

언어과학연구 27 (2003)

한영번역기에서의 주어 복원 문제에 대하여*

최 정 혜(고려대학교)

Choi, Jung-Hye. 2003. *Subject Recovery in Korean-English Machine Translator*. *The Journal of Linguistic Science* 27, 385-406. The primary purpose of this paper is to examine the patterns that some subjects are omitted in Korean-English machine translation, and to suggest a way of subject recovery. The strategy of subject recovery could be roughly divided whether the sentence is simple or not. In the substrategies of subject recovery, both the lexical items and the grammatical items are used; psychological verbs, suffixes, endings, etc. And most of these items are represented by virtue of features which are so flexible and easily addable in FCG(Feature Computation Grammar). Together with these, I will compare 'TransMaster' (Korean- English machine translator made by Machine Translation Lab in Korea Univ. and Prolangs Co., Ltd.) with another Korean- English machine translators('Han guide 2.0', 'Smartran'). (Korea University)

Key words 한영기계번역(Korean-English machine translation), 주어 복원(subject recovery), 한영번역기(Korean-English machine translator)

*이 논문은 언어과학회 2003년 하계학술발표회에서 발표한 글을 기초로 다듬은 글이다. 논문 심사 과정에서 도움말을 주신 익명의 심사위원 세 분께 감사를 드립니다.

이 논문은 2002년도 BK21 고려대학교 한국학 교육·연구단 연구지원비에 의하여 연구되었음.

1. 머리말

한국어와 영어는 서로 다른 언어 조상을 가지고 있는 언어이다. 따라서 번역을 할 때 라틴어를 기반으로 하는 언어들 사이에서 비교적 단순하게 대응시킬 수 있는 형태소나 단어, 구 따위의 대응 관계가 한국어와 영어 사이에서는 어려운 경우가 많다. 그 중 하나가 주어와 관련된 것이다.

(0) 가. 어디 가세요?

가'. Where are you going?

나. 많이 바쁘구나.

나'. You are so busy.

위의 예문에서도 알 수 있듯이, 한국어 문장에서 주어는 빈번히 생략되지만, 영어 문장에서는 일반적으로 주어를 명기한다.¹⁾ 따라서 한국어 문장을 영어 문장으로 번역하고자 할 때, 주어를 복원하는 문제가 필연적으로 제기된다.

본고는 한국어를 영어로 번역할 때 한국어 문장에서는 나타나지 않는 주어가 영어로 번역된 문장에서 어떤 형태로 나타나야 할 지를 생각해 보고, 한국어와 영어의 어떤 정보들을 통해 영어 번역문에서 주어를 복원해 낼 수 있는지에 대해 알아 보고자 한다. 그리고 한영번역기를 설계할 때 이런 복원 전략들을 어떻게 활용할 수 있을지 검토해 보고, 이를 바탕으로 만든 한영번역기를 통해 그 타당성을 검증해 보고자 한다. 더불어 본고를 바탕으로 만든 한영번역기의 성능을 좀더 객관적으로 검증하기 위해 기존의 다른 한영번역기들에서 주어가 복원되는 양상도 함께 살펴볼 것이다.

2. 주어 복원의 유형

2.1 단문에서의 주어 복원

1) 물론 구어체 영어에서는 주어가 생략되는 현상이 빈번하기는 하지만, 일반적으로 활자화되어 나타나는 문장을 비교해 볼 때, 한국어 문장이 영어 문장에 비해 주어가 생략되는 경우가 많다는 것은 확연한 사실이다.

한국어 문장에서 주어가 필수 요소인지에 대한 의문은 북한 문법에서 무주어문을 상정하고 있는 것을 보아도 잘 알 수 있다. 이것은 달리 말하면, 문장 안에서 그 문장의 주어를 예측할 수 있기 때문에 주어가 굳이 필수적으로 나타나지 않아도 된다는 뜻으로도 이해할 수 있다. 그만큼 주어라는 문장 성분은 문장 안에서 어느 정도 예측 가능한 요소라는 것을 말해 준다.²⁾

2.1.1 문형 정보를 이용한 주어 복원

우선, 문장에서 생략되는 주어의 유형을 살펴 보고자 한다. 첫 번째로 문장 유형에 따른 주어의 출현 양상을 살펴 보자. 국어의 문장 유형은 평서문, 의문문, 청유문, 감탄문, 명령문 5가지로 나눌 수 있고 이에 따른 예문은 다음과 같다.

- (1) 평서문 : 비가 옵니다. / 학교에 갑니다.
- (2) 의문문 : 비쌌니까? / 많이 늦었습니까?
- (3) 청유문 : 도서관에 갑시다.
- (4) 감탄문 : 얼마나 아름다운가! / 얼마나 높은지!
- (5) 명령문 : 나가세요!

모두 8개의 문장 중에서 주어가 있는 경우는 (1)의 첫 번째 문장뿐이지만 모두 자연스러운 한국어 문장임에는 분명하다. 이들 문장을 영어로 번역하면 대략 다음과 같다.

- (1) ' 평서문 : It rains. / I go to school.
- (2) ' 의문문 : Is it expensive? / Were you late a lot?
- (3) ' 청유문 : Let's go to the library.
- (4) ' 감탄문 : How beautiful she is! / How high it is!
- (5) ' 명령문 : Go out!

2) '어느 정도'라고 기술한 것은 문장 하나만을 보고 생략된 주어를 알 수 없는 경우가 있기 때문이다. 이 때에는 문장 단위 이상의 정보 즉, 문맥 정보가 필요하게 된다. 문맥 정보를 이용할 경우에도 기준 문장을 중심으로 어디까지를 참조할 것인지의 문제가 대두된다. 참조 문장의 범위에 관한 문제는 또 다른 연구의 차원이므로 이에 대해서는 본고에서 다루지 않기로 한다.

한국어의 경우와는 달리 영어로 번역했을 때에는 청유문과 명령문을 제외하고는 모두 주어 가 있는 문장이 된다. 따라서 평서문과 의문문, 감탄문에 대해서는 주어 복원의 전략이 필요하다. 그런데, 감탄문의 경우에는 평서문으로 그 문형만을 바꾸어 보면 결과적으로 영어 문장으로 번역했을 때 주어 가 같다는 것을 알 수 있다.

