

문헌학연습: 한국어 자료의 활용

1 주차 문헌의 연구

김미경

2025.9.1

문헌 사용 연구

왜 문헌을 쓰는가

유일한 옵션일 때가 있다

- ▶ 과거의 언어
- ▶ 용례의 사용량
- ▶ 용례의 다양성
- ▶ 화자가 의식하지 못하는 특성

화자를 모을 필요가 없다

- ▶ 가설을 즉각 확인할 수 있다
- ▶ 시행착오를 몇번이고 거칠 수 있다

문헌을 사용한 언어 연구

경험적 연구 (Empirical study)

- ▶ 언어표현 X 가 어떻게 쓰이는가?
- ▶ 언어표현 X 가 얼마나 쓰이는가?
- ▶ 언어표현 X 가 무엇과 함께 쓰이는가?
- ▶ 언어표현 X 가 어떤 환경에서 자주 쓰이는가?
- ▶ 언어표현 X 의 출현에 영향을 미치는 요소는 무엇인가?

문헌 관련 분야들

문헌 연구 분야

고전 연구

관심사	목표	학문 분과
-----	----	-------

사상	원전 복원	문헌학/고전학 (Philology/Classical studies)
----	-------	--

텍스트	조어 재구	역사비교언어학 (Historical Comparative Linguistics)
-----	-------	---

매체	선본 확립	서지학 (Bibliography)
----	-------	--------------------

문헌 연구 분야

모든 기록 연구

관심사	목표	학문 분과
-----	----	-------

매체	수집/분류	문헌정보학 (Library and Information Science)
----	-------	--

텍스트	진본 보존	기록관리학/기록학 (Archival Science/Studies)
-----	-------	---

문헌학과 역사언어학의 관계

문헌학의 도구로서의 과거 언어 연구

18 세기 문헌학 내에 비교 문법 개설

언어학의 도구로서의 과거 문헌 연구

문헌학은 19 세기 역사비교언어학 연구자의 상식

언어학의 하위 연구 분야

- ▶ 역사비교언어학이 언어학의 하위 분야로 밀려남
- ▶ 문헌학이 더 이상 상식이 아님
- ▶ 과거의 언어연구 + 문헌 연구가 언어학과에서 ‘문헌학’으로 통용

서지학과 역사언어학의 관계

언어학의 도구로서 과거 문헌 연구

매체의 특성을 통해 텍스트의 생산 시기와 작성자의 계통 판별

한중일

‘서지학’활용 (서지학과 문헌학의 분리)

서양

‘문헌학’활용 (서지학이 문헌학의 일부)

종합: 왜 문헌을 공부하나?

기록을 용례로 변환

문자 해독

용례의 메타정보 이해

- ▶ 용례의 시기 판별
- ▶ 용례의 발화자 판별
- ▶ 용례의 장르 판별

관련 분야: 이론/역사언어학 연구

분류

연구 목적에 따른 언어학의 한 분야

목표

언어 표현의 공시적/통시적 특성을 설명함

- ▶ 언어표현 X 는 Z 라는 성질을 가진다
- ▶ 언어표현 X 의 성질 Z 는 시간이 흐르면서 변화했다

문헌과의 관계

문헌에서 추출한 용례를 증거로 활용

문헌의 확장과 관련분야 확대

‘문헌’개념의 확장

《서지학개론》 서문 중, 한울아카데미, 2004

서지학은 문헌을 대상으로 과학적으로 연구하는 학문이다. 서양에서는 기원전 2 세기 고대 그리스에, 동양은 기원전 1 세기 전한시대에 비롯되었다. 인류문화의 발전에 따라 문헌의 개념이 변화되어 오늘날은 **문헌이란 기록된 정보원을 총칭하는 개념**이 되었다. 따라서 문헌이란 기록된 정보를 전달하는 모든 수단—도서, 잡지, 소책자, 마이크로필름, 오디오 및 비디오테이프, 슬라이드, 점자, 영화필름, 디스켓, 시디롬, 광디스크, 전자책, 데이터베이스 등—을 의미한다.

‘문헌’개념의 확장

새로운 분야의 발전

디지털인문학 (Digital Humanities)

역사언어학의 ‘문헌’ 확대

- ▶ 고문헌
- ▶ 고문헌 및 20 세기 이후 언어의 전산입력자료 (new!)
- ▶ 전산입력자료로 편찬한 말뭉치 (new!)

