

2021년도 21세기 신규장각 자료구축사업 연구용역 지원신청서

구 분	접수번호	※	과제번호	※
	과제구분	자유과제		
연구과제명	국문	규장각 소장 조선시대 중국어 역학서의 언어 정보 시멘틱 데이터 구축		
	영문	*		
연구책임자	성명		전공(학위)	
	소속		직급	
	이메일		전화	
연구참여자	공동연구원(연구책임자 제외)		2명	연구보조원 1명
총 연구기간	2021년 계약체결일 ~ 2022년 1월 15일			
연구비신청액	오천만원(일금 ₩50,000,000)			
<p>본인(등)은 귀 연구원의 2021년도 21세기 신규장각 자료구축사업에 의한 연구비를 지원받고자 소정의 서류를 첨부하여 신청서를 제출합니다.</p> <p style="text-align: center;">2021년 3월 25일</p> <p>연구책임자 소속 _____ 성명 (인)</p> <p>공동연구자 소속 _____ 성명 (인)</p> <p>공동연구자 소속 _____ 성명 (인)</p> <p style="text-align: center; font-size: 1.2em; font-weight: bold;">서울대학교 규장각한국학연구원 귀중</p>				

◎ 주의

※표시는 기입하지 말 것.

* 연구내용을 별도로 영문으로 발표할 경우만 기재

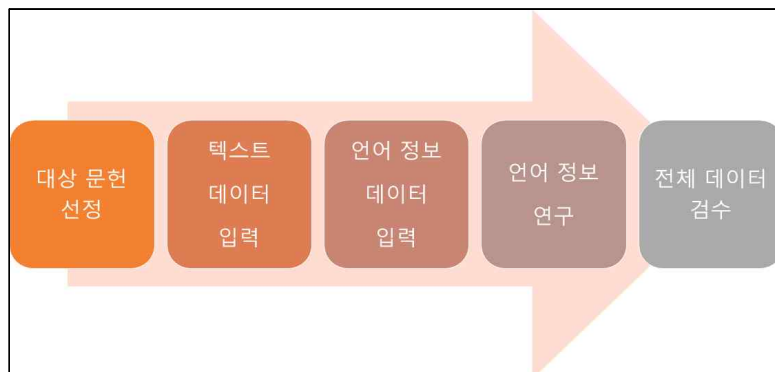
연구계획서 요약

연구과제명	(국문) <규장각 소장 조선시대 중국어 역학서의 언어 정보 시멘틱 데이터 구축>
	(영문) <A Construction of Semantic Data on Linguistic information in Chinese-Korean translation books of Chosun Dynasty in Kyujanggak>

◆ 계획 내용 요약

조선시대 중국어 역학서는 당시 음운 및 어휘 변화, 사회·문화상과 지리학적 정보를 알 수 있는 寶庫라고 할 수 있다. 각 문헌의 원본 혹은 해제가 저서로 출간되었고, 이에 대한 공시적, 통시적 어학 연구의 성과도 상당히 축적되어 있다. 그러나 그 연구 방법과 환경에 있어 우리나라는 아직 초보적인 단계에 머물러 있다. 규장각이나 국립중앙도서관과 같은 사이트에서는 원본의 스캔본만을 열람할 수 있고, <조선시대 외국어 학습서 DB> 웹에서는 2차 데이터 가공이 된 언해문을 확인할 수 있으나 특정 글자나 단어를 기준으로 한 정보는 얻을 수가 없어 검색의 효율성이 떨어진다.

이와 같은 상황에서 본 연구는 고도화된 데이터 가공 기술을 통해 양질의 언어 정보를 제공하는 <조선시대 중국어 역학서 언어 정보의 시멘틱 데이터>를 구축하고자 한다. 이는 원본과 번역문, 원문을 동시에 열람할 수 있도록 하여 규장각 웹사이트에서 아직 제공하지 않고 있는 ‘이미지+번역문’, ‘이미지+원문’ 부분에 대해 즉각적인 서비스를 제공할 수 있게 할 뿐 아니라, 낱글자 어음과 각 단어의 품사별, 주제별 분류 및 태깅을 통해 기존의 국내외 사이트의 한계점을 뛰어넘는 종합적이고 고도화된 중국어 역학서 언어 정보 데이터베이스로 역할 할 수 있을 것이다. 본 연구가 계획하고 있는 시멘틱 데이터 구축 과정은 다음과 같이 요약할 수 있다.



◆ 예상 연구 결과 정보

본 연구는 중국어 역학서의 번역문, 원문 및 어음과 어휘 정보를 XML형식의 기계가독형 데이터로 제출할 것이다. 먼저 번역문, 원문의 데이터는 현재 규장각 원문검색 DB의 ‘원문보기’의 ‘이미지+번역문’과 ‘이미지+원문’의 탭에 연계시켜 바로 대중들에게 서비스를 제공할 수 있다. 어음과 어휘 정보 데이터의 경우 현 규장각 웹사이트에서 ‘중국어 역학

서 언어 정보 아카이브'와 같은 독립 영역을 구축하여 여기에 데이터를 연계시켜 서비스 제공이 가능하다.

이와 같이 일신(一新)된 규장각 웹사이트는 국내외 모든 사람에게 조선시대 역학서 및 중국어에 대한 지식을 편리하게 제공하는 **고문헌과 언어 정보의 종합적 허브**로 자리매김하게 될 것이다. 이는 어학, 사회학, 민속학 등에 다양한 연구 분야에 대한 새로운 파장을 불러일으킬 것이며 교육 분야에 있어 새로운 메카로서 역할 할 것이다.

[규장각] 조선시대 중국어 역학서 언어 정보 시멘틱 데이터
<ul style="list-style-type: none"> ◦ 원본과 번역문, 원문의 동시 열람 ◦ 해당 문헌의 번역문, 원문의 검색 ◦ 해당 문헌의 어음, 단어 검색 ◦ 검색 부분과 원문의 연계(하이퍼링크) ◦ 해당 글자, 단어의 자유로운 재가공

(※텍스트 데이터 파일 샘플은 계획서 9~10페이지에 첨부)

연구 계획서

- 연구과제명 : 규장각 소장 조선시대 중국어 역학서의 언어 정보 시멘틱 데이터 구축

1. 연구의 목적과 필요성

동아시아 여러 나라는 오랜 기간 한자 문화와 유교, 불교문화를 공유하면서도 각 국가의 고유한 전통을 발전시켰다. 동아시아 여러 나라 중에서도 특히 우리나라와 중국은 가장 정치, 문화적으로 긴밀한 관계를 유지해왔다. 중국의 영향을 ‘수용’하는 위치에 있었던 우리나라는 1443년 한글의 창제와 더불어 한자의 불완전한 표음 기능을 극복하고, 고유의 문자를 바탕으로 한 문화를 형성하기 위한 첫 발걸음을 내디뎠다. 그러나 중국과의 외교, 경제적인 요구로 인한 중국어 학습의 필요성은 늘 강하게 있어 왔으며, 이는 사역원에서의 다양한 중국어 교재의 편찬이라는 결과를 낳게 되었다. 이 교재에 쓰인 ‘한글’은 한국어와 중국어 간 소통의 벽을 넘어 정확한 중국어를 발음하고 중국어 단어, 문장, 문화와 지리를 알 수 있는 매개체가 된 것이다. 사역원에서 편찬된 다양한 중국어 역학서의 언해본은 다음과 같은 당시의 어음, 어휘, 문화적 정보를 담고 있다.

