

## 검색 엔진을 위한 '백과 명사' 전자 사전의 구축 (I) < 인명 관련 백과 명사의 연구 >

남 지순, 이 주호, 최 기선

한국 과학 기술원 인공 지능 연구 센터  
대전시 유성구 구성동 373-1, 우:305-701

nam@world.kaist.ac.kr / leejh@csone.kaist.ac.kr / kschoi@world.kaist.ac.kr

### Korean Electronic Dictionary of Encyclopedic Nouns I. - Nouns related to Human Names -

Jee-Sun Nam, Ju-Ho Lee, Key-Sun Choi  
CAIR KAIST

#### 요 약

정보 검색 시스템에서 가장 문제가 되는 어휘 클래스는, 소위 '고유 명사'와 '합성 명사'로 분류되는 명사 유형이다. 이들 클래스는, 기존 대사전 및 전자 사전 (MRD)류에서, 그 어휘 목록을 체계적으로 제공하지 못하는 가장 대표적인 부류들인데, 실제 검색 시스템에서는 많은 경우 정보의 핵심어 (Key Word)가 된다. 본 연구에서는 신문, 잡지등 시사 문서류에서 가장 빈번히 발견되는 명사 유형의 하나인, '인명 관련 고유 명사' 유형에 대한 문제에 그 논의의 초점을 두고, 이들 명사들의 체계적인 처리를 위해서 어떠한 형태로 사전을 구성해야 하는지를 검토할 것이다. '고유 명사'라는 개념 자체가 지니고 있는 외연적 정의상의 문제점을 극복하기 위해서 우리는 '백과 명사 (Encyclopedic Noun)'라는 용어를 사용하기로 하며, 이는 좁은 의미의 고유 명사 및, 전문어, 고유 명사 관련 파생-복합어류등을 포함하는, 보다 확장된 개념으로 이해되어야 한다. <인명> 관련 백과 명사류의 하위 유형 분류 및 그 특징적 결합어 형태(Appropriate Particle)에 대한 연구 결과들이 소개된다.

#### 1. 문제의 제기

본 연구<sup>1</sup>는, '한국어 전자 사전 (Korean Electronic Lexicon)' 구축을 위한 일환으로, 정보 검색 시스템에서 가장 문제가 되는 유형의 하나인 '백과명사 (Encyclopedic Nouns)'류 집합에 대한 하위 유형 분류 및 그 사전의 구성에 목표를 둔다. 검색 엔진에서 사전 미등록어로 발견되는 대부분의 형태들이 백과 명사류와 합성명사류인데, 실제 검색 시스템에서는 많은 경우, 바로 이들이 정보의 핵심어(Key Word)가 된다. '백과 명사'에 대한 이 연구의 결과를 통해, 우리는 다음과 같은 효과를 기대할 수 있다.

- 백과명사 사전의 체계적인 완성을 위한 기초 작업
  - 기본 대분류 및 하위 유형 분류
  - 각 하위 유형을 구성하는 기본 어휘데이터 구축
  - 아직 등재되지 않은 백과명사류의 문맥 추정

명사를 분류하는 방법은 그 분류 각도에 따라 여러 가지가 가능하다, 전산 처리를 위한 사전 구축의 관점에서 고려할 때, 가장 우선적으로 고려되어야 할 사항은 다음의 두 가지 분류 형태이다.

- '보통 명사'와 '비보통 명사'의 구분
- '단순 명사'와 '비단순 명사'의 구분

위의 같은 분류가 우선적으로 요구되는 이유는, 무엇보다도 '비보통 명사' 및 '비단순 명사'로 판단되는 명사들의 목록이 현행 사전들에 체계적으로 수록되어 있지 못하기 때문이다. 이러한 상황은, 모든 자동 처리 시스템에서 첫 단계로 거쳐야 하는 '사전 검색' 단계의 취약성을 유발시키는 가장 큰 원인의 하나가 된다. 위의 두 유형들의 예를 들어 보면 다음과 같다.

- '황제, 정치가' vs '나폴레옹 1 세, 빌 클린턴'
- '관공서, 집단' vs '합자 회사, 전문 대학'

우리가 여기서 '비보통 명사', '비단순 명사'로 정의하고 있는 부류들은, 전통적으로 각각 '고유 명사 (Proper Noun)', '합성 명사 (Complex Noun)'류로 불리어 온 형태들이다.

#### 1.1. 백과 명사 (Encyclopedic Noun)

'보통 명사 (Common Nouns)'류와 구분되어야 할 것으로 판단되는 일정 명사류의 집합을, 보통 통용되고 있는 '고유 명사 (Proper Nouns)'라는 개념으로 총칭하기에는 부적절하다.

일반적으로, '세계에 존재하는 유일한 지시 대상물을 지칭하는 하나의 음운 연쇄'로 정의되는 '고유 명사'라는 개념은, 실제 화자들이 이 개념에 대하여 가지고 있는 언어적 직관과 반드시 일치하지 않는다. 가령 '화성', '토성'등과 같은 언어 형태들이, 유일한 대상을 가리키는 하나의 고유 명사로 판단된다면, '달', '태양'등은 이와 같은 의미적 속성을 공유하면서도, 보통 명사류에 속하는 형태로 취급된다. 고유 명사에 대한 개념적 정의와 언어 표현들에 대한 언어 습관이 일치하지 않기 때문이다.

또한, '프랑스', '프랑스어', '이탈리아', '이탈리아어'등이 고유 명사로 분류된다면, 분명 유일 개체를 지칭하는 명사로 볼 수 없는 '프랑스인', '이탈리아인' 등은 고유 명사류인가, 그렇지 않은가? 마찬가지로, '기독교', '불교'와 같은 명사가 고유 명사류가 되어야 할지, 보통 명사가 되어야 할지는, 그와 같은

<sup>1</sup> 본 연구에는, <다국어 정보 검색 시스템용 전자 사전의 구축> 및 <지능형 멀티미디어 통합 정보베이스> 과제 수행을 위한 정보 통신부 연구비 지원을 바탕으로, 한국 전자 통신 연구원(ETRI)과의 공동 연구로 진행된 연구 결과가 반영되었음을 밝혀둔다.

종교가 다시 여러 분파로 나뉘게 되면, 그 판단이 모호해 진다.

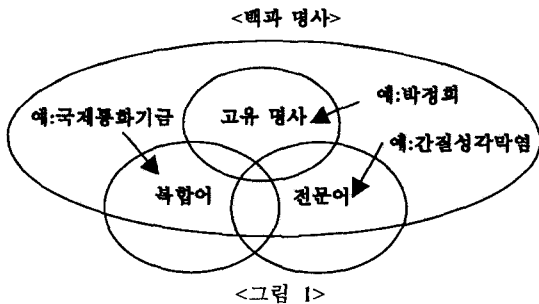
동식물 이름중에는, ‘도둑놈의 갈고리’, ‘키카데오이데아’ 등과 같은 학명들이 존재하는데, 이들이 보통 명사와 다른 속성을 가진 일종의 ‘고유 명사’ 류로 처리되어야 할 것 같은 직관적 느낌은, 이들을 ‘백합’, ‘장미’와 같은 일반 꽃이름들과 비교할 때, 어떠한 기준에 의해서 과연 전자들만을 고유 명사로 처리할 수 있을까 하는 의문을 제기한다.

이러한 문제들을 고려해 볼때, 전통적 개념의 ‘고유 명사’라는 용어는, 현행 사전들에 누락된 많은 유형의 명사 엔트리를 체계적으로 구축하고자 하는 우리들의 작업에 있어서, 그리 만족스럽지 못하다. 우리가 목표로 하는 작업의 대상은, 오히려 ‘백과사전적 정보를 담고 있는 용어들’이라 할 수 있으며, 사실, 이와 같은 이유로 대부분의 현행 언어 사전속에 체계적으로 총망라한 어휘 목록이 구성될 수 없었던 것이다.

우리는 앞으로 이 연구에서 ‘백과명사 (Encyclopedic Nouns)’라는 용어를 사용하기로 하며, 이는 대체로 전통적 개념의 ‘고유 명사’류보다 더 확대된 의미로 해석되어야 한다. 즉, 이 명사들은 다음과 같은 전통적 유형들을 내포하는 개념이다.

- 유일 개체를 지칭하는 **고유 명사 유형**
- 이러한 고유 명사류를 내포한 **파생복합어 유형**
- 전문적 분야에서 사용되는 **전문어 유형**

다음 <그림 1>에서 보이는 바와 같다.



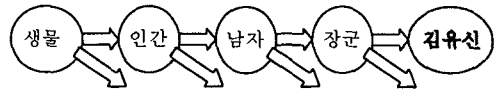
<그림 1>

그런데, 우리가 일컫는 ‘백과 명사’류라는 개념이, ‘보통 명사’류와 명시적으로 구분될 수 있는가 하는 문제는 사실상 그대로 남는다. 그러나, 여기서 우리는 ‘백과 명사’라는 부류를 어떤 일반적 정의에 의해 이론적으로 규정하지 않고, 이 부류에 속하는 명사의 유형 목록을 귀납적으로 제시함으로써, ‘백과 명사’라는 부류를 정의하고자 한다. 다시 말해, 이것은, 현행 사전류에서 비체계적으로 다루어지고 있는 모든 어휘 목록들을 체계적으로 점검하고 완성하는데 연구의 목표가 있기 때문에, 최대한으로 확장된 개념의 ‘백과 명사’류를 설정하는 것이 우선은 필요하기 때문이다.

따라서, ‘보통 명사’류와의 경계 설정은, 분야에 따라서는 상당히 자의적인 방법으로 이루어질 것이며, 이와 같은 경계의 구분은, 언어 이론적 논의의 대상이 아니라, 작업 과정상 잠정적으로 거쳐야만 하는 중간 단계의 개념으로 이해되어야 한다. 만일 모든 ‘백과 명사’류의 목록이 완결되었다고 가정한다면, ‘보통 명사’류와의 ‘이분적 대립 관계’의 설정은 더이상 필요하지 않게 된다. ‘유일 개체’를 지칭하는 속성은, 명사의 계층 구조상에서 단지 가장 말

단에 놓이는 단계일 뿐으로, 이들 계층 구조는 이와 같은 ‘유일’과 ‘비유일’의 이분 관계가 아니라, 하나의 연속된 다층 단계를 보이기 때문이다.

