

다국어 대화체 음성언어번역 시스템을 위한 IF(interchange format)와 IF 태깅

최운천

ETRI 음성언어팀

IF(interchange format) and IF Tagging for the Multilingual Spoken Language Translation System

Un-Cheon Choi

Spoken Language Processing Team, ETRI

ucchoi@etri.re.kr

요약

미국의 카네기 멜론 대학과 일본의 ATR 및 한국의 전자통신연구원 등이 가입한 CSTAR에서는 99년 국제간 음성언어번역 시스템 데모를 위해 IF(interchange format)를 이용하여 데이터를 주고 받기로 합의하였다. IF는 크기는 인터넷을 통해 다른 나라의 음성언어번역 시스템과 연결하여 데이터를 주고 받는데 사용되고, 작게는 음성언어번역 시스템 내의 해석 시스템과 생성 시스템 사이에 데이터를 주고 받는데 사용된다. IF는 중간언어 표현의 한 가지 방법으로 간단하면서도 단순한 표현으로 특정 영역 내에 나타나는 의미를 표현할 수 있도록 정의되었다. 대상으로 하는 영역은 여행 안내로 호텔 예약, 비행기 예약, 여행지 안내 및 예약 등을 포함하고 있다. IF의 가장 큰 특징은 표현방법의 단순화에 있다. 즉, 의미를 가장 잘 나타낼 수 있는 표현을 골라 IF를 정의하여 언어 종속적인 요소를 가능한 배제하였다. IF 태깅은 발화에 대해 적절한 일로 태깅을 수행하는 사람은 IF 태깅 요령에 따라 태깅을 수행하여야 한다. 현재 ETRI에서는 200 대화 이상의 한국어 데이터에 대해 IF 태깅을 완료하였으며 해석 시스템과 생성 시스템 개발을 계속하고 있다.

1. 서론

대화체 문장을 대상으로 하는 다국어 음성언어번역 시스템에서 언어번역 시스템은 IF(interchange format)를 이용하여 정보를 주고 받는다. IF는 크기는 인터넷을 통해 다른 나라의 음성언어번역 시스템과 연결하여 데이터를 주고 받는데 사용되고, 작게는 음성언어번역 시스템 내의 해석기와 생성기 사이에 데이터를 주고 받는데 사용된다. 미국의 카네기 멜론 대학과 일본의 ATR 및 한국의 한국전자통신

연구원 등이 가입한 CSTAR(Consortium for Speech Translation Advanced Research)에서는 99년 국제간 음성언어번역 시스템 데모를 위해 IF를 이용하여 데이터를 주고 받기로 합의하였다.

IF는 6개 언어(일본어, 영어, 독일어, 한국어, 이탈리아어, 프랑스어)를 사용하여 모아진 여행 안내 영역 즉, 호텔 예약, 비행기 예약, 여행지 안내 및 예약 등을 포함하고 있다. 여행 안내 영역에 대한 데이터베이스로부터 6개 언어의 전문가들에 의해, 공통의 의미를 갖는 간단하면서도 명료한 표현 방법을 정의한 것이다. 간단하면서도 명료한 표현 방법은 여행안내 및 호텔예약 영역이라는 특정 영역을 한정하기 때문에 가능했다. 한정된 영역에서 대부분의 발화는 제한된 수의 언어 행위(dialog action or dialog act), 즉 발화자의 의도로 구분할 수 있다. 예를 들면, 호텔 예약이 가능한지 여부를 묻는 행위, 호텔의 가격에 대한 정보를 주는 행위 등이다. 이런 언어 행위들이 모여서 IF를 만든다. IF는 언어 행위에 바탕을 두고 있기 때문에 발화 내의 세세한 시제나 양상 정보 등의 구문정보 등은 표현하지 못한다. 그런 정보들은 언어를 해석하여 IF로 만드는 과정에서 모두 무시된다. IF는 발화자가 수행하고자 하는 의도만을 표현한다.

