

국립국어원, 에스케이텔레콤과 인공지능 한국어 모델 개발 업무협약



국립국어원 온라인 소식지
쉽표, 마침표. | 2021. 4.

실전 국어 표기법(1)

“어찌나 궁금하던지”
‘던지’와 ‘든지’

국어원 30년

바른 국어 생활을 위한 어문 규범 관리

국어원 소식(1)

2021년 한국어 예비 교원 국외 실습 참가자 모집

우리말 다듬기(1)

검색의 기술-핵심어 표시

100년 전 우리말 풍경

신선한 언어 감각으로 그려낸 새로운 시대상

실전 국어 표기법(2)

“네가 있는 데로 갈게.”
‘대로’와 ‘데로’

읽기 좋은 글, 듣기 좋은 말

눈으로 하는 언어생활

사전 두 배로 즐기기

“표준국어대사전”, 아는 만큼 보여요!

우리말 다듬기(2)

세계 무대에 이름을 올리다-후보 지명

우리말, 그리고 사람

인공지능, 한문 고전을 자동으로 번역한다.





목차

쉽표, 마침표. | 2021. 4.

국어 배우기

- 2 실전 국어 표기법 “어찌나 궁금하던지”-‘던지’와 ‘든지’
- 4 “네가 있는 데로 갈게.”-‘대로’와 ‘데로’
- 6 읽기 좋은글, 듣기 좋은말 눈으로 하는 언어생활
- 10 사전 두 배로 즐기기 “표준국어대사전”,아는 만큼 보여요!

국어 알리기

- 14 국어원 30년 바른 국어 생활을 위한 어문 규범 관리
- 18 국어원 소식 2021년 한국어 예비 교원 국외 실습 참가자 모집
- 20 국립국어원, 에스케이텔레콤과 인공지능 한국어 모델 개발 업무협약
- 22 우리말 다듬기 검색의 기술-핵심어 표시
- 24 세계 무대에 이름을 올리다-후보 지명

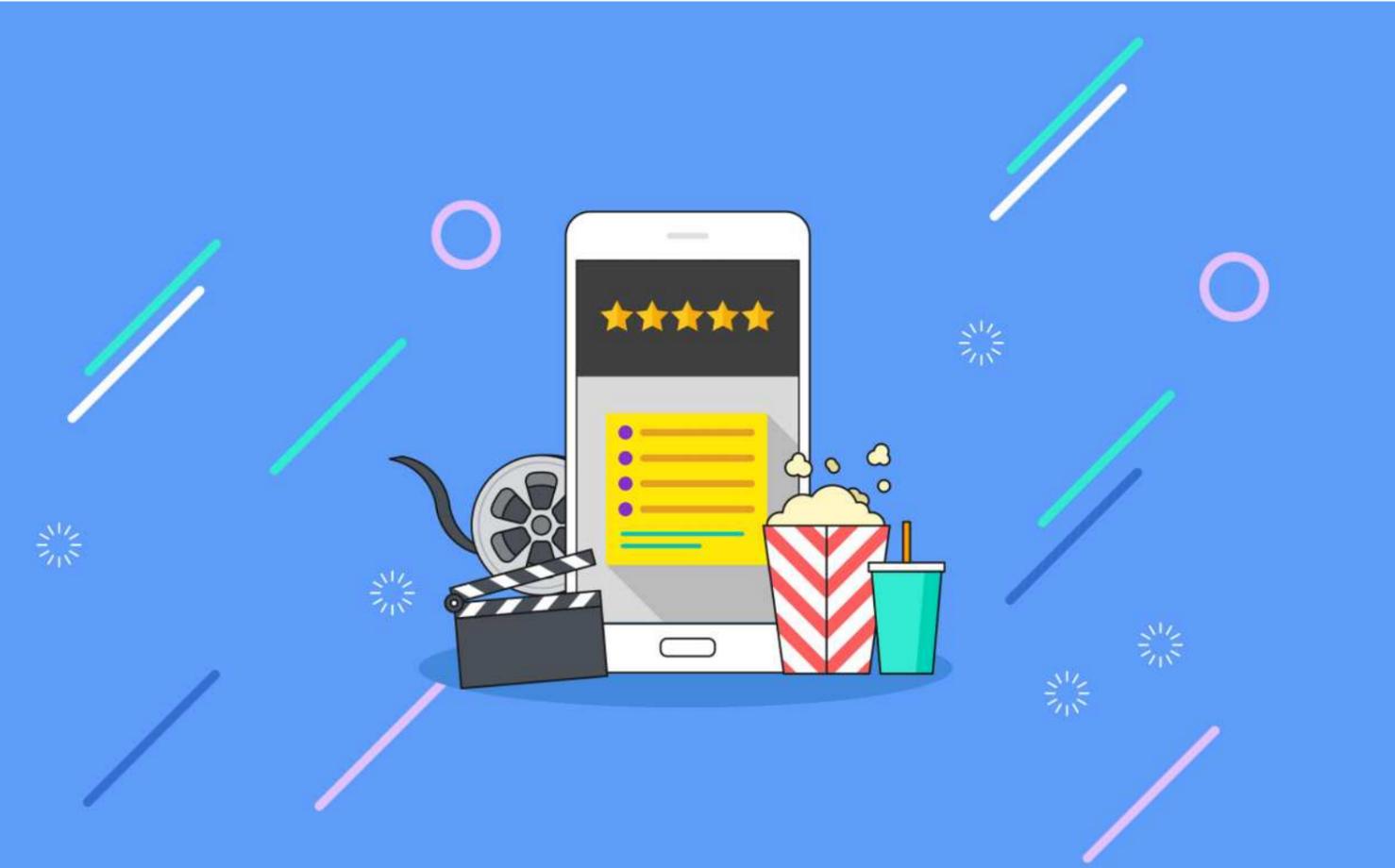
국어로 바라보기

- 26 100년 전 우리말 풍경 신선한 언어 감각으로 그려낸 새로운 시대상
- 32 우리말, 그리고 사람 인공지능, 한문 고전을 자동으로 번역한다.

국어로 함께하기

- 40 우리말 풀기 독자 참여 행사
- 44 독자 참여 행사

“어찌나 궁금해하던지” ‘던지’와 ‘든지’



어찌나 궁금하던지 우리 부모도 곧장 영화표를 예매했어요.

다행히 개봉관이 많아 쉽게 표를 구할 수 있었답니다.

주말 낮 시간대 예약이 모두 마감됐으면 어찌지?

그럼 심야에 보든지 평일에 하루 연차를 내든지 하자.

저 배우는 어떤 배역을 맡든지 잘 소화해.

‘던지’와 ‘든지’

《심포, 마침표》에서는 언제든지 쉽고 재밌는 우리말 정보를 얻을 수 있습니다.

-던지
지난 일을 나타내는 어미
아이가 얼마나 밥을 많이 먹던지 배탈 날까 걱정이 되었다.
동생도 놀이가 재미있었던지 더 이상 엄마를 찾지 않았다.

-든지
선택이나 무관의 뜻을 나타내는 어미
배든지 사과든지 마음대로 먹어라.
싫든지 좋든지 간에 따를 수밖에 없다.

무인발권기

마지막 장면이 어찌나 감동스럽던지, 아직도 마음이 편해.

나가는 곳

“어찌나 궁금하던지” ‘던지’와 ‘든지’

이 주의 화제작

이미 영화를 본 친구들이 감탄을 금치 못하더라고요.

얼마나 재밌던지, 시간 가는 줄 몰랐어.

배우들의 연기가 정말 뛰어나.

누리소통망에서는 지난 주말에 개봉한 영화가 큰 화제였습니다.

“네가 있는 데로 갈게.” ‘대로’와 ‘데로’



#우리_집_강아지와
#오늘도_계획대로
#매일_산책하기

보내준 약도대로 따라와.
그래, 알려준 대로 곧 갈게.

매일 산책하겠다는 연초의 다짐을 말한 **대로** 실천하는 친구였어요.

조금 귀찮기도 했지만 친구가 늘 산책하는 **데로** 나갔답니다.

‘대로’와 ‘데로’

대로

의존명사 '대로'는 '어떤 상태와 같이' 또는 '그 즉시'의 뜻을 나타내는 말로 동사나 형용사 뒤에 오며 앞말과 띄어 씁니다.

- 네가 본 **대로** 말해라.
- 연락이 오는 **대로** 즉시 알려 주세요.

대로

조사 '대로'는 '앞에 오는 말과 같음'을 나타내거나 '따라따로 구별됨'을 나타내는 말로 명사 뒤에 오며 앞말과 붙여 씁니다.

- 문제가 있다면 **범대로** 해라.
- 큰 것은 큰 것**대로** 따로 모아 두렴.

데로

'데로'는 '곳으로'의 뜻을 나타내는 말로 보통 동사나 형용사 뒤에 오며 앞말과 띄어 씁니다.

- 지금 네가 있는 **데로** 갈게.
- 좀 더 높은 **데로** 올라가 보는 게 어때?

“네가 있는 데로 갈게.” ‘대로’와 ‘데로’의 구분

평일에 일이 많아 지칠 **대로** 지친 저는 지난 주말 늘어진 채로 하루를 보내고 있었어요.

주말이라고 누워만 있지 말고 밝은 **데로** 좀 나와.

때마침 친구에게서 연락이 왔죠.

오랜만에 걸으니까 좋은데?
그래, 듣나는 **대로** 걸기만 해도 건강해진다고.

자주 걸으면 몸과 마음이 튼튼해지는 것처럼 기회가 생기는 **대로** 우리말 지식을 익히면 말글살이도 더 건강해질 거예요.

눈으로 하는 언어생활



글: 이미향(영남대학교 국제학부 교수)

옆 차선에서 나란히 달리는 버스의 옆면에 큼지막하게 써진 광고가 눈에 들어온다. “사랑... 해 보셨습니까?” 새로 나온 영화를 홍보하는 문구 같다. 단순한 의미인데, 스치는 찰나에도 보는 이에게 여운을 남긴다. 그저 묻는 것이 아니라, 누군가의 가슴 아픈 사랑을 곁에 앉아 전해 듣는 기분이라 할까? ‘광고주의 성공은 소비자의 실패’란 말을 알면서도 글에다가 큰따옴표까지 써 가며 전하는 메시지에 모든 경계를 풀어 버렸다. 문장 부호의 힘을 그렇게 크게 느껴본 적은 처음이었다.

