

기계가독형사전과 코퍼스에서 추출한 의미정보를 이용한 명사열의 의미해석

Interpretation of Noun Sequence using Semantic Information Extracted from Machine Readable Dictionary and Corpus

이 경 순* 김 도 완* 김 길 창* 최 기 선*
(Kyung-Soon Lee) (Do-Wan Kim) (Gil-Chang Kim) (Key-Sun Choi)

요 약 명사열의 의미해석은 명사들 사이의 의미적인 관계를 찾는 것으로, 한국어에서 명사열의 출현은 보편적인 현상이며, 그 생성 또한 비교적 자유롭다. 본 논문에서는 기계가독형사전과 코퍼스로부터 명사 사이의 <목적>, <물건-재료>, <원인> 등과 같은 의미관계 정보를 자동으로 추출한다. 추출한 의미관계정보에 기반하여 의미망을 구축하고, 의미 정보와 서술성 명사의 하위 범주 정보를 이용하여 명사열을 해석하는 방법을 제안한다. 본 논문에서는 명사열의 의미 해석의 대상을 한국어 명사열의 대부분을 차지하는 '수식 명사' + '핵심 명사' 형태로 한정하였다. 기계가독형사전과 코퍼스로부터 추출한 의미 정보와 하위 범주를 이용한 명사열의 의미 해석은 기존의 기계가독형사전 기반의 의미 해석보다 정확률 +40.30%, 적용률 +12.73%의 성능 향상을 나타내었다.

주제어 명사열 의미해석, 기계가독형사전, 의미관계, 의미망, 하위범주화

Abstract The interpretation of noun sequence is to find semantic relation between the nouns in noun sequence. To interpret noun sequence, semantic knowledge about words and relation between words is required. In this thesis, we propose a method to interpret a semantic relation between nouns in noun sequence. We extract semantic information from an machine readable dictionary (MRD) and corpus using regular expressions. Based on the extracted information, semantic relation of noun sequence is interpreted. And, we use verb subcategorization information together with the semantic information from an MRD and corpus. Previous researches use semantic knowledge extracted only from an MRD, but our method uses an MRD, corpus, and subcategorization information to interpret noun sequences. Experimental result shows that our method improves the accuracy rate by +40.30% and the coverage rate by +12.73% better than previous researches.

1 서론

명사열은 명사와 명사가 연결되어 하나의 의미 단위

를 이루는 구성으로 형태상으로는 복합어에서 구 구성 까지를 명사열로 볼 수 있다[14]. 한국어 문장에서 명사의 연결로 이루어지는 '명사열'의 출현은 보편적인 현상이며, 그 생성 또한 비교적 자유롭다.

명사열의 의미 해석은 명사열을 이루는 각 명사들이 서로 어떠한 의미 관계로 연결되어 하나의 의미 단위를 이루는가를 밝히는 것이다. 예를 들어, '모퉁이 가

* 한국과학기술원 전산학과
Division of Computer Science
Department of Electrical Engineering & Computer Science
KORTERM/KAIST
kslee, kschoi@world.kaist.ac.kr
dwkim, gckim@csonline.kaist.ac.kr

주소 : 대전시 유성구 구성동 373-1, 우:305-701

373-1 Kusong-dong, Yusong-gu, Taejeon, 305-701

연구분야 : 정보검색, 한국어정보처리, 자연언어처리

투고분야 : 한국어 정보처리

Tel : 042-869-5565

Fax : 042-869-8790

1) 본 논문에서는 두개의 명사가 연속으로 나타나는 것을 '명사열'이라 정의한다. [21]는 영어에서 Noun Sequences를 Noun Compounds와 같은 의미로 사용했고, [14]은 '명사 연결 구성'이라 하였다.

계'라는 명사열의 경우 수식 명사인 '모퉁이'는 핵심 명사인 '가게'가 존재하는 위치를 알려준다. 따라서 '모퉁이'는 수식 명사가 핵심 명사의 위치를 알려주는 의미 관계로 해석된다.

명사열의 의미 해석에 관한 기존의 연구는 명사의 의미 자질을 이용하는 방법과 명사 사이의 의미 관계를 이용하는 방법으로 크게 나뉜다. 명사의 의미 자질을 이용하는 방법은 같은 의미 자질의 결합을 가지지만, 서로 다른 의미 해석을 가지는 경우 그 의미 해석이 어렵다. 또 명사 사이의 의미 관계를 이용하는 방법은 명사열의 의미 해석에 필요한 의미 관계를 기계가독형사전(MRD: Machine Readable Dictionary)에만 의존하므로 다양한 의미 관계의 추출이 어렵다.

본 논문에서는 기계가독형사전과 코퍼스(corpus)로부터 의미 정보를 추출하고, 추출한 의미 정보와 하위 범주 정보를 이용하여 '명사열'을 해석하는 방법을 제안한다. 명사열의 의미 해석에 필요한 의미 정보의 추출을 위해 기계가독형사전과 코퍼스를 동시에 이용하여, 의미 정보의 부족을 보완하고자 한다. 또, 명사열의 의미 해석에는 명사열을 이루는 명사의 특성에 따라 의미 정보에 더불어 하위 범주 정보를 이용하여 좀더 정확한 의미 해석이 가능하게 한다. 본 논문에서는 명사열의 의미 해석의 대상을 한국어 명사열의 대부분을 차지하는 '수식 명사' + '핵심 명사' 형태로 한정한다 [8,14].

2 관련 연구

2.1 의미 정보의 추출

의미 정보란 단어의 개념을 나타내는 것으로서, 의미 자질(semantic feature)정보와 의미 관계(Semantic Relation)정보로 나눌 수 있다. 의미 자질 정보는 개념의 상하위어 관계(hypernym / hyponym)로 이루어진 IS_A 계층 구조로 표현되며, 의미 관계 정보는 목적 관계, 위치 관계, 시간 관계, 원인-결과 관계 등 사용 목적에 따라 다양하게 정의된다. 이러한 다양한 의미 정보를 이용하여 구축한 지식 베이스로 워드넷(WordNet)이 있다[17]. 워드넷은 단어 개념(synset) 사이의 상하위 개념 뿐만 아니라 반대 개념(antonym), 구성 개념(meronym/holonym) 등 다양한 의미 정보를 사용하여 구축한 지식 베이스로 자연언어처리에서 중요한 지식원 역할을 한다. 워드넷과 달리 다양한 의미 정보를 자동으로 구축하기 위한 연구로 [19]가 있다. 이 연구에서는 단어 사이의 의미 정보를 16개의 의미 관

계²⁾로 정의하고, 구문 분석기를 이용하여 기계가독형사전으로부터 자동으로 추출하였다. [21]은 명사열의 의미 해석에 필요한 의미 정보를 [19]와 동일한 방법으로 추출하였다. [21]은 의미 정보를 의미관계 정보와 의미자질 정보로 나누고, 구문패턴을 이용하여 MRD의 표제어와 뜻풀이말(definition)로부터 의미 관계 정보와 의미 자질 정보를 추출하였다.

한국어에서는 기계가독형사전을 이용하여 명사 의미 계층 구조를 구축한 연구가 [13]에 의해 이루어졌다. 이 연구는 서술성 명사, 복합 명사 및 외래어를 제외한 명사를 작업 대상으로 선정하고, 선정한 명사들의 사전 뜻풀이말을 이용하여 상위어를 결정하였다. 이것은 기존의 연구들이 상위 개념부터 하위 개념을 정해가는 방식을 사용한 것과는 달리, 각 명사들의 단계별 상위어를 정해주는 bottom-up 방식을 사용함으로써, 객관성을 지닌 명사 의미 계층 구조를 구축할 수 있게 하였다. 또, [10]은 영어 워드넷을 이용하여, 한국어 워드넷을 설계하고 구현하였다. 한국어 워드넷을 구축하기 위해 5,000개의 명사를 선정하고, 선정된 명사의 사전 뜻풀이말을 이용하여 한국어 명사 워드넷의 단말 노드와 상위 레벨을 개발하였다. 그 다음 번역된 영어 워드넷을 이용하여 개발된 한국어 명사 워드넷의 레벨을 수정하고 확장하는 작업을 수행하였다. 이 논문에서 구축한 한국어 명사 워드넷은 [13]이 구축한 것과 같은 명사 의미 계층 구조이지만, 계층을 이루는 것이 하나의 단어가 아니라 동의어 집합으로 표현된 어휘개념(synset)이라는 점에서 의미가 있다. 최근에는 [3,4,12]가 [19]의 연구를 한국어에 적용시키기도 하였다.

