[38 ~ 41] 다음 글을 읽고 물음에 답하시오.

최근 스마트폰이나 자동차 등에서 인공지능 음성 언어 비서 시스템이 사용되고 있다. 이 시스템이 제대로 작동하기 위해서는 사용자의 음성이 올바르게 인식되어야 한다. 그런데 불분명하게 발음하거나 여러 단어를 쉼 없이 발음하는 경우 시스템이 어떻게 이를 올바른 문장으로 인식할 수 있을까? 이럴 때는 입력된 음성 언어를 문자 언어로 변환한 다음, 통계 데이터를 활용하여 단어나 문장의 오류를 보정하는 자연어 처리 기술이 사용된다. 이러한 기술에는 철자 오류 보정 방식과 띄어쓰기 오류 보정 방식이 있다.

[A] 철자 오류 보정 방식은 교정 사전과 어휘별 통계 데이터를 ㉠ 기반으로 잘못된 문자열*을 올바른 문자열로 바꿔 주는 방식이다. 철자 오류 보정은 '전처리, 오류 문자열 판단, 교정 후보 집합 생성, 최종 교정 문자열 탐색' 과정을 거친다. 먼저 '전처리'는 입력 문장에서 사용자의 발음이 불분명하게 입력되어 시스템에서 처리가 불가능한 문자열을 처리가 가능한 문자열로 바꿔 주는 과정이다. 가령, '실크'가 '싵'으로 인식될 경우, '싵'이라는 음절이 국어에 쓰이지 않으므로 '실크'로 바꿔 준다. 이렇게 전처리가 끝나면 다음 단계인 '오류 문자열 판단' 단계로 넘어간다. 이 단계에서는 입력된 문장을 어절 단위의 문자열로 ㉡ 구분하여, 각 문자열이 교정 사전의 오류 문자열에 존재하는지 여부를 확인한다. 교정 사전이란 오류 문자열과 이를 수정한 교정 문자열이 쌍을 이루어 구축되어 있는 사전이다. 예를 들어 사람들이 자주 틀리는 어휘인 '할려고'의 경우, 교정 사전의 오류 문자열에 '할려고', 이를 수정한 교정 문자열에 '하려고'가 들어가 있다.

처리된 문자열이 교정 사전의 오류 문자열에 존재하지 않을 경우 바로 결과 문장으로 도출되지만, 존재할 경우 '교정 후보 집합 생성' 단계로 넘어간다. 이 단계에서는 오류 문자열과 교정 문자열 모두를 교정 후보로 하는 교정 후보 집합을 ㉢ 생성한다. 예컨대 처리된 문자열이 '할려고'일 경우, '할려고'와 '하려고' 모두를 교정 후보로 하는 교정 후보 집합을 생성한다. 그런 다음 '최종 교정 문자열 탐색' 단계로 넘어간다. 여기서는 철자 오류가 거의 없는 교과서나 신문 기사와 같은 자료에서 어휘들의 사용 빈도를 추출한 어휘별 통계 데이터를 활용하여, 교정 후보 중 사용 빈도가 높은 문자열을 최종 교정 문자열로 선택하여 결과 문장을 도출한다. 만일 통계 데이터에서 '할려고'의 사용 빈도가 1회, '하려고'의 사용 빈도가 100회라면 '하려고'를 최종 교정 문자열로 선택하는 것이다.

