

수사학적구조리론과 그에 기초한 영어본문구조모형화

장 총 혁

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《과학기술의 종합적발전추세와 사회경제발전의 요구에 맞게 새로운 경계과학을 개척하고 발전시키는데 큰 힘을 넣어야 합니다.》(《조선로동당 제7차대회에서 한 중앙위원회사업총화보고》 단행본 40페이지)

정보기술이 급속히 발전하고있는 오늘의 현실은 컴퓨터에 의한 자연언어처리체제를 비롯한 여러가지 언어처리체제들을 더 높은 수준에서 연구개발할것을 요구하고있다.

지난 시기 컴퓨터언어학과 자연언어처리연구에서는 문장을 주되는 연구대상으로 하여 문장의 구조와 의미를 해석하기 위한 연구사업을 깊이있게 진행하여 일정한 성과를 거두었다. 그러나 문장에 대한 분석만으로는 의미적으로 정확하고 이해하기 쉬운 번역문을 얻을수 없으며 문장이상의 언어단위 즉 본문에 대한 해석과 본문정보의 리용이 필수적인 과제로 제기되고있다. 본문에 대한 해석을 통하여 문장들사이에 맺어지는 구조적 및 의미적연관성을 밝힐수 있으며 본문정보를 리용하여 대명사지시해석이나 단어의미모호성 해소와 같은 해석처리의 정확성도 높일수 있다.

이 글에서는 본문의 수사학적구조리론과 그에 기초한 본문구조모형화방법을 제기하고 수사학적구조모형을 자연언어처리를 위한 본문분석에 리용하는데서 제기되는 문제들을 론하려고 한다.

무엇보다먼저 수사학적구조리론에 대하여 론하기로 한다.

1980년대말에 제기된 수사학적구조리론(Rhetorical Structure Theory)은 문장보다 큰 언어적단위인 본문의 구조에 존재하는 규칙을 연구하는 리론으로서 전통수사학에 기초하고있다. 수사학적구조리론은 자연언어본문의 구성에 대하여 론리의미적연결의 측면에서 연구하고 본문의 구조를 본문부분들사이에 맺어지는 수사학적관계에 의하여 특징짓는다. 수사학적구조리론은 본문구조에 대하여 다음의 3가지 기본사상을 내세우고있다.

(1) 중심요소-종속요소구조(nucleus-satellite structure)는 보편적인 특성이다.

(2) 본문계층구조의 기초는 본문의 기능적특성이다.

(3) 본문구조는 통신적역할을 한다.

수사학적구조리론은 본문이 구조적계층성을 가지며 본문부분들사이의 관계를 기능적 측면에서 직관화할수 있다는것을 밝히었다. 그러므로 수사학적구조리론으로는 각이한 주제와 길이의 본문들을 분석할수 있다. 수사학적구조리론은 본문에서 절들사이의 의미적관계를 서술하는 일반적인 방법을 제시하고있는것으로 하여 접속사들의 의미를 표현할수 있는 효과적인 수단으로도 된다.

수사학적구조리론은 다음의 4가지 기본요소 즉 관계(Relations), 구조도식(Schemas), 구조도식적용(Schema applications), 구조(Structures)로 구성된다.

수사학적구조리론의 기본요소에는 우선 수사학적관계가 있다.

본문의 두 부분사이에는 중심요소(nucleus)와 종속요소(satellite)로 이루어지는 일정한 관계가 존재한다. 중심요소는 저자의 목적에서 보다 중요한 요소로서 그것만으로도 이해될 수 있는 본문부분이지만 종속요소는 부차적인 요소로서 중심요소와의 관계속에서만 이해될 수 있는 본문부분이다. 수사학적관계정의는 다음과 같은 네개의 요소(중심요소에 대한 제한, 종속요소에 대한 제한, 중심요소와 종속요소의 결합에 대한 제한, 영향)를 포함한다.

수사학적관계들중의 하나인 양보관계(Concession)를 실례로 이 요소들을 설명해보자.

(1) 중심요소에 대한 제한: 저자는 중심요소에 반영된 상황에 대하여 긍정적으로 평가한다.

(2) 종속요소에 대한 제한: 저자는 종속요소에 반영된 상황이 성립되지 않는다고 주장하지 않는다.

