

# 2021년 봄학기 연구계획 메신저 채팅 대화 특성에 관한 연구

---

ALDE Lab.

이 다 영

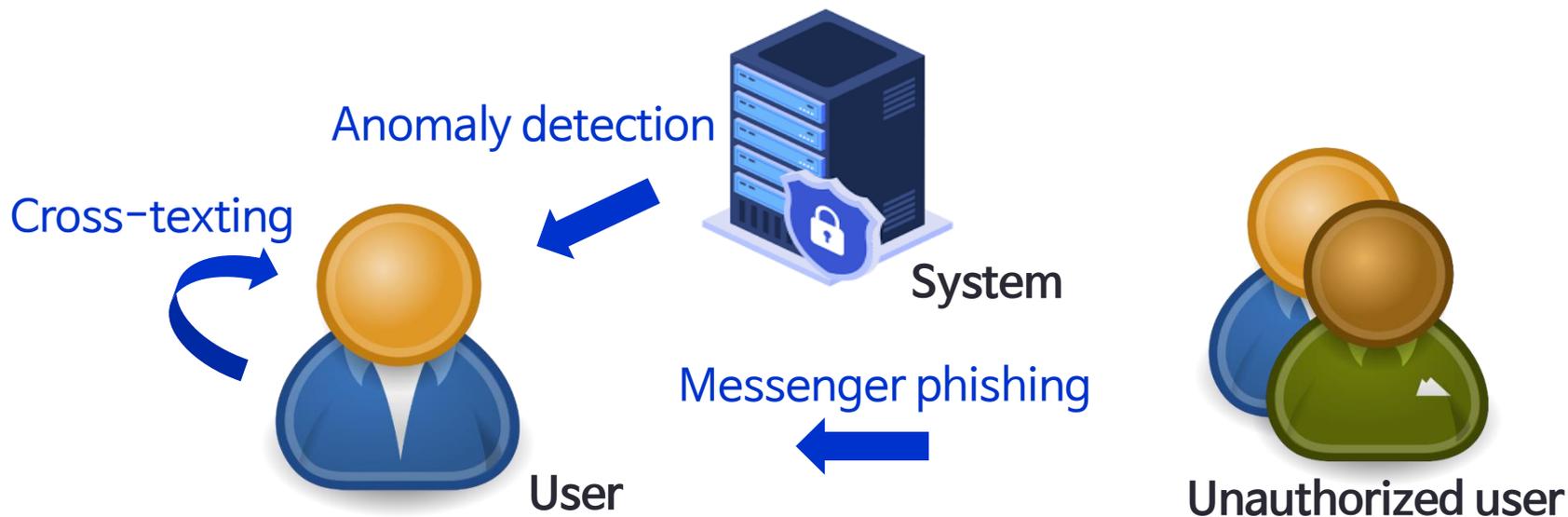
schematique@pusan.ac.kr

# 1. 연구 목표

- 메신저 채팅 대화에서 말투에 해당하는 특성 정의
  - 타인과 구별할 수 있는 특성
  - 대화 주제의 변화에도 유지되는 특성
  - 특정 관계/상황일 때 공통되는 특성
- 말투를 추출해 낼 수 있는 모델 고안
  - 한글 채팅 문장에 적합한 분석단위, 임베딩 방법
  - 채팅 데이터 구조에 적합한 분석 모델 고안

## 2. 응용분야

- Cross-texting detection
  - 사용자가 의도하지 않은 상대방에게 실수로 잘못 메시지 전송
- Phishing message detection
  - 상대방이 사용자의 개인정보를 부정하게 얻기 위해 메시지 전송
- Anomaly detection
  - 사용자가 평소와 다르게 메시지 전송



# 3.1 관련연구 : 메신저 대화-정성적 연구

- SNS 메신저 ‘카카오톡’ 언어현상 연구(2012, 최명원)
  - 말하는 것과 가까운 속도로 글로 표현하기 위한 축약/변형된 형태의 언어
  - 소리나는대로 표기
  - 감정, 상태, 동작을 전달하기 위한 표현
  - 명사형으로 끝나는 문체 사용, 띄어쓰기 생략
- 모바일 메신저 대화에서 나타나는 공손 전략(2014, 윤상석)
  - 표현의 다양한 형태 : 문어체, 구어체 특성 + 온라인 언어의 특성
  - 친근감, 동질감의 표현 : 문장부호, 자음, 어미 변형, 이모티콘 등
  - 자유로운 대화 형태 : 대화의 시작/끝, 순서, 응답시간이 음성 대화에 비해 제약 적음