(1)

문헌 예자	『翻譯老乞大』	『翻譯朴通事』	『老乞大諺解』	『朴通事諺解』	『朴通事新釋諺解』	『重刊老乞大諺解』	『華音啓蒙諺解』
學	(左音)향 (右音)효	(左音)향 (右音)효	(左音)향 (右音)효	(左音)향 (右音)효	(左音)향 (右音)효	(左音)향 (右音)효	쑤

(2)

문헌 예자	『翻譯老乞大』	『翻譯朴通事』	『老乞大諺解』	『朴通事諺解』	『朴通事新釋諺解』	『重刊老乞大諺解』	『華音啓蒙諺解』
點心	-	-	-	-	點心	점심	菓子

(3) 順天府官: 順天府官과, 『朴通事諺解』

順天府官員與欽天監衆官們: 順天府官員과 다뭇 欽天監 모든 관원들이, 『朴通事新釋諺解』

順天府: 北京, 『方言集釋』

(1)은 각 문헌에서 ‘學’ 발음의 변화과정을 보여준다. 이 글자는 匣母字로 『華音啓蒙諺解』 이후로 구개음화된 것을 알 수 있다. (2)는 ‘點心’의 언해나 번역문을 비교한 것으로, 『朴通事新釋諺解』와 『重刊老乞大諺解』에서는 우리말에서 쓰이는 뜻과 같이 ‘끼니’인 ‘점심’의 뜻으로 쓰이다가, 이후의 책에서는 ‘간식’으로 번역되어 어휘 의미의 변화를 보여준다. (3)에

나온 ‘順天府’는 北京으로, 北京은 明清시대에 府制상의 명칭이다. 조선시대 중국어 역학서에서는 공통적으로 北京을 ‘順天府’으로 흔히 칭하고 있다.

이와 같이 조선시대 중국어 역학서는 당시 음운 및 어휘 변화, 사회·문화상과 지리학적 정보를 알 수 있는 寶庫라고 할 수 있다. 각 문헌의 원본 혹은 해제가 저서로 출간되었고,¹⁾ 이에 대한 공시적, 통시적 어학 연구의 성과도 상당히 축적되어 있다.²⁾ 그러나 그 연구 방법과 환경에 있어 우리나라는 아직 매우 초보적인 단계에 머물러 있다. 규장각이나 국립중앙도서관과 같은 사이트에서는 원본의 스캔본만을 열람할 수 있으며, <조선시대 외국어 학습서 DB>³⁾웹에서 2차 데이터 가공이 된 언해문을 확인할 수 있지만, 특정 글자나 단어를 기준으로 한 정보는 얻을 수가 없어 검색의 효율성이 떨어진다. 중국 百度 웹에서 『노걸대』나 『박통사』의 원본을 디지털라이징하고,⁴⁾ 중국의 <東方言語學>(www.eastling.org)나 대만의 <小學堂>(xiaoxue.iis.sinica.edu.tw)에서 각 글자에 대한 문헌별, 시대별, 글자별 음운 정보를 집대성한 것을 보면 국내에서는 이 분야에 대한 관심이 얼마나 소홀했는지를 알 수 있다. 또한, <小學堂>이 대만 정부 산하의 연구 기관인 ‘中央研究院’에 의해 구축된 음운 정보 종합 데이터베이스임을 감안할 때, 우리나라도 역시 ‘규장각한국학연구원’과 같은 국내 대표 연구 기관이 선두 주자가 되어 국내 언어 자료 데이터베이스를 개발해야 할 필요성을 보여준다.

따라서 본 연구는 먼저 기존의 규장각 웹사이트의 원문보기에서 아직 제공하지 않고 있는 ‘이미지+번역문’, ‘이미지+원문’ 부분에 대한 데이터베이스를 구축하여 즉각적인 서비스를 제공할 수 있게 할 것이다. 여기서 더 나아가 각 글자에 대한 어음, 어휘 및 이에 대한 품사, 주제별 분류에 대한 정보를 정리하여 종합적인 ‘중국어 역학서 언어 정보 시멘틱 데이터’를 구축하여 향후 규장각 웹사이트에서 서비스를 제공할 수 있는 자료를 제공할 것이다.

본 연구는 다음과 같은 목적을 갖는다. 첫째, 기존 규장각한국학연구원 웹사이트의 데이터를 보충하고, 언어 정보에 대한 데이터베이스를 구축함으로써 대중에게 한 단계 진화한 양질의 서비스를 제공하고 명실공히 우리나라 학술 연구의 허브인 규장각에 대한 이미지를 쇄신한다. 둘째, 중국어 역학서 언어 정보 데이터를 웹을 통해 쉽게 접함으로써 대중들이 고문헌 및 중국어학에 대한 지식을 제고하게 하고, 연구자에게 있어 연구의 효율성을 도모한다. 셋째, 디지털 환경이라는 현시대에 맞추어 고문헌과 전산언어학을 접목한 새로운 융복합 연구의 예를 보여줌으로써 문헌학, 언어학 연구의 지평을 확대한다.

1) 정광(2010, 2011)에 의해 원본 『老乞大』 및 『老朴集覽』에 대한 역주가 저서로 편찬되었고, 그 외 양오진(2008)의 『老乞大朴通事研究』 이태수(2003) 『老乞大四種版本語言研究』와 같은 저서가 있다.