가령, ‘김유신’은 보통 명사에 대립되는 하나의 백과 명사가 아니라, 다음 네트워킹에서 보듯이 ‘생물 > 인간 > 남자 > 장군 > 김유신, 등...’과 같은 계층 구조속에서 마지막 단계를 보이고 있는 개념어이며, 이러한 명사 계층 구조는 얼마든지 다른 방법으로 더 자세히 분할될 수 있다는 것을 상기해야 한다.



<그림 2>

## 1.2. 합성 명사 (Complex Nouns)

모든 유형의 ‘비단순 명사’들은 넓은 의미의 ‘합성 명사’들을 가르킨다. 여기서, ‘합성어’라는 개념을 사용하기 전에, 다음과 같은 두가지 문제를 선결해야 한다.

첫째는, 현재 혼용되어 쓰이고 있는 ‘합성어’, ‘복합어’, ‘파생어’ 등의, ‘용어상의 혼란’뿐 아니라, 그들 사이의 분류가 실제적으로 매우 불분명한 경우가 많다는 점이다. 가령, ‘은세계’는 하나의 파생어인가, 복합어인가? 즉, ‘은’을 접두사로 간주하면, ‘파생어’의 하나가 될 것이고, 관형사로 간주하면, 두개의 자립 성분이 굳어져 한 덩어리로 실현된 ‘복합어’의 하나가 될 것이다.

‘실뿔줄’과 같은 경우, ‘실’을 접두사로 간주하면 하나의 파생어가 되나, 명사로 간주하면, 의미의 전성이 일어난 하나의 복합어가 된다. ‘비계살’, ‘구김살’과 같은 경우도 마찬가지이다. 이때, ‘살’은 접미사로 간주될 수도 있으며, 또는 명사로서 다른 명사와 결합하여 뜻이 굳어진 것으로 분석될 수도 있다.

둘째는, ‘합성 명사 (Complex Noun)’와 ‘명사구 (Noun Phrase)’의 구분 문제이다. 대체로, 의미 전성이 발생하는, 굳어진 복합 형태를 ‘합성 명사’로 간주하고, 비교적 자유로운 결합 관계를 보이는 형태를 ‘자유 명사구’로 분석하는데, 문제는 의미의 전성이 그렇게 명시적이지 않은 경우가 훨씬 더 많다는 것이다.

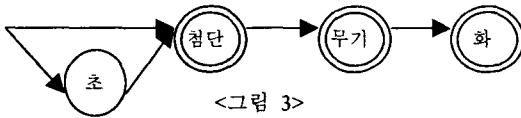
가령, ‘고무장갑’, ‘고무줄’ 등은 하나의 합성 명사로 분류되어야 하는가, 또한 ‘고무끈’, ‘고무지우개’, ‘고무제품’과 같은 형태는 어떠한가? ‘유령 회사’는 ‘주식 회사’에 비해 그 의미 전성의 정도가 더 심한 것은 사실이나, 그렇다고 ‘주식 회사’가 아주 자유로운 결합의 하나라고 간주할 수는 없다.

언어학적으로 이것들이 어떻게 처리되건간에, 실제 한국어 문서를 처리할 때, 이와 같은 복합 형태들을 하나의 단위로 인식하지 못하면, 명사 키워드 (Key Word) 구축시 상당히 많은 잉여적 정보 및 부적절한 정보들을 포함하게 될 뿐아니라, 정작 중요한 정보를 갖추고 있는 올바른 색인어들은 누락되고, 일반 단순어들이 별 의미없이 키워드로 등재되는 경우들이 발생한다.

자동으로 인식할 수 있는 시퀀스 (즉 사전의 기본 정보 단위)의 유형이, 복합형으로 확대되면 될수록, 중의성의 발생률은 현저히 저하하며, ‘진정량의 의미의 키워드 (즉 전문어, 백과 명사 등 정보량이 많은 어휘)’가 될 확률은 증가하게 된다. 따라서, ‘단순 명사’와 ‘비단순 명사’의 경우에도, 이와 같은

이분된 대립 관계보다는, 단순 명사는 명사의 복합 구성의 계층 구조상에서 가장 말단의 하위 구성 성분으로 분석하는 것이 더 합당한 것으로 보인다.

단순 명사로부터 생성될 수 있는 복합체의 형태들은, 몇가지 예를 들면 다음 전이 그래프에서 보이는 바와 같다.



<그림 3>

그러므로, 여기서 일컫는 '비단순 명사'류는, 단순 명사 형태를 제외한 모든 복합체를 포함하는 개념으로 이해되어야 하며, 사전 구축시 백과 명사류와 함께 가장 문제가 되는 어휘 목록을 구성한다는 점을 주목해야 한다. 이 부류에 대한 더 자세한 논의는 다음 기회로 미루기로 한다. 이 연구에서는 '백과 명사'의 문제에 초점을 맞추고, 다음 장에서 좀 더 세부적으로 살펴보기로 한다.

2. 백과 명사의 유형 분류

백과 명사의 유형은 다음과 같이 대분류된다.

<1>	인명 및 단체 관련 백과 명사류
<2>	지명, 자연 및 유적 관련 백과 명사류
<3>	날짜 및 역사 사건 관련 백과 명사류
<4>	예술, 사상 및 문화 관련 백과 명사류
<5>	병명 및 의학 증상 관련 백과 명사류
<6>	제도, 사상 및 법률 관련 백과 명사류
<7>	발견 및 발명, 기물 관련 백과 명사류

<표 1>

위의 각 유형들에 대한 예를 들어 보면 다음과 같다.

<1>인명 및 단체	이승만/일본인/감사원
<2>지명, 자연 및 유적	프랑스/태백산맥/만리장성
<3>날짜 및 역사 사건	강화도조약/ 현충일
<4>예술, 사상 및 문화	아상곡/ 선사상/ 지동설
<5>병명 및 의학 증상	간세포암/ 결합중양
<6>제도, 사상 및 법률	계엄법/ 관리통화제도
<7>발견, 발명 및 기물	그로브전지/니콜리튬

<표 2>

위에서 알 수 있듯이, '백과 명사'류는 전통적으로 고유 명사로 간주되어온 형태들을 대부분 포함함은 물론, 전문적 지식 정보를 담고 있는 분야별 전문어들을 상당 부분 함유하고 있으며, 이들 형태는 단일어는 물론, 파생어, 복합어 구성 형태를 취하는 경우가 빈번하다. 여기서 <1>의 인명 관련 명사류는, 특정 전문 분야 문서가 아니더라도, 일반 잡지나 뉴스, 신문등의 출판물에서 가장 많이 출현하는 '미등록어'의 유형을 이룬다.

본 연구에서는 <1>의 유형에 초점을 맞추어, 과연 어떠한 원칙에 의해서 이들 명사들에 대한 체계적인 전자 사전을 구축할 수 있는지에 대하여는 의해 보기로 한다.

3. <인명> 관련 백과 명사 사전의 구성

현행 한국어 대사전류에 등재된 '인명 관련' 백과 명사 유형은 다음과 같이 하위 분류된다.

인명 관련 백과 명사 (코드 HUM)	
<3-1>	개인 인명 (Indiv)
<3-2>	유도된 인물 명사 (DerHum)
<3-3>	단체 및 인물 집단 (Collect)

<표 3>

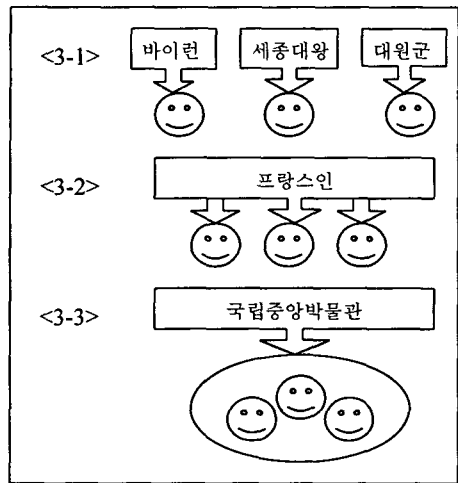
위의 3클래스는 각각 다시 '로마자 한글 표기'를 내포한 형태 (코드 ENG)와 '한자어 유래 형태 (코드 CHN)'로 나뉜다. 예를 들면 다음과 같다.

HUM			
	CHN	ENG	
<3-1>	Indiv	세종/대왕	아인슈타인
<3-2>	DerHum	만주/족	프랑스/인
<3-3>	Collect	국립/박물관	라이온즈/클럽

<표 4>

위와 같은 <CHN>과 <ENG> 클래스가 분리된 것은, '로마자 한글 표기'의 경우, '한자어 유래 형태'와는 달리, 그 표기법에 있어서, 실제로는 여러가지 형태가 발견되기 때문이다. 가령 프랑스 대통령 'Jacques Chirac'의 표기는, '작크 시락', '작끄 시락', '작끄 쉬락', '작끄 시라크' 등 여러 형태가 가능하다. 따라서 이러한 이형태 표기들이 잠정적으로 허용될 <ENG> 클래스의 설정이 요구된다.

위의 <표 4>에서, <3-1>은 대체로 '언어 표현 (Linguistic Expression)'과 '지시 대상물 (Extralinguistic Referent)'이 '1대1'의 대응 관계를 보이고 있는 반면, <3-2>와 <3-3>은 이와는 다른 대응 관계를 보이고 있다. 즉, 다음 <그림 4>와 같이 설명될 수 있다.



<그림 4>

위의 그림은, 각 클래스의 명사들을 도식화해서 보여주지만 모든 엔트리들이 반드시 위의 그림과 일치하는 것은 아니다. 가령, <3-2>에 속하는 '프랑스인'은 위의 그림처럼 여러 개체와 대응될 수 있으나, '대법원장'의 경우는 국가 조직상 단 한사람만이 존재한다. 따라서 <3-2>의 그림은, 이 클래스가 기타 다른 백과 명사류와는 다르게, 여러 개체와 대응될 수 있다는 '가능성'을 보여주는 것으로 해석되어야 한다. <3-3>의 그림은, 그 지시대상물로서 '인명'이 둘이상의 집단으로 구성되어 있어야 함을 보인다.