IF는 CSTAR 회원국들끼리 공통의 중간언어로 정의한 것이다. 각 회원국은 자국의 언어를 음성인식한 후 해석하여 IF를 만드는 해석시스템과, IF로부터 자국의 언어를 생성하여 음성으로 들려 주는 생성 시스템을 책임진다. 그리고 필요하면 자국의 언어에 대한 해석시스템과 생성 시스템을 다른 회원국에 제공할 수 있다. 이런 방식으로 한 회원국은 단지 자국의 언어에 대한 해석시스템과 생성 시스템 개발만으로, 다른 회원국들의 언어로 번역이 가능하게 된다. 그림 1은 6개 언어를 상호 번역해 주는 다국어 번역시스템을 그린 것이다. 각 언어를 IF로 변환하는 것을 해석이라 하고, IF로부터 각 언어를 만들어 내는 것을 생성이라고 한다. 1999년에 있을

다국어 대화체 음성언어번역 시스템을 위한 IF(interchange format)와 태깅

국제간 데모에서는 미리 정해진 두 나라 사이에 인터넷을 통하여 IF를 주고 받음으로써 대화를 하게 된다. 실제 데모에서는 IF는 데모 당사자들에게는 보이지 않는다. 시스템 사용자들은 단지 자국의 음성으로 대화를 하면 된다.

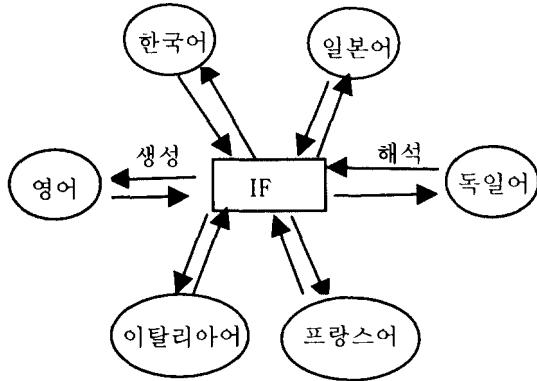


그림 1. 6개 언어를 이용한 다국어 번역시스템

IF를 이용한 언어번역의 간단한 예를 들어보자. 한국어로 "안녕하십니까"는 인사말을 의미하는 "greeting"이라는 IF로 변환된다. 그리고 "greeting"이라는 IF는 영어 생성기를 통해 "Hello"로 변환된다. 다른 예를 들어 보면, "혹시 바로 런던으로 가는 비행기는 없습니까"는 "request-information+availability+flight (destination=london, connection-type=nonstop)"으로 변환된다. 그리고 영어 번역의 경우엔 "Is there any flight direct to London"란 문장을 생성한다.

2. IF란 무엇인가

IF(interchange format)는 CSTAR 회원국들이 공통으로 이용하는, 개념이나 의도를 중시하는 언어 표현의 한 방법으로, 기존의 기계번역 방법 중 중간언어 방식을 이용한 번역에서 사용된 중간언어의 일종이라고 볼 수 있다. IF는 다음과 같은 세 가지의 원칙에 의거하여 디자인되었다. 첫째는 언어행위에 바탕을 둔다. 둘째는 조합에 의해 만들어진다. 즉, 언어행위는 화행(speech act)과 개념(concept), 그리고 속성(argument)의 조합으로 이루어진다. 셋째는 CSTAR 모든 회원국들의 언어에 배치되지 않아야 한다는 것이다. CSTAR IF는 1996년 독일의 Block 박사에게 의해 처음 제안되었다[1]. 그리고 97년 5월 CSTAR 회의에서 받아들여졌으며, 97년 6월 미국 카네기 멜론 대학(CMU)에서 IF 위원회 회의에서 구체적인 표현 방법이 결정되어, 97년 8월 최초의 버전이 공개되었다. 그 후 9월에 2차 버전이 98년 5월에 3차 버전이 공개되어 지금에 이르고 있다.

IF는 DA(dialog acts or dialog actions)와 속성(argument)과 값(value)으로 구성된다. DA는 항상 나타나야 하지만, 속성과 값은 생략이 가능하다.

DA는 화행(speech act)과 개념(concept)들의 조합으로 이루어진다. 화행은 화자가 정보를 받고, 주고, 구하는 등의 행위를 수행하는 것을 말한다. 개념은 제기된 말화 내용의 초점을 나타낸다. 개념은 객실 이용의 가능성 여부와, 여행, 비행기 등에 대해 정보 제공 등의 보다 명확한 의미를 나타내는 것들이다. DA는 화행만으로 이루어 질 수도 있고, 화행에 한 개 이상의 개념들을 조합하여 이루어 질 수도 있다. 속성은 개념의 구체적인 내용을 나타내는 것이다. 속성은 싱글 롬이나 더블 롬, 여행에 참가하는 인원이 몇 명인지 등 화행과 개념에 나타나지 않은 보다 구체적인 것을 나타낸다. 속성은 화행이나 개념에 따라 정해진다. 그리고 값은 각각의 속성에 해당되는 실제 값들로 모두 영어 단어 표기법에 의해 값의 이름을 정의한다.