띄어쓰기 오류 보정 방식은 잘못된 띄어쓰기를 통계 데이터와 비교하여 올바른 띄어쓰기로 바꿔 주는 방식이다. 이를 위해서는 입력된 문장의 띄어쓰기를 시스템에서 처리할 수 있도록 이진법으로 변환하는 과정이 요구된다. 이 과정에서 음절의 좌나 우, 혹은 음절의 사이에 공백이 있을 때 1, 공백이 없을 때 0으로 표기한다. 가령 '동생이 밥 을 먹었다'라는 문장에서 '밥'은 음절의 좌, 우에 모두 공백이 있으므로 이를 이진법으로 나타내 '1밥1'이 되는데, 이를 편의상 '밥(11)'로 나타낸다. 같은 방법으로 '밥 을'은 두 음절의 좌, 사이, 우에 모두 공백이 있으므로 '밥을(111)'이 되고, '밥 을 먹'은 '밥을먹(1110)'이 된다. 이때 문장의 처음과 끝은 공백이 있는 것으로 처리한다. 이렇게 띄어쓰기를 이진법으로 변환한 다음, 올바르게 띄어쓰기가 구현된 문장에서 ㉣ 추출한 통계 데이터와 비교한다. 그 결과 빈도수가 높은 띄어쓰기 결과에 맞춰 띄어쓰기 오류를 보정한다. 만약 통계 데이터에서 '밥을(111)'의 빈도수가 낮고 '밥을(101)'의 빈도수가 높을 경우, 이에 따라 '밥 을'은 '밥을'로 띄어쓰기가 보정된다.

이러한 방법들은 모두 올바른 단어나 문장에서 추출된 통계 데이터를 기반으로 보정이 이루어진다는 공통점이 있다. 보정의 정확도를 ㉤ 향상시키기 위해서는 통계 데이터의 양을 늘리는 것이 요구되지만, 이 경우 데이터 처리 속도가 감소하게 된다는 단점이 있다. 이러한 문제점을 해결하기 위해 최근 보정의 정확도와 데이터의 처리 속도를 모두 향상시키기 위한 방안이 지속적으로 연구되고 있다.

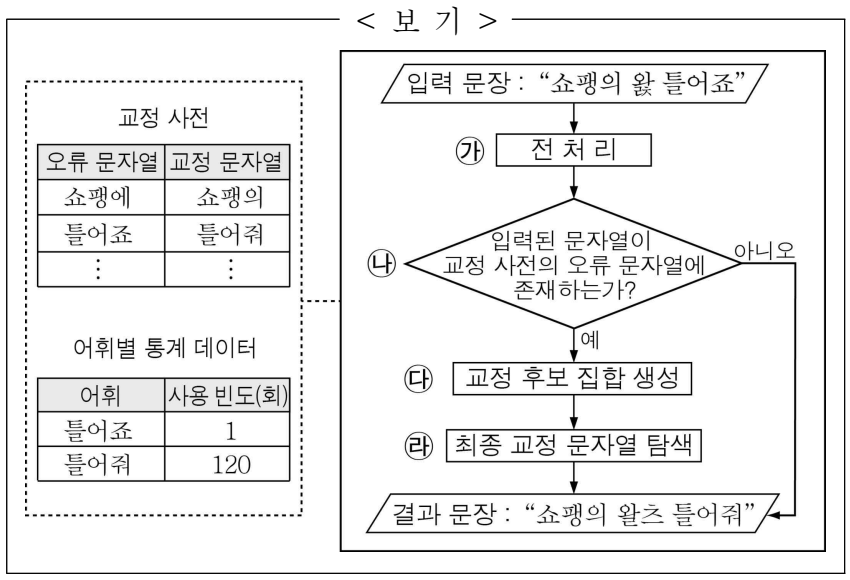
* 문자열 : 데이터로 다루는 일련의 문자.

38. 윗글에서 알 수 있는 내용으로 적절하지 않은 것은?

① 잘못 입력된 문장이 보정되지 않으면 음성 언어 비서 시스템이 제 기능을 발휘하지 못한다.
② 음성 인식 오류를 보정할 때는 사용자의 음성 언어를 문자 언어로 변환하는 과정이 선행된다.
③ 철자 오류 보정 방식은 각 단계마다 입력된 문장을 음절 단위로 구분하여 데이터를 처리한다.
④ 띄어쓰기 오류 보정 방식에서 입력된 문장의 처음과 끝은 공백이 있는 것으로 처리된다.
⑤ 통계 데이터에 포함된 데이터의 양을 늘리면 보정의 정확도는 증가하지만 처리 속도는 감소한다.

