

한·중 인터넷 신문 기사 표제 병렬말뭉치 연구 - 기계번역을 위한 시험적 연구

황 은 하
(연세대)

1. 들어가는 말

기계번역(machine translation) 기술은, 의존하는 언어 지식의 유형과 그 활용 방법에 따라 크게 규칙 기반 기법과 말뭉치 기반 기법으로 나누어진다. 후자의 경우 다시 예문 기반 기법, 패턴 기반 기법 등의 지식 기반 기법과 통계 기반 기법으로 나누어 볼 수 있다. 어떤 기법이든 원문 텍스트와 번역문의 텍스트쌍으로 구성된 병렬말뭉치(parallel corpus, translational corpus)의 활용을 우선 고려하지 않을 수 없다. 병렬말뭉치를 통해 직관으로는 발견할 수 없었던 번역의 규칙을 귀납하고 사전에 없는 대응 어휘나 패턴을 추출하며 번역 메모리(translation memory)로 활용하기도 한다. 단, 기사 표제의 번역은 일부 도메인의 경우 목표어 사용권 사회 그리고 언론기관의 담화관행 중심으로 재맥락화되는 경우가 있다(강지혜 36). 원문과 재맥락화된 번역문의 텍스트쌍은 일반적인

의미에서의 병렬말뭉치로 볼 수 없으며, 기계번역을 위한 재활용 가능성이 현저히 떨어진다¹⁾).

본고는 사설과 칼럼을 중심으로 한·중 기사 표제의 소규모 병렬말뭉치를 구성하고 가공하여 재구성 번역 유형이 차지하는 비율을 밝힘으로써 기계번역을 위한 병렬말뭉치의 재활용 가능성 여부를 가리고자 한다. 그 연장선상에서, 재활용이 가능한 원문과 원문에 충실한 번역문 표제쌍의 경우, 한국어 기사 표제가 중국어로 번역될 때 규칙적으로 나타나는 첨가, 생략, 치환, 변조 등 여러 가지 번역 현상들을 추출해 살펴본다. 마지막으로, 앞서 발견한 번역 현상들을 기계번역을 위한 컴퓨터 가독형 규칙으로 형식화할 수 있는지에 대해 약간의 논의를 시도해 본다.

이 연구의 필요성과 중요성은 주로 다음과 같은 세 가지 측면에서 논의될 수 있다.

우선, 신문 기사 표제 번역의 수요가 날로 늘어나고 있으며, 이 과중한 업무를 보조하기 위해 최근에는 기계번역 분야에서 온라인 기사의 자동 번역을 목표로 한 웹번역 기술에 대한 개발이 진행 중이거나 또는 이미 상용화되었다²⁾. 한국의 주요 언론 매체들인 조선일보, 동아일보, 중앙일보, 연합뉴스 등에서 이미 인터넷으로 부분 기사의 중국어 번역문을 제공하고 있으며, 일부 매체의 경우 중국어 뉴스의 매일 조회 수가 수만 건 이상으로 집계되었다³⁾. 따라서 대량 신문 기사의 실시간 번역은 번역 전문가들에게 있어서 중요한 과제 중의 하나이다.

다음으로, 신문 기사에서, 특히 신문 기사 표제는 기사의 핵심 내용에 대한 개요라는 점에서 정확한 번역의 중요성이 두드러진다. 일찍이 1940년대에 표제 위주로 신문을 읽는 독자라는 뜻의 ‘표제 독자(headline reader)’, ‘표제 소비자(a shopper of headlines)’라는 전문용어가 생길 만큼, 신문 기사 표제는 언론 방송계에서도 중요한 연구 과제의 하나이다. 라디오와 TV 뉴스에서 ‘오늘의 헤드라

-
- 1) 병렬말뭉치가 아닌 비교말뭉치(comparable corpus)로 활용되기도 한다.
 - 2) 파란 웹번역 서비스(URL: <http://trans.paran.com>)(2009년 7월 31일부로 서비스가 종료됨.)는 영한, 한영, 중한, 일한 웹번역을, 구글 웹 번역 서비스(URL: <http://translate.google.co.kr>)는 더욱 다양한 언어들 간의 웹번역을 제공하고 있다.
 - 3) 2007년 조선일보 중국어팀 팀장과의 전화 인터뷰에서 조사된 수치이다.

인'이라는 프로그램이 생겨난 것도 신문 기사 표제가 중요한 정보를 담고 있음을 시사하는 한 단면이다. 기사 표제의 중요성은 인터넷을 통한 뉴스 서비스에서 특히 두드러지는바, 독자는 인터넷 뉴스 사이트의 초기 화면에서 기사를 읽을지를 선택하게 된다(김경희 58). 즉 독자가 종이 신문에서는 표제 외에도 부표제, 전문 등에 의존하기도 하지만, 온라인 신문에서의 기사 선택은 협소한 초기 화면 상에서 기사 표제를 먼저 읽은 후 클릭을 통해 기사 내용에 접근할 수 있게 된다. 따라서 온라인 신문 상에서 특정 기사를 이해하고 평가함에 있어 표제는 종이 신문에서보다 더 중요한 기능을 한다고 할 수 있다.

마지막으로, 신문 기사 표제는 텍스트 특유의 개요 기능과 일반 텍스트와는 현저히 다른 어휘와 통사, 담화적 특성 때문에 언어학계의 관심을 받아 왔다. 송경화 외(255-69)는 신문 기사 말뭉치에 대한 계량적인 분석을 통해 형태소, 품사 분포 및 종결형에 있어서 일반 텍스트의 그것과 상당한 차이를 보임을 밝혀냈다. 기존의 한국어 신문 기사 표제에 대한 연구는 주로 언어 규범이라는 잣대를 적용하여 신문 기사 표제에 나타나는 과도한 생략, 축약 등을 문제점으로 인식해 왔으나(박갑수 133-152, 김희진 어법 50-54, 김희진 신문 38-41), 이런 현상이 기사 표제 일반적이며 불가피한 현상으로 진단되면서 최근에는 표제 특유의 문법으로 보는 경향을 보인다(尹世超 1-195). 특히 기계번역의 경우, 입력문을 일일이 기존 문법에 적격한 표제 형식으로 바꾸는 전처리를 거친다는 건 불가능하거나 굉장히 소모적이 일이다. 때문에 컴퓨터가 일반 텍스트와는 다른 한국어 신문 기사 표제를 이해하고 번역하며, 적격의 중국어 기사 표제를 생성할 수 있도록 하기 위해서는 더더욱 신문 기사 표제에 대한 번역학적인 연구가 선행되어야 할 것이다.

본 연구는 2장에서 언어간 기사 표제에 대한 기존의 국내외 연구를 살펴본다. 3장에서는 본 연구를 위한 병렬말뭉치의 구축과 가공 방법에 대해 기술하며, 4장에서는 한·중 신문 기사 표제의 번역과 비번역 비율을 살펴보고, 번역에 있어 첨가, 생략, 치환, 종결형의 변조 등 번역 현상들을 살펴본다. 5장에서는 4장에서 발견된 번역의 규칙이나 지식의 기계번역에의 적용 가능성에 대해 약간의 논의를 시도한다. 마지막으로 6장에서는 연구에 대한 요약과 더불어 남은 과제를 기술한다.

2. 앞선 연구

국내외의 언어간 신문 표제에 대한 앞선 연구는 그 연구 목적과 방법에 따라 주로 다음과 같은 세 가지로 나눌 수 있다. 첫째, 외국어교육을 위한 대조 연구, 둘째는 모국어에서 외국어로의 번역, 또는 외국어에서 모국어로의 번역을 위한 연구, 마지막으로, 기계 번역을 위한 연구이다.

