

2023년 1월 호

국립국어원 온라인 소식지 《쉼표, 마침표.》



## 2023년 1월호

## 목차

국어 배우기_뭉치가 알려주는 국어 말뭉치	
대화 잘하는 인공 지능을 만들기 위한 지식구축 기본 원칙	3쪽
국어 배우기_문장 다듬기	
실제 공문서의 문장을 다듬어 봐요!	···· 15쪽
국어 알리기_살아 숨 쉬는 지역어	
문학 작품 속 평안북도 방언	···· 17쪽
국어 알리기_이음이가 전하는 국어 소식	
국민이 고른 가장 적절하게 다듬은 말은 '열린 쉼터'	···· 19쪽
국어 바라보기_닮은 말,다른 뜻	
'바치다'와 '받치다'	 ···· 23쪽
국어로 함께하기	
우리말 풀기	 ···· 53쪽

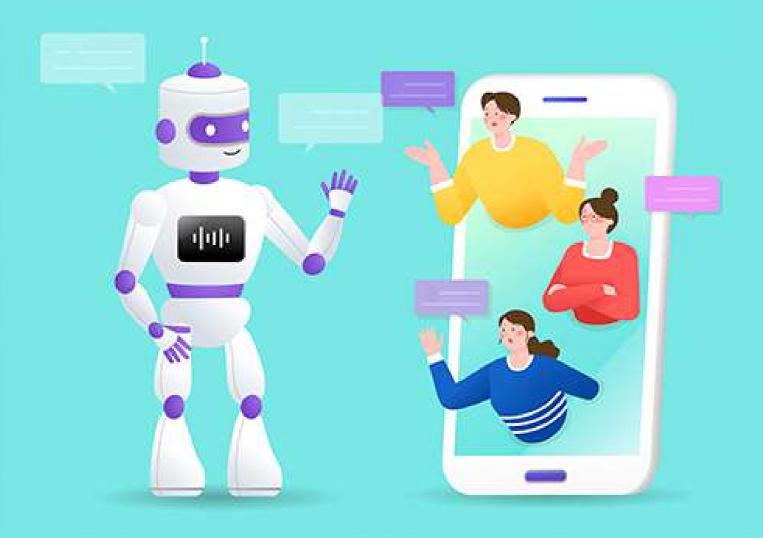
44.36

자기와 자신과 자기 자신	 25≊
국어 배우기_문장 다듬기 실제 공문서의 문장을 다듬어 봐요!	 29쪽
국어 알리기_살아 숨쉬는 지역어 문학 작품 속 경상도 방언	 31쪽
국어 알리기_이음이가 전하는 국어 소식 '한국어-외국어 병렬 말뭉치 구축 쟁점과 활용 방안' 워크숍 개최	 33≊
국어 바라보기_한국어 교육, 그리고 우리 고맙습니다	 35쪽
국어 바라보기_우리말, 그리고 사람 우리 일상 속에서 살아 숨 쉬는 사전을 꿈꾸며	 39쪽
국어로 함께하기	57 <b>조</b>

### 뭉치가 알려주는 국어 말뭉치

# 대화 잘하는 인공 지능을 만들기 위한 지식구축 기본 원칙

(주)와이즈넛 에이비엠(ABM)사업부 김정민 이사



### ○ 우리 생활 속 다양한 인공 지능

생활 속에서 나의 말에 귀 기울이는 존재들이 정말 많다. '사람'이야기가 아니다. 주위를 돌아보면 최신 자동차에서, 텔레비전에서, 스피커에서, 무심코 귀에 착용한 이어폰에서, 냉장고와 같은 가전기기에 탑재된 이른바 인공 지능 비서들이 사용자인 우리의 말을 듣기 위해 귀를 기울이고 있다.



아마존의 가정용 에이아이(AI) 로봇 아스트로(Astro)(사진 출처: 아마존 공식 누리집)

또, 스마트폰에 깔려 있는 금융이나 택배 등 다양한 서비스 앱을 켜면 이제는 챗봇이 상담을 기다리고 있고, 최근 몇 년간 콜센터 상담 또한 인공 지능 상담으로 대체되고 있다. 이제는 단순한 질문에 대한 답변, 통지나 안내를 위한 통화는 인공 지능으로 대체하고 '사람 상담원'은 조금 더 심화된 상담에 집중하는 식으로 역할이 구분되고 있다. 제반 기술의 고도화와 가용 인력의 부족, 24시간 365일 운영 요구로 인해 유통시장 역시 무인화되어 가고 있으며, 무인화의 중심에는 에이아이 비서가 있다.

뭉치가 알려주는 국어 말뭉치



세븐일레븐 디티(DT) 랩 스토어, 아이스 지오(GO)24 에이아이 무인 주류 판매기(사진 출처: AI 타임스, 2021)

그러나, 실제 인공 지능 비서들에게 조금 복잡한 문장 형태로 질의를 하게 되면 잘 이해를 하지 못하거나 엉뚱한 답변을 하기도 하여 사용자들이 당황하는 경우가 종종 발생하게 된다. 특정 인공 지능 비서 시스템에서 질문이력을 분석해 보면 업무 처리 시스템임에도 일상적인 질문을 던지는 경우가 발생하는데, 이는 비서와의 대화시작 시 해당 기업 또는 기관의 서비스 운영 목적과 이용 가이드를 명확하게 제시하지 않는 경우가 많기 때문이다.

### ○ 인공 지능 비서의 지식 유형

아마존 부사장이자 통계학 교수인 Matt Taddy는 인공 지능은 비즈니스 전문성과 빅 데이터 자산, 그리고 범용 머신러닝(General Purpose Machine Learning)으로 구성되어 있다고 정의했다.

#### AI = Domain Structure + Data Generation + General Purpose ML

Business Expertise
Structural Econom[etr]ics
Relaxations and Heuristics

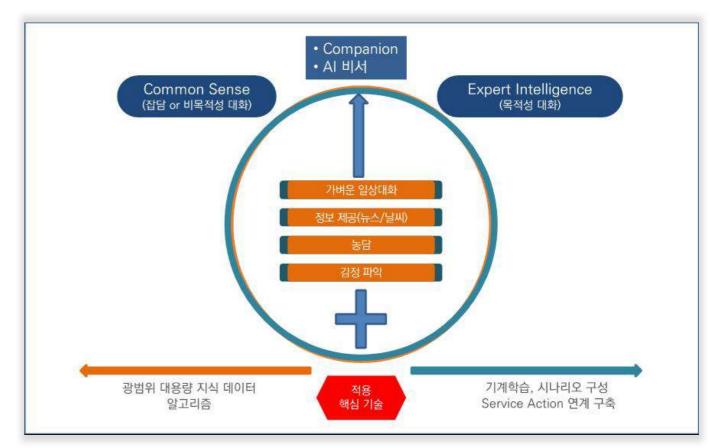
Reinforcement Learning
Big Data Assets
Sensor/Video Tracking

Deep Neural Nets Video/Audio/Text OOS + SGD + GPUs

Al definition (Matt Taddy, 2018) 타임스, 2021)

인공 지능은 잘 구현된 범용 알고리즘(General Purpose ML)과 비즈니스 전문성(Domain Structure)이 잘 반영된 빅 데이터(Data Generation)로 구성되어 있다는 것이 Matt Taddy 교수의 설명이다.

비즈니스 전문성(Domain Structure 하에 Business Expertise)이란 결국 목적 시스템이 활용되는 분야에 대한 깊은 이해를 의미한다. 이런 이해가 담보되지 않을 경우 해당 서비스는 최종 사용자가 효용성 이슈로 외면할 수도 있으므로 주의해야 한다. 가령, 택배회사의 챗봇이라면 사용자가 가장 관심 있어 할 사항들은 택배회사의 주요 업무인 택배 접수, 배송 현황, 반품 신청 등일 것이다. 만약 택배 서비스의 챗봇을 만드는 데 배송 현황에 대해 문의할 수 없다면 챗봇 서비스의 활용도는 현저히 낮아질 것이다.



대화의 유형 및 특징 (Wisenut Analysis, 2022)

스탠퍼드대학에서 출간한 'Speech and Language Processing'에서는 챗봇 대화의 유형을 Goal-oriented (목적성 대화)와 Chit-Chat(일상 대화)로 구분 하였다. 목적성 대화(Goal Oriented Conversation)는 주로 기업 및 기관에서 업무 서비스 제공 목적으로 축적한 지식으로 특정 범위를 선정하여 대화 주제와 이에 대한 학습 데이터를 구축한다. 주로 우리가 이용하는 금융 앱 내 챗봇 서비스가 이에 해당한다고 할 수 있다. 목적성 대화는 알고리즘으로 추천된 답변에 대한 정확도가 상당히 중요한 서비스이다.

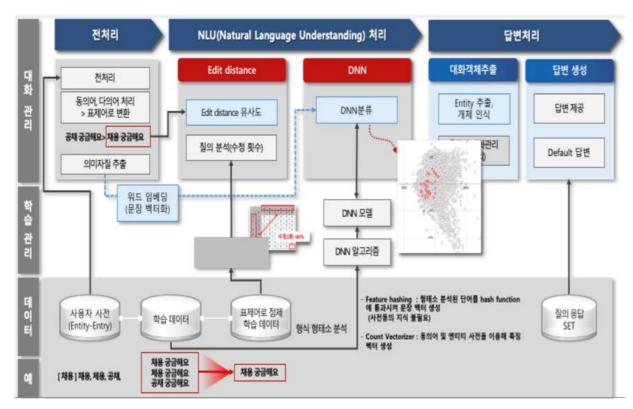
일상 대화(Chit-chat Conversation)는 주로 목적성 대화와 달리 사람 간의 흥미 및 감정 교류 위주의 대화로서 정확도에 대해서 주관적인 판단이 개입되는 것이 특성이다.

뭉치가 알려주는 국어 말뭉치

#### ○ 인공 지능 비서의 답변 의도 파악하는 방법

#### - 목적성 대화 챗봇

첫봇은 주로 질문에서 주요 특징점을 찾아 기존에 구축된 특징점과의 유사도를 판단하여 답을 내 주는 규칙기반 방식의 서비스이다. 주요 특징점은 질의가 예상되는 질문을 구성하는 주요 단어 및 표현이다. 보통 인공 지능 저작 도구를 제공하는 구글과 카카오, 와이즈넛 등의 기업들은 이런 예상되는 사용자의 질의 의도를 인텐트(Intent)로, 사용자 질의 의도에 포함되어 있는 단어 및 표현을 엔티티(Entity) 또는 피처(Feature)로 기술한다. 각 기업마다 최적화 방법은 다르기 때문에 다수 알고리즘을 앙상블하거나 다른 알고리즘을 사용할 수 있지만 기본 원리는 유사하다.