- (4) 감탄문 : 얼마나 아름다운가! / 얼마나 높은지!
- (4) ' 감탄문 : How beautiful she is! / How high it is!
- (6) 평서문 : (정말) 아름답습니다. / (무척) 높습니다.
- (6) ' 평서문 : She is (really) beautiful. / It is (very) high.

따라서 감탄문에 대한 주어 복원의 전략은 평서문의 주어 복원 전략에 통합하여 논의하기로 한다.

그렇다면 주어 복원에 있어 문형별 전략은 평서문과 의문문에 대한 전략 두 가지의 경우로 압축된다. 먼저 의문문의 경우, 문형의 특성상 의문은 상대에 대한 것일 수밖에 없다. 문제는 그 상대가 사람인지 아닌지가 다를 뿐이다. 따라서 (2) '의 예문에서 본 것처럼, 상대가 사람일 경우에는 주어를 you로, 사람이 아닐 경우에는 it으로 복원한다고 가정해 볼 수 있다.

(가) 의문문의 주어 복원 전략

의문의 상대에 따라 즉, 사람에 대한 것인지 아닌지에 따라 주어를 you 또는 it으로 복원할 수 있다.

가정 (가)가 옳은 것인지를 알아 보기 위해 몇몇 다른 문장들을 살펴 보자.

(1), (3), (4), (5)의 예문을 모두 의문문으로 바꾸어서 번역해 보면 다음과 같다.

- (1) "비가 옵니까? / 학교에 갑니까? → Does it rain? / Do you go to school?"
- (3) "도서관에 갑니까? Do you go to the library?"
- (4) "아름답습니까? / 높습니까? → Is she beautiful? / Is it high?"
- (5) "나갑니까? Do you go out?"

상대가 사람이 아닐 경우에 it으로 복원하는 전략은 위의 예문들에서도 옳은

것 같다. 다만, 상대가 사람일 경우에는 (4)의 첫 번째 문장에서는 you가 아닌 she로 복원되는 것이 좀더 자연스러워 보이기도 한다. 이것은 '아름답다'라는 어휘 특성에 의한 것으로 볼 수 있는데, 사람에 대해 '아름답다'라고 말할 때는 대개 여성에 대해 말하는 것으로 여겨지기 때문이다. 마찬가지로, '용감합니까?'라는 문장에 대한 영어 번역문은 'Is he brave?'가 될 것이데, 역시 '용감하다'라는 술어의 특성으로 인해 주어가 'he'로 복원될 수 있다. 이러한 예문들을 통해 의문문의 주어 복원 전략은 다음과 같이 수정될 수 있다.

(가) '의문문의 주어 복원 전략

의문의 상대에 따라 즉, 사람이 아닐 때에는 it으로, 사람일 때에는 술어에 따라 you 또는 she나 he와 같은 인칭대명사로 복원할 수 있다.

한편 영어로 번역되는 문장에서 선택된 단어에 따라 주어가 달리 선택되어야 하는 경우도 있다. 즉,

- (7) 가. 당신의 사무실에 들러도 될까요? → May I drop in your office?
 나. 이제 식사를 할까요? → Shall we eat now?

(7)은 모두 의문문이지만, (7가)에서는 번역된 문장에서 조동사가 'May'로 선택될 때 주어를 I로 복원하고, (7나)에서는 조동사가 'Shall'로 선택될 때 주어를 we로 복원한다는 전략을 세울 수 있다. 이러한 경우는 문형 정보와 더불어 번역되는 문장에서의 영어 단어 정보를 같이 활용해야 하는 경우라 할 수 있다.

평서문의 주어 복원 전략의 방향도 의문문과 크게 다르지 않은 듯하다. 의문문에서와 마찬가지로 (2), (3), (4), (5)의 문장들을 평서문으로 바꾸어 번역해 보면 다음과 같다.

- (2)'' 비쌉니다. / 많이 늦었습니다. → It is expensive. / I was late a lot.
 (3)'' 도서관에 갑니다. I go to the library.
 (4)'' 아름답습니다. / 높습니다. → She is beautiful. / It is high.
 (5)'' 나갑니다. I go out.

의문문에서와 마찬가지로 (4)의 첫 번째 문장처럼 ‘아름답다’와 같이 여성성을 가진 술어가 올 경우에는 주어 that로 복원되는 것이 자연스럽게 나머지의 경우 인칭대명사로 복원되는 경우는 모두 I로 복원되는 것이 자연스럽게. 또한 주어 that가 사람이 아닐 경우에는 의문문의 경우와 마찬가지로 it으로 복원할 수 있다. 따라서 다음과 같은 평서문의 주어 복원 전략에 대한 가정을 세울 수 있다.

(나) 평서문의 주어 복원 전략

서술의 상대에 따라 즉, 사람이 아닐 때에는 it으로 사람이 때에는 술어에 따라 I 또는 she나 he와 같은 인칭대명사로 복원할 수 있다.

문형 정보를 기반으로 할 때 주어 복원은 결국 평서문과 의문문의 문제로 집약되고, 이제 (가)와 (나)에서 문제가 되는 것은 ‘술어에 따라’라는 표현인데, 이것은 문장 유형에서 해결할 수 있는 범위에 있지 않으므로 더 작은 단위의 문제 즉 단어의 문제로 넘어 가게 된다.

2.1.2 단어 정보를 이용한 주어 복원

2.1.1의 예문 (4) ‘얼마나 아름다운가!’나 ‘얼마나 높은지!’와 같은 경우에는 문형 정보 외에 술어에 대한 정보가 필요하게 된다. 즉, 한국어 문장의 술어 안에 번역될 영어 문장의 주어를 she로 복원할 수 있는 정보가 있어야 하고, 이는 it으로 복원되는 경우에도 역시 마찬가지이다. 우선, 영어 번역문에서 주어 that가 it으로 되어야 하는 문장들을 생각해 보자. 즉, 영어문장에서 주어를 it으로 복원해야 하는 한국어 술어들을 범주화해 주어야 하는데, 생각해 볼 수 있는 것은 대략 다음과 같다.

(8) [날씨류]³⁾ : 화창합니다. → It is bright.

(9) [거리류] : 너무 멉니다. → It is too far.

(10) [시간류] : 3시간 걸립니다. → It takes 3 hours.

3) 의미 속성을 나타내는 전통적인 방식을 따라, 이 글에서도 대괄호 []를 사용하여 의미속성을 나타내도록 하겠다. []는 이후에 공통 자질을 할당하는 경우에도 동일하게 사용하게 된다. 참고로 본 논문에서 소개할 번역기는 총 66개의 술어 관련 자질을 설정하고 있다.

