

실시간 동시통역 시스템 개발을 위한 통역 분절단위 연구

구영은^o, 김지연, 홍정표, 홍문표, 최승권*

성균관대학교, 한국전자통신연구원*

{sarah8835, kite92, jphong2800, skkhmp}@skku.edu, choisk@etri.re.kr

A Study on Segmentation Unit for the Real-time Simultaneous Interpretation System

Youngeun Koo^o, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, Sung-Kwon Choi*

Sungkyunkwan University, ETRI*

요 약

동시통역에서는 번역이 즉각적으로 빠르게 이루어지면서 원천텍스트의 의미가 정확히 전달되는 것이 핵심이다. 따라서 실시간 동시통역 시스템의 개발을 위해서는 번역정확도와 번역속도가 균형적으로 최적을 이루는 지점에서 분절하는 방법론이 필요하다. 이를 위해 본 연구에서는 운율 정보, 문법·통사 규칙, 의미 단위, 담화구조 표지, 분절단위의 길이 등 다양한 언어학적 자질을 제시하였다. 또한 본 논문에서 제안한 방법론을 검증하는 실험을 진행하였으며 그 결과 영한 데이터는 82%, 한영 데이터는 90%의 정확도를 보였다.

주제어: 동시통역, 분절, 분절단위, 언어학적 자질

1. 서론

동시통역은 원천언어(Source language)로 된 표현이 발화됨과 동시에 그것을 목표언어(Target language)로 정확하게 번역하는 것을 목표로 한다. 따라서 인간 동시통역사들은 원천발화자의 발화를 문장이 끝날 때까지 전부 듣고 번역하는 것이 아니라, 적절한 지점에서 나누어 번역하는 전략을 사용한다. 이를 분절(Segmentation)이라 한다.

분절 전략은 실시간 동시통역 시스템에도 적용할 수 있다. 동시통역 시스템이 좋은 성능을 내기 위해서는 입력문을 적절한 지점에서 분절하여 타당한 분절 단위를 기준으로 번역해야 하는데, 이때 분절 단위가 길어질수록 번역의 정확도(Accuracy)는 높아지지만, 번역 속도는 낮아진다. 반대로 분절 단위가 짧아질수록 번역 지연(Latency)은 방지할 수 있지만, 번역의 정확도는 상대적으로 낮아지게 된다. 따라서 동시통역 시스템 개발을 위해서는 번역 지연은 최소화하면서 동시에 높은 번역 정확도를 가질 수 있는 최적의 분절점을 찾아내는 것이 매우 중요하다.

본 논문에서는 최적의 분절점을 찾기 위해 음운·음성, 통사, 의미, 화용·담화 등의 언어학적 접근을 통해 다양한 분절 자질을 상정하였다. 더불어 인지·심리언어학적 관점에서 동시통역 시스템에서의 적절한 분절 길이를 분석해보았다. 번역 정확도와 번역 속도를 모두 고려한 분절 방식은 동시통역 시스템의 성능 향상에 도움이 될 수 있을 것이다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 동시통역 단위에 대한 기존연구들을 통해 시스템에 적용 가능한 동시통역 단위에 대해 알아보겠다. 3장에서는 본 연구에서 제안하는 동시통역 단위를 분절하기 위해 사용

할 수 있는 언어학적인 자질들을 소개한다. 4장에서는 3장에서 제안한 분절 자질을 검증하는 실험을 수행하고 그 결과를 분석한다. 마지막으로 5장에서는 이 글을 정리하고 향후 연구 방향에 대하여 논의하고자 한다.

2. 관련 연구

적절한 동시통역 단위를 찾기 위해 다양한 관점에서 연구가 수행되었다. 먼저, 실시간 동시통역 시스템에 입력되는 문장의 끝을 예측하여 분절을 시도하는 연구가 있다 [1-2]. 특히 Sridhar et al.(2013)[2]의 연구에서는 구두점이 포함된 텍스트를 기계학습 시켜 음성으로 입력되는 문장에 구두점을 예측하여 삽입하는 방식으로 문장의 경계를 추정하였다. 또한, 시스템에 입력되는 문장을 단순히 일정한 길이로 분절해 최적의 분절 단위를 찾으려는 시도도 찾아볼 수 있다 [2-3]. 특히 Rao et al. (2007)[3]의 연구에서는 문장 전체를 번역하는 것보다 분절을 통해 문장을 나누었을 때 번역 품질이 더 좋다는 결과를 얻었다. 이러한 결과를 바탕으로 시스템에 입력되는 문장들을 어떠한 자질도 고려하지 않고 일정한 길이로 분절하여 결과를 비교하였다. 뿐만 아니라, 원천발화와 목표발화 사이의 구문 구조를 비교하여 분절의 경계를 예측하는 방식을 사용한 연구도 있다 [4-5].