(3) 중심요소와 종속요소의 결합에 대한 제한: 저자는 중심요소와 종속요소에 반영된 상황들사이에 잠재적인 또는 명백한 배제관계가 존재한다는것을 인정하지만 양립할수 있는것으로 생각한다. 그리고 중심요소와 종속요소들에 반영된 상황들이 양립한다는 사실이 독자로 하여금 중심요소에 반영된 상황에 대하여 보다 긍정적인 태도를 가지게 한다는것을 인정한다.

(4) 영향: 중심요소에 반영된 상황에 대한 독자의 긍정적인 태도가 증가한다.

본문의 표층에서 양보관계는 접속사 although가 이끄는 절을 통하여 나타난다.

Title: Dioxin

1. Concern that this material is harmful to health of the environment may be misplaced.

2. Although it is toxic to certain animals,

3. evidence is lacking that it has any serious long-term effect on human beings.

이 본문의 저자는 2와 3사이에 잠재적인 배제관계가 존재하는것을 인정하면서도 호상 양립할수 있다는것을 보여주었다. 이것을 도식으로 나타내면 아래와 같다.

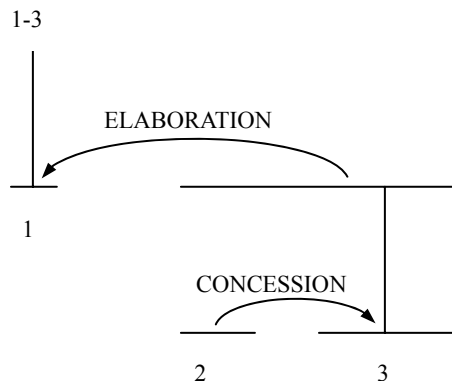


그림 1. 실례본문의 수사학적구조도식

이러한 수사학적관계분류의 기초로 되는 특징들이 연구자들의 관심에 따라 달라지기 때문에 관계들을 정확히 분류하기가 힘들다. 초기의 수사학적구조리론연구자들(Mann, Matthiessen 등)이 1991년에 제기한 관계모임은 수백개의 본문들에 대한 분석에서 성공적으로 리용되었는데 그들이 제기한 관계모임은 다음과 같다.

- | | |
|-----------------------------------|----------------------|
| ① 구체화관계(Elaboration) | ⑫ 평가관계(Evaluation) |
| ② 정황관계(Circumstance) | ⑬ 재진술관계(Restatement) |
| ③ 해결관계(Solutionhood) | ⑭ 요약관계(Summary) |
| ④ 의지적원인관계(Volitional Cause) | ⑮ 순서관계(Sequence) |
| ⑤ 의지적결과관계(Volitional Result) | ⑯ 대조관계(Contrast) |
| ⑥ 비의지적원인관계(Non-volitional Cause) | ⑰ 증거관계(Evidence) |
| ⑦ 비의지적결과관계(Non-volitional Result) | ⑱ 반대관계(Antithesis) |
| ⑧ 목적관계(Purpose) | ⑲ 양보관계(Concession) |
| ⑨ 조건관계(Condition) | ⑳ 동기관계(Motivation) |
| ⑩ 배타적관계(Otherwise) | ㉑ 가능화관계(Enablement) |
| ⑪ 해석관계(Interpretation) | ㉒ 정당화관계(Justify) |
| | ㉓ 배경관계(Background) |

수사학적구조이론에 의하면 정의되는 관계들은 그 어떤 문법적표식에도 의존하지 않으며 전적으로 본문부분들사이의 의미적관계나 기능적관계에 기초한다.

수사학적구조이론의 기본요소에는 또한 구조도식과 구조도식적용이 있다.

구조도식은 본문의 구조적성분을 정의하며 몇개의 본문부분을 구성요소로 하는 추상적인 본문구조모형이다. 구조도식은 구조도식적용조건과 함께 쓰이면서 본문의 수사학적 구조를 결정한다.

수사학적구조이론에는 수사학적관계를 나타내는 아래와 같은 5가지 형태의 구조도식이 있다.

