

의문문 질의 시스템을 위한 한국어 문장의 의미적 동일성 분석

박홍원

서울대학교 기술정책대학원과정
우:151-742 서울시 관악구 신림동 산56-1
watts123@hitel.net

An Analysis of Identity of Meaning in Korean Sentence For Questions-Query System

Hong-won, Park
Techno-Economics & Policy Program, Seoul National University

요약 본 논문은 변형된 한국어 문장에 대해 변형 이전의 문장과 의미적 동일성을 분석하여 한국어 의문문 질의어의 문형과 상이한 문형의 한국어 문장도 정보검색시 검색 대상문에 포함시켜 검색 정확도를 높임으로써 의문문 질의 시스템의 성능을 향상시키는 것에 연구의 초점을 맞추고 있다. 한국어 문장에서 주로 나타나는 피동화에 의한 변형, 분열문에 의한 변형, 명사화에 의한 변형, 어순 재배치에 의한 변형 등의 특성에 대해 알아보고 의문문 질의 시스템에서 이들 각각의 변형을 인식하여 변형 이전의 문장과 동일한 의미의 문장으로 처리하는 방법에 대해서 자세히 살펴보았다.

1. 서론

1-1. 연구 배경

최근 한국어 정보검색 시스템에 대한 연구가 활발히 진행되면서 기존의 단순한 키워드 검색 방식에서 한단계 발전한 자연어 질의어 검색 방식이 적지 않게 시도되고 있다.[1] 이러한 추세는 인터넷상의 각종 검색엔진뿐 아니라 각종 백과사전 CD-ROM 타이틀에서도 자연어 검색을 지원할 정도로 보편화되어가는 추세이다.

그러나, 이들 자연어 검색이 대부분 특정 품사에 대한 불용어 처리를 통해 색인어를 추출하는 수준 정도로만 활용되고 있어 사실상 키워드 검색에 비해서 그다지 개선된 검색 방법이라고 보기 힘들다.

이러한 문제점을 해결하여 검색자의 편의를 최대한 보장하고 검색율도 높이기 위한 목적으로 '의문문 질의 시스템'이 제안되었다.[2] 이 모델은 한국어 의문문 자체를 질의어로 입력받아 질의어 분석을 통해 주격주제어, 서술격주제어, 의문의 초점, 초점 관련 어구 등을 추출·분석하여 자연 상태의 문장을 검색결과로 보여주는 정보검색 모델이다.

본 논문에서는 [2]에서 이미 제안된 의문문 질의 시스템을 한단계 더 발전시켜 문장 검색시 검색 대상이 되는 한국어 문장의 의미적 동일성을 분석하도록 하여 상이한 문형의 문장도 동일한 의미의 문장일 경우 질의에 부합하는 문장으로 간주하게 함으로써 보다 정확하고 효과적인 검색이 이루어질 수 있도록 하였다.

1-2. 연구 방법

먼저, 한국어 문장에서 나타나는 문장 변형에는 어떤 것이 있는지에 대해 자세히 알아보았다. 그리고, 이들 문장 변형에 대해 [2]의 의문문 질의 시스템에서 앞서 제시한 주요 3요소(주격 주제어, 서술격 주제어, 의문의 초점(검색 대상문에서는 해당 속성의 대응어))가 동일하게 유지되어 나타나는지에 대해 알아봄으로써 의미적 동일성 분석 방법을 통해 변형 이전의 문장과 동일한 수준의 검색이 가능한지에 대해 살펴보았다.

마지막으로, 문장 변형이 일어날 때 주요 3요소 각각이 어떤 위치로 이동되고 문법적으로는 어떤 변화를 일으키는지에 대해 문장 변형별로 자세히 알아봄으로써 변형 이후 문장과 변형 이전 문장의 의미적

동일성을 분석하여 각각의 문장이 동일한 의미의 문장을 판단하는 과정에 대해서도 살펴보았다.

