

문장의 구획단위－《구》에 대한 이해

강 승 일

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《우리는 대담하게 목표를 높이 세우고 최신과학기술을 연구도입하는데 적극 달라붙어 최단기간에 중요분야의 과학기술을 세계적수준에 올려세워야 합니다.》(《김정일선집》 증보판 제22권 23페이지)

과학기술이 비약적으로 발전하는 오늘 폭발적으로 쏟아져나오는 정보자료들을 신속정확히 처리하는것은 인민경제발전과 인민생활향상, 과학기술발전에서 중요한 의의를 가진다.

이 글에서는 자연언어처리를 원만히 진행해나가는데서 중요한 문제의 하나인 문장의 구획단위에 대하여 밝히려고 한다.

문장론적자동분석에서 출발점은 문장을 보다 작은 단위들로 가르는데 있다. 그것은 문장에 대한 이해가 문장의 개별적단위들에 대한 분석과 종합을 거쳐 진행되기때문이다. 그러므로 문장을 무엇을 단위로 하여 어떻게 구획짓는가 하는것은 문장론적자동분석에서 선차적이고 관건적인 고리로 된다.

지금까지 문장구획짓기는 여러 측면에서 진행되어왔다.

언어학에서 문장은 형태－의미적측면, 어음－의미적측면, 논리－의미적측면, 의미－기능적측면에서 구획지어졌다. 이때 문장의 구획은 각이한 관점에서 각이하게 진행되며 거기서 얻어지는 문장의 구획단위도 역시 동일하지 않다.

여기서 흔히 리용하는 방법은 문장을 문장성분으로 나누는것이다. 그러나 이것을 자연언어처리분야에 받아들이자면 그 모형화에서 어려운 문제들이 많이 제기된다. 그것은 컴퓨터가 의미나 문장론적관계와 같은 추상적인 개념들을 리해하지 못하기때문이다.

자연언어처리분야에서의 문장의 구획단위는 조선어의 특성을 정확히 반영하면서도 컴퓨터가 리해하기 쉽게 정의되어야 한다.

이러한 난관을 극복하기 위하여 이 글에서는 위의 두가지 조건을 만족시키는 문장의 구획단위로서 《구》를 제기한다. 구는 문장을 용언적단어를 표식으로 하여 가른 문장의 구획단위이다. 여기서 용언적단어는 동사, 형용사, 체언의 용언형을 의미한다.

무엇보다먼저 구를 단위로 문장을 구획하는 근거에 대하여 보기로 한다.

그것은 우선 용언적단어가 문장속에서 다른 단어들과 맺는 관계의 중심, 논리적의미의 중심이기때문이다.

체언은 어떤 대상에 대한 일정한 명명에 지나지 않으므로 그것이 사건, 현상에 대한 설명을 나타내자면 반드시 용언적단어와 결합되어야 한다. 즉 대상을 나타내는 체언은 용언적단어를 통하여서만 사건, 사실, 현상을 설명할수 있다.

실례로 《나는 학교로 간다.》라는 문장에서 《나》라는 대상이 《간다》라는 동사를 통하여

《학교》라는 목표를 향하여 움직이는 현상을 들수 있다.

이처럼 용언적단어는 문장에서 문장성분들을 의미-문법적으로 통솔하는 단위인것으로 하여 문장을 가르는 표식으로 된다.

그것은 또한 컴퓨터에서 용언적단어에 대한 판단이 그리 어려운 문제가 아니기때문이다.

이미 형태론적자동분석이 상당한 정도로 진척된 결과 컴퓨터가 임의의 문장에서 품사의 유형과 토의 형태를 밝히는것은 별로 힘든 일로 되지 않고있다.

실례로 《나는 학교로 간다.》라는 문장이 제시되면 컴퓨터는 형태론적자동분석알고리즘에 의하여 《나-대명사, 는-도움토, 학교-명사, 로-조격토, 가다-동사, ㄴ 다-종결토》라는 형태구조분석정보를 쉽게 얻어낸다.

구에 의한 문장의 구획은 형태-의미적관점도 아니고 어음-의미적관점도 아니며 론리-의미적관점도, 의미-기능적관점도 아니다. 그것은 오직 자연언어처리를 위하여 용언적단위를 표식으로 가르는 문장의 구획단위이다.

문장을 구획하는데서는 용언적단어의 형태가 무의미하다. 다만 품사정보가 용언적단어라는 내용에 속하는가 속하지 않는가 하는것이 기본표식으로 된다.

다음으로 구의 유형에 대하여 보기로 한다.

