Preprocessing KSS Data (Detail)

≔ 태그

1️⃣ Download KSS
    압축 해제
2️⃣ Directory
    .wav 파일
    .txt 파일
3️⃣ script 편집
    ① transcript.v.1.4.txt을 Excel로 열기
        1) 구분 기호로 분리됨 선택 → 다음
        2) 구분 기호 : 기타 선택 → | 입력 → 다음
        3) 열 데이터 서식 : 일반 선택 → 다음
    ② 파일 편집
        1) 필요 없는 열 삭제 : C, D, E, F 열 삭제
    ③ 데이터 나누기 : 1 / 2 / 3 / 4
        1) A열 선택
        2) 데이터 탭 → 필터
        3) 필터 드롭다운 클릭
        4) 텍스트 필터 → 시작 문자
        5) 시작 문자에 1/1 입력 → 확인
        6) 본문 내용 전체 선택 → 복사
        7) 새로운 sheet 생성 후 붙여넣기
        8) 새로운 sheet 내용 전체 선택 → 복사
        9) 메모장에 붙여넣기
        10) 편집 탭 → 바꾸기(R)...

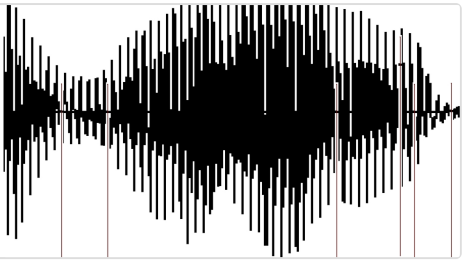
        11) 찾을 내용에 공백 복사해서 붙여넣기
        12) 바꿀 내용 : | 입력 → 모두 바꾸기(A) → X
        13) 메모장 저장
        14) 5) 부터 13) 반복 : 2/2, 3/3, 4/4
        15) 최종 생성 파일 : 4개
    ④ Directory
4️⃣ Preprocess Data
    가상환경 활성화
    경로 이동
    preprocess

4개 : 1, 2, 3, 4
2개 : 1, 2

1️⃣ Download KSS

Korean Single Speaker Speech Dataset
KSS Dataset: Korean Single Speaker Speech Dataset

k https://www.kaggle.com/datasets/bryanpark/korean-single-speaker-speech-dataset



- Kaggle에 로그인 후 다운로드
- Archive.zip 다운로드
- Archive.zip 구조
  - kss
    - 1
    - 2
    - 3
    - 4
  - transcript.v.1.4.txt

