

## 영조기계번역에서 구조적모호성의 해소방법

박 명 철

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《나라의 과학기술을 세계적수준에 올려세우자면 발전된 과학기술을 받아들이는것과 함께 새로운 과학기술분야를 개척하고 그 성과를 인민경제에 적극 받아들여야 합니다.》

(《김정일선집》 증보판 제11권 138~139페이지)

나라의 과학기술을 발전시키고 세계선진과학기술을 제때에 받아들이는데서는 나라들 사이의 언어장벽을 제거하는 문제가 중요한 과제로 나서며 여기에서 기계번역프로그램이 차지하는 위치와 중요성은 매우 크다.

이 글에서는 영조기계번역체계개발에서 중요한 위치를 차지하는 입구어의 구조적모호성을 해소하기 위한 방법론적문제에 대하여 분석하려고 한다.

영조기계번역에서 구조적모호성은 입구되는 문장구조가 구조적으로 가능한 두개이상으로 해석되는 경우를 말한다.

례: ① Flying planes can be dangerous.

(비행기를 조종하는것은 위험할수 있다.)

가능한 구조: - It can be dangerous to fly planes.

- Planes which are flying can be dangerous.

② Time flies like an arrow.

(세월은 류수같다.)

가능한 구조: - The passage of time is as quick as an arrow.

- A species of flies called 'time flies' enjoy an arrow.

③ The man saw the girl with the telescope.

(그 사람은 망원경으로 그 소녀를 보았다.)

가능한 구조: - The man saw the girl who possessed the telescope.

- The man saw the girl with the aid of the telescope.

위의 문장들에서의 모호성을 사람이 인식하는것은 실례 ②에서와 같이 언제나 쉬운것은 아니다. 특정한 정황이나 본문의 문맥에서 문장들은 모호하지 않을수 있다. 만일 실례 ③이 어떠한 이야기책에 나오게 된다면 이야기줄거리로부터 망원경을 가진것이 man(남자)인가 아니면 girl(처녀애)인가가 명백해질것이다.

문장론적인 해석이 어떠한 문장에 대하여 하나이상의 해석결과를 만들어낼 때 정확한 해석결과를 선택하는 방법을 찾는것이 필요하다. 그 리유는 목표언어로의 번역은 종종 선택된 해석결과에 따라 달라지기때문이다. 즉 실례 ③의 조선어번역은 ④-1 아니면 ④-2로서 모호성이 없는 문장으로 되여야 한다.

례: ④-1 그 사람은 망원경을 가지고있는 소녀를 보았다.

MAN subj TELESCOPE obj HOLDING GIRL obj SAW pred  
The man saw the girl who was holding the telescope.

④-2 그 사람은 망원경으로 그 소녀를 보았다.

MAN subj TELESCOPE inst GIRL obj SAW pred

The man, using the telescope, saw the girl.

영조기제번역에서 구조적모호성을 해소하는 방법에는 의미론적 및 다른 언어학적정보의 리용, 문맥적실마리의 리용, 비언어학적인 현실세계에 대한 지식의 리용, 대화형참조 등의 방법들이 있다.

영조기제번역에서 영어문장들의 구조적모호성은 무엇보다먼저 언어학적지식으로 불리는 정보들을 리용하여 해소될수 있다.

언어학적지식에는 여러가지가 있지만 공통적인것은 모두 문장들이 서술하는 실생활사건이 아닌 단어들과 단어들의 결합방식에 대한 정보를 리용한다는것이다.

이러한 방법은 공기제한에 대한 정보 즉 어떠한 구조에서 일정한 요소들의 출현이 어떻게 다른 요소들에 영향을 미치는가 하는 정보를 해석기에 제공하는 방법이다.