>> 채팅 문장 표현은 정형화 된 특성이 존재하지 않고, 개인/상황에 따라 다양한 형태로 나타날 수 있음

>> 변형된 표현, 불분명한 문장 경계 등 기존의 문장 분석 모델을 바로 적용하기 어려울 것으로 예상됨

## 3.2 관련연구 : 채팅 대화 주제/사용자 식별

- 인스턴트 메시징에서의 대화 주제 및 주제 전환 탐지(2008, 최윤정)
  - 사용자, 상대방의 발화에서 키워드 추출, 어느 주제에 속하는지 계산
  - 이전 대화 이력에서 키워드 추출, 어느 주제에 속하는지 계산해 결과 보정
  - 현재 대화와 이전 대화의 키워드 비교해 주제 유지/전환 탐지
- 메신저 피싱 대화 분석을 통한 탐지 방안(2019, 배지효)
  - 피싱 대화 데이터 형태소 분석 후 명사 추출
  - 피싱 대화의 진행 4단계 분류(유대감 생성 > 의심 해소 > 부탁의 이유 > 금전적 부탁)
  - 메시지에 사용된 단어들에 대해 점수를 계산해 피싱으로 탐지하는 방안 제안
- Transformer 모델을 이용한 채팅 사용자 식별(2020, 박윤보)
  - 1:1 대화 내역(불용어 포함) 형태소 분석 후 Transformer로 학습해 사용자 식별

>> (특정) 주제 탐지를 위해서는 추출한 명사(키워드)를 분석, 사용자 식별을 위해서는 키워드 외에 불용어도 포함시켜 분석

## 3.3 관련연구 : 한국어 문체/어체 관련 연구

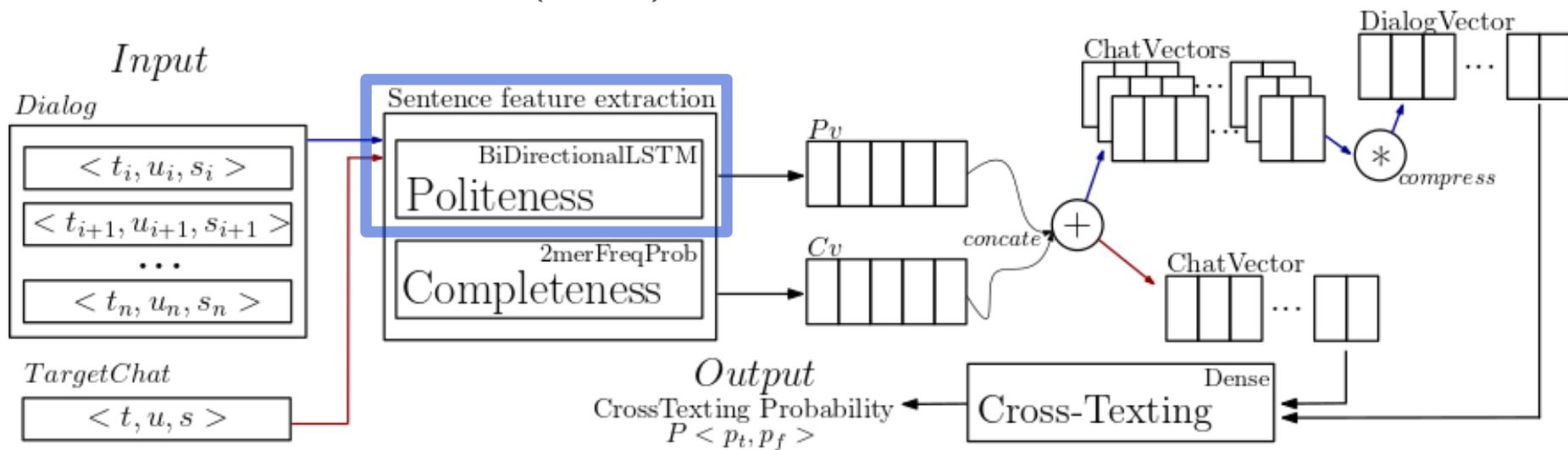
- Attention-Based Seq2Seq 모델을 이용한 한국어 어체 변환(2018, 홍태석)
  - 규칙을 사용하지 않고 임의 길이 해요체 어절을 합쇼체 어절로 변환하는 모델 제안
  - 문장의 어체를 결정하는 마지막 어절을 음절 단위로 입력, 해요체-합쇼체 pair 데이터로 학습
- 한국어 문체 변환을 위한 자소 기반 주의 집중 Seq2Seq 모델(2019, 김보은)
  - 위 2018 논문에서 입력 단위를 자소 단위(음절당 길이3 고정)로 변경해 성능 개선
- TGST : 트랜스포머 기반의 한국어 생성적 스타일 변환(2020, 박다솔)
  - 일반 문장을 낮춤말 문체로 변환한 모델 제안
  - 위 2018, 2019 논문과 차이
    - 대용량의 일반 문장으로 사전 학습된 모델 위에 문체 변환 모델 학습 진행
    - 마지막 어절만 보는게 아니라 문장 전체를 학습
    - 변환된 결과도 문장 곳곳에 단어의 치환, 표현 변경이 나타남