- (11) [척도류] : 깊습니다. → It is deep.
 (12) [통화류] : 싸다. → It is cheap.

한영번역기를 설계할 때 이런 특징은 결국 영어나 한국어에서 개별 단어들의 의미정보로 할당된다. 따라서 [날씨류], [거리류], [시간류], [척도류], [통화류]를 의미속성으로 가지고 있는 술어들에 공통의 자질을 할당하여 한국어 문장에서 주어가 없는 경우, 영어 문장에서는 it으로 복원하도록 해야 한다.⁴⁾

한편, 생략된 주어를 she로 복원하는 경우는 상당히 제한적이다. 앞에서도 언급했듯이, '아름답다'와 같이 다소 명확하게 여성(혹은 '여성성'을 가진 대상)을 기술하는 술어들에만 해당된다. 이런 술어들을 들어보면 다음과 같다.

- (13) [여성류] : 아름답다, 예쁘다, 가냘프다, 다정다감하다, ...

she로 복원되어야 하는 경우를 제외하고, 3인칭 주어로 복원되어야 하는 경우에는 he로 복원하도록 한다. she의 경우와 마찬가지로, '용감하다, 씩씩하다' 따위는 소위 [남성류]라고 할 수 있겠는데, 이런 부류의 술어들에 대해 특별히 주어를 he로 복원하는 전략을 세우고, 나머지를 she로 복원하는 것이나, 반대로 [여성류]의 술어들에 대해 따로 주어를 she로 복원하는 전략을 세우고 나머지를 he로 복원하는 것은 서로 상보적인 분포를 보이므로 이 글에서는 후자의 방법을 취하도록 한다.

한편 술어 외의 다른 단어로 인해 주어가 구별되어야 하는 경우가 있다.

- (14) 가. 그러더니 나가 버렸다. → Then he went out.
 나. 방긋방긋 웃었다. → The baby(He) smiled.

4) 본고의 논의에서 사용되는 자질은 자질 연산 문법의 자질 개념에 기초하고 있다. 자질 연산 문법(Feature Computational Grammar: FCG)은 어휘부에 등재된 표제어들의 자질과 이러한 자질 간의 연산을 핵심적인 개념으로 하여 운용되는 문법이다. FCG는 실용적인 시스템의 이론적 근거가 되어 자료체의 정보를 분석하고 그 결과로 특정한 결과물을 도출해 내는 것을 목표로 한다. FCG에서 자질은 '특정한 어휘 항목의 외연을 표시하기 위한 내항'으로 정의된다. 원리적으로 자질 체계는 개방되어 있으므로 다양한 어휘 항목과 적절한 자질들을 관련지어 주며 문법의 운용을 유연하게 해 준다.(김원경 2000: 19, 50, 52) 자질 연산 문법에 대한 자세한 논의는 고창수(2002), 김원경(2000)을 참고하기 바란다.

(14가)에서는 접속사인 ‘그러더니’를 통해 주어 복원하는 전략이 필요하다. ‘그러더니’는 1인칭이나 2인칭 주어와는 어울리지 않는다. 따라서 이 경우 he(또는 she)로 번역하는 것이 자연스럽다. 다른 예로, (14나)와 같이 부사에 의해 주어가 선택되어야 하는 경우가 있다. ‘방긋방긋’이라는 부사는 대개의 경우 아기가 웃는 모습을 표현할 때 사용한다. 따라서 구체적으로 baby라는 명사를 주어로 복원하든지 또는 무표적인 3인칭 단수인 he로 표현하는 전략이 필요하다.⁵⁾

2.1.3 형태소 정보를 이용한 주어 복원

주어를 복원해 내는 데 필요한 정보를 얻는 단위를 단어보다 더 작은 단위인 형태소에서 찾아야 하는 경우가 있다. 이 때에도 단어에서와 마찬가지로 대개의 경우, 술어를 이루고 있는 형태소를 통해 주어 복원의 실마리를 찾게 된다. 그리고 서술어를 구성하는 어간과 어미 중에서는 어미를 이루고 있는 일부 형태소를 통해 주어 복원의 정보를 얻을 수 있다. 한편, 계사 구문에서 서술어의 명사 자질을 보고 주어를 달리 복원해야 하는 경우도 있다.

먼저, 서술어의 어미 정보를 보고 주어를 복원하는 경우를 살펴 보자. 생략된 주어를 you로 복원해야 하는 경우가 있는데, 예문을 들어 보이면 다음과 같다.

(15) 한국말을 잘 하는군요. → You speak Korean well.

(15)는 어미 ‘-군요’를 보고 주어를 you로 복원하는 전략을 세울 수 있다.⁶⁾ 또한 존칭형태소 ‘-시-’도 주어 복원 전략의 지침이 되는데, 다음의 문장들을 살펴 보면,

-
- 5) ‘방긋방긋 웃었다’에 대한 번역문으로 ‘The baby(He) smiled’를 만드는 데에는 부사 ‘방긋방긋’을 보고 술어를 ‘smile’로 선택하고 부사는 영어 대역어를 할당하지 않는 절차가 따로 필요하다.
- 6) 번역기에서 주어복원을 위해 참조하는 정보 중 하나인 형태소 정보는 순수 국어학에서 분석하는 형태소 정보와는 다소 차이가 있다. 예문(15)의 경우처럼 ‘-군요’는 엄격하게 따지자면 ‘-군-’과 ‘-요’로 분석할 수 있겠지만, 영한번역기에서 주어복원을 위해 필요한 형태소는 이러한 낱말의 형태소가 아닌 ‘-군요’라는 통합된 형태소 단위이다. 따라서 기계번역에서 형태소는 순수 국어학에서의 ‘최소 유의미 단위’인 형태소의 개념보다는 전자 사전의 표제어 혹은 등재소(listeme)의 개념에 가깝다고 볼 수 있다.(황화상, 2001. 참고)

(16) 가. 회사에 갑니다. → I go to the company.

나. 회사에 가십니다. → He goes to the company.

와 같이 (16가)는 평서문에서 무표적으로 주어 I를 복원해 내면 되지만, (16나)는 주어를 he(또는 she)로 복원해야 자연스럽다. 따라서 술어에 '-시-'가 포함되어 있다면 주어를 he(또는 she)로 복원하는 전략을 세울 수 있다.⁷⁾

다음으로 주어가 없는 기사문의 경우를 살펴 보자. 기사문에서 주어는 It과 This의 두 가지로 복원할 수 있다.

