A Study on Translation of 'Hata' Verb Based on Relation between 'X-hata' and 'X-lul hata' in Korean-Chinese Machine Translation

Young-Ae Seo*, Yinxia Huang, MuXyo Hong, Sung-Kwon Choi

*Natural Language Processing Team, ETRI

요  수  약

패턴에 기반한 자동번역 시스템에 있어서, 번역율을 결정하는 가장 중요한 요소는 패턴의 양과 질이다. 필요한 대량의 패턴을 단기에 구축하기 위해서 패턴 정보 기술상의 업여성을 줄일 필요가 있다. '하다' 동사가 한국어에서 가장 번역하게 사용되는 동사 중의 하나로, 'X를 하다'와 'X하다'의 2가지 구문 구조를 갖는다. 본 논문에서는 동사구 패턴에 기반한 한국 자동 번역 시스템에서 'X를 하다'와 'X하다' 구문간의 상관성을 이용하여 두 구문이 'X하다' 동사에 대한 동사구 패턴을 공유하도록 함으로써 동사구 패턴 구축 시에 정보의 중복 기술을 피하고, 패턴의 적용 범위를 넓힌다. 패턴의 공유로 인해 발생하는 대역문 생성 시의 문제점은, 'X를 하다' 구문에 대응하는 중국어 대역 형태 별로 분류하여 해결함으로써 패턴을 공유하면서도 번역의 질은 유지시킬 수 있는 방법을 제안한다.

1. 서론

현재 한국전자 통신 연구원에서 개발 중인 한국 자동 번역 시스템 Tramie-KC는 데이터에 기반한 자동 번역 시스템으로서, 동사구 패턴을 기반으로 하여 한국 자동 번역을 수행한다.[1] 한중 동사구 패턴의 구축은 고용용고 난이도의 작업이므로, 동사구 패턴이 가능한 많은 한국어 용언의 쓰임을 처리할 수 있도록 패턴의 적용 범위(coverage)를 높일 필요가 있다.

본 논문에서는 Tramie-KC에서 한국어에서 많이 사용되는 '하다' 동사에 대한 동사구 패턴의 적용 범위를 높이는 방법에 대하여 제안한다. 이를 위해 'X하다'와 'X를 하다' 구문간의 상관성을 이용하여 두 구문이 'X하다' 동사에 대한 동사구 패턴을 공유한다. 패턴의 공유로 인해 발생하는 대역문 생성 시의 문제점은 'X를 하다' 구문에 대응하는 중국어 대역 형태 별로 분류하여 동사구 패턴이나 형태소 사전에 필요한 정보를 기술하여 처리함으로써, 정보 기술 시의 잉여성(redundancy)을 줄이면서 번역의 질은 유지시킬 수 있는 방법을 제안한다.

2. 관련 연구


3. ‘하다’ 동사의 한중 동사구 패턴
본 장에서는 한중 자동변역 시스템, Tramie-KC에서 사용하는 동사구 패턴과 ‘하다’ 구문의 특성에 대해 간략히 기술하고 ‘하하다’와 ‘를 하다’ 구문이 동사구 패턴을 공유할 수 있음을 보인다.

3.1 한중 동사구 패턴

가다1 : A=사람!가 B=건축물!에 가다! 
> A 빠/ B [그는 범인에 갔다]

[예1] ‘가다’의 한중 동사구 패턴 예

이러한 한중 동사구 패턴은 매우 고비용, 고난이도의 작업이므로, 의미적으로 같은 동사구일 경우 동사 구 패턴 정보를 공유함으로써 패턴 구축비용을 줄일 필요가 있다.

3.2 ‘하하다’와 ‘를 하다’ 구문의 특징

(1-1) 그는 마침내 결정을 했다.
(1-2) 그는 마침내 결정했다.


결정하다3 : A=사람!가 B=방법!을 결정하다! 
> A 결정!+ B [그가 방침을 결정하다]

[예2] ‘결정하다’의 한중 동사구 패턴 예

하하다26 : A=사람!가 B=방법!을 결정!를 하다! 
> A 결정!+ B [그가 방침을 결정한다]

[예3] ‘하하다’의 한중 동사구 패턴 예


4. 동사구 패턴 공유 시의 문제점 및 해결 방안
동사구 패턴 공유로 인해 대역문 생성 시에 발생하는 문제점과 이에 대한 해결책에 대해 논한다.