>> 어체를 결정짓는 요소가 주로 문장의 끝에 위치한다고 가정하고 있음

>> 어체/문체 관련 국내 연구는 주어진 문체 pair 데이터를 이용한 변환 위주, 문체를 단서로 사용자 식별로 진행된 연구는 부족

# 3.4 이전 연구: 형식적 특성 기반 Cross-texting

- How to Prevent Cross-texting in Mobile Messengers Using Honorifics Features in Korean Text(2021)



- 포털/웹 커뮤니티 댓글, AI-hub 문장으로 학습한 LSTM 기반 문장 5등급 분류
- 사용자가 5가지 말투 중 어떤 것을 얼마나 사용했는지를 학습
  - 채팅 문장 -> (판단) 문장 타입 -> (판단) 대화 타입
  - 채팅 문장 -> (판단) 문장 타입 -> (판단) 크로스 텍스트링 여부

# 3.4 이전 연구: 형식적 특성 기반 Cross-texting

## • 한계점

- 높임등급 - 데이터 출처 의존적, 노이즈 포함
  - 데이터 규모상 모든 문장을 수동으로 분류할 수가 없음
  - 데이터 출처와 실제 높임 문장 등급이 일치하지 않는 경우 존재
- Cross-texting case – 가상의 상황을 기반으로 생성된 데이터
  - 실제로 사용자가 실수한 데이터가 아니라 문장을 교차로 섞은 가상의 데이터
  - 실제 상황이라면 가능했을 분석 요소들을 활용하지 못함
    - 실제 동일한 사용자의 메시지
    - 메시지 전송 시간
    - 응답시간
- 5가지 유형에 해당하지 않는 채팅 대화 형태 존재
  - LSTM 학습에 사용된 문장들이 실제 채팅 대화 유형을 충분히 커버하지 못함
    - 변형된 높임 문장, 이모티콘이 포함된 높임 문장

## 3.5 소결론

- 채팅 대화에서 말투를 추출하기 위한 접근
  - 불용어도 분석 범위에 포함
  - 내용에 독립적인 말투 분석을 위해 형태소 분석을 진행하지 않고 음절 단위로 분석
  - 문장의 경계가 정확하지 않는 특징 고려
  - 말투는 술어부에서 주로 나타남을 고려
- 이전 진행 연구의 한계점 보완
  - 채팅 문장이 특정 유형에 속하는지를 분석하지 않고, 문장으로부터 특성을 추출하는 방법으로 접근

# 4.1 사전실험 - 말투만 보고 알 수 있을까?

- 메시지에서부터 말투를 추출하는 모델 만들기 전 사전 실험 진행
- 가설
  - 말투는 주제와 무관하게 계속해서 반복되므로 빈도가 높을 것
  - 말투에 해당하는 글자는 문장의 끝에 위치할 것
  - 말투에 해당하는 음절의 구성요소로 특정 화자 또는 상대방을 구별할 수 있을 것
- 실험 방법
  - 연구실에서 발표자와 {교수님, 연구실원들}의 대화 중 발표자의 메시지만 추출
  - Tr 80% + Te 20% 분리 후, 각 Tr, Te의 일치 비율을 아래와 같이 계산
  - 누적 빈도 단어 확인 및 일치하는 비율(Jaccard metric) 계산
  - 1) 두 대화의 각 화자 A, B로부터 추출한 상위 누적 빈도 단어의 일치비율
  - 2) 상위 누적 빈도 단어의 빈도 비율의 일치비율

$$1) \frac{|W_A \cap W_B|}{|W_A \cup W_B|} \quad \rightarrow \quad 2) \frac{\sum_w^{(W_A \cap W_B)} \min(\text{freq\_rate}(w_A), \text{freq\_rate}(w_B))}{\sum_w^{(W_A \cup W_B)} \max(\text{freq\_rate}(w_A), \text{freq\_rate}(w_B))}$$

