

개념기반 대화체 언어번역시스템의 번역부평가

최운천*, 한남용, 김재훈

ETRI, 음성언어연구실

Evaluation of the Translation Part of the Concept-based Spoken Language Translation System

Un-Cheon Choi*, Nam-Yong Han, Jae-Hoon Kim

Spoken Language Processing Section, ETRI

ucchoi@zenith.etri.re.kr

요약

이 논문은 개념기반의 대화체 언어번역시스템의 번역부의 평가에 대해 기술한 것이다. 대상언어는 한국어와 영어로 한국어를 해석하여 영어로 번역하는 시스템이다. 개념기반 시스템은 개념을 기준으로 입력된 문장을 해석하고 그 개념을 이용하여 번역한다. 개념기반 시스템은 개념에 기반을 두기 때문에 자유로운 간투사의 사용, 빈번한 단어 생략 등의 특성을 가지는 대화체 번역에 유리하다. 시스템의 평가는 입력문에 대한 번역문의 결과가 의미적으로 어느 정도 전달되었는지를 평가자의 주관적인 판단에 의해 평가한다. 현재 개발된 시스템은 여행안내 영역(domain)을 대상으로 하고 있다. 개발된 시스템에 대한 평가는 대화체를 전사한 문장과 음성인식의 결과 두 가지의 입력에 대해 하였다.

1. 서론

대화체 음성언어번역 연구는 언어장벽을 해소하고 보다 편리하게 외국인과 대화할 수 있는 시스템 개발을 목표로 세계 여러나라에서 진행 중이다. 전자통신연구소의 음성언어연구실에서는 지난 해의 만남약속 영역에 대한 시스템 개발과 연계하여 올해는 여행안내 영역에서의 대화체 음성언어번역시스템을 개발 중이다.

이 논문에서는 음성언어번역시스템의 한 부

분인 번역부의 평가에 대해 기술한다. 아직까지 대화체 언어번역에 대한 평가기준이 따로 없는 실정이라 가장 보편적인 방법인 평가자의 주관에 따른 평가방법과 그 결과를 기술하고자 한다.

2. 음성언어번역시스템의 번역부 개요

음성언어번역시스템 번역부의 핵심은 개념파서이다. 개념파서는 의미를 바탕으로 발화(utterance)를 해석하고 번역과 생성에 이용된다[1,3,4]. 개념파서는 의미를 바탕으로 한 구조로 입력된 발화 전체를 대상으로 그 발화에서 구체적으로 말하고자 하는 의미만 추출하여 번역하고, 상대국어로 생성하는 방법이다[3,4]. 현재 시스템은 한국어를 해석하여 영어와 일본어로 생성해 준다. 입력으로는 대화를 전사한 문장과 음성인식의 결과를 사용한다. 전사된 문장은 음성언어연구실에서 보유하고 있는 여행안내 영역의 600 대화 중 한국인이 발화한 부분에 해당되는 것을 전사한 것이고, 인식결과는 각 발화에 대한 음성인식의 결과를 의미한다.

개념파서는 시스템에 대해 잘 모르는 초보자라도 쉽게 문법을 작성하고 테스트할 수 있다. 그리고 우리가 실생활에서 사용하는 표현을 그대로 문법파일에 기술할 수 있어 영역 내에서의 확장이 매우 쉽다. 그리고 개념을 기반으로 하기 때문에 다국어 번역에 유리하다. 또한 개념위주로 해석을 하고, 간투사들은 무시할 수 있는 기

능이 있어 대화체 처리에 유리하다.

그러나 개념이 중요하기 때문에 영역이 바뀌면 새로운 개념을 설계해야 하기 때문에 다른 영역으로의 확장은 어렵다. 그리고 한국어와 같이 교착어인 경우는 단어수가 많고 문법이 복잡해진다. 또한 개념위주의 번역이므로 수의 일치나 시제의 일치 등 정밀한 번역은 어렵다. 본 시스템에서는 한국어처리를 위해 기준단어[4]를 이용한 전처리기의 사용으로 교착어이기 때문에 야기되는 문제점을 크게 해소하였다.

