# 공학 프로젝트 기획
# 번역기 성능 향상

## 자모 단위 변환 & 높임말, 낮춤말 변환

허재무, 김준태, 김주환, 김정희 2021.11.26(금)

# 피드백 진행 상황

# 피드백 진행 상황

**-피드백 list**

1. 처리못하는 문장 유사도 검사
2. Optimizer, activation function
3. 영어의 높임말 처리

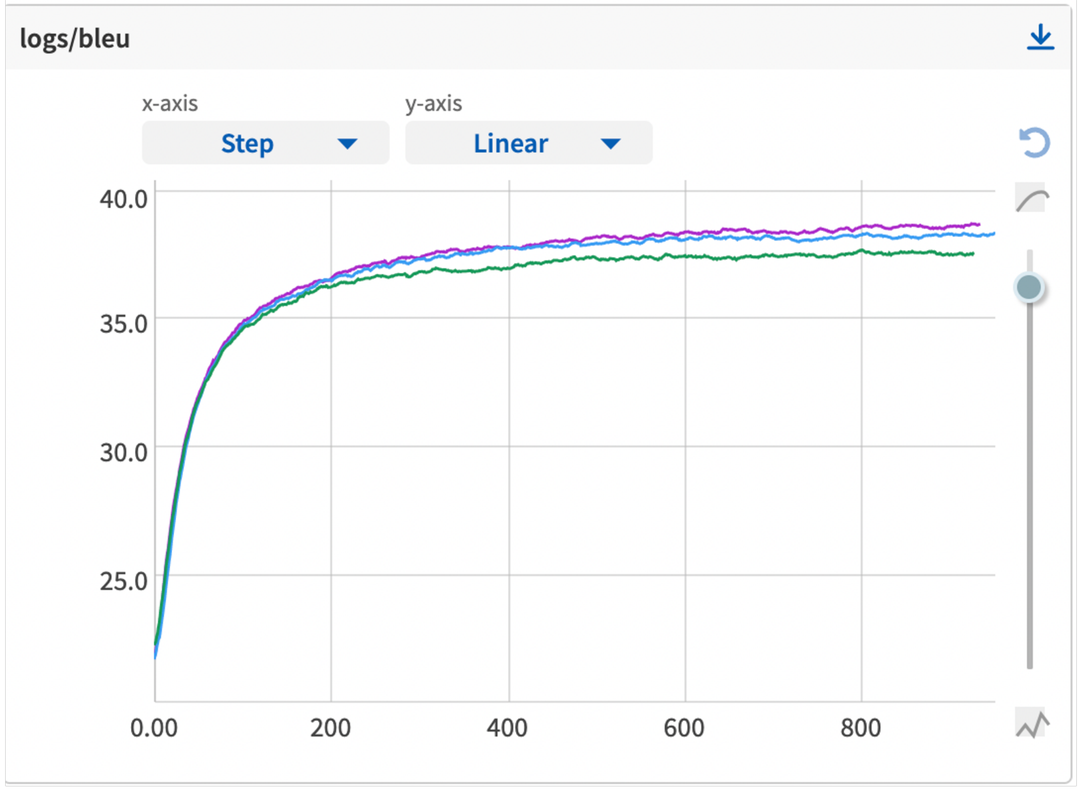
# 피드백 처리 현황

-optimizer

adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

# 피드백 진행 상황

## -optimizer(stop)




adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

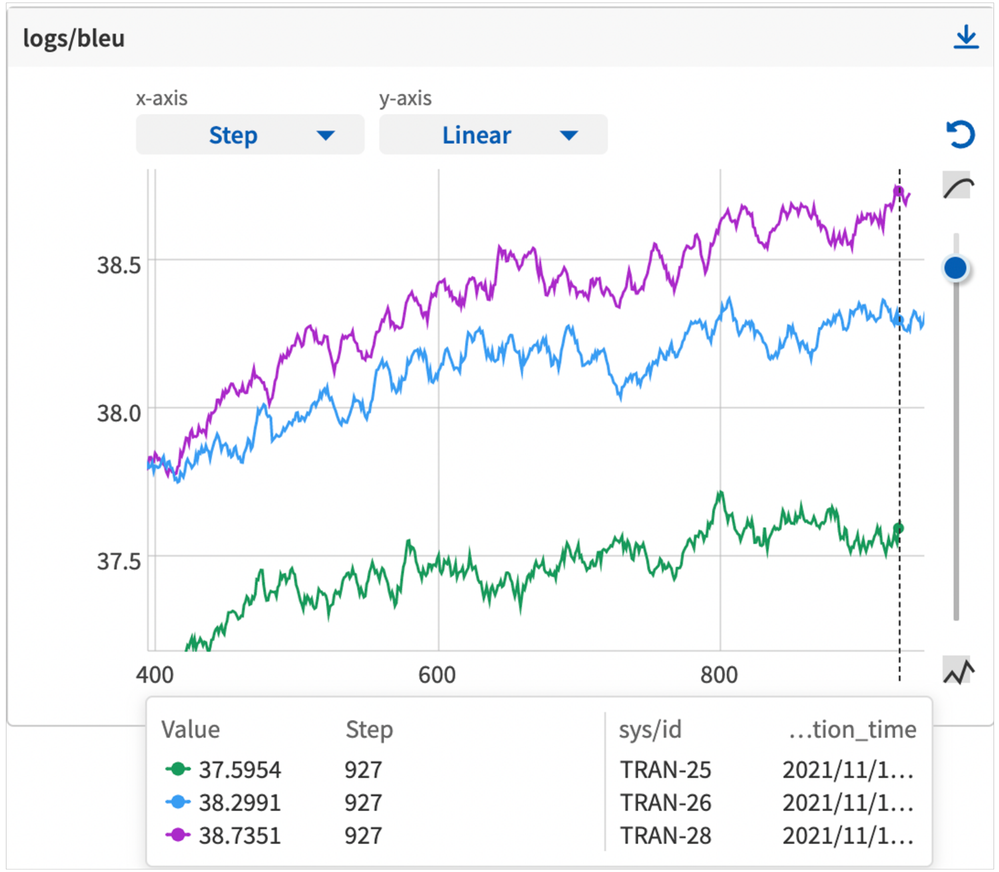
초록 - AdamW

파랑 - Radam

보라 - Adam

# 피드백 진행 상황

## -optimizer(stop)




초록 - AdamW

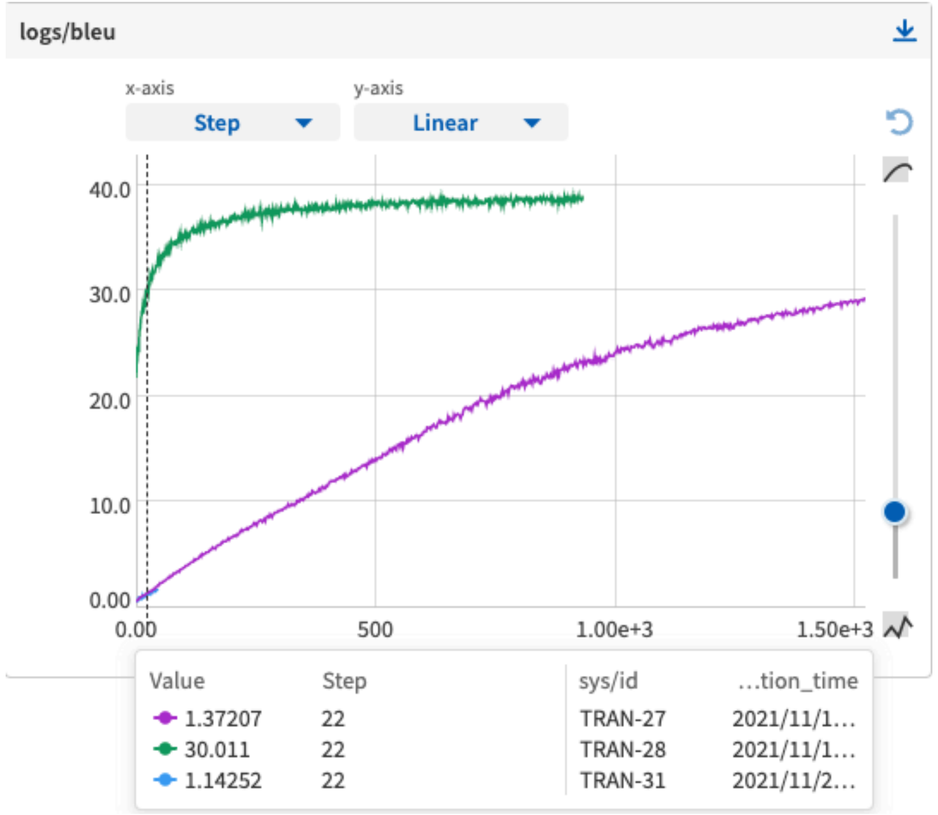