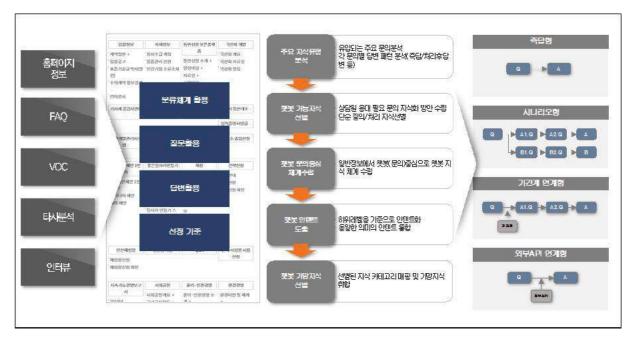


목적 지향 대화의 답변 의도 파악 방법(Wisenut Analysis, 2022)

일상 대화는 '배고파', '뉴진스 최신 앨범에 수록된 곡이 너무 좋아' 등과 같이 개인의 경험, 감정의 상태 등 주관적인 특징점들로 주로 구성되어 있다.

우리가 챗봇을 구축할 때 종종 이런 일상 대화와 목적성 대화를 한데 모아 지식을 구성하게 되는데, 서로 비슷하게 표현될 수 있어 오매칭을 유발하게 된다. 예를 들어 '너네 집은 어디 있어?'라는 일상 대화지만 위치에 대한 문의는 '지점 문의'와 동일하게 매칭되어 오매칭이 될 수 있다. 주로 소수의 피쳐로 구성된 질의가 오매칭을 유발하는 요인이 된다. 그래서 전처리 과정에서 소수의 피쳐로 구성된 질의의 경우 포괄적인 답변 또는 정확성에 대한 임계값을 높이는 방식으로 처리되어야 한다. 그래서, 보통 답을 찾는 과정에 있어 목적성 대화와 일상 대화의 모델을 분리하거나, 목적성 대화에 우선순위를 두어 매칭되는 지식이 없을 경우 일상 대화를 통해 답을 구할 수 있도록 한다.

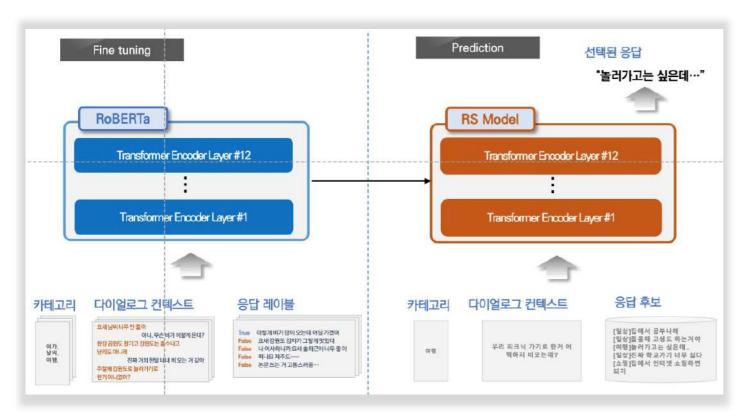
하지만, 이보다 더 우선적으로 실사용자에게 고지해야 하는 사항은 해당 챗봇의 '목적성과 한계'이다. 실제로 특정 공공기관의 경우 챗봇임을 별도 고지하지 않아 사용자가 이를 인간 상담사의 채팅 상담으로 오인하여 몇 천 자수준의 민원을 제기하는 질의가 지속 유입되기도 했다. (결국 민원 접수 기능을 챗봇에 부여하였다.) 목적성 대화는 해당 업무 분야의 사용자 관심도를 기준으로 예산을 감안하여 구축한다. 브이오시(VOC)나 콜센터를 통해 유입되는 질의에서 주로 단순 즉답 및 처리를 할 수 있는 문의를 우선순위로 구축한다.



목적 지향 인공 지능 비서 서비스의 지식 도출 절차(Wisenut Analysis, 2022)

#### -일상 대화 챗봇

일상 대화 챗봇은 지식을 특정하기에 상당히 방대하므로 주로 인공 신경망 기반으로 모델을 구축한다. 대규모의 학습 데이터를 기반으로 스스로 공통 성질을 추출하는 방식은 모델링을 위한 고사양의 서버 리소스와 방대한 데이터가 요구된다. 주로 사전 학습된 언어 모델인 비이아르티(BERT)와 같은 피엘엠(PLM, Pretrained Language Model)을 기반으로 구축한 학습 데이터를 미세 조정(Fine Tuning)하는 방식으로 구현한다.



일상 대화 답변 예측을 통한 추천 모델(Wisenut Analysis, 2022)

일상 대화는 상당히 포괄적이기 때문에 학습 데이터 구축에 앞서 어떤 데이터를 구축할지를 정해야 한다. 그런 부분에서 선행적으로 일상 대화의 분류 체계 수립이 중요하다.

국립국어원은 일상 대화를 지난 2019년부터 구축해 왔다. 2019년에는 메신저 대화 말뭉치를 구축했는데, 일상 대화의 포괄적인 범주의 주제 분류를 참고하고자 '국제 통용 한국어 표준 교육과정' (국립국어원, 2017; 김정숙. 이정희, 2018)의 주제 분류 체계를 준용하여 적용함으로써 일상 대화 유형 분류의 초석을 다졌다.

지능정보사회진흥원 역시 인공 지능 학습 데이터 구축 사업의 일환으로 일상 대화에 대한 구축을 지난 2020 년도부터 진행해 왔는데, 일상 대화에 대한 편향성을 염두하고 일상 대화 범주의 주제 분류를 정하는 데 있어 선행 과제인 국립국어원 '메신저 대화 자료수집 및 말뭉치 구축' 과제의 일상 대화 범주를 참조했다. 그러나 국립국어원의 말뭉치 사전은 약 1만 개 수준에 적용된 기준을 약 200만 건의 대규모 데이터를 적용하는 과정에서 예외 사항이 다수 발생했다.

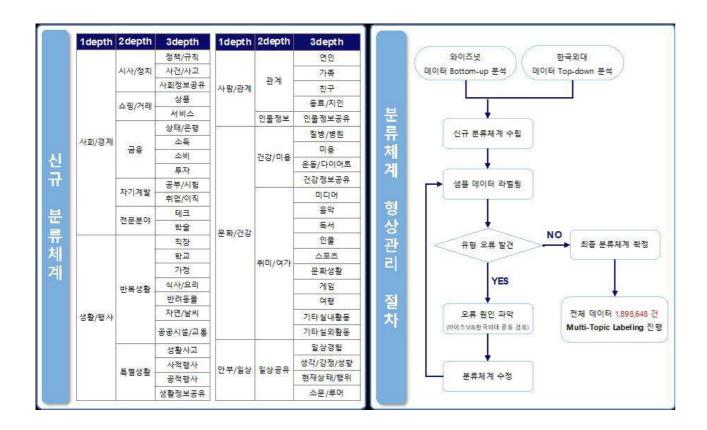
No	분류	주요정보		No	분류	주요정보
1		이름, 전화 번호, 가족, 국적, 고향, 성격, 외모, 종교, 반려동물, 연애관, 결		1	개인 신상	이름, 전화번호, 가족, 국적, 고향, 성격, 외모, 연애, 결혼, 직업, 종교
_		혼관, 개인의 기호(선호), 이상형, 인간 관계, 일상대화, SNS, 반려식물		2	주거와 환경	장소, 숙소, 방, 가구·침구, 주거비, 생활 편의 시설, 지역, 지리, 동식물
	숙소, 방, 가구, 점구, 우거비, 생활 편의 시성, 지역, 지리, 가전 제품, 자위, 잡인원, 육이,부동산, 주거시설, 이사, 생활비, 자동자, 날씨, 계절, 위치, 거리, 일, 이동 수단, 이동 경로, 대중교통(지하철,버스, 택시), 우편, 전화, 1 동신, 휴대전화, 인터넷서비스, 은행, 관공서		3	일상생활	가정생활, 학교생활	
2			4	쇼핑	쇼핑 시설, 식품, 의복, 가정용품, 가격	
			5	식음료	음식, 음료, 배달, 외식	
3	후일, 취미, 전시, 스포즈, 영화, 드라마, 예능, 동아리(동호회), 관심사, 방 역가 생활 책, 휴가, 술, 웹서핑, 문학, 음악, 미술, 공연, 스포츠 관람, 엔터테인먼트, 게임		6	공공 서비스	우편, 전화, 은행, 병원, 약국, 경찰서	
3			7	여가와 오락	휴일, 취미·관심, 라디오·텔레비전, 영화·공연, 전시회·박물관, 독서, 스포	
4	식음료	음식, 음료, 배달, 외식, 식사, 맛집, 식사 메뉴, 야식, 디저트, 요리, 간식	ZX.	8	일과 직업	취업, 직장생활, 업무
	초대, 방문, 소개팅, 약속, 가족(진척) 형사, 공적 형사, 사적(진목)모임, 여 형사 형 장소(경로), 여행 계획(일정, 숙소, 교통편, 경비), 여행탑, 기념품, 여행 사, 여행삼품	AV	9	대인 관계	친구·동료·선후배 관계, 초대, 방문, 편지, 모임	
5			10	건강	신체, 위생, 질병, 치료, 보험	
_		쇼핑 시설(장소), 식품, 의복, 가정용품, 가격, 물건, 택배, 중고거래, 서비	M	11	기후	날씨, 계절
6	상거래(쇼핑)	스 교환(환불), 구매 후기		12	여행	관광지, 일정, 짐, 숙소
7	일과 직업	취업, 직장, 업무, 회식, 급여, 계약, 협상, 회의, 스펙		13	교통	위치, 거리, 길, 교통수단, 운송, 택배
8	마용과 건강 신제, 위생, 질병, 치료, 보험, 부상, 수술, 병원, 운동, 마용, 다이어트, 건 강감진, 약품(건강보조식동 및 약품), 전료		14	교육	학교 교육, 교과목, 진로	
			15	사회	정치, 경제, 범죄, 제도, 여론, 국제 관계	
9	하고교육, 교과목, 진로 학원, 과제, 진학, 입시, 시험, 자격증, 성적, 자기 제발, 외국어 하습, 스타디, 하금(학술) 분야, 학회(제마나), 정지, 경제, 사 회, 사건(사고), 법과 제동, 여론, 국제, 관계, 재해(제안)	1	16	예술	문학, 음악, 미술	
			17	전문 분야	언어학, 과학, 심리학, 철학	

국립국어원 메신저 대화 말뭉치와 엔아이에이 (NIA) 한국어 누리소통망 데이터 분류 체계 비교)

와이즈넛은 이번 지능정보사회진흥원의 2020년도 한국어 누리소통망 데이터에 대한 고도화를 진행하면서, 지능정보사회진흥원에서 2021년도 기구축한 주제별 텍스트 일상 대화를 통합하여 고도화하는 과제를 수행하고 있다. 본 과제의 주요 목표는 기존 데이터에 대한 고도화도 있지만, 일상 대화 모델링 시 학습 데이터 구축에 참조할 수 있는 분류 체계를 정립하고자 했다. 그래서 한국외국어대학교 언어 인지학과와 함께 분류 체계를 정의하는 데 있어, 선행 연구 분석을 기반으로 한 하향식 분류 체계와 현재까지 축적된 약 190만 개의 데이터에서 샘플링한 데이터를 분석하여 상향식 체계를 수립하여 이를 통합하는 방법을 고려했다.