1-3. 연구 범위

본 연구에서는 한국어 문장의 문장 변형에 대해 살펴보되 [2]에서 제안된 의문문 질의 시스템에서 다른 의문문 질의 문장의 범위에 해당하는 변형에 대해 우선적으로 살펴보았다.

따라서, 의문문 질의 시스템에서 다루지 않은 문형 범주인 내포문, 복합문 등은 본 논문에서도 다루지 않았을 뿐만 아니라 검색문의 범위도 설명의문문의 응답이 될 수 있는 한국어 문장에 대해서만 언급하였다.

한편, 본 연구는 자연 상태의 한국어 문장 중에서 질의에 부합하는 문장을 검색하는 것이므로 언어학적 고찰에서와는 달리 조사나 어순에 따라 문장의 의미나 뉘앙스 등이 조금씩 달라질 수 있는 경우라 할지라도 설명의문문의 응답으로서 사실에 부합하는 내용이라면 사소한 의미나 뉘앙스의 차이는 무시하였다.

1-4. 국내 관련 연구

국내에서는 주로 기계번역 시스템에서의 한국어 역문 생성을 위해 문장 변형에 관한 관련 연구가 진행된 바 있다.[3][4] 영어의 의존 구조에서 한국어 의존 구조로의 변환 문제, 한국어 표현 특성에 맞는 문장 변환 등에 대해 다루었다. 그러나, 이러한 연구가 대부분 자연스럽지 못한 한국어 문장을 보다 자연스러운 한국어 문장으로, 혹은 보다 의미에 맞는 한국어 문장으로 바꾸기 위한 것일 따름이지 정상적인 한국어 문장들간의 문장 변형에 대해 다루지는 않았으므로 본 연구와는 연구 목적과 범위면에서 큰 차이를 보인다.

2. 문장 변형과 의미적 동일성 분석

2-1. 한국어의 문장 변형과 처리 범위

2-1-1. 한국어 문장 변형

일반적으로 한국어 문장에서 나타나는 문장 변형

에는 능동 혹은 피동과 같이 태에 따른 변화, 분열문에 의한 강조, 명사화, 어순 재배치에 따른 논항의 위치 변화 등이 있을 수 있으나 본 논문에서는 이들을 통칭해서 문장 변형이라고 부르기로 한다.

문장 변형은 일반적으로 다음과 같은 이유 때문에 일어나게 된다.

(1) 동작주와 피동작주를 명확히 하기 위해

예) 피동화

(2) 특정어나 특정어구를 강조하기 위해

예) 분열문, 어순 재배치

(3) 간결하고 함축적인 표현을 사용하기 위해

예) 명사화

이와 같은 문장 변형은 자연어 상태의 한국어 문장을 검색하는데 적지 않은 어려움을 준다. 문장의 형태가 다르다 할지라도 동일한 의미의 문장일 경우 모두 검색 대상에 포함시켜야 하기 때문이다. 이러한 필요에 의해 변형 가능한 문장 형태를 모두 살펴보고 그 처리 방법을 본 논문에서 모색하였다.

2-1-2. 문장 변형 처리의 범위

국어학에서는 문장 변형에 대해 매우 광범위하게 다루고 있지만 본 연구에서는 의문문 질의에 대한 검색 대상문만을 대상으로 해당 문장의 의미적 동일성을 분석하는 것을 목표로 한다. 따라서, 변형이 이루어지고 난 이후에 문장의 뜻이 상당히 달라지는 경우에 대해서는 연구 범위에 포함시키지 않았다.

한편, 의문문 질의 시스템에서 이미 언급한 바 있는 의문사의 변형에 대해서도 문장 변형 처리 범위에 넣지 않았다. 이는 ‘몇 년’ 혹은 ‘몇 일’이 ‘언제’와 문장 내에서 항상 대응적으로 사용되고 분리되어 사용될 수 없기 때문이다. 또, ‘-에 의해’등이 ‘한테’, ‘에게’ 등으로 바뀌어 사용되는 어휘적 변형에 대해서도 ‘문장 변형’의 범위에 포함시키지 않았다.