인간 통역사의 통역 과정에서 발생하는 다양한 발화의 특징들을 사용하여 분절 단위를 찾아보고자 한 연구들도 진행되었다. Goldman-Eisler(1972) [6]의 연구에서는 원천 발화와 목표 발화의 병렬 전사 자료를 활용하여 분절 단위와 구문 구조 간의 관련성을 찾고자 하였다. Kashioka(2002)[7]에서는 영-일 동시통역에서 목표발화에 등장하는 휴지 정보를 기반으로 동시통역 단위를 알아내고자 하였다. 이를 위해 병렬 전사 자료를 활용하여

목표발화에서 발생하는 긴 휴지의 끝과 대응되는 원천 발화의 품사를 조사하였다.

3. 방법론

본 논문은 실시간 동시통역 시스템에 적용할 분절 방법론을 개발하기 위한 언어학적 연구이다. 이를 위해 언어학의 세부 분야별로 분절과 관련된 자질을 제시하고자 한다. 본 연구는 한국전자통신연구원(Electronics and Telecommunications Research Institute, ETRI)으로부터 제공받은 영한, 한영 동시통역 데이터를 기반으로 한다. 이는 ‘TED’와 ‘세상을 바꾸는 시간, 15분’ 강연데이터로, 각 파일은 약 1~2분 길이에 해당한다. 또한, ETRI 데이터는 강연자가 발화한 원천텍스트, 그것의 번역 결과인 목표텍스트, 인간 동시통역사가 표시한 동시통역 분절점, 원천텍스트가 발화된 시간, 휴지 등의 다양한 정보를 담고 있다. (표 1)은 본 연구에서 제시하는 분절 자질을 정리한 것이다.

구분	분절 자질	
일반언어학	음운/음성	운율 정보, 휴지
	통사	문법·통사 규칙
	의미	의미 단위
	화용/담화	담화 구조 표시
인지언어학	분절단위의 길이	

표 1 동시통역 단위의 분절을 위한 언어학적 자질

3.1. 운율 정보

소리의 높낮이, 크기, 휴지 등의 운율 정보는 적절한 분절 지점을 예측할 수 있게 하는 음성적 정보이다. 그러나 동시통역학에서 운율 정보와 동시통역 단위의 분절 간의 관계를 다룬 연구는 거의 없다. 대신 대화분석론(Conversation analysis)에서 인간이 대화할 때 어떻게 상대방 발화의 ‘전이 적합 지점(Transition Relevance Place, TRP)’을 잘 파악하여 말을 주고받는지 포착하기 위해 운율 정보를 분석한 연구는 많다. 대화의 구조와 내용에 대해 분석하는 대화분석론에서 TRP는 ‘말차례 구성 단위(Turn Construction Unit, TCU)’의 끝 지점을 가리키고, TCU는 대화분석론에서 다루는 발화의 기본 단위를 말한다 [8]. 동시통역이 대화와 유사하다는 점을 고려할 때 [9], 동시통역에서의 분절단위는 대화분석론에서의 TCU와, 분절점은 TRP와 대응된다.

Ishimoto et al.(2011)[10]은 일본어 대화에서 TRP 지점을 예측할 수 있게 하는 운율적 자질을 연구하였고, 그 결과 TRP에 가까워질수록 TCU의 주파수(Pitch)와 크기(Power)가 점차 감소한다고 밝혔다.

실제로 이를 본 연구에서 다루는 영어와 한국어에 적용해본 결과, 비슷한 양상이 나타났다. 본 연구에서는 WaveSurfer를 이용하여 ETRI 영어 데이터 10개, 한국어 데이터 9개의 pitch와 power의 컨투어(Contour)를 추출하고, 컨투어의 모양에 따라 톤별로 0(none),

1(rise), 2(fall), 3(flat) 레이블을 부착하여 분석데이터를 만들었다. 영어는 단어 단위로, 한국어는 어절 단위로 레이블을 부착했다. 그 후 동시통역사가 표시한 분절점 앞의 운율 정보를 분석하였다.

