(진언)이라/ 세 닐굽 번 넓고 저스오라          &lt;1496진언권,2b&gt; 供養(공양)을 조케하는 眞言(진언)이라/ 세 닐굽 번 넓고 저스오라          너비 供養(공양)하습는 眞言(진언)이라/ 세 번 넓고 저스오라          &lt;1496진언권,3a&gt; 和云(화운)/ 모다 니르라          香幢(향뿔) 위워 供養(공양)하습노니 慈悲(즈비)를 브리디 가르샤 供養(공양)을 바드쇼셔          &lt;1496진언권,3b&gt; 혼 번 저스오라          불 혀 供養(공양)하습노니 慈悲(즈비)를 브리디 마르샤 供養(공양)을 바드쇼셔 [note] 혼 번 저스오라 [/note]          &lt;1496진언권,4a&gt; 仙間(선간)샤 고즈로 供養(공양)하습노니 慈悲(즈비)를 브리디 마르샤 供養(공양)을 바드쇼셔 [note] 혼 번 저스오라 [/note]          &lt;1496진언권,4b&gt; 仙間(선간)샤 菓과로 供養(공양)하습노니 慈悲(즈비)를 브리디 마르샤 供養(공양)을 바드쇼셔 [note] 혼 번 저스오라 [/note]          &lt;1496진언권,5a&gt; 仙間(선간)샤 다로 供養(공양)하습노니 慈悲(즈비)를 브리디 마르샤 供養(공양)을 바드쇼셔 [note] 혼 번 저스오라 [/note]          &lt;1496진언권,5b&gt; 香味(향미)供養(공양)하습노니 慈悲(즈비)를 브리디 마르샤 供養(공양)을 바드쇼셔 [note] 혼 번 저스오라 [/note]</p>
---

## 5. 감작새(SynKDP)

감작새(SynKDP)란 전주대학교 소강춘 교수와 그 제자(김진규, 박진양)들에 의해 개발된 프로그램이다. 통합형 한글 데이터 처리기인 SynKDP(Synthesized Korean Data Processor)를 일명 ‘감작새’라 부른다. ‘감작할 새에 자료를 처리해 준다’는 뜻으로 붙인 이름이다. 감작새는 구축된 말뭉치에서 필요한 언어 정보를 빠르고 정확하게 추출하기 위하여 개발되었다. ‘윈도 2000’을 기본 운영체제로 ‘윈도 98’과 ‘윈도 XP’에서도 실행 가능하며 한글97의 2바이트 파일(2b 파일)만을 처리한다.

이 프로그램은 한국학중앙연구원의 이근식 선생이 개발한 Morph 프로그램을 DOS 상태가 아닌 윈도상에서 보완하여 실현시킨 프로그램으로 지금까지 나온 국어자료처리 프로그램 중 가장 훌륭한 것으로 평가되고 있다.

### (1) 감작새의 기능

이 프로그램에서 실행할 수 있는 기능은 매우 다양하지만, 주된 기능은 대체로 다음과 같은 것이다.

- ① 빈도 분석 및 조사 : 어절 빈도, 음절 빈도, 음소 빈도
- ② 문장의 평균 어절(음절, 음소) 개수
- ③ 어절 색인
- ④ 용례 만들기
- ⑤ 검색 (KWIC 방식과 KWOC 방식)
- ⑥ 정렬

## (2) '깜짝새' 이용방법

프로그램을 이용하기 위해서는 몇 가지 전제 작업이 필요하다. 하나는 프로그램을 장착시킬 때에 반드시 '옛한글 입력기'를 설치하여야 하며, 옛한글을 사용하기 위해서는 한글 서체 중 새굴림체를 설치하여야 한다. 이 프로그램들을 설치하여도 검색할 말뭉치가 필요하다.

### (가) 말뭉치의 구조 및 종류

① 말뭉치 파일은 2 바이트 파일로 되어 있어야 한다.

② 출전은 < >로 표시되어 있어야 하며, 그 출전은 그 페이지의 본문이 시작되는 앞에 표시되어 있어야 한다.