3. 평가 방법

평가는 두 가지 종류의 입력에 대해 번역된 영어문장과 비교하여 토큰단위의 의미전달률과 발화단위의 의미전달률을 계산하였다. 발화단위의 의미전달률은 발화전체의 의미가 어느 정도 전달되었는지를 평가하고, 토큰단위의 의미전달률은 입력 발화 중 중심어(key-word)가 어느 정도 번역되었는지를 평가한다. 즉, 한 발화 내에 5개의 토큰이 있다면 5개 중 몇개의 토큰이 정확하게 영어로 번역되는지를 보면 된다.

3.1. 토큰단위의 의미전달률

토큰단위의 의미전달률은 여행안내영역의 전사된 텍스트 1500 발화 중 임의의 300 발화를 대상으로 하였다. 평가는 2명의 문법 개발자와 3명의 음성언어연구실원이 담당하였다. 이 중 3명은 300 발화 전체를 나머지 두명은 300 발화를 나누어서 평가하였다.

평가의 척도는 의미전달의 정도로 하였다. 평가는 입력된 한국어 문장에 대해 번역된 영어 문장이 어느 정도의 의미를 전달했는지를 수치(0~100)로 나타낸다.

의미전달의 핵심은 중심어를 어느 정도 잘 번역했는지를 보는 것이다. 예를 들면 한 발화에 5개의 중심어가 들어 있는데 이 중 4개가 제대로 번역이 되었다면(즉, 의미전달이 되었다면) 80%의 의미를 전달했다고 본다. 번역결과를 0~100으로 표시한다. 즉, 의미전달이 완벽하면 100을 주고, 의미전달이 되지 않을 경우는 0을 준다.

중심어는 각 발화의 의미 전달에 중요한 요소가 되는 의미단위를 말한다. 예를 들면 지명, 인명, 날짜, 시간, 의도를 표현하는 단어들(있다, 없다, 알고 싶다, 있습니까 등)이다. 번호나 날짜는 통째로 한 의미단위가 된다. 즉, 전화번호 일곱 내지 아홉자리가 통째로 한 의미단위가 된다. 위에 언급하지 않은 부분은 주관적으로 판단하여 점수를 매기도록 하였다.

3.2. 발화단위의 의미전달률

발화단위의 의미전달률은 전사된 텍스트 600 발화(그림 1)를 대상으로 하였다. 평가는 문법개발자 2명과 음성언어 연구실원 6명이 담당하였다.

평가의 척도는 아래와 같이 3가지로 구분하였다.

A : 의미전달이 잘 된 경우

B : 의미전달에서 부분적인 오류가 있지만 발화의 취지는 전달된 경우

C : 의미전달이 실패한 경우

위 평가척도의 기준에 따라 평가하고 번역결과가 없을 경우는 무조건 C로 한다. 이 중 A와 B를 번역성공으로 본다.

3.3. 음성인식 결과를 이용한 발화단위의 의미전달률

음성인식 결과를 이용한 발화단위의 의미전달률은 음성인식의 결과 283 문장(그림 2)을 대상으로 하였다. 평가는 문법개발자 2명과 음성언어 연구실원 6명이 담당하였다.

평가는 입력문이 올바른 경우만 위의 평가척도에 따라 평가한다. 입력문이 틀린 경우는 X 표시를 한다.

여기서 입력문이 올바른 경우란, 문장을 읽고 내용을 이해할 수 있는 경우, 문장의 일부에서 하나의 의미단위를 찾을 수 있는 경우, 문장의 일부에서 찾아진 의미단위 내에 간투사가 낀 경우이다. 단 다른 단어가 낀 경우는 제외한다. 그리고 다음과 같은 경우는 입력문이 잘못되었다고 판단한다. 즉, 문장을 읽어봐도 정확히 무엇을 말하는지 애매할 경우, 문장 전체에서 어떤 의미도

못 찾을 경우, 문장의 일부에서 대략의 의미를 찾았는데 그중 중요단어가 오인식되어 다른 의미가 사이에 끼는 경우이다.