분석 결과에 따르면, 분절점 직전에 pitch와 power가 모두 감소하는 톤이 한국어는 409개의 분절점 중 109개, 영어는 299개의 분절점 중 153개로 가장 많은 수를 차지하였다. 본 연구에서는 이러한 분석 결과와 기존 연구에 근거하여 운율 정보를 동시통역 분절점을 찾는 자질로 보고, 발화의 pitch와 power가 감소하는 지점에서 분절이 이루어진다는 가설을 세웠다.

소리의 유무 즉, 휴지 역시 동시통역 분절점을 유추할 수 있는 음성적 자질이다. 본 연구는 ETRI 데이터에 부착된 휴지 정보를 이용하여 휴지와 분절 간의 관계를 살펴보았다. ETRI 데이터에는 지나치게 짧아서 중요도가 높지 않은 경우를 제외하고는 묵음구간에 대해 2가지의 휴지 태그 중 하나가 부착되었다. ‘SENT_STR’는 묵음 이면서 음성인식 시스템이 언어적 관점에서 문장의 시작으로 예측한 구간이고, ‘SENT_END’는 문장의 끝으로 예측한 구간이다. 이에 따라, 본 연구에서는 시스템이 분석한 ‘SENT_STR’와 ‘SENT_END’를 동시통역 단위의 분절 자질로 설정하였다.

3.2. 통사/문법 규칙

통사 구조는 언어의 특성에 따라 다르게 나타난다. 본 연구의 대상 언어인 영어와 한국어의 경우 절과 구의 경계를 암시하는 통사적 자질이 존재한다. 영어의 경우 접속사가, 한국어의 경우 종결어미와 연결어미가 그러한 역할을 한다.

영어 접속사의 역할 중 하나는 절과 절을 이어주는 것이다. 그 때문에 접속사인 톤 앞에서 분절하게 되면 앞의 문장과 뒤에 새로이 등장하는 문장을 구분 지어줄 수 있다.

한국어의 경우, 주어-목적어-술어(SOV) 구조의 언어이기 때문에 어미를 이용하여 분절하는 것이 가능하다. 종결어미의 경우 문장의 끝에 등장하기 때문에 가장 명확한 분절 자질이다. 연결어미 역시 문장의 의미를 훼손하지 않은 채 분절할 수 있게 하는 자질이다. 특히 연결어미는 문장의 길이가 긴 경우 효과적으로 사용될 수 있는 분절 자질이다. 문장의 길이가 길어지는 것은 대부분 하나의 문장 안에 복수 개의 절이 등장하는 것에서 기인하는데, 이때 연결어미는 복수 개의 절을 연결해주는 역할을 한다. 따라서 종결어미와 연결어미의 경우 해당 톤 뒤가 경계를 나누는 지점이 된다.

3.3. 의미 단위

앞서 설명한 자질들과 더불어 단어가 가지고 있는 의미적 정보를 분절을 유도 또는 제한하는 자질로써 활용이 가능하다. 특히 영어는 동사, 한국어는 용언 및 조사의 의미적인 역할을 고려하여 분절 제한 규칙을 설정하는 것이 가능하다.

먼저, 영어는 동사가 갖는 의미 단위를 파악하기 위해 버브넷(VerbNet) 데이터베이스를 활용할 수 있다. 버브넷은 동사의 통사와 의미가 효과적으로 연결된 정보를 제공하고 있으며 [11], 2019년 9월 기준으로 8,537개의 동사의 통사 및 의미정보를 담고 있다. 본 연구에서는 버브넷에 기술된 동사의 의미가 자동사로만 사용되는 경우 분절을 유도하는 자질로 사용될 수 있다. 반면 동사의 의미가 타동사로만 사용되는 경우 의미적으로 동사 뒤에 논항(Argument)이 존재해야만 문장의 의미가 완전해지기 때문에 분절을 제한하는 자질로써 활용할 가능성이 있다.