압축 해제

```
unzip archive.zip
```

2️⃣ Directory

.wav 파일

- `fodong/vits/1/`
- `fodong/vits/2/`
- `fodong/vits/3/`
- `fodong/vits/4/`

.txt 파일

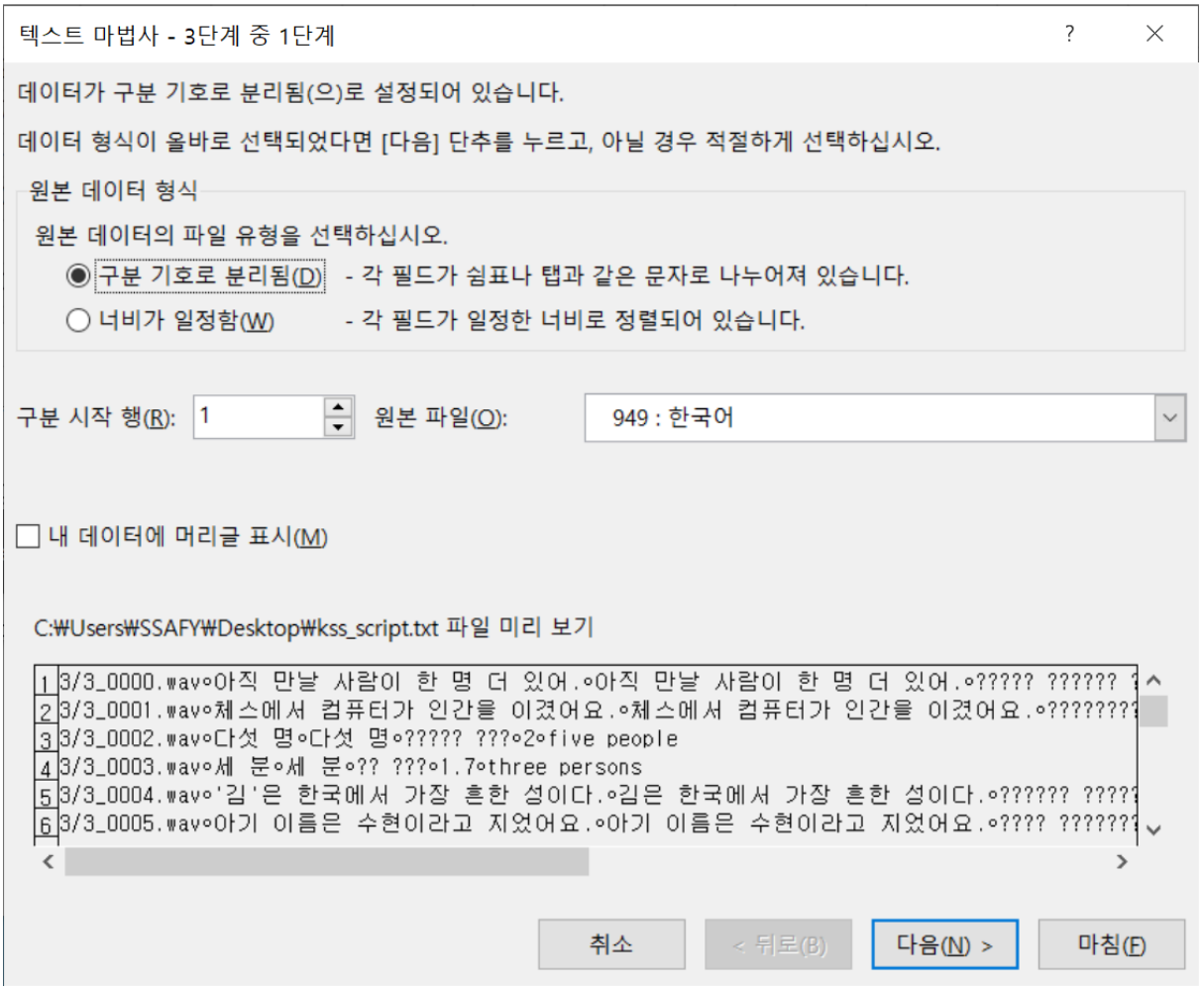
- `fodong/vits/filelists/transcript.v.1.4.txt`