39. [A]를 참고로 하여 <보기>의 ㉮ ~ ㉰를 설명한 내용으로 적절하지 않은 것은? [3점]

< 보 기 >



① ㉮ : '왃'를 '왈츠'로 교정하여 처리가 가능한 문자열로 바꿔 준다.
② ㉯ : '쇼팽의'를 교정 사전에서 확인한 결과 오류 문자열에 해당하지 않으므로 결과 문장으로 바로 보낸다.
③ ㉯ : '틀어죠'를 교정 사전에서 확인한 결과 오류 문자열에 해당하므로 '교정 후보 집합 생성' 단계로 보낸다.
④ ㉰ : '틀어죠'가 교정 사전의 오류 문자열에 있으므로 '틀어줘'만을 교정 후보로 하는 교정 후보 집합을 생성한다.
⑤ ㉱ : 어휘별 통계 데이터를 적용하여 사용 빈도가 높은 '틀어줘'를 최종 교정 문자열로 선택한다.

40. 윗글을 바탕으로 할 때, ㄱ ~ ㅁ에서 <보기>의 띄어쓰기 오류 보정이 일어난 이유로 가장 적절한 것은?

< 보 기 >

입력 문장	→	결과 문장
ⓐ 나는 학생 이다		ⓑ 나는 학생이다

(통계 데이터 빈도수 비교 결과)

ㄱ. ⓐ의 '생(01)' > ⓑ의 '생(00)'
ㄴ. ⓑ의 '학생(100)' < ⓐ의 '학생(101)'
ㄷ. ⓐ의 '이다(101)' > ⓑ의 '이다(001)'
ㄹ. ⓑ의 '생이다(0001)' < ⓐ의 '생이다(0101)'
ㅁ. ⓑ의 '학생이(1000)' > ⓐ의 '학생이(1010)'

① ㄱ ② ㄴ ③ ㄷ ④ ㄹ ⑤ ㅁ

41. 문맥에 맞게 ㉠ ~ ㉤을 바꿔 쓴 것으로 적절하지 <u>않은</u> 것은?

① ㉠: 바탕으로
② ㉡: 나누어
③ ㉢: 만든다
④ ㉣: 고친
⑤ ㉤: 높이기

[42 ~ 45] 다음 글을 읽고 물음에 답하시오.

[앞 부분 줄거리] 선녀였던 월영은 호원의 딸로 태어나 최 상서 아들 희성과 정혼하고 월귀탄 귀걸이를 징표로 준다. 모해로 부모를 잃은 월영은 상을 치르려고 소주에 이르는데, 월영의 현숙함을 듣고 소주 자사 위현은 차인을 보내 혼인하려는 뜻을 전한다.

"낭자의 말씀이 그른지라. 이제 낭자의 부모 친척이 없고 천리원정에 최생 소식을 통할 길이 없거늘, 헛되이 신의를 지키고 평생을 그르게 하니 어찌 아깝지 아니하리오. 또한 위 자사는 청춘에 부귀영화 일국에 진동하니 이제 낭자 결혼하여 빛난 가문에 아름다운 부인이 되어 생남생녀하시며 부귀영화 누리다가 백년해로하시고 위로 부모의 제사를 받들고 아래로 평생을 온전케 할 것이니 어찌 즐겁지 아니하리오. 사생을 돌아보지 아니하고 쓸데없는 최생을 따르고져 하시나이까. 낭자는 깊이 생각하소서. 불연즉 도리어 큰 화가 있을지라. 후회하여도 미치지 못하리다."

하거늘 낭자 변색 대로 왈,

"비록 규중에 있어 배운 것은 없으나 인륜대절은 아나니, 어찌 불측한 말로 감히 욕되게 하느뇨? 그대는 자사의 형세를 자세히 알거니와 나도 사대부 여자로 도리가 있거늘, 비례를 행하라 희롱하니 어찌 방자치 않으리오."

즉시 노복을 불러 등을 내치니 차인이 무료하여 돌아와 낭자의 화용월태며 수작하던 말을 자세히 고한대, 자사 듣기를 다하고 차탄 왈,

"이 여자는 짐짓 군자호구(君子好逑)라. 천만금이라도 달라지 못하거니와, 내 만일 이 여자를 구치 못하면 맹세하고 이 세상에 살지 못할지라."