평가치	비율	성공률
A	53.6%	76.5%
B	22.9%	
C	23.5%	
계	100%	

4. 결과 및 분석

4.1. 토큰단위의 의미전달률

표 1 토큰단위의 의미전달률

번호	개발자1	개발자2	평가자1	평가자2	평균
총점	23260	23755	24845	23560	
평균	77.53	79.18	82.82	78.53	79.52

전사된 텍스트를 대상으로 한 토큰단위의 의미전달률은 표 1 과 같이 79.52%이다. 문법 개발자 2 명을 포함한 5 명의 평가자들의 평가결과가 큰 차이가 없음을 위 표를 보면 쉽게 알 수 있다. 이 결과로 문법개발자가 생각하는 각 토큰들의 의미와 일반인이 생각하는 의미가 큰 차이가 없음을 알 수 있다. 의미라는 것이 다분히 주관적인데도 비슷한 결과가 나온 것은 한정된 영역인 경우 그 영역에서 나타날 수 있는 개념(의미)은 일반적인 상식으로 어렵지 않게 찾을 수 있다고 말할 수 있다.

결과가 80% 정도인데 이것은 문법을 보완하면 더 나아질 수 있을 것이다. 토큰단위의 의미전달률은 중심어 위주의 평가이므로 발화전체의 의미전달률과는 전혀 다른 의미지만 대화체 번역에서 중심어만의 번역으로도 어느 정도 대화가 가능하다는 가정을 따른다면 의미있는 평가방법이라 할 수 있다.

4.2. 발화단위의 의미전달률

표 2 발화단위의 의미전달률

평가치	1	2	3	4	5	6	7	8
A	141	111	95	99	113	107	89	103
B	26	45	37	51	38	55	58	56
C	33	44	68	50	49	38	53	41
계	200	200	200	200	200	200	200	200

전사된 문장에 대한 발화단위의 의미전달률은 표 2 와 같이 76.5%을 보였다. 여기서는 평가치 A, 와 B 인 경우를 성공으로 계산하였다. 위 표를 보면 각 평가자 별로 평가치가 약간씩 다르게 나타나 있다. 이유는 평가의 척도가 주관적이라 평가자별로 주관적 척도가 달라서 그런 것 같다.

발화단위의 의미전달률과 토큰단위의 의미전달률이 차이를 보이는 이유는 평가기준이 달라서이다. 토큰단위는 전체의 의미보다는 중심어라 불리는 토큰만이 대상이므로 발화단위의 의미번역을보다는 더 높게 나타나는 것이 당연하다.

4.3. 음성인식 결과를 이용한 발화단위의 의미전달률

인식의 결과인 입력문에 대한 평가결과는 75% 가 올바르다는 결과가 나왔다. 현재 인식시스템의 단어인식률이 72%임을 감안하면 인식에서 오인식된 단어가 번역에서는 올바르다고 판단할 수 있기 때문에 3% 정도 높게 나타난 것은 당연한 결과라고 생각된다. 표 3 은 입력이 올바르다고 판단된 부분에 대한 평가결과이다.

표 3 입력문이 올바른 경우의 음성인식결과에 대한 발화단위 의미전달률

	1	2	3	4	계	비율
A	95	98	112	91	396	46.4%
B	58	67	51	53	229	26.8%
C	66	53	41	69	229	26.8%
계	219	218	204	213	854	

위 표에서 보듯이 입력문장이 올바른 경우의 의미전달률은 73.2%이다. 이 결과는 전사된 문장의 의미전달률 76.5%와 별 차이가 없다. 입력문이 잘못된 결과를 제외한 것이기 때문에 전사된 문장과 거의 비슷한 결과를 내었다. 하지만 음성인식결과를 반영한 경우 즉 입력이 잘못되었다고 판단된 발화를 번역실패로 볼 경우 전체 의미전

달률은 55.2%로 크게 떨어졌다. 이유는 인식결과 자체가 잘못되면 그것을 이용해서는 의미를 찾아 내지 못하기 때문이다. 따라서 인식의 성능이 번역의 성능에 크게 영향을 미친다는 것을 알 수 있다.