파랑 - Radam

보라 - Adam

adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

# 피드백 진행 상황

## -optimizer(stop)




초록 - Adam 파랑 - SGDP 보라 - SGDW

adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

# 피드백 진행 상황

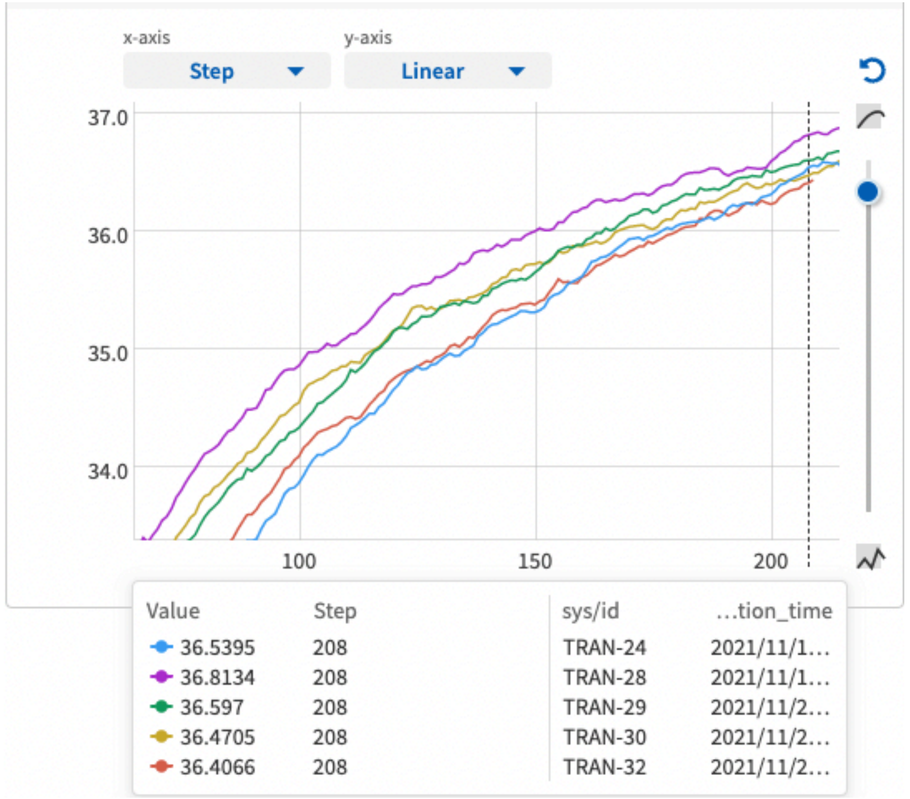
## -optimizer(running)


adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

초록 - AdaBelief  파랑 - AngularGrad  보라 - Adam  노랑 - AdamP  주황 - DiffGrad

# 피드백 진행 상황

## -optimizer(running)

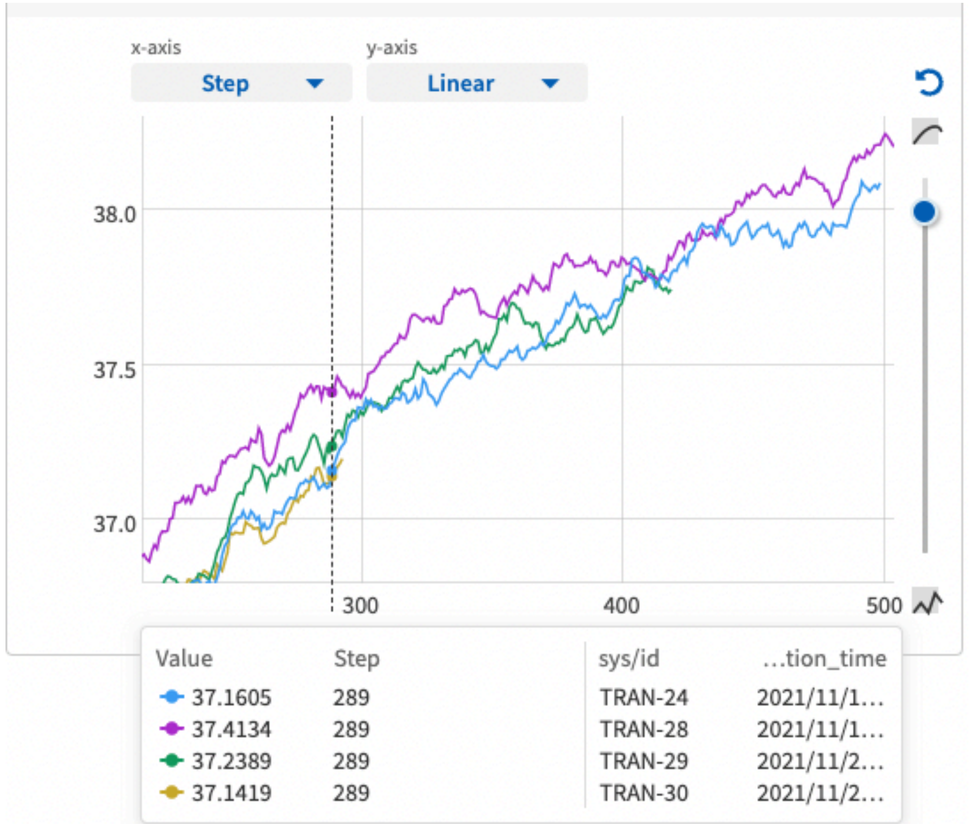



adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

초록 - AdaBelief  파랑 - AngularGrad  보라 - Adam  노랑 - AdamP  주황 - DiffGrad

# 피드백 진행 상황

## -optimizer(running)

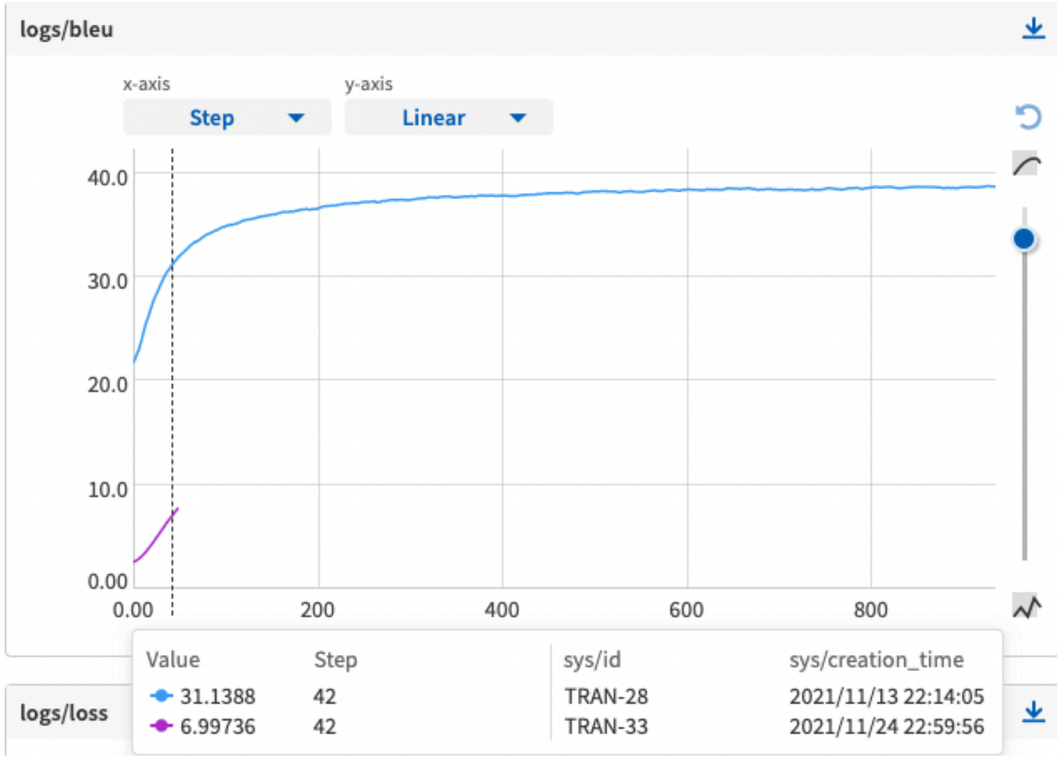



초록 - AdaBelief 파랑 - AngularGrad 보라 - Adam 노랑 - AdamP

adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

# 피드백 진행 상황

## -optimizer(running)




adaBelief - 2020
adamp - 2020
SGDP - 2020
diffgrad - 2019
Lamb - 2019
Radam - 2019
Adamw - 2019
SGDW - 2016
Adam - 2014
AngularGrad