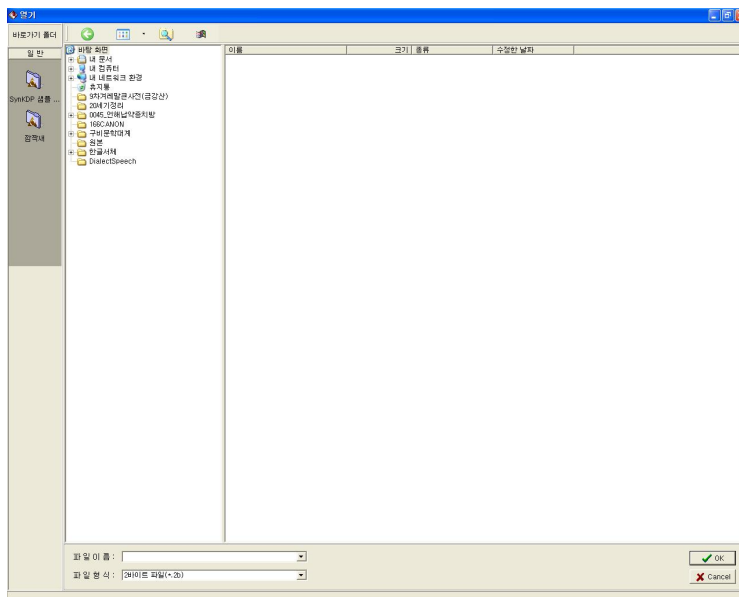
③ 말뭉치 파일은 성격이 다른 말뭉치이면 각각 달리 만들어 두어도 좋다. 예컨대 '현대소설' 말뭉치와 '고소설' 말뭉치를 따로 만들어 두어도, 이를 합쳐서 검색할 수도 있고, 따로 따로 검색할 수도 있다. 이용자의 편리성에 따라 만들어 두는 편이 좋다. 발표자는 세기별로 말뭉치를 만들어 사용하고, 이에 다시 '신소설', '독립신문'과 '제국신문'등을 합쳐 놓은 신문 말뭉치를 따로 이용하고 있다.

이 프로그램은 세종 말뭉치 태그를 인식하지 못하므로 세종 말뭉치를 그대로 사용할 경우에는 부정확한 결과를 제시할 수도 있으니 주의를 요한다.

### (나) '깜짝새'의 실행

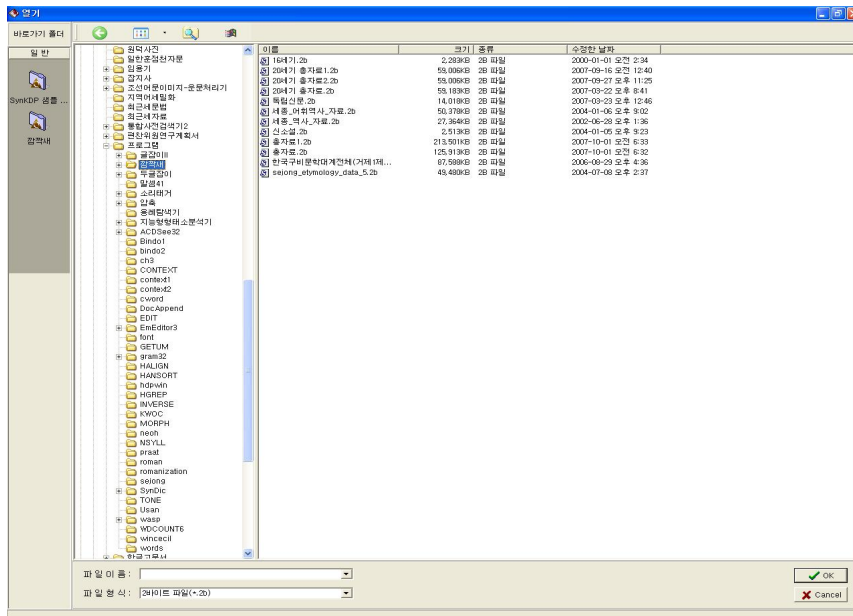
'깜짝새'를 설치하면 컴퓨터의 바탕화면에 다음과 같은 아이콘이 생길 것이다.