(제 10 회 한글 및 한국어 정보처리 학술대회)

이 세 클래스들은, 통사적으로 모두 '인물성(Human Feature)'의 자질을 가진 명사류로서, 일반 인물성 명사들, 가령 '학생', '소년'등과 같은, 소위 전형적인 '인물 보통 명사(Human Common Nouns)'류와 매우 유사한 성질을 보인다. 즉 모두, 일정 문맥 속에서 '누구?(Who?)' 의문사에 대응될 수 있는 통사적 자질을 가지고 실현될 수 있는 유형이다.

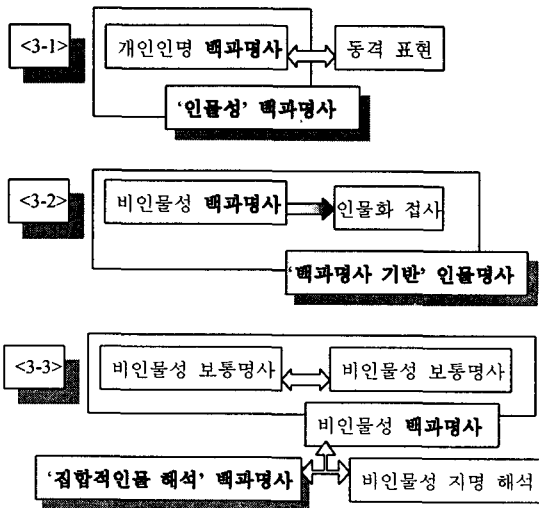
- (1) 누가 그 일에 개입했는가?
- (2) (학생들+김철수+조선족+고려대학교)가 개입했다

위와 같은 통사적인 검증은, 특히 <3-3 Collect> 유형을 일반 '지명 관련' 명사류와 구별시켜 주는 중요한 기준이 된다. 가령, '감사원', '롯데백화점'과 같은 경우, 그것이 하나의 장소를 나타내는 지명 관련 명사류로 간주되어야 할지, 인물 집단의 의미로 실현된 일종의 단체 인명 관련 명사류로 간주되어야 할지 문제가 된다. 그런데, 이들은 문맥에 따라 이와 같은 두 가지 의미적 해석을 모두 허용하는 형태들로서, 가령 순수 지명 관련 명사류인 '독립문', '다보탑'등과는 바로 이점에서 구별된다. 예를 들면 다음과 같다.

누가 그 일에 개입했다?	[인명]
감사원에서 그 일에 개입했다	
*독립문에서 그 일에 개입했다	
이 건물은 어디에서 가깝다?	[지명]
이 건물은 감사원에서 가깝다	
이 건물은 독립문에서 가깝다	

<표 5>

이와 같이 '인물성'의 자질을 공통적으로 가지는, 위의 세가지 유형의 '인명 관련 명사'류는, 형태적 구성 관점에서 살펴보면 다음과 같이 분석된다.



<그림 5>

<3-1>유형은, <개인 인명>을 나타내는 백과 명사류가 단독으로 실현되거나, 직책이나 신분등을 표현하는 동격의 특정 명사류(가령 '법사', '대왕', '공주'등)가 결합하여 형성된 표현들로서, 본질적으로 '인물성'을 지닌 '백과 명사'류이다. 단, 여기서 행정 기관 및 사회 단체, 학교, 회사등의 직책을 나타내는

명사류(가령 '회장', '장관', '교수', '전무', '총장'등)가 결합하는 형태는 별도로 처리한다. 뒤에서 다시 논의된다.

<3-2> 유형은, 본질적으로 '비인물성'을 가진 백과 명사류에, 인물성의 자질을 부여하는 특징어(가령 '프랑스인'이나 '감사원장'등에서 나타난 '인'이나 '장'와 같은 접사류와 '코카서스인종'에 나타난 '인종'과 같은 명사류)가 결합하여 '인물성' 명사로 유도된 형태들을 가르킨다. 이와 같이 유도된 명사들은 대부분 '백과 명사'류로 판단되기 어려운 경우들이나, 이들 목록의 체계적인 구축을 위해서는 이와 같은 프로세싱을 허용하는 '비인물성 백과 명사'류의 목록이 완결되어야 하므로, 현실적으로 '백과 명사' 사전에서 다루어져야 하는 형태들이다. 이 부류에서도, 위에서 제외시킨 '회장', '장관'등과 같은 명사류가 결합한 합성어는 별도로 처리한다. 5장에서 다시 언급된다.

<3-3> 유형은 대개, 본질적으로 '비인물성'인 '보통 명사'들이 여러개 결합하여 이루어진 복합어의 형태를 취하고 있으며, 이때 이 보통 명사들의 결합체는 비인물성 '백과 명사'류가 된다. 이들 결합체가 <3-2> 유형처럼 본질적 '인물 명사'류로 전환된 것이 아니라, 문맥에 따라 '집합적 인물 명사'의 해석을 받을 수 있다는 점에서 특징지어진다. 이들은, 앞서 언급한 바와 같이, 대부분의 경우, 비인물 백과 명사류로서 '지명'에 관련된 해석을 받을 수 있다. 그러나, 뒤에서 보게될 <지명> 관련 백과 명사류와는 달리 이와 같은 <인명> 해석을 받을 수 있는 속성은, 구문 연구에서 중요한 통사 정보를 제공할 것이다. 이들은, '국민', '대중'등과 같은 보통 명사류처럼, 반드시 '집단적 인물성'의 속성을 가진다.

백과 명사류는 대체로 특정 개체를 지칭하므로, 복수의 표지나 수량사의 삽입이 어렵다. 그러나 이와 같은 검증은, 백과 명사류(전통적 개념의 고유명사의 경우에도)를 결정적으로 구별시켜 주는 절대적 지표는 되지 못하며, 특히 <3-2> 클래스류의 경우, 위의 검증에 대해서, 보통 명사류와 매우 인접해 있는 것을 관찰할 수 있다.

인물성명사	수량사 동반	복수 표지
학생	다섯 학생	학생들
김유신	*다섯 김유신	*김유신들
프랑스인	다섯 프랑스인	프랑스인들
더블류비시	*다섯 더블류비시	*더블류비시들

<표 6>

마지막으로 한가지 지적할 점은, '백과 명사'의 대부분의 형태가, 단순어'가 아닌, 복합어, 파생어, 또는 파생복합어 형태를 취하는 일종의 '합성어'들이라는 점이다. 가령 <3-1>의 개인 인명 명사류의 경우에도 단순히 '성+이름'의 형태로 실현된 경우들 보다는 다음과 같은 합성 형태들이 더 많이 눈에 띈다. 몇가지 예를 들면 다음과 같다.

(단순형)	신 율복
<법사> 결합	강장/법사
<대제> 결합	콘스탄티누스/대제
<여왕> 결합	선덕/여왕
<세자> 결합	사도/세자
<대감> 결합	천륜/대감

<표 7>

‘신윤복’과 같이 단순어 형태로 실현된 개인 인명 명사들의 경우는, 국내의 전화번호부등을 참고하면 쉽게 기본 목록을 구성할 수 있다. 그러나, 전세계 인구수만큼 이를 규모의 개인 인명 사전을 갖출다는 것은 현실적으로 별 의미가 없다. 정보 검색 시스템이나 기계 번역 시스템등에서 다루어져야 하는 문서에 있어서, 주로 등장하는 개인 인명 명사류는 무엇보다도, 정치가, 예술가등 역사적으로 유명했거나, 세계적으로 인지도가 높은 인사들이 우선적으로 등장하기 때문에 백과 대사전류에서 다루고 있는 인명들을 우선적으로 처리하는 것이 효율적이다.

이와 같은 단순 형태가 아닌 경우들에 있어서, 각 클래스별 하위 유형을 분류한다는 것은, 결국 각 클래스에 특징적으로 결합하는 ‘본연적 형태소 (Appropriate Particle: 앞으로 AP 로 표기)’의 목록을 체계적으로 구성한다는 것을 의미한다. 개념적, 의미적 자질에 의해서만 명사의 하위 속성을 분류한다는 것은, 객관성과 명시성을 유지하기 어렵기 때문에, 실제 결합된 개별 ‘형태소(또는 단어)’들의 구체적 목록에 의해, 이들 백과 명사류들의 유형을 분류하는 것이 필요하다.

이와 같은 작업은, 의미적 직관을 형식적으로 검증할 수 있게 만들어 주며, 실제 자동화된 시스템에서 유용하게 사용될 수 있다. 가령 기계 번역 시스템에서, 위에서처럼 <여왕>등이 결합한 백과 명사 시퀀스들은 모두, ‘번역되지 않는 부분’인 순수 백과 명사 부분 (예: 선덕, 엘리자베스)와, 영어 표기 ‘Queen’으로 ‘번역될 수 있는 부분’, 이렇게 두 부분의 결합으로 이루어진 형태들임을 위의 사전 정보를 이용해서 쉽게 찾아낼 수 있다.

그러면, 다음에서, 이들 3 클래스의 내적 결합 구성 관계를 좀더 자세히 살펴보기로 한다.

### 3.1. 개인 인명 (코드 Indiv)

이 클래스에 속하는 백과 명사들의 유형을, 결합되는 ‘본연적 형태소(Appropriate Particle)’들의 유형에 따라 분류해 보면 다음과 같다. 현재 이 연구에서 제시하고 있는 AP 들은, 일단 기존의 한국어 대사전류를 중심으로 구성된 것(1차 버전)으로, 앞으로 계속 확장되고 보완될 예정이다.