이와 같이 단어 사이의 다양한 의미 정보를 추출하고 지식 베이스를 구축하는 작업은 모두 의미 정보를 추출하는 지식원으로 기계가독형사전을 사용한다는 공통점이 있다.

2.2 명사열의 의미해석

2.2.1 의미 자질을 이용한 의미 해석

명사의 의미 자질을 이용하여 명사열의 의미를 해석하는 방식에서는 의미의 계층 관계를 이용하여 명사를 분류한 뒤, 이를 이용하여 의미 자질 값을 추출하고, 추출한 의미 자질 값의 결합을 통하여 명사열을 해석

2) [19]에서는 단어 사이의 의미 관계를 Cause, Domain, Hypernym, Location, Manner, Material, Means, Part, Possessor, Purpose, Quespernym, Synonym, Time, Typical-Object, Typical-Subject의 16개로 정의하였다.

하는 순서를 따른다. 한국어를 대상으로 한 것에는 [6.11.5] 등의 연구가 있다. 이 방식에서 사용하는 명사의 의미는 계층 관계로 분류되며, 이 계층 관계는 하위 개념이 상위 개념의 속성을 물려받는 계층적인 트리 구조인 IS_A계층 구조를 구성한다. 명사의 IS_A계층 구조를 통한 의미 자질값의 추출은 하나의 명사가 여러 개의 의미를 지니는 다의어인 경우 문제가 될 수 있다. 의미 자질 값의 추출은 용언의 구분 패턴이나 격 분류에서 요구하는 정보를 기준으로 하며, 명사의 의미 분류에서 계층의 일부를 의미 자질 값으로 정의한다. [6]에서는 의미 자질 값으로 행위 주체, 동작성 명사, 신분, 장소, 도구 등을 정의하였다.

〈표 1〉 명사열의 의미론적 관계

의미 관계	수식 명사 의미자질	핵심 명사 의미자질
sort_of	종류	조직
	식물	나무
compose_of	인간	조직
measure_of	장소, 인공물	온도
made_from	재료	제품

명사의 의미 자질을 이용한 명사열의 의미 해석은 명사열을 이루는 수식 명사와 핵심 명사의 의미 자질의 결합 가능 여부를 이용한다. 〈표 1〉은 [11]에서 정의한 명사열 의미 분류의 일부이다.

이와 같이 의미 자질을 이용하여 명사열의 의미를 해석하는 방법은 매우 많은 의미 자질들의 결합 가능성을 모두 고려하여야 하고, 하나의 명사열이 복수개의 의미를 가질 때 이를 해석하기 어렵다는 문제점이 있다. 또, 다음과 같이 수식 명사와 핵심 명사가 같은 의미 자질을 갖는 명사열이지만 전혀 다르게 의미 해석되는 경우를 올바르게 처리하지 못한다.

- 아버지(사람) 책(책)
 - 〈소유 관계〉 (아버지가 소유한 책)
- 사용자(사람) 설명서(책)
 - 〈목적 관계〉 (사용자를 위한 설명서)

2.2.2 의미 관계를 이용한 의미 해석

의미 관계(semantic relation)를 이용하는 방법은

명사열을 이루는 수식 명사와 핵심 명사 사이의 의미 관계의 연결로부터 추론을 통하여 명사열의 의미를 해석한다. 이러한 방식을 이용하여 명사열을 해석한 연구로는 [21]이 있다. 이 연구에서는 기계가독형사전으로부터 명사가 가질 수 있는 의미 정보를 추출하고, 명사열을 이루는 수식 명사와 핵심 명사가 추출한 의미 관계에서 어떻게 연결되는가에 따라 적절한 의미 해석을 내린다. 명사열의 의미 해석은 구축된 단어의 의미 정보를 이용하여 이루어지며, 명사열이 가질 수 있는 가능한 의미 해석에 따라 명사열을 분류한 후, 추론 과정을 거쳐 적절한 의미 관계를 찾는 방식으로 수행된다. 이러한 명사열의 의미 해석 방법은 명사열의 의미 해석에 필요한 의미 정보의 추출을 전적으로 기계가독형사전에 의존하는 문제가 있다. 기계가독형사전은 단어의 의미를 정의하는데 사용하는 어휘와 문형의 한계를 가지며, 기계가독형사전만을 사용하여 의미 정보를 얻는다면 명사열의 의미 해석에 필요한 다양한 의미 정보의 추출이 어렵다.

기존 연구에서 살펴본 것과 같이, 명사의 의미 자질을 정의하고 정의된 의미 자질의 결합으로 명사열을 의미 해석하는 방법은 같은 의미 자질을 갖지만 다른 의미를 지니는 명사열의 해석이 어렵다는 문제점이 있다. [21]은 명사열의 의미 해석에 필요한 의미 정보의 추출을 기계가독형사전에 의존한다. 따라서, 명사열 의미 해석에 필요한 많은 어휘와의 다양한 의미 관계 추출에는 어려움이 있다.

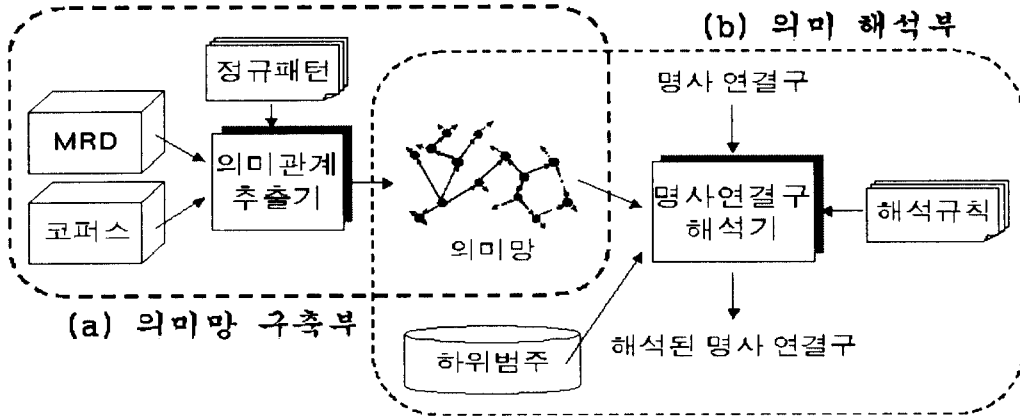
3. 기계가독형사전과 코퍼스에서 추출한 의미정보를 이용한 의미망 구축

본 논문에서 제안하는 명사열의 의미 해석 시스템은 (그림 1)과 같은 구조이다. 그림의 왼쪽 (a)는 의미망 구축으로, 기계가독형사전과 코퍼스로부터 정규 패턴을 이용하여 의미 정보를 추출하여 의미망을 구축하는 부분이다. 기계가독형사전과 코퍼스로부터 〈표 2〉와 같이 17개의 의미 관계와 7개의 의미 자질을 추출한다. 그림의 오른쪽 (b)는 하위 범주와 (a)에서 구축한 의미망을 이용하여 명사열의 의미를 해석하는 부분이다.

본 논문의 실험에서 기계가독형사전과 코퍼스에서 의미관계 정보를 추출하기 위해 정의한 대표적인 정규 패턴의 일부를 〈부록 1〉과 〈부록 2〉에 정리하였다.