(17) 가. 우리 어머니입니다. → This is my mother.

나. 금강산입니다. → This is Kungang Mountains.

다. 좋은 책입니다. → It is a good book.

사람이나 사물을 소개하는 경우에는 This로 주어를 복원하고, 그 외에는 It으로 복원하는 것이 자연스러워 보인다. 즉, 서술어가 사람이거나 고유명사이면 This로, 그 외의 경우는 It으로 주어를 복원하는 전략을 세울 수 있다.

2.1.4 구 정보를 이용한 주어 복원

형태소나 단어 이상의 표현들이 주어 복원의 정보를 제공하기도 한다. 다음의 문장들을 살펴 보자.

(18) 가. 아픈 것 같습니다. → You seem to be ill.

나. 공공장소에서는 담배를 피우면 안 됩니다. → You must not smoke in a public-space.

(19) 창문을 닫아 드릴까요? → Shall I close the window?

(18)은 평서문에서 주어를 you로 복원해야 하는 예들이다. (18가)의 '-ㄴ 것 같

7) 물론 존칭형태소 '-시'가 있는 모든 문장의 주어를 he(또는 she)로 복원할 수 있는 것은 아니다. 즉, '괜찮으십니까?'와 같은 문장은 주어를 you로 복원하는 것이 자연스럽다. 이러한 문장은 2.1의 문형 정보를 이용한 주어 복원 단계에서 처리가 되므로 형태소 정보를 이용하는 단계의 적용을 받지 않는다. 처리 수순과 관련된 문제는 3장에서 다시 논의된다.

습니다'라는 표현과 (18나)의 '-면 안 됩니다'의 표현이 있는 문장에서 주어가 없을 때, 그 주어를 you로 복원하는 것이 자연스럽다. 이것은 번역된 결과치인 영어 문장을 고려한 주어 복원 전략이라고 할 수 있다. '-ㄴ 것 같습니다'와 '-면 안 됩니다'는 영어에서 seem to와 shall이라는 조동사로 각각 대응된다. 즉, 이런 부류의 조동사가 있는 문장에서 주어가 없을 때에는 you로 복원하는 전략이 가능하다. (19)는 의문문이지만 생략된 주어가 you로 복원되지 않고 I로 복원된다. 이것은 '-아/어 드릴까요'라는 표현에 기인한 것이다.

이처럼 단어 이상의 특정 표현에 대해 한 문장의 주어가 결정되는 경우들이 있다.

2.2 복문에서의 주어 복원

복문에서 주어를 복원하는 절차를 마련하는 것은 단문에서 주어를 복원하는 절차를 마련하는 것과는 그 양상이 사뭇 다르다. 이는 단문에서 예측할 수 있는 경우의 수가 복문으로 확장되면서 상당히 복잡다단한 양상을 띠기 때문이다. 따라서 복문의 종류에 따라 단문에서의 복원 전략을 적용할 수 있는 경우도 있지만 일정 부분은 단문의 전략을 그대로 수용하기 어려운 경우도 있다.⁸⁾

본고에서 규정하는 복문은 서술어가 2개 이상 포함된 문장을 의미한다. 따라서 다음과 같이 본동사가 2개 나타나는 문장이나 '하다'로 이어지는 일부 문장, 서술어인 수식어를 가진 문장도 모두 복문으로 간주한다.

- (20) 가. 학교에 가서 공부를 한다.
 나. 어머니는 아이에게 우유를 먹이게 했다.
 다. 좋은 책입니다.

2.2.1 문형 정보를 이용한 주어 복원

본 소절에서는 주어가 없는 종속절을 가지고 있는 명령문, 청유문, 의문문의 경우를 살펴 보고자 한다. 우선, 명령문은 직접적인 상대를 가정하는 경우가 일

8) 문장이 길어지면서 생겨나는 변수는 단지 주어 복원의 문제에만 국한되는 것은 아니다. 대표적인 문제로 조용(anaphora)을 들 수 있다.

상적이며 또 그럴 경우에 주어는 담화 상황에 존재하는 사람이 경우가 일반적이므로 주절이 명령문일 경우, 종속절의 주어는 you로 복원할 수 있다.

(21) 모르는 것은 나한테 물어세요. → Ask me the thing that you don't know.

문형의 특성상 청유문일 경우에는 말하는 이를 포함한 '우리'가 서술의 주체가 되므로, 주절이 청유문일 경우, 종속절의 주어는 we로 복원할 수 있다.

(22) 지난 번에 갔던 다방으로 갑시다. → Let's go to the tearoom that we went to the last time.

한편 주절이 의문문이고 종속절의 주어는 없는 경우에는, 종속절의 동사 유형에 따라 주어 복원 전략이 두 가지로 달라진다. 즉, 종속절의 동사가 타동사인 경우에는 주어를 복원하게 되고 이 경우 복원되는 주어는 you가 된다. 그러나 종속절의 동사가 자동사인 경우에는 주어를 복원하지 않는 전략을 취한다.

(23) 가. 어제 만난 사람이 누구입니까? → Who is the man who you met yesterday?

나. 저기 가는 학생이 누구입니까? → Who is the student that ___ goes over?

2.2.2 단어 정보를 이용한 주어 복원

단어 정보를 이용해서 생략된 주어를 복원하고자 하는 경우에는 입력문인 한국어 문장뿐만 아니라, 출력문이 될 영어 문장에서의 단어도 참고를 해야 한다. 영어 문장의 단어를 참고해야 하는 이유는 역시 번역이라는 작업이 항상 1대1 대응이 되는 것은 아니라는 속성에서 비롯한다. 예를 들어, 한국어 '~어서'로 연결된 복문은 영어로 번역될 때, 때로는 so로, 때로는 because로, 또 때로는 to로도 대응될 수 있기 때문이다. 얼핏 생각해 보아도 so나 because로 번역될 때에는 주어가 있어야 하겠지만, to로 번역될 때에는 주어가 복원되지 않아야 할 것이다. 이러한 점을 감안한다면 영어 문장의 단어 정보도 주어를 복원하는 데 필요한 정보임에 틀림이 없다.

먼저, 한국어 문장에서 참고할 수 있는 단어 정보로는 주절의 서술어를 들 수 있다. 주절의 서술어가 [말하다]류이고 주절의 주어는 생략되어 있을 때 그 주어

는 he또는 she로 복원될 수 있다.