파랑 - Adam  보라 - Lamb

# 피드백 진행 상황

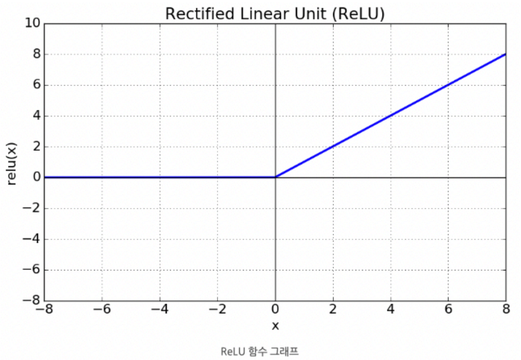
## -activation function

Relu
Softmax
Log_softmax

# 피드백 진행 상황

## -activation function

# Drying Relu



ReLU 함수 그래프

**Mish – 2020**

**Swish - 2017**

**ELU - 2015**

**Maxout - 2013**

**Leaky ReLU**

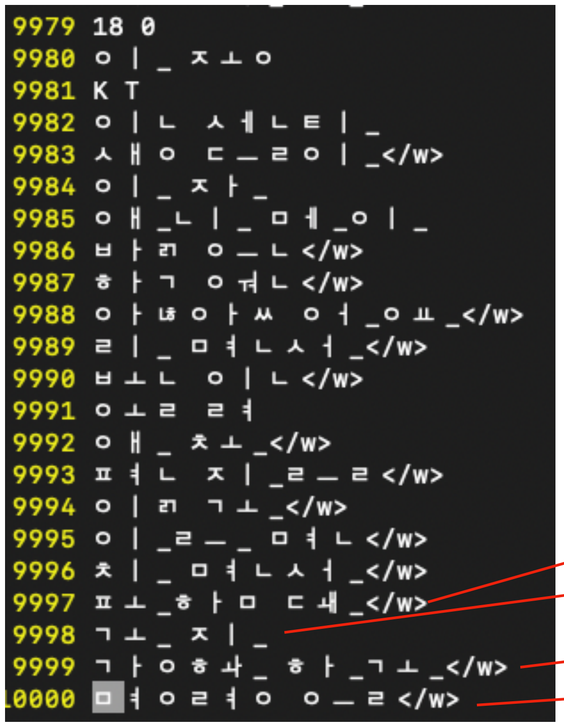
**Parametric ReLU (PReLU)**

# BPE 차이 비교

# BPE 차이 비교

## -BPE 비교(size : 10000)

자모



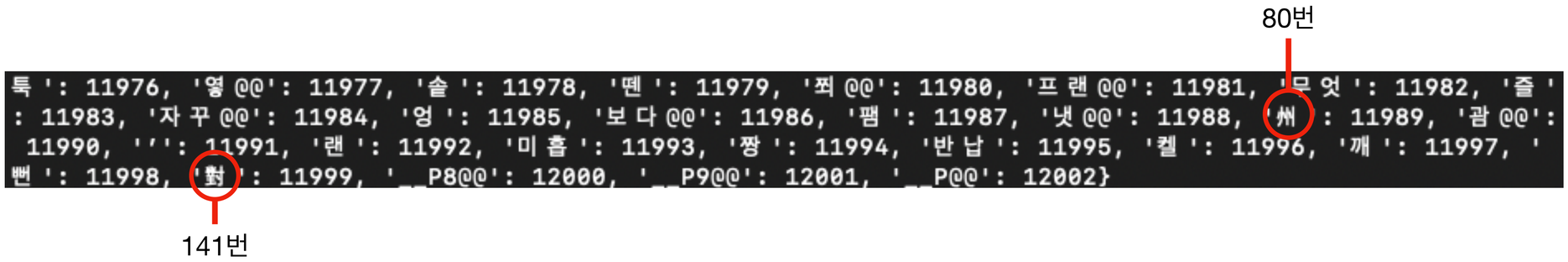

Kr




# BPE 차이 비교

## -vocab 비교(size : 12000)

kr




자모

# 높임말 반말 변환

## 데이터 전처리

# 지난 미팅 후 나왔던 개선 및 추가사항

- 수정하고 추가해야 했던 부분들

1. 약 3000개의 반말, 높임말 구분이 안되는 문장은 어떻게 처리를 할 것인가?

2. 반말-> 높임말 변환의 개선

3. 영어의 격식 표현 처리 이슈

# 3000개의 Noise 문장에 관하여

## - 문장의 종류와 처리 결과

계약에 의하여 발생한 채권으로서 그 이행기간이 경과하였으나 변제되지 아니한 채권
중점 감사분야에 대한 심층 검토 및 현장 동향 파악
운영자의 사정이나 천재지변 등 불가항력으로 시설 사용을 못하거나 강좌개설을 하지 못한 경우
그 밖에 위탁운영을 계속할 수 없는 사유가 발생한 때
자치분권 촉진 및 지원에 관하여 협의회에서 회의에 부의하는 사항
뱀의 무늬와 거북이 등껍질까지 음각선으로 세세하게 표현했다. 구
지역 내 모범모델의 발굴 및 확산 지원
직무태만, 품위손상이나 기타 그 밖의 사유로 인하여 직무수행에 적합하지 않다고 인정되는 경우