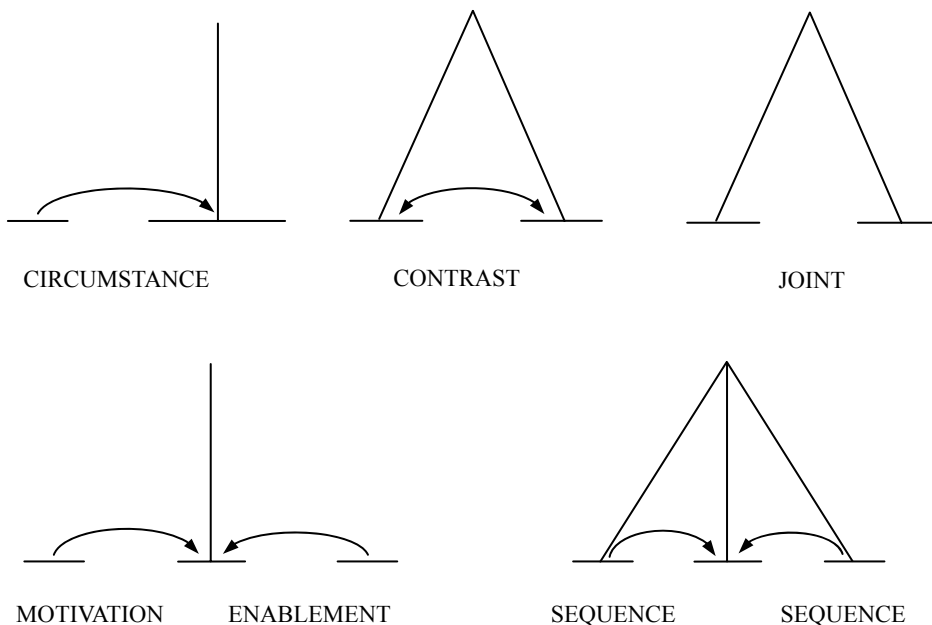


그림 2. 수사학적관계를 나타내는 구조도식

우의 도식에서 곡선은 관계를 나타내며 수직선은 중심요소에 의하여 표현되는 본문부분을 가리킨다. 여기서 본문부분은 중심요소나 종속요소로 되는 본문의 부분이다. 우의

도식에 반영되지 않은 관계들은 정황관계(Circumstance)의 도식과 같은 형태로 표시된다. 대조관계를 나타내는 도식은 두개의 중심요소를 가지며 순서관계도식은 순서를 이루는 요소들을 나타내는 여러개의 중심요소를 가진다. 합동관계(Joint)도식은 많은 중심요소를 가질수 있지만 대응되는 종속요소는 가지지 않는다.

다음으로 수사학적구조리론에 기초한 본문구조모형화에 대하여 론하기로 한다.

수사학적구조리론에 기초한 본문구조모형화에서 첫 단계는 본문을 여러 단위로 토막화하는것이다. 초기의 수사학적본문구조모형화에서 분석의 단위들은 주로 절단위였다. 그러나 자연언어처리의 견지에서 본문의 수사학적구조모형을 실현하는데서는 절단위를 모형화단위로 하기보다는 본문토막화에 의하여 얻어진 본문토막을 모형화단위로 하는것이 보다 타당성이 있다. 그것은 본문토막화에 의하여 분할된 본문토막들이 문법적관계에 의하여 구분되는 절단위와는 달리 정보적내용에서 일정한 공통성을 가진 본문부분들이기때문에 토막들사이의 의미적관련성을 확정하기가 쉽기때문이다.

수사학적구조리론에 기초한 본문구조모형화는 다음의 제한조건에 따르는 도식적용의 모임으로 실현된다.

(1) 완전성(Completedness): 모임은 전체 본문을 구성하는 본문부분들을 포함하는 하나의 도식적용으로 이루어진다.

(2) 련관성(Connectedness): 전체 본문이 하나의 본문부분으로 되는 경우를 제외하고는 분석에 리용되는 개별적인 본문부분이 다른 도식적용의 최소단위 또는 구성요소로 된다.

(3) 특이성(Uniqueness): 모든 도식적용은 서로 다른 본문부분들의 모임으로 이루어지며 다중관계도식에서 개별적인 관계들은 서로 다른 본문부분모임에 적용된다.

(4) 린접성(Adjacency): 개별적인 도식이 적용된 본문부분들이 린접하여 하나의 본문부분을 이룬다.

완전성, 련관성, 특이성의 결합은 수사학적구조모형이 나무구조로 표현될수 있게 한다. 위에서 본 수사학적구조도식에서 활동선에는 성립하는 관계이름을 밝혀주며 본문부분으로부터 아래방향으로 그어지는 수직선들은 대응하는 중심요소에 대한 도식적용에 의하여 갈라진다.