이 아이콘을 클릭하면 다음과 같은 화면이 나타날 것이다.

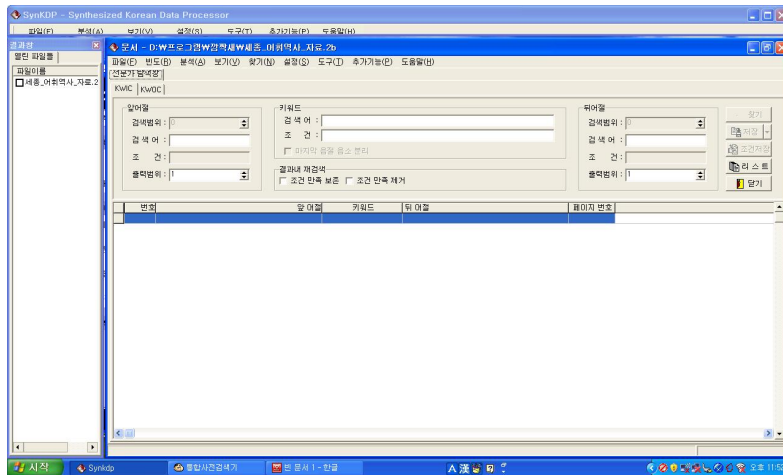


그러면 왼쪽의 상자 안에서 ‘내 컴퓨터’를 클릭하여 앞에서 말한 말뭉치 파일(즉 2바이트 파일)이 있는 서브디렉토리를 지정해 준다. 그러면 그 서브디렉토리에 있는 2 바이트 파일을 찾아서 오른쪽 상자 안에 그 파일 이름이 자동적으로 보일 것이다. 이 프로그램에는 ‘세종 어휘 역사 자료.2b’의 이름을 가진 말뭉치가 기본적으로 제공되어서, 말뭉치를 만들지 않아도 ‘세종 어휘 역사 자료.2b’의 파일이 보일 것이다. 이 말뭉치는 정제되어 있지 않아서 여러 가지 문제가 있으나, 지금까지 구축된 자료를 상당수 모아 놓은 것이라는 점에서 그런 대로 활용할 만하다. 그러나 이 말뭉치는 15세기부터 19세기말까지의 자료는 충분한 반면에, 20세기 이후의 말뭉치는 적은 편이기 때문에 현대국어 연구 자료로 활용할 때에는 보완이 필요하다.

우선 말뭉치를 선택한 화면을 보이면 다음과 같다.



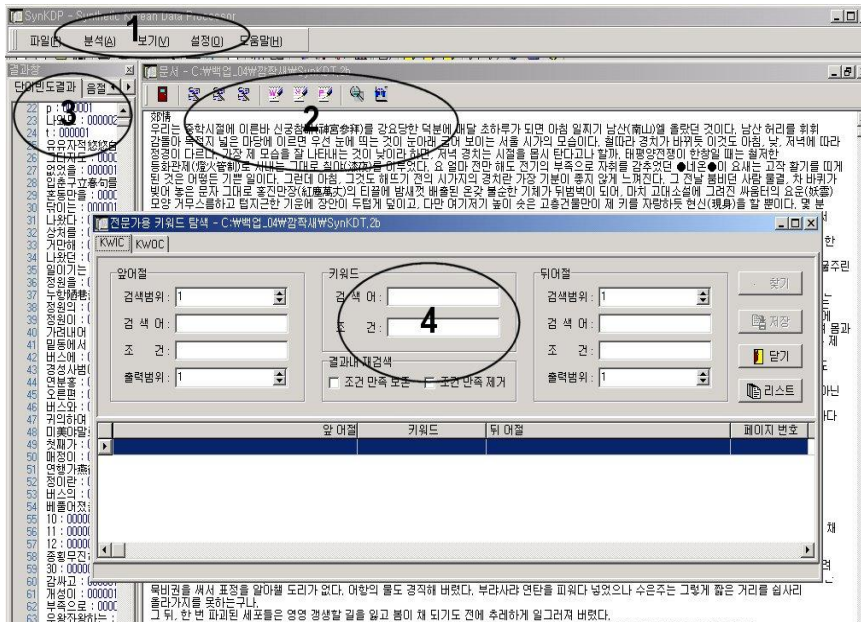
이 중에서 활용하고 싶은 말뭉치를 클릭한다. 그러면 다음 그림과 같이 왼쪽 상단에 열린 파일들 아래에 그 파일 이름이 나타난다. 그 파일을 클릭하면 이제부터 이 프로그램을 활용할 준비가 된 셈이다.