문장 부호를 잘못 쓰거나 안 써서 문제가 생긴 경험이 있는가? 영어 받아쓰기를 처음 할 때 겪은 일이다. 문장 끝에 마침표를 안 찍어서 애써 외워 쓴 10문제에서 0점을 받았다. 소릿값도 없는 마침표가 점수를 결정하다니! 그렇지만 모국어든 외국어든 문장 부호도 그 언어의 규칙이므로 번거롭다고 하여 이견을 내세울 수 없다. 선생님의 단호한 채점으로 문장 부호도 맞춤법의 일종이라는 것을 알게 되었다.

우리는 귀로 하는 언어생활과 말로 하는 언어생활, 두 가지 방식으로 살고 있다. 현대인에게 글을 보는 일은 더 늘었다. 굳이 도서관에서 책을 보지 않더라도, 스마트폰에서 정보를 얻고 각종 표지판과 간판에서 필요한 정보를 찾으며 하루를 보낸다. 그런데도 우리는 눈으로 하는 글살이를 잇을 때가 많다. 당연한 듯 쓰고 있는 글자 한 자 한 자는 처음부터 인류에게 허락된 것이 아니라 필요에 의해 인류가 고안한 발명품이다. 그 글자로 공적 자료를 남기고, 잊으면 안 될 경제 활동도 기록하고 누리소통망에 하루 일과도 기록한다. 사람에게는 ‘눈으로 하는 언어생활’이라는 신세계가 하나 더 있다.



“우리 아이가 받아쓰기를 잘 못해요. 어떻게 하면 좋을까요?” 학부모를 대상으로 하는 특강에서 자주 듣는 질문이다. 받아쓰기의 고민을 들어보면 집집마다 상황이 비슷하다. 아이 말이, 선생님이 ‘감니다’라 해서 ‘감니다’라고 받아썼는데 부르는 대로 쓴 것이 무슨 잘못이냐는 것이다. 이 말에 학부모들은 말문이 막힌다고 하지만, 그때는 ‘귀로 듣는 말도 있지만, 눈으로 보는 말도 있다고 아이에게 말해 주세요.’라고 조언한다. 우선 언어생활이 두 가지로 구분된다는 것과, 그러므로 눈으로 보는 언어생활을 하기 위해 들은 것을 고정된 형태로 바뀌어서 적는 능력이 필요하다는 것을 알게 하자는 것이다.

글에는 읽을 사람을 전제로 한 약속이 많다. 표기법은 쓰는 사람의 뜻을 잘 전하면서 읽는 사람이 임의대로 해석하는 낭패를 막는 일종의 장치이다. 고대 암각화를 예로 들어보자. 단단한 암벽에 한 자 한 자 새겨 간 사람은 자신의 뜻을 표현하기 위해 어떤 그림을 어느 자리에 넣을지 무척 고민했을 것이다. 그러나 정작 더 힘든 사람은 그것을 마주하는 이들이다. 표현하려는 쪽과 이해해야 하는 쪽 중에서 소통에 실패할 확률이 더 높은 이는 누구일까? 남의 생각을 읽어 내야 할 사람들이 아니겠는가? 쓴 사람의 마음을 정확하게 읽어 내기란 쉽지 않을뿐더러, 둘 이상으로 해석될 때 판단할 근거마저 없다면 읽기란 풀리지 않을 숙제가 될 것이다.

읽기 좋은 글은 글살이에서 지킬 약속을 지켜 쓴 글이다. 흔히 표기, 맞춤법, 문장 부호 등의 규칙을 따르는 일은 고통으로 기억되기 쉽다. 그러나 표기법은 글쓴이를 제약하는 쇠사슬이 아니라, 글쓴이의 뜻을 명확히 전하는 데 기여할 통로이다. 예를 들어 영어에서 ‘8’은 ‘eight’라 적는다. 실제 발음 [eit]와 비교해 보면, 소리 때문이 아니라 눈으로 ‘8’이라 알아보는 데 필요한 문자가 들어가 있다. ‘know(알다), knight(기사)’에서 소리 나지 않는 ‘k’를 써야 하는 까닭도 각각 ‘no(아니), ‘night(밤)’와 눈으로 구분하기 위해서다.

한자를 쓸 때는 획이나 점 하나도, 그 획을 긋는 방향 하나도 개인이 함부로 할 수 없다. 문자는 인상을 담은 풍경화가 아니며 상상은 더욱 아니다. 문자가 제 역할을 하려면 단순함이나 편리함보다 더 중요하게 앞세울 것이 많다. 그런 이유로 받아쓰기의 고통은 어디에도 있는 것이다.

다시 마침표까지 꼭꼭 찍던 이야기로 되돌아가 본다. 스페인어 문장 가장 앞에는 맑은 호수에 비친 듯 뒤집혀 그려진 느낌표나 물음표가 먼저 찍힌다. 스페인어 글살이에는 감탄문, 명령문, 의문문에서 ‘!(느낌표)’와 ‘?(물음표)’를 문장의 앞과 뒤에 다 쓰는 규칙이 있다. 그 이유를 물으니, 필자가 하려는 감탄, 명령, 질문 등을 문두에서부터 독자가 안다면 소통이 더 잘 되지 않겠느냐고 답한다. 그렇게 생각을 바꾸니 제약으로 보이던 표기법이 표현의 든든한 지원군이 된다.

‘쓴다’는 것은 언제 어디서나 읽을 이를 전제한다. 간판과 상품명도, 식품과 약품에 적힌 유통 기한이나 주의 사항 모두 읽을 사람에게 메시지를 전하고 있다. 용돈을 더 올려 달라는 아이의 손편지, 학교에 게시판에 올리는 누군가의 하소연 등은 더 길고 중요한 내용을 담고 있을지도 모른다. 이 중 어떤 글도 혼자만 알아볼 비밀 일기가 되어서는 안 된다. 개인의 소통에서나 공적 언어 생활에서나 똑같이, 독자를 방황하지 않게 배려하며 글 세상의 약속을 지켜 쓰는 것이 좋은 글을 만든다. 읽기 좋은 글은 쓰는 사람의 마음 밭에서 시작되는 것이다.

“표준국어대사전”, 아는 만큼 보여요! - 첫 번째 이야기



글: 이윤영(국립국어원 학예연구관)

“사전”이라고 하면 흔히 모르는 단어의 뜻을 알고 싶을 때 찾아보는 것이라고 생각한다. 실제로 단어의 의미를 정확하고 쉽게 풀이해 주는 것은 사전의 중요한 기능 중 하나이다. 그러나 사전에 ‘단어 뜻’만 있는 것은 아니다. 사전에는 흔히 생각하는 것보다 훨씬 많고 다양한 정보가 들어 있다. “표준국어대사전”도 그러하다. “표준국어대사전”은 1999년 책자로 처음 발간되었는데, 당시 국어학 관련 전문가들이 모여 많은 논의 끝에 사전에 담을 정보를 결정하였고, 그러한 만큼 “표준국어대사전”에는 발간 당시의 국어학 논의의 결과물들이 많이 담겨 있다. 또한 국가에서 발간하는 사전인 만큼 언어생활의 지침이 되는 어문 규범과 관련한 사항도 담고자 하였다. 따라서 일반적으로 사전이 담고 있는 원어나 뜻풀이, 용례 등의 정보 외에도, 단어의 문법적 쓰임이나 단어가 쓰인 문장의 문형 정보, 규범과 관련한 부가 정보 등이 훨씬 다양하게 제시되어 있다. 이제 “표준국어대사전”에 실려 있는 정보를 항목별로 살펴보고, 실제로 이러한 정보가 어떤 의미가 있고, 어떻게 활용될 수 있는지 살펴보겠다.



표준국어대사전에 실려 있는 항목을 모두 나열하면 다음과 같다.

표제어, 원어, 발음 정보, 활용 정보, 품사, 전문어 영역, 문형 정보, 문법 정보, 뜻풀이, 용례, 부가 정보, 관련 어휘, 참고 어휘, 최초 출현형, 다중 매체 자료, 관용구, 속담

이제 위에 제시한 항목에 대해 구체적으로 살펴보겠다.

(1) 표제어

1999년 발간 당시 “표준국어대사전”에는 약 510,000개의 표제어가 실려 있었다. 기존 국어사전에 담긴 표제어, “조선말대사전” 표제어가 모두 검토 대상이었고, 그 결과 북한어 38,000여 개도 사전에 등재되었다. 2019년 개편을 거치면서 북한어, 방언, 옛말은 “우리말샘”에서 제공하고 있고, “표준국어대사전”에는 현재 약 42만 개의 표제어가 실려 있다. 여기에는 일부 잘못된 표기도 실려 있는데, 많이 쓰는 잘못된 표기에 대한 바른 표기를 제공함으로써 규범 사전으로서의 역할도 하고자 함이다.

“표준국어대사전”을 찾아보면 표제어의 형태가 크게 3가지로 나온다. ‘하늘’처럼 아무 표시 없이 한글로만 구성된 것이 있고, ‘발생-하다’처럼 중간에 불임표가 들어간 것이 있으며, ‘모음^조화’처럼 중간에 ‘^’ 기호가 들어간 것이 있다. 무심코 지나치는 사람이 많겠지만 여기에는 단어 구성 및 띄어쓰기에 대한 중요한 정보가 포함되어 있다. ‘하늘’처럼 아무 표시가 없는 것은 단일어라고 볼 수 있다. 즉 더 이상 분석이 되지 않는 단어이다. 반면 ‘발생-하다’처럼 불임표가 들어간 것은 복합어이다. ‘발생-하다’는 어근 ‘발생’에 접미사 ‘-하다’가 붙은 파생어이고, ‘봄-바람’은 어근 ‘봄’과 ‘바람’이 결합한 합성어이다. 표제어만 보고도 단어의 부류를 알 수가 있는 셈이다. 간혹 불임표를 띄어 쓰는 기호로 오해하여 ‘봄-바람’을 ‘봄 바람’으로 띄어서 써야 하나냐는 질문을 받는데, 그렇지 않다.

붙임표는 어디까지나 단어의 내부 구성을 보여 주는 것이고, 글에서 쓸 때에는 ‘발생하다’, ‘봄바람’처럼 붙여서 써야 한다. 반면 ‘ㅅ’ 기호는 띄어 쓰는 것이 원칙이되 붙여 쓸 수 있음을 보여 준다. 한글 맞춤법 제50항에 따르면, 전문어는 띄어 쓰는 것이 원칙이되 붙여 쓸 수 있다. 바로 이러한 전문어의 띄어쓰기 원리를 보여 주는 것이 ‘ㅅ’ 기호이다. 따라서 ‘모음^조화’는 ‘모음 조화’라고 띄어서 쓰는 것이 원칙이되 ‘모음조화’처럼 붙여 쓰는 것도 가능하다. 이처럼 표제어 항목도 단순한 표기만 보여 주는 것이 아니라 단어 분석, 띄어쓰기 같은 중요한 문법적 내용을 담고 있다.