외국어 교육에서는 고급 학습자들의 신문 기사 읽기, 쓰기 교육을 위해 모국어와 외국어 간의 신문 기사 표제에 대한 대조 분석 연구의 필요성이 제기되고 있다. 언어별로 살펴보면, 영어와 한국어(이영옥 150-170), 영어와 중국어(易立梅 264-265, 陳麗昭 1-59), 영어와 페르시아어(파자네 코다반데, 91-127)에 대한 기사 표제의 대조 분석 연구가 대표적이다. 이 연구들은 출발어와 목표어의 기사 표제 각각의 어휘·통사·수사학적인 특징들을 기술하고 그 차이를 밝히는 데 주안점을 두고 있다. 파자네 코다반데(91-127)은 영어와 페르시아어의 기사 표제 병렬말뭉치를 구축하고, 이를 토대로 어휘·통사적 특성에 대해 세밀하고 체계적인 대조 분석을 진행하였는데, 계량학적인 분석의 선례를 열었다는 점이 특히 짚고 넘어갈 만하다.

최근 들어 언어간 신문 표제의 번역학적인 연구에 대한 국내 학계의 관심이 늘고 있다. 김한식(51-68), 조지은(75-91)은 모두 한일 신문 기사 표제의 번역을 다룬 논문으로, 한국과 일본의 대표적인 신문들의 기사 표제의 종결형 사용 경향을 각각 살펴보았다. 이를 바탕으로 김한식은 한일 번역 전략에 대해 고찰하고 있는데, 적지 않은 분량에 대한 기사 표제를 대상으로 계량적인 접근을 했다는 점에서 주목할 만하다(51-68). 조지은은 사설에 한정하여 한일 번역문이 목표어인 일본어 원문 표제와 얼마나 다른지를 고찰함으로써 효율적인 한일 기사 표제의 번역 방법을 제시하려는 시도를 했다(75-91). 영한 신문 기사 표제에 대한 번역 연구로, 김영신(Kim, Young-Shin 145-66), 강지혜(7-43)가 대표적이다. 김영신은 신문 기사 표제가 기사를 압축하고 요약하는 기능 외에도 다양한 수사적 장치를 이용하거나 원문 독자와의 공통 지식에 의존하여 해석을 유도하는 경우, 번역사가 사용하는 번역 전략을 변화가 없는 경우(no change), 첨가(addition)와 세분화(specification), 문체의 변이, 자국화(domestication) 외에 성구와 은유의 번역 전략 등으로 나누어 기술하고 있다(145-66). 강지혜는 『뉴스

위크 한국판』의 북한 관련 번역기사의 제목과 원문기사의 제목이 어떤 차이를 보이는지에 대해 분석하였다(7-43). 『뉴스위크 한국판』의 번역기사는 제목을 번역하기보다는 주로 새롭게 구성된다. 그럼에도 불구하고 저지는 이 역시 일종의 ‘번역’으로 보고 북한 관련 영어 기사가 한국 사회와 언론기관의 담화관행 중심으로 실현되는 재맥락화 현상을 조명하였다. 번역 전략에 있어서 주로 주체화, 명명, 지시의 전환 및 첨가, 서술의 전환 및 첨가, 명시화, 화행의 전환, 직접인용의 첨가 및 전환 등과 같은 특징을 분석해냈다.

온라인 기사 번역에 대한 수요가 늘면서 신문 기사 표제의 기계번역에 대한 연구 결과도 잇달아 발표되었다. 황금하 외는 뉴스 타이틀 번역이라는 특정 도메인에 초점을 맞추어 이의 언어적 특성을 살펴보고, 중한 언어적 유사성에 기반한 뉴스 타이틀 번역을 위한 중한 기계번역시스템에 대하여 설명하였다(350-57). 타케히코 요시미(Takehiko Yoshimi 496-500), 켄지 오노(Kenji Ono 1-8)는 영어-일본어에 대해, 마헤시 신하 외(R.M.K. Sinha 346-53)는 영어-힌디어에 대해, 디간타 사하와 시바지 만디오파드히요(Diganta Saha, Sivaji Bandyopadhyay 125-133)는 예문 기반 기법의 영어-벵갈리어에 대해 각각 기사 표제의 기계번역에 대해 다룬 바 있다.

결론적으로, 언어간 기사 표제의 대조 분석 연구와 번역학적인 연구들은 외국어 교육이나 번역 교육에 주안점을 두고 있고, 연구 대상에 있어서 대부분 소규모 데이터에 한정돼 있으며 언어 공학을 염두에 둔 논의는 이루어진 바가 없다. 한편, 신문 기사 표제의 기계번역 연구, 즉 언어 공학적인 연구들은 직관으로 파악이 가능한 많지 않은 몇 가지 번역의 규칙을 작성하고, 이를 기계번역 엔진에 적용하여 그 전후의 번역의 정확도에 대한 평가를 하고 있다. 즉, 신문 기사 표제의 언어적인 특성이나 번역의 특징에 대한 깊은 논의는 이루어지지 않았거나, 또는 언어학계의 관련 연구 결과를 활용하지 않은 것이 아쉬운 점으로 남는다. 더욱이, 한국어와 중국어의 언어간 신문 기사 표제 연구는 전혀 이루어진 바가 없는 것으로 조사되었다.

3. 한·중 신문 기사 표제의 병렬말뭉치의 구축과 가공

한·중 신문 기사 표제의 병렬말뭉치의 재활용 가능성 연구를 위해 우선, 표제마다 원문과 번역문을 정렬시킨 소규모의 실험용 병렬말뭉치를 구성한다. 다음으로 기계번역에 적용 가능한 대응 정보의 추출을 위해 원문과 번역문에 대해 어휘, 통사, 문체 단위의 대응 또는 비대응 관계에 대한 주석 작업을 수행한다.

3.1 병렬말뭉치의 구축

병렬말뭉치의 구축 대상 텍스트를 선정하기 위해 국내의 한·중 신문 기사 번역 서비스에 대해 조사하였다. 조선일보, 동아일보, 중앙일보 등의 일간지와 통신사 연합뉴스에서 자사 인터넷을 통해 일부 기사들의 중국어 번역문을 서비스하고 있는 것으로 조사되었다. 조선일보, 동아일보, 중앙일보는 기사별로 대응되는 원문 클릭 버튼이 있는 반면, 연합뉴스는 원문과 번역문의 대응쌍 확인이 불가능하거나, 또는 가능하더라도 많은 수작업을 필요로 하였다. 따라서 자료의 자동 수집이 불가능한 연합뉴스는 본 연구 대상에서 제외하고, 자동 수집이 가능한 조선일보, 동아일보, 중앙일보를 대상으로 2009년 1월분 중국어 번역문이 있는 모든 사설과 칼럼⁴⁾ 기사 191쌍을 수집하였고, 그 중에 기사의 표제 대응쌍만 추출하여 하위 병렬말뭉치를 구성하였다. 이로써 한·중 표제어 대응쌍 191개, 한국어 원문 889어절, 중국어 번역문은 2,651자로 구성된 소규모의 병렬말뭉치가 완성되었다. 신문사별 해당 기간의 중국어 번역문이 있는 사설과 칼럼 수는 다음의 표와 같다.