아래의 예를 보면서 좀 더 구체적으로 알아보자.

(1) 십이일에는 싱글과 더블 룸이 가능합니다.

a:give-information+availability+room (room-type=(single & double), time=md12)

(1)에서 볼 수 있는 것처럼 "십이일에는 싱글과 더블룸이 가능합니다."와 같은 발화에 지정된 IF는 세 단계로 구성된다. 즉, 정보제공을 의미하는 화행('give-information')과 가능성을 의미하는 'availability', 방을 의미하는 'room'이라는 개념, 그리고 개념 'room'과 'availability'에 의해 각기 정해진 속성인 'room-type(방의 종류)'과 'time(시간)'이다. IF는 각기 여행사 직원(agent)과 고객(client)을 나타내는 축약어인 'a:', 'c:'로 시작된다. 이 정보는 화자가 여행사 직원인지 고객인지를 나타내 주는 것으로 발화자가 누구냐에 따라 똑같은 IF로부터 다른 문장을 생성해 낼 수 있도록 한다.

화자에 대한 정보는, 표면상으로 같아 보이지만 의미상으로는 다른 발화로 번역되는 DA를 정확하게 번역하기 위해 삽입되었다. (2)의 예를 보자.

(2) a. 신용카드로 하시겠어요

request-information+payment (method=credit-card)

b. 신용카드로 할 수 있나요

request-information+payment (method=credit-card)

DA 'request-information+payment (method=credit-card)'는, 여행사 직원과 고객간 대화에 대응된다. 즉, 같은 DA에 대해 여행사 직원의 입장에서 보면, (2a)가 되어야 하고, 고객의 입장에서 보면 (2b)가 되어야 자연스러운 번역이 된다. 따라서 화자에 대한 정보를 삽입하는 것은 번역의 질을 유지하는데 필수적이다.

IF는 형식적으로 다음과 같이 나타낸다.

(3) speaker : speech act [+concept*] [argument*]

'*'는 화행을 개념과 구별하게 하거나, 개념들 사이에서의 구별에 사용된다. 괄호('[]')는 생략 가능성을 의미하며, '*'은 안 나타나거나, 하나 이상 나타날 수 있음을 의미한다.

2.1. 기본적인 화행(Speech Act)

IF에서 사용하는 화행에는 어떤 것이 있고, 그 의미가 어떤 것인지를 표 1에 보였다. 표 1에서는 전체 26개의 화행 중 일부만을 제시하였다. 여기에서 제시한 화행들은 2,111개의 발화를 포함하고 있는 56개의 영어 및 한국어대화(dialog)에 기반을 두고 있다. 이탈리아어, 일본어, 독일어 대화들도 참조하였다. 그렇다고 하여, 다음의 목록들이 일상에서 가장 빈번히 일어나는 화행에 대한 예는 아니다. 단지, 56개의 대화에 나타난 화행의 예를 보여 준 것이다.

accept	: 상대방의 요구를 받아들이는 것
acknowledge	: 대화 진행상의 단순 표현
apologize	: 미안함
closing	: 맺는 말
give-information	: 정보제공
greeting	: 처음 만나서 하는 인사
introduce-self	: 자기 소개
introduce-topic	: 대화를 시작하는 표현
negate	: 부정
request-information	: 정보 요구
.....	

표 1. 화행의 종류와 그 의미

2.2. 개념(Concept)

개념은 화행과 결합되며, 제기된 발화의 의미의 초점을 나타낸다. 아래의 두 문장을 보자.

- (3) a. "저는 워싱턴에 여행을 가고 싶은데요"
- b. "저는 제 가족과 함께 여행을 하려고 하는데요."

위 두 문장의 speech act는 'give-information'이다. 그러나 두 문장이 가지고 있는 의미의 초점은 서로 다르다. 앞의 문장은 여행지(워싱턴)가 초점이고, 뒤의 문장은 동행자(제 가족)가 초점이다. 이런 정보를 구분해 주는 것이 개념이다. 현재까지 정의된 개념은 62개이다. 하지만 앞으로 처리해야 할 데이터가 늘어날수록 계속 증가할 것이다.

