

'와/과'구문의 중의성 연구

유 혜 원

고려대학교 국어국문학과 (민족문화연구원 기계번역실 연구원)
yhw72@chollian.net(HP:011-662-2414)

The Study of ambiguity in the 'wa/kwa'

Hye Won Yoo

Dept. of Korean Lang & Lit, Korea University

요 약

본고는 한영번역기 개발을 위한 기초 작업으로 '와/과'구문에 나타나는 여러 가지 패턴을 정리하고 이들 구문에서 보이는 중의성 문제를 해결하고자 하였다. 이러한 작업을 위해서는 자료 수집 및 분석이 우선이기 때문에 코퍼스에서 '와/과'구문을 뽑아서 분석하여 규칙을 마련하였다. 여기에서 사용된 자질연산문법(FCG)은 자연언어처리를 위한 문법으로 변형규칙과 수행도의 개념 없이 자질을 이용한 연산 체계로서 언어처리를 하고자 하는 문법이다. 이 이론을 바탕으로 규칙을 세우고 실제 언어 자료를 뽑아서 테스트를 하여 95%의 성공률을 보여주었다. 그러나 여기서의 연구는 '와/과'구문의 처리를 위한 가장 뼈대가 되는 기초연구이며, 앞으로 좀 더 많은 처리가 이루어져야 하리라 생각된다.

1. 서론

한국어의 다양한 중의성 양상은 한국어 정보처리에 관련된 모든 작업에서 어려움을 초래하는 요인이 되고 있다. 특히 기계번역기를 개발하는 작업에서 이러한 중의성 문제는 먼저 해결되어야 하는 부분이다.

본고는 한영번역기 개발을 위한 기초 작업을 위한 연구로, '와/과'구문에 나타나는 중의성에 초점을 맞추었다. 여기서의 중의성은 문장이 본래 두 가지 이상의 뜻을 가지는 구조적 중의성의 경우를 다룬다기 보다는 한국어 분석에서 컴퓨터가 가지는 적은 정보와 규칙으로 인해 야기되는 분석적 중의성을 해결하는 방안을 마련하고자 한다. 한국어 구문분석을 위한 문법으로 언어 처리를 위해 고안된 자질연산문법(FCG)를 이용하여 이러한 문제를 해결하는 방법론을 보이겠다.

2. 와/과 구성에 나타난 분석적 중의성 및 유형

한국어의 '와/과' 구성은 영어의 and와 유사하게 여러 가지 양상을 보여준다. 즉, 단어와 단어를 접속하기도 하고, 구와 구, 문장과 문장을 접속하기도 한다. 문장에서 동일한 역할을 하는 성분이 '와/과'로 연결이 되어 있다면, 문장의 성공적인 논항분석을 위해서는 그 두 성분을 하나로 묶어주는 것이 이후의 구문분석의 절차를 성공적으로 수행하기 위해서 필요한 절차이다.

(1) 나는 사과와 배를 좋아한다.

위의 문장에서 성공적인 분석이 되어서 각 성분을 표시하기 위해서는 다음과 같은 분석이 요구된다.

(2) [나는]_{주어} [사과와 배를]_{목적어} [좋아한다]_{서술어}

즉, 논항분석이 이루어지기 전의 단계에서 이처럼 '와/과'

로 연결된 요소들을 묶어주는 절차가 필요함을 보여준다. 그러나 앞에서 잠깐 언급했듯이, '와/과'가 반드시 이처럼 동일한 요소를 묶을 때만 나타나는 것이 아니기 때문에 이러한 다기능적 성격이 분석적 중의성을 유발시킨다고 볼 수 있다.

'와/과'가 나타날 수 있는 환경을 크게 나누면 네 가지 정도로 요약해 볼 수 있다. 우선 하나의 단위로 처리가 가능한 '-와 + 부사류'가 있고, 두 번째로는 '와/과'를 논항으로 요구하는 동사 구성, 세 번째로 'N+ N'구성 중에서 묶일 수 없는 명사류, 마지막으로 둘을 묶을 수 있는 명사류 구성이 있다.

우선 '와+ 부사류'의 구성을 보일 수 있는 것은 다음과 같은 예문이 있다.