# 말투만 보고 알 수 있을까? [대화1]

{교수님, **선배1**, 선배2, 후배1}

```

2018/8/02(목)오후08:04:55 보 습니다 ~
2018/8/03(금)오전11:00:50 서 보내드립니다 ~
2018/8/06(월)오후03:04:35 보내드립니다 ~
2018/8/06(월)오후03:05:35 보 습니다
2018/8/07(화)오후07:01:41 서드립니다
2018/8/13(월)오후05:39:42 보내드립니다 ~
2018/8/14(화)오후08:13:29 보 습니다 ~
2018/8/14(화)오후09:05:54 니다
2018/8/21(화)오전10:01:29 보내드립니다 ~
2018/8/21(화)오전10:02:12 니다
2018/8/23(목)오후01:32:53 서 보내드립니다 ~
2018/8/29(수)오전10:13:36 보내드립니다 ~
2018/8/29(수)오전10:13:55 니다
2018/8/30(목)오후12:33:18
2018/8/31(금)오전09:36:52 서
2018/9/04(화)오전11:06:33 니다
2018/9/04(화)오전11:06:39 보내드립니다 ~
2018/9/07(금)오전09:59:35 서 습니다
2018/9/10(월)오후04:31:39 보내드립니다 ~
2018/9/10(월)오후04:32:11 니다
2018/9/11(화)오후05:38:41 니다
2018/9/17(월)오후03:31:12 보내드립니다 ~
2018/9/17(월)오후03:33:18 니다
2018/9/17(월)오후03:34:05 드 습니다
2018/9/17(월)오후03:36:51 다 보내드립니다 ~
2018/9/27(목)오후05:24:51 보내드립니다
2018/9/27(목)오후05:25:31 내 드
2018/9/27(목)오후05:25:58 습니다
2018/10/01(월)오후02:24:19
2018/10/01(월)오후02:25:53
2018/10/01(월)오후02:26:04 보 보내드립니다 ~
2018/10/01(월)오후03:03:26 보내드립니다 ~
2018/10/05(금)오전10:12:23 서
2018/10/08(월)오후03:19:28 보내드립니다 ~
2018/10/12(금)오전10:08:39 서 습니다
2018/10/12(금)오전10:10:18 내 니다
2018/10/12(금)오전10:13:16
2018/10/12(금)오후04:56:15
2018/10/12(금)오후04:56:28
2018/10/15(월)오후03:54:11 보내드립니다 ~
2018/10/19(금)오전08:29:43 서
2018/10/19(금)오전08:29:53 습니다
2018/10/19(금)오전09:43:51
2018/10/19(금)오전10:39:27 니다
2018/10/22(월)오후03:41:58 다
2018/10/22(월)오후03:42:07 니다
2018/10/22(월)오후03:57:48
2018/10/22(월)오후04:46:49 보내드립니다 ~
2018/10/29(월)오후05:56:24 보내드립니다 ~
2018/11/05(월)오후05:34:20 보내드립니다
2018/11/06(화)오전11:53:43 서드립니다 습니다
2018/11/06(화)오후05:46:12 보내드립니다 ~

```

상위 누적 빈도 단어만 표시  
회색명암 : 하위 빈도 단어, 호칭에 관한 내용

# 말투만 보고 알 수 있을까? [대화2]

{교수님, 선배1, 선배2, 후배1}



상위 누적 빈도 단어만 표시  
회색명암 : 하위 빈도 단어, 호칭에 관한 내용

# 말투만 보고 알 수 있을까? [대화3]

{교수님, 선배1, 선배2, 후배1}

2018/12/27(목)오후05:23:12 // . . . / / / . . . / / /

2018/12/27(목)오후05:23:24 이 // 에 서 주 가 . . . / / / 면 . . . 주 시 면 니 다

2018/12/27(목)오후05:23:25 // . . . . . ? . . . . .

2019/1/16(수)오전10:33:46 아 , . . . 니 다 ~

2019/2/12(화)오후05:26:41 // . . . / / / . . . ?