(2) 원어

원어는 해당 단어가 어디에서 온 것인지를 알려 준다. “표준국어대사전”에 실린 항목은 크게 4가지 부류로 나뉘는데, 고유어, 한자어, 외래어, 그리고 이들이 섞인 혼종어이다. 해당 항목이 이 중 어디에 속하는지는 원어를 보면 알 수 있다. 원어 정보가 제공되지 않은 것은 고유어, 원어 자리에 한자가 제공된 것은 한자어, 로마자가 제공된 것은 외래어이다. 실제로 원어는 로마자가 아닌 다른 문자를 사용하는 언어권에서 온 것도 있으나 로마자 외의 문자를 사람들이 이해하기는 어렵기 때문에 모두 로마자로 표시하였다.



원어는 표제어 바로 옆 괄호 안에 제시되는데, 역삼각형을 누르면 세부 정보를 알 수 있다. 위에서 ‘니르바나’는 산스크리트어이지만 표기는 로마자로 제시하고 있는 것을 볼 수 있다.

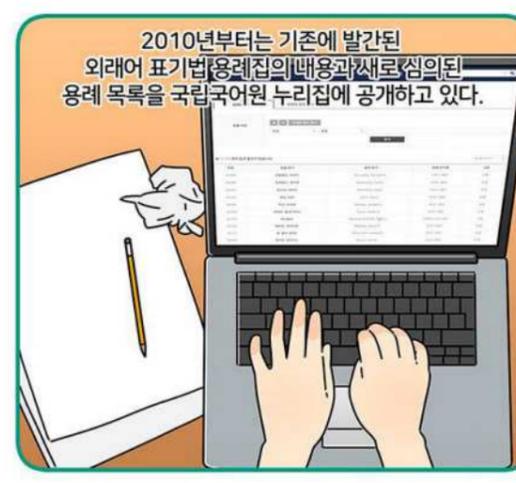
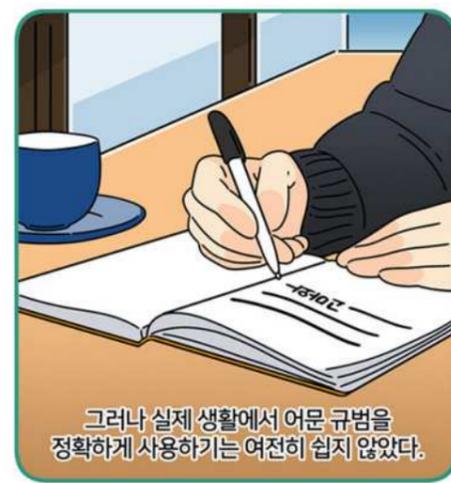
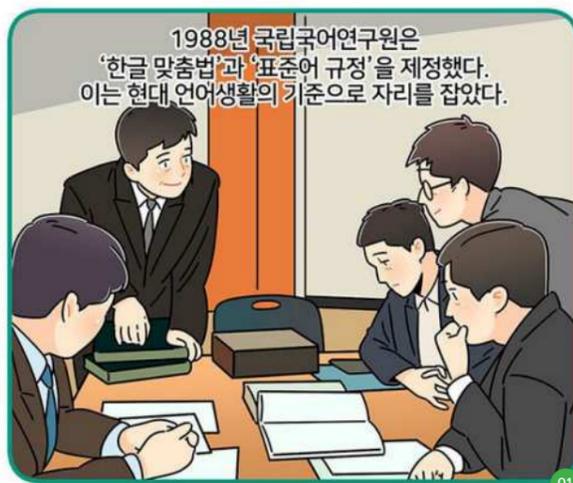
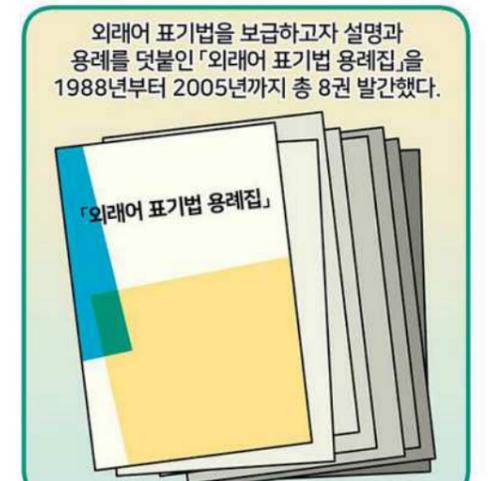
(3) 발음과 활용 정보

우리말은 표기한 대로 발음하는 것이 원칙이지만, 실제로 표기와 발음이 다른 경우도 상당히 많다. 또한 단어에 조사나 어미가 붙어 활용을 하게 되면 발음은 더 다양해진다. 이러한 발음 정보도 사전에서 확인할 수 있다. 표제어 아래 발음과 활용 정보가 나란히 나타나며, 활용한 형태의 발음 정보도 표시해 주고 있다.



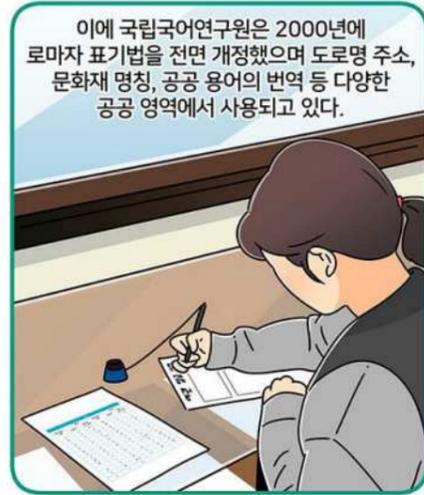
사전의 발음과 활용 정보를 보면, ‘수탕’은 단독으로 쓰일 때에는 [수탁]으로 발음하지만 조사 ‘이’, ‘만’이 붙었을 때에는 [수탈기], [수탕만]으로 발음한다는 것을 알 수 있다. 확장기 표시를 누르면 실제 발음도 들 수 있다. 용언은 활용 정보가 특히 중요하데, “표준국어대사전”에서는 어미 ‘-어/니’가 붙은 활용형을 대표로 보여 주고 있다. 이에 따라 ‘주다’에는 ‘주어/주니’가 제시되어 있다. 그런데 위에 보인 ‘머무르다’에는 ‘머무르어*’가 아닌 ‘머물러’가 제시되어 있고, ‘머물다’에는 ‘머물어*/머물니*’가 아닌 ‘머무니’만 제시되어 있다. 이는 ‘머무르다’와 ‘머물다’가 특이하게 활용하는 동사이기 때문이며, 이런 내용을 활용 정보로 보여 주고 있는 것이다. 이는 한글 맞춤법 제18항과 표준어 규정 제16항에서 규정하고 있는 것으로, 이처럼 어문 규범의 주요 내용도 사전에서 확인할 수가 있다.

바른 국어 생활을 위한 어문 규범 관리





09



10

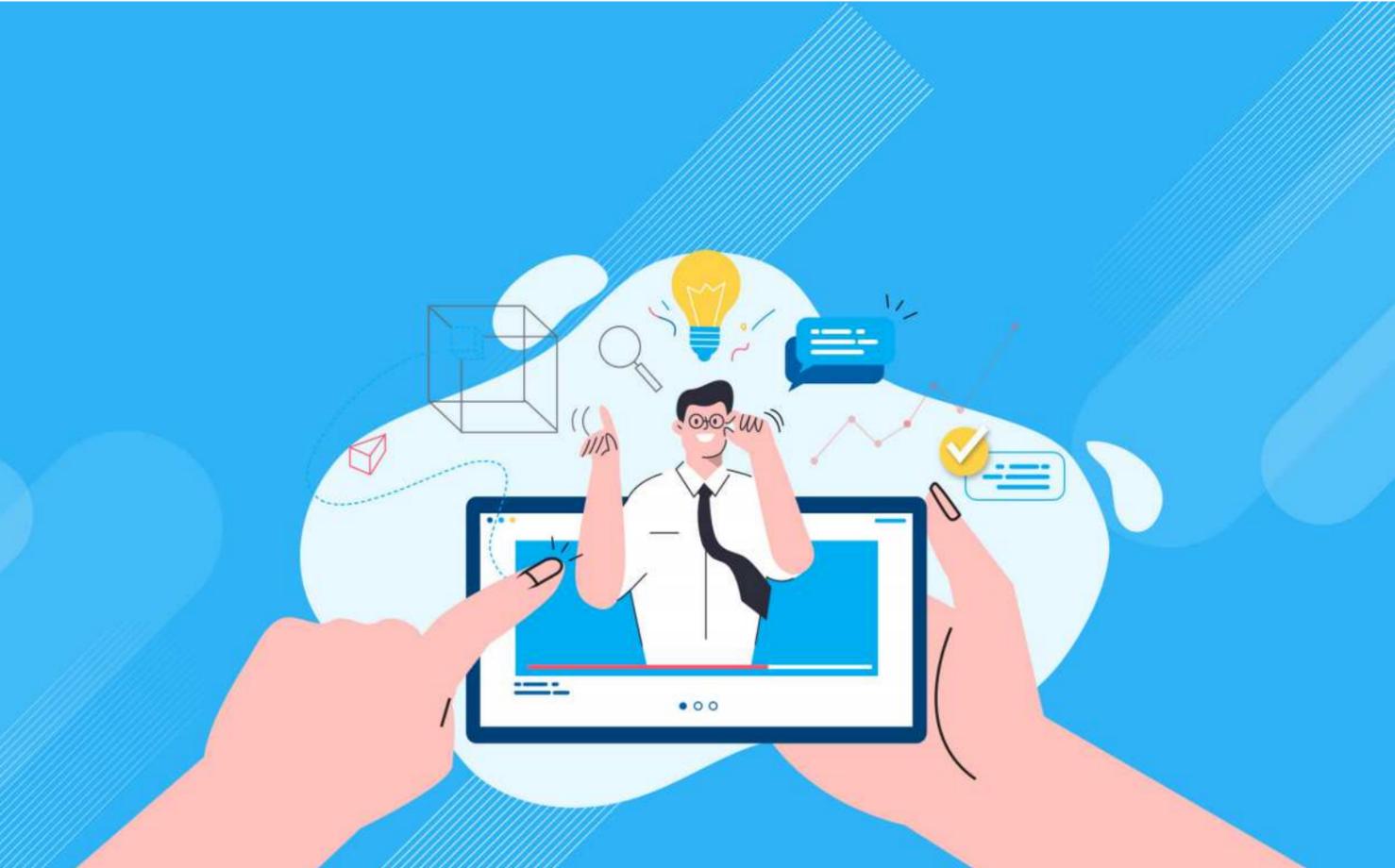


11



12

2021년 한국어 예비(신규) 교원 국외 실습 참가자 모집



실습 목적

한국어 예비(신규) 교원의 교육 현장 경험 확대 및 역량 강화
국내외 한국어 교육 기관의 우수 교원 확보에 기여

실습 지역

러시아, 우즈베키스탄, 카자흐스탄, 키르기스스탄

실습 내용 및 기간

내용	온라인 사전 연수		온라인 국외 실습
	개인 과제	합동 연수	
기간	8~9월 중		9~10월 중
	4주 80시간		3주 60시간
방식	비실시간	비실시간+실시간	비실시간+실시간

※ 실습 일정 및 기간은 현지 사정에 따라 변경될 수 있음.