통합안 수립 원칙은 첫째, 크라우드 워커를 활용하는 만큼 분류가 용이하도록 직관적 명칭을 사용하려 했고 둘째, 에이아이 모델 활용성을 고려한 범용성과 셋째, 최대 3단계에 걸친 계층적인 분류 체계를 통해 구체화를 고려하였다.

 $\Theta$  10



분류 체계를 수립하는 것 이상으로 형상 관리가 중요하다. 따라서, 190만 개의 데이터의 고도화 과정에서 발생되는 이슈는 대학과, 주관 기관, 가공 기관의 담당자로 구성된 형상 관리 위원회를 통해 의사 결정을 진행하여 체계를 형상 관리 과정을 진행하고 있다. 본 분류 체계와 관련된 분석 보고서는 본 과제 완료 후 최종 보고서로 공유할 계획이다.

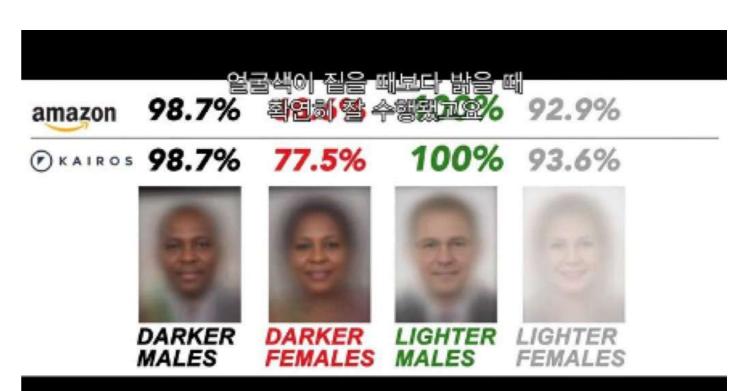
#### ○ 인공 지능 학습데이터의 중요성

인공 지능 비서를 만드는 데 있어 학습 데이터의 중요성은 아무리 강조해도 지나치지 않다. 스탠퍼드대학교 부교수이자 코세라 공동 창업자였던 앤드류 응(Andrew Ng) 교수는 Bigdata보다 Good Data가 중요하다고 강조한다. 앤드류 응 교수가 지목한 좋은 데이터란 첫째, 일관성이 있어야 하고(Defined consistenty) 둘째, 모델을 적용할 에이아이 서비스에는 해당 분야의 중요한 영역이 포함되어야 하며(Cover of important cases) 셋째, 생산 데이터로부터 시기적절한 피드백을 받으며(has timely feedback from production data) 넷째, 크기가 적당한 데이터(sized appropriately)라고 한다.

인공 지능 학습 데이터를 구축하는 데 있어 의미의 정확성, 구분의 정확성 그리고 통계적 다양성을 감안하여 가공하는 것이 중요하다. 다만, 통계적 다양성도 실제 사용성을 감안하는 것이 중요하다. 실제 사용하지 않는 대화를 만들어서 활용성이 떨어지게 되면 그만큼 비용 투자 대비 손실이 발생하기 때문이다.

#### ○ 마무리 - 인공 지능 학습데이터 구축 유의사항

최근 들어 인공 지능 서비스에서 가장 문제가 발생하는 것 부분은 바로 데이터의 편향성이다. 엠아이티(MIT) 미디어 랩의 연구원 조이 부울 란위니는 아마존의 얼굴 인식 에이아이 서비스에서 흑인 여성의 얼굴이 제대로 인식되지 않는 점을 과제를 수행하는 과정에서 우연히 알게 되었고, 이에 대해 원인을 파헤쳐 갔다. 이를 후에 다큐멘터리로 만든 것이 넷플릭스에서 방영된 알고리즘의 편견(Coded Bias)이다.

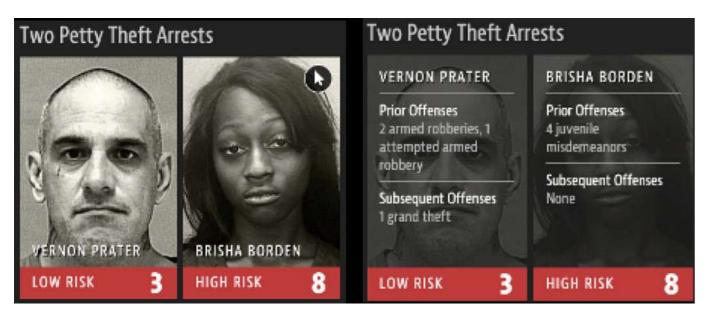


Performance on Facial Analysis Task of Gender Classification

아마존 서비스에 대한 얼굴 인식 서비스 테스트 결과

실제 아마존에서 직원 채용 시 활용하는 에이아이 이력서 검증 프로그램을 통해 여성이 한 명도 채용되지 않는 상황이 발생되는 상황이 있었다. 결국 20년 6월 10일에 아마존은 자사 얼굴 인식 알고리즘을 경찰이 사용하는 것을 1년 중단시켰고, 20년 6월 25일 미국 국회의원들은 얼굴 인식을 연방 차원에서 사용하는 것을 금지하는 법안을 도입하게 되었다.

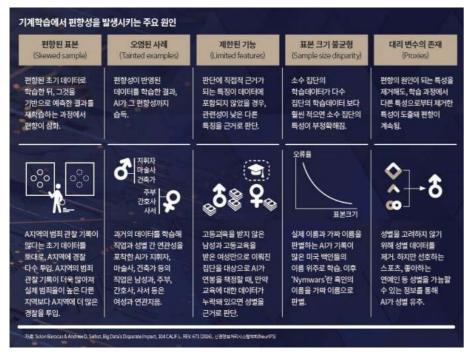
또한, 미국 탐사 보도매체 '프로퍼블리카'는 미국 내 많은 법원에서 사용 중인 에이아이 재판 지원 시스템인 COMPAS에 대한 기사를 보도했는데, 미국 플로리다에서 체포된 범죄자 1만 명을 대상으로 재범 가능성을 예측해 보니 흑인의 재범 가능성이 백인보다 2개 이상 높게 나오는 결과를 보여 줬다. 그러나 실제 현실에서는 흑인의 재범률은 백인보다 높지 않은 결과를 보여 줬다고 한다.



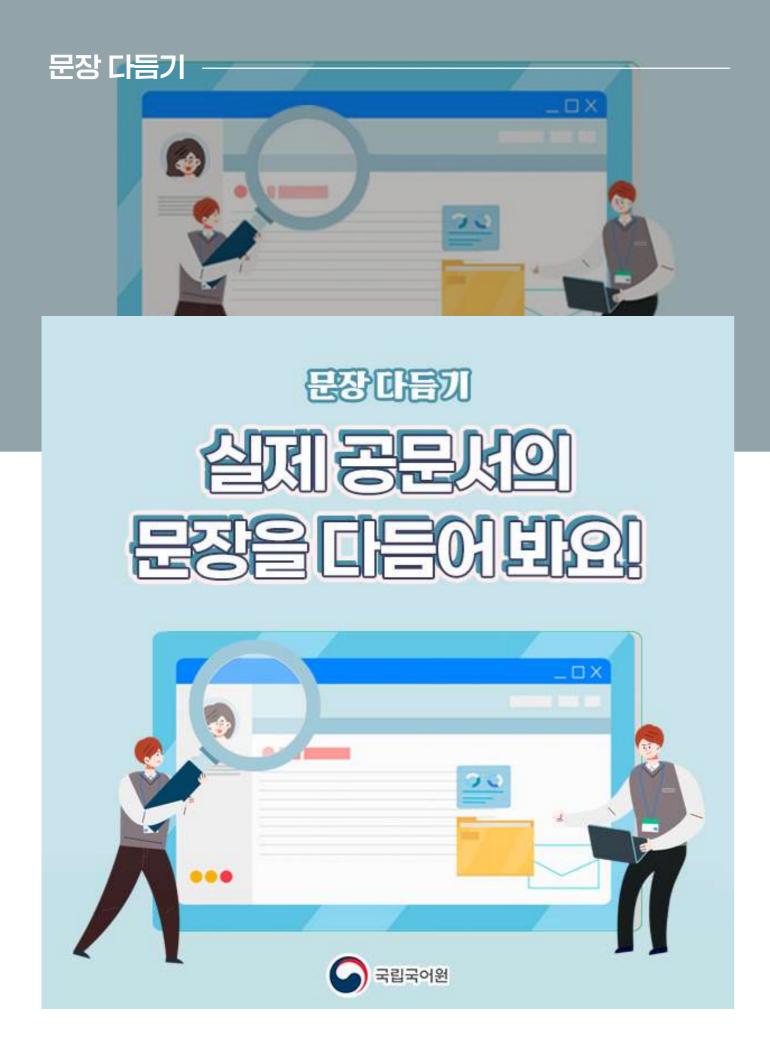
데이터 편향성에 따른 오예측 사례(사진 출처: propublica)

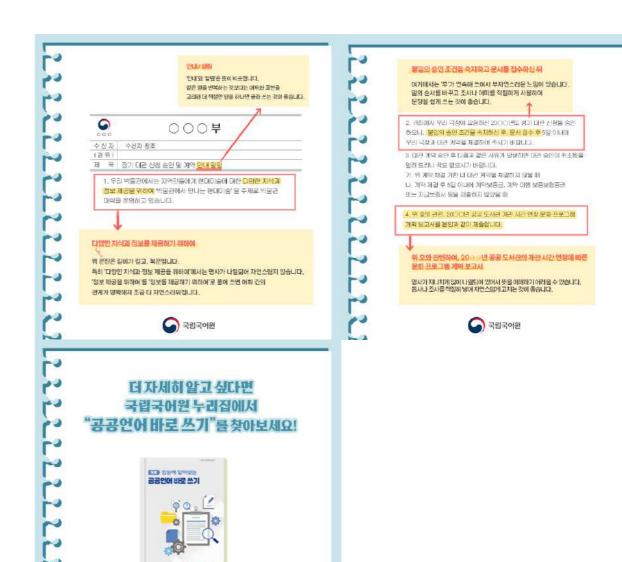
데이터의 편향성을 발생시키는 주요 원인 중 하나는 데이터의 불균형이다. 데이터가 국소적인 분야만 적용될 경우 범용적으로 활용이 어렵다는 것이다. 이는 집합의 개념과 비슷하다. 편향된 데이터와 이미 편향성이 반영된 데이터를 학습하거나, 관련성이 떨어지는 데이터를 학습하거나, 표본 크기가 균형을 갖추지 못하는 등의 원인을 통해 기계 학습은 편향성이 작용하게 된다.

이처럼, 학습 데이터 구축 시 많은 부분을 고려해야 하지만. 명심 또 명심해야 할 사항은 구체적인 목적과 한계치를 정확히 인지하고, 편향성이 발생하지 않도록 더욱 신중해야 한다는 점이다.



기계 학습에서 편향성을 발생시키는 주요 원인(그림 출처: 동아사이언스, 2021)





쉼표, 마침표.