개념은 화행들과 자유롭게 결합되는 것이 아니라, 화행과 개념간의 허용되는 결합을 나타내는 계층 구조에 따라 결합된다. 이렇게 DA를 만들 때 제한을 두는 이유는 의미 전달에 필요한 최소한의 DA만을 허용하기 위함이다.

2.3. 속성(Argument)

IF의 속성은 주어진 발화의 내용에 세밀한 특성을 부여한다. 다시 말하면, DA에서 구체적으로 명시하지 못한 것들을 속성을 이용하여 명시한다. 아래 (4)의 예를 살펴 보자. 발화 (4a)와 (4b)의 DA는 같은 화행인 'give-information'으로 할당이 되고, 같은 개념인 'temporal' 'arrival'에 의해 DA가 만들어졌다. 그러나 정보의 구체적인 내용을 담고 있는 속성이 서로 다른 까닭에 이 두 발화의 번역은 달라진다. (4a)에서의 속성 'with-whom'은 화자가 혼자 여행하는 것이 아님을 나타낸다. 또한 속성인 'time(may, md5)'과 'time(february, md12)'는 각각 다른 value를 가지고 있어 보다 정확한 번역을 할 수 있다

- (4) a. I and two associated arrive may fifth.
a:give-information+temporal+arrival (who=I, with-whom=(associate, quantity=2), time=(may, md5))
- b. I'll be arriving february twelfth.
a:give-information+temporal+arrival (who=I, time=(february, md12))

속성은 기본적으로 속성-값의 쌍(feature-value pair)으로 이루어지는 구조를 만든다. 속성은 값을 하나만 가질 수도, 여러 개의 값을 수도 있다. 또한 또 다른 속성-값의 쌍을 값으로 가질 수 있다. 그리고 값의 리스트나, 속성-값의 쌍의 리스트도 가질 수 있다. 값의 리스트는 콤마(,)를 사용한다. 그리고 AND(&)나 OR(:)로 연결될 수도 있다. 속성은 아래와 같이 정의할 수 있다.

- 속성은 F = V 형태를 갖는다. 여기서 F는 feature 이고, V는 그것의 value 이다.
- 만약 P와 Q가 속성이면, P,Q (list), P;Q (disjuncts), P&Q (conjuncts) 모두 속성이다.
- 만약 V1과 V2가 값이면, V1,V2(list), V1;V2 (disjuncts), V1&V2 (conjuncts) 모두 값이다.
- 만약 v가 값이면, (v, quantity=n)은 값이다.
- 속성과 그들의 값은 일정한 순서가 있는 것은 아니다.

3. IF의 부가 방법

이 장에서는 실제 IF를 부가하는 방법에 대해 알아본다. 먼저 아래의 예를 보자.

- (1.1 서울의 홀리데이 인입니다)
a:introduce-self (affiliation=holiday_inn_seoul)
- (1.2 저는 김성자입니다)
a:introduce-self (person-name=sungja_kim)
- (1.3 무엇을 도와드릴까요)
a:offer*help
- (2.1 호텔 방을 예약하고 싶습니다)
c:request-action+reservation+features*room

다국어 대화체 음성언어번역 시스템을 위한 IF(interchange format)와 태깅

(1.1)과 (1.2)는 자기가 누구인지를 알리는 것으로 (1.1)은 회사를, (1.2)는 담당자 이름을 나타낸다. (1.3)은 무엇을 도와 줄지를 묻는 표현이다. (2.1)은 'give-information+reservation+ features+ room' 대신, 'request-action+ reservation+ features+room'으로 덧붙여진다. 왜냐하면, 화행 'request-action'은 내가 직접 하는 것이 아니라 상대방에게 해 달라고 요구를 하는 것이고, 화행 'give-information'은 단순히 정보를 알려주는 것이기 때문이다. 그래서 "당신은 더블 룸으로 예약하셨습니다" 같은 발화에 적당할 것이다. 개념 'room'은, 그 발화에 명확한 'features'에 해당되는 표현이 없지만, 꼭 'features'를 앞에 사용해야 한다. 여기서 사용되는 'features'는 대응되는 개념 'price, temporal, location, availability' 등과 같이 발화 내에 그에 해당되는 명확한 표현이 없을 때 사용한다. 앞에서 개념들은 엄격한 길서를 가진다는 것을 언급했듯이, 개념 'features'는 'reservation' 뒤에 오고, 'room' 앞에 와야 한다.