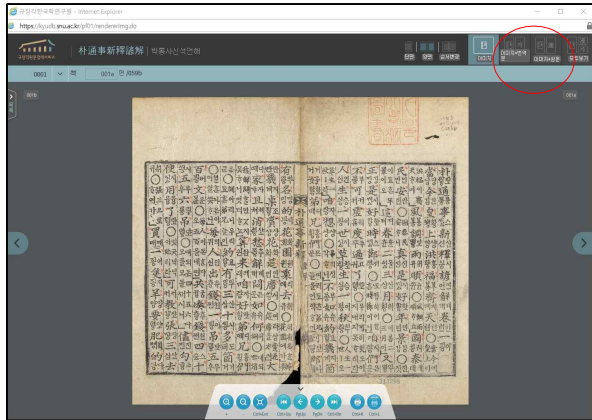
2) 노걸대의 음운 체계에 대한 연구는 鄭光(1974), 「翻譯老乞大 朴通事の 中國語音 表記 研究：四聲通解 歌韻內 諸字의 中聲表記를 中心으로」, 『국어국문학』 64집, 강신항(1978), 「노걸대(老乞大) 박통사산해내자음의 (朴通事諺解內字音) 음계」 등이 있으며, 어법에 대한 연구는 김광조(1998), 「『번역노걸대(翻譯老乞大)』의 조사 "료(了)", "래(來)", "야(也)"의 의미와 문법적 기능에 관하여」, 『중국문학』 30집, 류성은(2000), 「『노걸대』, 『박통사』에 보이는 "的"와 "着"의 용법 고찰」, 『중국어언어연구』 10집, 어휘에 대한 연구는 유성은(2003), 「근대중국어 유의어 소고 - 『노걸대』, 『박통사』에 보이는 상용동사를 중심으로-」, 『중국어언어연구』 16집, 신용권(2003), 「『老乞大』에 나타난 後置詞에 대하여」, 『중국어문학』 42집 등이 있으며 이외에도 판본학, 문화학 등의 방면의 논문이 다수 있다. 통계적으로 보았을 때 음운 방면 42편, 어휘 분야 29편, 문법 41편, 판본학 11편, 사회문화 방면 21편, 교육학 방면 14편, 번역학 방면 3편, 화용론 분야 9편의 논문이 있다.

3) 한국학중앙연구원에서 2011년에서 2016년 이루어진 사업으로, 해당 사이트는 다음과 같다.
http://waks.aks.ac.kr/rsh/?rshID=AKS-2011-AAA-2101

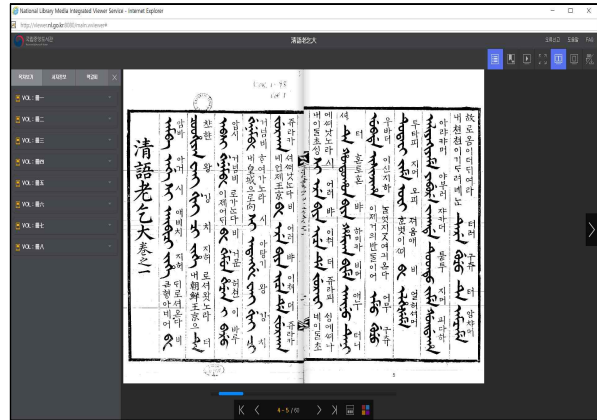
4) 老乞大와 朴通事の 원문은 https://yuedu.baidu.com에서 검색이 가능하다.

2. 국내외의 연구 동향 또는 연구 배경

국내에서 중국어 역학서를 스캔본의 형식으로 1차 데이터가공한 웹은 규장각이나 국립중앙도서관을 들 수 있다. 규장각한국학연구원의 ‘원문검색서비스’의 ‘자부’중 ‘역학류’에 이들 원문이 나타나 있다. 국립중앙도서관에서는 서지사항에 대한 분류가 따로 되어 있지 않고, 일정한 문헌명을 검색하여 원문서비스를 제공하는 문헌에 한 해 그 원본을 확인할 수 있다.



[그림1] 규장각 웹사이트 內
『朴通事新釋諺解』 원문

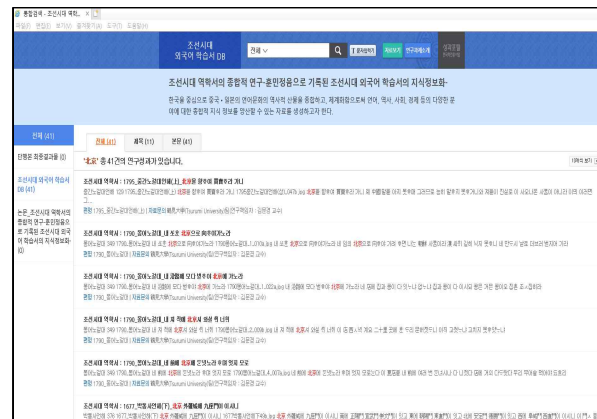


[그림2] 국립중앙도서관 웹사이트 內
『清語老乞大』 원문

이들은 전체적인 원본의 내용을 훑어볼 수 있다는 장점이 있다. 그러나 규장각 사이트 오른쪽 상단 (붉은 동그라미 부분)의 “이미지+번역문”, “이미지+원문” 서비스는 실제 제공하지 않고 있어, 여전히 원문의 내용을 이해할 수 없다는 한계가 있다. <조선시대 역학서 DB>는 이보다 한 단계 높은 2차 데이터 가공으로, 언해문 부분을 웹상으로 입력해 놓았다.



[그림3] <조선시대 외국어 학습서 DB> 중
『蒙語老乞大』의 예



[그림4] <조선시대 외국어 학습서 DB> 중
‘복경’ 검색 결과 예

국내에는 언어를 기록한 고문헌을 데이터베이스로 가공한 웹사이트가 상술한 몇 개 사이트에 지나지 않는다. 이는 국내에서 언어 자료를 바탕으로 한 데이터베이스 가공에 대한 관심이 얼마나 미흡했는지를 여실히 보여준다. 특히 <조선시대 외국어 학습서 DB>의 경우 중국어, 일본어, 몽골어, 만주어 학습서를 총망라했음에도 불구하고, 다양한 언어 정보를 보여줄 수 있는 온톨로지 설계의 부족으로 인해 언해문 검색 서비스만 제공한다는 점이 매우 안

[illegible][illegible][illegible]

- 7 -

특 징	해당 사이트	국내 사이트		국외 사이트	
		규장각 국립중앙도서관	조선시대 외국어 학습서 DB	小學堂	Baidu 『老乞大』, 『朴通事』 디지털 라이징본
원본의 연속적인 확인		○	△	×	×
해당 언어 원문 낱글자 검색		×	×	○	×
해당 언어 원문 단어 검색		×	×	×	×
해당 언어의 발음 검색		×	×	○	×
언해문 부분 단어 검색		×	○	-	×
검색 부분에 해당하는 원문확인		×	○	×	×
해당 웹사이트 데이터의 자유로운 재가공		×	○	○	○