AP 대분류	AP 목록	예
불교 관련	<법사>	강장/법사
	<국사>	보각/국사
	<대사>	달마/대사
	<율사>	자장/율사
	<존자>	아난/존자
	<선사>	보조/선사
	<거사>	향산/거사
	<역사>	금강/역사
	<보살>	구탈/보살
	<동자>	금강/동자
	<화상>	대구/화상
	<여래>	다보/여래
	<관음>	유희/관음
	<주지>	내진/주지
	<본존>	대원/본존
	<승>	공부/승
	<선>	가비라/선
	<불>	노자나/불
	<천>	나찰/천

왕족 관계	<왕>	의자/왕	
	<대왕>	세종/대왕	
	<용왕>	덕차가/용왕	
	<성왕>	동명/성왕	
	<대제>	콘스탄티누스/대제	
	<천왕>	나카무라/천왕	
	<황제>	나폴레옹/황제	
	<제>	도무/제	
	<왕후>	단경/왕후	
	<후>	여/후	
	<부인>	라파예트/부인	
	<황후>	기/황후	
	<황태후>	태/황태후	
	<왕비>	인수/왕비	
	<비>	명/비	
	<위>	대원/위	
	<천자>	불심/천자	
	<세자>	사도/세자	
	<왕자>	니콜라우스/왕자	
	<태자>	기타/태자	
<공주>	평강/공주		
<대군>	능창/대군		
<군>	대원/군		
<수사+세>	니콜라이/오세		
관직, 칭호	<대감>	친륜/대감	
	<대신>	귀척/대신	
	<공신>	호성/공신	
	<태공>	게오르크/태공	
	<재상>	철혈/재상	
	<백작>	몽테크리스토/백작	
	<후작>	랑부예/후작	
	<장군>	최일/장군	
	<박사>	노반/박사	
	<선생>	백결/선생	
	<호군>	친중/호군	
	<공>	목면/공	
	<당>	사임/당	
	<자>	한비/자	
	<님>	금성/님	
	<랑>	효종/랑	
	<거서간>	남해/거서간	
	<차차용>	남해/차차용	
	기타 유형	<교주>	대은/교주
		<형제>	술라긴바이트/형제
<천녀>		길상/천녀	
<선인>		아라라/선인	
<노인>		남극/노인	
<천신>		좌품/천신	
<신>		제우스/신	
<여신>		비너스/여신	
<대신사>		친중수운/대신사	
<녀>		하백/녀	
<성+씨>		해경궁/홍씨	
<성모>		성모/마리아	

<표 8>

위와 같이 특정 AP 류가 결합하여 이루어진 합성어 형태외에 발견되는 유형은, 대부분 <성+이름>의 내부 구성을 가진 명사류이고(예: 김유신), 그의, ‘호’나 ‘법명’, ‘필명’등이 바로 엔트리로 기재된 경우들도 존재한다 (예: 결재, 백범, 추사). 또한 ‘성모/마리아’

(제 10 회 한글 및 한국어 정보처리 학술대회)

처럼 AP(성모)가 합성어의 좌측에 실현되는 경우도 나타난다. '인명' 관련 백과 명사류는, 의문사 '누구?'에 대응되는 형태들로 특징지어졌기 때문에, 이 클래스에는 '제우스신', '나찰천', '길상천녀'와 같은 신화적 인물 또는 귀신을 나타내는 명사류들도 포함된다.

거듭 강조하거나, 위의 목록은 현행 대사전류에서 다루어지고 있는 백과 명사 유형을 분류, 검토하여 얻어진 결과이다. 역사적으로 의미가 있다고 간주되는 백과 명사류는 대체로 수록되었으나, 그 선정 기준이 명시적이지 않기 때문에 앞으로 더 보완되어야 할 부분들이 눈에 띈다. 또한 실제 신문, 잡지 등의 문서에서 많이 발견되는 다음과 같은 AP 결합 명사류들은 다루어지지 않고 있는데,

클링턴/대통령, 처어칠/수상,

엄밀히 말해서, 위의 <3-1> 부류의 '나폴레옹/황제'나 '세종/대왕'등의 구조와 사실상 동일한 구조를 보이고 있다. 이들 목록은 계속 변화하며, 또한 그 한계를 설정하기가 어려운 현실적인 문제가 존재한다. 따라서, 사전에 수록되더라도, '클링턴', '처어칠'와 같은 개인 인명의 형태로 등재되는 것이 일반적이다. 실제로, 이들 개인 인명과 그 직책 사이의 관계는, <3-1> 클래스에서 다루어지고 있는 AP 결합 형태들과는 성격이 다소 다르다. <3-1>의 인명은 대부분 역사적 인물들이며, 결합된 AP를 '(전+현) AP'와 같은 형태로 대치하기 어렵다 (예: 세종대왕/\*세종전대왕 vs 클링턴대통령/클링턴현대통령). 후자 유형들은 뒤에서 다시 언급된다.

3.2. 유도된 인물 명사 (코드 DerHum)

이 클래스에 속하는 명사들은, 모두 '비인물 백과 명사'에, '인물성'의 자질을 갖는 접사 (또는 특정 어휘) AP가 결합하여 합성어 형태를 이룬다. 이때 AP들과 결합하여 <3-2> 클래스 명사류를 허용하는 '비인물성 백과 명사'류를 분류하면 다음과 같다.

<DerHum>유도하는 백과명사	예
학문, 사상 및 종교명	황금만능주의
지명, 왕, 집안 이름	이집트
기관, 모임 및 단체명	대법원

<표 9>

이 클래스는 우선적으로, 인물성 명사를 유도하는 '비인물 백과 명사'류에 대한 체계적인 연구가 선행되어야 한다. 가령, '종교 관련' 명사들에 일련의 접사들이 결합하여 '종교 관련자'들을 나타내는 명사들이 형성된다면, 이때 이러한 '종교 관련 명사'들의 목록이 우선 빠짐없이 구성되어야 할 것이다. 그 다음, 어떤 접사 및 특정 어휘들이 결합하여 이와 같은 합성어를 구성하는지가 연구되어야 할 것이다.

위 <표 9>에서 제시된 세 가지 백과 명사류중에서 세번째 유형은, 뒤에서 살펴계 될 <3-3 Collect> 클래스를 구성한다. 따라서 이 경우, <3-3> 클래스 목록을 체계적으로 구축하는 작업이 선행되어야 한다.

그러면 여기서, 이 클래스를 구성하는 AP 유형들을 살펴 보면 다음과 같다.

<DerHum>을 유도하는 백과명사	AP	예
---------------------	----	---

학문,사상,종교명	<자>	마르크스주의/자
	<도>	전산학/도
	<가>	규수문학/가
	<인>	천주교/인
	<스트>	페시미스트
지명,왕,집안이름	<인>	프랑스/인
	<가>	에스테/가
	<족>	만주/족
	<족속>	가나안/족속
	<제족>	나가/제족
	<민족>	라틴/민족
	<사람>	나사렛/사람
	<인종>	코카서스/인종
<왕조>	스튜어트/왕조	
역사적 조직기구	<협판>	군부/협판
	<사령>	금란/사령
	<판서>	공조/판서
	<부사>	광정원/부사
	<사>	군전/사
	<관>	군시/관
	<장>	금군/장
기관,모임,단체명	<장>	대법원/장
	<원>	부산재향군인회/원

<표 10>

<기관, 모임 및 단체명> 백과 명사에 결합하는 AP 유형에 있어서도, '대통령', '장관' 등과 같이 직책을 나타내는 명사 유형들은 제외되었다. 가령 다음을 보자.

재무부/장관, 삼성물산/영업부/전무

이들은 하나의 백과 명사에 직책을 나타내는 '명사'가 결합하여 구성된 합성 명사의 하나로, 백과 명사 '대법원'에 직책을 나타내는 접사 '장'과 결합된 <3-2> 파생 명사 '대법원장'과 유사한 구조를 하고 있다.

그런데, 위와 같은 구성으로 형성되는 합성 명사류의 목록은, 각 백과 명사가 가르키는 '조직 체계'에 의존적이어서, 특히 일반 관공서나 국가 기관이 아닌 사기업 단체의 경우, 즉 <대리>, <전무>등과 같은 직책 명사 결합이 가능한 목록의 구성은, 각 해당 단체의 조직도 구성에 대한 정보가 없이는 불가능하다. 이들 직책 명사류는, 앞서 <3-1> 클래스의 AP유형을 논의할 때 검토되었던 형태들로, '개인인명' 고유 명사와도 직접 결합한다. 즉 직책 명사 '장관'은 다음과 같이 개인 인명 '강창희'와 결합하여 <3-1> 유형과 유사한 구조로 실현될 수도 있고, '과학기술부'라는 단체 인명 명사 (<3-3> 유형)와 결합한 <3-2> 유형과 유사성을 보이기도 한다.

강창희/장관  
과학기술부/장관

더우기, 위의 형태들이 다시 결합한 다음과 같은 복합 시퀀스가 관찰되기도 하는데,

강창희/과학기술부/장관  
과학기술부/강창희/장관

이와 같은 형태들은 별도의 연구를 필요로 한다. 5

장에서 다시 논의된다.

<3-2> 클래스를 구성하는 AP 들중 상당부분은, '백과 명사'류가 아닌 '보통 명사'류와도 결합하여 하나의 합성어를 구성할 수 있다. 이들은 본질적으로 <인물성> 자질을 부여하는 접사 또는 단어들이므로, 가령 접사 '자'의 경우를 보면, 다음과 같은 합성어들이 가능하다.

운전/자, 택시/운전/자, 대림/택시/운전/자

이때, '운전/자'는 하나의 보통 명사 파생어로, 그리고 '택시/운전/자'는 복합-파생어로 간주된다. 그러면 '대림/택시/운전/자'는 어떻게 되는가? 이것은 분명히 하나의 백과 명사 시퀀스이다. 이때, AP 를 분리해낸 형태가 하나의 백과 명사를 이루고 있어야 한다는 것이, 이 <3-2> 클래스에 대한 형식적 조건이므로, 이 경우에는 '자'가 홀로 AP 로 분석될 수 없고, <대림택시>를 분리해 내고 남은 형태 '운전/자'가 AP 유형으로 등재되어야 한다. 그러나, 이때 '운전/자'와 같은 파생어류를 AP 로 간주한다면, '운전/기사', '운전/경험/자' 등과 같은 복합 명사류에 대한 방대한 어휘 자료가 연구되지 않는한, 작성될 수 없는 목록이 된다.