하고 한 꾀를 내어 일봉 서간을 만들고 봉채를 차려 시비를 주며 왈,

"호부에 가서 ㉠ <u>여차여차</u>하라." 하고 보내니라.

각설, 낭자가 차인을 보내고 울울한 마음과 혈혈한 일신을 진정치 못하여 차탄함을 마지아니하더니, 시비 들어와 고하되,

"경성 최 상서 댁 노복이 서간을 드리나이다."

하고 서간과 금함을 드리거늘, 낭자 시비를 명하여 함을 열고 보니 명주 십여 필과 황금 채단이 들었는지라. 낭자 미소하고 시비로 하여금 서간을 보라 하니, 그 서간에 왈.

"경성 최생은 두 번 절하고 호 낭자 좌하에 올리옵나니 슬프다. 세월이 여류하여 벌써 상공의 삼년상을 지낸 지 오랜지라. 전일 언약을 굳게 지키어 지금까지 실가를 정하지 아니함은 이유 없도다. 낭자를 저버리지 아니함이니, 이제 십여 노복과 조그마한 보배를 보내나니, 이것이 소소하나 행장을 차리어 전일 정한 언약을 이룸이 또한 아름답지 아니하리오. 낭자는 빨리 돌아와 고대하옵는 마음을 저버리지 마옵소서. 허다한 말씀을 다 못하나이다." 하였더라.

낭자 듣기를 다하매 위 자사가 보낸 줄 알고 냉소하기를 이윽히 하더니, 시비 등을 불러 왈,

"최생의 서간을 보시고 냉소하시니 어쩐 일이시오니까."

[A] "봉서를 보니 의심이 많도다. 최생이 나를 데려가려 할진대, 천리 원정에 노복만 보내지 아니할 것이요, 또한 서간의 말씀이 심히 허소하니*, 의심이 두 가지요, 최생의 글씨는 사람마다 칭찬하는 바이나 글씨 이같이 무식하니, 의심이 세 가지요, 나의 월귀탄은 보내지 아니하였으니 의심이 네 가지요. 최 상서는 본대 정직한 군자라, 어찌 원로(遠路)에 이렇듯 보배를 보내리오. 의심이 다섯 가지라. 이는 위 자사가 나를 반드시 속이고져 하는 일이라. 어찌 경솔히 발행하리오."

유모와 시비 등이 이 말을 듣고 탄복함을 마지아니하더라.

낭자 즉시 봉서를 담아 그 노복으로 금백 채단을 도로 금함에 넣어 보내니 그 노복 하직하고 가는지라.

각설, 이때 자사 묘한 계고를 내어 보내고 내념에 생각하되,

'내 비밀한 계교는 유식한 남자라도 속을진대, 또한 어린 여자가 어찌 의심할 바가 있으리오.'

하고 기다리더니, 문득 노복이 헛되이 옴을 듣고 대경하여 발을 구르며 문 왈,

"네 어찌 헛되었는가"

노복이 가로되,

"㉡ <u>여차여차</u>하옵기로 봉서와 금함을 도로 올리나이다."

(중략)

"우리 등은 위 자사의 명을 받아 낭자를 모시려 왔사오니 낭자는 바삐 가시면 좋거니와 불연즉 이 비수 아래 놀란 혼백이 될 것이니, 어찌 청춘이 아깝지 아니하리오. 후회하여도 미치지 못하리니 낭자는 길이 생각하소서."

낭자가 정색 대 왈,

"내 비록 여자나 너희 등 비수는 두렵지 아니하나, 어찌 죽기를 저어하리오마는 지금까지 목숨을 보전하기는 이유 없도다. 부모의 유언도 있을뿐더러 후사를 근심함일러니, 이제 너희 등의 핍박을 보니 어찌 소소한 일을 생각하고 잔명을 구차히 살아 무엇에 쓰리오. 또한 내 벌써 죽어 너희 자사의