3️⃣ script 편집

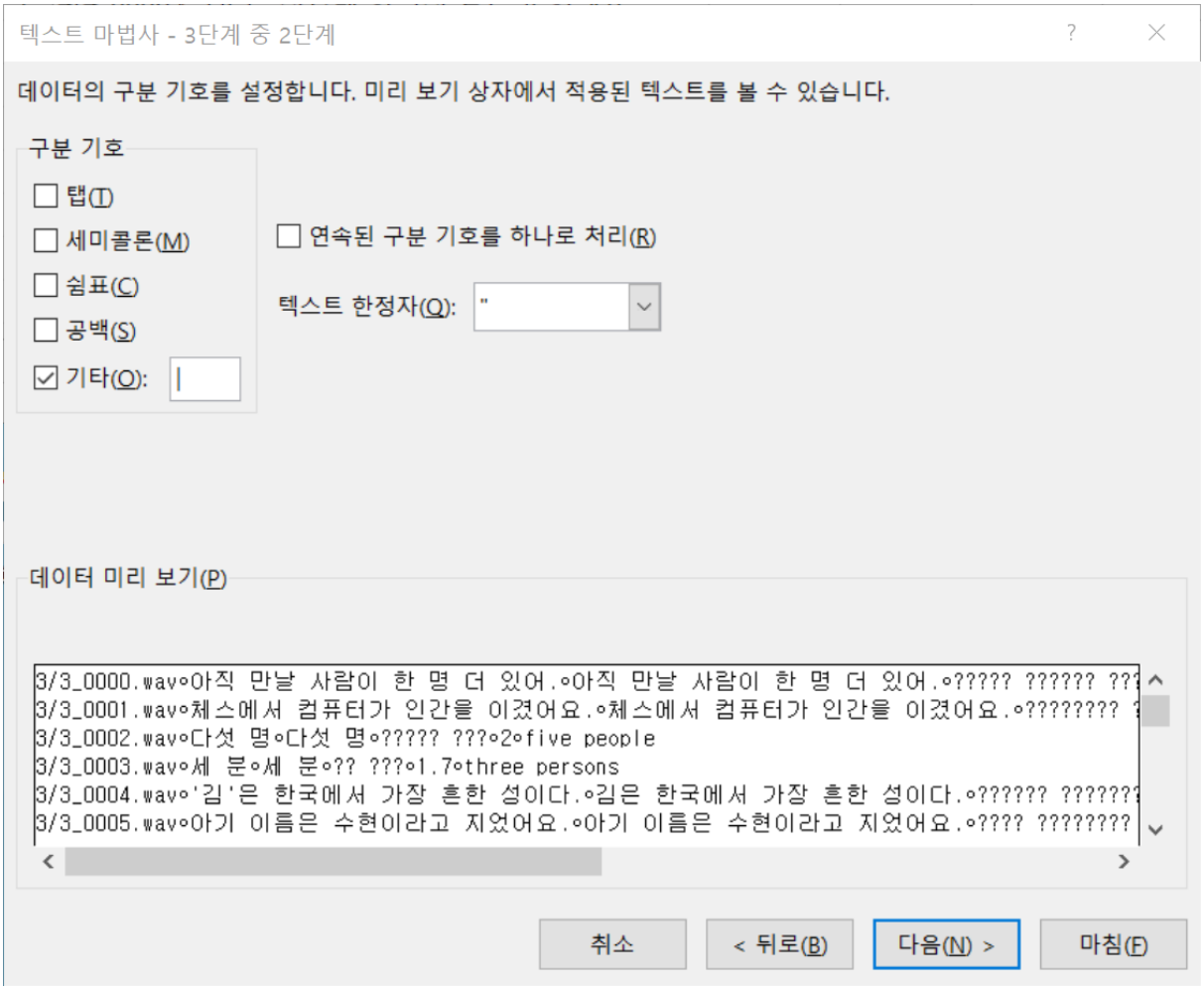
> Excel 활용

① transcript.v.1.4.txt을 Excel로 열기

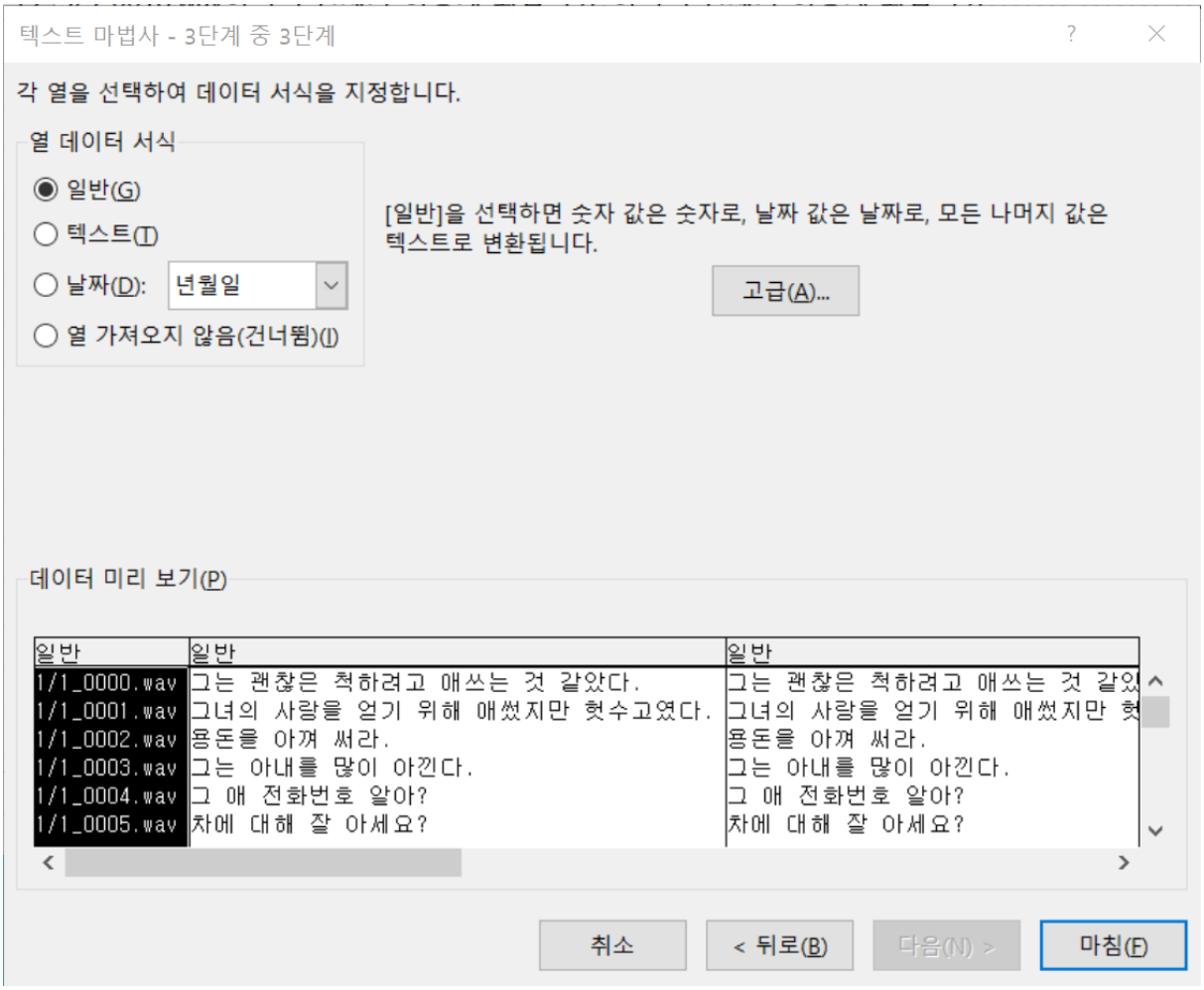