즉, 다음 <표 11>의 예들은 모두 <표 12>의 백과 명사 합성 시퀀스를 구성하므로, 이들도 <3-2> 클래스의 AP 유형으로 처리해야 하는 부담을 안게 된다.

보통 합성 명사	
기획실/이사	
총무과/계장	
환경/연구/원	

<표 11>

백과 명사 복합 시퀀스 구성	
보통명사	백과명사
기획실/이사	삼성전자/기획실/이사
총무과/계장	내무부/총무과/계장
환경/연구/원	극동아시아지부/환경/연구/원

<표 12>

다음의 경우들도 마찬가지이다.

한총련/주도/자, 민주노총/가담/자

위의 경우들도, 현재의 <3-2> 클래스의 AP '자'의 결합에 의한 백과 명사의 한 형태로 볼 수 없다. '한총련주도'라는 백과 명사는 존재하지 않기 때문이다. 따라서, 이는 '한총련'이라는 백과 명사가 보통 파생 명사 '주도/자'와 결합하여 이루어진 형태이다. 이와 같은 보통 파생-복합 명사류에 대한 목록이 제공되어야 처리 가능한 백과 명사류 형태들이다.

그러므로, 현재 <3-2> 클래스에서 '인물성' 자질을 유도하는 AP 유형의 목록 구성에 있어서 앞서 살핀 '장관', '전무', '대표', '운전자' 등과 같은 성분들은 포함되지 않는다. 이 형태들은 앞으로 <3-3> 클래스에서 다루게 될 '기관'이나 '단체', '모임' 관련 백과 명사류에 결합하여 복합어 형태를 구성하는 단어들이므로, 별도의 연구가 뒷받침되어야 한다. 한가지 지적할 점은, 앞의 <표 10>에서 다루어진 AP '장'은

다음에서 나타나는 형태들을 모두 포함하는 형태들이라는 점이다.

부장, 과장, 회장, 사장, 교장, 학장, 소장, 원장,...

이들은, 다음과 같은 시퀀스에서 실현되는 형태들로, <기관>을 나타내는 백과 명사에 AP <장>이 결합된 것으로 분석된다. 따라서 위의 '가담자', '주도자'의 경우와는 다른 형태이다.

<장>	안기부/장
	내무부총무과/장
	간도국민회/장
	마이크로소프트사/장
	경기고등학교/장
	문과대학/장
	국립환경연구소/장
한국과학기술원/장	

<표 13>

반면, '총장'은 '장'의 하위 유형으로 분류되기 어렵다. '장'을 분리해낸 형태가 하나의 '백과 명사' 형태를 취할 수 없기 때문이다. 가령 '연세대학교총장'에서 '장'을 제거하면 '연세대학교총'과 같이 비정상적인 시퀀스가 구성되므로, 이때 접사 '장'이, 어떤 명사 유형에 바로 결합한 것으로 분석될 수 없다. 따라서, '총장'은 위의 AP '장' 결합 백과 명사류에서 제외된다.

끝으로 덧붙일 점은, 이 클래스에 속하는 명사들은, 하나의 '비인물성 백과 명사'에서 유도되어 형성된 '인물성 명사'류, 다시 말해서, 기저에 하나의 백과 명사를 내포하고 일정한 형태의 접사(또는 단어)가 결합한, 합성어 형태들인데, 이들중에는, 이와 같은 명시적인 결합 구조를 보이지 않는 경우들도 발견된다는 점이다. 가령, 다음과 같이 민족, 인종등을 표현하는 형태들이 이 클래스에 해당된다면,

만주/족, 아프리카인, 몽고/인종

다음과 같은 경우도, 비록 형태적으로는 이들과 연관지을 수 있는 '백과 명사'가 명시적으로 존재하지 않으나, 위의 형태들과 함께, 같은 부류로 분류되어야 할 것으로 보인다.

니그로, 황인종, 다키아

### 3.3. 단체 및 인물 집단 (코드 Collect)

이 클래스에 속하는 백과 명사류는, 반드시 최소한 두 명 이상의 '복수 인물'을 대표하는 집합 명사류를 가르킨다. 이 부류에 속하는 명사들은, 앞서 언급한 바와 같이, 대부분의 경우, 문맥에 따라 이와 같은 '복수 인물'의 해석을 받을수도 있고, '장소 명사'의 해석을 받을수도 있다는 특징이 있다. 예를 들어, '모임'이나 '협회', '사회적 기구', '대학'등이 포함되는데, 그 유형의 목록을 구축하기 위해서는 백과 사전적 정보가 요구된다. 이 클래스에서 발견되는 단체 및 기관명들은 대부분의 경우, 일정 접사와의 결합을 통해 <3-2 DerHum 코드> 명사들의 유도를 허용한다.

<3-3 Collect> 클래스에 속하는 백과 명사들의 유형을 살펴보면 다음과 같다.

(제 10 회 한글 및 한국어 정보처리 학술대회)

<3-3 Collect 코드> 백과 명사 유형	예
3 음절어 AP 결합	국립중앙/박물관
2 음절어 AP 결합	국제통화/기금
1 음절어 AP 결합	공인회계사/회
로마자 AP 결합	내셔널/뱅크

<표 14>

이 클래스에 속하는 백과 명사들은, 국제적 기구로부터 국가 공공 기관, 중소기업체, 소규모 상점, 소규모 친목 동우회등에 이르기까지 매우 다양하다. 따라서, 개인 인명 명사의 경우에서와 마찬가지로, 인지도가 높은 중요 기관부터 사전 처리를 하는 것이 합리적이다.

이 클래스에 속하는 백과 명사 유형은, 이를 구성하는 AP의 음절별 특징에 따라 크게 세가지로 분류되었다. 그것은, 이 클래스의 의미적인 속성 분류는 지나치게 다양해서 큰 의미가 없는 것으로 판단되었기 때문이다. 여기서 가령 3 음절 AP 유형으로 분류된 '위원회'는, 1 음절 '회'의 하나로 분류될 수 있으나, 여기서는 최대한 명시적으로 그 유형을 연구하기 위하여, '우측 3 음절어'가 특징지어지지 어려운 경우 (이 경우 물론 2 음절 AP를 구성하지도 못한다)에 한해서만 '1 음절 AP' 유형으로 분류하였다. 즉, 예를 들어, '공인회계사/회'와 같은 경우, 다음과 같은 시퀀스 분리가 불가능하므로,

\*공인회/계사회, \*공인회계/사회

해당 AP는 '회'로 설정되어야 한다. 여기서 특정 AP를 3 음절까지로 한정지은 것은, 자의적인 결정이나, 우선은 가장 효율적인 경계선으로 파악된다. 4 음절, 5 음절이상 확대시킨다면 그만큼 더 세분화된 목록이 구성될 것이고, 결국은 각 백과 명사들이 모두 제시된 전체 목록을 의미하게 될 것이다.

기존의 대사전류에는 대체로 중요한 기관 및 단체들만이 등재되어 있기 때문에, 우선은 신문, 잡지등에서 무작위로 얻어지는 목록보다는 더 의미있는 기본 정보를 제공해준다. 실제로, 신문, 잡지등에서는 사회, 정치면에 중요 기관들이 등장한다면, 모임이나 회합을 알리는 여러 작은 단체들의 광고들도 많은 비중을 차지한다.

다음은 기존 대사전류에서 얻어진 결과를 정리한 것인데, 대체로 중요도가 높다고 판단된 것들에 초점을 맞추어 사전에 등재하고 있음을 알 수 있으나, 그 선택 기준이 어떤 엄밀한 근거에 의해 제시되어 있지 못하기 때문에, 누락된 것들에 대한 충분한 논거를 찾기 어려운 경우들이 발견된다. 다음 표는 현재 우리가 구축하고 있는 백과 명사 사전 (1차 버전)의 AP 목록의 일부를 보인다. 이 클래스를 구성하는 AP 유형은, 앞의 경우들과는 달리, 그 수가 폭발적이어서 지면 관계상 일부만을 다음에서 제시한다.

<3-3>	AP	예
3 음절 AP	<위원회>	경제과학/위원회
	<대학교>	연세/대학교
	<연구소>	국립환경/연구소
	<연구원>	한국정보통신/연구원
	<도서관>	국립중앙/도서관
	<박물관>	국립중앙/박물관
	<미술관>	호암/미술관

2 음절 AP	<회의>	경제과학심/회의
	<협회>	건축사/협회
1 음절 AP	<과>	남만/과
	<당>	공화/당
	<청>	관세/청

<표 15>

현재 버전에 포함된 <3-3 Collect> 클래스 구성 AP의 수를 음절별로 살펴 보면 <표 16>에서와 같다.

1 음절 AP 수	44
2 음절 AP 수	75
3 음절 AP 수	227
<3-3 Collect> 전체 AP 수	346

<표 16>

로마자 한글 표기를 포함한 유형은, AP를 위와 같이 음절별로 분류하지 않았다. 몇가지 예를 들면 다음과 같다.

로마자 한글 표기 AP 결합 형태	
<클럽>	라이온즈/클럽
<뱅크>	내셔널/뱅크
<그룹>	로열티치셀/그룹
<소사이어터>	로열/소사이어터
<아카데미>	로열/아카데미
<인스티튜션>	로열/인스티튜션
<에디션>	유니버설/에디션
<센터>	메디컬/센터

<표 17>

또한 로마자 한글 표기 백과 명사에는, 위와 같이 특정 AP의 결합없이, '약자 (Acronym)'의 형태를 띤 것도 있고, 단일어 형태나, 특정어 결합이 나타나지 않는 복합어 형태를 띤 것들도 포함된다. 즉 다음과 같다.

약자 형태	더블류비시(W.B.C.)
복합어 형태	걸/스카우트
단일어 형태	페러마운트

<표 18>

4. 자동 문서 처리 시스템과 백과 명사 사전

영어나 불어 문서에서 실현되는 백과 명사류에 대한 자동 처리는, '대문자'로 시작되는 어휘 형태들에 대한 형식적인 접근에 의해 가능하다. 그러나, 한국어의 경우에는 그 어떠한 형식적 표지도 존재하지 않아서, 이와 같이 기존 사전류에서 체계적으로 다루어지지 않는 '백과 명사'류의 처리 문제가, 보다 심각하게 대두될 수 밖에 없다. 예를 들어 다음을 보자.

