

본문의 형태단어처리에서 토 《이》의 분리와 관련한 문제

박 정 혁

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《새로운 과학기술분야를 개척하기 위한 사업도 전망성있게 밀고나가야 합니다. 나라의 과학기술을 세계적수준에 올려세우자면 발전된 과학기술을 받아들이는것과 함께 새로운 과학기술분야를 개척하고 그 성과를 인민경제에 적극 받아들여야 합니다.》(《김정일선집》 증보판 제11권 138~139페이지)

언어학에서 새로운 분야를 개척하며 그 성과를 언어생활에 적극 받아들이는데서 나서는 당면한 문제들중의 하나는 프로그램적방법과 수학적방법을 리용하여 컴퓨터로 본문의 의미적단위들을 자동분리하는 문제이다. 특히 형태단어에 들어있는 문법적형태부를 과학적으로 판단하고 그것을 말줄기와 말뿌리에서 정확히 갈라내는것이 중요하다.

여기에서 선차적인것은 발음과 맞춤법이 같고 뜻과 기능이 다른 단위들가운데서 토들을 갈라내는 문제이다.

이 글에서는 본문의 형태단어처리에서 토 《이》를 분리하는 문제에 대하여 논의하려고 한다.

《조선말대사전》(사회과학출판사 2007년)에 의하면 글자 《이》가 명사, 대명사, 불완전명사, 수사와 같은 단어로 판정되는것은 11개, 여러가지 뒤붙이로는 4개, 토로 쓰이는 경우는 5가지이다. 그러므로 본문의 형태단어처리에서 토 《이》를 분리하자면 먼저 글자 《이》가 품사나 뒤붙이로 쓰인것을 분리해낸 다음 토 《이》를 기능에 따라 분리하여야 한다. 이것은 형태단어들에서 토 《이》를 분리하는 문제가 대단히 복잡하고 어려운 언어학적 과제의 하나라는것을 보여준다.

모임론의 견지에서 보면 형태단어들에서 토 《이》를 분리하는 과정은 단어, 뒤붙이 및 토들의 모임에서 먼저 단어를 분리한 다음 뒤붙이를 떼어내고 끝으로 토들의 본질과 기능을 갈라보는 공정으로 이루어져있다고 말할수 있다.

무엇보다먼저 형태단어에서 단어 《이》를 분리하여 토분리에서 제외시켜야 한다.

글자 《이》가 단순히 토로만 쓰이는것이 아니므로 토 《이》의 분리에서 품사로서 쓰인 단어 《이》를 분리하여 제외시키는것은 선차적인 공정으로 된다. 그러자면 글자 《이》의 명사적쓰임을 잘 알아야 한다.

우선 우리 말에서 글자 《이¹》은 글자 《ㅣ》를 의미하여 그 이름으로서 명사이다.

이것은 우리 글자 《ㅣ》에 대한 상징적인 올림말이므로 본문에서는 대체로 쓰이지 않으며 쓰이는 경우에도 《글자 <ㅣ>》, 《문자 <ㅣ>》와 같이 단어 《글자》, 《문자》가 입력되거나 글자의 형태가 자음 《ㅇ》없이 모음자 《ㅣ》로만 등장한다. 그러므로 이러한 경우들은 모두 《명사》로 처리한다.

또한 글자 《이²》는 《이발》이라는 뜻을 가진 명사로 쓰인다.

언어생활에서는 명사 《이》가 널리 쓰인다. 《이가 맞다.》, 《남이야 전보대로 이를 쭈시든 말든》, 《사발 이빠진것》, 《삶은 무우에 이 안들어갈 소리》와 같은 단어결합이나 성구들에서 글자 《이》는 단어 《이발》의 의미이며 형태단어의 시작에 《이》가 오므로 토분리에서 제외시킨다. 《금이, 송곳이, 어금이, 버드령이, 옥이》와 같은 합성단어들에 들어있는 《이》는 여러가지로 분석된다.