한국어의 경우 술어가 갖는 의미가 온전하게 유지될 수 있도록, 연결어미가 등장하더라도 보조용언 앞에서는 분절을 제한하는 규칙을 설정할 수 있다. ‘싶다, ‘내다, ‘간다, ‘하다’ 등은 독자적으로 문장의 서술어가 되지 못하고, 분용언 뒤에 함께 쓰여 분용언의 의미를 도와주는 보조용언에 해당한다. 보조용언이 함께 사용되는 경우에는 보조용언 앞에 연결어미와 같은 자질이 존재하더라도 분절하지 않아야 한다. 격조사 또한 그 의미적인 역할을 고려할 때, 분절을 제한하는 자질로써 활용할 수 있다. 한국어의 SOV 어순은 입력문으로 격조사와 목적격조사를 갖는 토론까지만 들어온 경우, 술어를 중심으로 하는 명제적 구조가 아직 완성되지 않았기 때문에 분절이 제한되어야 하는 지점일 가능성이 크다. 관형격조사 또한 분절을 제한하는 자질로서 역할을 할 수 있다. 관형격조사 ‘의’의 경우, 뒤에 등장하는 명사와 하나의 의미적인 덩어리를 이루기 때문에 분절을 제한해야 하는 지점이다.

3.4. 담화 구조 표지

인간의 발화는 단순한 언어 표현의 나열이 아니다. 인간은 발화할 때 응결성(Coherence)을 갖추어 발화하려고 노력한다. 담화구조 이론(Rhetorical Structure Theory)은 이러한 텍스트의 응결성을 설명하기 위해 만들어진 텍스트 구조 분석 이론이다 [12]. 이를 처음 제기한 Mann & Thompson(1987)[13]은 하나의 텍스트 안에서 여러 부분의 텍스트들이 일정한 관계(Relation)로 연결된 구조를 이룬다고 하였다. 이때 문장들 간의 담화적 성격 변화를 표현하기 위한 언어적 표현이 존재하는데, 이를 담화구조 표지(Rhetorical Structure (transition) Marker, RSM)라고 한다.

담화구조 표지가 등장한다는 것은 그 앞에서 등장한 문장과 그 뒤에 등장하는 문장의 성격이 구분될 수 있음을 의미한다. 따라서 담화구조 표지는 전체 텍스트의 수사 관계를 파악하게 해주는 효과적인 분절 자질이다. 본 연구에서는 영어 180개, 한국어 140개의 담화구조 표지를 자체 수집하였으며, (그림 1)은 그 예시이다.

번호	유형	영어	한국어
1	추가	additionally	게다가, 더구나
2	추가	again	또
3	추가	also	또한
4	추가	and (then)	그리고
5	추가	as well as	~외에도
6	추가	at the same time	동시에, 또한
7	추가	besides	게다가
8	추가	further	게다가, 더 나아가
9	추가	furthermore	게다가
10	추가	in addition (to)	~외에도
11	추가	indeed	실로, 사실상
12	추가	likewise	유사하게
13	추가	moreover	게다가, 더구나
14	추가	next	다음으로

그림 1 구축한 담화구조 표지의 예시

3.5. 분절 단위의 길이

앞서 언급된 자질들이 원천텍스트에서 목표텍스트로 높은 번역 정확도를 이끌어낼 수 있는 자질이라면, 분절 단위의 길이는 적절한 번역 속도를 유지하기 위한 자질이다. 번역 속도를 유지하기 위해서는 적절한 분절 단위 길이를 설정해주는 것이 필요하다. 동시통역학 및 인지·심리언어학적 기존연구의 결과와 ETRI 동시통역 데이터의 분석 결과를 바탕으로, 본 연구에서는 최적의 동시통역 단위의 길이를 4.5초로 설정하였다.

먼저, 4.5초라는 분절 단위의 길이는 EVS(Ear-Voice Span) 시간에 관한 동시통역학 기존연구의 연구결과를 기반으로 한다. EVS란 원천언어로 발화된 연사의 말을 통역사가 들은 이후 이 발화를 통역사가 통역하여 목표언어로 된 발화를 생성하는 데까지 걸리는 시간을 의미한다. 즉 EVS는 동시통역가가 원천 발화를 들은 뒤, 이를 이해하고 해석할 때 필요로 하는 정보를 모두 얻는 데에 걸리는 시간을 의미한다.

Lederer(1978)[14]은 영-프 동시통역에서 발생하는 EVS가 평균적으로 3초에서 6초 사이로 측정되었음을 보여주었다. Ono et al. (2008)[15]의 연구는 일-영, 영-일 동시통역에서의 EVS를 분석하였고, 그 결과 각각 평균 4.532초, 2.446초라고 밝혔다. 또한 Lee(2002)[16]에 따르면, 영-한의 EVS는 평균 3초로 측정되었다. 언급된 연구들을 통해 정확한 번역값을 위한 분절 단위는 3초 이상인 것을 알아낼 수 있었다.