- (3) a. 나의 시간은 멈췄다, 그녀와 함께.¹⁾
- b. 안방문은 평소와 마찬가지로 열려 있었다.

위의 패턴들은 생산성도 높은 부류들로 이러한 구성이 가능한 항목들을 모아보면 다음과 같다.

(4) 와+ 부사류 패턴

-와 함께, 와- 단들이/들이서, -와 마찬가지로, -와 같이, -와 가깝게, -와 맞물려, -와 관련, -와 별도로, -와 달리, -와 나란히, -와 더불어

이는 '-와/과' 구성에서 부사류와 공기하는 것으로 후행하는 요소들의 품사적 자질만으로 이들을 다른 구성과 변별할 수 있다.

두 번째는 동사가 그것의 논항으로서 '와/과'를 가지는 문장들이다.

- (5) a. 그때 석훈이 아이와 눈웃음을 맞추며 돌아왔다.
- b. 효철은 다회와 처음 사귀기 시작하면서 가끔씩 이 아파트에서 자고 가곤 했다.
- c. 그녀는 나와 이름이 같았을 뿐 아니라 나이도 똑같았다.

위의 예문들은 동사가 '와/과'를 논항으로 취하는 동사들이다. 물론, '와/과' 성분이 논항인지 부가어인지의 여부가 논란의 여지가 있긴 하지만, 여기서 중간적 성격을

1) 위의 예문들은 민족문화연구원에서 구축한 150만 어절 코퍼스에서 '와/과'가 나오는 문장들 중에서 추출한 것이다.

보인다는 것이 그다지 중요하지 않을 수도 있다. 왜냐하면, 한국어를 다루는데 있어서 이론적으로 접근하는 것과 자연언어처리의 입장에서 언어를 처리하는 것에 있어서 중요한 정보가 때론 다를 수 있기 때문이다. 자연언어처리의 관점에서는 언어적 현상과 그것의 규칙을 찾는 것이 가장 중요하기 때문에 동사가 논항인지 아니면, 부가어적인 성격을 보이는 것인지에 대한 명확한 판단보다는 동사가 수의적이든 필수적이든 그 논항을 요구할 수 있다는 것이 더 중요하기 때문에 일단 동사의 특성상 나타날 수 있는 요소는 논항으로 볼 수 있다. 다음은 이러한 '와/과'를 다소 넓은 의미의 논항으로 취하고 있는 동사 목록들이다.

(6) 동사 목록

가깝다 / 가지다(인터뷰)/ 같다/ 전주다/ 결부되다/ 결합하다/ 결혼하다/ 경쟁하다/ 계약하다/ 공존하다/ 공통되다/ 관계없다/ 관련되다/ 교대하다/ 교배시키다/ 교환하다/ 구별하다/ / 분하다/ 나누다(인사를)/ 놀다/ 놀러가다/ 다르다/ 달라지다/ 담소하다/ 닿아 있다/ 대립되다/ 대면하다/ 대조되다/ 대화하다/ 동떨어지다/ 동업하다/ 동일하다/ 동참하다/ 떨어지다/ 같다/ 마주보다/ 마주치다/ 만나다/ 맞대다/ 맞먹다/ 맞붙다/ 맞추다/ 맺다(동맹을)/ 멀다(거리가)/ 멀어지다/ 무관하다/ 밀착되다/ 바꾸다/ 부딪치다/ 부합되다/ 분리되다/ 비교하다/ 비슷하다/ 사귀다/ 살다/ 상봉하다/ 상이하다/ 섞여있다/ 실망이하다/ 싸우다/ 쓰다/ 씹름하다/ 약속하다/ 어우러지다/ 어울리다/ 연결되다/ 연결시키다/ 연결하다/ 연계되다/ 연관시키다. / 유사하다/ 융합시키다/ 이야기하다/ 이웃해있다/ 인접하다/ 인터뷰하다/ 일치하다/ 절친하다/ 접속하다/ 접촉하다/ 직결되다/ 짝이되다/ 차단하다/ 차이나다/ 충돌하다/ 친하다/ 친해지다/ 통하다/ 합의하다/ 합쳐지다/ 헤어지다/ 협의하다/ 흉장난을 하다/ 흡사하다