(24) 서울에 친척이 있다고 그랬어요. → He said that there is a relative in Seoul.

한편, 주절의 서술어가 [말하다]류이고, 종속절의 주어가 생략되어 있을 경우에는 주절의 주어를 복사해서 종속절의 주어로 복원할 수 있다.

(25) 철수는 사과를 먹었다고 말했어요. → Cheol-su said that Cheol-su ate an apple.

많은 경우 그러하지만, 두 가지 이상의 정보가 더하여서 복원될 주어를 결정하게 되는 경우가 있다. 예를 들어, 주절의 서술어가 [주다]류일 때 그리고 의문문인 경우, 그 주어는 I로 복원된다.

(26) 김치 만드는 방법을 가르쳐 드릴까요? → Shall I teach the method that I make Kimchee?

다음으로, 영어 문장에서 참조할 수 있는 단어 정보로는 접속사를 들 수 있다. 영어 접속사의 경우, and와 but으로 연결된 문장은 어느 한 쪽 절에 주어와 있다 면 이 주어를 복사해 사용할 수 있다. 이는 and와 but으로 이어지는 절이 내용에서 서로 대등하기 때문이다.

(27) 가. 그 사람은 멋있고 친절합니다. → He is wonderful and he is kind.

나. 그는 성공하지 못했지만 열심히 일했다. → He did not succeed, but he worked hard.

If나 unless로 이어지는 문장은 대개의 경우 주어를 you로 복원할 수 있다.

(28) 물건을 싸게 사려면 큰 시장에 가야 합니다. → If you want to buy things cheaply, you have to go to the big market.

2.2.3 형태소 정보를 이용한 주어 복원

형태소 정보를 이용하는 대부분의 절차는 단문에서 이루어지는데, 복문의 경우에 특정 형태소 정보가 명시되어야 하는 경우가 있다. 예를 들어 다음 문장에서

(29) 또 한번 거짓말을 하는 날에 용서하지 않겠어. → If you tell a lie again, I will not forgive you.

단어 정보를 참고하자면 접속사가 if인 복문이므로 생략된 주어는 모두 you로 복원될 것으로 기대되지만, 주절의 주어는 I로 복원되었다. 이것은 주절의 서술어에 포함되어 있는 형태소 '-겠-' 때문이다. 즉, 주체의 의지를 나타내는 '-겠-'으로 인해 주어는 you가 아닌 I로 복원된다.

3. 주어 복원의 순서

2장에서는 문형, 단어, 형태소, 구 정보에 의한 주어 복원의 경우들을 살펴 보았다. 이들을 각각 따로 떼어서 생각해 보면 기실 그렇게 어려운 문제도 아닌 듯하다. 한국어 화자로서의 직관만 제대로 있다면 특정 단어나 형태소, 구 정보를 이용해 생략되어 있는 주어를 예측해 내는 것은 그리 특별한 것이 아닐지도 모른다.

문제는 이들 정보의 순서를 매겨야 한다는 데 있다. 컴퓨터가 이해할 수 있는 스크립트를 짜자면 이분법적이 방법론을 취할 수밖에 없다. 따라서 조건이 주어지면 그 조건에 대한 판단이 yes 혹은 no의 둘 중 하나로 귀결되어야 하고, 조건과 판단에 따른 선적인 흐름을 만들어야 한다.

입력 문장인 한국어로부터 출력 문장인 영어를 만들어 내는 한영번역기는 번역기⁹⁾

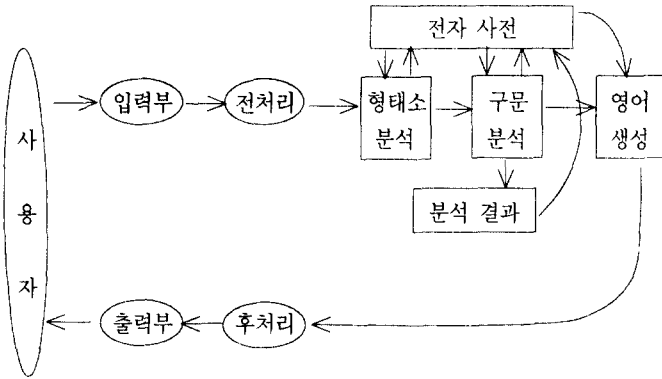
9) 기존의 기계번역 방법론은 대략 다음의 4가지로 나눌 수 있다: 규칙기반, 패턴기반, 예제기반, 확률기반. 본고에서 채택하고 있는 주요 방법론은 규칙기반이다. 각각의 방법에 대한 장단점을 간략히 제시하면 다음과 같다. (최승권 2003:95-113 참조)

- 규칙기반 : 장점 : 1. 초기 체계를 형성하는 데 용이하다
2. 언어학적 이론에 기반하고 있다
3. 주요 현상을 설명하는 데 효과적이다

단점 : 1. 규칙들이 전문가(언어학자, 프로그래머)에 의해 만들어진다.

에 따라 어느 정도 차이가 있겠지만 대략 다음과 같은 절차를 거친다고 볼 수 있다.

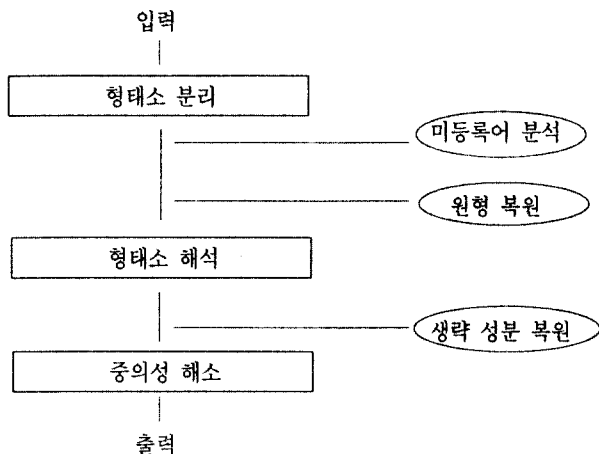
< 표 1 > 한영번역기의 구조



그리고 생략된 성분을 복원해 내는 절차는 형태소 분석의 단계에서 이루어지는데, 이 단계도 개별 한영번역기 내의 형태소 분석기에 따라 다소 차이가 있을 것지만, 대략 다음과 같은 과정들을 포함하고 있다.