본 연구에서 제안한 분절 단위의 길이는 목표 언어로 통역된 결과를 청각적으로 듣는 청중의 심리적 상태도 고려하여 설정되었다. Sridhar et al.(2013)은 동시통역 과정에서 4~5초 이상의 정적이 발생하면 청자가 심리적으로 답답함을 느낀다는 연구결과를 밝혔다. 즉, 좋은 동시통역이 되기 위해서는 5초 이상의 정적이 발생하지 않는 상태를 유지하며 원천텍스트가 목표텍스트로 통역되어야 한다.

4.5초의 분절단위 길이는 ETRI 데이터로도 확인 가능하였다. 기계가 문장의 끝으로 예측하여 부착하는 ‘SENT_END’ 태그를 영어 파일 19개, 한글 파일 10개에서 조사한 결과, ‘SENT_END’와 다음 ‘SENT_END’ 사이의 원천텍스트 길이는 영어의 경우 평균 4.55초, 한국어의 경우 평균 3.57초였다.

3.6. 제안하는 분절 방법론

지금까지 동시통역 분절과 관련이 있는 다양한 언어학적 자질들을 살펴보았다. 본 논문은 분절과 어떤 언어학적 자질이 관련이 있는지를 밝히고, 그러한 언어학적 자질들을 이용한 분절 방법론의 실효성을 검증하는 예비 연구라고 할 수 있다. 따라서 본 논문에서는 앞서 언급한 다양한 자질들 중에서 분절에 큰 영향을 끼치는 주요 분절자질들을 위주로 동시통역 분절 방법론을 구상해보고¹⁾, 4장에서 이를 평가하고자 한다.

본 연구에서 제안하는 분절 방법론은 다음과 같다. 원천텍스트가 입력문으로 들어오면, 분명하게 분절을 유도하는 자질이 등장하는 지점에서 최우선으로 분절한다. 영어는 담화구조 표지, 한국어는 담화구조 표지와 종결어미로 설정한다.

그러나 만일 그러한 자질이 최적의 분절단위 길이인 4.5초까지도 등장하지 않는다면, 여러 분절자질들을 이용하여 4.5초 근방인 4.5초±1초 내에서 가중치를 반영한 자질값의 합이 가장 높은 지점에서 분절한다. 영어 분절자질은 접속사, 문장 끝('SENT_END'), 휴지('SENT_STR'), 운율이고, 한국어는 연결어미, 문장 끝, 휴지, 운율이다. 그러나 운율은 그 자체로는 강력한 분절자질이 못하고, 문법·통사 규칙이나 의미 단위 등의 다른 자질과 함께 적용했을 때 두 자질이 상호작용하여 좋은 분절자질이 된다. 따라서 문장 끝이나 종결어미처럼 그 자체로 강력한 자질인 경우를 제외하고는 운율 정보를 고려한 경우와 아닌 경우를 각각 자질로 상정하였다.

(그림 2)는 본 연구에서 제안하는 동시통역 분절단위의 분절 방법론을 나타낸 것이다.

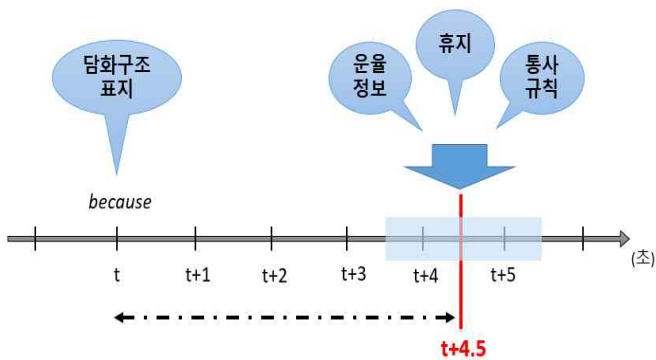


그림 2 제안하는 분절 방법론

4. 실험

본 연구에서 제안하는 동시통역 분절 방법론을 검증하기 위해 앞서 언급한 ETRI 데이터를 이용하여 실험을 진행하였다. 실험에 사용된 코퍼스는 총 영한 10개, 한영 9개 파일이며, 영한은 단어 기준 3,242개의 토큰, 한영은 어절 기준 2,602개의 토큰으로 구성되어 있다. 그리고 수집된 코퍼스에 동시통역 분절자질의 자질값을 부착하여 실험데이터를 구축하였다. 영한 데이터는 단어 단위로 분석하였고, 한영 데이터는 어절 단위로 분절점을 분석하였다.

(표 2)는 실험 결과를 정리한 표이다.