# 국민이 고른 가장 적절하게 다듬은 말은 '열린 쉼터'

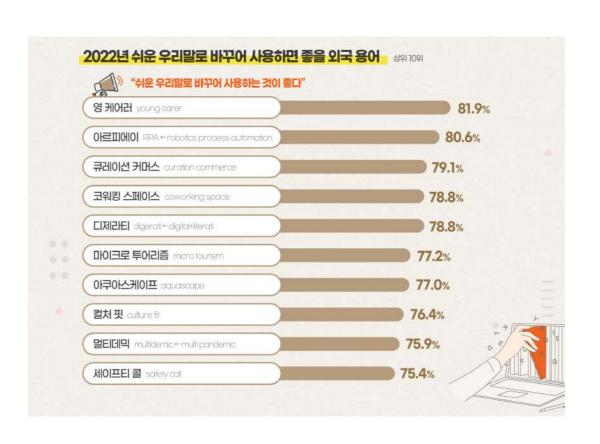


국민은 문화체육관광부(장관 박보균, 이하 문체부)와 국립국어원(원장 장소원, 이하 국어원)이 2022년에 어려운 외국어를 쉽게 다듬은 말 중 가장 적절히 다듬어진 말로 '열린 쉼터'를 골랐다.

문체부와 국어원은 '쉬운 우리말 쓰기 사업'의 하나로 우리 사회에 유입된 낯선 외국 용어를 알기 쉬운 다듬은 말로 바꾸는 일을 진행하고 있다. 지난해에는 1월부터 11월까지 22차례의 전문가 논의와 국민 수용도 조사를 거쳐 46개의 외국 용어를 다듬었다.

#### 지난해 국민이 선택한 쉬운 말로 꼭 바꿔야 할 외국 용어는?

지난해 다듬은 외국 용어 중 쉽게 바꾸어야 한다는 응답률이 가장 높은 용어는 '영 케어러' (81.9%)였다. 이는 '장애, 질병, 약물 중독 등을 겪는 가족을 돌보고 있는 청년(다듬은 말: 가족 돌봄 청년)'을 뜻한다. 두 번째로 응답률이 높게 나타난 '아르피에이' (80.6%)는 '인간을 대신해 수행할 수 있는 단순하고 반복적인 업무를 알고리즘화하고 소프트웨어로 자동화하는 기술(다듬은 말: 업무 처리 자동화)'을 말한다. 다음으로 '큐레이션 커머스(79.1%, 다듬은 말: 소비자 맞춤 상거래)', '코워킹 스페이스(78.8%, 다듬은 말: 공유 업무 공간)', '디게라티(78.8% 다듬은 말: 디지털 지식인)' 등을 쉬운 우리말로 바꿔야 할 외국 용어로 보았다.



#### 다듬은 말 중 가장 적절하다는 평가를 받은 말은 '열린 쉼터'

지난해 다듬은 말 중 적절하다는 평가를 가장 많이 받은 말은 '열린 쉼터'로 응답자의 93.1%가 긍정적으로 평가했다. '열린 쉼터'는 도시 계획에서 사람들에게 놀이 활동을 하게 하거나 마음의 편안함을 줄 목적으로 마련한 공간을 이르는 말로 '오픈 스페이스'를 알기 쉽게 대체한 말이다. '열린 쉼터' 외에도 응답자의 90% 이상이 '민간 우주 개발' (뉴 스페이스: 민간이 주도 하는 우주 개발), '가상 모형' (디지털 트윈: 가상 공간에 실물과 똑같이 만든 물체, 또는 그러한 것을 만드는 기술), '현지 조달' (로컬 소싱: 상품을 제작하거나 생산할 때 국내에서 만들어진 물자를 활용하는 전략)을 적절하게 다듬은 말로 평가했다.



### 다듬은 말은 전문가 논의와 수용도 조사를 거쳐 마련

문체부와 국어원은 공공성이 높거나 국민 생활과 밀접한 낯선 외국 용어를 <mark>새말모임\*에서 논의한다. 논의 결과 마련된 다듬은 말 후보는 국민 수용도 조사를 거쳐 다듬은 말로 선정된다. 올해에도 쉬운 우리말 사용을 위해 새로 유입되는 외국 용어를 다듬은 말로 바꾸는 사업을 더욱 활발하게 진행할 계획이다.</mark>

\* 어려운 외국 용어가 널리 퍼지기 전에 일반 국민이 이해하기 쉬운 다듬은 말을 제공하기 위해 국어 유관 분야 전문가로 구성한 위원회

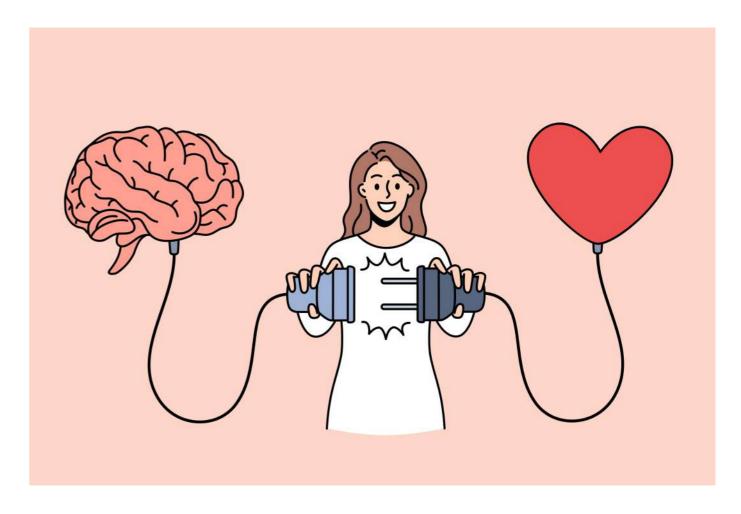




## 자기와 자신과 자기 자신

같은 듯 다른 쓰임새





우리말에서 재귀칭은 '자기, 저, 당신' 등이 있다. 재귀칭(또는 재귀 대명사)이란 문장에서 주어 등의 성분이 되풀이되는 것을 피하기 위해 쓰이는 대명사를 가리킨다. 가령 "철수는 자기를 아낀다."에서 '자기'는 '철수'를 되가리키는 재귀칭으로, 한 문장 안에서 명사 '철수'가 되풀이되는 것을 피할 수 있게 해 준다. 한 문장에서 같은 명사(특히 유정 명사)가 반복되는 것은 번거로울 뿐 아니라 부자연스럽기도 하다.

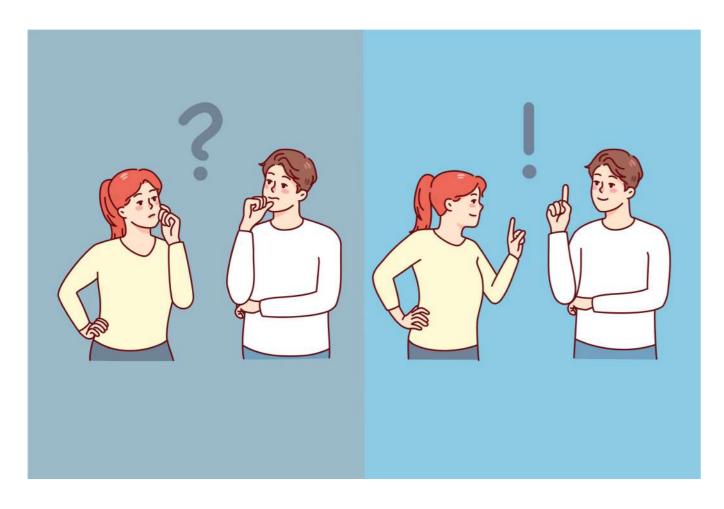
'저'와 '당신'은 높임의 정도만 다를 뿐 '자기'와 기능이 같다. "우리 아이는 늘 저만 위해 달라고 한다."에서 '저'는 아이를 낮추는 뜻이 있고, "할아버지는 당신의 사업을 위해 평생을 헌신하셨다."에서 '당신'은 할아버지를 높이는 뜻이 있지만, 둘 다 '자기'로 바꾸어도 크게 문제가 되지는 않는다. 자기는 낮추지도 높이지도 않는 명칭일 뿐 '저, 당신'과 근본적으로 같은 말이다.

'자신'과 '자기 자신'도 앞에 나온 말을 도로 가리킨다는 점에서 자기와 다를 바 없다.

아 다르고 어 다른 우리말

② 철수는 자기를/자신을/자기 자신을 아낀다.

⑦에서 '자기, 자신, 자기 자신'은 모두 철수를 가리킨다는 점에서 같다. 굳이 셋의 차이를 꼬집자면 자연스러움의 정도가 서로 미세하게 다르다고 할 수 있다. 곧 '자기 자신'이 가장 혀에 잘 붙고 '자신'은 무난하며 '자기'는 다소 덜 매끄럽다.



그런데 다음의 예에서는 자기와 자신과 자기 자신의 지시 대상이 서로 다를 수 있다.

- 만 민수는 철수가 자기를 아낀다고 생각한다.
- @ 민수는 철수가 자신을 아낀다고 생각한다
- @ 민수는 철수가 자기 자신을 아낀다고 생각한다.

(마)의 '자기'는 민수를 가리키고 (마)의 '자기 자신'은 철수를 가리키는 데 반해, (마)의 '자신'은 우선적으로 철수를 가리키지만 철수를 가리킬 가능성도 없지 않다. 일반적으로 복문 구성에서 자기는 상위 문의 주어를 가리키고, 자신이나 자기 자신은 내포 문의 주어를 가리키는 경향이 강하다. 다만 자신은 자기 자신에 비해 가리키는 대상이 다소 유동적인 것으로 보인다.

자기는 주로 3인칭 주어를 가리킬 뿐, 1인칭과 2인칭 주어를 가리키기는 어렵다. 자기와 달리 자신과 자기자신은 3인칭 주어는 물론 1, 2인칭 주어를 가리킬 수 있다. (⑩와 ⑩의 경우)

- ⊕ 나는 자신을/자기 자신을/자기(X)를 사랑한다.
- (☆) 너는 자신을/자기 자신을/자기(△)를 응원해야 한다.

그런가 하면 '자기'는 다른 명사와 결합하여 합성어나 구를 이룰 수 있지만 '자신'은 그럴 수 없다. (㈜의 경우)

④ 자기 기만/자신(X) 기만, 자기도취/자신((X) 도취, 자기만족/자신(X) 만족, 자기변명/자신(X) 변명, 자기 자본/자신(X) 자본, 자기 최면/자신(X) 최면

반면 '자신'은 강조적 표현으로 사용할 수 있지만, '자기'는 그럴 수 없다. (⑩와 逊의 경우)

- 그 순간 나 자신도/자기(X)도 모르게 소리를 지르고 말았다.
- (자) 그 문제는 너무 어려워 선생님 자신도/자기(X)도 풀 수가 없었다.

위 문장의 경우 '나도 모르게', '선생님도'라고 할 수도 있지만 '자신'을 덧붙여 '나 자신도 모르게', '선생님 자신도'로 표현함으로써 좀 더 강조하는 느낌을 나타낼 수 있다. 그렇지만 '자기'는 이렇게 쓰일 수 없다.