4.1 동사구 패턴 공유 시의 문제점
(1-1)과 (1-2)가 동일한 동사구 패턴을 이용하여 번역 가능함을 보였다. 그러나, (2-1)과 같이 ‘하하다’와 결합된 명사 ‘X′가 관련이나 관형절의 수식을 받는 경우 대역문의 생성에 문제가 발생한다.

(2-1) 그는 마침내 여러분 결정을 했다.
(2-2) 그는 마침내 여러분 결정하는 것을 했다.
4.2 'X하다'와 'X를 하다' 대역어의 상관관계에 대한 분류

관형어/절 수식 문제의 근본적 해결책은 'X'의 대역어가 나타나는 자연스런 대역문을 생성하는 것이 다. 이를 위해서는 'X하다'의 중국어 대역문이 '중국어 기능동사 + X의 중국어 대역어'의 형태로 변
역되거나, 중국어의 대역어가 동병(動賓) 구조로 이루어질 경우 가능하다. 동병 구조와 '강제'와 같이
한 중국어 동사의 내부 구조가 동사와 목적어의 결
합에 해당하는 경우를 말한다.

본 논문의 배치에 맞추어 한국어의 'X를 하다' 용
연구에 도움이 되는 중국어의 변형을 [표1]과 같이 분
류하였다. 중국어 변형 결과에 따라, 'X'의 중국어 기
능동사와 결합하는 유형과 결합하지 않는 유형으로
분류하고, 기능동사와 결합 불가능한 동사는 그 내부구조에 근거하여 다시 2가지 유형으로 하위분류
하여 유형화하여 총 3가지로 분류하였다.

<table>
<thead>
<tr>
<th>주형</th>
<th>중국어 대역어</th>
<th>한국어 'X를 하다'의 예</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 A</td>
<td>加(加, 增)</td>
<td>우대를 하다</td>
</tr>
<tr>
<td></td>
<td>给予(予)</td>
<td>给予과 결합 가능</td>
</tr>
<tr>
<td>2 B</td>
<td>作(作, 行)</td>
<td>연설을 하다</td>
</tr>
<tr>
<td></td>
<td>進行과 결합 가능</td>
<td></td>
</tr>
<tr>
<td>3 C</td>
<td>A, B 기능동사와 모두 결합 가능</td>
<td>비평을 하다</td>
</tr>
<tr>
<td></td>
<td>동신구조</td>
<td>강의를 하다</td>
</tr>
<tr>
<td></td>
<td>기타</td>
<td>공부를 하다</td>
</tr>
</tbody>
</table>

[표1] 'X를 하다'에 대한 중국어 변형 주형

유형1의 처리를 위해서는 'X'의 중국어 대역어의 정보에 대하여 대응하는 기능동사 정보를 범도로 형
태소 사전에 기술한다. 중국어의 기능동사는 그의
미적 특성에 근거하여 A형식-'가, 给予, 予, 给予'와 B형식-'作, 進行'으로 나뉜다[6]. 중국어 명사
와 결합 가능한 기능동사의 선택은 각 언어의 개별
적 특성에 의해 결정된다. 형태소 사전은 한국어 형
태소에 편, 형태소 코드, 의미코드(SEM), 중국어 대
역어(CROOT) 정보, 중국어 기능동사 정보 (CSVERB) 등
을 담고 있다. [예4]는 '결정'에 대한 형태소 사전 정보의 예이다.

유형2는 동사구 패턴에 중국어 동사의 내부 구조로
보를 기술하여 'X'의 중국어 대역에 대한 린크정보
를 기술한다. [예5]의 '식사하다'에 대한 동사구 패턴
에서 중국어부의 'v'는 동병 구조에서의 중국어 동
사(吃), 'n'은 목적어(飯)임을 표시한다.

120012 479
{(SEM 선택) (EROOT decision) (CROOT 결정) (CSVERB 作, 加以)} ......}

[예4] '결정'의 형태소 사전 정보의 예

식사하다1 : A=사람!가 식사하다! 
>n A 作:v飯'n [아비지께서 식사하시다.]

[예5] '식사하다'의 한중 동사구 패턴의 예

유형3은 유형1, 2에 속하지 않는 그 외의 경우이다.