4) 신문 기사는 사건을 객관적으로 기술하는 보도·해설 기사와 사측 또는 필자의 주관적인 의견을 담은 사설·칼럼 기사로 나눌 수 있다. 이와 같은 신문 기사의 두 가지 유형은 담고 있는 내용과 목적뿐만 아니라, 언어 형식에 있어서도 차이를 드러낸다. 본고는 사설·칼럼이 신문 기사로서 담고 있는 정보의 중요성과 언론사별로 매일 한 두 건씩 필수로 번역되어 서비스되는 중요한 기사라는 점을 염두에 두어 특별히 고찰 대상으로 삼았다.

	사설	칼럼	합계
조선일보	24	18 ⁵⁾	209
중앙일보	49	52	101
동아일보	24	24	48
합계	97	261	358

3.2 병렬말뭉치의 가공

병렬말뭉치의 가공은 언어간 기사 표제에 대한 선행 연구를 검토하고, 한·중 기사 표제의 번역 양상을 살펴, 주로 번역 유형, 형태, 통사적 대응 양상을 일괄적인 추출이 가능하도록 태그(tag)의 집합을 결정하여 주석한다. 우선, 번역 양상에 있어서는 번역인지 비번역이 아닌 개작(rewrite)⁶⁾인지의 여부를 구분할 수 있도록 한다. 다음으로, 번역의 경우 번역문을 원문과 비교하여 번역 과정에서 첨가, 생략, 치환⁷⁾이 이루어지는 어휘, 구 등의 언어 단위에 적당한 태그를 부착하기로 한다. 이외에 기사 표제의 번역에서 관찰된 종결형의 변조에 대해서도 원문과 번역문에 대해 각각 태그를 부착한다. 기사 표제 병렬말뭉치의 가공을 위한 태그의 유형과 각각의 설명을 보이면 다음의 표와 같다.

- 5) 해당 기간 조선일보의 중국어판에 게시된 중국어 칼럼은 23개였으나, 그중 5개의 칼럼 기사는 중국의 기고가가 중국어로 작성한 글로, 대응되는 한국어 기사가 존재하지 않았다.
- 6) 기사 표제의 번역은 직역, 의역 또는 개작인지를 물론하고 모두 동일한 기사에 대한 요약이라는 담화적 기능은 동일하며, 번역학계에서도 등가로 보는 경향이 지배적이다. 단, 기계번역이 현 단계에서는 어휘·통사적인 차원의 변환 수준에 머물러 있기 때문에 본 연구에서는 번역과 구분하여 비번역으로 분류한다.
- 7) 번역 전략은 앞선 연구들에서 첨가, 생략, 치환 외에도 여러 가지 번역 전략들이 보고되었으나 본 연구에서는 한·중 신문 기사 표제의 번역에서 가장 자주 관찰이 되는 이 세 가지 번역 현상에 대해서만 다루고자 한다.

분류	전략		태그			설명	
			시작	끝	기타		
비번역					re	개작 rewrite	
번역	첨가		<a>			첨가 addition	
	생략		<o>	</o>		생략 omission	
	치환		<t>	</t>		치환 transposition	
	완전종결	의문형				in	의문형 interrogative sentence
		평서형				d	평서형 declarative sentence
		청유형				p	청유형 propositive sentence
		명령형				im	명령형 imperative sentence
	비완전종결	명사로 끝나는 경우	논항을 취하지 않음			np	일반 명사구 noun phrase
			논항을 취함 ⁸⁾			pn	논항을 취하는 서술성명사 (predicate nominal)로 끝나는 경우
		연결어미로 끝나는 경우				ce	연결어미로 끝나는 경우 connective ending
		동사구				vp	동사구 verb phrase
		조사로 끝나는 경우				au	조사로 끝나는 경우 auxiliary word

표제의 원문 속에 주석할 태그는 일반 텍스트와 혼동되지 않도록 <>로 구분하며 시작과 끝을 구분할 수 있도록 끝은 </>로 달리 표기한다. 각각의 태그와 그것이 뜻하는 바는 다음과 같다.

- 첨가(addition) 태그: '<a>, '는 원문과 비교하여 번역문에 새롭게 첨가되는 어휘·통사 단위에 표기한다.

- 생략(omission) 태그: '<o>, </o>'는 원문에 주석을 하게 되는데 번역문에 번역되지 않은 어휘나 구 단위에 단위에 표기한다.

- 치환(transposition) 태그: '<t>, </t>'는 원문과 번역문에 각각 주석을 하되, 원문에서는 치환되기 이전의 어휘나 표현을, 번역문에서는 치환된 이후의 어휘나 표현에 부착한다.

8) 표제의 종결형에서 명사로 끝나는 경우, 앞선 연구들과 달리 논항을 취하는 서술성명사로 끝나는 경우를 구분하였다. 그 이유에 대해서는 4장에서 자세히 기술한다.

- 개작(rewrite) 태그: ‘r’은 번역문을 다시 한국어로 번역했을 때 어휘 의미적으로 원문의 추출이 전혀 불가능한 경우를 제목을 새로 뽑은 것으로 보아, 다시 쓰기 유형으로 분류하였다.

- 종결형의 변조: 한·영, 한·일 기사 표제 번역에서 관찰된 것과 마찬가지로 한·중 기사 표제 번역에서도 표제의 종결형의 변조가 관찰되었는데 이 정보를 주석하기 위하여 별도의 칼럼을 두 개 두어 번역 전후의 종결형에 대한 정보를 표기하였다.

다음은 태그 집합을 이용하여 가공한 한·중 기사 표제 병렬말뭉치의 샘플이다.

ID	한국어 원문	중국어 번역문	종결형		기타
			SL	TL	
c090131101	<t>북한</t>은 협박 중지하고 대화에 응하라	<t>朝鲜</t>应终止<a>对韩要挟政策<a>积极参与对话	i	d	
j090116102	한·미 FTA 새 국면 <o>… 지혜 모아 최선책 찾아야</o>	韩美FTA<a>面临新局面	c	d	
c090101201	당신은 세계 최고입니까?	您是“世界第一”吗?	q	q	
c090108201	소련을 무너뜨린 진짜 힘	摧毁前苏联的真正力量<a>是什么?	np	q	
c090116202	‘스포츠 경제’에 눈 돌려야	“姚明进军NBA”给我们的启示	c	np	re
c090102101	경제운용, <o>악재탓보다</o>호재살려야	经济运行<a>机制, 应当利用有利条件	c	d	
d090122101	임금 낮춰 일자리 나누기 본격 추진해보자	正式推进“降低薪水、分享工作岗位”的<a>政策	p	d	
c090115101	번지수 잘못 잤은 북한의 미·북 관계정상화 주장	北韩主张美北关系正常化打错了算盘	pn	d	

여기서 id는 모두 10개의 영어 문자와 숫자로 이루어지는데 위치별로 뜻하는 바가 다음과 같다.

번호	1	2	3	4	5	6	7	8	9	10
예	c	0	9	0	1	3	1	1	0	1
설명	신문사 이름의 약자	년		월		일		1-사설 2-칼럼	해당 날짜의 기사 번호	

위의 표가 보이는 바와 같이, ID의 첫 번째 자리는 신문사 이름의 약자이다. 조선일보는 c, 중앙일보는 j, 동아일보는 d로 구분한다. 왼쪽 두 번째에서 일곱 번째 자리는 기사 표제가 생산된 날짜이다. 여덟 번째 자리는 사설인지, 칼럼인지를 구분하는 코드다. 사설은 1, 칼럼은 2이며, 추후 신문 기사 분야를 추가할 경우 확장이 가능하다. 마지막 두 자리는 해당 날짜의 해당 신문사, 해당 분야의 기사 표제의 번호 코드다. 즉 위의 ID c090131101은 조선일보 2009년 1월 31일 사설 중의 첫 기사를 뜻한다. 각각의 표제에 ID를 부여함으로써, 데이터의 관리가 용이하도록 하며 파일명을 같은 ID로 저장한 기사의 본문을 참고해야 할 경우 접근이 수월해진다.