위의 화면은 여러 가지 화면이 복합적으로 구성된 것이다. 이 프로그램의 화면은 메뉴창, 문서창, 결과창, 탐색창의 4 가지 화면으로 구성되어 있다.

- ① 메뉴창 : 화면의 최상위에 위치해 있는 창
- ② 문서창 : 열린open 파일의 내용을 보여주는 창
- ③ 결과창 : 열린 파일의 경로와 파일명 그리고 처리된 빈도결과를 보여줄 창
- ④ 탐색창 : KWIC와 KWOC로 검색할 수 있고 그 LIST를 저장 사용할 수 있는 창

이를 그림으로 보이면 다음과 같다.



이 중에서 우리가 가장 많이 사용하는 부분은 ④ 탐색창이라고 할 수 있다. 왜냐 하면 이 프로그램 중에서 가장 중요한 기능은 '검색'이라고 할 수 있기 때문이다.



이 자료를 마우스로 지정하여 복사한 후에 문서작성기에 그대로 옮길 수도 있다. 그러나 이보다도 ‘어절 관련 저장’등에서 파일로 저장하는 방식을 택하는 것이 좋다. 그 부분을 클릭하면 다음과 같은 창이 화면에 뜨게 될 것이다.



‘순서’와 ‘비교방향’에서 원하는 방법을 클릭한 후 ‘저장’을 누르면 그 파일을 저장할 서브 디렉토리를 묻는 창이 나올 것이며, 그것을 선택하고 저장할 파일 이름을 타자하고 ‘저장’을 클릭하면 작업이 모두 끝난다. 그 파일은 2 바이트 파일로 저장된 것이므로, 반드시 2바이트 파일로 불러 와야 활용할 수 있음은 당연한 것이다.

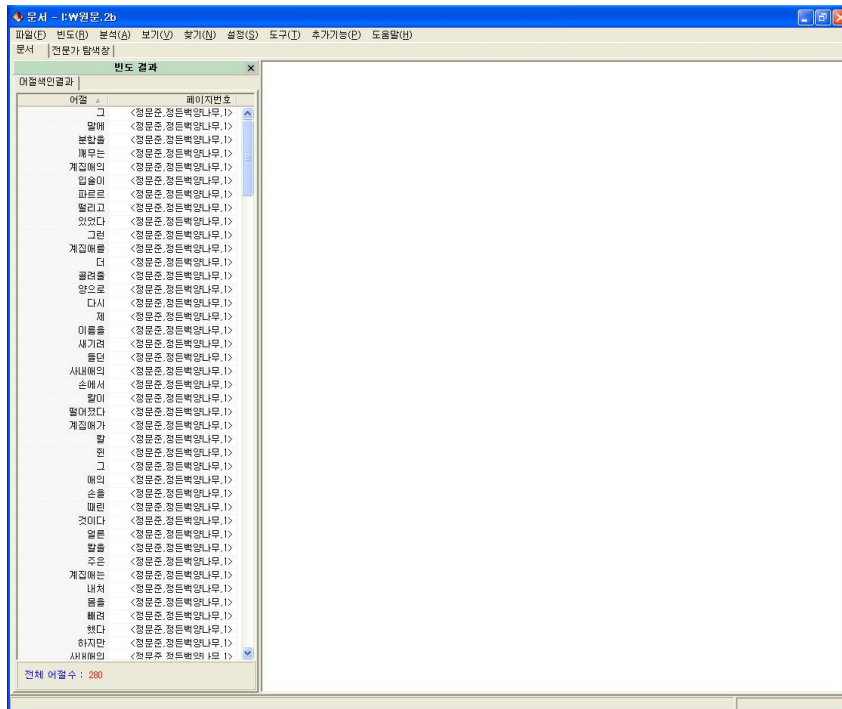
다른 파일들을 다시 분석하려면 왼쪽 ‘결과창’의 상단에 보이는 ‘열린 파일들’을 클릭하면 지금까지 조사한 결과물은 보이지 않고 파일 이름들이 다시 열거될 것이다. 그 파일을 선택한 후 같은 작업을 하면 된다.