아래의 신문기사본문을 실례로 수사학적구조모형을 작성해보자.

City police had to help control traffic recently when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Marriott Hotel. The hotel's help-announcement – for 300 openings – was a rare opportunity for many unemployed.

The people waiting in line carried a message, a refutation, of claims that the jobless could be employed if only they showed enough moxie. Every rule has exceptions, but the tragic and too-common tableau of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs, not laziness.

(최근에 도시경찰은 이제 문을 열게 될 매리오토호텔에 직업을 먼저 신청하기 위해 수백명의 사람들이 줄을 지어선 사정으로 교통지휘를 방조하지 않으면 안되었다. 300개의 일자리가 필요하다는 호텔측의 발표는 수많은 실업자들에게 혼치 않은 기회였다.

줄을 서서 기다리는 사람들은 실업자들도 열성만 있다면 직업을 가질수 있다는 주장에 대한 반박을 보여주는 격이 되었다. 모든 규정에는 례외가 있지만 수백, 지어

수천명의 사람들이 정규직을 얻으려고 뱀꼬리처럼 늘어서있는 비극적이고도 흔히 볼수 있는 모습은 게으름이 아니라 일자리가 부족한 현실을 보여주고있다.)

이 본문을 담화표식어에 의한 본문토막화를 통하여 7개의 본문부분으로 토막화할수 있다.

1. City police had to help control traffic recently
2. when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Marriott Hotel.
3. The hotel's help-announcement – for 300 openings – was a rare opportunity for many unemployed.
4. The people waiting in line carried a message, a refutation, of claims that the jobless could be employed if only they showed enough moxie.
5. Every rule has exceptions,
6. but the tragic and too-common tableau of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs,
7. not laziness.

다음 본문부분들사이의 수사학적관계를 정의한다.

본문부분 1과 2사이에는 비의지적결과관계가 존재한다. 본문부분 1-2는 본문부분 3이 담고있는 내용의 의지적결과로 된다. 본문부분 6과 7사이에는 부사 not에 의하여 표현되는 반대관계가 존재한다. 그리고 본문부분 5와 6사이에는 접속사 but에 의하여 표현되는 양보관계가 존재한다. 본문부분 5-7은 4가 나타내는 내용의 증거로 된다. 전체적으로 볼 때 본문부분 1-3은 4-7이 담고있는 내용에 대한 배경으로 된다. 이러한 관계정의에 기초하여 아래와 같은 수사학적구조도식을 얻을수 있다.

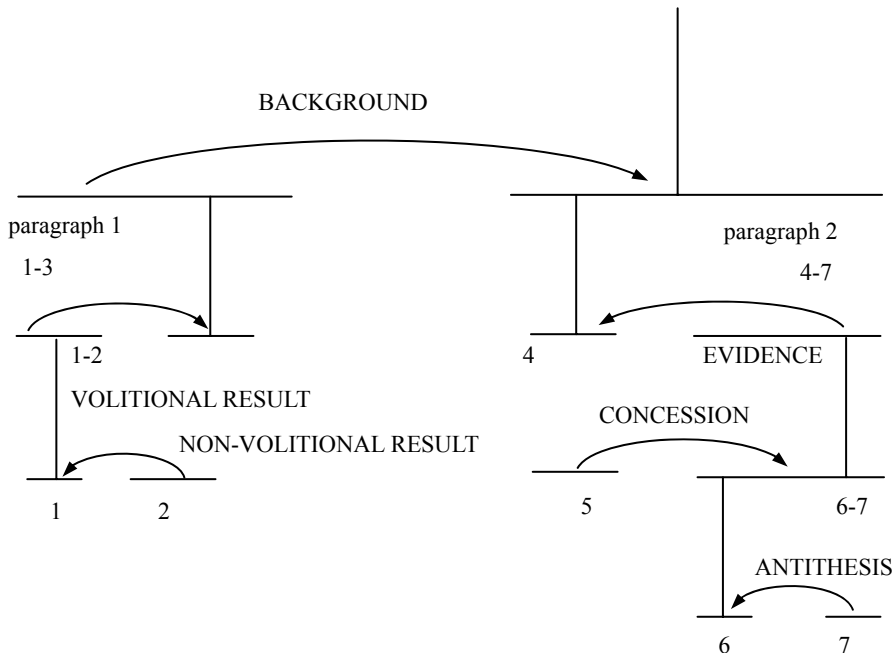


그림 3. 실례본문의 수사학적구조도식

수사학적구조이론과 그에 기초한 본문모형화에 대한 연구는 최근에 더욱 깊이있게

진행되어 여러가지 수사학적관계모임들이 정의되고 본문의 수사학적구조를 자료기지화한 코퍼스들도 존재하고있다. 실례로 RST TreeBank는 대규모본문구조코퍼스로서 16개 부류에 78개의 수사학적관계들을 정의하고 그에 기초한 수사학적구조분석자료를 제공한다.