1) 구분 기호로 분리됨 선택 → 다음



2) 구분 기호 : 기타 선택 → | 입력 → 다음



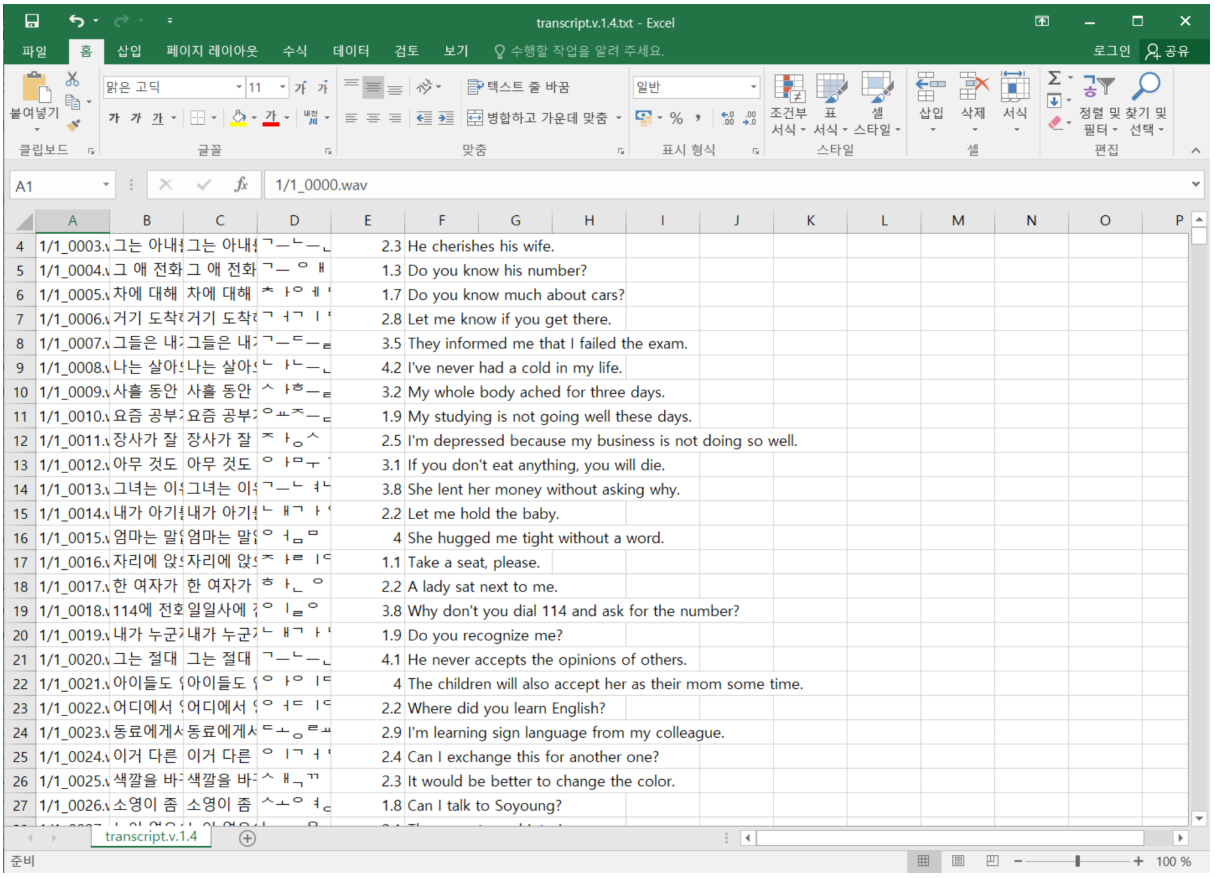