선발 인원 및 선발 방법

인원: 총 48~54명

방법: (1차) 서류 심사, (2차) 모의 수업 및 면접 심사

접수 방법

우편 접수만 가능

07511 서울특별시 강서구 금남화로 154 국립국어원 한국어진흥과

<한국어 예비(신규) 교원 선발> 담당자 앞

※ 접수 마감일(4.9.) 우체국 소인까지 유효함.

지원 자격 요건

신규 교원 (자격증 취득자)	예비 교원(자격증 취득 준비자)			비학위과정 (양성과정)
	대학	대학원	학점은행제	
취득 후 1년 이내의 무경력 교원	한국어교육 (부)전공 3학년 이상	한국어교육 전공 석사 과정 2학기 이상	외국어로서의 한국어학 전공 표준교육과정 (국가평생교육진흥원 학점은행 공표)에서 전공필수 7과목 이수	한국어교원자격 취득에 필요한 영역별 필수 이수 시간 (120시간 이상) 수료

자세한 사항은 '국립국어원 누리집>공지 사항'에서 확인할 수 있다.

국립국어원, 에스케이텔레콤과 인공지능 한국어 모델 개발 업무협약



국립국어원(원장 소강춘)은 4월 7일(수) 에스케이텔레콤(대표이사 박정호)과 업무협약을 체결하고, 한국어에 최적화된 차세대 인공지능(AI) 모델을 개발하는 데 적극 협력하기로 하였다.



▲ 2021. 4. 7. 업무 협약 체결식

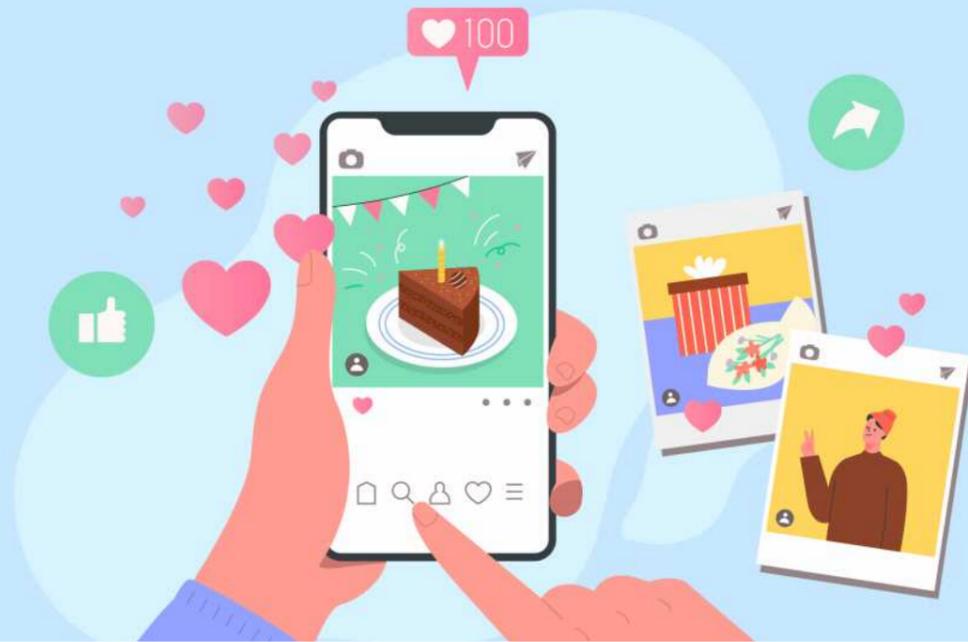
국립국어원은 인공지능 한국어 학습 자료인 한국어 말뭉치, 한국수어 말뭉치 등 다양한 언어 정보 자원을 에스케이텔레콤과 공유하고, 에스케이텔레콤은 이를 활용하여 사람처럼 자연스럽게 대화하는 차세대 인공지능 한국어 범용 언어 모델을 개발하기로 하였다. 새로 개발되는 모델은 내부 검증을 거쳐 다양한 서비스에 적용되며, 한국어 언어 모델과 데이터의 품질 평가에 관한 연구도 추진된다. 또한 국립국어원이 주최하는 ‘2021년 국어 정보처리 시스템 경진대회’를 에스케이텔레콤의 기존 인공지능 언어 모델을 활용하여 인공지능의 언어소통 능력을 겨루는 방식으로 개편하여 개최하기로 하였다.

정회원 국립국어원 어문연구실장은 “앞으로도 인공지능 개발에 필요한 언어 정보 자원을 지속적으로 발굴하여 관련 산업계와 학계에서 유용하게 활용될 수 있도록 시스템을 구축해 나가겠다.”라고 말했다.

문화체육관광부와 국립국어원은 2018년부터 인공지능 학습용 한국어 말뭉치를 구축하여 ‘모두의 말뭉치’에 공개하고 있다. 이 말뭉치는 지난해 10월 에스케이텔레콤이 개발한 인공지능 언어 모델 ‘코바트(KoBART)’를 비롯하여 관련 산업계와 학계에서 유용하게 활용되고 있다.

검색의 기술 핵심어 표시 해시태그

우리말
다들기



우리말
다들기

우리말
다들기

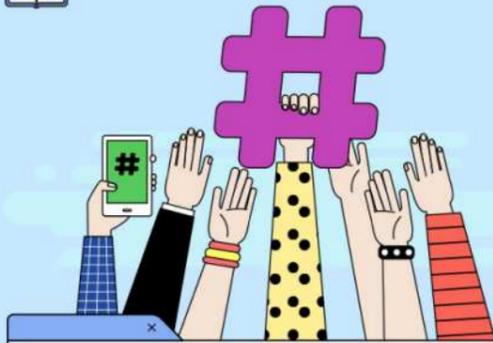
검색의 기술 핵심어 표시

해시태그

#해시태그_대체어_핵심어_표시

#소셜_네트워크_서비스_대체어_누리_소통망_서비스

핵심어 표시를 사용하면 누리 소통망에서 원하는 정보를 쉽게 찾을 수 있어요.



검색하고 싶은 단어 앞에 샵프 기호(#)를 붙여 쓰면 관련된 내용만 골라볼 수 있죠

인기 화제일수록 핵심어 표시가 더 많이 달리는데요
#핫토픽_대체어_인기_화제

자주 언급되는 핵심어 표시를 눈여겨보면 사람들의 관심사나 유행을 알 수 있어요

요즘에는 핵심어 표시가 의사 표현의 한 방식으로 널리 쓰이고 있어요

핵심어 표시를 달아 특정 의견을 지지하거나 참여 잇기에 동참할 수 있죠
#챌린지_대체어_참여_잇기

생각을 표현하는 또 다른 방법, 핵심어 표시. 다들기 말로 더 쉽게 표현해 보세요!

오늘의 다들기 말

- 해시태그 → 핵심어 표시
- 소셜 네트워크 서비스 → 누리 소통망 서비스
- 핫토픽 → 인기 화제
- 챌린지 → 참여 잇기

세계 무대에 이름을 올리다 후보 지명



우리말 다듬기

세계 무대에 이름을 올리다
후보 지명
노미네이트

매해 4월이면
전 세계가 주목하는 영화인들의 축제가 있죠.

우리말 다듬기

황금빛 트로피로 유명한
'아카데미 시상식'입니다.

우리말 다듬기

#노미네이트 대체어 후보 지명

아카데미상은 영화 산업의 중심인 미국에서도
가장 영예로운 상이라
후보 지명만으로도 큰 의미가 있죠.

지난해에는 우리 영화가
아카데미 시상식에서 4관왕을 차지하며
많은 사람을 놀라게 했는데요.

우리말 다듬기

#신 스틸러 대체어 명품 조연

작품과 연출력은 물론, **명품 조연**의 연기까지
세계인의 극찬을 받았습니니다.

우리말 다듬기

올해도 한국 배우가 여럿 출연한 영화가
이미 **후보 지명**을 받았다고 하네요.