*I know where Bill is  
I know where the bill is*

영어의 경우, 대개 '백과 명사'류로 간주될 수 있는 명사들은, 대문자로 시작하는 형식적인 특징을 갖고 있을 뿐 아니라, 관사나 수량사등이 결합할 수 없는 통사적인 특징등이 두드러지게 나타나는데, 한국어의 경우 이와 같은 구별 장치는 크게 도움이 되지 않는다.



그가 이승만을 도왔다  
그가 이재민을 도왔다

<3-2 DerHum>	24
<3-3 Collect>	346

< 표 19 >

위의 예에서, 사전에 '이승만'이라는 형태가 부재한 경우, 이것은 미등록어의 하나의 처리될 수 있는데, 이때 이것이 생산성 높은 복합 명사의 하나라서 누락된 것인지(가령 '이자/율', '이불/장' 등과 같은), 아니면 백과 명사로서 '지명'을 나타내는 형태인지(가령 '이리/시', '이문/동' 등과 같은), 구별할 길이 없다. '백과 명사' 사전이 체계적으로 구축되지 않는 한 해결될 수 없는 문제의 유형이다.

이와 같은 '미등록어' 문제는, 실제 형태소 분석 단계에서 가장 문제가 되는 부분의 하나이다. 여기서 분명한 사실은, 이와 같은 미등록어 목록의 사용은, 결국 '백과 명사' 및 '합성 명사' 사전을 구축하기 위한 어휘 자료의 역할을 할 뿐, 그이상의 아무런 해결 방안도 제시해 줄 수는 없다는 점이다. 이들 '미등록어' 목록을, 그 형태 자체로써 사전 엔트리로 처리한다면, 그것은 단지 그와 같은 시퀀스들이 한국어 문서상에서 발견된다는 정보만을 가지고 있는 것과 같다. 이때 무작위로 등재된 이 목록들의 내부 유형을 알지 못하는 한, 그 목록의 형태적, 의미적, 또는 구문적 정보를 부여하는 것은 전혀 불가능할 뿐 아니라, 극히 일부에 불과한 이 목록을 체계적으로 완결시키기 위한 작업이 더 진행될 수가 없다. 즉 이와 같은 미등록어의 하위 유형 분류가 이루어져야 하며, 이것은 앞에서 우리가 언급했던 대로, 미등록어 목록이 '백과 명사 사전' 및 '합성 명사 사전' 구축을 위한 원시 데이터로 사용될 뿐이라는, 논의의 원점으로 다시 돌아가는 일이 된다.

위의 두번째 예에서 나타난 '이재민'의 경우, 사실 인명의 백과 명사로 실현되었을 가능성도 배제할 수 없는데, 이 경우에도 인명의 '백과 명사' 사전이 구성되어 있다면 이와 같은 분석의 가능성을 고려할 수 있을 것이다. 그러나, 앞서 논의한 대로, 무작위로 모든 인명 명사를 일반 명사류와 함께 처리한다면 발생하는 중의적 해석의 수는 폭발적으로 증가하게 될 것이다. 따라서, 백과 명사 사전은, 보통 명사류와는 달리, 몇가지 '층위(Hierarchy)' 및 '특정 영역(Specific Domain)'에 의한 변별력을 가지고, 우선으로 사용될 수 있는 장치가 반드시 필요하다. 가령 정치인 인명, 예술가 인명, 역사적 지명, 현존 지명 이름등, 그 관련 정보의 유형에 따른 우선적인 사전 매칭 작업이 필요하다.

그러나, '백과 명사'류는 엄격히 말해서, 그 자체가 열려 있는 집합을 이루고 있으므로, 계속적으로 확장, 보완되어야 한다. 따라서, 이와 같은 체계적인 확장 작업을 위해서는, 어떠한 어휘 형태가 '인명'과 관련된 백과 명사류를 구성할 수 있는지, 바로 그 모든 형태의 '유형(Pattern)'을 파악하는 것이 중요하다. 이것은 사실상, 이때 결합되는 '특정 어휘 AP'에 대한 빠짐없는 목록 구성으로 가능하다. 이와 같은 '특정 어휘 AP' 목록을 통해, 백과 명사 유형 분류의 틀이 잡히게 되면, 어떠한 백과 명사들의 목록이 사전에 일차적으로 우선 다루어져야 하고, 그렇지 않은지에 대한 좀더 효율적인 판단이 가능해질 것이다.

우리가 구축하고 있는 '백과 명사' 사전에서, '인명 사전'의 경우에 국한해 볼때, 이들 클래스를 구성하는 AP의 수(1차 버전)는 다음과 같다.

인명 백과 명사	관련 AP의 수
<3-1 Indiv>	73

그리고, 현재 구축중인 '인명 백과 명사 사전'은, 현행 대사전류에서 다루어지고 있는 역사적 인물이나 단체, 국제적 인지도가 높은 형태들에 그 초점이 맞추어져 있으며, 그 수는 다음과 같다. <부록>에 현재 구축중인 인명 백과 명사 사전의 일부 발췌 부분이 등재되어 있다.

인명 백과 명사	원어 형태	사전 엔트리 수
<3-1 Indiv>	CHN	3 870
	ENG	7 060
<3-2 DerHum>	CHN	80
	ENG	260
<3-3 Collect>	CHN	1 970
	ENG	70
Total		13 710

< 표 20 >

'인명 백과 명사'를 구성하는 명사들의 AP 유형의 연구는, 사전이 아직 완결되지 않은 상태에서, 전적으로 사전에 의존적일 수 없는 시스템의 효율성을 높이는 데에도 매우 유용한 정보를 제공한다. 즉, 인명 명사들의 AP 목록을 이용하면, 비록 그 인명 백과 명사 목록 자체를 아직 구축하지 못한 단계에서도, 그와 같은 명사들이 실현될 수 있는 '문맥(Context)'을 매우 효율적으로 추적할 수 있다.

'인명 백과 명사'류가 실현될 수 있는 시퀀스 구조를 살펴 보면 대체로 다음과 같은데, 여기서 이들 '형식 구조(Formal Structure)'들에 대한 연구만으로는 충분한 인식 결과를 기대하기 어렵다는 것을 확인하게 될 것이다. 즉, 각 구조에 특징적으로 사용되는 결합어 AP들의 '어휘 정보(Lexical Information)'가 병행되어야 한다. (다음의 'EN'은 백과 명사(Encyclopedic Noun)를, 'N'은 명사(Noun)를 나타낸다).

<1>	<인명 EN-(후치사)> 유형
	<이승만이> 후보로 나왔다
<2>	<인명 EN (분야 N)'직책 N'-(후치사)> 유형
	<이회창한나라당명예총재는> 오후에 기자회견을 가졌다
<3>	<인명 EN-(의)'관계 N'-(후치사)> 유형
	<서태지의 형이> 미국에 갔다
<4>	<인명 EN '동격 N'-(후치사)> 유형
	<김민우 선배가> 나를 불렀다
<5>	<인명 EN '의존 N'-(후치사)> 유형
	<김민우씨가> 왔다

< 표 21 >

위의 표에서, <1>은 일반 명사들이 나타나는 가장 기본적인 문맥 형태로서, '인물성' 자질을 가진 보통 명사류와 백과 명사류가 모두 실현될 수 있다. <2>의 유형은 신문, 잡지등 시사성있는 문서에서 가장 많이 나타나는 유형으로, 우리가 지금까지 살펴본 모든 유형의 인명 백과 명사들이 실현될 수 있다.

<3>과 <4>는 의미 해석상으로는 분명히 구별되나, <3>에서 속격 후치사 '의'가 생략되면, 형태상 동일한 구조를 취한다. 이때, 동반된 명사의 유형에 따라, 서로 변별력을 잃어 두 경우 모두 중의적 해

식을 받을 수가 있는데 (가령 '서태지 형'은, 서태지의 형을 가르킬 수도 있고(<3>의 해석), 화자가 형으로 부르는 사람의 이름이 바로 서태지인 경우(<4>의 해석), 모두가 가능하며, '김민우 선배'의 경우도 마찬가지이다), 이와 같은 중의적 해석의 발생은 '정보 검색' 시스템에서 중요한 문제의 하나가 된다. 즉, 특정 인명 중심으로 일정 정보를 얻어내고자 할 때, 위의 <3>과 <4>와 같은 문맥을 어떻게 해석했느냐에 따라서 해당 정보가 들어있는 문서의 검색 결과에 중요한 영향을 미칠 것이기 때문이다. 여기서 이 두 유형을 비교해 보면 다음과 같은 통사적 차이를 발견할 수 있다.

인명 EN N = 인명 EN-의 N		
해석<3>	+	서태지형 = 서태지의 형
해석<4>	-	서태지형 = *서태지의 형
인명 EN N = N 의 이름이 인명 EN-이다		
해석<3>	-	서태지형 = *형의 이름이 서태지다
해석<4>	+	서태지형 = 형의 이름이 서태지다
인명 EN N = 인명 EN[성] N		
해석<3>	-	서태지형 = *서형
해석<4>	+	서태지형 = 서형
인명 EN N = 인명 EN[이름] N		
해석<3>	+	서태지형 = 태지형
해석<4>	+	서태지형 = 태지형
인명 EN N = (인명 EN) N!! (호격)		
해석<3>	-	서태지형 = *(서태지) 형!!
해석<4>	+	서태지형 = (서태지) 형!!

<표 22>

그러나, 이와 같은 중의적 해석의 발생 가능성은, <3>과 <4>의 모든 경우에 나타나지는 않는다. 동반되는 개별 명사의 속성에 의존적이어서, 가령 다음과 같은 경우,

서태지 (형님 + 선배)

이와 같은 중의성이 발생한다면, 다음과 같은 관계 명사들(<3>유형)이 실현된 경우, <4>와 같은 동격 명사로의 해석은 불가능하다.