3) 열 데이터 서식 : 일반 선택 → 다음



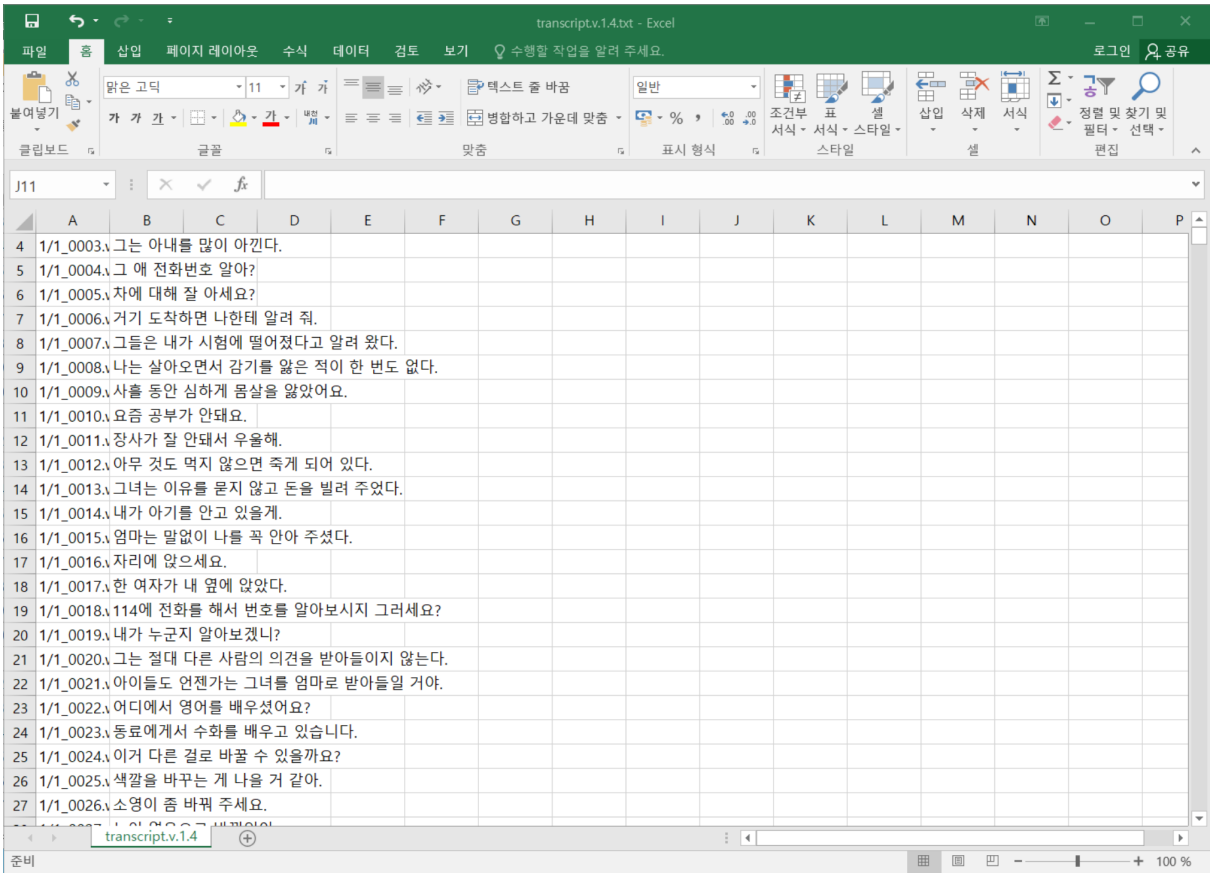
② 파일 편집

1) 필요 없는 열 삭제 : C, D, E, F 열 삭제

[삭제 전]



[삭제 후]