4.3 변역 유형별 관형어/절의 처리

유형1에 속하는 'X를 하다'의 'X'가 관형어/절의 수식을 받는 경우, 'X하다' 동사구 패턴을 가진은
후, 중국어 변형부의 동사 부분을 'X'의 형태소 사
전에서 가져온 CVERB와 CROOT 정보를 이용하
여 중국어 '형식동사+서술성명사'의 형태로 처리한
다. 그리고 나서 기존의 Tramie-KC의 관형어/절 처
리 부품을 호출하여 관형어/절이 'X'의 중국어 대역
어를 수식하도록 하여 변역을 완료한다. [그림1]은
(2-1)의 문장이 [예2]의 동사구 패턴과 [예4]의 형태
소 사전 정보를 이용하여 처리되는 과정을 보여준다.

유형2의 경우는, 유형1에서의 치환과정을 거치지
'X하다' 동사구 패턴의 중국어 대역부에 'n'의 맥상정보로 가중된 'X'의 중국어 대역어를 수시하도록 관행어/설 치러 무단을 허용한다.

유형3의 경우 상피보면 한국어에서의 관행어 혹은 관행어가 중국어 문장에서는 숭부를 수시하는 보로 나타나는 것이 일반적이다. (3-1)에 대한 대역 문 (3-2)가 바로 그 예이며, 혼역된 중국어 문장 (3-2)가 한국어로 변역하면 (3-2')와 같다.

(3-1) 그는 늦은 출발을 했지만 실망하지 않았다.
(3-2) 그는 늦게 출발했지만, 실망하지 않았다.
(3-2') 그는 늦게 출발했지만, 실망하지 않았다.

4.4 실험 결과

경제, 뉴스, 코퍼스로부터 169개의 'X를 하다' 구문이 포함된 문장을 무작위로 추출하여 본 논문에서 제안한 방법으로 변역하였다. 실험 결과, 89.3%를 차지하는 112개의 문장은 제안된 방법으로 변역하였을 경우 자연스러운 변역 출력문을 얻을 수 있었다.

그러나, 유형 3의 경우 중 제안한 방법으로 변역할 경우 원문은 의미적 차이가 있는 출력문을 얻음으로써 좋은 결과를 얻을 수 있는 경우가 18개, 즉 10.7% 발견되었다. 이 경우를 기타로 분류하여 실험 결과를 표로 나타내면 아래와 같다.

<table>
<thead>
<tr>
<th>유형</th>
<th>출력 변도(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A 19 (11.2%)</td>
</tr>
<tr>
<td>1</td>
<td>B 39 (23.1%)</td>
</tr>
<tr>
<td>1</td>
<td>C 52 (30.8%)</td>
</tr>
<tr>
<td>2</td>
<td>17 (10.1%)</td>
</tr>
<tr>
<td>3</td>
<td>24 (14.2%)</td>
</tr>
<tr>
<td>기타</td>
<td>18 (10.7%)</td>
</tr>
<tr>
<td>계</td>
<td>169 (100%)</td>
</tr>
</tbody>
</table>

[표2] 실험 결과

기타의 경우는 유형 3의 처리방법으로 변역할 의 미적인 차이를 남았다.

(4-1) 그가 어려운 공부를 한다.
(4-2) 그는 어렵게 공부를 한다.
(4-3) 그는 어려운 공부를 한다.

변역된 중국어 (4-2)를 한국어로 그대로 직역하면 (4-2')가 된다. 어학적적인 견지에서 볼 때 (4-1)에 대한 가장 중국어타운 표현은 (4-3)이다. (4-3)을 다시 한국어로 변역하면 원문인 (4-1)과는 구조가 다른 (4-3')의 구조를 갖게 된다. 기타 경우의 해결을 위해서는 동사구 패턴과 형태소 사전 정보 이외의 보다 심화된 중국어 의미론적, 동사론적 그리고 화용론적 연구와 그에 따른 연구가 추가적으로 이루어져야 할 것이다.

5. 결론

본 논문은 동사구 패턴에 기반한 한중 자동변역 시스템 Tramie-KC에서 'X를 하다'와 'X하다' 구문의 상관성을 이용하여 'X하다' 동사구 패턴을 공유하여 처리함으로써 정확도 높은 결과를 얻을 수 있다. 패턴의 공유 시 관행어나 관형절의 수식 문제 가 발생함을 보이고, 이를 해결하기 위해 'X를 하다'에 대응하는 중국어 변역 유형을 3가지 형태로 분류하여 각 유형에 대한 변역 방법을 제안하였다. 제안된 방법은 다른 기능동사 구문 등에도 확대 적용 가능하다. 그러나, 유형 3의 경우는 보다 자연스러운 변역문의 생성을 위해서는 중국어의 표현 특성을 반영할 수 있는 추가의 연구가 이루어져야 할 것이다.

참고문헌


610