[규장각] 조선시대 중국어 역학서 언어 정보 시멘틱 데이터

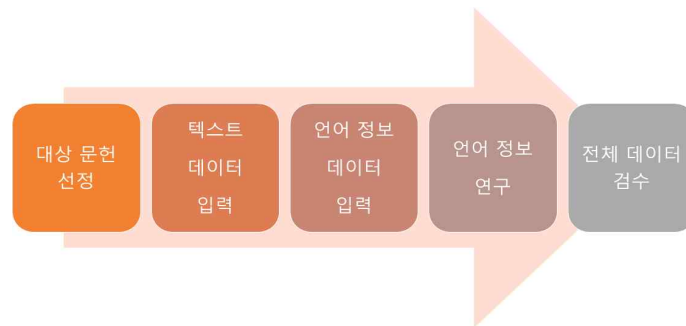
- 원본과 번역문, 원문의 동시 열람
- 해당 문헌의 번역문, 원문의 검색
- 해당 문헌의 어음, 단어 검색
- 검색 부분과 원문의 연계(하이퍼링크)
- 해당 글자, 단어의 자유로운 재가공

규장각, 국립중앙도서관의 데이터베이스와 <조선시대 외국어 학습서 DB>는 원본의 연속적인 확인이 가능하며, <조선시대 외국어 학습서 DB>는 언해문 부분에 대한 특정 정보 검색이 가능하다는 특징을 가진다. <조선시대 외국어 학습서 DB>과 『老乞大』, 『朴通事』 디지털라이징본은 각각 웹 상에 입력된 데이터에 한해서 자유로운 재가공이 가능하다. 그러나 ‘小學堂’ 사이트를 제외하고 대부분 모두 특정 글자나 단어에 대한 정보의 검색이 불가능하다는 공통점을 갖고 있다. ‘小學堂’ 역시 원본을 연계하여 볼 수 없다는 단점을 갖고 있다.

이와 같은 한계점을 극복하고, 좀 더 고도화된 데이터 가공 기술로 양질의 언어 정보를 제공하기 위해 본 연구는 <조선시대 중국어 역학서 언어 정보의 시멘틱 데이터>를 구축하고자 한다. ‘시멘틱 데이터(Semantic Data)’란 ‘시멘틱(semantic)’이란 말 그대로 데이터가 갖고 있는 ‘의미’ 요소에 대해 태그를 부여한 데이터를 말한다. 이는 원문을 그대로 타이핑한 것보다 글자나 단어에 대한 상세한 정보를 함축함으로써 대중들로 하여금 고도로 정련된 정보를 얻게 해준다. 본 연구는 원본과 번역문, 원문을 동시에 열람할 수 있도록 데이터베이스를 구축할 뿐만 아니라 낱글자, 단어에 대해서도 각각의 어음과 단어의 품사, 주제를 태깅하여 기존의 국내외 사이트의 한계점을 뛰어넘는 종합적이고 고도화된 중국어 역학서 언어 정보 데이터베이스를 구축할 것이다.

3. 연구 방법 및 내용

본 연구는 다음과 같이 1) 대상 문헌 선정, 2) 텍스트 데이터 입력, 3) 언어 정보 데이터 입력, 4) 입력한 데이터를 토대로 한 언어 정보 연구, 5) 연구 결과를 바탕으로 한 전체 데이터 검수 및 의미를 내재한 시멘틱 데이터 최종 구축의 순서로 진행한다.



본 연구는 규장각에서 원문보기 서비스를 제공하는 중국어 역학서 중, 한글로 발음이 표기되어 있는 다음 문헌을 대상으로 한다. 『老乞大諺解』, 『朴通事諺解』, 『譯語類解』, 『五倫全備諺解』, 『譯語類解補』, 『朴通事新釋諺解』, 『重刊老乞大諺解』, 『華音啓蒙諺解』, 『華語類抄』, 『漢語抄』의 총 10종의 문헌이며, 동일 책名の 다른 판본을 합하여 총 30권에 대한 데이터베이스를 구축할 것이다.

데이터는 텍스트 데이터와 언어 정보 데이터의 2종을 구축할 것이다. 텍스트 데이터는 문헌 내의 비정형 텍스트 정보를 처리하기 위해서 XML(Extensible Markup Language)로 구축하며, “규장각 원문 검색 서비스”에서 “이미지+원문”과 “이미지+번역문” 정보를 제공하기 위한 기본적인 메타데이터와 원문, 좌음, 우음, 언해문, 번역문 정보를 포함하는 콘텐츠 정보를 포함하고 있다. 또한 문장 내에 등장하는 언어 정보를 처리하기 위한 “어휘” 엘리먼트(Element)를 부여하고, 이를 언어 정보 데이터와 연계한다. 『老乞大諺解』에 해당하는 텍스트 데이터 샘플은 다음과 같다.

```

<?xml version="1.0" encoding="UTF-16"?>
<문헌>
  <메타데이터>
    <제목 언어="cn">老乞大諺解</title>
    <제목 언어="kr">노걸대언해</title>
    <고유번호>奎2044-v.1-2</title>
  </메타데이터>
  <text>
    <contents page="002a">
      <paragraph id="001">
        <원문>老乞大諺解上</원문>
        <좌음>랄깁따연계샹</좌음>
        <우음>란키다연계샹</우음>
        <번역>노걸대언해(상)</번역>
      </paragraph>
      <paragraph id="002">
        <원문>大哥<어휘 id="000001">你</어휘><어휘 id="000002">從</어휘><어휘 id="000003">那裏</어휘><어휘 id="000004">來</어휘></원문>
        <좌음>따거니쥬나리래</좌음>
        <우음>다거니쥬나리래</우음>
        <언해>큰형아네어드러로서브터온다</언해>
        <번역>형님, 당신은 어디서오셨습니까?</번역>
      </paragraph>
      <paragraph id="003">
        <원문><어휘 id="000005">我</어휘><어휘 id="000006">從</
  
```

```

어휘><어휘 id="000007">高麗</어휘><어휘 id="000008">王京</어휘><어휘 id="000009">來</어
휘></원문>

<좌음>어쫑갈리왕깡래</좌음>
<우음>오충갸리왕깡래</우음>
<언해>내高麗王京으로서브터오와</언해>
<번역>저는 고려 왕경에서 왔습니다. </번역>

</paragraph>
</contents>
</text>
</문헌>

```

위 데이터 스키마는 “규장각 원문 검색 서비스”의 데이터 스키마에 부합하도록 수정, 보완이 가능하다. 다음으로, 언어 정보 데이터는 텍스트 데이터에서 출현하는 다양한 언어 정보에 대한 데이터를 구축하기 위해서 어휘 ID, 어휘, 현대중국어 발음, 좌음, 우음, 품사, 분류, 의미 정보를 핵심으로 하는 언어 정보 데이터와 이를 문헌과 연결하기 위한 어휘 ID, 출현 문헌, 출현페이지, 출현 문장을 포함하는 언어-문헌 연결 데이터를 구축할 것이다. 예시는 다음과 같다.