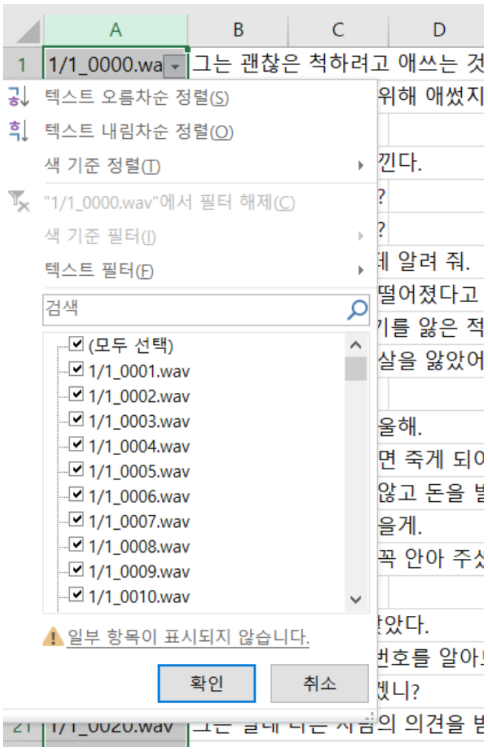


③ 데이터 나누기 : 1 / 2 / 3 / 4

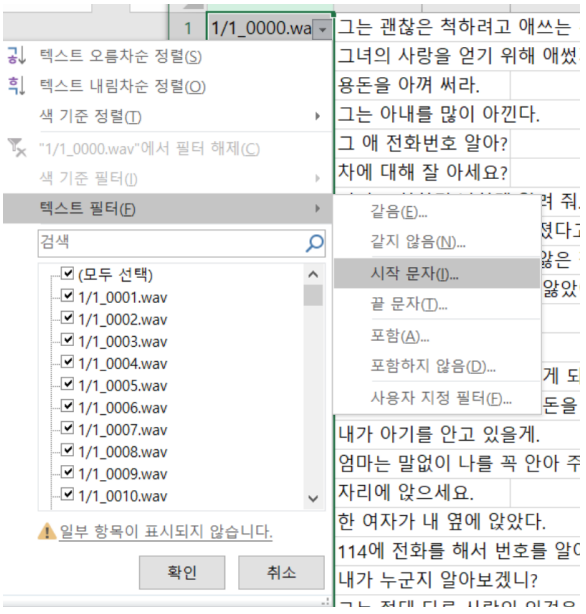
1) A열 선택

2) 데이터 탭 → 필터

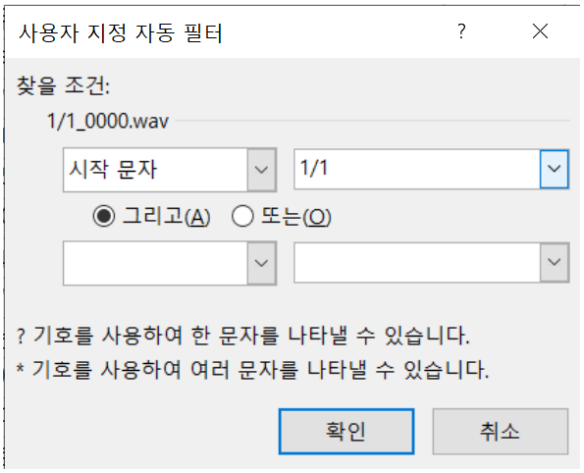
3) 필터 드롭다운 클릭



4) 텍스트 필터 → 시작 문자



5) 시작 문자에 1/1 입력 → 확인