1982년 4월 8일 이전에 사실상 건립된 연면적 85㎡이하의 주거용건물로서 1982년 제1차 촬영항공사진에 수록되어 있거나 또는
이전에 건립하였다는 확증이 있는 무허가 건물
고등학교 이상의 학교에서 동물보호 또는 동물복지과목이 포함된 동물관련 분야를 전공한 자
그 밖에 협동조합 생태계 조성에 필요한 시책에 관한 사항
그 밖에 구청장이 센터의 이용을 제한할 필요가 있다고 인정하는 경우: 이용료 등 전액 반환
재활용 가능자원의 분리수거체계 구축과 이에 필요한 장비·인력 확보계획 및 보관용기의 설치
사. 식품영양 및 조리 전문가 각 1명
영 제43조에 따른 보존기간이 경과한 기록물의 보존기간 재책정, 보류 또는 폐기
질병이나 해외여행 등으로 6개월 이상 임무를 수행하기 어려운 경우
「5·18민주유공자예우에 관한 법률」 제4조에 따른 광주민주유공자와 그 유족이나 가족이 신청하는 증명
청원경찰 : 「청원경찰법」에 의하여 채용된 경비업무에 종사하는 상근인력
서울특별시 노원구(이하 "구"라 한다) 지역의 기관·단체들과의 상시 협력체계 구축
비밀(대외비를 포함)이 해제되지 않은 기록물을 열람·대출하고자 하는 경우
지정문화재의 보존 및 관리 등을 위한 효율적인 방안 심의
공사 전·후 보행환경 및 교통사고 예방에 관한 사항
「국민건강보험법」에 따른 국민건강보험료를 최근 6개월 이상 체납으로 인하여 생계가 어려운 경우
「지방공무원 임용령」(이하 "영"이라 한다) 제21조의2에 따른 견습직원의 견습근무 기간 종료 후의 임용
「상훈법」에 의한 상순위의 유공서훈을 받은 통장의 자녀
치료비: 재난으로 부상당한 사람의 치료에 소요되는 통상적 비용
그 밖에 장애인의 건강증진 및 장애인 건강보건관리를 위하여 필요한 사항
법 제8조 제1항의 규정에 의한 규제의 존속기한
학계 및 교육계에서 인권 관련 연구 및 경험이 있는 사람
노인복지법시행령 제11조 규정에 의한 노인의 날 등 행사참여자
청년 고용 확대를 위한 일자리 창출 사업
제2항에 따라 지원되는 정보통신제품 및 정보통신서비스 사용에 따른 통신비 등 운영비의 일부 또는 전부 지원
의료기사·의무기록사 및 안경사 지도 등에 관한 사항
해당 동에 소재한 각급 학교, 기관, 단체에 속한 사람