서태지 (아버지 + 제자)

여기서, 위와 같이 형식 구조만으로 그 유형을 분류한다는 것은 큰 의미가 없음을 다시 확인된다. 즉, 각 구조속에 실현될 수 있는 동반 명사들의 어휘 목록이 구축되어야, 비로소 이들 개별 명사들을 중심으로, 함께 공기할 수 있는 '인명 백과 명사'들의 주장이 가능해질 것이다.

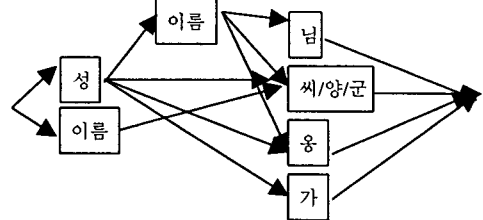
위의 표에서, 끝으로 <5>의 유형은, 인명 백과 명사가 실현될 수 있는 가장 특징적인 문맥 형태로, 이때 동반되는 '의존 명사'의 목록도 매우 제한적이다. 다음은 그 의존 명사들이다.

종류	성과 결합	이름과 결합	전체와 결합
씨	+	+	+
양	+	+	+
가	+	-	-
군	+	+	+
님	-	-	+
용	+	-	+

<표 23>

위에서 볼 수 있듯이, 이들 의존 명사들은 성이나 이름 각각에 결합할 수 있는 형태가 있는가 하면 (예: 김민우씨/민우씨/김씨), 어느 특정 형태만 결합 가능한 것도 있다 (예: 김가/\*민우가/\*김민우가).

이와 같은 제약 관계를 '유한 오토마타 (Finite State Automata)'로 표현하면, 다음 전이 그래프에서 보이는 바와 같다.



<그림 6>

### 5. 향후 연구

앞서 논의한 바와 같이, 신문, 잡지등에서 가장 빈번히 발견되는 '인명 관련 백과 명사 시퀀스' 형태는, 다음 <표 24>에서 보이는 바와 같다.

김병관/동아일보/회장
권근술/한겨레신문사/장
조세형/국민회의/총재권한대행
미국인/선교사/언더우드

<표 24>

위에서 나타난 시퀀스들은 모두, 일종의 합성어를 형성하는 '복합체' 형태를 띠고 있다. 첫째 예는 개인 인명 백과 명사 '김병관'과 단체를 나타내는 백과 명사 '동아일보', 그리고 직책을 나타내는 보통 명사 '회장'의 세 부분으로 구성되어 있다. 둘째 예는 개인 인명 백과 명사 '권근술', 단체 백과 명사 '한겨레신문사', 그리고 인물성 명사를 유도하는 접사 '장'의 결합으로 이루어졌다. 셋째 예는 첫번째 예와 동일한 구성을 보이고 있고 (권근술+국민회의+총재권한대행), 네번째 예는 지명 관련 백과 명사 '미국'에 인물성 접사 '인'이 결합하여 파생 백과 명사 '미국인'이 나타났고, 보통명사 '선교사', 그리고 개인 인명 '언더우드'의 결합으로 구성되었다. 즉, 다음과 같은 구성 성분들의 조합으로 이루어졌다.

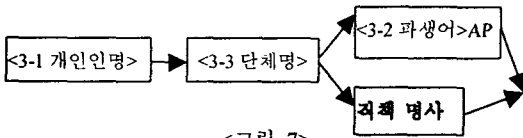
유형	의미	예
백과명사	개인 인명	김병관, 권근술, 조세형, 언더우드
	인물 단체	동아일보, 한겨레신문사, 국민회의
	지명 관련	미국
보통명사	직업, 직책	회장, 총재권한대행, 선교사
파생접사	직책 관련	장(한겨레신문사/장)
	국민	인(미국/인)

<표 25>

이때, '인물성 (Human Feature)'의 속성을 가지는 백과 명사류는 위에서 회색으로 표시한 두가지 '백과 명사'류와, 인물성 '파생 접사'에 의해 유도된 '파생 백과 명사'류가 해당된다.

그러면 <표 24>에서 나타난 백과 명사 시퀀스는 어떻게 처리되어야 하는가? 이들 시퀀스는, 현

작업에서 진행중인 <백과 명사 사전>의 각 하위 어휘 클래스들의 내적 결합 관계를 밝힘으로써 처리할 수 있다.



<그림 7>

이와 같은 결합 경로를 갖는 모든 백과 명사 시퀀스의 목록을 구축하기 위해서는, '인명 관련' 백과 명사 세 클래스 모두에 대한 빠짐없는 연구와, '직책 관련 보통 명사' 목록에 대한 연구가 동반되어야 함이 확인된다.

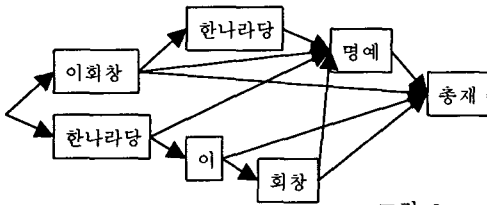
이와 같은 특정어들에 대한 치밀한 목록 구성과 함께, 또하나 고려되어야 할 사항은, 이들 시퀀스들이 '변형(Transformation)'되어 실현되는 문맥들 사이의 '동의 관계'를 파악하는 일이다. 가령 다음 시퀀스는,

이회창/한나라당/명예총재

문서상에서 다음과 같이 여러 형태로 변환되어 실현될 수 있다.

한나라당/이회창/명예총재  
한나라당/이총재  
이회창/총재, etc...

따라서, 이와 같은 변형 관계의 문맥들에 대한 통사적인 연구가 동반되어야 하며, 이러한 관계는 '부분 문법 (Local Grammar)'의 형태로 구현되어 '유한 오토마타 (FSA)'로 변환될 수 있다. 다음 그래프는 몇 가지 변형 관계를 보이는 유한 오토마타를 나타낸다.



<그림 8>

백과 명사 사전의 구축과 함께, 이와 같은 백과 명사 어휘들이 실현될 수 있는 '문맥'의 유형 및 그 변형 가능성에 대한 통사적인 연구는, 검색 시스템의 구현에 있어서 매우 중요한 부분이다.

백과 명사 시퀀스가, 문서속에서 다양한 형태로 변형되어 실현될 수 있다는 사실은, '합성 명사' 사전 구축시에도 그대로 재현되는 문제의 유형이다. 가령, '수도함락'과 같은 형태를 확장된 의미의 합성어로 간주한다면, 이때 이 시퀀스의 내부 구조를 아는 것은 매우 중요하다. 그것은, 이와 같은 합성어가 합성 명사 사전에 등재되어 있다 하더라도, 문서상에서는 다음과 같이 다양한 형태로 출현할 수 있기 때문이다.

수도함락  
= 수도가 함락되었다  
= 함락된 수도  
= 수도를 함락한 적군이,...

여기에, 합성 명사 사전의 구축 작업에, 구문 정보 연구가 반드시 병행되어야 하는 필요성이 확인된다.

한국어 '어휘-문법 (Lexicon-Grammar)'의 통사적 연구결과 및 유한 오토마타로 변환되는 '부분 문법 (Local Grammar)'의 구축 결과들은, 이와 같은 '보통 명사 합성어 사전'뿐만 아니라, '백과 명사 합성어 사전'에도 매우 중요한 정보를 제공하게 된다.

참고 문헌

- [1] Gary-Prieur, Marie-Noëlle, Le nom propre constitue-t-il une catégorie linguistique ?, *Langue française* N-92, Larousse, 1991
- [2] Gross, Maurice, La construction de dictionnaires électroniques, *Annales des Télécommunications*, Tome 44 N-1:2, CNET, 1989
- [3] Gross, Maurice, Lexicon based Algorithms for the Automatic Analysis of Natural Language, in *Theorie und Praxis des Lexikons*, Walter de Gruyter: Berlin, 1993
- [4] Nam, Jee-Sun, *Dictionnaire des noms simples du coréen*, RT N-46, Laboratoire d'Automatique Documentaire et Linguistique, University Paris 7, France, 1994
- [5] Nam, Jee-Sun, *Construction of Korean electronic lexical system DECO*, Papers in Computational Lexicography Complex '96, ed. by Ferenc Kiefer, Gabor Kiss and Julia Pajzs, Budapest, Hungarian Academy of Sciences, 1996
- [6] Nam, Jee-Sun, *A Local Grammar-based Approach to Recognizing of Proper Names in Korean Texts*, Papers in Fifth Workshop on Very Large Corpora (WVLC-5), Beijing/Hong Kong, 1997
- [7] Silberztein, Max, *Dictionnaire électronique et analyse automatique de textes, Le système INTEX*, Paris: Masson, 1993
- [8] 이회승, <국어대사전>, 민중서림, 1988.
- [9] 한글학회, <우리말큰사전>, 어문각, 1991
- [10] 정래정, 김준태, 고유 명사 출현 패턴을 이용한 색인의 성능 향상에 관한 연구, 한글 및 한국어 정보 처리 학술 발표 논문집, pp 68-72, 1996.
- [11] 양장모, 김민정, 권혁철, 언어 정보를 이용한 한국어 미등록어 추정, 정보 과학회 봄 학술발표 논문지, pp 957-960, 1996

<<부록 : <인명 백과 명사 사전>의 일부 발췌문>>

\*\*\*\*\*  
 <3-2 DerHum/ENG 코드> 클래스 일부  
 \*\*\*\*\*  
 :말라가시/족, HUM,DerHum,ENG,Malagasy/族,N/H2,  
 :말라야/인, HUM,DerHum,ENG,Malay/人,N/H2,  
 :말레이/인, HUM,DerHum,ENG,Malay/人,N/H2,  
 :멜라네시아/인, HUM,DerHum,ENG,Melancsia/人,N/H2,  
 :모로/족, HUM,DerHum,ENG,Moro/族,N/H2,  
 :모아브/인, HUM,DerHum,ENG,Moab/人,N/H2,  
 :모이//제족, HUM,DerHum,ENG,Moi/諸族,N/H2,  
 :모조케르토/인, HUM,DerHum,ENG,Modjokerto/人,N/H2,  
 :몬/족, HUM,DerHum,ENG,Mon/族,N/H2,  
 :몽골리안, HUM,DerHum,ENG,Mongolian,N,  
 :무스코기/족, HUM,DerHum,ENG,Muskogean,N/H2,  
 :무어/인, HUM,DerHum,ENG,Moor/人,N/H2,  
 :미낭카바우/족, HUM,DerHum,ENG,Minangkabau/族,N/H2,  
 :미스테카/족, HUM,DerHum,ENG,Mixteca/族,N/H2,