글: 강은혜

※ 참고 자료 안상순, 『우리말 어감 사전』, 도서출판 유유, 2021.

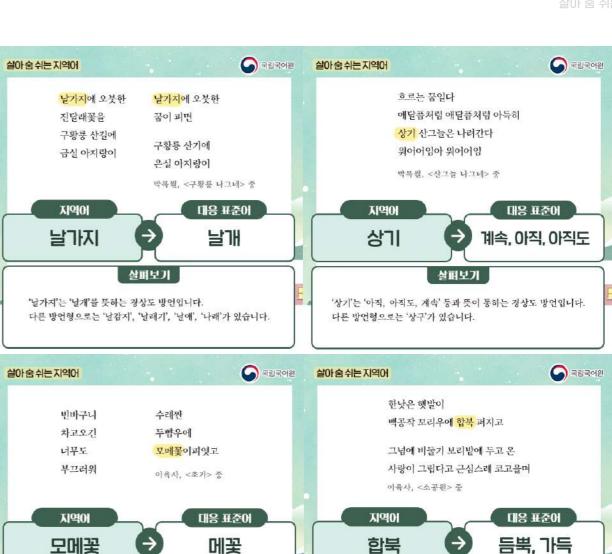




**○** REROR

국립국어원 온라인 소식지 쉼표, 마침표.





살펴보기

다른 방언형으로는 '항거', '항거석', '항거시' 등이 있습니다.

'합북'은 '듬뿍', '가득'을 뜻하는 경상도 방언입니다.



살펴보기

메꽃은 주로 밭 어귀에서 흔히 볼 수 있는 나팔꽃 모양의 꽃입니다.

'모메꽃'은 '메꽃'을 뜻하는 경상도 방언입니다.

다른 방언형으로는 '들보라', '모메' 등이 있습니다.

### 이음이가 전하는 국어 소식

# '한국어-외국어 병렬 말뭉치 구축의 쟁점과 활용 방안' 워크숍 개최



국립국어원(원장 장소원)은 1월 14일(토) 오전 9시 코지 모임 공간 강남역 2호점에서 '한국어-외국어 병렬 말뭉치 구축의 쟁점과 활용 방안'을 주제로 워크숍을 개최하였다. 22년 한국어-외국어 병렬 말뭉치 구축 사업단(연구책임자 이정희)이 주최하는 이번 워크숍에서는 국립국어원이 구축하고 있는 한국어-외국어 병렬 말뭉치의 활용 방안 및 기대 효과에 대한 전문가의 특별 강연과 구축 과정에서 발생하는 주요 쟁점에 대한 토론이 진행되었다.

최근 한국 문화 콘텐츠의 글로벌 확산에 따라 한국과 한국 문화에 대한 관심이 높아지면서 한국어 교육 및 자동 통 · 번역 기술 개발을 위한 언어 자원 수요가 증가하고 있다. 하지만 영어, 일본어, 중국어 등을 제외한 태국어, 베트남어 등 외국어의 경우 확보된 언어 자원이 매우 부족한 실정이다.

이러한 상황을 극복하기 위해 국립국어원은 2021년 8월부터 한국과 정치•경제•문화 교류가 활발하지만 기구축 **언어 자원이 부족한 8개 언어**\*를 대상으로 한국어-외국어 병렬 말뭉치 구축 사업을 추진해 왔다. 한국어를 8개 언어로 번역한 병렬 말뭉치는 언어 자원의 불균형을 해소하고 관련 산업의 부가 가치를 높이는 데 이용된다.

\* 태국어, 베트남어, 인도네시아어, 캄보디아 크메르어, 필리핀 타갈로그어, 인도 힌디어, 러시아어, 우즈베크어

국립국어원은 2021년 1차 사업에서 약 800만 어절의 병렬 말뭉치를 성공적으로 구축 완료하였으며, 2022 년 이후 1,000만 어절 규모로 확대하여 지속적으로 병렬 말뭉치를 구축하고 있다. 1차 사업에서 구축한 한국어-외국어 병렬 말뭉치는 연구 및 기술 개발의 기초 자료로 누구나 활용할 수 있도록 '모두의 말뭉치(https://corpus.korean.go.kr)'를 통해 올해 상반기에 공개될 예정이다.

이번 워크숍에서는 각계 전문가들을 초청하여 한국어-외국어 병렬 말뭉치가 향후 관련 기술 연구와 산업에 미칠 영향을 전망하고, 병렬 말뭉치 번역 과정에서 발생하는 다양한 쟁점에 대해 작업자들이 서로의 경험과 생각을 나눔으로써 발전적 병렬 말뭉치 자원 구축과 활용 방향을 점검하였다.

### 한국어 교육, 그리고 우리

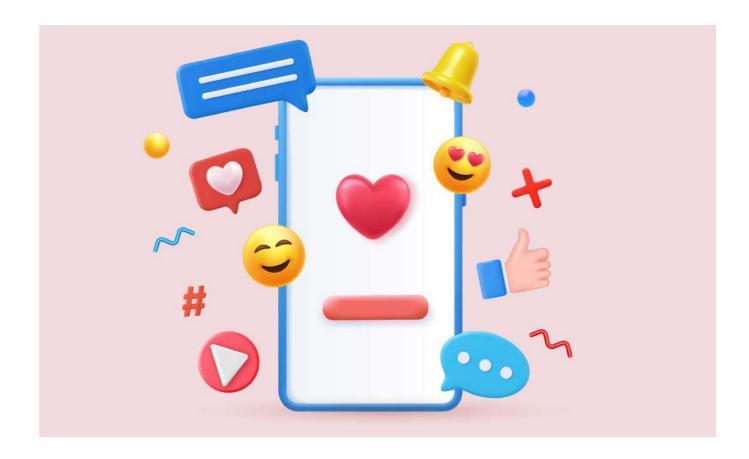


## 고맙습니다

이미향(영남대 글로벌교육학부 교수)

나는 한국어 선생님이다. 선생님으로서 학생들로부터 가장 많은 연락을 받는 날은 아무래도 새해 벽두와 스승의 날이다. 한 해의 마지막과 시작 때, 우리는 주위를 돌아보며 사람들과 인사를 나눈다. 지난날, 연말이 되면 연례행사처럼 서점을 들르고 카드를 골라, 한 해의 안부와 고마움을 담아 글을 썼다. 오늘날 한국 사회는 인사를 하는 방법이 많이 바뀌었다. 이런 때 나는 오히려 외국인 제자들로부터 한국어로 된 카드 또는 문자메시지를 받는다. 한편으로 뿌듯하면서도 한편으로는 멋쩍다.

올해도 연말과 새해에 인사를 받았다. 이메일과 문자메시지로 형식은 달라졌지만, 새해 안녕과 건강을 비는 인사말은 정감이 넘친다. 그런데 그런 마음을 담아 도착한 메시지의 대부분이 '새해를 축하합니다.'로 시작한다. 새해가 된 것을 축하한다니…. 한국어 선생님의 습관인 양, 처음 몇 통에 대해서는 바른 표현으로 고쳐 주고 싶은 마음이 굴뚝과 같다가 곧 인사로 이해하고자 마음을 굳힌다. '축하한다'는 인사는 스승의 날에도 많이 받는다. '스승의 날을 축하한다'는 문구가 하루 종일 들어오는데, 특별한 날에 축하한다는 말을 직역한 것이라 어쩔 수 없다. 단지 어색한 표현에 내가 불편할 뿐이다.

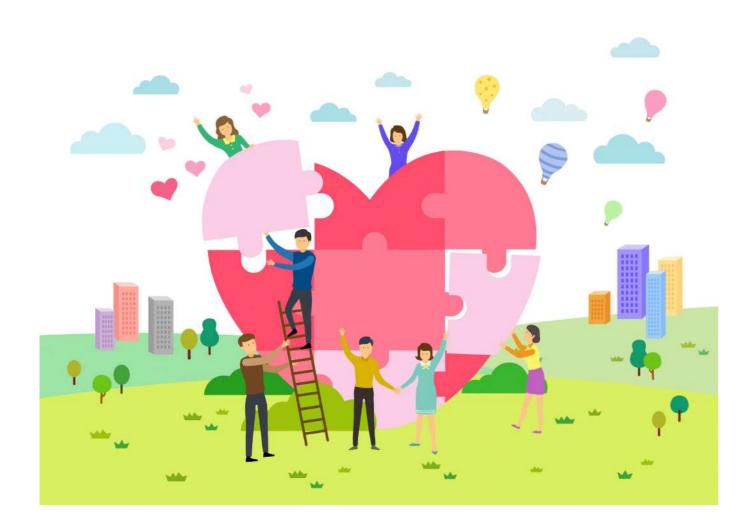


한국어 교육, 그리고 우리

여러 언어에서 축하한다고 하는 저 상황에서 우리는 어떻게 말하고 있는가? 새해 인사로는 '새해 복 많이 받으세요.'라고 하고, 어버이날이나 스승의 날에 보내는 인사로는 '고맙습니다'라고 한다. 외국인 학습자들은 '고맙습니다'를 완전히 초급 단계에서 배우기 시작한다. 남이 베풀어 준 호의나 도움에 대해 좋은 마음을 드러내는 '고맙다'는 말은 어떤 언어에서든 가장 기본이 되는 표현이며, 배움의 동기도 확실하다. 그런데 한국학습자들이 '고맙습니다'를 알아갈 때 겪는 몇 가지 어려움이 있다.

우선은 발음의 문제이다. 한국어 '고맙습니다'는 '감사합니다'에 비해 외국인에게 어렵다.

입에서 막히는 한국어의 받침소리 때문이다. 그보다 더 큰 고비는 한자어와 고유어에 대한 한국 사람들의 편견이다. 많은 한국인이 한자어로 된 '감사합니다'를 더 공손한 표현이라고 생각한다. 심지어 '감사합니다'가 상대를 더 높이는 표현이고, '고맙습니다'는 편한 사람에게 쓰는 것이라고 알려 주는 오류도 범한다. 설령 자신이 그렇게 쓰고 있다고 하더라도 누군가에게 교육적 모범으로 보일 답은 아니다. 또 다른 난관은 친구뻘인 한국인 20대들이 무수히 보여 주는 '고맙습니다'의 오용례이다. 현지인의 표현은 직관이 없는 외국인 학습자의 언어 사용에 영향을 크게 미친다. 수업을 마친 강의실에서, 택시에서 내리면서, 사무실에서 볼일을 보고 돌아서는 많은 장면에서 20대들은 흔히 고맙다는 인사 대신 '수고하세요'라고 하고 나가 버린다.



과연 '고맙습니다'가 한자어인 '감사'에 비해 격이 떨어질까? '고맙다'는 말은 남이 베풀어 준 호의로 마음이 즐거운 것이고, '감사하다'는 고맙게 여기거나 고마운 마음이 있다는 것을 보이는 말이다. 혹 그 두 말이 어떤 위계 관계에 놓여 있는 말이라면, 뉴스를 마치는 진행자가 전 국민 앞에서 '고맙습니다'라고 인사할 수 없다. '고마'는 '공경'을 뜻하는 옛말이다. 다시 말해, 상대방을 공경하는 마음에서 하는 인사이다.