"이전에 건립하였다는 확증이 있는 무허가 건물"

**Noise로 그대로 corpus에 두기로 결정**

# 반말 -> 높임말 개선사항

## - 문장의 예외적인 경우 처리 및 사전 개선

### 예외처리 사전의 도입

```
EXC_4_deal_2 = [

    ['ㄴㅣ',['ㅆ','ㅇㅓㅇㅛ'],'ㅇㅛ']
]

EXC_4_deal_3 = [

    ['ㄴㅡㄴㄱㅏ',['ㅗ','ㅗ','ㅏㅇㅛ'],['ㅜ','ㅜ','ㅓㅇㅛ'],['ㅏ','ㅏ','ㅣㅇㅛ'],'ㅇㅓㅇㅛ'],
    ['ㄷㅏ',['ㅆ','ㅆ','ㅅㅡㅂㄴㅣㄷㅏ'],['ㄴ','ㅂ','ㄴㅣㄷㅏ'],['ㄹ','ㅂ','ㄴㅣㄷㅏ'],'ㅂㄴㅣㄷㅏ']
]

EXC_4_deal_4 = [

    ['ㄱㅔ','ㄱㅔㅇㅛ','ㄱㅔ'],
    ['ㄷㅔ','ㄷㅔㅇㅛ','ㄷㅔ'],
    ['ㄱㅓㄹ','ㄱㅓㄹㅇㅛ','ㄱㅓㄹ']
]
```

~다

앞 글자 종성에 'ㅆ'이 있다면 '습니다'로 변경

시험을 봤다 => 시험을 봤습니다.

앞 글자 종성에 'ㄴ'이 있다면 'ㅂ니다'로 변경

도자기를 만든다. => 도자기를 만듭니다.

•
•
•

형태의 변화가 많은 형태소에 대해서는 예외적인 경우를 두어서 처리

# 반말 -> 높임말 개선사항

## - 문장의 예외적인 경우 처리 및 사전 개선

예외사항에 대한 구별

```
def isExcept(self, input):
    if input[1]=='special':
        return 1
    elif input[1]=='special-':
        return 2
    elif input[1]=='-special':
        return 3
    elif input[1]=='specialx':
        return 4
    else:
        return 0
```

1. -세- 와 같이 변칙성이 무작위인 경우
2. 앞에오는 것을 신경써야하지만, 같은 형태소 안에서 해결이 되는 경우
3. 앞에 오는 것을 신경써야하며 하나의 형태소 안에서 해결이 안되는 경우
4. 같은 형태소의 같은 단어이지만, 문장의 특정 위치에서만 변환이 이루어져야 하는 경우