먼저 형태단어자료기지에 《금이, 송곳이, 어금이, 버드령이, 옥이》와 같은 단어들이 있는가 없는가를 보고 자료기지에 이미 등록된 단어가 있으면 그것을 명사로 규정한다. 형태단어자료기지에 없을 때에는 《이》의 뒤에 토가 오면 명사로 하며 토가 오지 않고 용언 또는 부정부사가 오거나 합성어로서 다른 명사말뿌리와 합해지면 역시 명사로 한다. 이런 경우들에 의하여 명사 《이》가 토 《이》의 분리에서 제외되게 된다.

또한 글자 《이³》은 해로운 벌레를 의미하는 명사, 글자 《이⁴》는 갈퀴의 끝부분을 의미하는 명사이며 글자 《이⁵》는 《이패》를 의미하는 명사로서 형태단어의 시작에 《이》가 출현하므로 토 《이》의 분리에서 제외한다.

이와 함께 《이(여기)까지 다치지 말고 놔두게나.》에서와 같이 장소를 가리키는 명사 《여기》의 말체형으로 쓰이는 글자 《이⁶》도 역시 제외하며 대명사 《이⁷》(이것)도 같은 리유로 토분리에서 제외한다.

또한 글자 《이⁸》은 감동사로서 《이, 다치겠다. 물러서라.》와 같이 혼자서 형태단어로 쓰이는데 이것도 역시 토가 아니므로 제외한다.

또한 글자 《이^{9, 10}》은 수사로서 형태단어의 첫시작에 올수도 있고 가운데서 나타날수도 있으며 불완전명사로 쓰이는 《이¹¹》과 같이 토 《이》처럼 형태단어의 마지막에 올수도 있는 글자는 따로 분석할 필요가 제기된다.

실례로 수사 《이천이백이십이》를 보면 수사 《이》는 형태단어 《이천이백이십이》에서 첫시작과 가운데, 마지막에서 출현하였다. 《이천》에서는 우의 글자들에서처럼 형태단어의 시작에서 출현하였기때문에 제외할수 있지만 《이백이십》에서는 형태단어의 가운데서 출현하였으므로 그것이 토인가 아닌가를 판정해보아야 한다.

그러자면 글자 《이》의 앞과 뒤에 오는 단어들과의 련관속에서 분석하여야 한다.

앞에 오는 단어 《이천》을 명사(보통명사 혹은 고유명사)로 보는 경우 글자 《이》가 문법적형태부로 되자면 반드시 그뒤에 다른 토가 오거나 형태단어로 끝나야 한다.

《이백이십》에서 《이》뒤에 온 단어 《백이십이》의 《백》은 토가 아니므로 이 단계에서 글자 《이》는 문법적형태부가 아니라 뒤에 온 단어와의 련계속에서 《수사》라는 결론을 내리게 된다. 그뒤에 온 단어 《이십이》에서의 첫 글자 《이》도 마찬가지이다.

마지막단어 《이》의 분석은 불완전명사 《이¹¹》과 문법적형태부들인 《이¹¹》-《이²⁰》들과 함께 살펴보기로 한다.

이 모든 과정을 종합하여 여러 품사로서의 《이》의 집단을 토 《이》의 모임에서 제외시키게 된다.

품사 《이》들을 분리하는 공정은 컴퓨터에 이미 구축되어있는 품사정보와 형태단어정보에 의하여 진행하게 된다.

다음으로 형태단어에서 뒤불이 《이》를 분리하여 토분리에서 제외시켜야 한다.

우선 명사조성뒤불이로서 뒤불이 《이》를 분리하여야 한다.

실례로 《고기잡이》, 《가을걷이》, 《해돋이》, 《모살이》들에서 《이》는 모두 동사말뿌리 뒤에 붙어서 그것을 명사로 만드는 명사조성의 뒤불이이다. 특히 《고기》, 《가을》, 《해》, 《모》는 명사에 포함되고 《잡(다)》, 《걷(다)》, 《돋(다)》, 《살(다)》는 동사에 소속된다.

여기서 두 소리마디로 된 《고기》와 《가을》을 각각 한개씩의 소리마디로 가를 때 그것들이 독자적인 의미적단위에 대응되지 않는다는 분해공정이 앞서게 된다는것을 놓치지 말아야 한다.