- 2. 규칙을 유지하고 확장하는 데 어려움이 있다
- 3. 부수적인 현상들을 설명하는데 효과적이지 않다.
- 패턴기반 : 장점 : 1. 번역 패턴에 기반하고 있다.
2. 패턴 규모가 증가하면서 번역의 질을 점진적으로 높일 수 있다.
단점 : 1. 대규모의 번역 패턴을 구축해야 하는 어려움이 있다.
2. 번역 속도가 저하된다.
- 예제기반 : 장점 : 1. 말뭉치로부터 지식을 추출할 수 있다.
2. 말뭉치에 있는 번역 패턴에 기반한다.
3. 사람이 조작해야 하는 수고로움을 덜어준다.
단점 : 1. 언어의 생산성을 지원하기 어렵다.
2. 대용량의 말뭉치가 필요하다.
3. 유사성 계산을 지원할 훌륭한 시소러스가 필요하다.
- 확률기반 : 장점 : 1. 별도의 규칙 없이 번역 모델을 구현할 수 있다.
2. 사람이 조작해야 하는 수고로움을 덜어준다.
단점 : 1. 개별적인 언어 현상을 기술하기 어렵다.
2. 대용량의 정렬된 대역 말뭉치를 구축하기 어렵다.
3. 문장이 길어질 경우 의존적인 언어 현상을 다루지 못한다.

< 표 2 > 형태소 분석기의 구조



형태소 분석의 단계에서 생략된 성분들을 추론하고 이에 따른 복원 전략을 마련하는 이유는 본고에서 주어 복원의 정보로 활용하는 문형이나, 구, 단어도 모두 그 기본 단위는 형태소에서 비롯되기 때문이다. 그리고 이러한 형태소 정보들은 그 유형에 따라 일정한 자질로 정의된다.

본 장에서는 단문을 대상으로 하여 주어를 복원하는 스크립트를 구상해 보고자 한다.¹⁰⁾ 다음 예문들을 살펴 보자.

- (30) 가. 아픈 것 같습니다. → You seem to be ill.
- 나. 비싼 것 같습니다. → It seems to be expensive.
- 다. 아프신 것 같습니다. → He seems to be ill.

영어로 번역했을 때 모두 'seem to'라는 표현으로 번역되지만, (30가)는 '-어 보인다'와 '아프다'라는 표현을 통해 you로, (30나)는 '비싸다'라는 표현을 통해 it으로, (30다)는 '-시-'라는 형태소 표현을 통해 주어를 he로 복원한다. 이 세 문장만을 두고 볼 때는 먼저 '-시-'를 보고 he로 복원하고, 다음으로 '비싸다'를 보고

10) 모든 문장을 포괄할 수 있는 주어 복원 스크립트는 단문과 복문을 아우르는 것이어야 한다. 그러나 복문을 포함하는 주어 복원 과정은 구문 연산의 일부 절차와 연계해서 고려해야 하는 부분이 있는 관계로, 본고에서는 단문만을 대상으로 한 주어 복원 과정을 보이고자 한다.

it으로 복원하고, 그런 후에 이들 조건에 걸리지 않는 경우에 you로 복원하는 따위의 간단한 스크립트를 짤 수 있을 것이다.

문제는 문형과, 단어, 형태소, 구 표현의 복합적인 정보를 계기적으로 순서지어 주어야 한다는 데 있다. 지금까지 생략된 주어를 복원하는 데 이용된 정보를 나열해 보면 다음과 같다.

- (31) 문형 정보 1 (평서문) → I 복원
 문형 정보 2 (의문문) → You 복원
 단어 정보 1 ([날씨/거리/시간/척도/통화]) → It 복원
 단어 정보 2 ([여성성]) → She 복원
 단어 정보 3 ([남성성]) → He 복원
 단어 정보 4 (May) → I 복원
 단어 정보 5 (Shall) → We 복원
 형태소 정보 1 (-시-) → He 복원
 형태소 정보 2 (-군요) → You 복원
 형태소 정보 3 (-이다) → This/It 복원
 구 정보 1 ([당위성]) → You 복원
 구 정보 2 ([주다류]) → I 복원

이제 이들을 순서 지어 주어야 하는 수순만이 남았다. 먼저 스크립트를 짜는데 처음 시작점이 될 기준이 필요하다. 문형정보를 기준으로 스크립트(1)를 시작해 보자.

문형정보가 맨 처음의 기준점이 될 수 있는지, 즉 문형정보보다 앞서는 조건은 없는지를 살펴보자. 문장유형이 명령문, 청유문이면 주어를 복원시킬 필요가 없는데, it으로 복원해야 하는 [날씨류/거리류/시간류/척도류]는 명령문과 청유문의 문장으로 쓰이는 것이 불가능하다.

- (32) 가. *화창해라, *화창하자
 나. *멀어라, *멀자
 다. *늦어라, *늦자
 라. *높아라, *높자

11) 기계번역에서 사용하는 스크립트의 용어를 그대로 사용할 필요는 없으므로, 스크립트에서 사용하는 다소 전문적인 용어는 피하고 되도록 풀어서 기술하도록 한다.

따라서 주어가 it으로 복원되는 경우가 스크립트 상에서 문형 정보의 앞에 오거나 혹은 다음 수순이 되어도 상관이 없다. 그러나 다른 조건들, 즉, 주어가 she나 he나 you로 복원되어야 하는 수순들은 문형 정보의 수순을 앞지를 수 없다. 특히, 형태소 정보 '-시-'에 의한 주어 he의 복원은 단어에 의한 she복원보다는 다음 단계에 와야 한다.

(33) 가. 아름다우십니다. → She is beautiful.

나. 바쁘십니다. → He is busy.

다. 용감하십니다. → He is brave.

이러한 주어 복원의 정보를 바탕으로 단문에서의 주어 복원 스크립트를 짜 보면 다음과 같은 순서가 된다.

(34) 단문에서의 주어 복원 스크립트

시작 : [문장유형 1]

문장 유형이 명령문이거나 청유문인가?

yes : 주어를 복원하지 말라.