공기제한에 대한 정보를 얻기 위하여 동사들의 하위범주화들을 리용할수 있다. 하위범주화들을 리용하면 어떤 형태의 보충어들이 특정한 동사와 결합할수 있는가 하는 정보를 얻을수 있다. 실례로 《주다》의 의미를 가지는 give 동사는 《주는 대상》을 가리키는 명사를 주어로, 《주어지는 대상》을 가리키는 명사를 직접보어로, 《받는 대상》을 가리키는 명사를 간접보어로 가지게 된다. 더우기 어떠한 형태의 명사들이 이러한 문장론적인 역할을 하는가 하는것을 《주는 대상》이 활동체여야 한다는것 등과 같이 그 명사들에 의미론적인 특징을 부여하여 어느 정도 규정지을수 있다.

다음의 실례를 들어보자.

례: ⑤a I read about the air crash in France.

(나는 프랑스에서 있는 비행기추락에 대한 기사를 읽었다.)

⑤b I read about the air crash in the paper.

(나는 신문에서 비행기추락에 대한 기사를 읽었다.)

실례 ⑤와 같은 모호한 문장쌍들이 주어졌을 때 해석기는 관련된 명사가 《읽기 가능한》(readable)이라는 의미표시가 달린 명사들의 모임 실례로 book, magazine, newspaper 등에 속해있을 때 전치사 in을 가진 전치사구에 의하여 동사 read가 수식될수 있다는것을 안다면 정확한 해석결과를 만들어낼수 있다.

다른 실례를 들어보자.

례: ⑤c Have you seen the story about the air crash in the paper?

(신문에 난 비행기추락에 대한 기사를 보았습니까?)

⑤d Have you seen the story about the air crash in the jungle?

(원시림에서 있는 비행기추락에 대한 기사를 보았습니까?)

실례 ⑤c와 ⑤d와 같은 문장들은 story와 in-전치사구에 대하여 알맞는 정보가 부호화된다면 정확히 해석될수 있다.

례: ⑥a He wrote the letter with a fountain-pen. (그는 만년필로 편지를 썼다.)

비교: He wrote the letter with the parcel. (그는 소포와 함께 편지를 썼다.)

⑥b He visited the museum with his brother.

(그는 자기 형과 함께 박물관을 참관하였다.)

비교: He visited the museum with the Elgin marbles.

(그는 고대그리스의 대리석조각이 있는 박물관을 참관하였다.)

⑥c He told the story with a funny accent. (그는 재미난 억양으로 이야기를 하였다.)

비교: He told the story with a funny ending. (그는 재미나게 끝나는 이야기를 해주었다.)

실례 ⑥의 문장들을 보면 ⑥a에서는 단어 write가 도구 with-구문을 취하고 도구는 물리적인 보어가 되어야 한다는 정보를, ⑥b에서는 단어 visit와 유사한 동사들의 동반자가 활동체가 되어야 한다는 정보를, ⑥c에서는 단어 tell이 방식 with-구문을 취하며 가능한 보충어들의 제한된 범위(stutter, accent 등)를 가진다는 정보를 해석기가 입수할수 있다면 정확히 해석될것이다.

영조기계번역에서 영어문장들의 구조적모호성은 다음으로 문맥적지식을 리용하여 해소될수 있다.

실천적으로 볼 때에는 실지로 모호한 문장들은 거의나 없다. 만일 모호성을 해소하는데 리용되는것이 없다면 보통 해당 문장이 출현하는 문맥이 정확한 해석을 제기하게 된다.

우의 실례 ③에서는 이야기줄거리가 두 대상중에서 어느 대상이 망원경(telescope)을 가지고있는가를 암시하게 된다는것을 보여준다. 그러한 정보는 해당 문장의 앞문장 또는 앞단락, 지어는 훨씬 앞에 있는 내용들에서 언급될수도 있다. 그런데 특정한 경우에 모호성해소에 도움이 되는 이러한 문맥적지식을 리용할수 있는 기계번역체계는 현재 개발되지 못하였다. 다른 측면에서 《the man has a telescope》와 같은 본문구동지식을 보관하는 효과적인 방법이 있다고 가정하여도 번역기계가 실지 필요한 정보가 어느것이고 그것을 언제 리용하며 그것이 필요할 때까지 얼마나 오래동안 보관하여야 하는가를 안다는것은 매우 힘들다. 그러므로 주어진 문장에서 또는 주어진 본문의 매개 문장들에서 추출되는 모든 사실들을 그 어떤 모호성을 해소하기 위하여 모두 추출하고 보관하는 다기능화된 기계번역체계를 개발하는 문제가 절실히 제기된다.