수사학적구조리론에 기초한 본문모형화는 자연언어처리체계개발에서 중요한 의의를 가진다. 수사학적구조리론에 토대한 모형화방법으로 수백개의 본문을 분석한 결과에 의하면 거의 모든 본문들이 수사학적구조를 가진다는것이 밝혀졌다. 그러나 수사학적구조리론도 본문모형화에서 일정한 부족점을 가지고있다.

우선 수사학적구조리론으로 구조를 설명할수 없는 본문들이 존재하는것이다. 일부 본문들, 실례로 법규, 계약서, 공보문 그리고 시와 같이 고정된 시작구조와 마감구조를 가지는 본문들은 수사학적구조리론으로 구조를 서술할수 없다.

또한 수사학적관계에 대한 정의를 이루는 네가지 요소중에서 영향이 중요한 역할을 수행하기때문에 수사학적구조분석은 저자가 의도하는 목적에 대한 설명으로 되며 결과적으로 수사학적구조분석은 본문전체에 대한 기능적분석으로 된다. 그러나 본문의 많은 특성들을 기능적분석만으로는 밝혀낼수 없기때문에 자연언어처리에서는 수사학적구조에 의하여 설명되지 않는 본문의 특성들을 처리할수 있는 방법이 있어야 한다.

또한 수사학적구조가 본문의 형태적구조가 아니라 기능적구조이며 본문의 형태적특성과 관련되지 않기때문에 자연언어처리에 리용하는 경우 표층에서 나타나는 구조적실마리가 없이 어떻게 본문에 대한 수사학적구조분석을 실현하겠는가 하는 문제가 제기된다. 그러므로 자연언어처리에서 수사학적구조리론을 효과적으로 응용하기 위해서는 수사학적구조모형화에서 본문의 표층정보를 리용하기 위한 방도가 있어야 한다. 그러한 방도의 하나가 수사학적구조모형을 실현하는데 담화표식어들을 리용하는것이다. 담화표식어들은 본문을 일정한 부분으로 구획지어주면서 부분들사이의 논리의미적연관관계를 나타낸다. 이러한 담화표식어들을 본문의 표층에서 나타나는 언어학적징표로서 수사학적구조분석에 리용한다면 자연언어처리를 위한 본문구조모형화를 실현할수 있다.

그러나 수사학적관계가 담화표식어에 의하여 나타나지 않는 경우도 많다. 실례로 2001년에 Carlson 등이 작성한 수사학적구조코퍼스에서는 238개의 대조관계중 61개, 307개의 설명-증거관계중 79개만이 명백한 담화표식어에 의하여 표시되는 관계들이었다. 이외에도 담화표식어를 수사학적관계를 나타내는 언어학적특징으로 리용하는데서 여러가지 문제들이 제기된다. 실례로 증거관계를 나타내는 담화표식어로서 since를 리용할수 있지만 이 경우에 언어처리체계는 이 단어가 담화표식어로서 증거관계를 나타내는 종속요소와 결합할뿐아니라 문장속에서 부사로서 쓰일수도 있다는것을 고려하여야 한다. 그렇지만 담화표식어를 수사학적구조모형화에 리용하는것은 초기에 제기된 수사학적구조모형을 일정하게 개선하고 자연언어처리의 목적에서 보다 효과성이 있다는데 의의가 있다.

수사학적구조리론과 수사학적본문구조모형을 자연언어처리에 효과적으로 리용하기 위한 문제는 앞으로 더욱 깊이있게 연구되어야 할 문제이다. 우리는 자연언어처리에 본문정보를 리용하기 위한 연구를 더욱 심화시켜 기계번역체계를 비롯한 여러가지 언어처리체계들의 성능을 더욱 높여나가야 할것이다.