2-2. 문장 변형의 종류와 그 특성

2-2-1. 피동화에 의한 변형

피동화는 능동문에서의 명사구 논항들이 재배치되어 피동문을 형성하는 현상을 말하며 이때 능동문에

서의 목적어는 피동문에서 주어로, 능동문의 주어는 사격 보어로 바뀌거나 완전히 삭제된다. 또한, 문장 변형을 문법 관계의 교체라는 측면에서 볼 때에도 피동화가 그 가장 대표적인 예라 할 수 있다.[5]

- [가-1] 안중근은 누구를 암살했나?
- [가-2] 누가 안중근에게 암살되었나?
- [가-3] 안중근은 히로부미를 암살했다.
- [가-4] 히로부미가 안중근에게 암살되었다.

[가-4]는 기본문형인 [가-3]으로부터 피동변형된 문장이다. [가-3]의 목적어인 '히로부미'가 [가-4]에서는 주어로 바뀌었다. 약간 의미적 차이가 있기는 하지만 [가-3]과 [가-4]는 각각 [가-1]과 [가-2]의 응답문이 될 수 있다.

2-2-2. 분열문에 의한 변형

분열문은 원래 강조 구문의 하나로서 월의 내용을 '구정보+신정보'의 짜임으로 나누어 신정보 요소를 특히 드러나게끔 조작한 의도적인 표현법으로 강조하기 위한 단어 혹은 구를 분열문의 초점 자리 '이'에 위치시켜 강조한다.[6]

- [나-1] 한국전쟁은 언제 일어났나?
- [나-2] 한국전쟁이 일어난 해는 언제인가?
- [나-3] 한국전쟁은 1950년에 일어났다.
- [나-4] 한국전쟁이 일어난 해는 1950년이었다.

[나-4]는 [나-3]의 문장으로부터 분열문 형태의 변형이 이루어진 것으로 [나-3]의 '1950년'이 강조되었다. [나-3]과 [나-4]는 [나-1]과 [나-2] 모두의 응답문이 될 수 있다.

이는 [나-4]가 문장의 형태에 있어서 [나-1]에 대한 응답문인 [나-3]의 그것과 완전히 동일하지는 않지만 [나-1]의 문형과 동일한 [나-3]과 의미적 동일성을 가지고 있기 때문에 사실상 [나-1]에 대한 응답으로 간주되어 해당 응답문으로 검색될 수 있음을 뜻한다. 한편, [나-3] 역시 문장 형태는 [나-2]의 그것과 다르지만 [나-4]와 의미적 동일성을 가지고 있기 때문에 [나-2]의 응답문이 될 수 있다.

2-2-3. 명사화에 의한 변형

명사화란 통사적으로 문장이나 동사구가 명사구의 자격을 갖게 되는 절차로 정의할 수 있다. 명사화 절차의 동기는 문장이나 동사구를 더 큰 구성 속에 내포시킬 필요성, 그래서 문장, 동사구를 명사구로 재범주화할 필요성에서 비롯되는 경우가 일반적이다.[7]

- [다-1] 임진왜란은 언제 일어났나?
- [다-2] 임진왜란의 발발은 언제의 일인가?
- [다-3] 임진왜란의 발발년은 언제인가?
- [다-4] 임진왜란은 1592년에 일어났다.
- [다-5] 임진왜란의 발발은 1592년의 일이다.
- [다-6] 임진왜란의 발발년은 1592년이다.

[다-5]와 [다-6]은 기본 문장인 [다-4]로부터 변형된 문장이다. 통사적으로 볼 때 [다-5]가 보다 적절한 명사화라고 할 수 있지만 실제 한국어 문장에서는 [다-6]이 더 많이 사용된다. 따라서, 본 논문에서는 [다-5]와 같이 순수하게 주어, 동사 논항만이 명사구를 형성하는 경우와 [다-6]과 같이 통사적으로 약간의 변형이 일어나지만 통사 구조가 순수한 명사구의 그것과 동일한 경우, 2가지 모두를 '명사화'로 보기로 한다.