5. 마무리

이 논문은 다국어 음성언어번역시스템의 일부로 구축된 한영번역시스템의 번역부의 평가에 대해 기술하였다. 평가결과는 대략 76% 정도의 의미전달률을 보이고 있다.

이 논문은 대화체 언어번역의 평가에 대한 한 가지 방법을 제시하였다. 대화체 언어번역에 대한 연구는 최근에 활발히 연구되는 분야라 아직 명확한 평가기준이 없는 실정이다. 이 논문에서 제시한 평가기준이나 의미전달률이라는 용어 자체가 처음 사용된 것들이라 검증이 필요한 부분이다. 그리고 평가 자체가 평가자의 주관에 따라 좌우될 수밖에 없는 실정이다. 보다 객관적인 평가방법과 평가기준을 정하는 것이 시급한 문제라고 생각된다.

감사의 글

이 연구는 정보통신부의 지원에 의해 이루어진 결과물입니다.

참고문헌

- [1] A. Waibel *et al.*, "JANUS-II: Translation of Spontaneous Conversational Speech," *Proc. of ICASST'96*, vol. I, pp. 401-404, 1996
- [2] Manny Rayner *et al.*, "Estimating Performance of Pipelined Spoken Language Translation Systems," *Proc. of ICSLP'94*, Yokohama, Japan, vol III, pp. 1251-1254, 1994.
- [3] 한남용, 최운천, 김영성, "대화체 음성인식 및 기계번역을 위한 언어처리 연구," '95 신호처리합동학술대회 논문집, 1995
- [4] 최운천, 한남용, 김영성, "개념파서를 이용한 대화체 음성언어번역", '95 한국정보처리학회 학술발표논문집, 1995

표 1. 전사된 문장의 평가지

I:아 안녕하세요 저 노병선입니다 그 동안 잘 지 내셨어요	O: Hello. I'm noh byeongsun. hello.
I:예 아주 저도 좋아요 어 이번에 전화한 게 다른 게 아니라 어 이버 우리 세계 잡이 이번에 웅변대회에서 수상을 했길래 올텐도 주변을 관광을 시켜 줄려고 하거든요 그래서 부탁이 있어서 전화 드린 거예요	O:Yes. that's good. telephone number I'm planning to make a trip orlando. I'd like to get some information please.
I:아 잘 됐군요 그러면 차액 요금은 얼마나 됩니까	O:That's good. how much is the difference in the fare?
I:아 잘 됐군요 팩스 번호를 알려 드리죠 어 팩스 번호는 팔이 에 삼 오 육 삼 팔 사 일 공 칠 입니다	O:That's good. I'll send the fax number the fax number eight two three five six three eight four one zero seven.

그림2. 음성인식결과와의 평가지

I:네 내용이 갈려고 하는데요 호텔방을 예약하려고 되나요	O:Yes. I'd like to reserve a room do you have?
I:예 저는 어 정인섭이구요 예 친구백구십육년 오월 사일부터 오월 십일까지 구박을 할려고 합니다	O:Yes. I'm einseop cheong. I'm planning to stay from the fourth of May nineteen ninety six to the tenth of May during nine night.
I:예 가족용으로 하구요 호텔에서 가 공항에서 호텔까지 리무진을 이용할 수 있습니까	O:Yes. I'll take a family room do you have?
I:어 그러면 유월 지불은 어떻게 하면 되죠	O:In June, how can I pay for it?