3.1 기계가독형사전을 이용한 의미 정보 추출

기계가독형사전에서는 표제어와 뜻풀이말로부터 의미 정보를 추출한다. 정규패턴(regular expression)을



(그림 1) 의미관계 정보 추출 및 명사어의 의미 해석 시스템 구조

<표 2> 추출하는 의미관계와 의미자질 정보

의미 관계	정의	의미자질
A (수단) B	B는 A의 수단	±추상물
A (시간) B	B는 A의 시간	±동물
A (원인) B	B는 A의 원인	±집단, 조직
A (행위-대상) B	B는 A행위의 대상	±사람
A (전체-부분) B	B는 A의 부분	±위치
A (행위-주체) B	B는 A행위의 주체	±물질
A (상위어) B	B는 A의 상위어	±시간
A (도구) B	B는 A를 위한 도구	
A (위치) B	B는 A의 위치	
A (물건-재료) B	B는 A의 재료	
A (목적) B	B는 A를 수행하는 목적	

(a) MRD의 뜻풀이말과 품사부착상태

상자
 나무나 두꺼운 종이 등으로 만든 네모난 그릇
 나무/ncn+나/jcj 두껍/paa+~/etm 종이/ncn 등/nbn+으로
 /jca 만들/pvg+~/etm 네모나/pvg+~/etm 그릇/ncn

(c) 정규패턴을 이용하여 추출한 의미관계정보

상자 — (상위어) —> 그릇
 상자 — (물건-재료) —> 나무
 상자 — (물건-재료) —> 종이

(b) 의미관계정보 추출을 위한 정규패턴

상하위어 관계
 조건: 뜻 풀이말의 끝에 있는 핵심어 (A)
 관계: [표제어] — (상위어) —> [A]
 재료 관계
 조건: B/ncn 으로/jca 만들/pvg ~/etm NP
 관계: [표제어] — (물건-재료) —> [B]

(그림 2) 단어의 뜻풀이말, 의미관계정보 추출 패턴 및 추출한 의미 정보

정의하여 다양한 의미 정보를 자동으로 추출한다. 이때 추출하는 의미 정보는 '단어 -(의미 관계) → 단어'와 같은 형태의 의미 관계 정보와 '+의미 자질'과 같은 형태의 의미 자질 정보이다.

의미 정보를 추출하는데 사용하는 정규 패턴은 품사 부착된 단어의 나열로 이루어지며, 이것은 사전에 있는 뜻풀이말의 일부를 분석하여 구축한다. 의미 자질은 기계가독형사전의 표제어가 가지는 상위에 따라 결정한다.

(그림 2)는 기계가독형사전의 구조, 정규 패턴의 일부, 그리고 추출할 수 있는 의미 정보의 예이다. (그림 2)의 뜻풀이말에서 '상자'는 '나무'나 '종이'를 재료로 만들어지며, 그 상위어는 '그릇'이다. 사전에 정의되어 있는 '상자'의 뜻풀이말과 이것의 품사 부착된 상태가 (a)부분이다. (b)는 기계가독형사전으로부터 <상위어 관계>와 <물건-재료 관계>를 추출하기 위한 정규 패턴이다. (a)의 표제어와 품사 부착된 뜻풀이말에 (b)의 정규 패턴을 적용하면 (c)와 같은 의미 정보를 추출할 수 있다.

3.2 코퍼스를 이용한 의미 정보 추출

사전의 뜻풀이말은 표제어를 설명하기 위해 표제어와 의미적으로 관련된 단어를 사용하므로 의미 정보 추출의 중요한 지식원이 된다. 그러나, 뜻풀이말에는 단어의 의미를 정의하는데 제한된 개수의 단어를 사용하며, 경우에 따라 뜻풀이말이 매우 짧아 명사열의 해석에 필요한 다양한 의미 정보의 추출에 어려움이 따르기도 한다.

실제 사전에서 정의하고 있는 '설명서'의 뜻풀이말과 이 뜻풀이말로부터 추출할 수 있는 의미 정보는 다음과 같다.

[뜻풀이말] **설명서** : 내용이나 이유·사용법 등을 설명한 글
 [의미정보] 설명서 -(상위어)→ 글

그러나, 이 정보는 '사용자 설명서'와 같이 <목적 관계>로 이루어진 명사열의 의미 해석에 필요한 정보로는 부족하다. 명사열의 의미 해석에 필요한 의미 정보의 추출을 사전에만 전적으로 의존할 때 발생할 수 있는 의미 정보의 부족 현상을 해결하기 위해 코퍼스를 사용한다.

코퍼스는 사전에 비하여 단어 사이의 의미 관계를 정의하는 정규 패턴의 출현 빈도가 적지만, 사전보다

다양한 단어와 문형이 나타나므로 다양한 의미 관계의 추출이 가능하다. 또, 코퍼스를 확장함에 따라 단어가 가지는 의미 정보를 확장시킬 수 있는 장점이 있다.

코퍼스에서 의미 정보를 추출하는 방법은 사전에서 추출하는 것과 유사하다. 의미 정보의 추출은 코퍼스로부터 추출할 수 있는 의미 정보를 정의하고, 정의된 의미 정보를 나타낼 수 있는 정규 패턴을 코퍼스에 적용하여 수행한다.

〈표 3〉 기계가독형사전과 코퍼스로부터 <목적관계> 추출을 위한 정규패턴

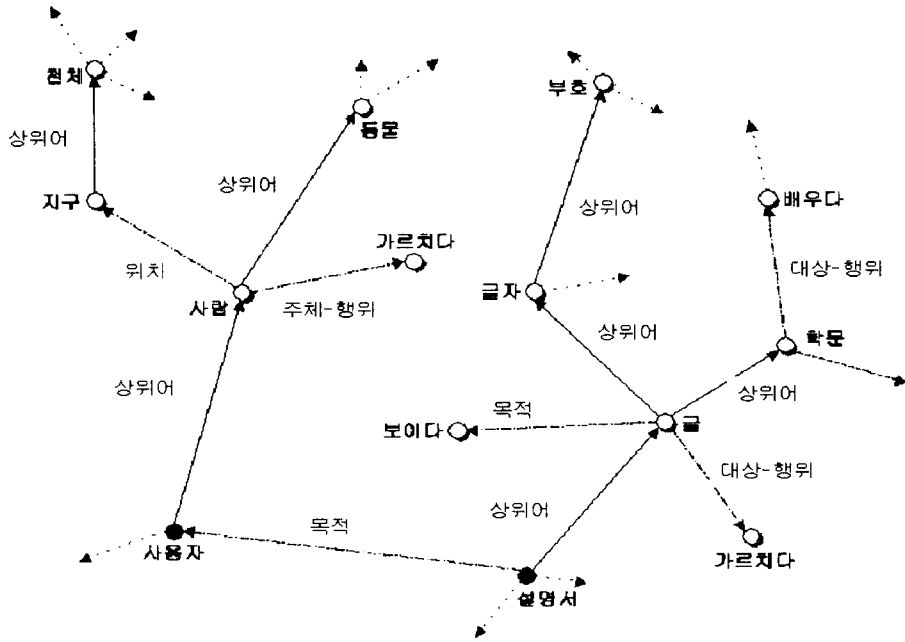
기계가독형사전	정규패턴	A/ncn (을 를)/jco 위하/pvg 어/ecsc
	의미관계	[표제어] -(목적)→ [A]
코퍼스	정규패턴	B/ncn (은 는)/jxt C/ncn (을 를)/jco 위하/pvg L/etm NP
	의미관계	[B] -(목적)→ [C]

〈표 4〉 기계가독형사전과 코퍼스에서 추출된 의미정보

기계가독형사전	설명서: 내용이나 이유·사용법 등을 설명한 글	설명서 -(상위어) → 글
코퍼스	Solaris 고급 사용자 설명서는 Solaris 시스템소프트웨어 사용자를 위한 책이다.	설명서 -(목적) → 사용자

의미 정보를 추출할 때 사용하는 정규 패턴은 코퍼스를 분석하여 구축하며, 사전에서 사용하는 것과는 그 형식이 다르다. 코퍼스에 사용하는 정규 패턴은 단문의 형태를 반영하는 형태로 이루어진다. 〈표 3〉은 기계가독형사전에서 사용되는 정규 패턴은 표제어가 주어로서 생략된 형태임에 반해 코퍼스에서 사용되는 정규 패턴은 'B는 C를 위한 NP'와 같이 하나의 단문 구조를 반영하고 있음을 보인다.