언어 정보 데이터							
ID	어휘	현대 중국어	좌음	우음	품사	분류	의미
000001	你	nǐ	니	니	대	2인칭	당신
000002	從	cóng	쫑	충	개		~로부터
000003	那裏	nàlǐ	나리	나리	대	장소	①어느곳 ②그곳
000004	來	lái	래	레	동		오다
000005	我	wǒ	어	오	대	1인칭	나, 저
000006	高麗	Gāoli	갈리	갸리	대	지명	고려
000007	王京	Wángjīng	왕깡	왕깡	대	지명	왕경

언어 - 문헌 연결 데이터			
ID	출현 문헌	출현 페이지	출현 문장
000001	老乞大 諺解	002a	0001
000002	老乞大 諺解	002a	0001
000003	老乞大 諺解	002a	0001
000004	老乞大 諺解	002a	0001
000005	老乞大 諺解	002a	0002
000006	老乞大 諺解	002a	0002
000007	老乞大 諺解	002a	0002
000008	老乞大 諺解	002a	0002

본 연구는 연구 대상이 되는 10권의 문헌에 실린 총 200,000여자(페이지당 400자, 권당 50쪽, 총 10권)에 대한 글자를 각 항목으로 지정하여 작업할 것이다. 또, 동일 서명의 다른 판본인데 규장각의 청구 기호가 다른 경우 전체 문헌에 대한 대조를 통해 데이터를 구축할 것이다. 이렇게 본 연구는 총 30권 문헌의 600,000여자에 대한 텍스트 데이터와 언어 정보 데이터를 구축할 것이다.

4. 신청 주제와 관련된 연구책임자의 연구 성과

본 연구의 연구책임자는 중국어 음운학 박사로, 「『奎章全韻』의 화음 체계 연구」(2008), 「明清官話音韻專題研究」(2012), 「入聲字 韻母 표기를 통해 본 朝鮮韻書의 官話方言 기초 연

구」(2018), 「『華東正音通釋韻考』의 운모 체계와 북경관화와의 관련성 연구」(2018), 「黃胤錫의 理藪新編 卷二十 「韻學本源」안의 「韻法直圖」 음운 특징 연구」(2019) 등과 같이 조선시대 운서 및 운도, 사역원의 중국어 교재 연구에 있어 풍부한 경험을 갖고 있다. 이외에도 한국연구재단에서 「동국정운에서 개음절 복수음을 갖는 입성자 연구」(2020)을 진행하였으며, 한, 중 음운자료를 이용한 「淸末民初 北京話 入聲字의 演變 연구」(2017)에서 한국중어중문학회의 우수논문상을 수상하기도 하였다. 따라서 조선시대 역학서나 이와 관련된 음운에 있어 상당한 지식과 식견을 갖추고 있다고 할 수 있다.

뿐만 아니라 중국의 근대음을 담고 있는 元代 이후의 음운 자료를 바탕으로 다량의 연구를 진행하였다. 예를 들어, 「《元聲韻學大成》에 나타난 入聲字 변화 연구」(2018), 「《中原雅音》과 《中原音韻》音系 比較 研究」(2019), 「『韻切指歸』안의 『中原音韻』注 研究」(2019), 「《中州音韻輯要》안의 中州音과 北音의 특징과 吳方言의 영향 연구」(2019)을 통해 근대음의 변화 및 방언과의 연관 관계에 대해 규명하였다.

최근에는 淸末 民初시기의 서양인, 한국인, 일본인에 의해 편찬된 중국어 회화서에 대한 음운, 어휘, 교육 방면으로 활발히 연구하고 있다. 「『A GRAMMAR OF THE CHINESE COLOQUIAL LANGUAGE COMMONLY CALLED THE MANDARIN DIALECT』의 음운 체계와 기초 방언 연구」(2014), 「淸末民初 漢語學習書에 나타난 음계 기초 연구 - 《ENGLISH-CHINESE DICTIONARY OF THE STANDARD CHINESE SPOKEN LANGUAGE》, 《官話萃珍》, 《支那四聲字典》, 《支那聲音字彙》를 중심으로」(2017), 「Calvin Wilson Mateer의 중국어 교육과 《官話類編》에 나타난 북경관화와 남경관화의 특징」(2019), 「20세기 초 중국어 학습서 안의 조사의, 呢, 了, 着的 어음 변화 연구」(2020), 「토마스 웨이드(Thomas Wade)의 저작 《尋津錄》에 나타난 19세기 중반 北京官話의 음운 체계와 특수 讀音字 연구」(2020) 등과 같은 연구에서 淸末 民初시기 외국인의 중국어 교육과 이들이 편찬한 중국어 교재 안의 음운, 어휘적 특징에 대해 규명하였다.

연구책임자는 조선시대 음운 자료에 대한 풍부한 연구 경험을 갖고 있으며, 근대 시기 다량의 중국의 언어 자료를 통해 이 시기 음운, 어휘에 대한 해박한 지식을 보유하고 있다. 이를 바탕으로 조선시대 중국어 역학서에 대해 전문적인 데이터베이스를 기획, 설계할 수 있으며 총 과정을 감독하고 세부 내용에 대해 검토할 수 있는 책임자라고 할 수 있다.

5. 연구 추진 일정 및 연구 분담 계획

본 연구는 중국어학 전공의 연구책임자 1명과 언어이론 및 응용언어학 전공 공동연구원 1명, 인문 정보학 공동연구원 1명 및 중국어학 전공의 연구보조원 1명이 협력하여 진행할 것이다. 각 구성원이 맡는 업무는 다음과 같다.

구성원	전공	주요 업무	세부 업무
연구책임자	중국어학	<ul style="list-style-type: none"> 연구 과정 감독 데이터베이스 구상, 오류 검토 	<ul style="list-style-type: none"> 월별 업무 수행 상황 확인, 중간·최종 보고서 작성 중국어 역학서의 원문, 번역문 입력 방식 구상 및 과정 감독, 내용 감수 중국어 역학서의 음운, 어휘 입력 방식 구상 및 과정 감독, 내용 감수

공동연구원1	언어이론 및 응용언어학	<ul style="list-style-type: none"> ◦ 데이터베이스 구상 ◦ 중국어 역학서 언어 정보의 분류, 입력, 검토 	<ul style="list-style-type: none"> ◦ 중국어 역학서의 원문, 번역문 입력 방식 구상 ◦ 중국어 역학서 어음, 어휘 정보(태깅)분류 ◦ 중국어 역학서 어음, 어휘 입력, 내용감수
공동연구원2	인문 정보학	<ul style="list-style-type: none"> ◦ 온톨로지 및 데이터베이스 설계, 구축 및 검토 	<ul style="list-style-type: none"> ◦ 중국어 역학서 데이터베이스 온톨로지 설계 ◦ 데이터베이스 설계 및 구축, 오류 검토
연구보조원	중국어학	<ul style="list-style-type: none"> ◦ 문헌 수집 ◦ 중국어 역학서 언어 정보 데이터 정리, 입력 	<ul style="list-style-type: none"> ◦ 중국어 역학서 및 관련 문헌 수집, 분류 ◦ 중국어 역학서 원문-번역문 데이터 입력

구체적인 월별 연구 추진 일정은 다음과 같다.