확장된 문헌의 자료 양태

고문헌

20 세기 이전 기록물의 원본 및 영인본

전산입력자료

- ▶ 고문헌 및 20 세기 이후 출판물의 전산화 자료
- ▶ 20 세기 이후 컴퓨터로 생산된 언어자원

말뭉치

- ▶ 목적을 가지고 선정된 전산입력자료
- ▶ 추가 정보가 부착된 전산입력자료

확장된 문헌의 특성 비교

문헌 비교

고문헌	전산입력자료	말뭉치
소수의, 거의 유일한 증거 특이 용례 중심 질적 연구 위주	대량의 다양한 증거 다수 용례 중심 질적 및 양적 연구	대표성이 있는 증거 분포 중심 양적 연구

관련 분야의 확장

과거의 문헌

- ▶ 유물로서의 고문헌을 하나하나 다룸
- ▶ 용례 하나 하나를 주의깊게 분석
- ▶ 고문헌의 특성과 취급요령: 서지학
- ▶ 용례의 분석: 이론/역사언어학 (질적) 연구

관련 분야의 확장

확장된 문헌

- ▶ 전산상에 존재하는 대량의 문헌을 동시에 다룸
 - ▶ 용례들의 분포와 빈도가 중요해짐
 - ▶ 연구자가 생각하지 못한 대상을 발견할 수 있게 됨
-
- ▶ 대량 문헌의 특성과 취급요령: 말뭉치 언어학, 컴퓨터 언어학
 - ▶ 새 패턴 발견: 말뭉치 언어학, 컴퓨터 언어학
 - ▶ 분포와 빈도 세기: (이론/역사언어학) 양적 연구

관련 분야: 양적 연구

분류

연구방법론의 일종

목표

- ▶ 언어 표현의 공시적/통시적 특성을 숫자로 표현된 증거로 설명
- ▶ 숫자의 많고 적음에 의미 부여 (항상 전체 수치 필요)

문헌과의 관계

숫자를 세기 위해 문헌 자료가 필요함

관련 분야: 말뭉치 언어학 연구

분류

연구 자료에 따른 언어학의 하위 분야

목표

실제로 사용된 언어의 패턴을 관찰하고 기술

- ▶ 화자들은 실제로 이런 표현을 사용한다 (corpus-based)
- ▶ 언어 자료에서 예상치 못했던 패턴이 발견되었다 (corpus-driven)

문헌과의 관계

문헌을 분석한다는 것 자체가 분야의 정체성

관련 분야: 컴퓨터 언어학 연구

분류

연구 목적에 따른 응용언어학의 한 분야

목표

컴퓨터가 인간의 언어를 이해하고 생성하게 함

문헌와의 관계

- ▶ 말뭉치 언어학의 결과물 (말뭉치, 분석) 을 핵심 재료로 사용
- ▶ 말뭉치 생산을 주도함

문헌학연습

문헌학연습의 목표

언어표현 X 를 고문헌에서 찾기

- ▶ 과거의 문자 학습
- ▶ 과거 언어의 문법 학습
- ▶ 과거 언어 텍스트 읽기 연습 (강독)
- ▶ 과거의 기록물 특성 익히기

언어표현 X 를 전산입력자료/말뭉치에서 찾기 (new!)

- ▶ 전산입력자료와 말뭉치 파악하기
- ▶ 말뭉치 편찬 이해하기
- ▶ 언어표현 추출하기

2025-2 문헌학연습의 목표

이론/역사언어학자인 나의 바람

연구 대상 표현을 관찰하고 싶다

실행상의 장벽

대응

문제

말뭉치에서 찾아보면 되겠지...?

파일이 너무 많음

코딩을 배우라고...?

파이썬이 두려움

말뭉치 파일 구조를 분석하라고...?

텍스트 아닌 요소가 너무 많음

이걸로 무슨 주장을 할 수 있지?

주장 유형을 모름

2025-2 문헌학연습의 목표

이 수업의 해결책

- ▶ 전산입력자료/말뭉치에서 연구할 표현 추출하기
- ▶ 언어 표현의 사용 양상을 토대로 할 수 있는 주장 살펴보기

목표하지 않는 것

- ▶ 파이썬 전문가
- ▶ 컴퓨터언어학자

전산입력자료 및 말뭉치 분석을 위한 능력

Gries & Newman (2014) '말뭉치 분석 소프트웨어에 필요한 기능'(일부)

- ▶ 여러 개의 파일 열기
- ▶ 여러 가지 인코딩 대응하기
- ▶ 단어, 품사, 단어 연쇄 등의 빈도 세기
- ▶ 정규식을 이용하여 지정한 패턴의 빈도 세기
- ▶ 검색 결과를 출현 환경과 함께 제시하기
- ▶ 검색어의 공기어 보여주기
- ▶ 단어간 관련성의 강도를 계산하기
- ▶ 결과를 저장하고 파일로 내보내기

수업 계획

수업 계획

1 주차 (9/1)

- ▶ 문헌과 언어학의 관계
- ▶ 실습: Miniconda, VS Code, Jupyter Notebook 설치

2 주차 (9/8)

- ▶ 선행연구 분석 (연구질문 & 데이터 중심)
- ▶ 실습: txt 파일 읽어오기

수업 계획

3 주차 (9/15)

- ▶ 선행연구 분석 (연구질문 & 데이터 중심)
- ▶ 실습: csv, tsv 파일 읽어오기

4 주차 (9/22)

- ▶ 선행연구 분석 (연구질문 & 데이터 중심)
- ▶ 실습: 데이터프레임 다루기

수업 계획

5 주차 (9/29)

- ▶ 선행연구 분석 (연구질문 & 데이터 중심)
- ▶ 실습: docx 읽어오기
- ▶ 숨겨진 실습: 우리가 자주 쓸 파이썬 코드 이해하기

6 주차 (10/6 추석)