그러나 초보자들이 조심하여야 할 것이 있다. 프로그램에서 기본적으로 제공되는 파일인 ‘세종 어휘 역사 자료.2b’를 선택하여 빈도 조사는 실험하지 않는 것이 좋다. 왜냐 하면 이 파일은 그 크기가 엄청나서 그것을 분석하는데, 많은 시간이 소요되기 때문이다. 그래서 이 자료를 선택하여 실행해 본 초보자는 이 프로그램이 무슨 ‘깜짝새’냐고 묻는 사람도 있었다.

#### (4) 어절 색인 만들기

‘어절 색인 만들기’는 ‘전문가 탐색창’을 이용한다.

- ① 어절 색인을 할 원문 파일(2바이트로 된 파일)을 왼쪽 결과창에서 클릭을 한다.
- ② 전문가 탐색창에서 위에 있는 ‘도구’를 클릭한다.
- ③ 그러면 ‘어절색인, 용례사전, 어절 리스트 저장’이라는 메뉴가 보일 것이다.
- ④ 이 중에서 ‘어절 색인’을 클릭한다.
- ⑤ 그러면 ‘현재창에서 보기, 한글에서 보기, 저장하기’의 메뉴가 보일 것이다.
- ⑥ ‘현재창에서 보기’를 선택하면 다음과 같은 화면이 순식간에 나타날 것이다.



여기에서 ‘어절’을 클릭하면 ‘가나다’순으로 정렬이 될 것이며 ‘페이지 번호’를 클릭하면 출전 순서로 정렬이 될 것이다.

‘한글’에서 보기를 선택하려면 ‘한글 97’이 기본적으로 설치되어 있어야 한다. 그리고 ‘저장하기’를 선택하면 역시 서브디렉토리와 저장할 파일 이름을 선택한 후 저장하여 문서작성기로 불러 오면 된다.

### (5) 용례사전 만들기

‘용례사전 만들기’도 역시 ‘전문가 탐색창’을 이용한다.

- ① 용례사전을 만들 원문 파일(2바이트로 된 파일)을 왼쪽 결과창에서 클릭을 한다.
- ② 전문가 탐색창에서 위에 있는 ‘도구’를 클릭한다.
- ③ 그러면 ‘어절색인, 용례사전, 어절 리스트 저장’이라는 메뉴가 보일 것이다.
- ④ 이 중에서 ‘용례사전’을 클릭한다.
- ⑤ 그러면 ‘어절별 용례사전 만들기’란 메뉴가 보일 것이다.
- ⑥ 서브디렉토리와 저장할 파일 이름을 물으면 그것을 선택한 후 저장하여 문서작성기로 불러 오면 된다. 그 결과를 보이면 다음과 같을 것이다.