위의 동사들을 보면 대부분은 어떤 대상이 다른 대상과 맺는 관계를 나타내거나, 행위자가 어떤 대상에 대해 맺는 관계를 나타내는 의미적 속성이 있는 동사들임을 알 수 있다. 이들 동사들은 '와/과'논항을 필수적으로 가지기도 하고, 이 논항을 수의적으로 요구하기도 한다. 그러나, 공통적으로 이들은 어떤 대상과의 관계를 가지는 동사들이기 때문에 이들 동사들의 공통성을 자질 공유로서 포착해 볼 수 있다. 우선 이들 동사들이 범주적으로 동사라는 공통성을 가지고 있기 때문에 [v] 자질을 가지고 있으며, 뿐만 아니라 앞서도 지적했듯이 의미적 공통성을 가지고 있다. 이들의 관계적인 속성을 고려해 보면 [relation]정도의 자질로 표시할 수 있다. 이 [relation]자

질은 동사 관계뿐만 아니라 명사들의 속성을 묶는데도 유용한 자질이기에 때문에 생산적이다. 따라서 이들 동사들의 공통적인 속성으로 다음과 같은 자질을 설정할 수 있다.

(7) '와/과'를 논항으로 취하는 동사의 자질 속성
[v, relation]

세 번째로 고려해야 하는 것이 'N+N' 구성에서 서로 묶일 수 없는 명사류의 출현이다. 이는 다음과 같은 예문을 고려해 볼 수 있다.

- (8) a. 또 조선 업계는 일본 조선 업계와 공동으로 국제적인 선박 수요를 평준화할 수 있는
b. 보험 회사와 합의가 안 돼 수리를 시작하지도 않았다.
c. 한때 켈리스트 로스트로포비치와 염문설이 나돌기도 했으나, (이하생략)

위의 예문들은 범주상으로만 본다면 '와/과'를 처리하는데 있어서 가장 어려움을 보여주는 것이라고 할 수 있다. 그러나 위의 예문들에 나타나는 명사들의 쌍을 비교해 보면 해결 가능성이 보인다. 다음은 그러한 목록들을 뽑은 것이다.

- (9) a. 회사와 합의/ 후보와 시소게임/ 민방 업계와 갈등/ 늑대와 춤/ 세가사와 라이선스계약/ 컴퓨터와 대화/ 정은이와 약속/ 남자와 결혼/ 운영체제와 대화/ 업자와 결탁/ 광역의회와 분리/ 이박사와 친교/ 개발 본부와 가계약/ 부처와 협의/ 교사와 교류/ 거머리와 대결/ 아저씨와 인사/ 여행업자와 특약/ 돌고래와 대화
b. 업계와 공동/ 문제와 때/ 로스트로포비치(인명)와 염문설/ 염소와 이마/경우와 차이/ 백담치 후보와 2백여표차/ 강우혁후보와 손/ 쓴 박사와 시간여행/ 쓰레기와 범벅/ 언니와 방학/ 권씨와 정반대/ 우리와 뜻
c. 나와 친구/ 옛 동료와 우정

위의 예문은 일정한 유형을 보여준다. (9a)의 문장은 앞에 있는 명사(N1)이 비서술성 명사인데 반해서 뒤에 나오는 명사(N2)가 서술성 명사인²⁾ 특성을 갖고 있다. 서

술성 명사는 다음과 같은 자질을 가질 수 있다.

(10) 서술성 명사의 자질 [n, action]

이러한 서술성 명사의 자질 정보를 이용하면 나머지 N+N구성과의 차별성도 보일 수 있다.

(9b)의 예문은 N2가 서술성을 가지지 않는다는 면에서 (9a)와 구별된다. 그러나 위의 예문에 나타난 N의 자질들을 생각해 보면 상이한 의미부류에 속함을 알 수 있다. 예를 들어 '로스트로포비치와 염문설'을 가지고 생각해 보면 다음과 같이 자질을 나타낼 수 있다.

(11) 로스트로포비치 [n, human]
염문설 [n, event]

이처럼 이들이 서술성이라는 면에서는 비관여적이지만 의미적 자질에서 큰 차이를 보이고 있음을 알 수 있다. 이는 뒤에 서술할 묶일 수 있는 'N+N'의 자질과 비교해 본다면 극명하게 드러나게 될 것이다.