우리말 다듬기

#네티즌 대체어 누리꾼

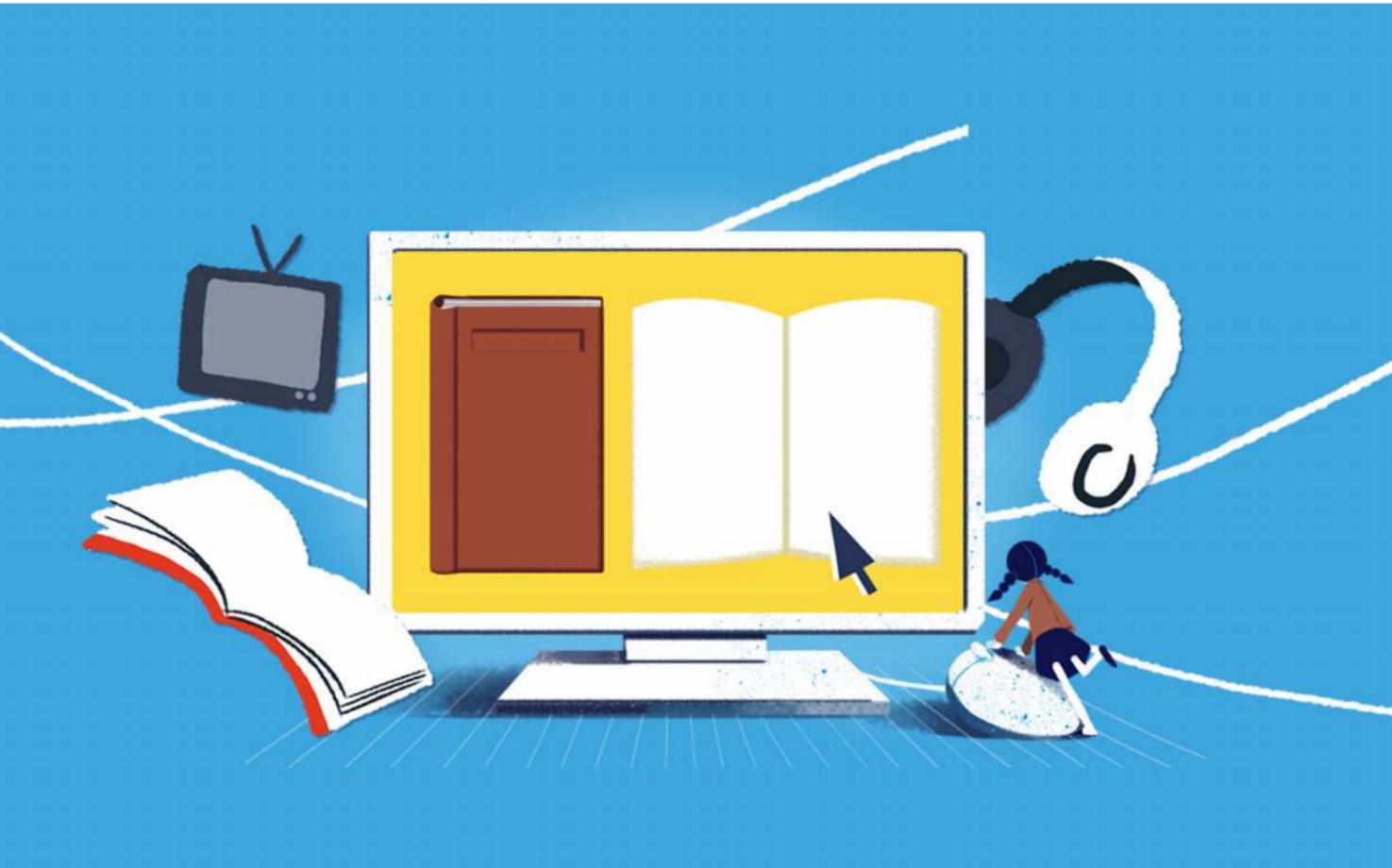
누리꾼들의 예상대로 좋은 소식이 들릴지,
한번 기대해 볼까요?

우리말 다듬기

오늘의 다듬은 말

노미네이트 → **후보 지명**
신 스틸러 → **명품 조연**
네티즌 → **누리꾼**

신선한 언어 감각으로 그려 낸 새로운 시대상



글: 안예리(한국학중앙연구원 부교수)

종이책에서 전자책으로, 그리고 오디오 북까지, 21세기의 독서 방식은 점점 다양해지고 있다. 이러한 변화는 컴퓨터, 태블릿, 킨들, 스마트폰 등 전자기기의 발달에 따른 것이다. 20세기가 시작되던 무렵에도 인쇄 기계와 연필자의 도입, 종이의 대량 생산과 같은 기술의 발달이 독서 문화에 큰 변화를 가져왔다. 여기에 한글 사용의 확대, 근대 교육의 보급 등 문화적, 제도적 변화가 맞물리며 대중적 독서물이 급격히 성장했다.



▲ <그림 1> 신소설 표지: 이인직의 『치악산』(1908), 안국선의 『금수회의록』(1908), 이해조의 『춘외춘』(1912)
(국립한글박물관 제공)

20세기 초에는 당대의 급변하던 사회상을 다룬 소설들이 특히 큰 인기를 끌었다. ‘신소설’로 불리던 이들 작품은 등장인물들의 생생한 목소리를 통해 당시 한반도 안팎에서 벌어진 사건들이 개인의 삶에 미친 영향을 그려냈다. 갑오개혁 이후의 신구(新舊) 갈등을 한 가정의 가족사로 풀어 가기도 했고, 청일전쟁으로 고아가 된 주인공이 천신만고 끝에 유학을 가며 벌어지는 인생사를 통해 근대식 교육의 중요성을 강조하기도 했다.

이인직, 『치악산(상)』(1908), 23면.

(서울 개화파 이 판서의 딸로 원주에 시집을 온 이 씨 부인이 남편 백돌에게 유학을 권하는 장면. 개화에 적극적인 친정과 보수적인 사댁의 분위기가 대조적이다.)

부인이 그 남편이 타국으로 공부하러 가되다 후는 말을 듣고 얼굴에 근심하던 빛을 감추고
(부인) 여보시오 그런 마음 잊거든 후로뱃비 셔물을 나가서 우리 아버지께 말삼만 후면 아버지께서는 당장이라도 치링도 후여 주실 거시오 몇히던지 공부후실 동안에 학비도 넉넉히 디여 주실 터이니 후로뱃비 떠나시오 그러나 여의 아버지께서 허락을 후실나구.....
(백) 허허허 마누라는 기화흔 친정 아버지를 자랑후는 말이오구려 우리 아버지는 완고의 마음이니니 아들더러 외국에 가서 공부후라 후실 리가 잇나

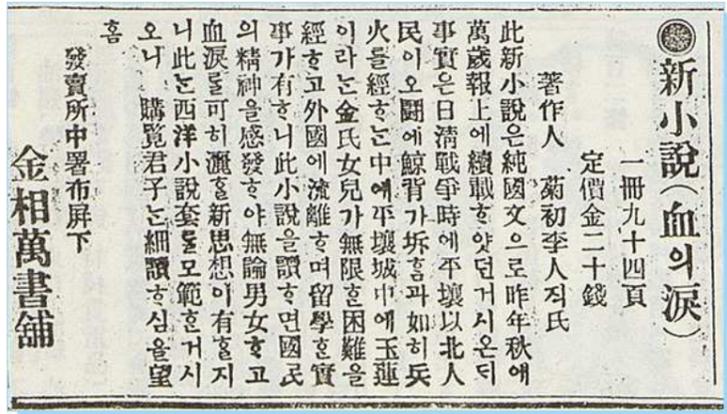
신소설 작가들은 주인공의 입을 통해 조혼의 폐습을 비판하고 신학문 도입의 중요성을 강조하였으며, 미신 타파 등 풍속 개량을 촉구하였다. 전문이 한글로 작성된 신소설은 당시 한자를 모르던 사람들이 즐길 수 있는 거의 유일한 독서물이었는데, 인구의 대다수가 한문을 모르는 비식자층이었던 당시의 상황을 고려할 때 신소설이 갖는 문화적 파급력은 결코 작지 않았을 것이라 생각된다.

물론, 신소설이 한글로 작성된 최초의 소설은 아니다. 신소설의 등장 이전에도 이른바 ‘고전 소설’을 통해 한글 소설의 전통이 이어져 왔다. 임진왜란 이후 중국의 한문 소설이 국내에 들어와 한글로 번역되어 사대부 여성과 중인들 사이에서 인기를 끌었고, 판소리계 소설 등 창작 소설도 조선 후기 동안 널리 읽혔다.



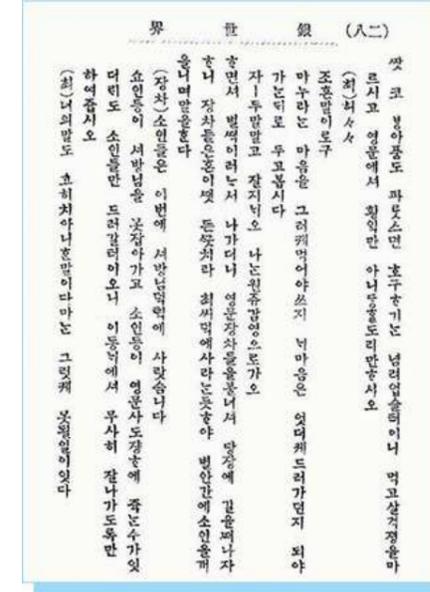
▲<그림 2> 조선 후기 고전 소설 『심청전』 필사본
(국립한글박물관 제공)

신소설, 즉 ‘새로운 소설’이라는 명칭은 바로 이 고전 소설과의 차별성을 강조하기 위한 것인데, 이 명칭은 후대에 붙여진 것이 아니라 당대에 붙여진 것이었다. 최초의 신소설 작품으로 알려진 『혈의누』의 신문 광고에서 ‘신소설’이라는 표현을 볼 수 있으며, 작품의 표지에도 ‘신소설’이라는 수식어를 붙인 경우가 많았다. 그렇다면, 1900년대 독자의 눈에 비친 신소설 작품들은 어떤 면에서 새롭다는 감각을 불러일으켰을까?



▲<그림 3> 1907년 6월 2일자 『만세보』에 실린 신소설 『혈의누』 광고:
줄거리를 소개한 뒤 이 소설이 새로운 사상을 담고 있고 서양 소설의 문투를 본떴다고 하였다.

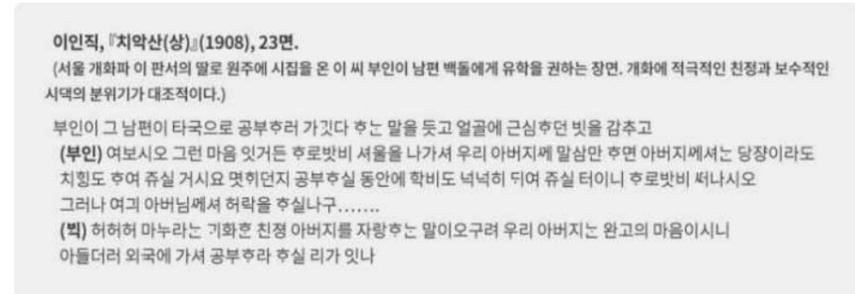
고전 소설과 신소설은 그 내용과 주제, 형식과 문체, 발행 형태 등 여러 측면에서 뚜렷이 구별되었을 뿐 아니라 언어 사용 양상에서도 큰 차이를 보였다. 고전 소설과 달리 신소설에서는 대사가 지문과 분리되어 등장인물의 목소리가 간접 화법이 아닌 직접 화법으로 제시되었다. 또한, 사건의 전개에 치중하던 고전 소설과 달리 신소설에서는 생생한 장면 묘사가 중시되었다.