4. 한·중 기사 표제의 번역 연구

기사 표제의 번역은 직역, 의역 또는 개작인지를 물론하고 모두 동일한 기사의 핵심 내용에 대한 요약이라는 점에서 번역학적으로 등가로 보는 경향이 지배적이다. 단, 기계번역 기술이 어휘·통사적인 차원의 변환 수준에 머물러 있는 점으로 미루어 보아 어휘·통사적 대응 정보의 추출이 불가능한 개작의 경우와 어휘·통사적 대응 관계 추출이 가능한 경우를 구분하여 각각의 비율이 어떻게 되는지를 따져 볼 필요가 있다.

‘직역’과 ‘의역’, ‘개작’의 개념 정의에 대한 의견이 다양하겠으나, 본고에서는 기계번역을 위한 어휘·통사적 대응 정보의 추출을 감안하여, 원문과 번역문의 어휘적 대응 양상을 기준으로 그 개념을 구분하였다. 따라서 ‘직역’은 축자적인 번역을, ‘의역’은 원문에 대체로 충실한 번역문이나, 첨가, 생략, 치환, 종결형의 변조 등 여러 가지 번역 전략이 구사된 번역을 지칭하기로 한다. 아래에

각각의 예를 보이면 다음과 같다.

- (1) j090112102 국회 개혁, 제도와 의식을 함께 바꿔야
国会改革需要改变制度和思想
(국회 개혁은 제도와 사상을 함께 바꿔야 한다)⁹⁾
- (2) j090112201 미국이나, 중국이나 이분법 벗어나야
韩国应从美国和中国的两分法中摆脱出来
(한국은 미국과 중국의 이분법에서 벗어나야 한다)
- (3) j090114101 법정 서는 부끄러운 교육감 그만 나와야
应该首先找回因教育委员会委员长而丧失的信任
(우선 교육감 때문에 잃은 신뢰를 되찾아야 한다)
- (4) c090123101 2008년 4분기 GDP 3분기 비해 5.6% 감소
韩经济危机时刻 迫需政府发挥作用
(한국 경제 위기 시각 정부가 작용을 발휘해야 한다)

(1)은 원문의 내용과 언어 형식상 모두 절대적으로 대응되는 번역 유형으로, 본고에서 말하는 ‘직역’의 예이다. (2)는 번역문이 원문과 비교하여 내용상 대체적으로 충실하나 밑줄 그은 부분, 즉 행위의 주체인 ‘한국(韓國)’이 첨가되었기에 본고의 분류에 따라 ‘의역’으로 분류한다. (3), (4)는 번역문이 원문과 비교하여 내용과 형식상 모두 거의 대응되지 않거나 아예 다른 것으로 판단되어 ‘개작’으로 분류한다. 또, 종결형의 변조에 대해서는 직역과 의역에 대해서만 고찰할 예정인데 직역의 예인 (1)과 의역의 예 (2)가 좋은 예들이다. 두 표제 모두 당위성을 뜻하는 어미 ‘-어야’로 일반적인 문장으로서 완전종결하지 않은 양상을 보이는 반면, 중국어 번역문은 모두 완전한 종결 형태로 문장의 형식을 갖

9) 기사 표제의 예는, 첫줄에 ID와 원문을, 두 번째 줄에 번역문을, 세 번째 줄에는 ()안에 번역문을 다시 한국어로 옮긴 2차 번역문을 각각 싣는다. 2차 번역문을 싣는 이유는 번역문이 원문에 비해 어떻게 달라졌는지를 보이기 위함이다.

추고 있다.

번역 유형별 비율 산출을 위해 3장에서 구축하고 가공한 병렬말뭉치를 토대로 번역의 양상을 유형화하고 각각의 비율을 표로 보이면 다음과 같다. 그 결과가 신문사별 번역 정책이나 지침에 기인한 것인지를 알아보기 위해 신문사별로 나누어 살펴본다.

분류1	분류2	번역 전략	조선일보	동아일보	중앙일보	합계
번역	직역	축자 번역	9x	25(x)	37x	7136.702128
	의역	첨가 생략 치환	17(40.47619)	20(40.47619)	4740.47619	22645.744681
		변조(종결형)				
비번역	개작		1645.744681)	345.744681	1445.744681	3317.553191
합계			2100	00	2525	27 ¹⁰⁾ (%)

표에서 보이는 바와 같이, 세 신문사는 의역 유형이 차지하는 비율이 모두 40%대를 차지하는 것 외에는 상당한 차이를 보인다. 특히, 직역과 개작에서 큰 차이를 보이는데, 축자적 번역의 비율이 동아일보가 가장 높고, 중앙일보가 그 뒤를 따르며 조선일보가 가장 낮다. 반대로, 개작의 비율은 동아일보가 가장 낮고, 다음으로 중앙일보, 조선일보 순이다. 신문사별로 번역의 정책이나 지침에 일정한 차이가 있다고 볼 수 있는 대목이다. 아울러, 세 신문사를 통틀어 보면 의역이 가장 많고 직역의 비율이 그 다음을 이으며 개작의 비율이 제일 낮다.

개작의 17.6%를 제외한 82.4%에 달하는 한·중 기사 표제쌍은 기계번역을 위한 병렬말뭉치로써의 활용이 가능할 것으로 판단된다. 직역, 즉 축자 번역의 경우 전체의 36.7%에 달하며, 원문과 번역문의 정렬쌍에서 어휘·통사적 대응 정보를 추출할 수 있다. 그렇다면 세 가지 번역 유형 중에 가장 높은 비율 45.7%(86개)를 차지하는 의역의 경우, 원문과 번역문에서 어떤 언어적 내용이 나 형식들이 대응을 이루지 않는 것일까. 계량적으로 살펴봤을 때 의역 표제 86

10) 수집된 기사 표제의 원문과 번역문은 191쌍이나, 세 개의 오역된 예는 집계에서 제외하였다.

개 중에 나타난 첨가, 생략, 치환의 빈도는 다음과 같다.

	표제	빈도
첨가	40	52
생략	9	9
치환	42	49

첨가 현상은 표제 40개에 걸쳐 빈도가 52회로 집계되었다. 다음으로, 치환 현상이 42개의 표제에서 49회 집계되어 뒤를 이었으며, 생략되는 경우는 9개의 표제에 걸쳐 9회 나타나 가장 낮은 빈도를 보였다. 아래에 첨가, 생략, 치환이 각각 어떤 언어 단위에 있어서 이루어졌는지를 살펴보고, 종결형의 변조 양상에 대해서도 자세히 다루어 보고자 한다.

4.1. 첨가

앞에서, 한중 기사 표제의 번역의 여러 번역 전략에서 첨가 전략이 흔히 사용되는 것을 알 수 있었다. 번역문에 첨가된 언어적인 단위는 주로 어휘 또는 구이며, 담화적 기능으로 봤을 때 행위나 사건의 주체, 사건에 대한 기사 생산자의 주관적인 태도, 사건의 행위나 동작 등이 가장 많았다. 그 예들을 살펴보면 다음과 같다.