보강 대신 자율학습

수업 계획

7 주차 (10/13)

- ▶ 선행연구 분석 (연구질문 & 데이터 중심)
- ▶ 실습: XML 과 그 변종들 읽어오기

8 주차 (10/20)

- ▶ 개인연구 브레인스토밍
- ▶ 실습: JSON 읽어오기

수업 계획

9 주차 (10/27)

- ▶ 개인연구 주제 발표
- ▶ 실습: 형태소 분석과 품사 부착 (koNLPy)

10 주차 (11/3)

- ▶ 실습: 연어 분석
- ▶ 개인연구 진도보고

11 주차 (11/10)

개인연구 중간발표

수업 계획

12 주차 (11/17)

- ▶ 그래프 그리기
- ▶ 실습: 빈도 분석과 시각화

13 주차 (11/24)

- ▶ 실습: word2vec
- ▶ 개인연구 진도보고

수업 계획

14 주차 (12/1)

- ▶ 시기별 데이터를 이용한 변화의 추적
- ▶ 개인연구 진도보고

15 주차 (12/8)

개인연구 발표

실습

상황 파악

나는...

- ▶ 말뭉치에서 표현을 세서 데이터프레임으로 나타낼 수 있다.
- ▶ 말뭉치에서 표현을 추출하여 용례만 파일로 저장할 수 있다.
- ▶ 말뭉치 파일 여럿을 자동으로 읽어올 수 있다.
- ▶ 파이썬을 다룰 수 있다.
- ▶ 내 컴퓨터에 파이썬이 설치되어 있다.

환경 갖추기: Miniconda 설치

1. 설치 파일 다운로드

- ▶ 접속: <https://www.anaconda.com/download/success>
- ▶ **Mac 사용자:** 본인 Mac 의 칩 종류 확인 후 다운로드
 - ▶ **Apple Silicon (M1/M2..):** 'Download for Apple Silicon' 클릭
 - ▶ **Intel:** 'Download for Intel' 클릭
- ▶ **Windows 사용자:** 제안된 파일 다운로드

2. 설치 진행

- ▶ **Mac:** 다운로드한.pkg 파일 실행
- ▶ **Windows:** 다운로드한.exe 파일 실행 (설치 시 옵션들 모두 체크)

환경 갖추기: Miniconda 설치

3. (매우 중요) 터미널/Anaconda Prompt 재시작

- ▶ 설치 완료 후, 열려있는 모든 터미널 창을 닫고 새로 실행
- ▶ 다음 명령어를 입력해서 설치가 되었는지 확인

```
1 python --version
2 conda --version
```

환경 갖추기: 파이썬 가상환경 만들기

1. 터미널 실행

- ▶ **Mac:** '터미널 (Terminal)' 앱 실행
- ▶ **Windows:** 시작 메뉴에서 '**Anaconda Prompt**' 검색하여 실행 (**일반 터미널 (cmd) 이 아님!**)

환경 갖추기: 파이썬 가상환경 만들기

2. 가상환경 생성 및 실행 (아래 명령어 순서대로 입력)

1. 파이썬 3.13 버전으로 'philology' 가상환경 생성

```
1 conda create -n philology python=3.13
```

2. 생성한 가상환경 활성화

```
1 conda activate philology
```

3. 주피터 노트북 연결 프로그램 설치 (프롬프트 맨 앞에 (philology) 확인 후 진행)

```
1 conda install ipykernel -y
```


환경 갖추기: VS code 설치하기

설치파일 내려받기 및 설치 진행

- ▶ <https://code.visualstudio.com>
- ▶ 제안된 파일 다운로드 및 설치
- ▶ 설치 마지막 단계에서 'PATH 에 추가' 옵션 확인

환경 갖추기: VS code 설정하기

VS code 에서 작업 폴더 열기

- ▶ 원하는 위치에 '언어학실습' 폴더 만들기
- ▶ VS code 실행 후 'File > Open Folder...' 선택
- ▶ 방금 만든 '언어학실습' 폴더를 열기
- ▶ 앞으로 모든 실습 파일은 이 폴더 안에 저장!

환경 갖추기: VS code Extensions 설치하기

- ▶ VS code 왼쪽 메뉴의 'Extensions' 아이콘 클릭
- ▶ "python" (Microsoft) 검색 후 install
- ▶ "Jupyter" (Microsoft) 검색 후 install

환경 갖추기: 첫 코드 실행

환경 갖추기: 파이썬 인터프리터 선택해보기

- ▶ 새 파일 > Jupyter Notebook 선택
- ▶ 새로 열린 파일의 첫 번째 셀에 아래 코드와 같이 입력하고 왼쪽의 실행 버튼 클릭
- ▶ 커널을 선택하라고 뜨면 목록에서 'philology (Python 3.13.5)' 선택
 - ▶ (커널 목록에 뜨지 않으면 VS Code 를 재시작)
- ▶ VS code 오른쪽 상단에 버전이 표시되면 성공!

```
1 print("Hello, world!")
```