이와 같이 명사화에 의해 변형된 문장 [다-5]와 [다-6]은 문장 형태와 관계없이 그들 문장이 가지는 의미적 동일성에 의해 [다-1], [다-2], [다-3] 모두의 응답문이 될 수 있다. 물론, [다-4]도 [다-2], [다-3]의 응답문이 될 수 있다.

2-2-4. 어순 재배치에 의한 변형

한국어의 어순은 SOV 형식을 취하는 부분 자유 어순이라고 할 수 있다. 즉, 문장의 모든 논항이 자유롭게 그 위치를 바꿀 수 있는 것이 아니라 필수적이 아닌 부사어등의 경우에만 비교적 자유롭게 위치를 바꾸는 것이 가능하다. 필수적이 기본 어순을 지키지 않을 경우 원래 전하고자 하는 의미와는 다른 의미의 문장이 되거나 의미의 뉘앙스가 달라질 수 있다.[3]

[라-1]이 일반 사실의 기술에 중점을 두고 있는 반면, [라-2]는 홍콩섬을 판 대상이 '영국'인 것을 강조하고 있다. 문장에서의 어순 재배치는 주제어 혹

은 강조어와 같은 화용론적 정보에 의해 많은 영향을 받는다.

- [라-1] 중국은 영국에게 홍콩섬을 팔았다.
- [라-2] 중국은 홍콩섬을 영국에게 팔았다.

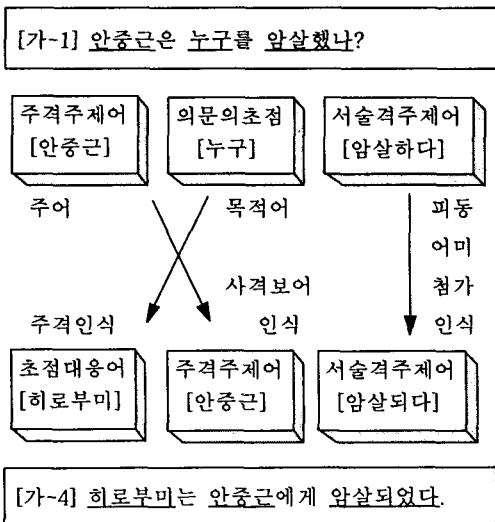
그러나, 본 논문은 의문문 질의 시스템을 울 전체로 하고 있으므로 특정 사실을 검색하고자 하는 질의의 목적에 크게 벗어나지 않는 한 검색 대상문에 포함시키고 있다. 따라서 앞 항에서와 마찬가지로 이러한 의미적 차이는 무시하기로 한다.

2-3. 검색 대상문의 의미적 동일성 분석

2-3-1. 피동화와 의미적 동일성

피동화는 앞에서 살펴본 바와 같이 능동문의 주어 가 사격보어로 나타나고 능동문의 목적어(질의문에서 의문의 초점에 대응하는 어구)는 주어로 나타난다. 아울러 능동문의 동사도 피동문에서 그 형태를 바꾸어 나타난다.

따라서, [2]에서 언급한 의문문 질의 시스템에서는 <그림1>에서와 같은 방법으로 피동변형된 문장에 대해 능동형의 응답과 동일한 의미의 문장으로 처리할 수 있다.