- (5) j090112201 미국이나, 중국이나 이분법 벗어나야
 <a>韩国应从美国和中国的两分法中摆脱出来
 (<a>한국은 미국과 중국의 이분법에서 벗어나야 한다)
- (6) j090106102 노인 범죄 막을 사회 안전망 시급하다
 <a>韩国急需建立阻止高龄犯罪的社会安全网
 (<a>한국은 고령범죄를 막을 사회안전망 구축이 시급하다)

- (7) j090109201 세계 보건위기는 막을 수 있다
<a>我们能够遏制卫生危机
(<a>우리는 보건위기를 막을 수 있다)¹¹⁾
- (8) j090117201 남자답지 않은 남자가 좋다
<a>我喜欢没有阳刚之气的男人
(<a>나는 남자답지 않은 남자가 좋다)

(5)~(8)는 모두 원문 기사에 언급되지 않은 행위의 주체가 번역문에 첨가된 예들이다. 한국어를 모국어로 하는 독자들은 (5), (6)의 행위의 주체가 다른 나라가 아닌 ‘한국’임을 당연하게 읽어낼 수 있으나, 목표어인 중국어로 읽는 독자들에게는 그렇지 않을 수 있다. 이는 독자를 배려한 명시화 전략으로 볼 수 있다. (7)과 (8)은 행위의 주체가 첨가되었다는 점은 동일하나, 그 이유는 다르게 읽힌다. 두 표제의 원문에서 생략되었던 1인칭 대명사로 이루어진 주어는 번역문에서는 명시화된 것이다. 이는 무주어문이 가능한 한국어의 특징과 무주어문이 불가능한 중국어의 언어학적 특징에서 비롯된 것이다.

- (9) c090105102 1951년 일본 법령 '독도는 일본 영토 아니다'
日本1951年法令<a>证明独岛不是日本岛屿
(일본 1951년 법령, 독도가 일본섬이 아님을 <a>증명하다)
- (10) j090116202 국립현대미술관 소격동 시대
国立现代美术馆<a>迎来昭格洞时代
(국립현대미술관이 소격동 시대를 <a>맞이하다)

(9), (10)은 원문에 없던 서술어가 번역문에 첨가된 예이다. (9)의 경우, 서술어를 첨가하지 않고, 원문의 직접인용문 형식대로 번역을 해도 중국어 문법에 적절한 번역문을 얻을 수 있다. 그러나 중국어 기사 표제가 서술어가 없는

11) (7)의 경우처럼(첨가, 생략), 하나의 기사에 두 가지 이상의 번역 현상이 공존하는 경우가 있지만, 예시에는 시선을 분산시키지 않도록 현재 다루고 있는 번역 현상의 주석만 보이도록 한다.

경우가 거의 없다는 점을 감안하면, 이는 목표어인 중국어다운 번역문을 생산하기 위한 의도적인 첨가였다고 할 수 있겠다.

원문 기사 표제가 다루고 있는 사건이나 행위의 주체나 객체를 좀 더 뚜렷하게 드러내기 위해 어휘가 첨가되는 경우도 적지 않게 발견되었다.

- (11) d090131201 한국에 온 클림트
 韩国展示克里木特<a>作品
 (한국에 클림트 <a>작품을 전시하다)
- (12) d090122201 통신 빅뱅
 通信<a>企业<a>之间的“世纪大战”
 (통신 <a>기업간의 <t>‘세기의 대전’</t>)

(11)은 원문에서 은유로 표시되었던 내용, 즉 한국에 오는 것이 ‘클림트’가 아닌 ‘클림트의 작품’임을 명시화하였다. (12)는 ‘빅뱅’이 일어난 주체가 ‘통신’에서 ‘통신 기업’들로 명시화되었다.

위와 같이 사건이나 행위의 주체, 서술어, 주체나 객체를 명시화하기 위한 언어 단위 등의 주요한 첨가 내용 외에도, 다른 목적 또는 이유로 첨가 전략이 사용된 예들을 볼 수 있다.

- (13) d090123101 성장도 일자리도 마이너스, 한국경제 벼랑 끝에
 섰다
 <a>经济比例增长、工作岗位均呈负值；
 韩国经济危在旦夕
 (<a>경제 비례성장, 일자리 모두 마이너스
 를 나타내고, 한국경제는 벼랑 끝에 섰다)
- (14) j090129102 포털의 사생활 침해 어디까지 갈 건가
 门户网站<a>“地图检索”服务<a>严重
 侵犯个人隐私
 (포털의 <a>‘지도검색’ 서비스가 사생활을
 <a>심각하게 침해하다)

(13)은 명사구 ‘경제 비례’가 첨가되었으나 그 이유는 (11), (12)와는 다른

경우로, 원문에서 축약하여 사용하던 ‘성장’을 번역문에서 ‘경제 비례 성장’으로 명시화해 준 경우라고 해야겠다. (14)는 ‘포털이 사생활을 침해하는’ 것이 구체적으로 어떤 서비스인지를 구체화시킨 예이다. 또한, ‘嚴重(심각하게)’라는 부사를 첨가하여 원문 기사 표제를 생산한 필자의 해당 사건이나 행위에 대한 주관적인 느낌까지 번역문에 명시화를 시켰다.

4.2. 치환

앞서, 치환 전략의 사용 빈도가 첨가나 생략보다 더 높게 나타난 것을 알 수 있었다. 목표어 독자의 기대에 부응하기 위해, 또는 사건이나 행위에 대한 시각을 목표어 독자의 그것과 맞추기 위해 원문의 단어나 표현을 번역문에서 다른 단어나 표현으로 치환하는 것은 자주 관찰되는 번역 전략이다.

- (15) j090131101 <t>북한</t>은 협박 중지하고 대화에 응하라
<t>朝鮮</t>应终止对韩要挟政策积极参与对话
(<t>조선</t>은 대한 협박 정책을 중지하고 대화에 적극 응하라)
- (16) j090130202 <t>북</t>·미, ‘9·19 성명’으로 돌아가야
<t>朝</t>美重回《9·19共同声明》
(<t>조</t>·미, ‘9·19 성명’으로 돌아가다)

한국에서 ‘북한’, ‘북’으로 지칭되는 북녘은 중국어 화자들에게는 ‘조선’이라야 옳다. (15), (16)은 원문의 ‘북한’과 그 축약형인 ‘북’을 번역문에서 각각 ‘조선’과 ‘조’로 번역하여 독자의 외교적·정치적 시각을 고려하여 치환한 예이다.

- (17) j090119201 <t>MB 정권</t> 홍보는 식은 피자
<t>李明博政府</t>的宣传就像一张凉透了的比萨饼
(<t>이명박 정부</t>의 홍보는 한 장의 식어 버린 피자와 같다)
- (18) j090114202 <t>다산</t>의 일본론

<丁若鏞>의日本论
(<정약용>의 일본론)

(17), (18)은 중국어 독자들의 배경 지식을 고려한 치환의 예이다. (17)은 원문에 사용된 별명 ‘MB’를 원문에서 ‘이명박’으로 치환하였는데, 이는 한국 언론이나 독자들에게 ‘MB’는 익숙하고 통용되는 대통령의 별명이나, 중국 독자들에게는 익숙하지 않은 점을 고려한 것이다. (18)도 같은 경우로, 중국 독자들은 ‘다산’보다는 ‘정약용’에 익숙한 점이 고려된 치환이다.