# 영어 격식 표현을 고려한 변환

## - 영어 격식 표현에 대한 논문을 찾아봤다.

### 참고 논문

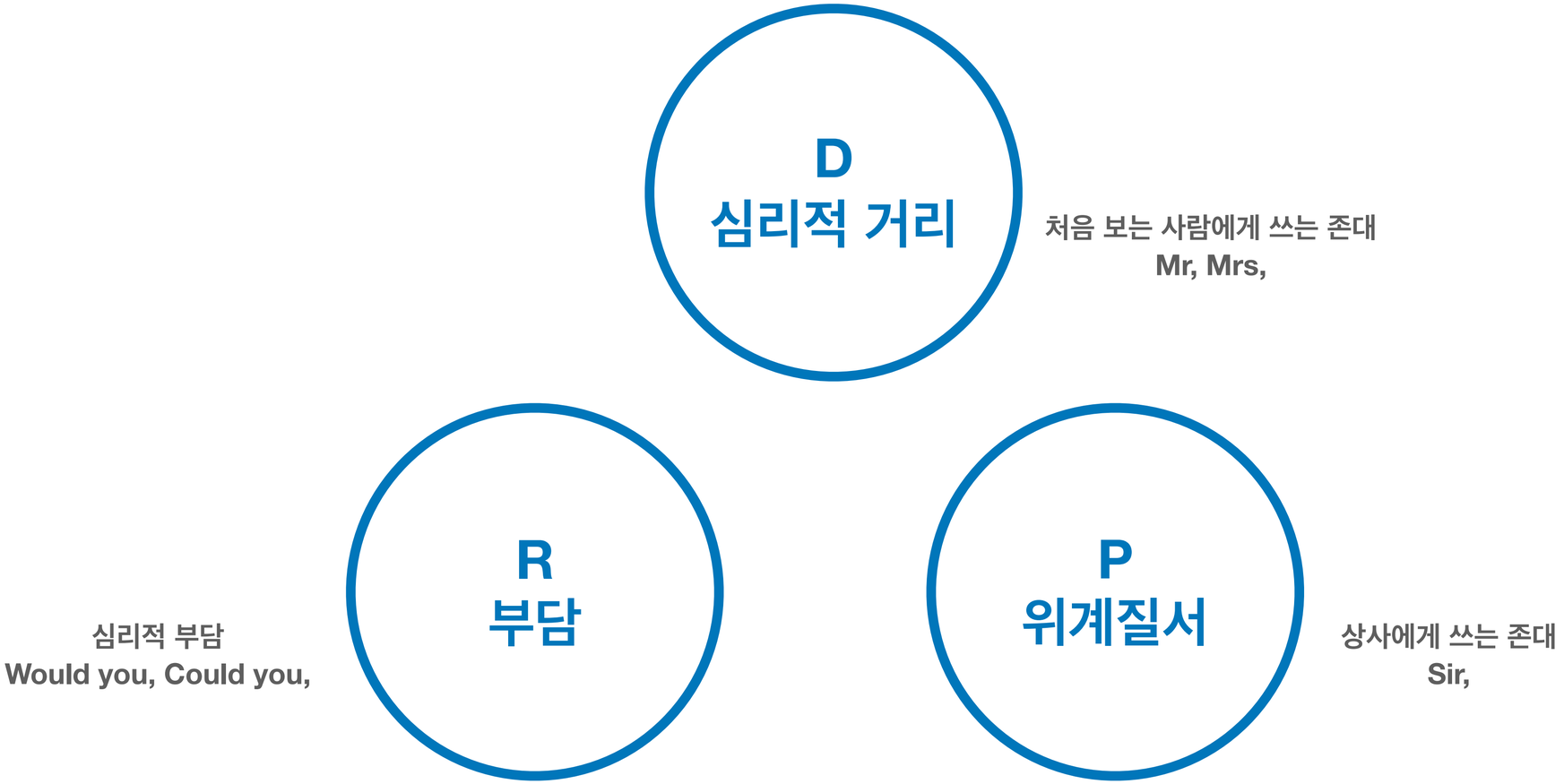
예를 들어, 'Mr' 'Mrs' 'Miss', 'James', 'Jim', 'Jimmy'와 같이 구분적 호칭이 있는 영어를 사용하는 사람은 의사소통에서 상대방과의 높고 낮음의 거리나 친소의 거리를, 이에 따라 구분하고 인식할 것이다. 연령이나 지위 등에 의한 존대 체계를 가진 한국어를 사용하는 사람은 연령이나 지위의 작은 차이라도 상대방과의 거리로 인식하고 구분하며, 이를 존대 체계에 따라 표현한다.

이원국 영어 공손의 거리 : 근접과 격원, 새한영어영문학회, 2005, 149 - 179

https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01057228

# 영어 격식 표현을 고려한 변환

## - 영어 격식 표현의 유형



# 영어 격식 표현을 고려한 변환

## - 영어 변환 시도

영어의 P type과 D type의 경우 한국어의 높임말과 상황이 맞아떨어지기 때문에 단순 치환으로 변환이 가능

```
convert_low_en('Mr. Kim invented gibbson\'s method')
```

```
" Kim invented gibbson's method"
```

# 영어 격식 표현을 고려한 변환

## - 영어 변환 시도

R
부담

a. Please open the window, Jen.
b. Do you think you could find the time to take those invitations to the printers, Jen?
c. Would it be all right if I left Honey with you on Wednesday morning- I'd like to shopping by myself for a change.

?