그 : 그 말에 분함을 깨우는 계집애의 입술이 파르르 떨리고 있었다. <정문준,정든백양나무,1>  
 그 : 그런 계집애를 더 골려줄 양으로 다시 제 이름을 새기려 들던 사내애의 손에서 칼이 떨어졌다. <정문준,정든백양나무,1>  
 그 : 계집애가 칼 권 그 애의 손을 때린 것이다. <정문준,정든백양나무,1>  
 그 : 그 칼 내놓지 못하겠니? <정문준,정든백양나무,1>  
 그 : 이 주먹맛 보고 싶어 그래? <정문준,정든백양나무,1>  
 그 : 그런데도 넌...봐, 상처 입은 백양나무가 몸부림치는 걸. <정문준,정든백양나무,1>  
 그 : 저 백양나무는 그들이 이곳 가압소학교에 붙던 날 심은 것이다. <정문준,정든백양나무,1>  
 그 : 그 애가 한창 백양나무에 가마니를 둘러주고 있는데 새끼줄 통구리를 뱀 사내애가 혈금씨금 달려왔다. <정문준,정든백양나무,2>  
 그 : 그들이 웃고 떠들며 가마니를 두른 백양나무에 새끼줄을 촘촘히 감아주었다. <정문준,정든백양나무,2>  
 말애 : 그 말에 분함을 깨우는 계집애의 입술이 파르르 떨리고 있었다. <정문준,정든백양나무,1>  
 분함을 : 그 말에 분함을 깨우는 계집애의 입술이 파르르 떨리고 있었다. <정문준,정든백양나무,1>  
 깨우는 : 그 말에 분함을 깨우는 계집애의 입술이 파르르 떨리고 있었다. <정문준,정든백양나무,1>  
 계집애의 : 그 말에 분함을 깨우는 계집애의 입술이 파르르 떨리고 있었다. <정문준,정든백양나무,1>  
 계집애의 : 당돌하게 내뱉는 계집애의 고운 눈초리가 매섭게 치며졌다. <정문준,정든백양나무,1>  
 계집애의 : 손같은 계집애의 손아귀에서 꼬리를 감추고 있었다. <정문준,정든백양나무,1>  
 <이하 생략>

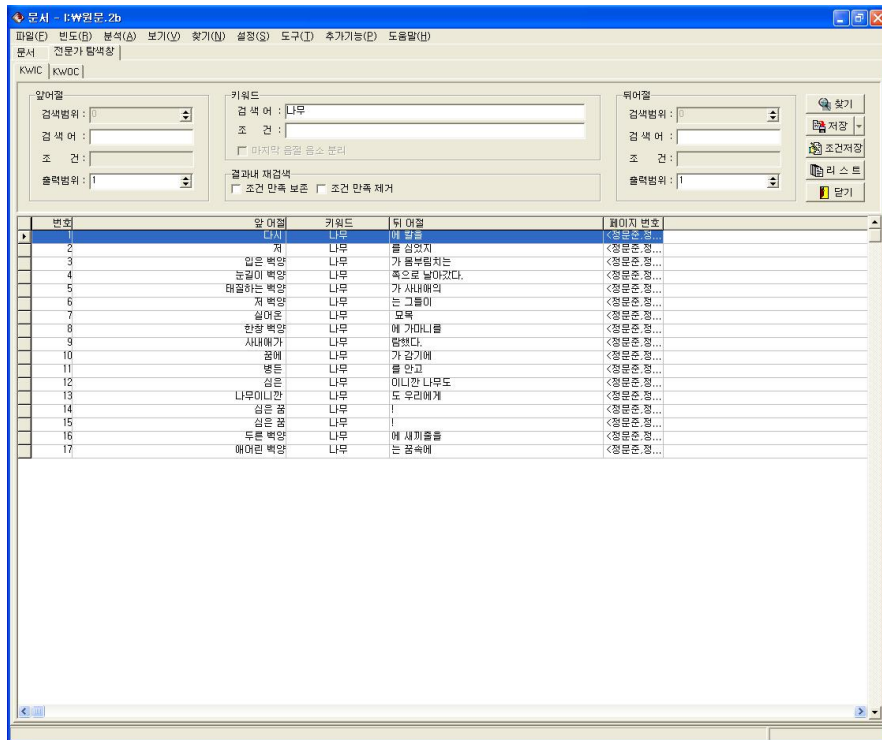
전문가 탐색창에는 많은 메뉴가 보이는데, 이중에 ‘도구’의 메뉴를 클릭하면

## (6) 검색

이 프로그램의 진수는 ‘검색’에 있다고 해도 과언이 아니다. 검색을 위해서는 ‘전문가 탐색창’을 이용하여야 한다.