(19) c090120101 <윤증현> 경제팀이 배우고 해야 할 일
<新的>经济班子应汲取上届的教训
(<새> 경제팀은 상기의 교훈을 섭취해야 한다)

(19)는 기사 표제에 담고 있는 정보를 일부 단순화하기 위해 치환 전략이 사용되었다. 중국어 독자들에게 자주 바뀌는 이웃나라의 장차관의 이름은 생소하며, 꼭 필요한 정보가 아니다. 오히려 ‘새’ 장관이라는 정보가 더 필요한 것이며, 따라서 원문의 기획재정부 장관 ‘윤증현’을 번역문에서는 ‘새’로 바꾸어 썼다.

(20) c090120201 <수퍼볼보다 인기 높은> 오바마 취임식
奥巴马就职仪式<人气空前火爆>
(오바마 취임식의 <인기가 전례없이 폭발적이다>)

(20)은 오바마 취임식의 인기를 설명하고 있는데, 미국에서 인기가 높은 ‘수퍼볼’과 비교하였다. 수퍼볼은 미국에서 최고의 인기 스포츠 종목인 반면에 중국에서는 잘 알려지지 않은 종목으로, 중국어 독자들은 그 인기의 정도에 대해서는 더욱이 실감이 나지 않을 수 있다. 따라서 번역문에서는 ‘전례없이 폭발적이다’로 치환하여 썼다.

- (21) j090107201 일본 ‘복주머니’에 <t>드리운</t> 불황의 그늘
 日本“福袋”<被</t>经济不景气阴影<所笼罩</t>
 (일본 ‘복주머니’가 경제 불경기의 그림자에 휩
 싸이다)

(21)은 중국어의 단어 결합 특성상 불가피한 치환의 예를 보인다. 한국어에서 ‘그늘’은 ‘드리우다’와 높은 공기관계(co-occurrence)를 보이는 반면, 중국어에서 ‘陰影(그늘)’은 흔히 동사 ‘휩싸이다’와 결합하여 ‘被...陰影所籠罩(...의 그늘에 휩싸이다)’의 피동 구조로 쓰인다.

4.3. 생략

첨가와 치환에 비해 생략 전략은 상대적으로 낮은 사용 빈도를 보인다. 말뭉치에서 관찰된 생략의 대상은 어휘에서 구, 절까지 다양하다¹²⁾.

- (22) j090116102 한·미 FTA 새 국면 ... <o>지혜 모아 최선책
 찾아야</o>
 韩美FTA面临新局面
 (한미 FTA 새 국면을 맞다)
- (23) j090122101 <o>미흡한 1차</o> 구조조정 ... 여기서 멈춰
 선 안 된다
 应将结构调整继续进行下去
 (구조조정을 계속 진행해 나가야 한다)
- (24) j090105102 민주당 <o>국회 본회의장</o> 농성 풀어야
 民主党应停止非法静坐
 (민주당은 불법 농성을 멈춰야 한다)

(22)~(24)는 구나 절 등의 언어 단위들이 생략된 예들이다. (22)는 중앙일보의 사설로, 자국민 독자들을 향해 ‘지혜를 모아 최선책을 찾아야’ 하는 당위성

12) 번역학에서 ‘생략’은 ‘누락’, ‘삭제’ 등 관련 개념과 비교하여 그 엄밀한 차이를 따지기도 하지만, 본고는 원문에 있으나 번역문에는 옮겨지지 않은 경우를 통틀어 ‘생략’으로 지칭하기로 한다.

을 피력하고 있다. 그러나 번역문에서는 불필요한 정보로 분류되어 생략되었다. (23), (24)는 상대적으로 중요도가 낮은 정보의 단순화 전략으로 생략이 일어난 예이다.

4.4. 종결형의 변조

신문 기사 표제의 연구에서 종결형 또는 문말 품사에 대한 연구가 가장 많은데, 이는 신문 기사 표제의 종결 형식이 일반 텍스트와 구분되는 가장 두드러진 언어적 특징이기 때문일 것이다. 표제는 내용과 형식이 완전한 초행완전(初行完全)을 원칙으로 하고 있으나(박갑수 1장), 실제로 완전종결형의 비율이 낮고 비완전종결형의 비율이 높다고 보고되었다(송경화 외 255-69). 본 연구의 말뭉치에 대한 조사 결과도 선행 연구들과 마찬가지로, 한·중 신문 기사 표제의 종결형은 크게 완전종결형과 비완전종결형으로 나눌 수 있었다. 전자의 경우 다시 의문형, 평서형, 청유형, 명령형의 네 가지 형식이 관찰되었고, 비완전종결형의 경우는 표제 문말에 명사, 연결어미, 부사, 조사 등 네 가지의 종결 형식이 관찰되었다. 표제 문말에 명사가 오는 경우는 논항을 취하는 서술성명사가 오는 경우와 그 외의 경우로 세분하였다. 병렬말뭉치에서 관찰된 한·중 신문 기사 표제의 종결형별 분포는 다음과 같다.

분류	한국어 원문					중국어 번역문						
	조선 일보	중앙 일보	동아 일보	합계	비율 (%)	조선 일보	중앙 일보	동아 일보	합계	비율 (%)		
완 전 종 결	명령형	3	6	1	10	5.2	0	1	2	3	1.6	
	평서형	9	23	7	71.8	20.4	29	71	20	6.6	62.8	
	청유형	0	3	2	358.4	2.6	0	0	0	0	0.0	
	의문형	3	3	6	731.4	7.3	6	2	3	6.6	5.8	
비 완 전 종 결	조사	0	4	1	5	2.6	0	0	0	0	0.0	
	명사	일반명사구	20	41	28	101.6	46.6	7	25	23	0	28.8
		논항을 취한 서술성 명사	4	2	0	339.6	3.1	0	0	0	0	0.0
		연결어미	3	17	3	705.3	12.4	0	0	0	0	0.0
	동사구	0	0	0	1,423	0.0	0	2	0	0	1.0	
합계	0	3	10	2	15	7.8	0	1	191	3		

1) 한국어 원문 표제는 ‘관형어 + 명사’의 통사적 구조를 가진 명사구가 46.6%로 가장 높은 비율을 차지하고, 이어 평서형과 연결어미로 종결하는 경우가 각각 20.4%, 12.4%로 그 뒤를 잇는다.

2) 중국어 번역문의 종결형은 원문의 종결형과는 아주 다른 양상을 보이는데, 평서형이 62.8%로 압도적 우위를 차지하고 명사구가 28.8%로 두 번째 순위를 차지한다.

3) 한국어 원문 표제의 종결형이 명사로 끝나는 경우, 앞선 연구들과 달리¹³⁾, 논항을 취하는 서술성 명사가 오는 경우와 그 외의 경우로 나누어 계량하고 분석하였다. 그 이유는 논항을 취하지 않는 명사로 종결되는 경우와 논항을 취하는 서술성명사로 종결하는 두 가지 경우에 따라, 번역의 양상이 판이한 차이를 보이고 있기 때문이다. 한·중 기사 표제 번역에서 논항을 취하는 서술성 명사로 종결되는 경우는 서술성명사가 중국어의 동사로 번역되고, 종결형은 평서형으로 변조되는 것이 일반적이다. 반면에 일반명사구의 경우, 중국어 번역문이 통사 구조의 변조가 없이 원문과 같은 통사 구조를 유지하게 되는 것으로 관찰되었다. 우선, 일반명사구의 예를 살펴보도록 한다.