우리는 매일 인사를 하고 산다. '고맙습니다. 미안합니다, 다음에 우리 차 한잔해요'와 같은 인사말들은 어찌보면 영양가가 없는 빈말 표현이다. 그러나 비록 주어와 서술어를 갖춘 문장이 아니라 해도, 사람들 사이에서 오히려 완벽한 문장보다 더 큰 기능을 한다. 한때 시청률 23%를 넘어선 인기 드라마에서 주인공이 한 말이 기억난다. 그 주인공은 열심히 공부해서 공무원이 되어, 기차역 분실물센터를 지키고 싶다고 말한다. 그곳에서는 '고맙습니다'를 자주 들을 수 있기 때문이란다. 고맙다는 인사는 자신의 호의를 드러내는 동시에, 타인의 자존감을 높여 살아갈 이유를 만들어 주기도 한다는 말이다.

한국말에는 '축하하다, 수고하다'마저 대체할 인사말 '고맙습니다'가 있다. 이 말의 힘을 빌려 나도 한국어 교육을 함께하는 '우리들'로 열두 달을 고민해 주신 여러 분들께 인사드리고자 한다. 한국말 한마디를 가르치는 어느 칠판 앞, 또는 한국어 교재 한 면을 채우느라 애쓰시는 어느 책상 앞의 모든 한국어 선생님, 고맙습니다.

# 우리 일상 속에서 살아 숨 쉬는 사전을 꿈꾸며

경북대학교 남길임 교수



《쉼표, 마침표》

만나 뵙게 되어 반갑습니다. 국립국어원 소식지 독자 여러분께 자기소개를 부탁드립니다.

남길임

안녕하세요? 경북대학교 국어국문학과 교수 남길임입니다. 저는 한국어 신어, 미등재어의 수집과 기술에 많은 관심을 갖고 있어요. 저서로 〈신어 2020- 코로나 펜데믹 시대의 새로운 언어〉, 〈신어 2021- 코로나 19 시대, 우리는 일상회복 중인가?〉가 있습니다. 작년부터 아시아사전학회(Asialex: The Asian Assocation for Lexicography) 부회장을 맡아서 올해 6월 서울에서 개최되는 Asialex 2023, 제16회 국제학술회의를 준비하고 있어요.



《쉼표, 마침표》 사전학을 전공하셨고 오랫동안 관련 연구에 매진해 오셨는데요. 사전학이란 무엇이고 왜 필요한 것인지 궁금합니다.

#### 남길임

어느 학문이든지 시대의 이념이나 문화, 역사적 사건에 따라 부침이 있기 마련이죠. 하지만 사전 편찬이나 사전학만큼 많은 변화가 있는 분야도 드문 것 같습니다. 저는 대학원 시절부터 ( 연세한국어사전》, (외국인을 위한 한국어 학습 사전) 등의 사전 편찬에 참여하고, 수년 동안 ' 신어 조사 사업'을 진행하며 사전과 사전학의 격동기를 체험해 왔는데요. 지난 수십 년간, 인쇄사전의 쇠퇴, 말뭉치의 도입과 사용자 참여의 확대 등 사전학자들이 이야기하는 전환기(transition period)"는 현재에도 진행 중이에요. 한동안 사전학의 변화는 계속되리라 생각합니다.

질문으로 돌아가서, 이러한 변화를 고려한다면 사전학을 정의하는 건 쉽지 않은 것 같아요. 간단히 정의를 내리자면, 사전은 인류 역사에서 4,000년 이상의 역사를 가져 왔는데요. 이러한 사전을 편찬하고 비평하는 학문 전체를 사전학이라고 합니다. 일반적으로는 실제 사전 편찬 과정을 다루는 사전편찬학, 사전에 대한 비평을 다루는 이론사전학으로 구분하기도 해요. 그런데 우리가 일반적으로 알고 있는 인간을 위한 사전도 있지만, 검색엔진이나 자동번역기 등 기계를 위한 사전도 있어서, 이러한 모든 사전을 포괄한다면 사전학의 범위는 매우 넓어질 것입니다. 우리의 눈에는 보이지 않지만, 우리의 일상생활 곳곳에 사전이 있고, 우리는 우리도 모르는 사이에 사전을 활용하고 있어요. 즉 '보이는 사전'과 '보이지 않는 사전'이 있죠. 사전은 이러한 정보의 원천이자 언어 정보의 가장 정교화된 자원으로서 일상생활의 편의에도 상당한 영향을 미치고 있습니다. 이러한 사전을 연구하는 일이야말로 한국어 자원 전체를 체계적으로 기술하고 연구하는 중요한 일이라고 생각해요.

《쉼표, 마침표》 웹사전이 등장하며 사전의 형태나 쓰임이 다양해졌습니다. 사전의 과거와 현재는 어떤지, 미래는 어떻게 전망하시는지 궁금합니다.

#### 남길임

사전 편찬의 흐름은 두 가지 관점에서 논의할 수 있습니다. 첫째는 인쇄사전의 비즈니스 모델이 더는 존재하지 않는 현실이고, 둘째는 뉴뉴미디어 시대의 언어 변화 속도입니다. 우리는 이제 사전을 서점에서 찾기보다는 온라인에서 검색하고 있죠. 전문가가 아닌 사람도 사전의 항목을 기술하고 있고요. 또 이중언어사전의 경우에는 온라인 사전을 검색하기보다는 번역기 등을 사전 대용으로 사용합니다. 언어사전은 인류가 가진 언어 자원의 가장 정교화된 체계라고 할 수 있는데요. 기존 상업적 모델의 변화 속에서 언어 자원의 축적이 어떤 주체에 의해서 어떻게 이루어져야 할 것인지는 전 세계 사전학자들의 고민이기도 해요.

언어 변화의 속도 역시 언어사전 편찬에 상당한 쟁점이 되고 있습니다. 누구나 글쓰기로 감정을 표현하는 뉴뉴미디어 시대에는 언어 변화의 속도가 이전의 속도와 비교할 수 없을 만큼 빠르죠. 일반 대중이 글쓰기의 주체가 되면서 생산해 내는 다양한 새로운 표현들과 비격식적 문어를 사전에 어떻게 담을 것인지 고민해야 하는 시점입니다. 국어사전이 한국인의 의사소통 언어를 전부 반영하는 것이라면 일상생활에 많은 부분을 차지하는 미등재 표현들을 어떻게 처리해야 할까요? 실제로 의사소통 언어의 많은 부분을 담당하는 온라인 언어의 처리는 자연언어처리 등 공학적 관점에서도 상당한 관심을 받고 있는데요, 전통적으로 사전 표제어의 기반이 되어 온 문어 어휘들에서부터 구어뿐만 아니라 온라인 언어의 편입까지도 고려해야 할 시점이라고 생각해요.

이런 흐름 속에서 최근 사용자 참여형 사전이 등장했는데요. 사용자 참여형 사전이란 전문 편찬가가 아닌 사용자가 직접 표제어를 제안하거나 뜻풀이까지 기술할 수 있는 사전을 말합니다. 그러나 사전 편찬 과정에서 사용자의 참여하더라도 현실 언어의 역동성과 포괄성을 담보할 수 없다는 문제는 여전히 존재해요. 가장 큰 문제는 몇몇 논문에서 보고되기도 했지만 사용자 참여형 사전에서 사용자는 대체로 충성스러운 소수 사용자에 국한된다는 거죠. 또한 이들이 자신의 관심사에만 맞추어 기술한다는 문제도 있고요. 따라서 사용자 참여형 사전의 최대 문제는 '균형성'과 '포괄성'의 한계입니다. 따라서 사용자 참여형 사전의 이상적인 모델을 만들어 가되, 사전 편찬 전문가와 사용자의 효율적 협업 모델을 모색할 필요가 있다고 봅니다. 이는 현대 사전학 연구자들의 주요 관심사이기도 해요.

《쉼표, 마침표》 요새는 웹사전이 더 많이 쓰이는 것 같습니다. 종이사전과 다른 웹사전만의 특징이 있다면 무엇일까요?

#### 남길임

이제 종이사전과 웹사전을 비교하는 것조차 어색한 시대가 된 것 같아요. 대학에서 인쇄사전을 한 번도 경험해 보지 못한 세대들을 가르치고 있으니까요. 종이사전 편찬에 참여했을 때는 지면의 한계를 고려하여 표제어의 수를 정하고, 뜻풀이, 예문, 삽화 등 미시구조 정보의 양도 조절했어요. 표제어의 선형적 배열순서도 남한의 체계와 북한의 체계가 다르듯이. 사전 디자인에서 중요한 부분이었습니다. 하지만 웹사전은 지면의 제약에서 벗어나기에 더 이상 표제어의 수나 미시구조 정보의 양을 고려할 필요가 없어요. 표제어의 추가나 삭제가 언제나 가능한 점은 언어의 역동성을 잘 반영할 수 있는 모델이라고 생각되고요. 뉴뉴미디어 시대의 특성을 반영한 소리, 텍스트, 영상을 다양하게 결합한 사전을 기획할 수도 있습니다. 물론 정보가 많다고 좋은 것은 아니죠. 유용한 정보에 걸맞은 시각화가 중요할 것 같습니다. 무엇보다도 웹사전의 장점은 활용이 검색의 메커니즘을 통해 이루어진다는 점이에요. 검색을 통해 찾고자 하는 단어를 무조건, 즉각적으로 찾아낼 수 있는 거죠.



《쉼표, 마침표》 표제어의 추가와 관련해서 말씀 주셨는데요, 사전에 단어를 새로 올릴 때에 필요한 기준이 있을까요?

#### 남길임

일반적으로 사전의 표제어로 등재되느냐 마느냐의 가장 중요한 기준은 '출현'과 '빈도'입니다. 즉 실제로 사용되느냐 아니냐, 사용되었다면 얼마나 사용되느냐이지요. 충분한 빈도로 출현한다면 당연히 사전에 등재되어야하는 것이고, 사전의 뜻풀이나 예문 등도 실제 사용된 용례 내에서 가장 전형적인 것을 우선적으로 기술해야 합니다. 사전의 용도나 목적에 따라 다르겠지만, 사전의 표제어 구성과 뜻풀이, 예문 등은 언어 현실을 가장 잘 반영하는 방식이어야 해요. 이것은 일반적으로 실제 사용 양상과 빈도로서 증명됩니다.

《쉼표, 마침표》 코로나와 관련된 신어 자료집을 출간하신 배경이 궁금합니다.