▲<그림 4> 신소설 『은세계』(1908) 본문: 지문과 대사가 분리된 것을 볼 수 있다.
신소설의 대사를 통해 20세기 초 구어의 모습을 엿볼 수 있다.

이러한 새로운 서술 기법 덕분에 신소설에서는 구어적 성격이 강한 *상징어의 쓰임이 두드러졌다. 코를 실룩실룩하고, 연기가 무력무력 나오고, 휘적휘적 걸어가고, 호독호독 통탕통탕 소리가 나고, 잔디를 부드부드 뜯고, 졸려서 눈이 푹푹 감기는 등 신소설의 지면에는 여타 자료들에서 보기 힘든, 생동감 넘치는 100년 전 한국어의 모습이 담겨 있었다.

*상징어: 소리나 모양, 동작 따위를 흉내 내는 말. 의성어와 의태어로 나뉜다.



격동의 시대를 살았던 인물들의 희로애락을 다룬 이야기에는 우는 장면이나 웃는 장면이 자주 등장한다. 그런데 1900년대의 울고 웃는 모습에 대한 묘사는 오늘날과는 다소 차이가 있어 보인다. 신소설에는 ‘킁킁’ 웃는 사람이나 ‘엉엉’ 우는 사람은 등장하지 않는다.

동사	상징어	현대 표준어
울다	비죽비죽	서운섭섭후야 비죽비죽 울며 <1912 추월색, 072>
	비죽비죽	옥남이가 그 말을 하며 비죽비죽 우니 <1907 고목화, 95>
	꿀꿀	치마자락으로 눈을 가리고 울울 늦겨 우니 <1908 빈상설, 36>
	질금질금	소견머리 업시 질금질금 울어서 <1912 추풍감수록, 35>
	쪽쪽	계집년이 먼저 쪽쪽 울며 권가 압흐로 가 안드니 <1907 고목화, 17>
	훌쩍훌쩍	훌쩍훌쩍 울다가 안방 압 툇마루 아리 가 오도카니 셋느디 <1908 빈상설, 36>
	훌적훌적	제 설름이 북받쳐 훌적훌적 울다가 <1911 목단화, 80>
웃다	빙글빙글	근 오십 된 여인이 빙글빙글 우스며 나오더니 <1908 빈상설, 79>
	빙긔빙긔	리치슈는 정숙의 말을 듯고 코우숨을 빙긔빙긔 웃고 <1911 목단화, 65>
	깔깔	거리에 잇는 우희들이 다 깔깔 웃는다 <1912 만인계, 53>
	꺄꺄	양덕이 그 말을 듣고서야 깨다라셔 꺄꺄 우스며 <1908 원양도, 30>
	덩싹덩싹	자긔 소싱 허송을 덩싹덩싹 웃는 낫헤 피눈물노 이별하고 <1912 명월정, 122>
	빙긔	상현이는 그 말을 듯더니 빙긔 웃고 외면을 후더라 <1914 안의성, 034>
	생긔	지혜 만코 약은 금년이 싱긔 웃스며 <1911 목단화, 22>
	쌍긔	셀자가 호외를 들고 보다가 쌍긔 웃더니 <1906 혈의누, 43>
	싱긔	금분이가 고기짓을 쌀낭쌀낭후며 싱긔 웃고 <1908 빈상설, 27>

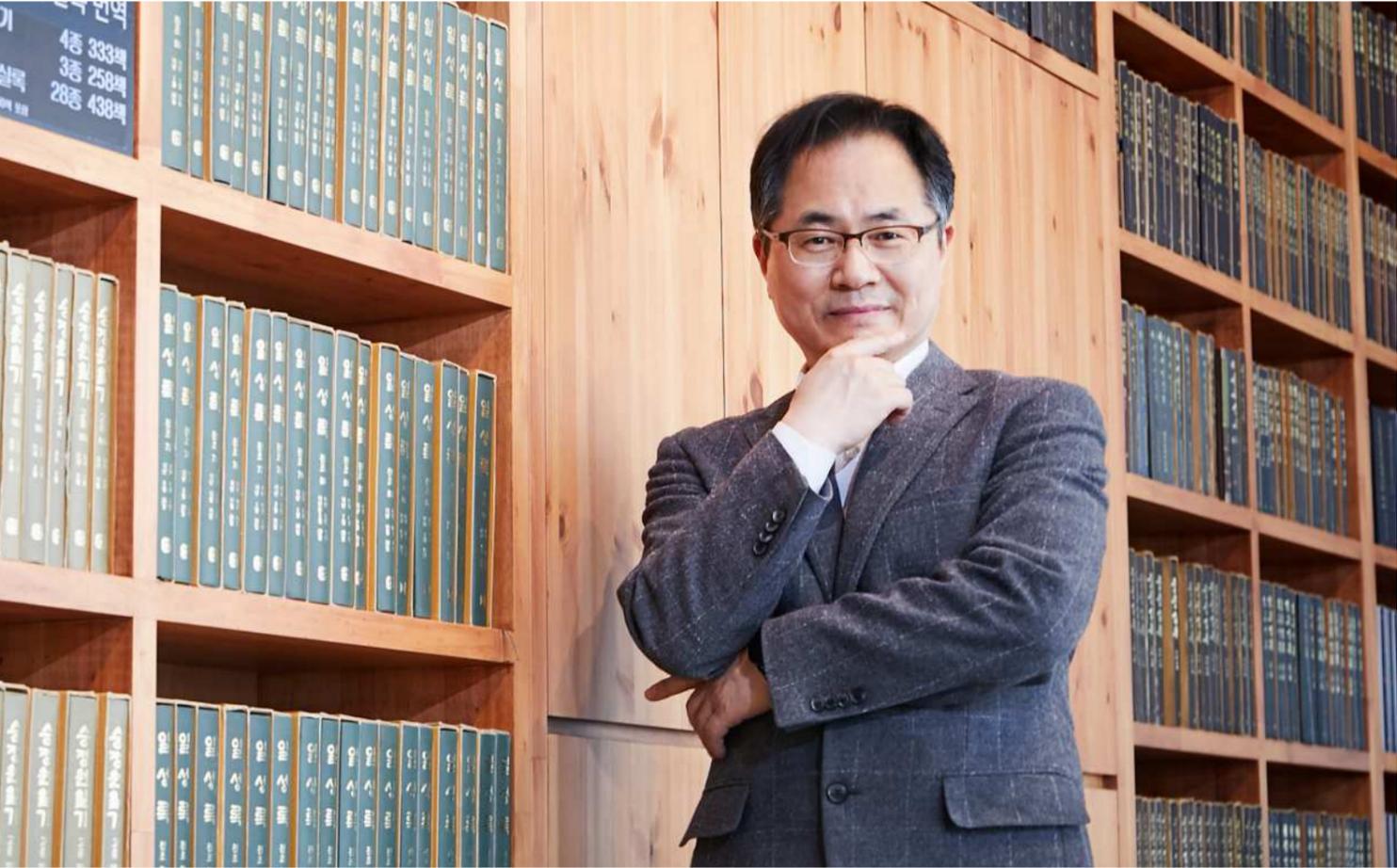
훌쩍훌쩍 우는 모습이나 빙긔 웃는 모습은 쉽게 상상할 수 있지만, *비죽비죽 우는 모습이나 *쪽쪽 우는 모습, *빙글빙글 웃는 모습이나 *덩싹덩싹 웃는 모습은 잘 상상이 되지 않는다.

- *비죽비죽: 언짢거나 비웃거나 울려고 할 때 소리 없이 입을 내밀고 실룩거리는 모양.
- *쪽쪽: 요망스럽게 자꾸 우는 모양.
- *빙글빙글: 입을 슬며시 벌릴 듯 말 듯 하면서 자꾸 소리 없이 부드럽게 웃는 모양.
- *덩싹덩싹: 팔다리를 춤추듯이 자꾸 가볍게 움직이는 모양.

꿀꿀 흐느끼고 싱긔 미소 짓는 인물들의 감정선을 따라가며 이야기에 몰입했을 20세기 초의 독자들을 마음속으로 그려 본다. 실제의 말을 옮겨 놓은 듯한 대사와 생동감 넘치는 장면 묘사는 작품의 호소력을 강화하는 언어적 장치였을 것이다. 그리고 21세기를 살아가는 우리에게 신소설은, 아직 음성으로 기록되지 않았던 20세기 초 한국어의 모습을 생생하게 전해 주는 소중한 기록 유산이다.

인공지능, 한문 고전을 자동으로 번역한다.

한국고전번역원 고전정보센터 백한기 센터장



한국고전번역원은 한국천문연구원과 함께 ‘한문 고전 자동번역 서비스’를 개발하고, 지난 1월부터 시범 운영하고 있다. 한문 고전 자동번역 서비스가 주목받는 것은 고전 번역에 인공지능(AI) 기술을 최초로 도입했기 때문이다. 2017년 개발을 시작한 한문 고전 자동번역 기술은 그간 발전을 거듭해 우리 역사의 최대 기록물인 『승정원일기』와 함께 천문 고전 원문까지 번역하는 수준에 이르렀다. 한문 고전 자동번역 기술은 고전번역 사업의 미래를 어떻게 바꿀까. 고전번역에 인공지능 기술을 도입한 배경과 그 의미를 한국고전번역원 고전정보센터 백한기 센터장에게 들어 본다.

국내 최초의 고전 자동번역 서비스

《선표, 마침표.》

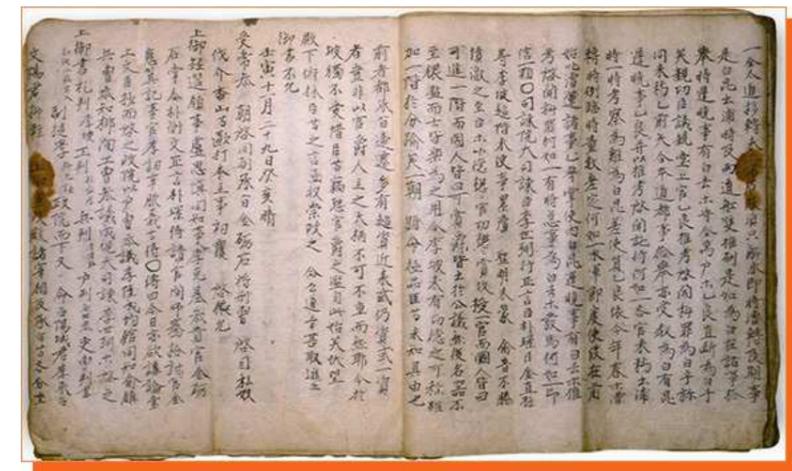
한국고전번역원에서 국내 최초로 고전 자동번역 서비스를 개발했습니다. 첫 번째 고전으로 『승정원일기』를 선택한 이유는 무엇입니까?