사전을 위한 정규 패턴과 코퍼스를 위한 정규 패턴을 이용하면 〈표 4〉와 같이 기계가독형사전에서는 '설명서'에 대한 <상위어 관계>를 코퍼스에서는 기계가독형사전에서 추출할 수 없었던 <목적 관계>의 의미 관계 정보를 추출할 수 있다. 이와 같이, 다양한 문형을 반영하는 정규 패턴을 이용하여 대량의 코퍼스에서 의미 정보를 추출하면 기계가독형사전만을 사용할 경우



(그림 3) 의미관계에 기반한 의미망의 일부

에 발생할 수 있는 의미 정보의 부족 문제를 완화할 수 있다.

를 이용한 명사열의 의미 해석 방법에 대해 설명한다.

3.3 의미관계정보에 기반한 의미망의 구축

정규 패턴을 이용하여 기계가독형사전과 코퍼스에서 추출한 의미 정보는 전체 단어가 의미 관계의 링크를 통해 연결된 의미망 형태의 지식 베이스를 이룬다. 이 의미망은 단어 사이의 상·하위어 관계로 단어 의미 사이의 계층이 형성되고, 형성된 계층 구조 내에서 단어들어 다양한 의미 정보로 다시 연결되는 형태가 된다. (그림 3)은 의미 정보 추출을 통해 형성된 의미망 중에서 '사용자'와 '설명서'를 중심으로 한 주변의 의미망을 보인다.

본 논문에서는 명사의 의미 정보를 명사열의 의미 해석에 필요한 것으로 한정하였지만, 계층을 이루는 각 단어가 가질 수 있는 의미 정보는 필요에 따라 더욱 다양하게 정의하고 추출할 수 있다. 의미 정보를 추출하는데 사용하는 코퍼스를 늘리고 정규 패턴을 추가한다면 지식 베이스를 이루는 의미망의 크기도 더욱 증가할 것이다.

4. 의미 정보를 이용한 명사열의 의미 해석

추출한 의미 정보와 서술성 명사의 하위 범주 정보

〈표 5〉 한국어 명사열의 의미 분류

명사열의 의미 관계	예제	해석 예제
주체-행위 관계	경찰 명령	경찰이 명령하다
대상-행위 관계	자동차 운전	자동차를 운전하다
목적 관계	사용자 설명서	사용자를 위한 설명서
소유 관계	개인 회사	개인 소유의 회사
시간 관계	아침 출근	아침에 출근하다
원인-결과 관계	가출 상치	가출로 인한 상치
위치 관계	구내 식당	구내에 위치한 식당
재료-물건 관계	가족 가방	가족으로 만든 가방
전체-부분 관계	자동차 바퀴	자동차의 부분인 바퀴

4.1 명사열의 의미 분류

본 논문에서 대상으로 하는 명사열은 '수식 명사' + '핵심 명사' 형태의 명사열이다. 이러한 형태의 명사열을 의미 관점에서 분석하려는 많은 연구가 있었으나, 모든 명사열을 설명할 수 있는 의미 분류는 정립되지 않았다. 또한 하나의 명사열이 둘 또는 그 이상의 여

러 의미로 해석될 수도 있다 [21.16].

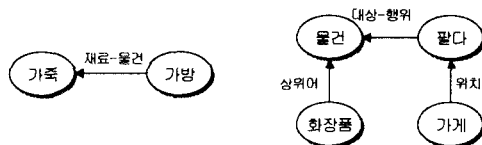
최근 [21]은 '수식 명사가 wh-의문에 대한 대담으로 핵심 명사와 이루는 가장 적당한 의미 관계'를 기준으로 명사어를 14개로 분류하여 의미 해석을 수행하였다. 또, [16]은 일본어 명사구의 한 형태인 N1 no N2 를 N1과 N2가 이루는 의미 관계에 따라 크게 8개로 분류하고 NTT 사전³⁾을 이용하여 그 의미 해석을 수행하였다. 한국어에서는 (6)과 (11)이 수식 명사와 핵심 명사가 갖는 의미 자질의 결합 가능성에 따라 30여개로 명사어를 분류하고, 의미 해석을 수행하였다. 본 논문에서는 임의로 선택한 한국어 명사어 100개의 분석 결과와 기존 연구의 명사어 분류를 참고하여, 한국어 명사어를 <표 5>와 같이 분류한다.

4.2 핵심 명사의 특성에 따른 의미 해석

명사어의 의미 해석은 명사어를 이루는 수식 명사와 핵심 명사가 <표 5>의 의미 분류 중 어떠한 것에 가장 어울리는가를 밝히는 작업이다. 본 논문에서는 명사어를 이루는 '핵심 명사'의 특성에 따라 다양한 의미 정보를 사용하여 의미를 해석한다. 핵심 명사가 비서술성 명사인 경우는 의미관계정보를 이용하여 의미 해석을 하고, 상태·동작을 나타내는 서술성 명사인 경우는 추출한 의미 정보와 서술성 명사가 가지는 하위 범주 정보를 같이 이용한다.

4.2.1 핵심 명사가 비서술성 명사인 경우

명사어를 이루는 핵심 명사가 비서술성 명사⁴⁾인 경우는 앞절에서 추출한 의미 정보를 이용하여 의미를 해석한다.



(그림 4) 가죽 가방: <물건-재료 관계>과 화장품가게: <목적 관계>

<표 6> 명사어의 의미 해석 규칙

의미 관계	규칙 번호	수식명사의 조건	핵심명사의 조건
주체-행위 관계	규칙 1	<주체>	의미망 연결
	규칙 2	의미망 연결	<행위>
대상-행위 관계	규칙 3	<대상>	의미망 연결
	규칙 4	의미망 연결	<행위>
목적 관계	규칙 5	의미망 연결	<목적>
	규칙 6	의미망 연결	<도구>
	규칙 7	의미망 연결	<위치>
소유 관계	규칙 8	+사람 or +동물	
시간 관계	규칙 9	<시간>	의미망 연결
	규칙 10	+시간 and -위치	
원인-결과 관계	규칙 11	<원인>	의미망 연결
	규칙 12	의미망 연결	<사고>
위치 관계	규칙 13	<위치>	의미망 연결
	규칙 14	의미망 연결	<위치>
	규칙 15	+위치	+위치 or 없음
재료-물건 관계	규칙 16	<재료>	의미망 연결
	규칙 17	의미망 연결	<물건>
	규칙 18	+재료	-추상물
전체-부분 관계	규칙 19	<전체>	의미망 연결
	규칙 20	의미망 연결	<부분>
	규칙 21	+조직	+사람 or +동물

예를 들어 '가죽 가방'이라는 명사어는 의미망에서 (그림 4)와 같이 연결되며, 이러한 의미 관계의 연결은 '가죽을 재료로 하여 만들어진 가방'이라는 뜻을 지니게 된다. 따라서, '가죽 가방'은 <재료-물건 관계>로 이루어진 명사어로 해석한다. '화장품 가게'도 '화장품'이라는 물건을 팔기 위한 장소인 '가게'와 같은 의미를 지니며, <목적 관계>로 이루어진 명사어로 해석한다. 의미망 내에서 명사어를 이루는 명사들의 상대적인 위치는 '가죽 가방'과 같이 하나의 링크로 직접 연결된 경우 뿐만 아니라, '화장품 가게'와 같이 여러 단어와 의미적 연관성을 지니면서 연결이 이루어지기도 한다.