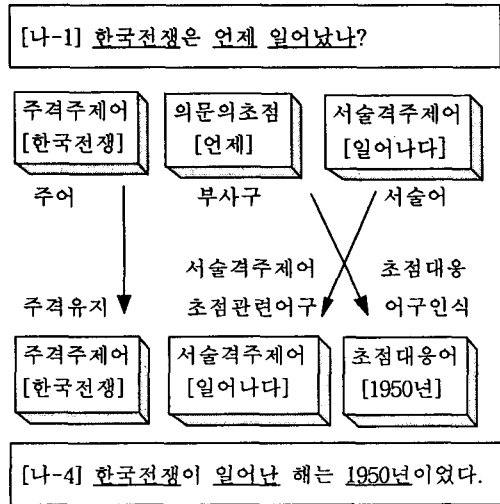
<그림1> 피동화에 의한 변형 처리

먼저, [가-1]의 목적어로 나타나고 있는 '누구'에 대한 대응어인 '히로부미'가 [가-4]에서 주어로 나타나고, [가-1]의 주어인 '안중근'이 [가-4]에서 사격보어로 나타나고, [가-1]의 서술격 주제어인 '암살하다'가 [가-4]에서 '암살되다'와 같이 피동 어미가 첨가된 형태로 나타나는 것을 통해 피동 변형으로 인식하고 변형되기 이전의 문장과 동일한 의미의 문장으로 처리할 수 있다.

한편, <그림1>의 방법을 역순으로 적용하면 전절에서 살펴본 [가-2]에 대한 응답문으로 [가-3]을 검색할 수 있다. 이러한 역순 적용법은 다른 변형에서도 동일하게 적용될 수 있으며 이후의 변형 처리부터는 역순 적용에 대해 언급하지 않기로 한다.

2-3-2. 분열문과 의미적 동일성

분열문은 앞에서 살펴본 바와 같이 특정한 어구를 강조하기 위해 특정 논항을 분열문의 초점 자리인 '-이다' 앞에 위치시키는 것이었다.



<그림2> 분열문에 의한 변형 처리

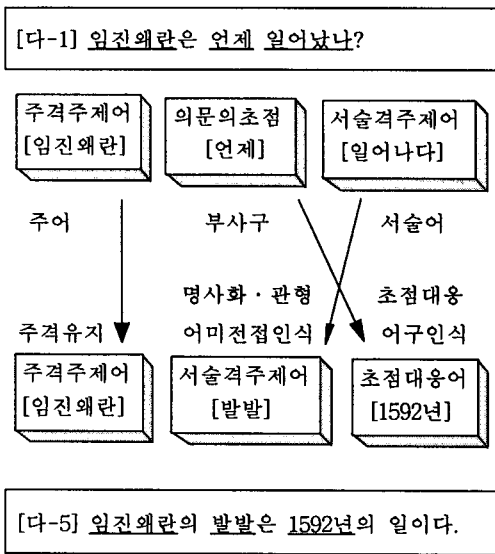
<그림2>에서 보는 바와 같이 [나-1]의 주어는 그대로 유지되고, 서술격 주제어인 '일어나다'도 유지되어 나타나고 있을 뿐만 아니라 그 뒤에 곧바로 '초점 관련 어구[QF2]'인 '해'가 후접되어 나타나고, [나-1]에서 초점인 '언제'의 대응어가 될 수 있는 시간 속성 명사(혹은 부사)인 '1950년'이 발견되고 그 뒤에 곧바로 분열문 변형 표현인 '-이다'가 후접하는

정보등을 통해 분열문에 의한 변형으로 인식하고 변형되기 이전의 문장과 동일한 의미의 문장으로 처리할 수 있다.

2-3-3. 명사화와 의미적 동일성

명사화는 앞에서 살펴본 바와 같이 문장이나 동사구가 명사구의 형태로 변형되어 나타나는 것으로 본 연구에서는 서술격 주제가 그와 동일한 의미의 명사로 나타남으로써 변형되는 명사화 문장에 주목하고 있다.

명사화에 의한 변형 처리는 분열문에 의한 변형 처리와 상당히 유사하다. 명사화에서도 분열문과 마찬가지로 변형이 일어났을 때 주어가 그대로 유지되며 서술격 주제어와 의문의 초점이 각각 위치를 바꾸기 때문이다.



<그림3> 명사화에 의한 변형 처리

<그림3>에서 보는 바와 같이 [다-1]의 주어가 그대로 유지되어 나타나고 서술격 주제어인 '일어나다'가 명사 형태인 '발발'로 바뀌어 나타나고 초점 대용어구인 '1592년'이 나타나고 있는 것을 통해 명사화 변형으로 인식하고 변형되기 이전의 문장과 동일한 의미의 문장으로 처리할 수 있다.