‘전문가 탐색창’ 바로 아래에는 ‘KWIC’와 ‘KWOC’로 검색할 수 있는 내용이 있다. 그리고 그 아래에는 ‘앞어절, 키워드, 뒤어절’의 세 부분으로 구성되어 있다. 그리고 각 어절의 아랫 부분에는 ‘검색어’와 ‘출력범위’ 등이 있고, 키워드에는 ‘검색어’와 ‘조건’이 들어 있다. 이용 방법을 보이면 다음과 같다.

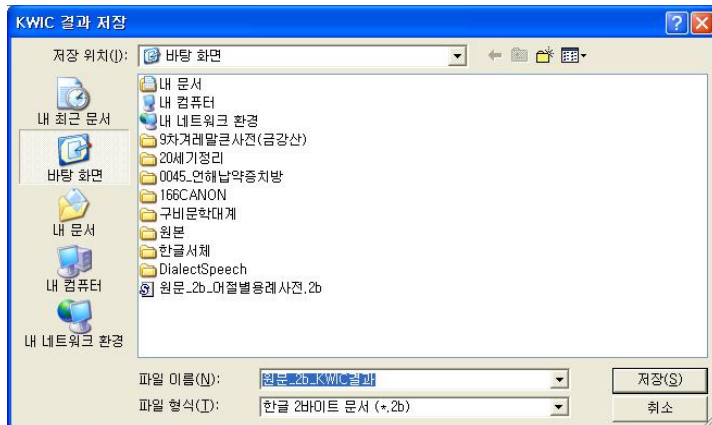
- ① ‘검색어’를 타자한다. 예를 들어 ‘나무’라고 한다.
- ② 오른쪽에 있는 ‘찾기’를 클릭하거나 ‘엔터키’를 누른다.
- ③ 화면에 검색하고자 하는 형태를 키워드로 하여 앞뒤에 어절들이 배열된 모습을 볼 수 있을 것이다.



그리고 '총 몇 개를 찾았습니다'란 정보가 나타날 것이다. 그 용례의 숫자를 제시한 것이다.



④ 오른쪽에 있는 '저장'을 클릭하면 역시 서브디렉토리와 저장할 파일 이름을 타자하라는 지시가 보일 것이다. 그것을 따라 저장한 후에 문서작성기로 읽어 오면 될 것이다.



이 프로그램 사용에는 여러 가지 조건이 있다.

① 옛한글 입력

옛한글을 입력하려면 'Alt + I'를 누르면 옛한글 입력기가 나타난다. 거기에서 입력하면 된다.

② 앞 뒤 어절 수 조절

키워드를 중심으로 앞뒤의 어절을 몇 개까지 검색할 것인가에 따라앞어절과 뒤어절의 '출력범위'를 지정해 주는 기능이다.

③ 조건 선택

이 조건 선택은 매우 중요하다. 어떤 음절의 '초성, 중성, 종성'만 검색할 수 있을 뿐만 아니라, '초성 + 중성, 초성 + 종성, 중성 + 종성', 그리고 '음절'과 '아무개'(즉 조건이 없는 것)를 선택할 수 있도록 하는 것이다. 이러한 내용은 '조건'에 마우스를 갖다 대고 마우스의 오른쪽을 누르면 그 내용들을 볼 수 있다. 그것을 숫자로 표시하였는데 그것을 보이면 다음과 같다.

- 초성 : 1
- 중성 : 2
- 종성 : 3
- 초성 + 중성 : 4
- 초성 + 종성 : 5
- 중성 + 종성 : 6
- 음절 : 7
- 아무개 : 8

예를 들어서 말뭉치 중에서 'ㅁ'이 나타나는 용례를 뽑으려면 검색어에서 'ㅁ'(물론 'ㅍ, ㅂ' 등을 입력하여도 마찬가지이다)를 입력하고 '조건'에서 '1'을 타자하거나 마우스의 오른쪽을 눌러 '초성'을 선택하면 된다. 그러면 'ㅁ' 초성을 가진 모든 용례들이 출전과 함께 보일 것이다.