(9c)의 예문은 N2의 자질이 관계성을 띤 명사일 때라는 특징을 가진다. 위에서 이미 동사를 서술하면서 [relation]의 자질을 설정했는데, 이 예문도 마찬가지로 그와 같은 방식으로 문제 해결이 가능할 것이라 생각된다.

묶일 수 있는 N1+N2의 구성은 [1]에서 그 유형이 잘 정리되어 있다. 이 연구는 N1과 N2의 구성의 하나의 의미단위를 이룰 것이라는 가설아래 이를 증명한 논의이며, 이 연구는 자연 언어의 기계처리에 적합한 모델을 만들고자 했다는 측면에서 본 연구와 밀접한 관련성을 갖고 있다고 할 수 있다. [1]에서 N1 + N2구성이 가능한 구조를 다음과 같은 네 가지로 나누고 있다.

- (12) N1과 N2 구성
- 반의적 관계 결합 구조
 - 동의적 관계 결합 구조
 - 상하의적 관계 결합 구조
 - 은유적 의미 생성

위의 연구는 이 두 가지 N 사이의 유사성을 발견했다는 측면에서 많은 의의를 가지지만, 실제로 자연언어를 처리하는데 있어서는 정밀하고 광범위한 워드넷(woronet)이 구축되어 있지 않다면, 그 관계를 포착하기

2) 서술성 명사는 비실체성 명사 중에서 행동성을 보여주는 동사들을 말하는 것이다. 따라서, '사랑, 동작, 인사, ...' 등이 이러한 서술성 명사에 해당한다고 할 수 있다.

가 무척 어려운 점이 있다. 그러나 이들이 단일한 의미 단위로 묶일 수 있다는 가능성은 자질 이론에서는 무척 많은 시사점을 주는 것이라 생각된다. 즉, 이들이 그러한 관계를 형성한다는 것은 좀더 큰 의미 자질 수준에서는 같은 자질을 공유하고 있다는 것을 의미하기 때문이다. 반의적 관계는 잘 알려진 대로, 자질로 표시하자면 모든 자질들을 공유하고 단 하나의 자질이 다를 경우를 나타내는 것이고, 동의적 관계란, 거의 모든 자질이 동일한 낱말들 사이의 관계를 지칭하는 것이며, 상하의적 관계라고 하는 것은 결국 둘이 공유하는 자질이 있다는 것을 의미하기 때문이다. 따라서, 묶일 수 있는 N1+N2 구성도 역시 자질로서 그 특성을 파악해 볼 수 있다는 결론에 도달하게 된다.

(13) a. 여름비는 세상의 모든 **환희와 비애, 사랑과 이별**을 일시에 잠재워 버릴듯 점점 더 세차게 쏟아졌다.

b. 의복은 **천과 천**의 봉합에 의하여 선이 생기게 되며 형태가 완성된다.

c. 인간 세계는 기독교에서 말하듯 버림받고 추방당함으로써 내던져진 **고난과 꺾박**의 저주받은 세계가 아니고, 자연과 동화하면서 무한한 조화와 질서의 가능성 속에서 살아가는 아름답고 찬란한 광명의 세계인 것이다.

d. 게다가 **왕실과 귀족**들의 지지를 받으며 귀족 불교로 자리잡은 많은 사원들은 세금을 내지 않는다는 특혜를 이용하여 (중략) 타락에 타락을 거듭하고 있었다.

e. 커피는 **설탕과 프림**의 배합도 중요하지만 섞는 속도와 방향이 무엇보다 중요한 겁니다.

위의 예문에서 (13a)는 반의 관계의 예라고 할 수 있다. 그러나 이 둘은 의미 자질에 있어서 공유를 나타낸다.

(14) 환희 [n, emotion]

비애 [n, emotion]

(13b)는 동일한 낱말이 반복되는 경우로 이런 경우는 아주 특별한 경우를 제외하고는 같이 묶어 주는 것이 확률적으로 성공률이 높다고 할 수 있다. (13c)는 동의관계에 있는 것으로서 이들도 역시 자질 공유가 일어남을 알 수 있다.

