

조각문을 고려한 음성 인식 문법 설계

윤승, 김상훈, 박준
과학기술연합대학원대학교 컴퓨터 소프트웨어 및 공학 전공
한국전자통신연구원 음성/언어정보연구센터
e-mail : syun@etri.re.kr

An Investigation of Grammar Design to Consider Minor Sentence in Speech Recognition

Seung Yun, Sang-Hun Kim, Jun Park
Computer Software & Engineering, UST
Speech/Language Information Research Center, ETRI

요 약

조각문이란 문장 성분을 온전히 갖추지 못한 문장으로 일반적인 문장과 달리 종결 어미로 문장을 끝맺지 못하는 문장을 말한다. 실험실 환경에서와 달리 실제 음성 인식 환경에서는 이러한 조각문이 비교적 빈번히 나타나므로 연속 음성 인식 시스템의 성능 향상을 위해서는 이러한 조각문에 대한 고려가 필수적이다. 본 연구에서는 음성 인식 문법 기술에 있어서 조각문을 반영한 경우와 그렇지 않은 경우의 커버리지를 비교해 봄으로써 조각문에 대한 고려가 음성 인식 성능 향상에 기여할 수 있음을 알아 보았다.

1. 서론

사람과 사람간의 대화에서는 조각문의 비율이 40% 이상 나타나는 것으로 보고되어 왔다[1]. 그러나 사람과 로봇 또는 사람과 컴퓨터 사이의 대화에서는 일반적으로 화자가 청자로 대우받는 시스템의 능력에 제약이 있음을 가정하고 대화를 진행하기 때문에 조각문의 비율이 사람 간의 대화에서처럼 높게 나타나지는 않는다. 그러나 이러한 경우에도 약 9% 정도의 조각문이 나타나고 있음이 보고된 바 있으며[2] 이는 연속 음성 인식 시스템의 설계에 있어 조각문에 대한 고려가 필요하다는 증거가 될 수 있다.¹ 특히 실용 시스템에서도 음성 인식의 대상이 기존의 고립어, 연결어 인식에서 연속어 인식으로 바뀌고 있는 동향을 고려한다면 조각문을 인식 대상에 포함시켜야 할 필요성은 더욱 커진다 하겠다. 대화체 연속 음성 인식 시스템의 경우에는 더욱 그러하다.

2. 조각문의 유형

본 연구에서는 [2]의 연구에서와 같이 온전한 문장의 형태를 갖추지 못하고 접속어미, 감탄사, 부사어, 체언, 체언+조사 등으로 종결되는 문장을 조각문²이라고 정의한다. 음성 인식 문법에 반영하고자 하는 조각문을 유형별로 간단하게 예를 들어 살펴보면 다음과

같다.

2.1. 체언 유형

체언 유형은 비교적 많이 등장하는 표현으로 이는 사용자가 경제성의 원리에 따라 발화하거나 시스템의 능력을 낮게 가정하고 사용하여 나타나는 유형이다.

- (1) ㄱ. 이전 채널로 변경
- ㄴ. 이 드라마 제목

2.2. 체언+조사 유형

체언+조사 유형은 담화 정보에 의해 유추가 가능한 부분이 생략되는 경우, 또는 질문에 대한 응답이나 부가 질문 등에서 나타난다.

- (2) ㄱ. 오늘 저녁에 방송하는 영화는
- ㄴ. 공중과 중에서

2.3. 접속어미 유형

접속어미 유형은 구어의 특성에 따라 종결어미로 실현되어야 할 것이 접속어미로 나타나거나 복문에 해당하는 것을 두 문장으로 나누어 발화하는 경우에 해당된다.

- (2) ㄱ. 오후 6 시 30 분 되면 OCN 으로 채널 변경해 주고
- ㄴ. 우선 MBC 를 녹화하고

¹ [1]과 [2]의 경우 기본 분석 단위 및 조각문의 정의가 완벽하게 일치하지는 않으므로 두 연구의 결과를 완전히 동등한 위치에서 비교할 수는 없다.

² 조각문은 발화의 수정, 재구성 등과는 구별되는 개념이다.

2.4. 기타 유형

기타 유형으로는 의미를 부가하거나 또는 말하기에 자연스러운 문장을 만들기 위해 쓰인 부사 유형, 그리고 주로 응답에 해당되는 표현인 감탄사 유형을 들 수 있다. 경우에 따라 빈도는 낮지만 관형사 유형이 나타나기도 한다.

- (2) ㄱ. SBS 자세히
- ㄴ. 우선 MBC 를 녹화하고

3. 조각문을 반영한 음성 인식 문법 설계

본 연구에서는 먼저 음성 인식 대상 영역을 TV 제어 및 TV 가이드 영역으로 결정하였다. 그리고 해당 영역에서 발화 가능한 문장을 WOZ 방식으로 자유롭게 100 대화 분량을 수집하였다. 이후 이를 바탕으로 인식기 및 대화 시스템에서 수용 가능한 문장 유형을 정의하고 해당 범주 내에서 450 대화를 추가로 수집하였다. 다음으로 수집된 사용자의 발화, 그리고 TV 및 TV 가이드 영역에서 추출된 지식을 확장이 가능하도록 소규모 온톨로지로 구축하고 이를 바탕으로 사용자의 발화를 E-BNF 형태의 문법으로 기술하였다. 이 때 유사 구문을 추가로 수집하였으며 또한 TV의 기능이나 TV 가이드의 정보를 고려할 때 발화가 가능하지만 수집된 말뭉치에 포함되어 있지 않은 문장은 추가로 포함시켜 문법을 작성하였다. 그리고 이 과정에서 수집된 발화 자체에 조각문이 포함된 경우에는 있는 그대로 반영하였다. 아래에 이러한 과정을 거쳐 기술된 음성 인식 문법의 일부 예를 보였다. <예시 1>의 문법은 ‘MBC 로 돌려’ 류의 문장을 수용 가능하도록 한다.