1단계 : [단어/형태소/구 정보]

no : 문장의 서술어가 [날씨류]이거나 [거리류]이거나 [시간류]이거나 [척도류]이거나 [통화류]를 가지고 있는가?

yes : 주어를 it으로 복원하라.

no : 문장의 서술어가 [여성성]을 가지고 있는가?

yes : 주어를 she로 복원하라.

no : 문장의 서술어 형태소가 '-시-'를 가지고 있는가?

yes : 주어를 he로 복원하라.

no : 문장의 서술어가 [남성성]을 가지고 있는가?

yes : 주어를 he로 복원하라.

no : 문장의 서술어가 계사이고 [사람]이나 [고유명사]를 가지고 있는가?

yes : 주어를 this로 복원하라.

no : 문장의 서술어가 계사인가?

yes : 주어를 it으로 복원하라.

no : 문장의 서술어가 [주다류]를 가지고 있는가?

yes : 주어를 I로 복원하라.

no : 문장의 서술어 형태소가 '-군요'를 가지고 있는가?

yes : 주어를 you로 복원하라.

no : 문장의 서술구가 [당위성]을 가지고 있는가?

yes : 주어를 you로 복원하라.

2단계 : [문장유형 2]

no : 문장 유형이 의문문인가?

yes : 문장의 조동사가 'May'인가?

yes : 주어를 I로 복원하라.

no : 문장의 조동사가 'Shall'인가?

yes : 주어를 we로 복원하라.

no : 주어를 you로 복원하라.

no : 주어를 I로 복원하라.(문장 유형이 평서문인 경우)

위의 스크립트를 바탕으로 다음 20문장을 번역해 보았다. 사용한 번역기는 "TransMaster"로 고려대학교 기계번역연구실과 (주)프로랭스가 같이 만들었던 번역기이다. 그리고 TransMaster와 비교하기 위해, 같은 문장을 기존에 나와 있는 두 종류의 번역기를 통해 번역한 결과를 같이 덧붙인다.¹²⁾

(35) "TransMaster"의 번역 결과

1. 꽤 춥네요. : It is quite cold.

2. 아프세요? : Are you ill?

12) 주어가 없는 문장으로 복문을 포함한 임의의 문장들을 대상으로 간단하게 번역기를 시험해 보았다. 대상 문장은 문형으로는 평서문, 의문문, 청유문, 명령문의 네 가지였고, 중3 교과서에 나올 수 있는 수준의 난이도를 가진 문장들로 택하였다. 총 200문장을 시험한 결과 75%의 성공률을 보였는데 이것은 전체 문장의 번역 완성도와는 무관하게 주어를 옳게 만들어졌는지에 따른 결과치이다. 그 중 몇몇 문장들을 예로 제시한다.

• 목이 자꾸 마른 걸 보니 점심을 짜게 먹었나 봐요.

→ It seemed to eat lunch since I am thirsty again and again.

• 한국말로 표현할 수 있는 대로 해 보세요.

→ Try to do according as you can express it in Korean.

• 아무 처방도 쓰지 않을 바에야 왜 병원에 있습니까?

→ If you do not decide to be also nobody RX, why are you at the hospital?

• 불우한 이웃을 도울 양으로 이야기를 하더군요.

→ He talked to help an unfortunate neighbour.

• 지금 심심해서 뭘 마시려던 참인데 같이 마십니다.

→ We are bored now what, so we are about to drink, let's drink together it.

- 3. 서두르지 마세요. : Don't hurry up.
- 4. 천천히 마셔야 합니다. : You must drink it slowly.
- 5. 이제 들어갈까요? : Shall we enter now?
- 6. 지금 질문을 해도 될까요? : May I ask a question now?
- 7. 제법 멀군요. : It is quite far.
- 8. 혼자 가나요? : Do you go alone?
- 9. 1,000원입니다. : It costs 1,000 won.
- 10. 무척 깊은 호수입니다. : It is a very deep lake.
- 11. 바쁘신가요? : Are you busy?
- 12. 항상 바쁘시군요. : He is always busy.
- 13. 금방 나가셨습니다. : He went out immediately.
- 14. 도와드리겠습니다. : I will help you.
- 15. 잊어버리지 마세요. : Don't forget it.
- 16. 얼마나 빠른지! : How fast it is!
- 17. 친하게 지냅니다. : Let's be on good terms.
- 18. 피카츄입니다. : This is Pikachyu.
- 19. 그 사람 이름을 잊어버렸어. : I forgot the man's name.
- 20. 2시간을 서점에서 보냈습니다. : I sent 2 hours in the bookshop.

< 표 3 > “한가이드 2.0”과 “스마트란”의 번역 결과

문장	한가이드 2.0	스마트란
1. 꽤 춥네요.	Quite cold .	The chubne fairly.
2. 아프세요?	Is it painful?	The apeuse.
3. 서두르지 마세요.	Do not hurry up.	Do not hurry up.
4. 천천히 마셔야 합니다.	Must drink slowly.	We must drink slowly.
5. 이제 들어갈까요?	May I come now?	Do you enter now?
6. 지금 질문을 해도 될까요?	May I ask a question now?	May you ask now?
7. 제법 멀군요.	Is far nicely.	We are far fairly.
8. 혼자 가나요?	Do you go alone?	Are you alone the edge?
9. 1,000원입니다.	It is 1,000 won.	We are a 1,000 won.
10. 무척 깊은 호수입니다.	It is the very deep lake.	We are very deep lake.
11. 바쁘신가요?	Are you busy?	Are you busy?
12. 항상 바쁘시군요.	Always busy .	We are busy always.
13. 금방 나가셨습니다.	Went out immediately.	We went soon.
14. 도와드리겠습니다.	Help me.	We will help.
15. 잊어버리지 마세요.	Do not forget.	Do not forget.
16. 얼마나 빠른지!	It is how fast!	How we are fast.
17. 친하게 지냅니다.	Let's get along well.	Let's get along.
18. 피카츄입니다.	It is Pikachyu.	The pikachyungibni.
19. 그 사람 이름을 잊어버렸어.	Forgot the person name.	We forgot his name.
20. 2시간을 서점에서 보냈습니다.	Spent 2 hours in bookstore.	We spent the 2 ■■■ in the bookstore.

기존의 한영 번역기에서 주어를 복원하는 경향을 살펴보면, 크게 주어를 복원하는 경우와 아예 주어를 복원하지 않는 경우로 나뉜다. 그러나 주어를 복원하지 않는 경우라 하더라도 비교적 명시적으로 주어를 예측할 수 있는 환경에서는 아주 부분적으로 주어를 복원하기도 하고, 주어를 복원하는 경우라도 일정 길이 이상의 문장이 입력될 경우 번역을 하지 않는다는지 하여 부분적으로는 주어를 복원하지 않는 경우도 있다.