(15) 고난 [n, experience]

꺾박 [n, experience]

(13d)는 상하의 관계를 나타낸 것이다. 이 역시 물론 서로 자질을 공유함을 알 수 있다.

(16) 왕실 [n, human, class]

귀족 [n, human, class]

(13e)는 [1]에서 포착하지 못한 것으로서 워드넷상에서 본다면 서로 자매 관계에 있을 법한 관계를 갖는 것들이다. 실제로 N1+N2의 관계에서 본다면 이러한 관계에 의한 구성이 많이 나타남을 알 수 있다. 다음은 이러한 부류의 예문이다.

(17) a. **우리나라와 중국**이 문화적으로 단절되었던 20세기는 중국뿐 아니라 동양 문화권의 각국이 서양 문화를 받아들이는 시대였다.

b. 한편에서는 **김동리와 김동석**의 순수문학 논쟁이 문단을 떠들썩하게 했다.

c. **아버지와 삼촌**은 늘 기쁨이 곁에 있었다.

위의 관계들은 [1]에서는 관계를 포착하기 힘들었던 예들이다. 그러나 이들도 자질로 나타내 본다면 다음과 같이 나타낼 수 있을 것이다.

(18) 우리나라 [n, nation] 중국 [n, nation]

김동리 [n, human] 김동석 [n, human]

아버지 [n, human] 삼촌 [n, human]

이처럼 이들의 관계를 자질로 포착해 본다면 이들을 처리할 수 있는 가능성이 열린다. 그러면 이러한 특성을 바탕으로 다음 절에서는 이들을 어떤 규칙 체계로서 처리할 수 있는지 보기로 하겠다.

3. '와/과' 구문의 연산

자연언어처리 분야에서 여러 가지 다양한 언어 이론들이 도입되었다. 최근 언어학에서 각광받고 있는 생성문법에서 출발한 HPSG, LFG, GPSG 등과 의존 문법 등이 많이 쓰이고 있는 언어 이론이다. 실제로 [2]에서는 '와/과'구문의 중의성을 해소하는데 있어서 통합기반 문법인 HPSG에 기반하여 처리 방안을 마련하였다. 그러나 이러한 모델들은 기본적으로 복잡한 구조를 상정하고 있기 때문에 유연성이 떨어진다는 단점을 가지고 있다. 그러나 최근 [3], [4] 등에서 제시되고 있는 자질 연산 문법 (FCG)는 이러한 단점을 보완하고자 하는 취지를 잘 보

여주고 있다. 여기서는 기본적으로 규칙 기반 문법을 지향하고 있으며, 각 항목들이 가지는 자질 연산을 통하여 자연 언어를 처리하고자 한다.³⁾

본고에서는 이러한 자질 연산 문법에 기초하여 위의 '와/과' 문제를 해결하고자 한다. 앞 장에서 '와/과'의 여러 가지 패턴을 살펴보고, 우리는 대충 다음과 같은 것으로 요약해 볼 수 있다.

(19) ① '와/과'를 논항으로 갖는 동사 자질:

[v, relation]

- ② 분리형 N1 + N2 : [n] + [n, action]
- ③ 분리형 N1 + N2 : N1과 N2의 의미 자질 상충
- ④ 분리형 N1 + N2 : [n] + [n, relation]
- ⑤ 결합형 N1 + N2 : [n + α] + [n + α]
- ⑥ N + 부사류 : [n] + [ad]

위의 유형이 앞 절에서 살펴본 '와/과' 출현의 모든 패턴들이다. 이들 패턴의 규칙을 잡기 위해서는 이들 자질을 매칭시키는 규칙의 순서를 정하는 것이 필요하다. 규칙의 순서는 가장 특징적인(specific) 규칙을 먼저 써야 한다. 따라서, (19⑥)의 규칙이 가장 먼저 적용되어야 한다. 그리고 두 번째로는 가장 가까이 있는 [v]의 자질을 확인하여 이 동사가 [v, relation]인지의 여부를 결정하는 것이 필요하다. 그리고 난 후에는 'N1+N2' 유형 중에서 묶어야 할 항목들과 묶지 말아야 할 항목을 결정하는 것이 필요하다. 이러한 규칙을 바탕으로 순서도를 작성해보면 다음과 같다.