명사어를 이루는 명사들이 여러 단어와 여러 단계의 의미적 연관성을 지니며 의미망 내에서 연결이 이루어진 경우, 추론을 통하여 의미 해석을 수행한다. 본 논문에서는 이와 같은 추론이 <표 6>의 의미해석 규칙을 통하여 이루어진다. <표 6>의 추론 규칙은 3절에서 추출한 의미 정보가 이미 단어와 단어 사이의 의미 관계를 나타내고 있기 때문에 가능하다. 예를 들어 '가죽 가방'의 핵심 명사 '가방'은 의미망 내에서 <재료-물건>의 의미 관계 정보로 '가죽'과 연결된다. 이것은 <재료-물건 관계>를 해석하는 추론 규칙 17에 적용된다.

의미해석 규칙은 의미 관계 정보를 이용한 것과 의미 자질 정보를 이용한 것으로 나눌 수 있다. 의미 자

3) NTT 사전은 일본의 NTT Communication Science Laboratory에서 구축한 의미 자질 트리로써, 3,000개의 노드와 의미 자질, 300,000개의 명사를 포함한다.

4) '-하다', '-되다'와 같은 접미어가 붙어서 상태·동작을 나타내는 명사를 서술성 명사라 하며, 그렇지 않은 경우를 비서술성 명사라 한다. 서술성 명사는 ncpa, ncps의 품사 태그를 갖으며, 비서술성 명사는 ncn의 품사 태그를 갖는다.

질 정보만으로 이루어진 규칙은 추출한 의미 정보가 부족하거나, 의미 자질의 결합만으로도 의미를 결정할 수 있는 경우를 위해 정의한 것이다. 특히, <소유 관계>의 추론 규칙은 3절에서 단어 사이의 소유 관계를 나타내는 의미 관계를 추출하지 않기 때문에 의미 자질의 결합만으로 추론 규칙을 정의한다.

명사열을 이루는 수식 명사와 핵심 명사가 의미망 내에서 의미해석 규칙을 만족하며 연결되었다 하더라도, 의미망 내에서 연결되는 링크는 여러 개일 수 있으며, 그 링크 또한 직·간접으로 연결될 수 있다. 이러한 모든 경우를 동일한 의미 해석으로 볼 수는 없으므로, 가능한 여러 해석 중에서 가장 적당한 것을 선택해야 한다. 본 논문에서는 명사열을 이루는 명사의 의미 관계 연결 상태에 따라 가중치를 주는 방법을 사용하여 가장 적절한 의미 해석을 선택한다. 수식 명사와 핵심 명사가 의미 관계로 직접 연결된 경우에 가장 큰 가중치를 주며, 단어의 의미 관계를 연결하는 고리 역할을 하는 단어의 수가 증가할수록 작은 가중치를 준다. 또, 하나 이상의 중간 노드를 거치는 경우에는 노드와 노드를 잇는 의미 관계의 종류에 따라 다른 가중치를 주며, 그 크기는 아래의 순서와 같이 <상위어>에 가장 큰 가중치를 부여한다.

1. 상위어
2. 전체-부분, 부분-전체, 물건-재료, 재료-물건
3. 대상-행위, 행위-대상, 주체-행위, 행위-주체

의미해석 규칙을 명사열의 의미 해석에 적용하기 위해서는, 해석하고자 하는 명사열의 수식 명사와 핵심 명사가 의미망 내에서 어떤 형태로든 링크를 가지고 연결이 되어 있어야만 한다.

- (기계가독형 사용자 : 사용하는 사람 사진) (1)
- (기계가독형 설명서 : 내용이나 이유·사용법 등을 설명한 글 사진) (2)
- (코퍼스) Solaris 고급 사용자 설명서는 Solaris 시스템 소프트웨어 사용자를 위한 책임니다. (3)

'사용자 설명서'라는 명사열을 의미 해석하기 위해 사전으로부터 추출할 수 있는 의미 정보는 각각 (1'), (2')이며, 이러한 정보만으로는 의미해석규칙에 의해 <소유 관계> (사용자가 소유한 설명서)로 해석이 된

다. 그러나, 코퍼스(3)으로부터 (3')과 같은 의미 정보를 추출할 수 있으며, 이 정보를 이용하여 사용자 설명서를 <목적 관계> (사용자를 위한 설명서)로 더 자연스럽게 의미 해석한다.

- 사용자 -(상위어)→ 사람 (1')
- 설명서 -(상위어)→ 글 (2')
- 설명서 -(목적)→ 사용자 (3')

코퍼스로부터 추출한 의미 정보를 통해서 보다 더 자연스러운 해석 결과를 얻을 수 있다.

4.2.2 핵심 명사가 서술성 명사인 경우

명사열을 이루는 핵심 명사가 서술성 명사인 경우는 의미 관계 정보 뿐만 아니라, 핵심 명사가 가지는 하위 범주 정보를 사용한다. 본 논문에서 사용하는 하위 범주 정보는 (7)에서 구축한 서술성 명사의 하위 범주⁵⁾이며, (그림 5)의 (b)와 같은 형태로 이루어져 있다.

(그림 5)의 (a)는 기계가독형사전에서 나타나는 '거래'의 뜻풀이말과 여기서 추출할 수 있는 의미 정보이며, (b)는 서술성 명사인 '거래'가 갖는 하위 범주 정보를 나타낸 것이다. (그림 5)의 (b)는 서술성 명사인 '거래'가 'N1이 N2를 거래하다'와 같은 단문 형태로 쓰이며, N1에는 사람, 단체 등을 나타내는 명사가, N2에는 물건, 돈, 상품, 증권, 서비스 등을 나타내는 명사가 쓰일 수 있음을 보인다.

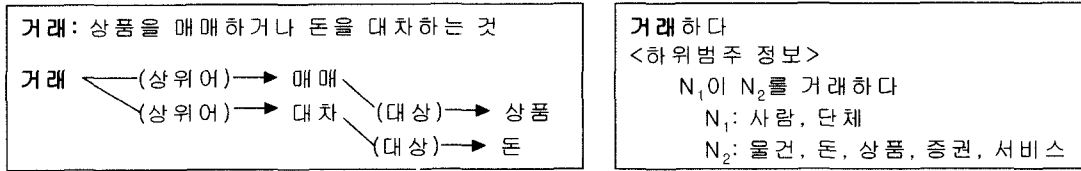
하위 범주 정보는 기계가독형사전에서는 추출할 수 없는 의미 정보(<행위의 주체>, <행위의 대상>)에 대한 의미 정보)를 포함하고 있음을 알 수 있다. 이와 같은 하위 범주 정보는 기계가독형사전, 코퍼스와 함께 명사열의 의미 해석에 필요한 의미 정보를 제공한다.

'서비스 거래'라는 명사열에서 '거래'는 서술성 명사이며 하위 범주 정보를 가지므로, (그림 5)에서 (b)의 N2 하위 범주 정보를 이용하여 해석이 가능하다.

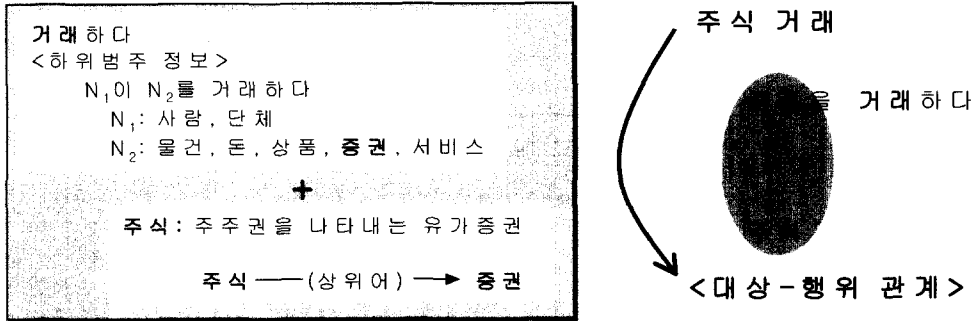
서비스 거래 : 서비스를 거래하다
 ⇒ 대상-행위 관계

그러나, '주식 거래'와 같이 명사열을 이루는 수식 명사가 핵심 명사의 하위 범주와 일치하지 않는 경우가 있다. 이러한 명사열의 의미 해석은 명사열을 이루는

5) '하다' 접미어가 붙어 상태·동작을 나타내는 명사 중 사용 빈도가 높은 것에 대하여, 하위 범주를 구축하였다.