여기로부터 {《(명사말뿌리+동사말뿌리)+이》+《Φ/토》}의 단어구조가 성립되면 이때의 《이》는 《명사조성뒤붙이》로 규정될 수 있다. 여기에서 《Φ》는 《이》뒤에 다른 아무것도 오지 않는다는 것이고 《토》는 《해돋이를, 해돋이가》와 같이 토가 붙는 경우를 의미한다.

물론 명사조성뒤붙이는 꼭 《명사+동사》형에만 붙는 것이 아니다.

《갓난이》나 《젊은이》에서 《갓난이》는 《(부사+동사+규정토)+이》로 볼 수 있고 《젊은이》는 《(형용사+규정토)+이》로 볼 수 있다. 여기서 공통점은 《용언+이》가 어떤 대상을 나타내는 것으로 된다는 것이다. 따라서 이런 경우의 《이》도 같은 항목의 뒤붙이로 규정할 수 있다.

이와 함께 뒤붙이 《이》가 형용사말뿌리에 붙어서 명사를 만드는 경우(넓이, 높이, 낮이)와 명사나 부사에 붙어서 명사를 만드는 경우(《더퉁이》, 《꿀꿀이》, 《팔랭이》)도 여기에 포함시킨다.

그러나 《매미》, 《빠꾸기》, 《깍뚜기》 같은 것은 이미 《이》가 말뿌리에 녹아붙은 것이기 때문에 《이》형태를 보존하지 못하고 형태부의 단순화과정이 일어난 것으로 처리하게 된다. 그리고 명사조성뒤붙이는 아니지만 명사에 붙은 뒤붙이로서 《이》의 경우도 여기에 포함시킨다.

례: 성철이, 예영이

우의 실례에서 《성철》, 《예영》은 보통명사나 다른 품사가 아니라 고유명사로서 사람이름이며 다 받침으로 끝났다. 이렇게 《이》가 닫긴마디로 끝나는 사람이름에 붙은 경우 뒤붙이로서 토분리에서 제외시킨다.

또한 부사조성뒤붙이로서 뒤붙이 《이》를 분리하여야 한다.

글자 《이》가 형용사의 말줄기에 붙어서 부사를 이루는 경우에는 그것을 뒤붙이로 규정하고 토분리에서 제외시킨다.

례: 자유롭다>자유로+이

가볍다>가벼+이

공교롭다>공교로+이

따뜻하다>따뜻+이

이것을 일반화하면 《형용사말줄기+맺음토》>《형용사말줄기+이》로 된다. 이때 《이》는 부사조성뒤붙이로 규정된다. 이와 함께 확대된 용언에 글자 《이》가 붙어서 부사를 이루는 경우에도 그것을 뒤붙이로 보고 토분리에서 제외시킨다.

례: 쓸데없+이, 감쪽같+이, 한결같+이

이때 《이》역시 부사조성뒤붙이로 규정된다. 여기에는 상징부사에 《이》가 붙는 경우도 포함된다.

례: 빙긋+이, 썩긋+이

또한 동사조성뒤붙이로서 뒤붙이 《이》를 분리하여야 한다.

형용사나 상징부사의 말뿌리에 글자 《이》가 붙어 동사를 만드는 경우 뒤붙이로 규정하고 토분리에서 제외시킨다.

례: 반짝+이+다>반짝이다

높+이+다>높이다

들먹+이+다>들먹이다

허덕+이+다>허덕이다

이상의것을 종합하면 동음이의적인 《이》의 모임속에서 단어와 뒤붙이로 결정되는 요소들은 다 제외되고 오직 토만이 남게 된다.

다음으로 토 《이》들을 기능에 따라 분리하여야 한다.

토 《이》들의 모임속에서 그것이 어떤 성격의 토로 되는가를 결정하는것은 토분리의 마감단계이다.

우선 주격토를 결정하여야 한다.

례: 책이 있다. 사람이 많다.

책들이 있다. 사람들이 많다.

여기에는 명사뒤에 주격토 《이》가 직접 음으로써 결정할수 있는 경우와 복수토 《들》이 분석된 다음 주격토를 분리할수 있는 경우도 있다. 이 토는 대상토이므로 명사 등 체언에 붙으며 주로는 받침으로 끝나는 체언뒤에 온다.

례: 춤이 절로 나온다.