(20) '와/과' 항목의 연산

※ ① : 두 항을 묶는다. ② : 두 항을 묶어주지 않는다

인접후행하는 [ad]가 있는가

1y : ②로 가라.

1n : 후행하는 첫 번째 [v]가 [v, relation]인가

2y : ②로 가라.

2n : [n]이 인접선행하고, [n]이 인접후행하는가

(※ [n] 인접후행 조건:[a]나 [ad]는 뛰어넘어라)

3y : 인접선행하는 [n]이 [n, action]이 아니고 인접후행하는 [n]이 [n, action]인가

4y : ②로 가라.

3) 김원경(2000)은 FCG의 특성을 잘 정리하고 있다. 첫째 변형 규칙과 수형도의 개념을 사용하지 않고, 둘째로 가장 중요한 개념으로 '자질'을 쓰며, 세 번째로 산출의 안정성과 경제성을 위해 잉여성을 필요한 개념으로 인정한다고 보고 있다.

4n : 인접후행하는 [n]이 [n, relation]인가

5y : ②로 가라.

5n : 인접선행하는 [n]과 인접후행하는 [n]의 의미 자질이 일치하는 것이 있는가

6y : ①로 가라

6n : ②로 가라.

3n : ②로 가라.

위의 연산을 통하여 우리가 '와/과' 패턴으로 잡은 문장들을 거의 해결할 수 있다. 위의 (※)의 부가 조건은 명사를 수식하는 요소들을 제외하기 위한 조건으로 '[a]+[n]'의⁴⁾ 구성일 경우에는 인접후행하는 것으로 본다는 것이다. 그러나 이러한 연산만으로 문제가 되는 것이 있다. 첫 번째 문제는 인접후행하는 [ad]자질 중에서 '와/과 함께'류와는 다른 다시 말해서, 실제로 후행하는 요소를 수식하는 부사가 있을 수 있다는 것이다. 따라서, 이러한 부사류에는 역시 [relation]의 자질을 주어서 '와/과'와 함께 나올 수 있는 부류를 묶어주는 것이 필요하다. 따라서 후행하는 [ad]를 [ad, relation]으로 바꾸어 주어야 한다. 두 번째 문제는 [n, relation]이라는 자질이 상위에 위치해 있기 때문에 묶일 수 있는 것 중에서 [n, relation]의 자질을 가질 수 있는 것이 있다. 예를 들어 다음의 문장을 보자.

(21) a. 오빠와 아버지는 목욕탕에 가셨다

b. 나는 그와 친구가 되었다.

위의 규칙대로라면 (21a)의 문장은 '오빠'와 '아버지'가 [human] 자질을 공유하고 있지만, 두 명사 모두 [relation]의 관계이기 때문에 위의 분석에서 오류가 나게 된다. (21b)의 문장은 성공적인 문장이긴 하지만, 이러한 경우의 예문을 좀 더 자세히 들여다본다면 위의 경우가 매우 제한적으로 쓰이고 있음을 알 수 있다. 즉, [n, relation] 다음에 나오는 동사가 '되다'류에 한정되기 때문이다. 따라서 위의 분석이 성공률을 높이기 위해서는 좀 더 많은 제약을 주어야 함을 알 수 있다. 즉, 동사 정보를 덧붙여야만 한다. 이를 반영한 알고리즘이 다음과 같다.

(22) '와/과' 항목의 연산

4) 여기서 [a]라고 하는 것은 명사를 수식하는 관형사를 말하는 것이다.

※ ① : 두 항을 묶는다. ② : 두 항을 묶어주지 않는다

인접후행하는 [ad, relation]이 있는가

1y : ②로 가라.

1n : 후행하는 첫 번째 [v]가 [v, relation]인가

2y : ②로 가라.

2n : [n]이 인접선행하고, [n]이 인접후행하는가

(※ [n] 인접후행 조건 : [a]나 [ad]는 뛰어넘어라)

3y : 인접선행하는 [n]이 [n, action]이 아니고 인접후행하는 [n]이 [n, action]인가

4y : ②로 가라.