	영한	한영
파일별 평균 분절점 개수	23개	36개
1) 정확도	82%	90%
2) 일치도	69%	70%
분절 단위별 평균 길이	4.8초	3.9초

표 2 제안하는 분절방법론의 성능

본 연구에서 제안하는 분절방법론으로 10개의 영한 데이터를 분절한 결과, 파일당 평균 23개의 분절점이 발생했다. 평가자가 제안하는 방법론으로 추출된 분절점들이 합당한지 평가한 결과 82%의 정확도를 보였다. 그리고 ETRI 데이터에 부착된 인간 통역사의 분절점과 비교해본 결과 69%의 일치도를 보였는데, 나머지 31%의 일치하지 않은 분절점들을 다시 살펴보았을 때 51%는 타당한 분절점으로 평가되었다.

다음으로, 제안하는 분절방법론으로 9개의 한영 데이터를 분절한 결과, 파일당 평균 36개의 분절점이 발생했다. 한영의 경우에도 영한과 동일하게 분절점의 정확도를 평가하였고, 90%의 정확도가 나타났다. 또한 제안하는 방법론의 분절점은 통역사의 분절점과 70%의 일치도를 보였고, 일치하지 않는 30% 중 68%는 타당한 분절점으로 평가되었다.

4.1. 실험 분석

본 장에서는 실제 실험 결과를 통해 본 논문에서 세운 동시통역 시스템 개발을 위한 분절 방법론이 타당한지 분석하고자 한다. 동시통역 시스템의 성능을 평가하는 가장 이상적인 방법은 인간 동시통역사의 동시통역 방법과 비교하는 것이다. 하지만 앞서 언급한 바와 같이, 실제 동시통역사의 동시통역 데이터는 완벽한 연구를 수행하기에는 충분하지 않다. ETRI 데이터에 통역사가 표시한 분절점이 모두 완벽하다고 할 수는 없다. 하지만 본 실험은 동시통역 단위 분절에 관한 언어학적 연구의 예비 실험의 역할이기 때문에, 현 단계에서는 통역사가 표시한 분절점을 정답으로 보고 실험 결과를 분석하였다.

1) 의미 자질은 분절 자질로서의 활용 방식을 추가 검토 중이며, 본 논문의 실험에서는 제외되었다.

	분절단위	시간
통역사	그래서 오늘은 여러분들하고 도대체 이동 통신이 어떻게 돌아가는지 그리고 어떻게 하면 이동 통신을 발전시킬 수 있는지 이런 얘기들을 좀 해보겠습니다.	11.5
제안	그래서 오늘은 여러분들하고	2.5
	도대체 이동 통신이 어떻게 돌아가는지	2.6
	그리고 어떻게 하면 이동 통신을 발전시킬 수 있는지 이런 얘기들을	4.9
	좀 해보겠습니다	1.4

표 3 분절 효과 예시 (1)

먼저, RSM은 텍스트의 통사·의미적 구조를 해치지 않으면서 지나치게 긴 동시통역 단위가 발생하지 않도록 해주는 역할을 하였다. (표 3)에서 알 수 있듯이 통역사는 약 12초 정도의 긴 분절 단위로 분절했지만, RSM을 이용하여 분절된 결과 적절한 길이의 분절 단위로 분절할 수 있었다. 그 뿐만 아니라, ‘그리고’와 ‘어떻게 하면’ 처럼 RSM이 연달아 등장하는 경우, 모든 RSM에서 각각 분절하는 것이 아니라, 분절단위의 끝을 예측할 수 있게 하는 운율 정보와 결합된 RSM에 더 큰 가중치를 두어 불필요한 분절은 제한되었다.

	분절단위	시간
제안	그냥 단순하게 내가 시간을 쓰기만 하면 되는 거죠	2.4

표 4 분절 효과 예시 (2)

운율 정보는 연결어미와 결합할 때에도 분절이 불필요한 지점에서의 분절을 제한해주는 역할을 하였다. (표 4)에서 밑줄 친 ‘하면’은 연결어미이다. 그러나 본 연구에서 제안하는 방법론에는 분절단위의 끝을 유도하는 운율 정보가 없는 경우의 연결어미는 가중치가 낮게 설정되어 있다. 그 결과, 문장의 끝과 더 관련 있는 종결어미 뒤에서 분절하여 바람직한 동시통역 단위로 분절할 수 있었다.