<예시 1> 음성 인식 문법의 예

```
<change> ::= 돌려 | 틀어...;
<channel> ::= MBC | KBS1 | KBS2 | SBS...;
<adverbial_particle> ::= 로 | 으로;
<sentence> ::= <channel><adverbial_particle><change>;
```

이후 2 절에서 언급한 4 가지 조각문 유형이 수용 가능하도록 문법을 확장하였다. 즉 조각문으로의 변형이 가능한 문들의 경우 생략이 가능한 인식 단위를 건너 뛴 수 있도록 기술하고 일부 조각문의 특성에 따라 부가될 수 있는 인식 단위는 수의적으로 나타날 수 있도록 표현하였다. 다음은 <예시 1>의 문법을 2.1 및 2.2.에 해당하는 조각문 유형 수용이 가능하도록 확장한 예이다.

<예시 2> 조각문을 반영한 음성 인식 문법의 예

```
<change> ::= 돌려 | 틀어...;
<channel> ::= MBC | KBS1 | KBS2 | SBS...;
<adverbial_particle> ::= 로 | 으로;
<sentence> ::= <channel>[<adverbial_particle>] [<change>];
```

확장된 문법을 따르면 기존의 문장 이외에 ‘MBC’, ‘MBC 로’, 등의 조각문도 수용할 수 있다.

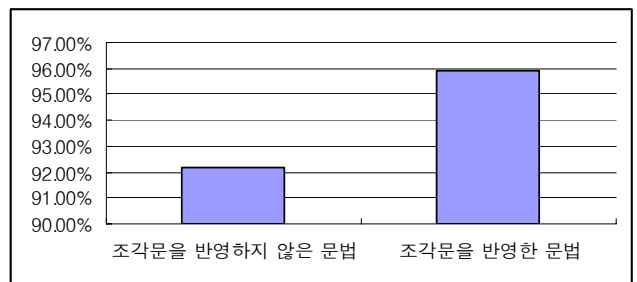
이러한 과정을 거쳐 550 대화에서 나타난 사용자의 발화와 이를 기반으로 한 확장 표현, 그리고 이에 기반한 조각문을 수용할 수 있는 270 여 개의 E-BNF 형태의 문법이 작성되었다.

4. 실험 및 결과

조각문을 반영하여 작성한 음성 인식 문법과 그렇지 않은 음성 인식 문법 간의 커버리지를 비교하기 위해 100 대화 규모의 실험 말뭉치를 앞서와 동일한 환경에서 수집하였다. 다음으로 실험 말뭉치에 나타난 사용자 발화를 음성 인식 시스템의 입력이라 가정하고 서로 다른 두 문법에서 어느 정도의 비율로 수용 가능한지에 대해 조사하여 보았다.

그 결과 조각문을 고려한 문법에서는 92.2%의 발화문이 수용 가능한 것으로 나타났고 조각문을 고려한 문법에서는 95.9%의 발화문이 수용 가능한 것으로 나타났다. 아래에 이를 표로 나타내었다.

<표 1> 조각문 반영 음성인식문법과 아닌 문법 간의 비교



두 문법 간의 커버리지 차이는 약 3.7%로, 차이가 아주 크게 나타나지는 않았는데 이는 조각문을 반영하지 않은 문법에도 기존에 수집된 발화가 대규모였던 탓에 일부 조각문 유형이 이미 자연스럽게 반영되어 있었기 때문으로 판단된다. 만일 소규모 말뭉치를 기반으로 작성된 문법이라면 차이가 더 크게 나타났을 것이다.

5. 결론

지금까지 조각문과 이를 반영한 음성 인식 문법에 대해 알아 보았다. 이 과정을 통하여 강인한 대화체 연속 음성 인식 시스템을 만들기 위해서는 조각문에 대한 고려가 필수적임을 알 수 있었다. 다만 이렇게 작성된 문법은 기존의 문법보다 확장성은 증대되나 이에 따라 복잡도 역시 함께 높아지므로 이를 적절한 비율로 최적화하여 효율적으로 반영할 수 있는 방법 또한 연구되어야 할 것이다.

참고문헌

[1] 서은아, 남길임, 서상규, “구어 말뭉치에 나타난 조각문 유형 연구”, 한글 264, 한글학회, pp123~151, 2004.

[2] 윤승, “음성 인식 성능 향상을 위한 조각문 양상 연구”, 제 23 회 음성통신 및 신호처리 학술대회 논문집 23 권 1 호, pp.233~236, 2006.