연구 목표	규장각 소장 조선시대 중국어 역학서의 언어 정보 시멘틱 데이터 구축		
연구 년·월	연구내용	비고	
		월별 보고서	자문단 회의
2021년 5월	<ul style="list-style-type: none"> ◦ 연구 대상 수집 및 연도, 성격별(회화서/어휘집) 목록화 ◦ 온톨로지 설계 및 데이터베이스 입력 방식 설계 	○	○
2021년 6월	◦ 중국어 역학서 어휘집류 원문, 번역문 데이터 입력 및 감수	○	○
2021년 7월	◦ 중국어 역학서 회화서류 원문, 번역문 데이터 입력 및 감수	○	○
2021년 8월	◦ 중간보고서(제목:「조선시대 중국어 역학서 원문·번역문 데이터베이스 설계와 구축」) 정리 / 오류 검토 및 수정	○	○
2021년 9월	◦ 중국어 역학서 어휘집류 어음 데이터 입력 및 감수	○	○
2021년 10월	◦ 중국어 역학서 회화서류 어음 데이터 입력 및 감수	○	○
2021년 11월	◦ 중국어 역학서 어휘집류 어휘 데이터 입력 및 감수	○	○
2021년 12월	◦ 중국어 역학서 회화서류 어휘 데이터 입력 및 감수	○	○
2022년 1월	◦ 최종보고서(제목:「조선시대 중국어 역학서 언어 정보 시멘틱 데이터 구축」) 정리 / 데이터 입력 내용 재검토, 수정	○	○

6. 연구 결과의 활용 계획 및 기대효과

본 연구는 중국어 역학서의 번역문, 원문 및 어음과 어휘 정보를 XML형식의 기계가독형 데이터로 제출할 것이다. 먼저 번역문, 원문의 데이터는 현재 규장각 원문검색 DB의 원문보기의 “이미지+번역문”과 “이미지+원문”의 탭에 연계시켜 바로 대중들에게 서비스를 제공할 수 있다. 어음과 어휘 정보 데이터의 경우에는 현 규장각 웹사이트에서 ‘중국어 역학서 언어 정보 아카이브’와 같은 독립 영역을 구축하고, 여기에 데이터를 연계시켜 서비스를 제공할 수 있다.

이와 같이 일신(一新)된 규장각 웹사이트는 국내외 모든 사람에게 조선시대 역학서 및 중국어에 대한 지식을 편리하게 제공하는 **고문헌과 언어 정보의 종합적 허브**로 자리매김하게 될 것이다. 이는 어학, 사회학, 민속학 등에 다양한 연구 분야에 대한 새로운 파장을 불러일으킬 것이며 교육 분야에 있어 새로운 메카로서 역할 할 것이다.

먼저, 본 연구의 활용 계획은 다음과 같다.

첫째, 규장각에 대한 홍보 및 우리나라 문헌의 해외 홍보에 활용한다. 본 연구를 통해 구축된 규장각의 중국어 역학서 언어 정보 데이터베이스는 국내 최초의 고문헌에 대한 가장 고도로 가공된 데이터베이스이며, 국외 어느 언어 정보 데이터베이스보다 완성도가 높은 것이다. 이렇게 우수한 데이터베이스로 탈바꿈된 규장각 웹사이트로 ‘규장각’ 및 ‘규장각한국학연구원’을 효과적으로 홍보할 수 있을 뿐만 아니라 자연적으로 우리나라 문헌을 홍보할 수 있는 효과도 거둘 수 있다. **둘째, 국어학 교육 및 연구에 활용한다.** 역학서는 기본적으로 이중언어 자료로, 당시 국어의 변천을 알 수 있다. 본 연구를 통해 구축된 당시 언해문과 각 어음에 대한 한글 표기, 한글 어휘는 국어학 연구나 교육에 즉각적으로 활용될 수 있다. **셋째, 중국어학, 동아시아 언어학 교육 및 연구에 활용한다.** 본 연구가 상세히 구축한 어음과 어휘, 번역문, 원문 내용은 중국어 음운, 어휘, 어법 및 이와 관계를 밀접하게 하고 있는 주변 동아시아 언어인 일본어, 만주어, 몽골어 등의 교육과 연구에도 활용될 수 있다. **넷째, 디지털 인문학 교육 및 연구에 활용한다.** 본 연구의 데이터베이스는 온톨로지 및 데이터베이스 기획 및 설계 과정에 대한 교육이나 국내 디지털 인문학의 연구 성과를 교육하는 데 활용될 수 있다. 또한 향후 본 데이터베이스와 연계한 사회언어학이나 동아시아 언어 등의 데이터베이스 구축에도 토대 자료를 제공해줄 수 있다.

사회적으로 얻을 수 있는 기대효과는 다음과 같다.

첫째, “실생활 속 인문학”의 구현한다. 대중이라면 누구나 쉽게 규장각 웹사이트를 통해 고문헌 및 근대 중국어에 대한 양질의 정보를 얻을 수 있다. 이는 대중들에게 “실생활 속 인문학”을 실현하게 해줄 것이다. **둘째, 한중(韓中) 관계 회복 및 통합을 위해 기여한다.** 최근 사드 및 코로나로 인해 중국에 대한 이미지가 쇠락하였고, 이는 정치, 경제, 교육 시장, 연구 분위기에 전반적으로 부정적인 영향을 미쳤다. 본 연구를 통해 구축된 데이터베이스로 대중 누구나 쉽게 중국어학에 대한 지식을 얻을 수 있으며, 이는 중국에 대한 이미지를 회복하게 해줄 것이다. 또한 이는 한중 정치 관계 완화나 경제적인 이윤 창출의 발판이 될 것이다. **셋째, 코퍼스 및 AI 기술에의 적용한다.** 본 연구에서 구축된 음소 단위의 어음 데이터베이스나 어휘 데이터베이스는 중국어나 동아시아 언어의 코퍼스를 구축하는데 직접적인 자료를 제공한다. 또한 이는 AI의 언어 데이터베이스를 제공해주어, 근세 중국어나 한국어와 관련된 모든 AI 기술에 두루 쓰일 수 있다.

본 연구로 인한 학문적인 기대효과는 다음과 같다.