론문구를 용언적단어가 어떤 형태로 끝났는가에 따라 종결형구, 접속형구, 규정형구, 상황형구로 나눈다. 구의 표식은 《/》으로 하며 용언적단어는 용언이라고 쓰기로 한다.

① 종결형구

용언을 표식으로 구획된 단위에서 용언의 형태가 종결형으로 끝난 구이다.

례: 나는 그가 준 책을 읽었다.

우의 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《나는 그가 준/책을 읽었다./》로 된다. 《책을 읽었다.》라는 구는 용언의 형태가 종결형으로 끝났으므로 종결형구로 된다.

② 접속형구

용언을 표식으로 구획된 단위에서 용언의 형태가 접속형으로 끝난 구이다.

례: 그는 대학에 와서 나에게 책을 주었다.

우의 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《그는 대학에 와서/나에게 책을 주었다./》로 된다.

《그는 대학에 와서》라는 구는 용언의 형태가 접속형으로 끝났으므로 접속형구로 된다.

③ 규정형구

용언을 표식으로 구획된 단위에서 용언의 형태가 규정형으로 끝난 구이다.

례: 나는 내가 공부하는 대학을 사랑한다.

우의 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《나는 내가 공부하는/대학을 사랑한다./》로 된다.

《나는 내가 공부하는》이라는 구는 용언의 형태가 규정형으로 끝났으므로 규정형구로 된다.

④ 상황형구

용언을 표식으로 구획된 단위에서 용언의 형태가 상황형으로 끝난 구이다.

례: 그는 힘차게 노래를 불렀다.

우의 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《그는 힘차게/노래를 불렀다.》로 된다. 여기에서 《그는 힘차게》라는 구는 용언의 형태가 상황형으로 끝났으므로 상황형구로 된다.

다음으로 자연언어처리분야에서의 구와 문법에서의 구의 차이점과 공통점에 대하여 보기로 한다.

지금까지 구에 대해서는 주로 문법에서 논의되었다. 그러나 자연언어처리분야에서 구는 문법의 확대성분에서 논하는 구와 엄연히 구별된다.

《조선어리론문법》(1986)에서는 구에 대하여 〈주어+술어〉의 구조를 가지면서 확대성분의 기능을 노는 문장론적단위라고 정의하고 구의 유형을 주어구, 술어구, 보어구, 상황어구, 규정어구로 보았다.

자연언어처리분야에서의 구와 문법에서의 구는 본질에서 차이난다.

자연언어처리분야에서의 구가 문장의 구획단위라면 문법에서의 구는 확대성분으로서 문장성분의 단위이다.

자연언어처리분야에서의 구와 문법에서의 구는 구분기준에서 차이난다.

자연언어처리분야에서의 구가 용언적단어 즉 동사, 형용사, 체언의 용언형을 구분기준으로 한다면 문법에서의 구는 〈주어+술어〉의 구조를 구분기준으로 한다.

자연언어처리분야에서의 구와 문법에서의 구는 유형에서도 차이난다.

자연언어처리분야에서의 구가 종결형구, 접속형구, 규정형구, 상황형구로 되어있다면 문법에서의 구는 주어구, 술어구, 보어구, 규정어구, 상황어구로 되어있다.

그렇다고 하여 자연언어처리분야에서의 구와 문법에서의 구사이에 공통점이 없다는것을 의미하는것이 아니다.

자연언어처리분야에서의 구는 어디까지나 문법에서의 구를 자연언어처리분야에 받아들인것으로서 그것들사이에는 일정하게 공통성이 존재한다. 그것은 자연언어처리분야에서의 구와 문법에서의 구가 모두 서술성을 가진 단어들을 단위로 하여 구획되며 단계적성격을 띠는것이다.

언어학에서는 확대성분으로 되는 단어결합내부에서 마지막위치를 차지하면서 술어와 같이 서술적기능을 수행하는 단어를 서술어라고 하며 서술어를 가진 단어결합이 확대성분으로 되어있는것을 구라고 하였다.

실례로 《설명절은 사람들이 마음속으로 새 희망을 꿈꾸는 명절이다.》라는 문장에서 확대규정어로 되어있는 《사람들이 마음속으로 새 희망을 꿈꾸는》은 그 마지막단어가 규정형으로 되어있으며 동시에 서술적기능을 수행하고있다. 따라서 우의 실례에서 《사람들이 마음속으로 새 희망을 꿈꾸는》은 규정어구이다.

자연언어처리분야에서의 구와 문법에서의 구의 공통점과 차이점을 유형별로 보면 다음과 같다.