영조기계번역에서 영어문장들의 구조적모호성은 다음으로 현실세계에 대한 지식을 리용하여 해소할수 있다.

현실세계에 대한 지식을 적용하여 구조적모호성을 해소하는 방법은 실례 ⑦과 같은 경우에 적용된다.

례: ⑦ The man saw the horse with the telescope.

(그 사람은 망원경으로 말을 보았다.)

실례 ⑦에서 우리는 with the telescope가 반드시 saw를 수식하고 특히 horse의 경우에는 그 선택대상으로 될수 없다는것을 현실세계에 대한 지식으로부터 알고있으며 이 방법이 영조기계번역체계에 도입되면 영어문장의 구조적모호성이 해소될수 있다.

현실세계에 대한 지식과 우에서 논의된 언어학적지식사이의 차이를 명백히 하기는 어렵다. 실례로 아래의 문장들을 들수 있다.

례: ⑧a We will meet the man you told us about yesterday.

(우리는 어제 당신이 우리에게 이야기한 그 사람을 만날것이다.)

⑧b We will meet the man you told us about tomorrow.

(우리는 래일 당신이 우리에게 이야기한 그 사람을 만날것이다.)

우의 실례 ⑧a에서는 동사 told에 부사적어구인 yesterday가 부착되었으며 ⑧b에서는

tomorrow가 반대로 will meet에 부착된다는것이 명백하다. tomorrow의 미래시칭이 will meet의 미래시칭과 일치하며 yesterday의 과거시칭이 told의 과거시칭과 일치하다는것은 언어학적으로 본 분석이다. 그러나 현시점에서 과거에 무엇인가를 하려고 계획을 짤수 없고 또 앞으로 무엇인가를 이미 말해버릴수 없다는것이 현실적인 사실인것이다. 다시말하여 언어학적인 지식은 흔히 현실세계에 대한 지식과 밀접히 련관되어있다. 이것을 다른 실례 ⑨로도 설명할수 있다.

례: ⑨ The man saw the girl with red hair.

(그 사람은 빨간 머리카락을 가진 소녀를 보았다.)

우의 실례에서는 사람들이 머리카락은 무엇인가를 보는데 리용될수 없다는것을 알기때문에 모호성이 제거되게 된다. 언어적으로 보면 hair는 동사 see에 대하여 허용되는 단어가 아니다. 그것은 《보다》의 의미를 가지는 동사 see에 해당되는 가능한 단어들에 대하여 의미론적인 특징부호화를 하여 알수 있다. 따라서 언어학적인 지식과 현실세계에 대한 지식은 넓은 범위에서 일치한다고 볼수 있다.

그러나 현재 기계번역체계개발에서는 문맥과 적용분야를 상대적으로 좁게 제한시킨다고 하여도 가능한 모든 모호성들을 해소하는데 리용할수 있는 현실세계에 대한 지식을 부호화하고 체계와 결합시키는것이 많은 시간과 노력이 드는 작업인것으로 하여 원만한 수준에 이르지 못하고있다.

인공지능과 컴퓨터기술의 발전에도 불구하고 이러한 문제는 단기간내에 해결되기 힘들며 따라서 현실세계에 대한 지식의 복잡성과 그 리용의 어려움은 모호성해소의 기본장애로 된다.

우리는 앞으로 영조기계번역체계의 성능을 높여나가는데서 제기되는 실천적문제들을 적극 해명해나감으로써 과학기술강국건설에 적극 이바지해나가야 할것이다.