(그림 5) 서술성 명사 '거래'의 의미 정보와 하위 범주 정보



(그림 6) 주식거래: 대상-행위 관계

수식 명사와 핵심 명사의 하위 범주에 나타나는 명사 사이의 의미 관계를 통하여 의미를 해석한다. 기계가독형사전에서 '주식'의 뜻풀이말과 의미관계 추출을 위한 정규패턴을 이용하여 의미정보를 추출할 수 있다.

주식 : 주주권을 나타내는 유가 증권
 주식 -(상위어)-> 증권

따라서, 핵심 명사 '거래'의 하위 범주 정보와 수식 명사 '주식'의 의미 정보를 같이 이용하여 '주식 거래'라는 명사어를 해석할 수 있게 된다 (그림 6).

주식 거래 : N1(주식)을 거래하다 ⇒ 대상-행위 관계

이와 같이, 해석하고자 하는 명사어가 기계가독형사전과 코퍼스에서 추출한 의미 정보만으로 의미 해석이 불가능한 경우 핵심 명사의 하위 범주 정보를 이용하여 해석한다. 또, 수식 명사가 핵심 명사의 하위 범주에 나타나지 않는 경우는 수식 명사의 의미 정보와 핵심 명사의 하위 범주 정보를 같이 이용하여 의미 해석을 가능하게 한다. 이처럼 하위 범주 정보는 핵심 명사가 서술성 명사인 명사어의 의미 해석에 필요한 의미 정보이다.

5. 실험 및 평가

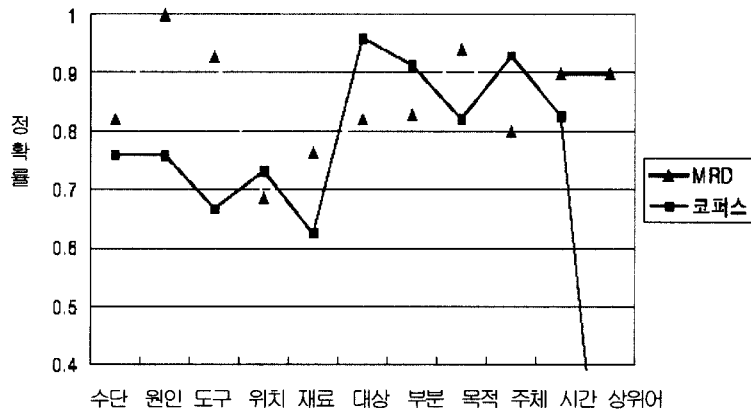
정규 패턴을 이용하여 기계가독형사전과 코퍼스로부터 의미 정보를 추출하는 실험과 자동 추출된 의미 정보와 하위 범주 정보를 사용하여 명사어의 의미 해석 실험을 수행한다.

(표 7) 실험에 사용된 기계가독형사전과 코퍼스

지식원	크기	비율	
기계가독형사전	뜻풀이말 8,780문장	3.22%	
코 퍼 스	소셜	251,279 문장	91.89%
	수필	13,385 문장	4.89%
전 체	273,444 문장	100%	

(표 8) 기계가독형사전과 코퍼스로부터 추출한 의미관계정보 개수

의미관 계정보	기계가독형사전	코퍼스	의미관 계정보	기계가독형사전	코퍼스
수단	317	576	위치	71	1,157
원인	18	32	재료-물건	3	16
주체-행위	2,054	11,242	물건-재료	35	45
대상-행위	5,523	28,177	목적	309	182
전체-부분	64	22	위치	358	1,174
도구	14	3	시간	624	211
상위어	7,947	0	전체	18,262	41,680



(그림 7) 기계가독형사전과 코퍼스의 의미 정보별 정확률

5.1 의미 정보 추출 실험

의미 정보 추출은 기계가독형사전과 코퍼스를 이용한다. 기계가독형사전의 표제어는 5,946개이다. 의미 정보 추출에 사용되는 데이터는 <표 7>과 같다. 기계가독형사전의 뜻풀이말은 품사 태깅 [1]을 수행하였다. 의미 정보 추출 실험은 수동으로 구축한 129개의 정규 패턴을 이용하였다.

실험 1

<표 8>은 기계가독형사전과 코퍼스로부터 추출한 의미 관계 정보의 개수를 보인다. 기계가독형사전에서는 총 5,235개의 단어에 연관된 18,262개의 의미 관계 정보를 추출하여, 단어 당 의미 관계 수가 3.49개를 이루었다. 또, 기계가독형사전과 코퍼스를 동시에 사용한 경우에는 10,255개의 단어에 연관된 53,644개⁶⁾의 의미 관계 정보를 추출하여 단어 당 의미 관계 수가 5.23개를 이루었다. 의미 자질 정보는 기계가독형사전에서만 추출하였으며, 표제어의 상위어에 따라 의미 자질을 결정하였다. 7개의 의미 자질에 대하여 총 996개가 추출되었으며, 정확률 97.32%를 나타내었다.

실험 2

기계가독형사전과 코퍼스에서 추출한 의미 정보 중에서 무작위로 고른 500개의 의미 정보에 대해 정확률을 측정하였다(표 9). <표 9>에서 보는 바와 같이, 기계가독형사전에서 추출한 의미 정보와 코퍼스에서

추출한 의미 정보의 정확률은 각각 80.6%와 82.6%의 결과를 나타내었다. 이것은 기계가독형사전 뿐만 아니라 코퍼스에서 추출한 의미 정보도 유용함을 나타낸다. (그림 7)은 추출된 의미 정보별 정확률을 보인다.

(표 9) 기계가독형사전과 코퍼스에서 추출한 의미 정보의 정확률 측정

의미관계 종류	기계가독형사전		코퍼스	
	샘플수	정확률	샘플수	정확률
수단	5	4 / 5	5	2 / 5
대상-행위	164	144 / 164	384	326 / 384
전체-부분	3	2 / 3		
주체-행위	69	43 / 69	104	82 / 104
상위어	221	182 / 221		
도구	1	1 / 1		
위치	12	6 / 12	4	2 / 4
재료-물건	1	1 / 1		
목적	8	8 / 8	1	1 / 1
시간	16	12 / 16	2	0 / 2
전 체	500	403 / 500 (80.6%)	500	413 / 500 (82.6%)

5.2 명사어의 의미 해석 실험

명사어의 의미 해석은 [7]의 하위 범주 정보와 자동어로 추출한 의미 정보를 이용한다. 실험은 [9]이 구축한 복합 명사 리스트⁷⁾ 중에서 무작위로 고른 450개를 대상으로 하였다. 실험에서는 5명의 실험자 중 과반수 이상이 선택한 의미 해석을 정답으로 간주하였다.

6) 3,160개의 단어와 6,298개의 의미 관계가 기계가독형사전과 코퍼스에서 중복되어 추출되었다.

7) 복합 명사는 명사어의 한 형태이며, 본 실험에서는 두개의 명사로 이루어진 복합 명사를 실험 대상으로 사용하였다.

실험 3

여러 의미 정보들을 사용하였을 때의 의미 해석의 정확률을 비교하여, 본 논문에서 제시한 여러 의미 정보가 명사어의 의미 해석에 끼치는 영향을 알아보았다.

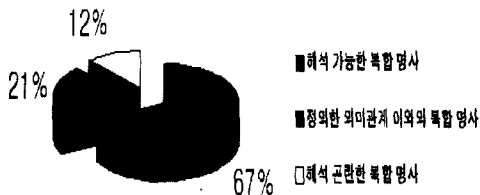
실험에 사용한 복합 명사는 (그림 8)과 같이 구성되어 있다. 450개의 실험 데이터 중 논문에서 제시한 의미 분류에 포함되지 않는 것이 93개, 의미 해석이 불가능한 복합 명사가 54개⁸⁾였다. 실험은 이 둘을 제외한 나머지 303개의 복합 명사에 대해서 수행하였다.