백한기

한국고전번역원에서 올해 예정한 고전 번역 목표량이 119책입니다. 그중 절반인 61책이 『승정원일기』죠. 『승정원일기』는 조선 시대 왕명 출납을 관장했던 승정원에서 국정 관련 내용을 매일 일기 형태로 기록한 책입니다. 안타깝게도 임진왜란과 이괄의 난 등을 겪으며 화재로 소실되어 현재는 1623년(인조 1년)부터 1910년(순종 4년)까지 총 277년의 기록만이 남아 있죠. 그런데도 그 분량이 책으로는 3,245권, 글자로는 2억 4,300만 자에 이를 만큼 단일 서종으로는 세계에서 가장 많은 양을 자랑합니다.

1999년 유네스코(UNESCO) 세계기록유산으로 지정된 후, 『승정원일기』 번역에 대한 국민적 관심도 커졌는데요. 국회에서 조기 번역 방안이 없는지를 물었을 정도입니다. 『승정원일기』가 이렇게 주목받는 궁극적인 이유는 문헌으로서의 가치 때문입니다. 왕명의 출납은 물론 각종 행정 사무와 의례(儀禮), 심지어 날씨에 이르는 모든 사실이 자세하게 기록되어 있다 보니, 한국 역사를 가장 방대하게 담은 사고(史庫)의 역할을 하죠.

『승정원일기』 번역은 이미 1994년도부터 진행되어 왔습니다. 물론 지금까지는 사람이 일일이 한문 문장을 번역했죠. 전체 번역서 예상 책 수 2,395책 가운데 현재 713책이 번역되었는데요. 27년 동안 약 30%가 진행된 것입니다. 사람이 번역하는 속도로는 앞으로도 약 30여 년이 지나야 전체를 번역할 수 있는 셈입니다. 이런 이유로 『승정원일기』에 가장 먼저 자동번역 기술을 적용하게 됐습니다. 그 밖에도 인공지능 학습 데이터로 삼을 수 있는 번역 성과물이 가장 많다는 점, 내용이 반복적이고 문장이 정형화되어 있어 자동번역에 적합하다는 점 등도 선정 이유입니다.



▲『승정원일기』(유네스코 한국위원회 소장)

《심표, 마침표.》

고전을 자동으로 번역하는 것은 외국어를 번역하는 것과 어떻게 다릅니까.

백한기

자동번역 모델을 만들기 위해서는 기본적으로 학습 데이터가 필요합니다. 원문과 번역문을 일대일로 일치시켜 병렬 구조의 문장을 만드는데요. 말뭉치(코퍼스)라고도 부릅니다.

포털의 외국어 자동번역 서비스는 우리가 현재 사용하는 언어를 대상으로 합니다. 이와 달리 한문 고전의 언어는 이미 사어(死語)가 되어 버렸죠. 그 때문에 고전 문헌을 자동으로 번역하려면 먼저 시기마다 사용했던 어휘나 문장 형식들을 모두 인공지능에 학습시켜야 합니다. 이 과정에서 실로 어마어마한 양의 말뭉치가 필요합니다. 더욱이 한문은 표의문자입니다. 하나하나의 글자가 낱말의 뜻을 가지고 있고, 한 글자가 다양한 뜻을 갖기도 하죠. 글자 하나에 서른 개가 넘는 뜻이 있는 것도 있습니다. 이런 글자는 어떤 글자와 조합 되느냐에 따라 전혀 다른 뜻으로 해석되죠. 그렇기 때문에 외국어 번역을 할 때보다 많은 양의 말뭉치가 필요합니다.

고전번역원에서 제공하는 자동번역 서비스도 아직은 『승정원일기』를 시범 번역하는 정도이고, 정확도에서도 개선해야 할 부분이 많습니다. 『승정원일기』 대신 한시나 출처 없는 비문의 글을 넣으면 번역이 잘 안 돼요. 인공지능이 『승정원일기』의 데이터로만 학습했기 때문이죠.

사실 한문 번역과 관련해 일반 대중이 더 궁금해할 만한 것은 할아버지 때부터 보관했던 문집이나 족자, 좋은 글귀나 문서의 내용 동일 겁니다. 이런 것까지 번역하는 수준에 이르려면 더 다양한 자료의 원문과 번역문을 모두 수집해 학습 데이터를 확보해야 합니다. 또한 한시, 산문, 일기 등 갈래별로 말뭉치를 모아 각각의 자동번역 모델을 만들어야 하죠. 그래야 더 정확한 번역 결과를 얻을 수 있습니다.



122만 개 말뭉치로 자동번역 모델 완성하다

《심표, 마침표.》

『승정원일기』 번역에 쓰이는 인공지능 기술을 소개해 주세요.

백한기

『승정원일기』 자동번역 서비스에는 국사편찬위원회에서 구축한 『승정원일기』 원문과 한국고전번역원이 번역한 영조·인조·고종 시대의 번역문 약 122만 개 병렬 말뭉치로 사용됐죠. 여기에 주요 포털사이트에서 채택하고 있는 ‘인공신경망 기반 번역(Neural machine translation, NMT)’ 기술을 적용해 번역 모델을 개발했습니다. 인공신경망 기반 번역 기술은 단어와 구문을 쪼개어 번역했던 기존의 ‘통계 기반 번역’(SMT)과 달리, 문장을 통째로 파악해 번역하는 최신 기술입니다. 인공지능이 어순과 문맥의 의미·차이가 지 반영하기 때문에 더 정확한 번역 결과를 얻을 수 있죠.

첫해에는 원문 300자 미만의 병렬 말뭉치를 만들어 인공지능에 학습시켜 보았습니다. 그랬더니 짧은 문장은 잘 번역하는데, 긴 문장은 번역하지 못하더라고요. 그래서 2차연도에는 원문 300자 이상의 병렬 말뭉치를 최대한 많이 만들고, 당시 소개된 기계학습 기술을 다 동원했습니다. 또한 한문 번역자와 자연어 처리 전문가들이 번역 결과를 반복해 평가하면서 한문 특성에 맞도록 조금씩 알고리즘을 변경해 가면서 번역 모델링 작업을 했죠. 그 결과 마지막 3차연도에는 『승정원일기』 자동번역 기술에 3만 개의 천문 고전 말뭉치를 더해 ‘천문 고전 자동번역 모델’까지 개발할 수 있었습니다.

《심표, 마침표.》

한문 고전 자동번역 기술을 개발하기까지 여러 시행착오도 있었을 것 같습니다.

백한기

사실 전문 번역자들도 한문 고전 번역에는 어려움을 느끼고 있고, 번역 결과에 대한 이견도 많습니다. 이런 이유로 한문 고전 자동번역은 불가능한 일이라고 생각한 적도 있었죠. 어쩌면 시도 자체가 무모했는지도 모르겠습니다. 모두가 반신반의하는 가운데 사업에 착수했고, 사업 첫해에 35만 개의 말뭉치를 구축했습니다. 하지만 이것만으로는 양이 절대적으로 부족했기 때문에 당시 개발 업체가 보유하고 있던 45만 개의 중한(中韓) 말뭉치를 활용했습니다. 중국어와 한자의 근간이 같으니 학습 데이터로 유용하지 않을까 하고 생각했던 거죠. 하지만 성능 향상에는 크게 도움이 되지 않았습니다. 당시 추가한 중한 말뭉치는 중한 번역기에 사용되던 것이라, 뉴스 기사 등 현대 문장 위주였죠. 그래서 45만 개라는 상당히 많은 데이터를 활용했음에도 고전에는 맞지 않았던 것입니다.

결국, 중한 말뭉치를 모두 걷어 내고 『승정원일기』의 말뭉치를 최대치로 활용했습니다. 그래도 부족한 학습 데이터를 확보하기 위하여 승정원일기와 한국 고전 종합 디비(DB)에 수록된 어휘들을 추출하여 학습 데이터로 활용하고, 한문이 옛날 중국 고전에서 비롯되었으니 『사서삼경(四書三經)』이 기계학습에 도움이 될 것 같아 말뭉치로 만들어 학습을 시켰습니다. 이렇게 했더니 미세하지만, 성능이 향상되었습니다. 2~3차 사업에서도 상당한 예산을 들여 누적 122만 건의 말뭉치를 확보했고, 총 101회의 모델링 작업을 거쳐 최종 ‘승정원일기 자동번역 모델’을 만들었습니다. 현재 대국민 서비스로 공개한 바로 그 모델이죠.

이뿐만이 아닙니다. 『승정원일기』가 약 300년의 역사를 다루고 있다 보니 시대마다 표현과 문장 구조가 다릅니다. 각 시대에만 쓰인 문장 구조가 있는 것이죠. 그래서 『승정원일기』가 어느 시대에도 번역될 수 있게 하려면 시대별 데이터가 필요했습니다. 이렇게 각 시대의 문장을 모두 학습 데이터로 사용하면서 시대 간극을 해소하기도 했습니다.



▲ 한국고전번역원 전경



▲ 한국고전번역원에서 번역 작업을 하는 모습

《심표, 마침표.》

인공지능 기술로 한문 고전을 번역하는 것이 학계에 어떤 의미가 있을까요?

백한기

한문 고전 번역자를 양성하려면 보통 6~7년이 걸립니다. 한문 고전의 문자적 특수성 때문이죠. 대학에서 한국학 관련 분야를 전공하고도 고전번역교육원에서 연구 과정 3년, 연구 과정 2년, 졸업 후 시역 과정 1~2년을 거쳐야 비로소 한문 고전 번역자로 활동할 수 있습니다. 사정이 이렇다 보니 전문 번역자가 많지 않습니다. 이번 『승정원일기』 번역에 참여한 전문 번역자가 70~80명 정도, 『조선왕조실록』을 비롯한 문집, 특수 고전을 번역하는 사람까지 통틀어도 150~200명 정도입니다. 이들이 연간 150~180권의 책을 번역하는 것입니다.