	분절단위	시간
통역사	즉 지난 백년 동안의 기온 상승보다 현재 십 년 동안의 기온 상승이 더욱 더 가파르게 올라가고 있습니다	10.2
제안	즉 지난 백년 동안의 기온 상승보다	4.3
	현재 십 년 동안의 기온 상승이 더욱 더 가파르게 올라가고 있습니다	5.9

표 5 분절 효과 예시 (3)

	분절단위	시간
통역사	저기 안에 있는 모든 내용을 다 이해하면	2.9
	저랑 자리를 바꾸시면 됩니다	1.6
제안	저기 안에 있는 모든 내용을 다 이해하면 저랑 자리를 바꾸시면 됩니다	4.5

표 6 분절 효과 예시 (4)

동시통역 분절단위의 길이 역시 효과적인 분절자질이였다. 4.5초라는 최적의 분절단위 길이를 통해, (표 5)에서는 분절단위가 너무 길어져서 번역 지연이 발생하는 것을 예방했음을 확인할 수 있다. 반대로 (표 6)에서는 최적의 분절단위 길이를 통해, 분절단위가 너무 짧아져서 불필요한 분절이 생기거나 번역정확도를 해칠 가능성을 줄여주었다.

다음의 예시들은 분절 오류가 발생한 경우들이다.

	분절단위	시간
통역사	우리 인간의 본질인 인간의 생명과 삶의 근간을 송두리째 위협하고 있습니다	5.7
제안	우리 인간의 본질인 인간의 생명과 삶의 근간을 송두리째 위협하고	5.2
	있습니다	0.5

표 7 분절 오류 예시 (1)

(표 7)의 경우 연결어미를 포함한 토큰 ‘위협하고’가 분절점을 예측하게 하는 운율 정보인 ‘pitch 하강 - power 하강’과 결합하여 분절을 일으켰다. 하지만 이 위치는 분절이 불필요한 지점이다. 이러한 결과는 주석자가 운율 정보의 레이블을 수동으로 부착하며 생긴 오류 때문일 수 있다. 혹은 ‘pitch 하강 - power 하강’이 분절점을 예측하게 하는 운율 정보라는 가설이 완벽하지 않아서 일 수도 있다. 예를 들어, (표 7)에서처럼 연결어미가 ‘pitch 하강 - power 하강’ 운율 정보를 갖기는 하지만, 이어서 종결어미라는 강한 분절자질이 등장하는 경우 연결어미의 토큰보다는 종결어미의 토큰이 더 큰 가중치를 가져야 한다. 이처럼 추후 연구에서는 분절점을 예측하게 하는 운율 정보에 대한 가설이 보완되어야 할 필요가 있다. 이때 3.3장에서 제안한 의미단위를 이용한 분절 제한 규칙을 분절방법론에 적용하면 (표 7)과 같은 오류는 방지할 수 있을 것으로 본다.

	분절단위	시간
통역사	오늘 제가 말씀드릴 내용은 적용 그리고 협력입니다	3.6
제안	오늘 제가 말씀드릴 내용은 적용	2.6
	그리고 협력입니다	1.0

표 8 분절 오류 예시 (2)

앞서 (표 3)에서는 RSM 자질을 통해 원천텍스트가 올바르게 분절되었음을 살펴보았지만, (표 8)을 보면 RSM의 분절 규칙이 여전히 보완되어야 함을 알 수 있다. ‘그리고’는 대표적인 RSM 중 하나이다. 그러나 (표 8)에서의 ‘그리고’는 명사와 명사를 연결하는 접속조사와 같은 역할을 하고 있다. 이러한 경우 번역정확도를 위해서는 ‘그리고’ 앞에서 분절되지 않는 것이 타당하다. 따라서 RSM 자질 역시 통사·의미적 분절 제한 규칙을 통해 보완되어야 할 것으로 판단된다.

5. 결론

본 논문에서는 실시간 동시통역 시스템 개발을 위한 동시통역 단위의 분절 방법론을 제안하였다. 적절한 동시통역 단위의 분절을 위해 운율 정보, 문법·통사 규칙, 의미 단위, 담화구조 표지, 분절단위의 길이 등 다양한 언어학적 자질을 제시하였다. 또한, 이를 바탕으로 분절 방법론을 구상하여 실험을 진행하였다. 본 논문에서 제안하는 방법론으로 분절된 결과, 영한 데이터의 분절점은 82%, 한영 데이터는 90%의 정확도를 보였다.