4n : 인접후행하는 [n]이 [n, relation]이고, 후행하는 [v]가 '되다류'인지 확인하라.

5y : ②로 가라.

5n : 인접선행하는 [n]과 인접후행하는 [n]의 의미자질이 일치하는 것이 있는가

6y : ①로 가라

6n : ②로 가라.

3n : ②로 가라.

위의 규칙을 '되다류'라고 한 것은 좀 더 많은 예문이 발견되었을 때, 이는 확장가능하기 때문에 일단 '되다류'로 명시했으며, 이러한 동사류가 하나의 자질로 묶일 수 있는 가능성을 열어 놓기 위해서 광범위한 표현을 하였다. 그리고 '의미자질'을 확인하는 부분에서 문제가 있을 수 있는데, 기본적으로 FCG의 자질은 계층적으로 구성된다기 보다는 선형적으로 구성되며, 이들이 동등한 자격을 갖는 것으로 이해한다. 그러나 여러 자질 목록들이 모두 다 결정된다면 이들 자질의 종류에 따라서 범주자질, 의미자질, 기타 여러 자질들로 나누어서 컴퓨터가 처리하도록 정보를 줄 수 있기 때문에, 이러한 진술은 FCG의 기본정신에도 부합하는 것이며, 처리시에도 아무런 문제를 야기하지 않으리라 생각된다. 위의 순서도를 보면 하나의 경우만을 제외하면 모두다 묶지 않는 경우가 되고 있음을 보여주는데, 그렇다면, 묶는 경우만을 특징적으로 보여주면 훨씬 간단해질 수 있으리라는 반론이 가능하다. 그러나 [n]항목이 인접후행하는 경우라 하더라도 동사의 특징 때문에 이들을 먼저 잡아 주어야 하는 경우가 생기기도 하고, 또 앞으로 더 많은 자료를 통해서 위의 알고리즘을 수정하기 쉽게 하도록 개방적 체계를 마련해 둔 것이다. [n] 앞에 수식어가 붙는 문장들을 모두 고려하였을 경우 더 복잡해질 수 있으며, 여기에서 마련한 규칙은 가장 기본적인 뼈대가 되는 것이다.

그러면 이러한 연산을 바탕으로 실제로 얼마만큼의 처리가 가능한지를 테스트해 보도록 하겠다.

4. 실험 및 결과

위의 연산을 바탕으로 실제 실험을 하기 위해서 150만 어절 코퍼스에서 뽑은 '와/과'가 나오는 문장 10,021개 중에서 500 문장을 무작위로 뽑아서 그것의 성능을 테스트 하였다. 여기서 실제 컴퓨터 상에서 테스트를 해 보는 것이 가장 이상적이지만, 그러기 위해서는 전체 구문 분석의 엔진이 뒷받침되어야 하는데, 아직 그 단계까지 작업이 진행되지 못한 관계로 실제 자료를 수작업을 하여, 하나 하나 자질을 대조하면서 테스트를 진행하는 방법을 택하였다. 위의 연산의 테스트 결과는 다음과 같다.

<표1>

문장수	오류문장	성공문장	성공률
500	25	475	95%

위의 결과표에서도 알 수 있듯이 위의 알고리즘만으로도 95%의 높은 성공률을 보여주었다. 오류가 난 25문장은 어떤 패턴을 보여주는 문장들이 있었다. 우선 복합명사와 관련해서 보여주는 오류이다.

(23) 자동휴식기능 내장 하이콤도 '개구장이 까치'를 수출용으로 새롭게 보완해 일본 세가사와 라이선스 계약을 체결키 위해 협의중이다.

위의 문장에서 '세가사와 계약'이 묶인다면 '계약'에 [n, action]자질로 묶일 가능성이 있지만, 여기서는 인접후행하는 [n]을 찾게 되므로 '세가사와 라이선스'가 묶이게 되어 오분석될 가능성이 있다. 만약 이 두 단어가 미등록어라면 미등록어에 배당되는 자질이 각각 배당되므로 묶일 가능성이 있다. 이는 복합명사 처리 순서를 '와/과' 처리보다 먼저 시행한다면 해결될 가능성이 있다.

다음은 다른 의미 자질을 가지고 있다고 하더라도 묶여서 나오는 유형이다.