2019/4/24(수)오후07:21:39 // . . . . . /

2019/4/24(수)오후07:24:07 , 사 니 다

2019/4/30(화)오후03:59:26 , . . . 거 아 서 |

2019/4/30(화)오후03:59:35 서 에 . . . 어 주 요

2019/4/30(화)오후03:59:39 아 이 주 요 ~

2019/4/30(화)오후04:06:12 . . . . .

2019/4/30(화)오후04:06:29 아 가 고 , 이 서

2019/5/14(화)오후06:26:05 , 사 니 다

2019/5/14(화)오후06:27:05 내 일 . . . 해 니

2019/5/16(목)오후03:38:27 , . . . 내 하 고 하 는 요 ~

2019/5/16(목)오후03:38:36 , . . . 면 주 요

2019/5/16(목)오후03:40:14 , . . . 니 다

2019/5/20(월)오후01:51:59 // . . . . . /

2019/5/22(수)오전11:10:53 , . . . 일 일 거 아 요

2019/5/22(수)오전11:11:20 , . . . 하 시 는 이 이 어 가 , 가 다 고 하 요 ||

2019/5/22(수)오후02:09:47 에 는 가 요 | . . . 가 로 가 요 ?

2019/5/22(수)오후02:10:03 아 는 가 , 가 로 가 요 ?

2019/5/22(수)오후02:12:01 아 가 다 에 이 주 요 ~

2019/5/29(수)오후03:57:50 시 , 로 이 주 요 ?

2019/5/29(수)오후03:58:25 , 요 일 서 일 내 주 요 ?

2019/5/31(금)오후03:06:04 사 니 다 ~

2019/6/07(금)오후04:45:58 , . . . 서

2019/6/07(금)오후04:46:04 , 다 에 서 하 시 고 ,

2019/6/07(금)오후04:46:16 . . . 에 해 서 시 면 니 다

2019/6/07(금)오후05:27:27 주 로 어 가 서 시 면 고 ,

2019/6/07(금)오후05:27:42 다 니 다

2019/6/11(화)오후09:33:54 요 가 요 ?

2019/6/11(화)오후09:34:04 , 로 해 니 다

2019/6/12(수)오후03:10:58 에 면 로 해 서 주 요 ~

2019/6/12(수)오후03:11:05 , 가 어 에 거 어 요 ||

2019/6/13(목)오후03:59:45 아 니 시 요 에 거 어 요 ||

2019/6/13(목)오후04:00:29 아 사 니 다 어 로 다 고 요 |

2019/6/13(목)오후05:05:15 아 . 이 고 하 ~ 해 가 요 ?

2019/6/13(목)오후05:12:08 에 는 이 해 가 요 ?

2019/6/13(목)오후05:21:46 가 니 다 요

2019/6/13(목)오후05:23:01 가 니 다 요

2019/6/13(목)오후05:23:45 이 서 니 다

2019/6/13(목)오후05:33:14 해 서 는 는 거 아 요 ~

2019/6/13(목)오후06:06:27

상위 누적 빈도 단어만 표시  
회색명암 : 하위 빈도 단어, 호칭에 관한 내용



## 4.2 사전 실험 결과

### • 가설확인

- 상위 누적 빈도를 차지하는 단어가 주로 말투와 연관이 있어 보임
- 말투에 해당하는 글자가 반드시 문장의 끝에 오는 것은 아님
- 자주 쓰인 표현으로도 상대방이 누군지 유추가 가능했음

### • 실험결과

- 표1) 2)는 상위 누적빈도 40% 단어에 대해 Jaccard 1), 2) 계산결과
- 1) 의 방법이 정답 쌍을 더 정확하게 찾아냈지만
- 2)의 방법이 1)보다 정답에 대해 더 높은 일치도 결과 도출
- $P=0.3$  이하에서는 해당하는 단어 개수의 부족(상위 단어가 차지하는 비중이 높음)으로 판별 성능이 낮았음
- $P=0.5$ 부터  $0.9$ 까지는 판별 성능이 점차 떨어짐
- 이 실험에선 호칭에 관한 정보가 지배적, 호칭이 없는 데이터를 처리하려면 더 정교한 모델 필요

P=0.4	교수님	선배1	후배1	선배2
교수님	0.875	0.6	0.16	0.3
선배1	0.444	0.75	0.125	0.333
후배1	0.182	0.174	0.25	0.233
선배2	0.294	0.533	0.2	0.65