남길임

사실 우연한 시기에 코로나 신어를 수집하기는 했지만, 매년 출현하는 한국어 신어를 정기적으로 수집하고 기술하는 건 언어학적으로나 사회 문화적 의미에서 매우 중요한 일이라고 생각해요. 국립국어원 신어 조사 사업은 1994년부터 2019년까지 거의 매년 지속되어 왔고요, 저는 2013년부터 2019년까지의 신어 조사 사업에 참여했어요. 코로나 세계적 유행이 본격화된 2020년에 신어 조사 사업이 중단이 되었었는데요, 코로나의 유행으로 새로운 신어는 계속 쏟아지는 상황이었죠. 아무리 〈우리말샘〉 등에서 사용자 참여를 유도한다고 해도 이전 포괄적인 신어조사만큼의 양과 질이 되지 못할 것이라는 생각이 들었어요. 웹 언어의 역동성과 신어의 저빈도 속성을 고려할 때, 이 시기에 코로나 신어를 수집하지 않으면 안 된다는 우려도 있었습니다. 웹 언어는 일분일초가 다르게 움직이고, 어제의 자원이 오늘의 자원과 동일하지 않아요. 따라서 신어 조사는 당대 온라인 자료를 기반으로 하는 것이 가장 신뢰성이 있죠. 대표적으로 독일, 스페인, 중국 등에서도 연도별, 월별 신어를 수집합니다.

마침 코로나 신어에 대한 국제학술회의 발표를 준비하던 연구진들은 코로나 신어만이라도 수집하자는 취지로 저서 작업을 시작했어요. 생각보다 코로나가 길어지는 바람에 두 권의 연도별 신어 자료집, 〈신어 2020- 코로나 펜데믹 시대의 새로운 언어〉, 〈신어 2021- 코로나 19 시대, 우리는 일상회복 중인가? )가 나왔습니다.

신어와 사전의 관계에 대해서라면, 세계 최대의 온라인 사전을 표방하는 Wordnik.com의 창시자인 에린 매킨(Erin Mckean)이 "세상의 모든 영어를 수집하겠다"라고 포부를 밝힌 것처럼 "세상의 모든 한국어"를 수집할 필요가 있다는 말씀을 드리고 싶어요. 그것이 꼭 ( 표준국어대사전》이나 〈우리말샘〉이 아니라 하더라도, 어딘가에는 한국인이 사용하는 모든 단어를 모아 둘 필요가 있지 않을까요? 단순한 목록이라도 좋고, 품사나 문법 정보, 맥락 정보가 있다면 더 좋겠지요. 신어뿐만 아니라 메신저, 댓글 등 온라인 언어를 포괄하는 모든 언어 자원에 대한 집적과 체계화는 향후 살아 있는 한국인의 일상 대화를 처리하는 자연언어처리에도 유용한 자원이 될 것입니다.

#### 《쉼표. 마침표》

현재 사용되는 언어를 기록하여 후대에 남기는 것도 중요하지만 과거의 어휘들을 발굴하고 수집할 필요도 있을 것 같습니다.

#### 남길임

언어는 곧 그 나라 문화의 원천이자 기반이에요. 당대 사용자들이 사용하는 언어 전체를 기록하고, 과거 어휘들을 발굴하고 수집하는 것 모두 매우 중요하다고 생각합니다. 기존 역사 사전을 디지털화하는 일도 필요하겠고요, 어휘역사사전 등의 편찬과 보강을 통해 미처 사전에 등재되지 않았던 표제어를 수집하는 일, 사전에 등재된 단어라도 어원이 밝혀지지 않았던 다양한 생활 속 어휘들의 어원을 밝히고 기술하는 작업도 필요하다고 생각합니다. 사실 역사 자료에서 어원을 발굴하고 수록하는 것은 현대어를 수집하고 기록하는 것보다 훨씬 품이 많은 작업일 거예요. 따라서 역사 자원에 대해서는 오랫동안 꾸준한 수집과 기술 체계가 필요합니다. 한편 거꾸로 생각해 보면, 당대 언어 자원 역시 미래에는 과거 어휘 역사 자원이 되겠지요? 따라서 당대 언어 자원이 사라지고 묻히지 않도록 잘 정리하고 채록하는 시스템을 갖추는 것도 균형적인 시각 속에서 다루어져야 할 듯합니다.

《쉼표, 마침표》 사람들의 의사소통에 필요한 새로운 말을 현실의 사전이 충분히 제시하지 못하는 모습을 봅니다. 사전은 사용자 언어의 어디까지를 담아야 하는 걸까요? 사전을 둘러싼 쟁점을 바라보는 교수님의 시각이 궁금합니다.

#### 남길임

일반 대중들이 사전에 대해 가지고 있는 생각은 다소 이중적인 측면이 있다고 봅니다. 예를 들자면, 신어의 경우에는 몇몇 품위 없는 유행어를 표제어로 등재하는 것이 적절치 않다고 보는 시각도 있어요. 하지만 막상 유행하는 단어를 찾았는데 사전에 없다면 사전이 이런 흔한 어휘도 싣지 않는다고 불평을 하기도 합니다. 사전의 규범성이나 권위에 대한 시각은 일반인과 전문가가 다르고, 전문가 사이에서도 의견이 상당히 다를 수 있습니다. 이러한 논쟁에도 불구하고. 저는 사전은 가치 판단 없이 현재 한국인이 사용하는 한국어 모두를 표제어로 등재해야 한다고 생각합니다. 개별 표제어가 가지고 있는 차별성이나 윤리성 등의 가치 판단은 등재 이후 뜻풀이나 참고 정보, 표제어의 시각화 등 미시구조나 사전의 구조적 체계를 통해서 얼마든지 해결할 수 있으리라고 생각해요. 규범사전은 사전 그대로의 존재 의의가 있는 것이지만, 현대 한국인의 의사소통을 모두 담은 사전 역시 사전의 기본적인 정의에 충실한 사전입니다.

특히 최근 논란이 되고 있는 차별과 혐오 표현, 비윤리적 표현과 관련하여 생각해 볼까요? 사용자는 특정 표현의 의의가 궁금해서 사전을 찾기도 하지만 그 윤리성의 여부에 대한 유권 해석이 필요할 수도 있습니다. 차별 혐오 표현이라고 해서 사전에서 제외하는 것보다는 적극적으로 등재해서 그 단어가 특정한 맥락에서 비윤리적으로 사용될 수 있다는 것을 알리는 것도 필요합니다.



#### 《쉼표. 마침표》

국립국어원의 〈우리말샘〉은 사용자가 직접 참여하는 국어사전입니다. 일반 국민과 전문가의 참여로 양적으로도 질적으로도 우수한 국어사전을 표방하고 있는데요. 이러한 참여형 사전의 올바른 역할과 주의해야 할 점은 무엇이라고 보시는지요?

#### 남길임

(우리말샘)은 '개방성'과 '실용성', 그리고 '연계성'을 주요 골자로 하여 편찬된 사전으로 알고 있어요. 그런데 이 중 사용자 참여의 '개방성'을 늘리는 것의 수용 범위를 정하는 것은 제안 정보의 양과 질을 고려할 때 어려운 문제인 것 같습니다. 〈우리말샘〉의 개통 초기에는 사용자 제안 항목의 등재율이 50%를 넘었는데, 2019년 정도를 기점으로 등재율이 30% 아래로 낮아집니다. 또한 3명의 사용자가 90,000여 개의 항목을 등록하는 등 소수의 사용자만이 사전의 제작에 참여하고 있는 듯한 모습을 보이기도 합니다. 사용자 참여형 사전의 전형적인 한계 중 일부를 〈우리말샘〉도 가지고 있는 셈이죠. 아직 사용자 참여형 사전으로 자리 잡지 못했다고도 말할 수 있겠네요. 물론 이상적인 사용자 참여형 사전으로의 여정에는 더 많은 논의와 시간이 필요합니다.

제가 생각했을 때, 사용자 참여형 사전이 사전학에 기여할 수 있는 가장 큰 부분은 일반 언어 화자가 일상생활에서 사용하고 있는 어휘 항목들을 지금보다 훨씬 더 자유롭게 수집하는 것입니다. 등재할 가치가 있는 수많은 어휘 항목을 놓치지 않도록 더욱 다양한 사용자 참여를 유도하는 것이지요. 다만 전문가의 적극적인 개입도 여전히 필요합니다. 특히 사용자 참여의 빈틈을 매울 수 있는, 출현의 순간을 놓치면 다시 수집하기 어려운 신어와 미등재어의 수집에 좀 더 노력을 기울였으면 합니다. 유사한 관점에서 새로운 어휘의 수집과 기술을 하나의 채널에서만 의존하는 것도 한계가 있습니다. 분명히 인지할 필요가 있습니다. 예를 들어, 2021년에 새로이 〈우리말샘〉에 등재된 항목과 동일 시간 웹 뉴스를 중심으로 수집된 〈신어 2021〉을 비교해 보면、〈우리말샘〉은 〈신어 2021〉 중 단 9개의 신어만을 등재했다는 것을 알 수 있어요. 42개의 제안 항목을 등재할 수 있다는 사실은 의의가 있지만 나머지 250여 개의 신어에 대한 정보는 찾을 수 없습니다. 하나의 방법론으로 채널을 고정하기보다는 다양한 사용자의 참여를 유도할 필요가 있고, 사용자와 전문가의 협업 구도를 좀 더 체계적으로 마련할 필요가 있다고 생각해요.

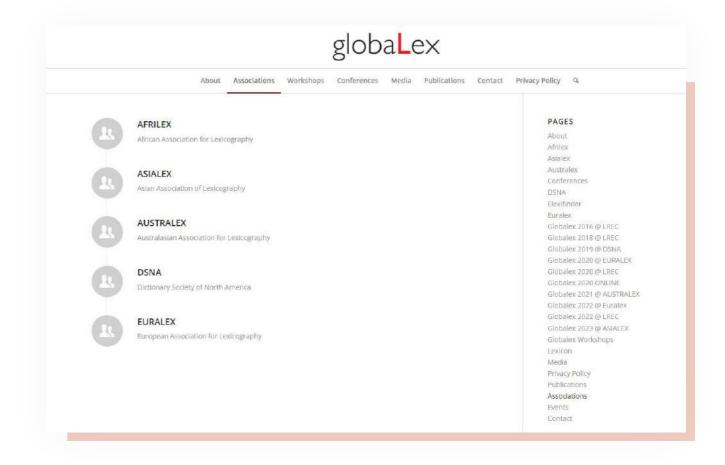
《쉼표. 마침표》 화제를 바꾸어 오는 6월에 아시아렉스(Asialex)에서 주관하는 학술대회에 대하여 질문드리고자 합니다. 먼저 아시아렉스와 관련해 소개를 부탁드립니다.

#### 남길임

일각 대륙마다 그 대륙을 대표하는 사전학회가 있다는 사실을 아시는 분은 그리 많지 않을 거예요. Asialex(The Asian Assocation for Lexicography) 역시 아시아를 대표하는 사전학회이고요, 아시아와 세계 사전의 연구와 실제 사전 편찬을 함께 연구하는 단체입니다. 실제로 아시아사전학회 외에 유럽사전학회(Euralex), 아프리카사전학회(Afrilex), 오스트레일리아사전학회(Ausatralex), 북미사전학회(DSNA) 등이 독립적으로 존재하고, 세계사전학회(Globalex)라는 협업 단체에서 정기 세미나와 학술회의 등을 통해 협력하고 있습니다.