'i 옴라우트'를 보이는 자료를 검색하려면 두 음절을 가진 단어를 타자하고 '조건'에서 첫 음절에서는 '애'를 보일 수 있는 음절을, 그리고 뒤음절에서는 '이' 모음을 보이는 음절(예컨대 '애기')을 타자한 후에 조건에서 중성, 중성 즉 22를 타자해 놓고 검색하면 된다. 이 경우에는 앞어절 출력범위와 뒤어절 출력범위를 각각 1로 해 놓아야 좋다.

세종 어휘 역사 자료 중에서 이렇게 검색한 자료를 보면 'i 옴라우트'에 해당되지 않는 자료도 부지기수로 나오지만, 그 중에서 이에 해당되는 항목만을 남기고 모두 삭제해 놓으면 검색자가 원하는 자료를 얻을 수 있다. 그래서 이렇게 얻은 자료로는 '댕기던, 애비의, (비가) 내리고, 월급쟁이, 붙잡힌' 등등의 자료를 볼 수 있을 것이다.

그리고 앞어절, 키워드 뒤어절, 페이지 번호 등을 클릭하면 그 자료들을 정렬할 수 있을 것이다.

이 프로그램에 대한 자세한 사용법은 이 프로그램 속에 들어 있는 '소강춘교수논문.hwp' 파일을 열어서 이용할 수 있을 것이다.

### III. 맺는 말

국어 연구를 위해 개발해 놓은 프로그램은 여러 가지가 있다. 여기에 소개하지 못한 프로그램 중에 ‘글잡이’나 ‘지능형 형태소 분석기’(21세기 세종 계획 개발), 그리고 ‘역사 자료 형태소 분석기’(연세대, 김상태, 황용주 개발), 방언지도 제작 프로그램(이상규 개발), 방언지도 해석기(북한 사회과학원 언어학연구소 개발) 등은 시간과 지면의 제약으로 그 설명을 하지 못했다. 국어 자료 처리 프로그램들은 아직 그 사용법을 구체적으로 제시하지 않아서, 사용자마다 각자 이용하고 있는 형편이다. 각자 어떻게 활용하고 있는지에 대한 대답의 기회라도 가져 볼 필요가 있으나, 어떤 학자는 이 방법을 자기만의 방법으로 알고서 공개하지 않는 경우가 많은 것으로 보인다. 특히 각 대학의 연구소에서 사용하는 것들이 그렇다. 그러나 이것은 옳은 방법이 아니다. 학문은 혼자 하는 것이 아니기 때문이다. 이론은 독자적일지 모르나 자료는 만인이 공유하는 것이다. 그 자료들을 어떻게 실험하는가 하는 방법의 문제도 공유하지 않으면 안 되기 때문이다.

그러나 마지막으로 남겨 둘 말이 있다. 컴퓨터는 앞으로 국어의 여러 현상을 찾아 주는데는 많은 공헌을 할 것이지만, 그 자체가 설명해 주지는 않는다, 따라서 자료를 보고 고민하고 해석하고 설명하는 몫은 컴퓨터의 몫이 아니라 연구자의 몫이다. 아무리 모음체계나 자음체계 등을 결정해 주지만, 어디까지나 학문은 사람이 하는 것이다. 사람은 컴퓨터로 더 나은 학문의 환경을 만들어 갈 뿐이다.

A Method of Korean Language Data Process using a Computer

Hong Yun Pyo

In digital age, Korean linguist must get full mileage out of the computer program. What ultimately determines the success or failure of the Korean language study is how well it is computing searching a Korean database. In actual fact, as soon as we get the hang of this computer, we will be able to study faster. So I introduce new ways to use the computer for Korean language data process into the Korean linguists. Especially I introduce them how to use the computer program stage by stage.