- | | |
|-----------------|---|
| (25) d090101201 | 공기업 사장의 선물
国有企业社长的“礼物”
(공기업 사장의 ‘선물’) |
| (26) c090122201 | 오바마 취임사의 한반도 구상 |

13) 송경화 외(255-69)에서는 기사 표제의 문말에 명사가 오는 경우를 하나로 묶어 계량화하였고, 조지은(75-91)에서는 체언과 동작성명사를 구분하였으나, 그 구분의 이유에 대해서는 ‘동작성 명사도 명사로 분류하여도 무방하여 [...] 선행연구와의 차이점을 부각시키기 위해 [...]’라고 하여 서술성 명사의 명사와의 구분의 언어학적·번역학적 필요성을 명확히 제시하지 못하고 있다. 김한식(51-68)은 기사 표제의 문말 품사 분석에서 일반 체언과 ‘동작성을 지닌 한자어 명사’를 구분하여 기존 연구들과 차이를 보이고 있으나, ‘[...] 이렇게 분류한 것은 헤드라인의 문말에 상당히 많이 사용되고 있기 때문’이라고만 밝혀, 그런 구분 방법이 언어학적으로나 번역학적으로 가지는 의미에 대한 논의는 충분히 이루어지지 않았다.

奥巴马就职演说中的韩半岛构想
(오바마 취임사의 한반도 구상)

(25)는 비록 서술성명사 ‘구상’으로 종결되었으나, ‘구상’이 해당 표제에서 논항을 취하지 않기 때문에 번역문의 통사적 구성은 원문과 같은 구조를 유지하게 된다. 이와는 달리 논항을 취한 서술성명사 종결형의 경우, 번역문에서 원문의 명사구와는 다른 평서형의 완전종결형으로 변조된다. 아래가 그 예이다.

(27) c090121202 열싸안은 인종·문화... 관용과 도전이 미(美)역사 창조
人种和文化相互融合 宽容和挑战创造美国历史
(인종과 문화가 서로 융합하고, 관용과 도전이 미국 역사를 창조한다)

위의 예에서 ‘창조’는 서술성명사로 ‘미 역사’를 목적어 논항으로 취하며, 따라서 번역문에서 ‘창조’는 동사 ‘創造’로 번역이 된 것을 확인할 수 있다.

4) 한국어 원문 표제는 명령형과 청유형이 각각 5.2%, 2.6%로 일정 비율을 차지하는데 반해, 중국어 번역문에는 명령형이 1.6%로 축소되고, 청유형은 전혀 나타나지 않았다. 이는 한국어 원문과 중국어 번역문의 독자가 달라지는 것을 고려하거나, 또는 중국어 표제 뽑기의 관행을 따라 번역자가 종결형의 변조 전략을 구사하기 때문인 것으로 분석된다.

아래는 원문 기사 표제의 명령형이 다른 종결형으로 바뀐 예이다.

(28) c090123102 오바마와 공화당의 '정치 협력 모델'을 보라
奥巴马和共和党“政治合作”给我们的启示
(오바마와 공화당의 ‘정치 협력’이 우리에게 주는 계시)
(29) j090130102 용산 참사를 거리시위 빌미로 삼지 말라
龙山惨剧不应该成为街头示威的导火索
(용산 참극은 ‘거리시위’의 도화선이 되지 말아

- (30) j090101201 야 한다)
 ‘기억상실 정치’부터 추방하자
 改变应从“失忆症政治”开始
 (‘기억상실 정치’부터 변화가 시작되어야 한다)

(28)~(30)의 원문은 한국 대표일간지 사설들로, 명령형 또는 청유형의 종결형을 통해 자국민을 향해 강력한 정치 메시지를 던지고 있다. 그러나 중국어 화자들은 그 대상이 될 수 없으므로, 명사구 또는 평서문으로 번역되었다.

5) 한국어 원문 표제에서 연결어미로 종결하는 경우가 12.4%를 차지하여 꽤 높은 비율을 보인 반면, 중국어 번역문에서는 이런 경우가 전혀 출현하지 않는다. 이는 중국어의 언어적 특성상 어미로 종결하는 비완전종결형이 있을 수 없고, 대부분 평서형으로 변조되기 때문이다.

- (31) c090107101 기름값 쌀 때 해외 자원 투자 서둘러야
 韩国应趁油价下跌积极确保海外资源
 (한국은 기름값이 내렸을 때 해외 자원 확보에 서둘러야 한다)
 (32) j090130202 북·미, ‘9·19 성명’으로 돌아가야
 朝美重回《9·19共同声明》
 (조·미, ‘9·19공동성명’으로 돌아가다)

4.5 기계번역을 위한 규칙 작성의 예

4.1~4.4의 논의를 바탕으로, 의역이 발생하는 원문의 문맥 조건과 출력문의 구성을 기계번역에 적용 가능한 규칙으로 형식화할 수 있는지에 대해서 약간의 논의를 시도해 보기로 한다. 이를 위해 첨가의 예를 다시 들어 본다.

- (33) c090121201 철거민 진압작전의 가슴 아픈 결말
 <a>警方强行驱赶迁户酿成悲剧
 (<a>경찰, 철거민을 강제로 내쫓아 비극을 만든다)

(34) c090109201 국회의원 숫자 줄이기 서명 운동 벌여야
<a>韩国国会也要进行“裁员”
(<a>한국 국회도 ‘감원’을 해야 한다)

위의 예문들에서처럼 기계번역에서도 원문에 출현하지 않은 행위나 사건의 주체를 번역문에서 명시화해 줄 수 있는 방법은 없을까? 우선, 2장에서 언급된 해당 표제들의 상위말뭉치, 즉 해당 표제를 포함한 기사를 살펴본다.

(33)이 추출된 기사 텍스트:

```
<file>c090121201kk.txt</file>
<text>13)
<body>
<div>
<head>철거민 진압작전의 가슴 아픈 결말</head>
<p><s>20일 경찰이 서울 용산4재개발구역에서 충분한 보상금을 지급하라며 건물 점거농성을 벌이던 철거민들을 강제 해산시키다가 불이 나 농성자 5명과 경찰관 1명이 사망했다.</s><s>경찰 18명, 농성자 6명은 크게 다쳤다.</s>
...
</div>
</body>
</text>
```

14) <txt>, <body>, <div> 등등의 태그는 텍스트의 구조를 보이기 위한 말뭉치 가공 태그 집합이다. 여기서는 ‘21세기 세종계획’의 말뭉치 가공 태그 집합을 그대로 사용한 외에 기사 텍스트의 특성상 부표제가 사용되며, 일반 기사 표제, 본문과는 구분되는 특징을 지닌 것을 감안하여 <subhead>, </subhead>의 태그를 추가하여 말뭉치에 주석했다.

(34)가 추출된 기사 텍스트:

```

<file>c090109201kk.txt</file>
<text>
<body>
<div>
<head>국회의원 숫자 줄이기 서명 운동 벌여야</head>
<subhead>"한국은 미국 인구 16%에 국회의원 숫자는 70% 넘어</subhead>
<subhead>국회의원 숫자 줄여야 국회 효율 높아지고 세금 가벼워져"</subhead>
<head>강천석·주필</head>
<p><s>대한민국</s> 국민의 새해맞이는 우스운 사람들의 우스운 짓거리로 엉망진창이 되고 말았다.</s><s>이스라엘과 이슬람 무장 정파(政派) 하마스가 대포와 로켓 포탄을 주고받는 가자 지구의 팔레스타인 난민(難民)과 더불어 세계에서 가장 처량한 신세가 됐다.</s><s>
...
</div>
</body>
</text>

```

위의 예시에서처럼, 두 경우 모두 표제에서 드러나지 않은 행위나 사건의 주체는 부표제 또는 본문에 나타나는 첫 번째 명사이자 주어인 것으로 나타났다. 그 확률이 얼마나 높을 지에 대해서는 보다 큰 규모의 데이터에 대한 계량적인 연구가 이루어져야 할 것이다.