(제 10 회 한글 및 한국어 정보처리 학술대회)

:미코로네시아족, HUM,DerHum,ENG,Micronesia/族,N/H2,  
 :바리새인, HUM,DerHum,ENG,Pharisees/人,N/H2,  
 :바리새교인, HUM,DerHum,ENG,Pharisees 敎人,N/H2,  
 :바스크인, HUM,DerHum,ENG,Basque/人,N/H2,  
 :바타크족, HUM,DerHum,ENG,Batak/族,N/H2,  
 :반달족, HUM,DerHum,ENG,Vandal/族,N/H2,  
 :반투족, HUM,DerHum,ENG,Bantu/族,N/H2,  
 :발트족, HUM,DerHum,ENG,Balt/族,N/H2,  
 :버마인, HUM,DerHum,ENG,Burma/人,N/H2,  
 :베두인족, HUM,DerHum,ENG,Bedouin/族,N/H2,  
 :베르베르인, HUM,DerHum,ENG,Berber/人,N/H2,  
 :베트콩, HUM,DerHum,ENG,Vietcong,N,  
 :벵골인, HUM,DerHum,ENG,BENGal/人,N/H2,  
 :보국족, HUM,DerHum,ENG,Vogul/族,N/H2,  
 :보도족, HUM,DerHum,ENG,Bodo/族,N/H2,  
 :보어인, HUM,DerHum,ENG,Boer/人,N/H2,  
 :보타크족, HUM,DerHum,ENG,Votyak/族,N/H2,  
 :붕고족, HUM,DerHum,ENG,Bungo/族,N/H2,  
 :부기족, HUM,DerHum,ENG,Bugis/族,N/H2,  
 :부눈족, HUM,DerHum,ENG,Bunun/族,N/H2,  
 :부르군트족, HUM,DerHum,ENG,Burgund/族,N/H2,  
 :부르야트족, HUM,DerHum,ENG,Burjat/族,N/H2,  
 :부시먼족, HUM,DerHum,ENG,Bushman/族,N/H2,  
 :비샤야족, HUM,DerHum,ENG,Visaya/族,N/H2,  
 :비하르인, HUM,DerHum,ENG,Bihar/人,N/H2,  
 :사라센인, HUM,DerHum,ENG,Saracen/人,N/H2,  
 :사르트족, HUM,DerHum,ENG,Sart/族,N/H2,  
 :사마리아인, HUM,DerHum,ENG,Samaria/人,N/H2,  
 :사모디족, HUM,DerHum,ENG,Samodii/族,N/H2,  
 :사모예드족, HUM,DerHum,ENG,Samoyede/族,N/H2,  
 :사이샤야트족, HUM,DerHum,ENG,Saisyat/族,N/H2,  
 :사카이족, HUM,DerHum,ENG,Sakai/族,N/H2,  
 :삼니움인, HUM,DerHum,ENG,Sannium/人,N/H2,  
 :색슨족, HUM,DerHum,ENG,Saxon/族,N/H2,  
 :산족, HUM,DerHum,ENG,Shan/族,N/H2,  
 :서고트족, HUM,DerHum,ENG,WestGoths/族,N/H2,  
 :세르비아인, HUM,DerHum,ENG,Serbia/人,N/H2,  
 :셈인, HUM,DerHum,ENG,Sem/人,N/H2,  
 :셀족, HUM,DerHum,ENG,Scel/族,N/H2,  
 :소러시아인, HUM,DerHum,ENG,小/Russia/人,PF/N/H2,  
 :소말리족, HUM,DerHum,ENG,Somali/族,N/H2,  
 :솔로인, HUM,DerHum,ENG,Solo/人,N/H2,  
 :솔론족, HUM,DerHum,ENG,Solon/族,N/H2,  
 :수르//왕조, HUM,DerHum,ENG,Sur//王朝,N/H2,  
 :순다인, HUM,DerHum,ENG,Sunda/人,N/H2,  
 :스와지족, HUM,DerHum,ENG,Swazi/族,N/H2,  
 :스웨덴족, HUM,DerHum,ENG,Sweden/族,N/H2,  
 :스코틀랜드인, HUM,DerHum,ENG,Scotland/人,N/H2,  
 :스튜어트//왕조, HUM,DerHum,ENG,Stuart//王朝,N/H2,  
 :스페인인, HUM,DerHum,ENG,Spain/人,N/H2,

:등문/고, HUM,Collect,CHN,登聞鼓,N/H3,  
 :등장/국, HUM,Collect,CHN,登局,N/H3,  
 :마곡/사, HUM,Collect,CHN,麻谷寺,N/H3,  
 :마미//사장, HUM,Collect,CHN,馬尾蹄匠,N/H3,  
 :마사/회, HUM,Collect,CHN,馬事會,N/H3,  
 :마산//발전소, HUM,Collect,CHN,馬山發電所,N/H3,  
 :만국//사관//학교, HUM,Collect,CHN,萬國士官學校,N/N/H3,  
 :만국//우편//연합, HUM,Collect,CHN,萬國郵便聯合,N/N/H3,  
 :만국//원자량//위원회, HUM,Collect,CHN,萬國原子量委員  
 會,N/N/H3,  
 :만의/사, HUM,Collect,CHN,萬儀寺,N/H3,  
 :만자/군, HUM,Collect,CHN,蠻子軍,N/H3,  
 :망문//상전, HUM,Collect,CHN,望門床頭,N/H3,  
 :망월/사, HUM,Collect,CHN,望月寺,N/H3,  
 :맹안/군, HUM,Collect,CHN,猛安軍,N/H3,  
 :명동//성당, HUM,Collect,CHN,明洞聖堂,N/H3,  
 :명화/적, HUM,Collect,CHN,明火賊,N/H3,  
 :무관//학교, HUM,Collect,CHN,武官學校,N/H3,  
 :무량/사, HUM,Collect,CHN,無量寺,N/H3,  
 :무역//협력//기구, HUM,Collect,CHN,貿易協力機構,N/N/H3,  
 :무위/사, HUM,Collect,CHN,無爲寺,N/H3,  
 :문수/사, HUM,Collect,CHN,文殊寺,N/H3,  
 :문혁/파, HUM,Collect,CHN,文革派,N/H3,  
 :문화//방송, HUM,Collect,CHN,文化放送,N/H3,  
 :문화//예술//진흥//위원회, HUM,Collect,CHN,文化藝術振興委  
 員會,N/N/H3,  
 :문화//예술//진흥원, HUM,Collect,CHN,文化藝術振興  
 院,N/N/H3,  
 :문화//혁명//소조, HUM,Collect,CHN,文化革命小組,N/N/H3,  
 :문화재//관리국, HUM,Collect,CHN,文化財管理局,N/H3,  
 :문화재//위원회, HUM,Collect,CHN,文化財委員會,N/H3,  
 :물가//안정//위원회, HUM,Collect,CHN,物價安定委員  
 會,N/N/H3,  
 :물장/성, HUM,Collect,CHN,物藏省,N/H3,  
 :물품/사, HUM,Collect,CHN,物品所,N/H3,  
 :미/군, HUM,Collect,CHN,美軍,N/H3,  
 :미/소//공동//위원회, HUM,Collect,CHN,美蘇共同委員  
 會,N/N/H3,  
 :미국//공보원, HUM,Collect,CHN,美國公報院,N/H3,  
 :미국//국제//교류처, HUM,Collect,CHN,美國國際交流  
 處,N/N/H3,  
 :미국//노동//총//동맹, HUM,Collect,CHN,美國勞動總同  
 盟,N/N/PF/H3,  
 :미국//도서//관//협회, HUM,Collect,CHN,美國圖書館協  
 會,N/N/SF/H3,  
 :미국//박물학//박물관, HUM,Collect,CHN,美國博物學博物  
 館,N/N/H3

\*\*\*\*\*

<3-3 Collect/CHN 코드> 클라스 일부

\*\*\*\*\*

:동/유럽//경제//상호//원조//회의, HUM,Collect,CHN,東/Europe/  
 經濟相互援助會議,PF/N/N/N/N/H3,  
 :동/인도//회사,  
 HUM,Collect,CHN,East/India/Company,PF/N/H3,  
 :동국//대학교, HUM,Collect,CHN,東國大學校,N/H3,  
 :동남//아시아//농업//개발//회의, HUM,Collect,CHN,東南/Asia/  
 農業開發會議,N/N/N/N/H3,  
 :동녕//로//총관//부, HUM,Collect,CHN,東寧路總管府,N/SF/N/H3,  
 :동녕//부, HUM,Collect,CHN,東寧府,N/H3,  
 :동림//당, HUM,Collect,CHN,東林黨,N/H3,  
 :동방//교회, HUM,Collect,CHN,東方教會,N/H3,  
 :동서//대비원, HUM,Collect,CHN,東西大悲院,N/H3,  
 :동성//파, HUM,Collect,CHN,桐城派,N/H3,  
 :동아//대학교, HUM,Collect,CHN,東亞大學校,N/H3,  
 :동양//방송, HUM,Collect,CHN,東洋放送,N/H3,  
 :동양//척식//주식//회사, HUM,Collect,CHN,東洋拓殖株式會  
 社,N/N/N/H3,  
 :동인//교회, HUM,Collect,CHN,同仁教會,N/H3,  
 :동포//교회, HUM,Collect,CHN,同胞教會,N/H3,  
 :동학//당, HUM,Collect,CHN,東學黨,N/H3,  
 :동학/사, HUM,Collect,CHN,東鶴寺,N/H3,  
 :동화/사, HUM,Collect,CHN,桐華寺,N/H3,  
 :등대/국, HUM,Collect,CHN,燈臺局,N/H3,