1) Tr-Te 상위 누적 빈도 단어 일치도

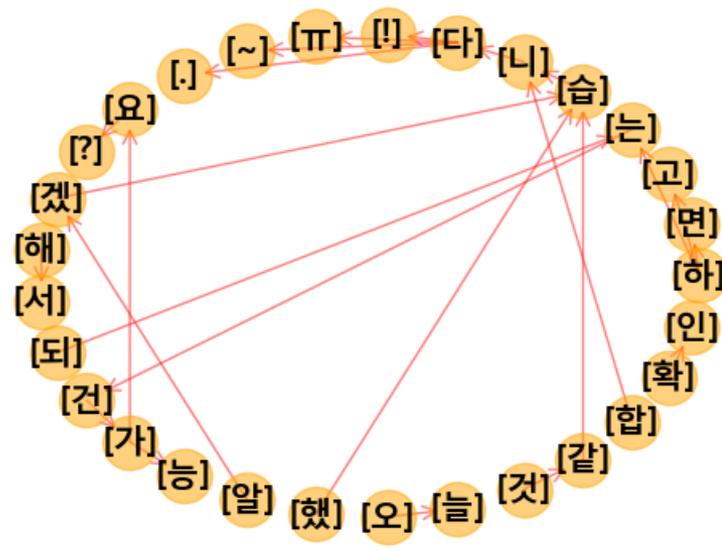
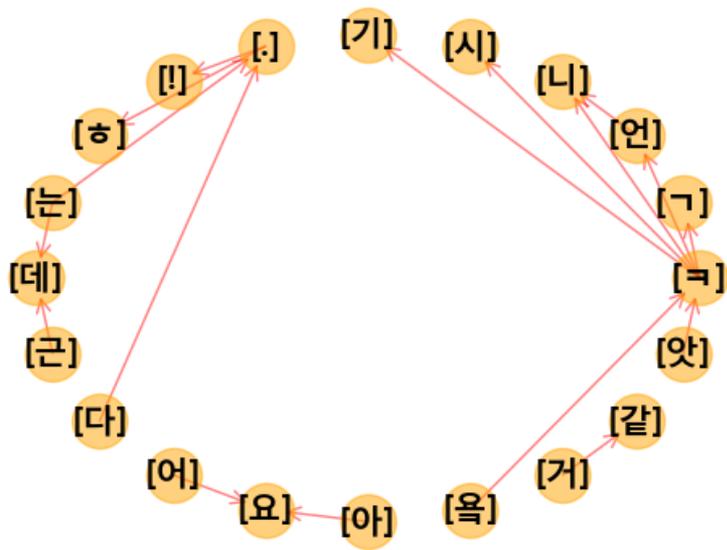
P=0.4	교수님	선배1	후배1	선배2
교수님	0.903	0.554	0.273	0.436
선배1	0.488	0.824	0.232	0.372
후배1	0.382	0.389	0.423	0.432
선배2	0.504	0.554	0.341	0.754

2) Tr-Te 상위 누적 빈도 단어 빈도비율 일치도

# 4.3 다음 구상 중인 모델(1)

## Transition Distance

- 음절과 음절 다음에 나타날 확률 그래프
- 모르는 문장을 그래프 상에서 이동시키면서 weight의 합을 구함
  - 인접한 두 음절이 그래프 상에 없는 경우
  - 그래프에 있지만 edge가 없는 경우
    - Shortest path
    - Lowest hop path



## 4.4 다음 구상 중인 모델 (2)

- 채팅 데이터를 음절 단위로 학습한 임베딩 모델
- 각 음절의 co-occurrence 정보로 그래프 구축
- 그래프 임베딩 ( 화자의 채팅 말투 특성 )
- 그래프 유사도 계산, 그래프 클러스터링

# 5. 진행 사항 및 이후 계획

## • 진행 사항

- 한 명의 유저가 서로 다른 (최소 3명) 상대와 대화한 데이터 확보
  - 실제 동일인의 데이터로 Cross-texting 데이터 생성
- 누적 상위 빈도 단어로 개별 dialogue 구분이 일부 가능함을 확인
- 그래프 임베딩 관련 사전 연구 조사
  - Graph2vec(2017)
- 그래프 기반 텍스트 분석 관련 사전 연구 조사
  - Graph Convolutional networks for Text Classification(2019)

## • 이후 진행 계획

- 다양한 방법으로 음절 단위 그래프 생성
- 채팅 대화 문장에 적합한 임베딩 방법 고안

감사합니다  
QnA