(5.4 어느 것을 원하십니까)

a: request-information+preference

(6.1 싱글이요)

c: accept+features+room (room-type=single)

(7.1 그리고 사흘 밤입니까?)

a: request-verification-give-information+temporal (duration=(time-unit=night, quantity=3))

IF의 함당은 엄격히 발화 자체에만 근거를 두고 있지 않다. 인접한 내용이나, 좀 떨어진 문맥 등도 고려해야 한다. 예를 들어, 발화(6.1)는, '어느 것을 원하느냐'에 응답했기 때문에, 'c: accept+ features+room'으로 태그가 붙여 졌다. 다른 문맥에서라면 이 똑같은 발화는 DA 'give-information+ features+room'이나 'request-action+reservation+ features+room'으로 붙여 졌을 것이다. 발화 (7.1)에 정확한 IF를 할당하기 위해선, 앞의 대화 내용(문맥)이 고려되어야 한다. 직원은 고객이 3일 방 동안 객실을 예약하길 원하는지 확인하고 있다. 확인은 항상 이미 언급된 정보를 가리킨다. 위의 예에서는 직원이 고객에게 예약 정보를 확인하는 것이기 때문에 'request-verification'이 되어야 한다.

현재 정의된 IF의 목록으로 만들 수 없는 경우는 'no-tag'로 주석을 달게 된다. 이러한 것들에는 특별히 복잡한 구조(즉, 진의가 깊이 숨겨진 복잡한 단계의 발화), 잘못된 시작, 영역에서 벗어난 발화(즉, 다른 어떤 새로운 정보를 주지 않는 발화나, 여행 계획과 관련되지 않은 정보)들이 있다.

위와 같은 원칙에 의해 구축되어 공개된 IF 데이터베이스는 110 대화이다. 110 대화에는 영어 40 대화, 한국어 60 대화, 나머지 일본어, 이탈리아어 10 대화이다. 이와는 별도로 한국전자통신연구원에서는 한국어 IF 데이터베이스를 계속 구축하고 있다. 현재 총 220 대화에 대한 IF 데이터베이스 구축이 완료되었다. 그리고 한국어 해석 시스템과 생성 시스템도 함께 개발하고 있다. 표 2는 CSTAR에서 공개

한 데이터베이스와 현재 한국전자통신연구원에서 구축중인 데이터 베이스를 비교한 것이다. 표 2에서 "값의 수"가 많이 차이가 나는 이유는 ETRI IFDB의 경우 많은 수의 고유명사(인명, 지명, 호텔명, 관광 상품명 등)가 포함되어 있기 때문이다.

	CSTAR IFDB	ETRI IFDB
대화수	110	220
발화수	2186	1969
IF의 수	4079	4906
화행의 수	26	26
DA의 수	395	333
개념의 수	58	56
속성의 수	77	80
값의 수	250	803

표 2. CSTAR 공개 DB와 ETRI DB의 비교

4. 마무리

대화체 문장을 대상으로 하는 다국어 음성언어번역 시스템에서 언어번역 시스템은 IF를 이용하여 정보를 주고 받는다. IF는 중간언어 표현의 한 가지 방법으로 간단하면서도 단순한 표현으로 특정 영역 내에 나타나는 의미를 나타낼 수 있도록 정의되었다. IF는 작은 영역에 대해서는 명확한 의미표현이 가능하므로, 인식오류를 내포한 음성인식의 결과를 입력으로 받아들여도 좋은 번역 성능을 낼 수 있다. 그러나 영역이 달라지면, 의미나 의도가 서로 달라지기 때문에 많은 부분을 새로 정의하여야 하는 부담이 있다. 그리고 시제나, 복수 등의 처리에 약하다는 단점도 있다.

IF 태깅은 지금까지 주로 IF를 잘 이해하는 언어학 전공자들의 수작업으로 이루어 지고 있다. IF DB가 구축되면 필수적 수작업 보다는, 자동으로 IF를 태깅할 수 있는 시스템 개발이 요구된다.

감사의 글

이 연구는 정보통신부의 지원에 의해 이루어진 결과물입니다.

참고문헌

[1] Dialogue Act Annotation, CSTAR Consortium, 1997