한국 대표자로서 이번 학술회의의 의의에 대해서 두 가지 정도를 말씀드릴 수 있어요. 한국은 지난 2001년도에 연세대학교에서 Asialex 국제학술회의를 개최한 바 있는데요, 이번 2023년 Asialex Seoul은 20여 년만의 두 번째 개최라는 의의가 있습니다. 지난 20여 년간 높아진 한국의 문화, 경제적 위상, 한류 등으로 인해 연구자들이 이번 학술회의에 대한 기대가 매우 높은 것으로 알고 있습니다. 이미 27개국의 연구자들이 100편 이상의 발표 초록을 접수해 주셨어요. 저는 안예리 교수님(한국학 중앙연구원)과 함께 2023년 Asialex 학술회의 개최를 준비하고 있는데요, 이러한 학자들의 관심을 반영해서 학술회의 주제를 "사전학과 인공지능 그리고 사전 사용자(Lexicography, Artificial Intelligence, and Dictionary Users)"로 기획하고, 현장 답사로 한글박물관 견학, 네이버(주) 사옥 답사 등을 포함했습니다.

두 번째 의의는 코로나 이후 최초의 대면 학회라는 시기적 적절성입니다. 이것은 매우 중요한 부분인데요, 2020년 이후 지난 3년 동안 한 번도 대면으로 진행하지 못했습니다. 이번 Asialex 학회에서는 본 학회 프로그램과 더불어 두 개의 워크숍, 한국사전학회의 전국학술회까지 함께 개최될 예정이에요. 두 워크숍은 각각 세계사전학회에서 개최하는 "사전학과 신어-세부 주제: 미등재어(Lexicography and Neology: unregistered words)"와 홍콩 폴리텍 대학 Amy Chi 교수가 주도하는 Lexteach 워크숍입니다. 자세한 사항은 Asialex 2023 홈페이지를 참고하시면 됩니다.

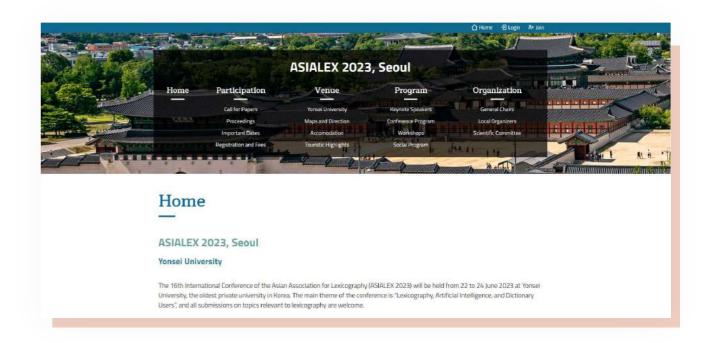


《쉼표, 마침표》 이번 아시아렉스를 통해 기대하시는 바가 있으신가요?

#### 남길임

이번 제15회 아시아사전학회의 기획 주제는 "사전학과 인공지능 그리고 사전 사용자(Lexicography, Artificial Intelligence, and Dictionary Users)"예요. 학술회의 기획 초기에는 주제를 '사전학과 인공지능'이라고 했다가, 결국 인공지능이 모사하는 언어는 인간의 언어 지능이고, 인공지능 시대의 사전이 지향하는 것은 언어 사용자의 언어 직관을 충분히 반영하는 것이라는 점을 좀 더 부각하기 위해 논의를 거쳐 '사전 사용자'를 포함했습니다. 이번 학술회의의 주요 주제는 다음 다섯 가지의 의문으로 요약할 수 있을 듯합니다. 즉 "인공지능 시대에 사전이 필요한 이유는 무엇인가?, 인공지능은 어떻게 언어를 마스터하는가?, 인공지능은 사전 편찬에 어떤 영향을 미치는가?,사용자들은 사전에서 무엇을 기대하는가?,인공지능과사용자사이에서 언어학자의 역할은 무엇인가? "입니다.

이러한 주제에 좀더 구체적으로 접근하고자 저희는 학계와 업계의 기획 발표자들을 모셨어요. 국내에서 The Oxford Guide to Practical Lexicography라는 사전학 개론서의 저자이자 현재 Lexical Computing(주)의 사전 편집장인 Michael Rundell, 사전 사용자 연구의 석학인 Yukio Tono, 세상의 모든 영어를 모은다는 세계 최대의 온라인 사전 Wordnik.com의 창시자 Erin Mckean, 네이버(주)의 글로벌 사전 센터의 리더 김종화, 이렇게 네 분입니다. 일반적으로 사전학 관련 국제 학술회의는 연구자뿐만 아니라 사전의 기획과 출판, 활용을 담당하는 다양한 산업계과 교육 업계에서 참여를 하는데요. 이번 학술회의를 통해 정교한 언어학적, 사전학적 접근과 산업계의 협업의 방향성이 모색되기를 기대하고 있습니다.



《쉼표. 마침표》 끝으로 오늘날 국내외 사전학계가 당면한 과제에 대해 말씀해주세요.

남길임

월터제이옹은 그의 책 〈구술문화와 문자문화〉에서, 사전이야말로 문자문화, 인쇄문화의 최대의 발명품이라고 기술했어요. 몇천 개의 어휘만으로 의사소통이 가능한 구술문화에 존재하지 않던 학술용어나 문학용어를 담은 몇만, 몇십만 어휘 규모의 사전은 문자문화, 인쇄문화의 꽃이라 할 만하지요. 그런데 인쇄문화의 전통적인 발명품이 그 구조와 형식의 측면에서 현재 뉴뉴미디어 시대에도 여전히 적절한지, 전통적인 표제어 등재 원칙이 현대에도 여전히 유효한지 고민이 필요합니다.

이제 글쓰기는 더 이상 전문가들의 전유물이 아니고, 모든 국민들이 글쓰기의 주체로 활약하는 비격식 글쓰기의 시대가 왔습니다. 비윤리적 표현에 대한 문제도 결국 비격식 글쓰기 현상과 깊은 관계가 있다고 생각해요. 이러한 언어 현실은 어떤 교육이나 홍보를 통해 이전으로 되돌릴 수는 없습니다. 사전학적 관점에서는 이러한 언어 환경의 변화를 고려한 새로운 사전 모형에 대한 연구, 전통 사전학의 틀을 벗어난 선제적인 연구가 필요한 시점이라고 봅니다.

우리말, 그리고 사람

불어 국내의 경우, 사전 편찬의 발전에 큰 역할을 해 왔던 대학 기관이나 출판사는 더 이상 사전 편찬의 사업을 지속하기 어려운 상황에 놓여 있어요. 이미 사전학회나 다른 토론회 등에서도 많은 논의가 있었지만 저 역시 사전학의 혁신을 위해서는 국립국어원 등 일부 기관이나 소수의 연구자가 이러한 역할이나 권위를 독점하는 일은 없어야 한다고 생각합니다. 언어와 사전에 대한 다양한 시각을 가진 연구자들이 경쟁과 협업을 통해서 사전을 편찬, 보완하고, 시대와 사용자가 요구하는 사전의 새로운 틀을 함께 고민해야 합니다. 무엇보다 이를 위해서는 더 다양한 분야의 전문 연구자를 양성할 필요가 있어요. 또한 인접 분야와의 소통을 강화하고 대중의 요구에도 충분히 귀를 기울일 필요가 있습니다. 이에 대해서는 제가 Asialex 2023년 기획 취지문에서 사전과 사전학의 사회적 역할에 대해 쓴 내용을 일부 인용하고자 합니다.

"사전은 우리 도처에 있고 우리가 인식하지 못할 뿐 일상 어디에서나 사용되고 있다. 말뭉치 혁명 등 인류 역사상 사전학의 수많은 성과에도 불구하고, 사전에 대한 연구는 여전히 대중에게는 낯선 분야이다. 대중에게 사전은 여전히 어렵고, 사전학은 미지의 영역이다. 사전학자들 스스로가 대중에게 다가감으로써 새로운 시대가 요구하는 사전의 정체성을 모색할 필요가 있다." (Asialex 2023 서문 중 일부)

글: 강은혜

## 우리말 풀기

## 문제 1

김 선배는 (뒤끝/뒤끗)이 있는 사람이다.

정답 찾기 Q

## 문제 2

실적이 우수한 사원에게 (성과급/성과금)을 지급하기로 했다.

정답 찾기 Q

# 문제 3 미운 사위에게는 (매생이/메생이)로 국을 끓여 준다고 한다. 정답 찾기 Q 문제 4 엿장수의 가위 소리가 (아스라이/아스라히) 들려왔다. 정답 찾기 Q 문제 5 할머니는 청국장 맛의 비결이 (해콩/햇콩)으로 만드는 것이라고 했다. 정답 찾기 Q

## '우리말 풀기'에 참여하기

#### 참여 방법

- 1. ['우리말 풀기'에 참여하기]를 눌러 주세요.
- 2. 문제를 풀고, 답을 적어 주세요.
- 3. 응모자의 이름과 연락처를 적어 주세요.
- 4. '제출하기'를 눌러 주세요.

#### 응모 기간

2023. 1. 3.~2023. 1. 16.

#### 당첨자 발표

개인별로 알리고 《쉼표, 마침표》 2023년 1월 31일에 공고함.

### 선물

당첨자 열 분께 커피 교환권을 드립니다.

## 2022년 12월 호 우리말 물기 당철짜

조\*현(6742) 박\*우(1117) 이\*진(1374) 고\*숙(3148) 박\*서(1443) 우\*진(7201) 진\*민(6171) 우\*용(8992) 하\*현(4429) 주\*희(5149)

## 문제 1

김 대리의 제안서를 본 적이 있는데, 일을 참 **(잘하데/잘하대)**.

정답 찾기 Q

## 문제 2

나는 네가 성공하기를 (바라/바래) 마지않는다.

정답 찾기 Q

# 문제 3 그녀는 부하 직원에게 (말발/말빨)을 세우며 업무를 넘겼다. 정답 찾기 Q 문제 4 (오래간만/간만)에 찾은 모교는 변한 것이 없었다. 정답 찾기 Q 문제 5 우리도 이제부터는 (여봐란듯이/여보란듯이) 살아 보자고. 정답 찾기 Q



## "早里里可"创春的初了

#### 참여 방법

- 1. ['우리말 풀기'에 참여하기]를 눌러 주세요.
- 2. 문제를 풀고, 답을 적어 주세요.
- 3. 응모자의 이름과 연락처를 적어 주세요.
- 4. '제출하기'를 눌러 주세요.

#### 응모 기간

2023. 1. 17.~2023. 1. 30.

#### 당첨자 발표

개인별로 알리고 《쉼표, 마침표》 2023년 1월 31일에 공고함.

#### 선물

당첨자 열 분께 커피 교환권을 드립니다.

## 2022년 12월 호 우리말 물기 당철짜

김\*애(3426) 김\*주(2846) 김\*윤(5257) 최\*길(4246) 김\*선(1915) 이\*준(6242) 조\*성(7797) 서\*우(0093) 조\*철(3038)

박\*숙(0608)