위의 한영번역기 “한가이드 2.0”은 기본적으로 주어를 복원하지 않는 전략을 세우고 있지만, 의문문에서는 주어를 복원한다는 것을 알 수 있다(문장 2, 5, 6, 8, 11 참조). 반면, 또 다른 한영번역기인 “스마트란”은 주어를 복원하는 전략을 갖고 있는데, 무표적으로 주어를 복원할 경우에는 ‘we’로 복원해 내는 것을 알 수 있다. 그리고 “스마트란”은 일정 길이 이상의 문장이 입력되면 ‘Too Long Input Sentence!’라는 메시지를 띄우고 아예 번역을 하지 않는다.

반면, 본고의 결과물이기도 한 한영번역기 “TransMaster”가 번역한 문장들(35)를 보면 기본적으로 주어를 복원하는 전략을 취하고 있으며, 영어 번역문에서 무표적인 주어로 “I”를 만들어 낸다는 것을 알 수 있다. 그리고 “TransMaster”는 문형, 특정 술어류, 어미, 구 등의 정보를 바탕으로 적절한 주어를 복원해 내고 있다.

4. 요약 및 남은 문제

이 글에서는 기계번역기를 통해 한국어를 영어로 번역할 때, 한국어 문장에서는 나타나지 않는 주어를 영어 문장에서 어떻게 복원할 수 있을 것인가에 대한 전략을 짜 보았다. 주어 복원의 전략은 목표 문장이 단문인지 복문인지에 따라 다소간 다른 양상을 보였고, 이 때 참조할 수 있는 정보로는 문형, 단어, 형태소, 구가 있었다. 단문을 대상으로 하는 영한번역기의 스크립트는 3장의 (34)와 같이 만들 수 있었다. 더불어 (34)의 스크립트를 바탕으로 (주)프로랜스와 고려대학교 기계번역실에서 만든 번역기로 간단한 문장들을 번역해 보았으며, 시중의 다른 기계번역기 2종류를 택해 동일한 문장들로 비교해 보았다.

(34)와 같은 스크립트는 대상 문장이 단문이므로, 비교적 간단하게 작성되었다. 그러나 복문을 고려한다면 스크립트는 상당히 복잡해 질 수밖에 없다. 또한 현재

의 기계번역기는 분석의 최대 단위가 문장이지만, 궁극적으로는 텍스트 혹은 담화를 분석 단위로 하는 번역기가 되어야 할 것이다. 이 경우 고려되어야 하는 것은 생략과 조응의 문제이다. 특히 조응은 조응의 참조 범위를 어디까지로 잡을 것인지, 또 각 조응 요소가 동일한 문장 성분인지 아닌지에 따라 상당히 복잡한 현상을 띠는 것이다. 이는 또다른 차원의 문제로 확장되므로 차후의 연구가 필요한 부분이다.

몇 가지 방법으로 기계번역기가 만들어지고 있지만, 그 어느 것도 자연언어의 유연성과 창조성을 100% 소화해 내지는 못하고 있는 것이 현실이다. 그래서 인간의 언어 능력을 모의한다는 일이 어쩌면 또다른 생명체를 탄생시키는 일인지 하는 의문마저 들게 된다. 그럼에도 초기의 우려와 달리 현재의 일한/한일 번역기가 90%이상의 번역률을 보이는 것을 볼 때, 영한/한영 번역기(혹은 중한/한중 번역기 등등)도 머지않아 실용화할 수 있는 단계에 이르게 되지 않을까 하는 기대를 가져보게 된다.

참고문헌

- 고창수(1993). “인공지능을 위한 국어 문법의 기술”, 『어문논집』 32, 459-472, 고려대.
- 고창수(2002). 『자질연산문법이론』, 도서출판 월인.
- 권혁철(1990). “자연언어 처리에 있어서의 형태소 처리 및 어휘 사전”, 『어학연구』 26-1, 253-259, 서울대 어학연구소.
- 김영택(1994). 『자연언어처리』, 신아사.
- 김원경(2000). “한국어 격 정보와 자질 연산 문법”, 고려대학교 박사학위논문.
- 백봉자(1999). 『외국어로서의 한국어 문법 사전』, 연세대학교 출판부.
- 서병락(1996). “한영기계번역을 위한 번역 패턴 기반의 영어 문장 생성기”, 서울대학교 컴퓨터 공학과 박사학위논문.
- 송도규(1996). 『인지언어학과 자연언어 자동처리』, 홍릉과학출판사.
- 유혜원(2002). “국어의 격 교체 구문의 연구: 한영 기계 번역 시스템을 중심으로”, 고려대학교 박사학위논문.
- 여상화(2003). “영한 기계번역의 이해”, 제4회 국어정보화 아카데미 발표자료집, 137-173.

이동혁(2003). “한영기계번역을 위한 언어구의 처리 방법”, 『언어과학연구』 24, 189-212, 언어과학회.

이홍식(2000). 『국어 문장의 주성분 연구』, 월인.

이희자·이종희(2001). 『한국어 학습용 어미·조사 사전』, 한국문화사.

정천영(2000). “의미 자질과 논항 구조에 기반한 대화체 한-영 기계 번역”, 충북대학교 컴퓨터공학과 박사학위논문.

최승권(2003). “기계번역 방법론”, 제4회 국어정보화 아카데미 발표자료집, 93-113.

최호철·홍종선·조일영·송향근·고창수(1998). “기계 번역을 위한 한국어 논항 체계 연구”, 『한국어 의미학』 3, 1-39, 한국어 의미학회.

황화상(2001). “형태소 분석을 위한 한국어 어절의 구성 양상 연구”, 제13회 한글 및 한국어 정보처리 학술대회 발표논문집, 25-32.

[번역기]

스마트란. (주)엘엔텍. 서울대학교. 자연언어처리 / 인공지능 연구실.
 한가이드 2.0. (주)엘엔아이소프트.
 TransMaster. (주)프로랭스. 고려대학교 기계번역연구실.

최정혜
 136-701 서울시 성북구 안암동 5가 1번지
 고려대학교 민족문화연구원 기계번역연구실
 전화번호: (02)3290-2499
 전자우편: eirsiris@korea.ac.kr

투고논문접수일	2003년 11월 15일
논문심사일	2003년 11월 17일
심사완료일	2003년 12월 15일