(24) 사실 마르쿠제는 이러한 자기승화의 상태는 사회와 인간의 전면적 개조가 없는 달성될 수 없는 것이라고 말한다.

위의 예문에서 '사회'와 '인간'은 의미 자질을 공유하지 않을 뿐만 아니라 [action]이나 [relation]의 자질과도 관련이 없기 때문에 이들은 묶이지 않는다. 그러나, '인간과 사회'는 언어적 관계에서 많이 묶이는 부류들이기 때문

에, 만약 언어 정보를 이용한다면 해결가능할 수도 있을 것이다.

그 다음의 문제는 여러 가지 면에서 해결하기 어려운 부분이다. 계속 이 논의에서는 선행 명사와 후행 명사와의 동질성이 이들을 묶게 하는 것이라는 논리를 펴왔다. 따라서 동일한 항목이 오는 경우는 반드시 묶이게 되어 있다. 그러나 동일한 항이 인접해서 온다고 하더라도 그것이 묶여서는 안 되는 경우가 있다.

(25) a. 원근법은 한편으로 그 안에서의 개개의 위치와 그 위치에서 가능한 시계(視界)의 통일이 객관적으로 규정된다는 것을 말하지만 다른 한편으로는 일정한 시점에서 있을 수 있는 시계의 통일은 상당히 자유롭게 선택될 수 있다는 것을 말한다.⁵⁾

b. 사랑은 매우 중요한 동기이지만, 그 형태와 형태의 완성을 위한 사회적, 제도적 표현이 전혀 고정되어 있지 않은 것으로 보인다.

위의 두 문장은 동일한 단어가 나온다고 하더라도 서로 묶어서는 안되는 문장들이다. 이는 현재로서는 처리하기 힘든 문제로 생각이 되며, 문장 전체의 구조를 좀더 세밀히 분석했을 때 이런 문장도 해결할 수 있으리라 생각한다.

이러한 문제가 있다고 하더라도 간단한 규칙과 몇 개의 자질로 '와/과' 패턴에서 보이는 여러 가지 분석적 중의성 문제를 어느 정도 해결할 수 있었다. 실제 언어처리에 있어서는 좀 더 복잡한 분석이 이루어져야 하리라 생각되지만, 여기서의 기본틀이 그대로 이용될 수 있으리라 생각된다.

5. 결론 및 남은 문제

'와/과'구문 중에서 일부 분석적 중의성의 해결 방법을 기술하였다. 이 연구는 자질연산문법(FCG)에 의한 규칙 기반으로 이루어졌다. 여기서의 작업은 현재 진행 중인 한국어 분석기를 위한 기초 연구의 일환으로 이루어졌고, 앞으로 계속 연구를 지속할 생각이다. 아직 '와/과'에 대해서도 묶어 주거나, 묶지 않은 후의 처리나, 더 많은 수식어가 붙을 수 있는 여러 가지 문장들을 처리해 주어야 하는 과제들이 남아 있긴 하지만, 좀 더 많은 연구들

을 통하여 미진한 부분은 후고를 기약하기로 한다.

■ 참고문헌

- [1] 한광익. 1998. {-와/-과}로 구성된 명사구의 의미 연구. 고려대학교 석사학위논문.
- [2] 서영훈. 1991. 의미 정보를 이용하는 중심어 주도의 한국어 파싱. 서울대학교 박사학위논문.
- [3] 고창수. 1993. "인공지능을 위한 국어 문법의 기술." 『어문논집』(고려대) 32.
- [4] 김원경. 2000. 한국어 격정보의 자질 연산. 고려대학교 박사 예비발표 발표문.
- [5] 황도삼외역. 1998. 자연언어처리. 홍릉과학출판사.(Makoto Nagao 저)
- [6] H.L.Somers. 1987. *Valency and case in computational linguistics*. Edinburgh University Press.

5) 여기서는 후행하는 [n] 앞에 '그'라는 관형사가 나오기는 하지만 앞에 조건에서 관형사가 인접후행하고 그 관형사를 [n]이 인접후행한다면 인접후행하는 것으로 부가 조건을 주었기 때문에 여기에 걸리게 된다.