그러나 분절 오류를 해결하고 분절 방법론의 정확률을 향상시키기 위해서는 각 자질을 보완해야 한다. 운율 정보의 경우 pitch와 power의 레이블을 주석자가 수동 부착하는 것보다 각각의 수치를 기반으로 자동 부착하여 정확성을 높여야 할 것이다. 또한, 언어 특수적인 문법·통사 규칙을 분석하여 분절 제한 규칙을 보완해야 한다.

뿐만 아니라, 분절단위의 길이인 4.5초를 기준으로 제시된 3.5~5.5초라는 분절범위 역시 그 타당성이 재검토되어야 할 필요가 있다. 시스템의 경우 2~3개의 단어 정도의 추가 입력문 하더라도 번역지연이 크게 증가하지는 않는다. 오히려 2~3개의 단어를 더 들음으로써 번역정확도가 크게 향상할 수 있다면 입력문을 조금 더 들어오는 것이 바람직할 것이다. 따라서, 추후 연구에서는 분절범위 내에서 자질값의 최대를 만족하는 지점에서 무조건적으로 분절하는 것이 아니라, 일정 임계치를 넘는 경우에만 분절하고 그렇지 않은 경우에는 분절을 보류하는 방법론을 검토할 예정이다.

마지막으로, 다양한 도메인의 동시통역 데이터를 충분히 많이 수집하여 본 논문에서 제안하는 방법론을 보완하고, 정확하게 평가할 계획이다.

감사의 글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(R7119-16-1001, 지식증강형 실시간 동시통역 원천기술 개발)

참고문헌

[1] E. Matusov, A. Mauser and H. Ney, Automatic Sentence Segmentation and Punctuation Prediction

for Spoken Language Translation. In International Workshop on Spoken Language Translation (IWSLT), 2006

[2] V. K. R. Sridhar, J. Chen, and S. Bangalore, Corpus Analysis of Simultaneous Interpretation Data for Improving Real Time Speech Translation. In INTERSPEECH, pp.3468-3472, 2013.

[3] S. Rao, I. Lane and T. Schultz, Optimizing Sentence Segmentation for Spoken Language Translation. In 8th Annual Conference of the International Speech Communication Association, 2007.

[4] T. Fujita, G. Neubig, S. Sakti, T. Toda, and S. Nakamura, Simple, Lexicalized Choice of Translation Timing for Simultaneous Speech Translation, In INTERSPEECH, pp.3487-3491, 2013.

[5] H. He, A. Grissom II, J. Morgan, J. Boyd-Graber and H. Daumé III, Syntax-based Rewriting for Simultaneous Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.55-64, 2015.

[6] F. Goldman-Eisler, Segmentation of Input in Simultaneous Translation. Journal of psycholinguistic Research. 1(2), 127-140, 1972.

[7] H. Kashioka, Translation Unit Concerning Timing of Simultaneous Translation, In LREC, 2002.

[8] H. Sacks, E. A. Schegloff, and G. Jefferson, A Simplest Systematics for the Organization of Turn-taking for Conversation, Language, 50, pp.696-735, 1974.

[9] G. V. Chernov, Inference and Anticipation in Simultaneous Interpreting: A Probability Prediction Model, John Benjamins, 2004.

[10] Y. Ishimoto, M. Enomoto, and H. Iida, Projectability of Transition-Relevance Places Using Prosodic Features in Japanese Spontaneous Conversation, In 12th Annual Conference of the International Speech Communication Association, 2011.

[11] M.T. Paziienza, M. Pennacchiotti and F.M. Zanzotto, Mixing WordNet, VerbNet and PropBank for Studying Verb Relations, In LREC, pp. 1372-1377, 2006.

[12] 정여훈, “수사구조이론과 한국어 텍스트 분석의 실제”, 언어사실과 관점, 32, pp.261-288, 2013.

[13] W.C. Mann and S.A. Thompson, Rhetorical Structure Theory: Toward a Functional Theory of Text organization. Text-Interdisciplinary Journal for the Study of Discourse, 8(3), pp.243-281, 1988.

[14] M. Lederer, Simultaneous Interpretation - Units of Meaning and Other Features. In Language interpretation and communication, Springer, pp.323-332, 1978.

- [15] T. Ono, H. Tohyama, and S. Matsubara, Construction and Analysis of Word-level Time-aligned Simultaneous Interpretation Corpus, In LREC, 2008.
- [16] T. H. Lee, Ear Voice Span in English into Korean Simultaneous Interpretation. Translators' Journal, 47(4), pp.596-606, 2002.