6. 나오는 말

본 연구는 우선 언어간 기사 표제에 대한 선행 연구들을 자세히 검토하여 그 장단점을 분석하였다. 이어서 소규모의 한·중 기사 표제 병렬말뭉치를 구축하고 가공하였다. 이를 대상으로 원문과 번역문을 비교하고 검토하여 기계번역을 위한 재활용이 가능한 직역과 의역, 그리고 현 단계에서 재활용이 불가능할 것으로 보이는 개작 세 가지의 유형으로 분류하여 각각의 비율을 산출하였

다. 아울러 원문과 번역문이 어휘적인 내용과 언어 형식적으로 대응이 안 되는 경우를 첨가, 생략, 치환, 종결형의 변조 등 네 가지로 나누어 어떻게 그리고 왜 다른지를 살펴보았다. 마지막으로, 위의 논의를 바탕으로, 기계번역에 적용 가능한 번역 규칙 작성의 예를 보였다.

본 연구를 통해 한·중 신문 기사 표제의 번역에서, 엄밀한 의미에서 번역으로 없는 ‘개작’의 비율이 17.6%에 그쳐 한·중 신문 기사 표제 병렬말뭉치가 기계번역을 위한 활용 가능성이 충분하다는 점이 증명되었다. 또, 종결형의 변조에 있어서, 한국어 원문 표제의 종결형이 명사로 끝나는 경우, 선행 연구들과 달리, 한·중 번역 규칙을 고려하여 논항을 취하는 서술성 명사가 오는 경우와 그 외의 경우로 나누어 계량하고 분석하였다. 논항을 취하지 않는 명사로 종결되는 경우와 논항을 취하는 서술성명사로 종결하는 두 가지 경우에 따라, 한중 번역의 양상이 어떻게 달라지는지를 예시를 통해 살펴보았다.

단, 한·중 기사 표제의 번역 전략의 기계번역을 위한 규칙화 작업은 지면 관계상 충분히 논의되지 못하였다. 본고에서 기술된 한·중 기사 표제의 여러 번역 전략들을 기계번역에 적용할 수 있도록 일일이 검토하고 규칙화하며, 실제로 기계번역 엔진에 적용하여 실효성 여부를 평가하는 작업은 다음으로 미룬다. 앞으로, 본고에서 다룬 사실·칼럼과는 사뭇 다른 언어학적 특징을 보이는 보도·해설에 대한 한·중 기사 표제 연구로 연구 대상을 넓히고 대규모 병렬 말뭉치로 연구를 확장하며, 그에 대한 보다 세밀한 번역학적인 연구가 이루어져야 할 것이다.

참고문헌

- 강지혜. 2008. 「번역기사의 제목에 관한 연구 : 「뉴스위크 한국판」의 북한 관련 번역기사를 중심으로」. 『번역학연구』 9(2): 7-43.
- 김경희. 2000. 「인터넷 뉴스 이용자의 정보처리 지속에 관한 연구」. 『언론과 사회』9(1): 47-97.
- Kim, Young-Shin. 2003. 「Translation of English News Headlines into Korean」. 『번역학연구』 4(1): 145-166.

- 김한식. 2001. 「한-일 양국 신문 헤드라인의 표현 및 그 번역」. 『통번역학연구』 5: 51-68.
- 김희진. 2000. 「어법 틀리고 단어선택도 부적절: 정치면 기사 표제의 문제」. 『신문과 방송』 358: 50-54.
- . 2005. 「신문·방송 기사 제목의 문제 :잘못된 단어, 사동법과 명령법 오용」. 『신문과 방송』 418: 38-41.
- 박갑수. 1998. 『신문·광고의 문체와 표현』. 서울: 집문당.
- 박갑수. 1994. 「언어 규범과 현실의 간극: 신문 표제어의 문제」. 『先淸語文』 22: 133-152.
- 송경화, 강범모. 2006. 「신문 기사의 언어 사용 양상: 코퍼스언어학적 접근」. 『인지과학』 17(4): 255-269.
- 이영옥. 2002. 「한영간 신문 표제어 표현의 차이: 동음이의어와 인용문」. 『커뮤니케이션학연구』 10(1): 150-170.
- 조지은. 2008. 「효율적인 한일번역을 위한 신문 사설 제목 유형 분석」. 『일어일문학』 39: 75-91.
- 황금하, 송희정 외. 2000. 「뉴스 타이틀 번역을 위한 중한 기계 번역 시스템」. 『한글 및 한국어 정보처리대회 논문집』 350-357.
- 陳麗昭. 2002. 中英新聞標題對比研究. 福建師範大學碩士學位論文.
- 易立梅. 2007. “漢英體育新聞標題對比研究”. 《科技信息(學術研究)》2007年20期: 264-265.
- 尹世超. 2005. 標題語法. 北京: 商務印書館.
- Farzaneh Khodabandeh. 2007. “A Contrastive Analysis of English and Persian Newspaper Headlines.” in *The Linguistics Journal*, April 2007 volume 2, Issue 1: 91-127.
- Kenji Ono. 2003. “Translation of News Headlines.” in *Proceedings of Machine Translation Summit IX*. 1-8.
- R.M.K. Sinha. 2002. “Translating News Headings from English to Hindi.” in *Proceedings of Artificial Intelligence and Soft Computing 2002*. 346-353.
- Saha, D. and S. Bandyopadhyay. 2005. “A Semantics-based English-Bengali EBMT System for Translating News Headlines.” MT Summit X, in

Proceedings of Second Workshop on Example-Based Machine Translation. 125-133.

Takehiko Yoshimi. 2001. "Improvement of Translation Quality of English Newspaper Headlines by Automatic Pre-editing." in *Machine Translation* Volume 16(4): 496-500.

[Abstract]

**A Study on Parallel Corpus of Korean-Chinese Newspaper Headlines
online: A Trial Study for Machine Translation**

Huang, Yinxia
(Yonsei University)

Studies on translation of newspaper headlines from Korean to Chinese are first conducted. To see if parallel corpus of Korean-Chinese newspaper headlines is applicable for machine translation, Korean newspaper headlines and their Chinese correspondents are analyzed through construction and annotation of parallel corpus. The corpus covers one-month newspaper headlines of three Korean major newspapers online(Dong-a Ilbo, Chosun Ilbo and JoongAng Ilbo). The translation strategies of Korean-Chinese newspaper headlines are classified into three categories, and proportions of each category are calculated: literal translations(36.7%) and free translations(45.7%) that can be applied for machine translation, the rewrite translations that can not be applied for machine translation(17.6%). The literal translations and free translations that can be applied for machine translation(82.4%), are further analyzed four translation strategies, such as addition, omission, transposition and change of mood. Finally, aforementioned strategies are tried to implement to Korean-Chinese machine-translation.

▶ Key Words: news headline, translate Korean into Chinese, parallel corpus, machine translation

황은하(Huang, Yinxia)

연세대학교 언어정보연구원 연구원

enha@lex.yonsei.ac.kr

관심분야: 기계번역, 한중 대조분석, 한중 번역, 말뭉치 언어학

논문투고일: 2009년 08월 05일

심사완료일: 2009년 08월 20일

게재확정일: 2009년 09월 05일