2-3-4. 어순 재배치와 의미적 동일성

어순 재배치는 다른 변형과는 달리 매우 간단하게 동일한 의미의 문장으로 인식하여 처리할 수 있다. 앞에서 전제한대로 본 연구에서는 약간의 의미적 차이는 무시하고 해당 질의에 부합하는 특정 사실을 포함한 문장만을 검색하는 것이 목적이므로 기본 문형의 주요 논항(주격 주제어, 서술격 주제어, 초점 대용어)이 발견되면 어순에 관계없이 동일한 의미의 문장으로 처리한다.

[마-1] 구본웅은 무슨 파의 화가였나?

[마-2] 구본웅은 야수파의 화가라기보다 차라리 낭만파의 화가였다.

[바-1] 신대륙을 발견한 사람은 누구인가?

[바-2] 아메리고 베스푸치가 신대륙을 발견했다는 사실은 뒤늦게서야 거짓으로 판명되었다.

그러나, [마-2]와 [바-2]의 경우에서와 같이 부정의 의미를 담고 있는 표현등과 같이 사실에 반대되는 표현에 대해서는 동일한 의미로 처리하지 않도록 하였다.

3. 결론

본 논문에서는 한국어 의문문 자체를 질어의로 선택하여 정보검색을 시도한 의문문 질의 시스템에서 한국어 검색 대상문의 문장 형태가 질의문의 문장 형태와 상이하더라도 동일한 의미 영역 내에 있을 경우에 동일한 의미의 문장으로 간주하고 검색 범위에 포함시키는 방법에 대해 알아보았다.

한국어에서 주로 나타나는 문장 변형을 피동화에 의한 변형, 분열문에 의한 변형, 명사화에 의한 변형, 어순 재배치에 의한 변형 등 크게 4가지로 분류하고 그 각각의 특성과 의미적 동일성 분석 가능성의 근거를 살펴보았다.

아울러, 이러한 4가지 대표적 문장 변형을 통해 변형된 한국어 문장을 의문문 질의 시스템에서 각종 격정보 인식, 피동 어미 첨가 인식, 서술격 조사 '-이다'의 후접 인식 등의 문법 정보 인식을 통해 변형 이전의 문장과 동일한 문장으로 처리하는 과정을 설명하였다.

이로써, 의문문 질의 시스템에서 보다 명확하고 효과적인 검색을 할 수 있게 되리라 기대하며 앞으로 계속된 연구를 통해 내포문, 복합문 등의 범위까지도 처리할 수 있는 더 나은 시스템이 개발되어야 할 것으로 본다.

참 고 문 헌

- [1] 안윤애, 류근호, “지식기반 정보검색 시스템의 질의처리 모델”, 컴퓨터 과학연구, 제3권 제1호, 1995.
- [2] 박홍원, “의문의 초점을 고려한 자연어 기반의 정보검색 시스템”, 제9회 한글 및 한국어 정보처리 학술발표 논문집, 1997.
- [3] 서영애, “Meaning-Text Model에 기반한 개념그래프에서의 한국어 문장 생성”, 포항공대 석사학위 논문, 1997.
- [4] 김명철, “자연스러운 표층문 생성을 위한 한국어 표현특성에 관한 연구 -PIVOT 방식의 기계번역에서-”, 한국과학기술원 석사학위 논문, 1987.
- [5] 연재훈, “문법 관계 교체 구문의 연구 -장소 보어 교체 구문과 소유주 인상 구문을 중심으로”, 한글 제232호, 한글학회, 1996.
- [6] 정재형, “초점과 그 표현 유형”, 국어국문학 제24집, 부산대학교 국어국문학회, 1987.
- [7] 김홍수, “속격 명사화, 명사 병치 명사화의 양상과 기능”, 어문학논총 제13집, 국민대학교 어문학연구소, 1994.