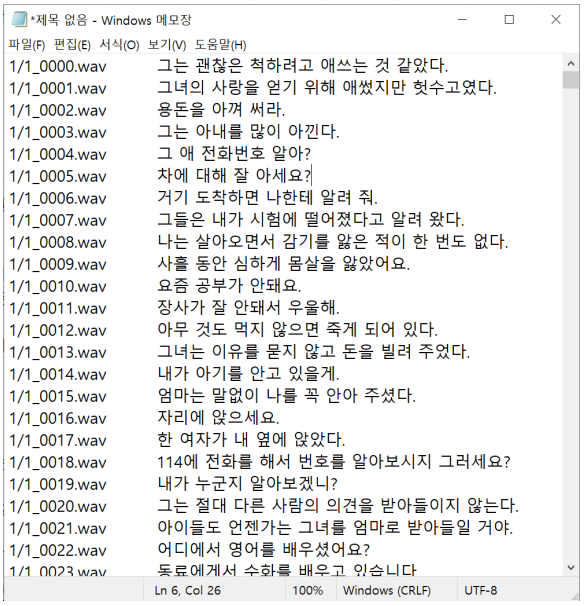


6) 본문 내용 전체 선택 → 복사

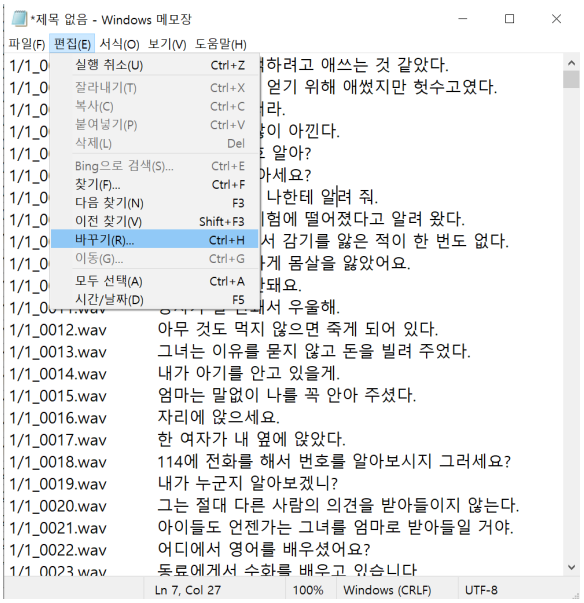
7) 새로운 sheet 생성 후 붙여넣기

8) 새로운 sheet 내용 전체 선택 → 복사

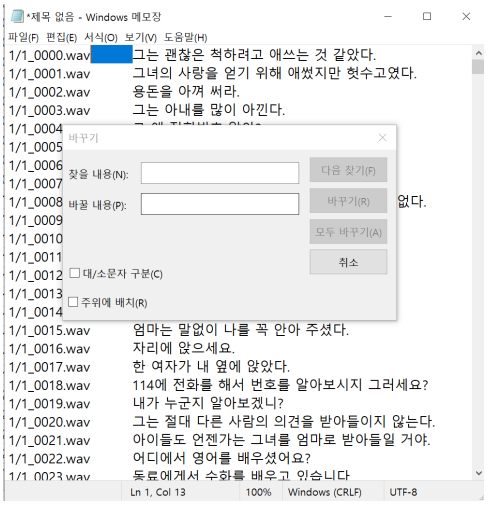
9) 메모장에 붙여넣기

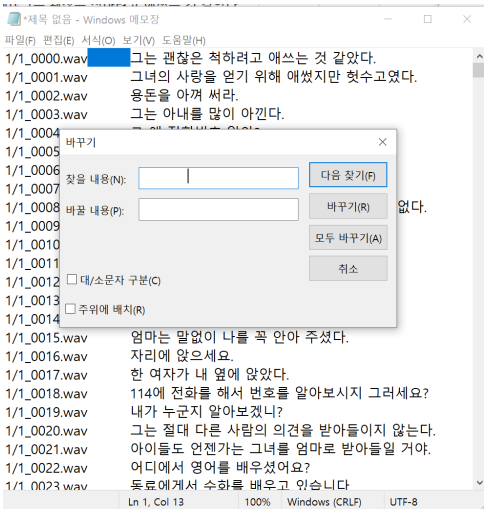


10) 편집 탭 → 바꾸기(R)...