여기서 사용하는 정확률과 적용률은 다음과 같이 정의한다.

$$\text{정확률} = \frac{\text{정답}}{\text{정답} + \text{오답}} \times 100$$

$$\text{적용률} = \frac{\text{정답} + \text{오답}}{\text{정답} + \text{오답} + \text{해석실패}} \times 100$$

실험 결과 <표 10>과 같은 정확률과 적용률을 나타내었다. 결과에서 코퍼스에서 추출한 의미 정보와 하위 범주 정보를 사용한 방법이 기계가독형사전만 사용한 방법보다 정확률은 40.30%, 적용률은 12.73% 증가한 것을 알 수 있다.



(그림 8) 실험에 사용된 복합 명사의 구성 비율

<표 10> 명사어의 의미 해석 실험

사용한 의미 정보	정답	오답	해석실패	정확률	적용률
기계가독형사전(MRD)	35	48	220	42.17%	27.39%
MRD+코퍼스	54	46	203	54.00%	33.00%
MRD+하위범주	55	41	207	57.29%	31.68%
MRD+코퍼스+하위범주	73	38	192	65.76%	36.63%

8) 모든 실험자가 무슨 뜻인지 모르거나, 불가능한 명사의 조합이라고 결정한 것들로 '불 출', '살 남자', '즉시 강제' 등이 있었다.

5.3 오류 분석

5.3.1 의미 정보 추출의 오류 유형

실험 문장 오류 및 품사 태깅 오류

실험에서 사용하는 정규 패턴은 어휘와 품사어를 고려한 것이다. 이러한 정규 패턴을 사용하기 위하여 [1]의 모델을 기계가독형사전의 뜻풀이말에 적용하여 품사 부착된 문장을 만든다. 이 때, 품사 태깅 자체의 오류로 인하여 정규 패턴이 적용되지 못하거나 잘못된 의미 정보를 추출하였다. 또, 의미 정보를 추출하는 문장 자체에 오류가 있어 잘못된 의미 정보를 추출하기도 했다.

병렬 구문 분석 오류

의미 정보를 추출하기 위한 정규 패턴은 단어 단위의 간단한 병렬 구문만을 분석하였다. 따라서, 복잡한 병렬문이나 구 단위의 병렬이 이루어진 경우에는 의미 정보의 추출이 어려웠다.

정규 패턴 오류

의미 정보를 추출하는 정규 패턴을 잘못 정의하거나 모호성을 지닌 정규 패턴이 있었다. 정규 패턴의 오류는 다음 예와 같이 기대하지 못한 명사 '인공', '모양'이 패턴의 한 부분을 차지하는 경우에 자주 발생하였다.

- 부대 : 종이, 피륙 등으로 부대 -(물건-재료)→ 종이 (O)
- 만든 큰 자루 부대 -(물건-재료)→ 피륙 (O)
- 바퀴 : 돌게 하려고 둥근 테 모양으로 만든 물건 바퀴 -(물건-재료)→ 모양 (X)
- 조화 : 인공으로 만든 꽃 조화 -(물건-재료)→ 인공 (X)

정규 패턴을 적용할 때 시소러스 등의 의미 정보를 함께 이용하면 다양한 의미 정보 추출의 정확률을 높일 수 있을 것으로 기대된다.

5.3.2 의미 해석의 오류 유형

의미 정보의 부족 및 의미망 오류

명사어의 의미 해석은 기계가독형사전과 코퍼스에서 자동으로 추출한 의미 정보와 하위 범주 정보를 사용한다. 실험에서 추출된 의미 정보의 오류로 인하여 의

미 해석이 틀리는 경우가 발생하였다. 앞의 예제 '바퀴-(물건-재료)-모양', '조화-(물건-재료)-인공'에서 보았듯이, 정규패턴의 오류는 의미망의 오류에 직접적인 영향을 미친다. 따라서, 정규패턴을 정교화하고, 자동으로 추출된 의미 정보를 수정·보완하는 작업을 거쳐 의미망을 구축하면, 오류를 줄일 수 있을 것으로 기대된다. 또한, 추출한 의미 정보가 부족할 경우에는 의미 해석을 할 수 없다. 명사열 '가스사고'에 대한 의미 해석을 하고자 하는 경우, '가스' 또는 '사고'에 대한 정보가 없다면, 의미 해석이 불가능하게 된다. 이런 의미 정보의 부족은 의미관계 추출을 위한 정규패턴을 확장하고, 시소러스 정보를 이용하고, 추론기법을 이용하는 방법이 필요하다.

해석 규칙의 오류

명사열을 이루는 수식 명사와 핵심 명사가 의미망 내에서 연결이 이루어지지만, 추론을 통한 의미 해석이 어려운 경우가 있었다. 이러한 것들은 아래와 같이 '행위'를 중심으로 '행위의 주체'나 '행위의 대상'이 연결되는 것이 대부분으로, 기계가독형사전과 코퍼스에서 추출한 의미 관계 중 상당수를 <행위-대상> 또는 <대상-행위>와 <행위-주체> 또는 <주체-행위>가 차지하기 때문에 발생하는 현상이다.

사고 --(대상-행위)

--> 일으키다 <--(행위-대상)-- 장애

이러한 의미 관계를 통해 의미망 내에서 연결된 명사 연결구는 의미를 해석하기가 어렵다. 본 실험에서는 이러한 형태로 연결된 명사 연결구에 대해서는 의미 해석 규칙을 작성하지 않았다.

또한, 해석 규칙의 오류로 인하여 잘못된 해석을 선택하는 경우가 있었다. 의미 자질을 이용한 의미 해석 규칙에 이러한 오류가 많았으며, 특히 의미 자질만을 이용한 <소유 관계>의 해석 규칙은 어휘 수준의 의미 해석이 불가능하여 그 결과가 더욱 좋지 못했다. 더욱 다양한 의미 정보의 추출과 해석 규칙의 보완이 요구된다.

5.4 토의

모든 명사열을 설명할 수 있는 의미 분류는 정립되지 않았으며, 본 논문에서는 기존의 연구를 참고하여 자주 등장하는 명사열의 의미 분류만을 실험대상으로 하였다(표 5).

실험 결과 약 21%에 해당하는 93개의 명사열이 의미 분류에 속하지 않았으며, 이들은 <대등 관계>(‘아들 딸’)나 <형상 관계>(‘검정 고무신’), <수단-방법 관계>(‘공동 판매’) 등 새로운 의미 관계의 정의를 필요로 하는 것으로 보인다.

실험결과에서 나타난 여러 가지 오류에서 우선적으로 해결해야 하는 오류들은 다음과 같은 순서이다. 의미관계정보를 추출하기 위한 정규패턴의 오류는 의미망 오류에 직접적인 영향을 미치므로, 정규패턴의 정확한 기술과 정교화가 필요하다. 명사열의 의미 해석을 하기 위해서는 의미망을 이용하는 데, 구축된 의미망의 크기가 작다면, 명사열에 나타난 명사에 대해 정보를 찾을 수 없을 가능성이 높다. 이는 정보의 부족으로 나타나는 현상인데, 정규패턴의 확장, 의미관계의 확장, 시소러스 등을 이용한 명사들 사이의 상하위관계 정보의 확장 등으로 대량의 명사에 대해서 의미관계 정보를 구축할 수 있도록 해야 한다.

6. 결론

본 논문에서는 기계가독형사전과 코퍼스로부터 명사의 의미 정보를 추출하고, 추출한 의미 정보와 하위 범주 정보를 이용하여 한국어 명사열의 의미를 해석하는 방법을 제안하였다. 명사의 의미 정보 추출의 지식 원으로 기계가독형사전과 함께 코퍼스를 사용하여 다양한 의미 정보 추출이 가능함을 보였다. 기존의 연구가 명사의 상·하위어 관계를 중심으로 의미 정보를 수작업으로 구축한 반면, 본 논문에서는 상·하위어 관계, 목적 관계, 시간 관계, 위치 관계 등 다양한 의미 정보를 자동으로 추출하였다. 기계가독형사전과 코퍼스로부터 추출한 의미 정보는 각각 80.6%와 82.6%의 정확도를 나타냈다. 기계가독형사전과 코퍼스로부터 추출한 의미 정보를 결합하여 단어와 단어가 의미 관계의 링크로 연결된 하나의 의미망을 구축하였다. 다양한 분야의 코퍼스로부터 의미 정보를 추출한다면, 자연언어처리 전반에 걸쳐 이용할 수 있는 지식 베이스를 구축할 수 있다.