이처럼 한문 고전 번역 자체가 특수한 분야이다 보니, 인공지능 기술로 자동번역을 한다는 사실을 쉽게 받아들이거나 번역의 결과를 온전히 받아들이지 못하는 사람들도 많습니다. 아직은 『승정원일기』의 자동번역 결과도 완벽한 수준이 아니니 앞으로의 가능성을 믿기가 더 어려울 수도 있죠.

그런데도 고전 번역 자동기술 개발을 위해 노력하는 것은 이것이 향후 한국학 분야의 학문 연구나 관련 콘텐츠 개발에 큰 도움이 될 것이기 때문입니다. 전문 번역자의 도움 없이는 접근하기 어려웠던 우리 역사 기록물을 대중이 더 쉽게 살펴볼 수 있고, 5천 년 우리 역사의 기록과 주옥같은 고전 작품 속에서 새로운 사실과 지식, 문화 정보를 얻을 수 있을 것입니다. 우리 후손은 물론 외국에까지 우리 문화의 우수성을 널리 알리고 손쉽게 입증할 수 있는 때가 올 것이라 기대합니다.

자동번역, 한문 고전 연구와 대중화에 기여할 것

《심표, 마침표.》

말씀하신 것처럼 『승정원일기』는 ‘세계 최대 규모의 역사 기록물’로 유네스코 세계기록유산에도 등재되었습니다. 하지만 그 내용이 대중에게는 생소할 것 같은데요. 『승정원일기』 속 재미있는 이야기 한 가지만 소개해 주세요.

백한기

『승정원일기』에는 조선 시대 정치, 경제, 사회, 문화와 관련한 내용이 풍부하게 실려 있어 조선 시대 연구나 역사 문화 콘텐츠 개발에 활용될 가능성이 매우 높습니다. 그중 잘 알려지지 않은 사건 한 가지를 소개하면 영조 52년(1776년) 2월 4일 기록에 당시 왕세손이었던 정조가 할아버지인 영조에게 생부 사도세자의 죽음에 대한 기록을 『승정원일기』에서 지워 달라고 요청한 일입니다. 당시 왕세손 자격으로 국정을 맡고 있던 정조는 아버지 사도세자의 일로 백관들 대하기가 불편할 뿐 아니라 많은 사람이 『승정원일기』를 읽고 내용을 전할 수 있으니 자신의 마음이 애통하고 곤궁하다며 기록을 삭제해 달라고 상소를 올렸죠. 조심스럽지만 강한 어조의 상소를 읽은 영조는 정조의 요청을 받아들여 기록을 지우도록 지시합니다. 이렇게 해서 사도세자의 죽음과 관련한 내용이 『승정원일기』에서 사라졌습니다.



▲ 정조의 상소문을 발췌해 자동번역기로 번역한 내용

《심표, 마침표.》

자동번역 서비스를 개발하는 과정에서 기억에 남는 일화가 있으신지요.

백한기

자동번역 서비스를 개발하는 과정에서는 일화라고 할 것이 별로 없습니다. (웃음) 오히려 사업을 시작하게 된 동기가 일화에 가까울 것 같네요. 2016년 이세돌과 알파고의 바둑 대결로 전 세계가 시끌시끌했을 당시, 고전번역원에서도 소위 ‘알파고 특강’을 받은 적이 있습니다. 강의를 해 주셨던 분이 서울과학종합대학원의 빅데이터 전문가인 김진호 교수님이었는데요. 강의 끝에 한문 고전도 인공지능 기술을 활용해 번역할 수 있을 것 같다는 말씀을 남기셨죠. 사실 그때까지만 해도 우리 모두 그게 가능할 거라고 생각하지 않았습니다.

이후 동아일보와의 인터뷰에서 김진호 교수님이 이 이야기를 다시 꺼내셨습니다. ‘한문 번역을 꼭 사람이 해야 한다는 것도 고정관념이다. 이미 번역된 내용과 그 원문을 컴퓨터에 학습시키면 고전도 자동으로 번역할 수 있다’는 거였죠. 당시 과학기술정보통신부의 사무관이 그 기사를 읽고 과제 공모에 참여해 볼 것을 고전번역원에 제안했고, 그것을 계기로 ‘인공지능 기반 고전 문헌 자동번역 시스템 구축 사업’이 시작됐습니다. 빅데이터 전문가가 우연히 던진 화두가 오늘날의 성과로까지 이어진 것이죠.

많은 사람이 삼국지를 비롯한 중국 고전과 서양 고전에는 익숙한데 정작 우리 고전은 잘 읽지 않습니다. 번역서가 상대적으로 적다 보니 접근이 어렵고, 번역서가 있더라도 문장이 난해하여 이해하기가 어려운 것이 대부분입니다. 그렇기 때문에 고전번역원에서도 한문 고전을 대중의 눈높이에 맞게 현대화해서 번역하려는 노력을 지속하고 있습니다.

자동번역 서비스를 개발하는 것 또한 이와 같은 노력의 일환입니다. 자동번역 서비스가 보편화하면 사람이 일일이 번역하기까지 걸리는 시간을 절약하고, 누구나 쉽게 궁금한 내용을 찾아볼 수 있죠. 고전번역원의 표어가 ‘우리 가슴에 우리 고전을’입니다. 국민의 가슴에 우리 고전을 심어 주는 것이 우리가 하고자 하는 사업의 궁극적인 방향이고, 이런 노력을 통해 그 목표를 완수해 나갈 수 있을 거라고 생각합니다.

그러니 어렵겠지만 우리 고전에 더 관심을 두고 많이 읽어 보시기를 바랍니다. 그러면 우리 조상들이 남긴 기록 문화가 얼마나 위대한지를 확인하고 자부심도 느낄 수 있을 겁니다.

우리말 풀기



성큼 다가온 봄이 바람결에서부터 느껴지는 요즘.
 춘곤증으로 힘든 오후를 보내고 있지는 않으세요?
 그럴 때는 간단한 스트레칭으로 몸과 마음을 깨워 보세요.
 우리말 풀기에 참여하고 음료 교환권 받는 것도 잊지 마시고요.

문제 1

신소설, 즉 ‘□□□□□’이라는 명칭은 바로 이 고전 소설과의 차별성을 강조하기 위한 것인데, 이 명칭은 후대 붙여진 것이 아니라 당대에 붙여진 것이었다.

단서 100년 전 우리말 풍경

문제 2

요즘에는 ‘□□□□□’가 의사 표현의 한 방식으로 널리 쓰이고 있어요.
 ‘□□□□□’를 달아 특정 의견을 지지하거나 참여 잇기에 동참할 수 있죠.

단서 우리말 다듬기

문제 3

아이가 얼마나 밥을 많이 먹□□ 배탈 날까 걱정이 되었다.
 동생도 놀이가 재미있었□□ 더 이상 엄마를 찾지 않았다.

단서 실전 국어 표기법

문제 4

1986년 국립국어연구원은 ‘□□□□□□□’을 제정했다.
 ‘□□□□□□□’은 혼란스럽게 쓰이던 외래어의 한글 표기를 통일함으로써 원활한 의사소통을 도모하고자 마련했다.

단서 국어원 30년

참여 방법

1. “우리말 풀기 참여하기”를 눌러 주세요.
2. 문제를 풀고, 답을 적어 주세요.
3. 응모자의 이름과 연락처를 입력해 주세요.
4. ‘제출하기’를 누르면 끝!

응모 기간

2021. 4. 6. ~ 2021. 4. 20.

당첨자 발표

개인별로 알리고 <<심표, 마칩표.>> 2021년 5월 호에 공고함.

선물

당첨자 열 분께는 음료 교환권을 드립니다.



2021년 3월 호 우리말 달인 당첨자

하*환(2154) 정*영(7378) 강*근(0091) 이*재(1178)
김*옥(1835) 안*연(7347) 김*기(4039) 정*성(3512)
정*석(2669) 조*영(0277)

우리말 풀기



아침 저녁으로 일교차가 커 감기에 걸리기 쉬운 요즘입니다.
 외출할 때는 따뜻한 외투 한 벌 챙기기.
 《심표, 마침표.》를 읽은 후에는 ‘우리말 풀기’에 참여하고 따뜻한 음료 선물 받기.
 꼭 기억하세요.

문제 1

2019년 개편을 거치면서
 북한어, 방언, 옛말은 “우리말샘”에서 제공하고 있고,
 “□□□□□□□□”에는 현재 약 42만 개의 표제어가 실려 있다.

단서 사전 두 배로 즐기기

문제 2

문화체육관광부와 국립국어원은
 2018년부터 인공지능 학습용 한국어 말뭉치를 구축하여
 ‘□□□□□□’에 공개하고 있다.

단서 국어원 소식

문제 3

조사 ‘□□’는 ‘앞에 오는 말과 같음’을 나타내거나
 ‘따로따로 구별됨’을 나타내는 말로
 명사 뒤에 오며 앞말과 붙여 씁니다.

단서 실전 국어 표기법

문제 4

□□가 제 역할을 하려면 단순함이나 편리함보다
 더 중요하게 앞세울 것이 많다.
 그런 이유로 받아쓰기의 고통은 어디에도 있는 것이다.

단서 읽기 좋은 글, 듣기 좋은 말

문제 5

아카데미상은 영화 산업의 중심인 미국에서도
가장 영예로운 상이라 □□□□만으로도 큰 의미가 있죠.
올해도 한국 배우가 여럿 출연한 영화가
이미 □□□□을 받았다고 하는데요.

단서 우리말 다듬기

문제 6

『□□□□□』가 약 300년의 역사를 다루고 있다 보니
시대마다 표현과 문장 구조가 다릅니다.
각 시대에만 쓰인 문장 구조가 있는 것이죠.

단서 우리말, 그리고 사람

2021년 3월 호 우리말 달인 당첨자

하*환(2154) 정*영(7378) 강*근(0091) 이*재(1178)
김*옥(1835) 안*연(7347) 김*기(4039) 정*성(3512)
정*석(2669) 조*영(0277)

참여 방법

1. “우리말 풀기 참여하기”를 눌러 주세요.
2. 문제를 풀고, 답을 적어 주세요.
3. 응모자의 이름과 연락처를 입력해 주세요.
4. ‘제출하기’를 누르면 끝!

응모 기간

2021. 4. 20. ~ 2021. 5. 4.

당첨자 발표

개인별로 알리고 <<심표, 마침표.>> 2021년 5월 호에 공고함.

선물

당첨자 열 분께는 음료 교환권을 드립니다.