첫째, 포스트 코로나 시대의 효과적인 연구 기반의 조성한다. 본 연구를 통해 구축된 데이터베이스는 기존에 종이 형태의 문헌으로 일일이 원하는 정보를 찾아야 하는 번거로움에서 벗어나, 원하는 글자나 단어에 대한 어음, 뜻을 찾을 수 있는 편리함을 제공한다. **둘째, 각종 사회학 분야 연구 연구에 토대 자료를 제공한다.** 본 연구로 구축한 언어 정보 데이터는 언어권별 역사, 종교, 문학, 민속학 등의 연구를 위한 기초 자료로 활용될 수 있으며, 이 방면의 연구를 진작시키는 계기가 된다.

7. 기타사항

본 연구는 국내외 문헌학, 중국어 음운, 어휘학 전공자를 자문위원으로 두고자 한다. 이를 통해 좀 더 정밀한 데이터베이스의 설계와 구축이 가능할 것이다. 또한 연구보조원을 통해 매월 자체 보고서를 작성하여 연구 과정의 내실화를 도모할 것이다.

연구참여자 인적사항

1. 연구책임자

소속				E-mail	
성명		(한글) (한자)			
생년월일			전 화	사무실 :	
				핸드폰 :	
주소					
학 력	기 간		학 교	전 공	학 위
경 력	기 간		기 관		직 위

◆최근 3년간 연구실적 및 진행·신청 중인 연구과제

- 연구실적

논문실적					
게재 년월	논문명	학술지명	발행처명	학술지 구분	전체 저자수
2020. 12	토마스 웨이드(Thomas Wade)의 저작 《尋津錄》에 나타난 19세기 중 반 北京官話의 음운 체계와 특수 讀 音字 연구	비교문화연구	비교문화연구소	국내전문학술 지(KCI급)	1
2020. 08	《A Mandarin Primer》에 나타난 음운 체계와 기초 방언 연구	중국인문과학	중국인문학회	국내전문학술 지(KCI급)	1
2020. 06	20세기 초 중국어 학습서 안의 조사 的, 呢, 了, 着의 어음 변화 연구	중국학논총	한국중국문화학회	국내전문학술 지(KCI급)	1
2020. 02	清末民初 北京官話 교재에 나타난 웨이드식 표기의 특징 연구	인문사회과학 연구	인문사회과학연구소	국내전문학술 지(KCI급)	1
2019. 09	Calvin Wilson Mateer의 중국어 교육과 《官話類編》에 나타난 북경관 화와 남경관화의 특징	중국학논총	한국중국문화학회	국내전문학술 지(KCI급)	1
2019. 06	《中州音韻輯要》 안의 中州音과 北 音의 특징과 吳方言의 영향 연구	중국어문학지	중국어문학회	국내전문학술 지(KCI급)	1
2019. 5	黃胤錫의 理藪新編 卷二十 「韻學 本源」 안의 「韻法直圖」 음운 특징 연 구	인문사회과학 연구	인문사회과학연구소	국내전문학술 지(KCI급)	1
2019. 04	『韻切指歸』안의 『中原音韻』注 研究	중국언어연구	한국중국언어학회	국내전문학술 지(KCI급)	1
2019. 04	《中原雅音》과 《中原音韻》音系 比較 研究	중국인문과학	중국인문학회	국내전문학술 지(KCI급)	1
2018. 12	《元聲韻學大成》에 나타난 入聲字 변화 연구	중국어문학지	중국어문학회	국내전문학술 지(KCI급)	1
2018. 08	《華東正音通釋韻考》의 운모 체계와 북경관화와의 관련성 연구	중국인문과학	중국인문학회	국내전문학술 지(KCI급)	1
2018. 03	入聲字 韻母 표기를 통해 본 朝鮮 韻書의 官話方言 기초 연구	중국어문학지	중국어문학회	국내전문학술 지(KCI급)	1

－ 진행·신청 중인 연구과제

연구과제명	지원기관	연구비 (단위 : 천원)	연구기간 (부터 ~ 까지)	역할 (연구책임자 또는 공동연구원)
『동국정운(東國正韻)』안의 개음절 복 수음을 갖는 입성자에 대한 연구	한국연구재단	20000천원	2020.05.01.~ 2021.04.30	연구책임자
청말민초 韓, 中, 日 語音 자료 데이 터베이스 구축을 통한 北京官話 형성 과 음운 특징 연구	한국연구재단	60000천원	2021.05.01.~ 2024.04.30 [신청중인 과제]	연구책임자

2. 공동연구원(1)

*공동연구원 각각 작성

소속		E-mail		
성명	(한글)	(한자)		
생년월일		전 화	사무실 :	
			핸드폰 :	
주소				
학 력	기 간	학 교	전 공	학 위
경 력	기 간	기 관		직 위

◆최근 3년간 연구실적 및 진행·신청 중인 연구과제

－ 연구실적

논문실적					
게재 년월	논문명	학술지명	발행처명	학술지 구분	전체 저자수
2020. 05	汉泰 언어 접촉과 성조 수용 연구 －태국 북부 방언을 중심으로	인문사회과학연구	인문사회과학 연구소	국내전문학술지 (KCI급)	1
2019. 12	單音節聲調和多音節聲調 －以漢語和日語為例	漢語史與漢藏語 研究	中國社會科學 出版社	국제일반학술지	2
2019. 06	基於優選論的漢韓聲調對應分析	現代語文	曲阜師範大學	국제일반학술지	1
2019. 02	試論音段體系與表層聲調形式的關係 － 韓國慶尚方言和中國北部吳語為例	중국어문학논집	중국어문학연 구회	국내전문학술지 (KCI급)	1

－ 진행·신청 중인 연구과제

연구과제명	지원기관	연구비 (단위 : 천원)	연구기간 (부터 ~ 까지)	역할 (연구책임자 또는 공동연구원)
음운규칙 프로그래밍을 통한 데이터 베이스 응용과 언어연구:중국 오방언 을 중심으로	한국연구재단	14,000	2020.09.01 ~ 2021.08.31	연구책임자
동남아시아권 언어의 漢字 차용어 통합 DB 구축과 漢語 성조 수용양 상 연구	한국연구재단	14,000	2021.09.01 ~ 2022.08.31 [신청중인 과제]	연구책임자

2. 공동연구원(2)

*공동연구원 각각 작성

소속		E-mail	
성명	(한글) (한자)		
생년월일		전 화	사무실 :
			핸드폰 :
주소			
학 력	기 간	학 교	전 공 학 위
경 력	기 간	기 관	직 위