— 종결형구와 술어구

례: 방안은 보기가 매우 깨끗하였다.

이 문장을 용언을 표식으로 구획하면 《방안은 보기가 매우 깨끗하였다.》로 되며 따라서 문장전체가 바로 종결형구이다. 종결형구는 문법에서의 술어구와 비슷한데 이 문장에서 술어구는 《보기가 매우 깨끗하였다.》로 된다.

이처럼 자연언어처리분야에서의 구는 순수 용언적단어를 기준으로 하여 갈랐기때문에 문장전체가 종결형구로 되지만 문법에서의 구는 《주어+술어》의 구조를 기준으로 하기때문에 《보기가 매우 깨끗하였다.》가 술어구로 된다.

례: 나는 영어를 하고/그는 로어를 한다./

그는 열심히 공부를 하였다./

나는 일터로 나가는/동무를 만났다./

이 문장들에서 종결형구는 《그는 로어를 한다.》, 《공부를 하였다.》, 《동무를 만났다.》이다. 그런데 이것들은 문법에서의 술어구로는 되지 않는다. 그것은 문법에서의 구가 확대 성분으로서 두개이상 단어들이 결합되어 하나의 성분으로 쓰이는것을 나타내는 개념이라면 종결형구는 자연언어처리분야에서 순수 용언적단어를 기준으로 가른 문장의 구획단위이기때문이다.

—접속형구

례: 그는 룡구를 하고 나는 배구를 하였다.

이 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《그는 룡구를 하고/나는 배구를 하였다.》로 된다.

《그는 룡구를 하고》라는 구는 용언의 형태가 접속형으로 끝났으므로 접속형구로 된다. 접속형구를 구획한 실례의 문장을 보면 다음과 같다.

례: 나는 대학에 가고/그는 집으로 갔다./

그는 열심히 훈련을 하고/또 하였다.

여기에서 《나는 대학에 가고》, 《훈련을 하고》라는 구는 용언의 형태가 접속형으로 끝났다. 이것이 바로 자연언어처리분야에서 가른 문장의 구획단위로서 접속형구이다.

그러나 문법에서는 접속형구와 비슷한 유형의 구를 설정하지 않고있다. 그것은 문법에서의 구는 문장의 구획을 위하여 설정한것이 아니라 확대성분의 단위를 정하기 위하여 설정한것이기때문이다. 다시말하여 문법에서의 구는 문장의 구획을 위한것이 아니기때문에 접속형구와 비슷한 유형의 구를 설정할 필요가 없다.

—규정형구와 규정어구

례: 나는 새로 이사온 동무를 만났다.

위의 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《나는 새로 이사온/동무를 만났다.》로 된다. 여기서 《나는 새로 이사온》이라는 구는 용언의 형태가 규정형으로 끝났으므로 규정형구로 된다.

문법에서의 규정어구는 규정어로 쓰인 단어결합 즉 확대된 규정어를 말한다.

위의 실례에서는 《새로 이사온》이 문법에서의 규정어구로 된다.

차이는 규정형구에서는 《나는》까지 포함하여 구로 보지만 규정어구에서는 《나는》은 포함하지 않고 《새로 이사온》만을 구로 보는것이다.

례: 그는 일하는/동무를 도와주었다./

나는 책을 읽는/동생에게 물었다./

여기서 《그는 일하는》, 《나는 책을 읽는》이라는 구는 용언의 형태가 규정형으로 끝났으므로 규정형구로 된다.

위의 문장들에서 규정어구는 《책을 읽는》이다. 즉 규정형구와 규정어구는 명백한 차이

가 있다.

그것은 우선 자연언어처리분야에서는 확대성분이든 단순성분이든 용언의 형태가 규정형으로 되어있으면 규정형구로 갈라내지만 문법에서는 용언의 형태가 비록 규정형이라고 해도 확대성분이 아니면 구로 보지 않는것이다.

레: 그는 일하는 동무를 도와주었다.

우의 실례에서 《그는 일하는》은 규정형구로는 되지만 규정어구로는 되지 않는다.

그것은 또한 자연언어처리분야에서는 구를 문장의 구획단위로서 규정형으로 끝난 용언뿐아니라 그 앞단위까지 다 포함하여 부르는 개념이지만 문법에서는 확대된 규정어만을 규정어구로 본다는것이다.

레: 나는 책을 읽는/동생에게 물었다.

규정형구-나는 책을 읽는

규정어구-책을 읽는

-상황형구와 상황어구

레: 그는 유희장에서 재미나게 놀았다.