명사열의 의미 해석은 한국어 명사열의 대부분을 차지하는 '수식 명사 + 핵심 명사' 형태를 대상으로 하였으며, 의미적 중심이 되는 핵심 명사의 특성에 따라 하위 범주 정보와 앞서 추출한 의미 정보를 이용하였다. 명사열의 의미 해석은 기계가독형사전에서 추출한 의미 정보만을 사용한 의미 해석보다 정확률 40.30%, 적용률 12.73%의 성능 향상을 나타내었다. 실험 결과를 통하여 코퍼스와 하위 범주 정보가 명사열의 의미

해석에 필요한 의미 정보를 제공함을 알 수 있었다.

향후 과제로 의미 정보의 추출에 시소러스 등을 사용하여 의미 정보 추출의 정확률을 더 높일 수 있는 방안이 필요하다. 또, 가능한 많은 명사열을 설명할 수 있는 의미 분류에 대한 연구가 요구된다. 의미 정보를 추출하기 위해 사용한 정규 패턴과 명사열을 의미 해석하기 위해 정의한 해석 규칙의 정확률을 높이기 위한 지속적인 데이터 분석이 필요하다.

참고문헌

[1] 강인호, 김재훈, 김길창. 1998. 최대 엔트로피 모델을 이용한 한국어 품사 태깅. 제 10회 한글 및 한국어 정보처리 학술대회.

[2] 금성출판사. 1992. 그랜드 국어사전. 금성출판사.

[3] 김도완. 2000. MRD와 코퍼스를 이용한 명사 연결구의 의미해석. 한국과학기술원 석사학위 논문.

[4] 김도완, 이경순, 김길창. 1999. 의미관계와 문형정보를이용한 복합명사 해석. 제11회 한글 및 한국어 정보처리 학술대회.

[5] 김수남, 원상연, 권혁철, 주종철, 이상기. 1998. 의미 정보를 이용한 한국어 복합명사 분석. 한국정보과학회 가을 학술발표논문집.

[6] 김영훈, 이상조, 한, 영 기계 번역을 위한 명사의 의미 분류와 수식 관계의 표현. 한국정보과학회 가을 학술발표논문집.

[7] 김인택. 1997. 한국어 문장 분석을 위한 용언의 하위범주화에 관한 연구. 시스템 공학 연구소-부산대학교 언어연구교육원 보고서.

[8] 김지영, 권혁철. 1992. 합성명사의 의미 관계 분석 시스템을 위한 지식베이스 구축 기법. 한국정보과학회 가을 학술발표논문집.

[9] 남지순. 1997. 한국어 전자 사전. 인공지능연구센터 기술보고서. CAIR-TR-97-72.

[10] 문유진. 1996. 한국어 명사를 위한 WordNet의 설계와 구현. 정보과학회 논문지 제2권 제4호.

[11] 박미성, 김정해, 이상조. 1991. 명사간의 결합 의미 관계 규명을 통한 명사 합성기 구현. 한국정보과학회 가을 학술발표논문집

[12] 이해중, 조정미, 문준혁, 서정연. 1999. 한국어 어휘지식베이스 구축시스템. 제11회 한글 및 한국어 정보처리 학술대회.

[13] 조평욱, 옥철영. 1997. 한국어 명사 의미 계층 구조 구축. 제 9회 한글 및 한국어 정보처리 학술대회.

[14] 최경봉. 1996. 명사 연결 구성의 해석 원리. 언어.

[15] Dolan, William B. and Vanderwende, L. and Richardson, S. 1993. Automatically Deriving Structured Knowledge Base from On-Line Dictionaries. In *Pacific Association for Computational Linguistics*.

[16] Kurohashi, Sadao and Sakai, Yasuyuki. 1999. Semantic Analysis of Japanese Noun Phrases : A New Approach to Dictionary-Based Understanding. In *ACL99*.

[17] Miller, G. and Beckwith, R. and Fellbaum, C. and Gross, D. and Miller, K. 1990. *Introduction to WordNet: an on-line lexical database*. International Journal of Lexicography 3.

[18] Montemagni, S. and Vanderwende, L. 1992. Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries. In *COLING92*.

[19] Richardson, S. 1997. *Determining Similarity and Inferring Relations in a Lexical Knowledge Base*. Ph.D. thesis, The City University of New York.

[20] Vanderwende, L. 1995a. Ambiguity in The Acquisition of Lexical Information. In *AAAI 1995*.

[21] Vanderwende, L. 1995b. *The Analysis of Noun Sequences using Semantic Information Extracted from On-Line Dictionaries*. Ph.D. thesis, Georgetown University.

〈부록 1〉 기계가독형사전에서 의미관계정보 추출을 위한 정규 패턴 예

의미관계 정보	정규 패턴
표제어 (수단) A	A (을 를)/jco 이용/ncpa 하/xsv 어/ecs
표제어 (원인-결과) A A (원인-결과) 표제어	A 로/jca 인하/pvg 어/ecs A 의/jcm 원인/ncn 이/jcc 되/pvg 어/ecs
표제어 (전체-부분) A A (전체-부분) 표제어	A 로/jca 구성/ncpa 되/xsv ㄴ/etm A (을 를)/jco 구성/ncpa 하/xsv 는/etm
표제어 (상-하위) A	A 의/jcm 하나/ncn 뜻풀이말의 끝에 있는 핵심어(A)
표제어 (도구) A	A 예/jca 사용/ncpa 하/xsv 는/etm 도구/ncn
표제어 (위치) A A (위치) 표제어	A 예/jca 위치/ncpa 하/xsv ㄴ/etm A (을 를)/jco 하/pvg 는/etm (곳 장소)/ncn
표제어 (물건-재료) A A (물건-재료) 표제어	A 로/jca 만들/pvg ㄴ/etm A (을 를)/jco 만들/pvg 는/etm 재료/ncn
표제어 (목적) A	A (을 를)/jco 목적/ncn 으로/jca
표제어 (시간) A	A (을 를)/jco 하/pvg ㄹ/etm 때/nbn

〈부록 2〉 코퍼스에서 의미관계정보 추출을 위한 정규 패턴 예

의미관계 정보	정규 패턴
A (수단) B	B (을 를)/jco 이용/ncpa 하/xsv 어서/ecs A 하/xsv 다/ef
A (원인-결과) B B (원인-결과) A	A (이 가)/jcs 원인/ncn 이/jcc 되/pvg 어/ecs B 하/xsv 다/ef B 로/jca 인하/pvg 어/ecs A 하/xsv 다/ef
A (전체-부분) B B (전체-부분) A	B (은 는)/jxt A 의/jcm 일부/ncn A 로/jca 구성/ncpa 되/xsv ㄴ/etm B
A (도구) B	A (을 를)/jco 사용/ncpa 하/xsv 어/ecs B (을 를)/jco 만들/pvg
A (위치) B B (위치) A	B 예/jca 위치/ncpa 하/xsv ㄴ/etm A A (은 는)/jxt B (을 를)/jco C 하/pvg 는/etm (곳 장소)/ncn
A (물건-재료) B B (물건-재료) A	B 로/jca 만들/pvg 어/ecx 지/px ㄴ/etm A A 로/jca 만들/pvg ㄴ/etm B
A (목적) B	A (은 는)/jxt B (을 를)/jco 위하/pvg ㄴ/etm A (은 는)/jxt B (을 를)/jco 목적/ncn 으로/jca 하/pvg
A (시간) B	A 무렵/ncn B 하/xsv 다/ef