우의 실례는 《추다》에 명사조성뒤붙이 《ㅁ》이 붙은것으로 볼수도 있으며 《춤》을 명사로 보고 거기에 《이》가 붙은것으로 볼수도 있지만 어떤 경우에도 주격토 《이》의 결정에는 지장을 주지 않는다. 그러므로 명사 등 체언에 붙은 토 《이》를 주격토라고 결정할수 있다.

또한 맺음토를 결정하여야 한다.

례: 내 가슴도 아프이.

정말 좋은 생각을 하였네그러. 용하이.

웁으이. 지금은 여기에 와있어.

우의 실례에서 토 《이》는 형용사의 말뿌리에 붙는데 그 말뿌리가 《아프(이)》, 《용하(이)》와 같이 열린마디로 끝나는 경우에는 그뒤에 붙은 맺음토 《이》가 직접 분리되지만 《웁으이》와 같이 말뿌리가 닫긴마디로 끝나는 경우에는 결합모음 《으》를 분리한 다음 맺음토 《이》를 분리해낸다.

또한 바꿈토를 결정하여야 한다.

토 《이》는 체언을 용언화하는 바꿈토로 되는 경우가 있다.

례: 작업장이면서, 벽이며, 지붕이며, 유원지이며

여기에서 형태단어는 《명사말뿌리+이+이음토》의 구조를 가지는데 이런 경우 토 《이》는 바꿈토로 규정된다.

례: 저수지이다, 저수지이니다, 저수지이구말구

여기에서도 토 《이》는 명사와 맺음토의 사이에 놓이는 바꿈토로 된다. 이 두가지 경우를 종합하면 《이》뒤에 오는 이음토와 맺음토는 풀이토로서의 공통성을 가지게 된다.

결국 토 《이》를 바꿈토로 분리해내기 위하여서는 그앞에 오는 품사결정과 함께 그뒤에 오는 토들의 공통성과 차이점도 옳게 밝혀내야 한다.

또한 상토를 분리하여야 한다.

상토 《이》는 동사의 말줄기가 주로 모음으로 끝났거나 자음 《ㄱ, ㄴ, ㄷ, ㄱ, ㅎ, ㅍ》로 끝난데 붙어서 피동형태나 사역형태를 나타낸다.

례: 놀이다, 덧놓이다, 덮이다, 짚이다, 먹이다, 깎이다, 붙이다, 끼이다, 쓰이다

이런 단어들이 실제로 글말문장에 쓰일 때에는 형태단어(실례로 《놀이다》가 형태단어 《놀입니다까?》, 놀이였어요, 놀이지만, …》등)로 실현되므로 상토 《이》의 분석은 다른 어

러 맺음토들에 대한 분석과 함께 진행되는것으로 된다.

끝으로 토 《이》가 주로 자음으로 끝난 체언의 말줄기가 자음으로 시작되는 도움토나 그밖의것과 결합할 때 자음끼리 잇닿는 어려운 발음을 극복하기 위하여 그사이에 결합모음과 같이 끼워들어가는 경우가 있다. 이것을 사전에서는 편의상 《토》항목에 넣었으므로 5가지의 토를 분리하는것에 포함시켰다.

례: 책이랑 연필이랑, 책이나마, 책이라도

우의 실례에서 물론 글자 《이》가 나타나는 경우가 우에서 논의한 순서대로만 주어지는것이 아니지만 그가운데 임의의 문맥에 놓인것이라고 하여도 이 체계속에 들어가서 비교되고 분석되면 그것이 토인가 아닌가 하는것이 분석판단될수 있다. 즉 글자 《이》가 품사와 덧붙이 및 토의 세가지 계열속에서 어느 한 갈래에 속하게 되고 그것이 만약 토에 속하면 우의 다섯가지 토의 항목에 비교되면서 그것이 무슨 토인가를 결정할수 있다.

이러한 분석공정은 곧 품사나 덧붙이 및 토사전을 구체적으로 정확히 작성하고 형태부분석공정과 그 방법을 치밀하게 작성하는것을 전제로 하고있다.

모든 언어학자들은 우리 말의 전체 토의 갈래를 분석할수 있는 정연한 자료기지를 구축하기 위한 연구사업을 심화시켜 언어학을 과학기술적토대우에 확고히 올려세워야 할것이다.