실례의 문장을 용언을 표식으로 구획하면 두개의 구로 된다. 즉 《그는 유희장에서 재미나게/놀았다./》로 된다. 여기서 《그는 유희장에서 재미나게》라는 구는 용언의 형태가 상황형으로 끝났으므로 상황형구로 된다.

문법에서의 상황어구는 두개이상의 단어로 이루어진 상황어 즉 확대된 상황어를 의미한다.

레: 천막안에서는 날이 새도록 등잔불이 꺼지지 않았다.

우의 레문에서 《날이 새도록》이 문법에서의 상황어구로 된다.

레: 그는 나는듯이/집으로 달렸다./

우리 대학 전자도서관은 아주 훌륭하게/꾸러졌다./

여기에서 《그는 나는듯이》, 《우리 대학 전자도서관은 아주 훌륭하게》라는 구는 구획된 단위들에서 용언의 형태가 상황형으로 끝났으므로 자연언어처리분야에서의 상황형구로 된다. 우의 레문을 문법적으로 분석하면 《그는 나는듯이》는 단순성분으로 이루어졌으므로 구로 되지 않으며 《우리 대학 전자도서관은 아주 훌륭하게》에서 《아주 훌륭하게》가 상황어구로 된다. 즉 상황형구와 상황어구는 명백한 차이가 있다.

우선 자연언어처리분야에서의 구는 확대성분인가 단순성분인가에 관계없이 용언의 형태가 상황형으로 끝나면 구로 구획하지만 전통문법에서의 구는 용언의 형태가 상황형으로 끝났다고 하여도 확대성분이 아니면 구로 되지 않는다.

레: 그는 나는듯이 집으로 달렸다.

실례에서 《그는 나는듯이》는 자연언어처리분야에서의 구 즉 상황형구로 되지만 문법에서의 구 즉 상황어구로는 되지 않는다.

또한 자연언어처리분야에서의 구는 상황형으로 끝난 용언뿐아니라 그 앞단위까지 포함하여 구로 보지만 문법에서의 구는 확대된 상황어만을 구로 본다.

레: 우리 대학 전자도서관은 아주 훌륭하게 꾸러졌다.

여기서 자연언어처리분야에서의 구는 확대된 상황어뿐아니라 《우리 대학 전자도서관은》까지 포함하여 《우리 대학 전자도서관은 아주 훌륭하게》가 구로 되지만 문법에서의 구

는 확대성분인 《아주 훌륭하게》만이 구로 된다.

자연언어처리분야에서의 구는 이미전에 나온 조선어문장론적자동처리연구에서의 분구와도 일정하게 차이를 가진다.

구도 분구와 마찬가지로 용언적단어를 기준으로 하여 가른 문장의 구획단위이지만 유형에서는 분구와 차이난다.

분구가 종결형분구, 접속형분구, 규정형분구, 상황형분구, 체언서술형분구로 이루어졌다면 구는 종결형구, 접속형구, 규정형구, 상황형구로 이루어졌다. 즉 구에는 체언서술형구라는것이 없다.

체언서술형분구는 용언을 표식으로 구획된 단위에서 용언의 형태가 체언의 상황형, 규정형, 접속형, 종결형으로 끝난 구획단위이다.

레: 여기가 나의 대학이다./

너는 최우등생일수록/ 학습을 더 잘해야 한다.

그는 나의 친구인/김동무를 만났다.

이것은 나의 책이고/저것은 너의 교과서이다.

이처럼 체언서술형분구인 경우에는 그 형태가 체언서술형의 상황형, 규정형, 접속형, 종결형으로 표현되며 이때의 구획은 각각 용언의 상황형, 규정형, 접속형, 종결형처럼 한다. 그리고 체언서술형의 분구류형은 체언서술형단어의 형태에 따라서 동사나 형용사에 의한 분구류형분류와 같이 취급한다.

그러므로 이 글에서는 체언서술형분구를 따로 설정하지 않고 종결형구, 접속형구, 규정형구, 상황형구에 체언서술형분구의 매 류형들을 각각 포함시켜 구의 류형을 4개로 하려고 한다. 그것은 위에서 본바와 같이 체언서술형분구가 구의 다른 류형과 같이 취급되는 조건에서 구의 류형을 4개로 하면 구의 모형화에서 복잡성을 피하고 자동분석의 속도와 정확성을 높일수 있기때문이다.

우리는 앞으로 문장의 구획단위로서의 구에 대한 연구를 더욱 심화시켜 자연언어처리를 보다 원만히 진행하도록 하는데 도움을 주어야 할것이다.