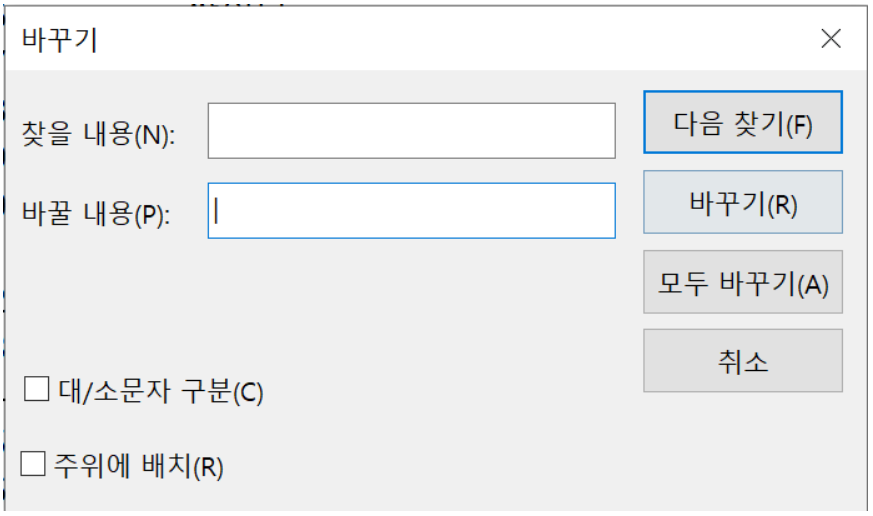


11) 찾을 내용에 공백 복사해서 붙여넣기





12) 바꿀 내용 : | 입력 → 모두 바꾸기(A) → X



13) 메모장 저장

- 파일 명 : kss_script_1.txt

14) 5) 부터 13) 반복 : 2/2, 3/3, 4/4

- 주의할 점
  - 7) 새로운 sheet에 붙여넣기 한 후 첫 번째 행이 1/1이면 해당 행 삭제하기

15) 최종 생성 파일 : 4개

- `kss_script_1.txt`
- `kss_script_2.txt`
- `kss_script_3.txt`
- `kss_script_4.txt`

④ Directory

- `fodong/vits/filelists/kss_script_1.txt`
- `fodong/vits/filelists/kss_script_2.txt`
- `fodong/vits/filelists/kss_script_3.txt`
- `fodong/vits/filelists/kss_script_4.txt`

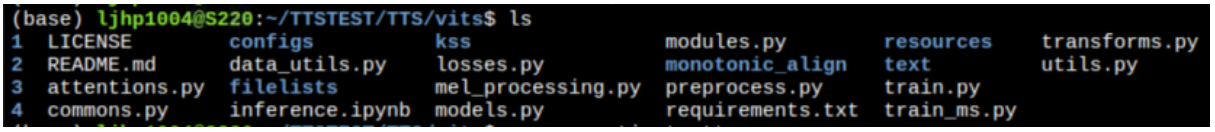
4️⃣ Preprocess Data

가상환경 활성화

```
source activate fodong
```

경로 이동

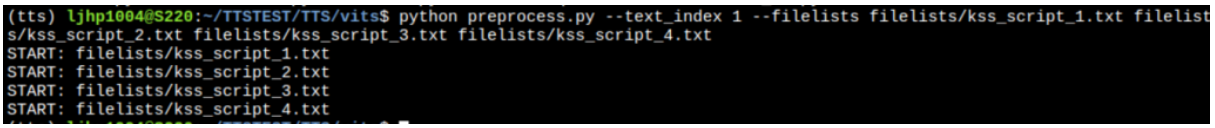
- 경로 : `fodong/vits/`
- `preprocess.py` 파일 확인



preprocess

4개 : 1, 2, 3, 4

```
python preprocess.py --text_index 1 --filelists filelists/kss
```



2개 : 1, 2

```
python preprocess.py --text_index 1 --filelists filelists/kss
```