◆최근 3년간 연구실적 및 진행·신청 중인 연구과제

- 연구실적

논문실적					
계재 년월	논문명	학술지명	발행처명	학술지 구분	전체 저자수
2020. 08.	디지털 분석 기법을 활용한 시조 연구 방법 탐색-『한국 시조대상전』 수록 단형시조 를 중심으로-	다문화콘텐츠연 구	문화콘텐츠기술연구원	국내 전문 학 술지(KCI급)	2
2020. 06.	한국어 감정 디지털 온톨로 지 구축에 관한 연구	한국어 의미학	한국어의미학회	국내 전문 학 술지(KCI급)	3
2020. 03.	딥러닝을 활용한 경판 방각 본 소설의 유형 고찰 -영웅중 소설과 애정소설 유형을 중 심으로-	국제어문	국제어문학회	국내 전문 학 술지(KCI급)	2
2019. 12.	고전소설에 대한 디지털 감 정 분석방법론 탐색 - <구운 몽>을 대상으로 -	동아시아고대학	동아시아고대학회	국내 전문 학 술지(KCI급)	2
2019. 10.	디지털 맵핑(Mapping)을 통 한<구운몽> 연구 및 교육적 활용	열상고전연구	열상고전연구회	국내 전문 학 술지(KCI급)	2
2019. 10.	디지털 분석 기법을 활용한 시조문학 연구 - 인공지능 (AI) 제재 창작 시조를 중 심으로 -	문화와 융합	한국문화융합학회	국내 전문 학 술지(KCI급)	2
2019. 06.	빅데이터와 고전문학 연구방법 론	어문론집	중앙어문학회	국내 전문 학 술지(KCI급)	2
2019. 06.	딥러닝으로 불경 읽기 - Word2Vec으로 CBETA 불 경 데이터 읽기	원불교사상과 종교문화	원불교사상연구원	국내 전문 학 술지(KCI급)	1
2019. 03.	계량적 문제 분석을 통한 <소현성록> 연작의 변이양 상 고찰 - 이대15권본과 규 장각21권본을 중심으로 -	국제어문	국제어문학회	국내 전문 학 술지(KCI급)	2
2018. 12.	소현성록 연작의 문체론적 고찰 - 컴퓨터를 활용한 계 층분석을 바탕으로	인문과학연구	인문과학연구소	국내 전문 학 술지(KCI급)	2
2018. 09.	부끄러움/창피함/쑥스러움/ 수치스러움/수줍음간의 관 계 고찰: 공기 명사 및 네트 워크 분석을 통한 문맥 분 석	언어	한국언어학회	국내 우수 전 문 학 술 지 (KCI급)	2
저역서					
발행 년월	저역서명	발행처명		저역서 구분	전체 저자수
2018. 12.	시맨틱 데이터 아카이브의 구축과 활용 - 제도와 인사의 관계에 대한 근대 (1895~1910) 학교 자료를 중심으로	보고사		저서	1

－ 진행·신청 중인 연구과제

연구과제명	지원기관	연구비 (단위 : 천원)	연구기간 (부터 ~ 까지)	역할 (연구책임자 또는 공동연구원)
한국 역사 인물 온톨로지 설계와 LOD(Linked Open Data) 서비스 구축 및 운용	한국연구재단	200,000	2020.07. ~ 2025.06.	연구책임자

3. 연구보조원

*연구보조원 각각 작성

소속		E-mail			
성명		(한글)	(한자)		
생년월일		전 화	사무실 :		
			핸드폰 :		
주소					
학 력	기	간	학 교	전 공	학 위

□ 예산서

한국학 연구용역 사업 연구비 산출내역서

(연구책임자 : 성명 (인))

<1차 년도 연구비 산출내역>

(단위 : 원)

항 목		신 청 액	산 출 내 역
① 인건비		21,600,000	1) 연구책임자, 공동연구원1: 월 400,000 * 9개월 * 2명=7200,000 2) 공동연구원2: 월 1000,000 * 9개월 * 1명=9000,000 3) 연구보조원: 월 600,000 * 9개월 * 1명=5400,000
소 계		21,600,000	
② 연구활동경비	· 회의비	360,000	월 40,000(연구책임자, 공동연구원, 연구보조원 총 4명, 인당 1만원) * 9개월
	· 여비		
	· 유인물비	101,818	관련 논문 및 자료 인쇄 비용
	· 교통통신비		
	· 기타비용	620,000	1) 사무용품(파일, 잉크 등) 구매 비용=170,000 2) 자문위원 자문료=1회 50,000*9개월(1개월당 1회)=450,000
소 계		1,081,818	
③ 직접성 경비	· 조사연구비		
	· 문헌구입비	500,000	조선시대 역사, 문헌, 중국어학 관련 저서 구매 비용
	· 전산처리비	20,000,000	1) 데이터 입력비용: 자당 40원 * 페이지당 400자 * 권당 50페이지 * 10권 = 800만원 2) 데이터고도화 비용 (음운 및 어휘 식별): 자당 60원 * 페이지당 400자 * 권당 50페이지 * 10권 = 1200만원
소 계		20,500,000	
④ 간접비	· 간접연구경비	2,272,727	
소 계		2,272,727	
⑤ 부가가치세		4,545,455	
소 계		4,545,455	
합 계		50,000,000	

<2차 년도 연구비 산출 내역>

항 목		신 청 액	산 출 내 역
① 인건비		21,600,000	1) 연구책임자, 공동연구원1: 월 400,000 * 9개월 * 2명=7200,000 2) 공동연구원2: 월 1000,000 * 9개월 * 1명=9000,000 3) 연구보조원: 월 600,000 * 9개월 * 1명= 5400,000
소 계		21,600,000	
② 연구 활동 경비	· 회의비	360,000	월 40,000(연구책임자, 공동연구원, 연구보조원 총 4명, 인당 1만원) * 9개월
	· 여비		
	· 유인물비	101,818	관련 논문 및 자료 인쇄 비용
	· 교통통신비		
	· 기타비용	620,000	1) 사무용품(파일, 잉크 등) 구매 비용=170,000 2) 자문위원 자문료=1회 50,000*9개월(1개월당 1회)=450,000
소 계		1,081,818	
③ 직접성 경비	· 조사연구비		
	· 문헌구입비	500,000	조선시대 역사, 문헌, 중국어학 관련 저서 구매 비용
	· 전산처리비	20,000,000	1) 데이터 입력비용: 자당 40원 * 페이지당 400자 * 권당 50페이지 * 10권 = 800만원 2) 데이터고도화 비용 (음운 및 어휘 식별): 자당 60원 * 페이지당 400자 * 권당 50페이지 * 10권 = 1200만원
소 계		20,500,000	
④ 간접비	· 간접연구경비	2,272,727	
소 계		2,272,727	
⑤ 부가가치세		4,545,455	
소 계		4,545,455	
합 계		50,000,000	