양국어 어휘분류망의 자동 구축

Automatic Construction of Lexical Classification Net for Two Languages

황금하*, 최기선**
한국과학기술원 전산학과 전문용어언어공학연구센터 (*,**)
중국 연변과학기술대학 겸 (*)
{hgh, kschoi}@world.kaist.ac.kr

요 약

본 연구에서는 이미 만들어진 양국어 단일 언어 어휘 분류체계를 이용하여 양국어 어휘 분류등급 간의 개념유사도에 의한 양국어 분류체계간의 연관 관계를 구축하고자 한다. 중국어 유의어사전과 한국어 분류어휘표를 이용하여 양국어 어휘 분류체계에서의 분류등급 간의 개념유사성 및 양국어간의 어휘 유사성에 의하여 어휘분류망을 자동 구축한다. 자동 구축된 어휘분류망을 통하여 한국어 분류어휘표의 어휘 구성 및 분류체계에 대한 분석 평가를 진행할 것이며 나아가 한국어 분류어휘표에 대한 어휘 및 분류체계에 대한 보완을 시도하고자 한다. 본 연구는 한국어 자체 어휘 분류체계의 구축 방법론의 연구에도 어느 정도 도움될 것으로 기대한다.

1. 서론

어휘 개념 유사도의 획득은 거의 자연언어처리 전반 분야에서 그 필요성을 보여주었다. 단일 언어 어휘 개념 유사도 계산에 대한 연구는 자연언어처리의 여러 세부분야에서 많이 진행되어 왔음에도 불구하고 양국어 어휘 개념 유사도 계산에 대한 연구는 많이 진행되어 오지 않았다. 양국어 어휘 개념 유사도 계산은 양국어 정렬시스템, 기계번역에서의 어휘 변환 및 다국어 정보검색 등 다양한 응용분야를 가지고 있다. 그러나 지금까지 제공되는 지식이 대부분 단일 언어 지식이기 때문에 양국어 어휘의 개념 유사도 획득에는 상당한 어려움이 따랐다.

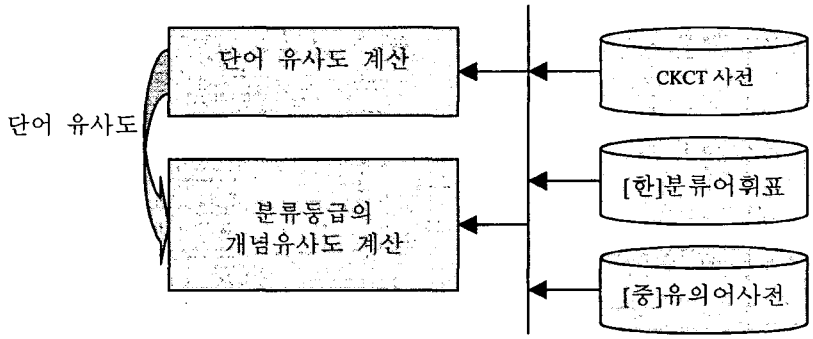
본 연구에서는 단일 언어 어휘 지식을 함유한 중국어 유의어 사전과 중국어 정보를 가진 한국어 분류어휘표 (Ver0.3)를 이용하여 양국어 개념 유사도를 반영하는 어휘분류망을 구축하고자 한다. 양국어 단어 유사도 계산에 기반하여 이미 분류되어진 중국어 어휘 분류등급과 한국어 어휘 분류등급 간의 개념유사도를 계산하고 이를 어휘분류망에 반영한다. 여기에서 말하는 분류등급은 어휘 분류체계의 최하위 분류등급과 최하위 분류등급들의 유형 집합인 상위 분류등급을 모두 가리키는 것인바 어휘분류망은 실용성을 고려하여 최하위 분류등급 간 유사 관계와 상위 분류등급 간 유사관계를 반영한다.

본 연구에서 사용하게될 한국어 분류어휘표는 일본어 분류어휘표의 한국어 버전으로 이에 대한

분석 평가가 아직 이루어 지지 않고 있다. 본 연구에서는 어휘분류망의 구축을 통하여 한국어 어휘분류표의 어휘 적용률, 어휘형평성 및 분류적합성에 대하여 정량, 정성 분석을 진행하고자 한다. 이 부분의 연구는 향후의 한국어 어휘분류표의 어휘량 및 분류체계에 대한 보완, 그리고 나아가 한국어 자체 어휘분류체계의 구축 방법론에 대한 연구에 어느 정도 도움될 것으로 기대한다.

2. 중-한 양국어 어휘분류망의 구축

구축에서 이용한 자원은 중국어 유의어사전 및 한국어 분류어휘표 내 중국어 정보 등이 있다. 본 연구에서는 양국어 단어 유사도에 근거하여 양국어 어휘분류망을 자동으로 구축하였다. 다음은 시스템 구성도이다.



[그림 1] 중한 어휘분류망 자동구축을 위한 시스템 구성도

- 양국어 단어유사도계산

다이스 계수 (Dice coefficient) [1]공식 및 일부 휴리스틱을 이용하여 양국어 단어의 형태적 유사도를 계산한다. 본 시스템에서는 중국어 유의어사전에서의 중국어 단어와 분류어휘표에서의 중국어 단어만 이용하였다. 분류어휘표의 일부 어휘에는 해당하는 중국어 단어가 없지만 얻고자 하는 최종 목적이 어휘 분류등급 간 유사도기 때문에 결과에 큰 영향을 미치지 않는다.

$$Sim(C_t, C_g) = \frac{2 \times |C_t \cap C_g|}{|C_t| + |C_g|} \quad (1)$$

식(1)에서

C_t: 중국어 유의어사전에서의 중국어 어휘

$|C_t|$: C_t 의 문자수

C_g: 한국어 분류어휘표에서의 중국어 어휘

$|C_g|$: C_g 의 문자수

$Sim(C_t, C_g)$: 두 단어 C_t와 C_g의 유사도

예: "[중]寬容 ⇔ [한]너그럽다/寬大":

$Sim(C_t, C_g) = Sim(\text{"寬容"}, \text{"寬大"}) = 0.5$

- 양국어 분류등급 간 유사도 계산

두 분류등급 내에 유사한 단어 (단어유사도가 임계치 이상인 단어) 가 많을수록 두 분류등급의 유사도도 높다는 가정하에 중국어 유의어사전과 한국어 분류어휘표의 분류등급 C와 K간 유사도 계산 과정을 다음과 같이 설계 한다:

1) 중국어 분류등급에서의 단어 C_c와 한국어 분류등급 K간의 거리 $Sim(C_c, K)$를 구한다.

단어 C_c와 C_k ($\forall C_k \in K$)의 유사도에서의 가장 큰 유사도 $MaxSim(C_c, C_k)$, ($\forall C_k \in K$)

를 취한다.

$Sim(C_c, K) = MaxSim(C_c, C_k), (\forall C_k \in K)$

2) 중국어 분류등급 C와 한국어 분류등급 K의 일차 유사도 $CountSim(C,K)$를 구한다. $CountSim(C,K)$는 $Sim(C_c, K) > t1$ ($t1$ = $threshold$, $\forall C_c \in C$) 인 레코드의 개수다.

$CountSim(C,K) = \Sigma |Sim(C_c, K) > t1|$,

$(\forall C_c \in C, t1 = threshold)$

3) 중국어 중간분류등급 Big$C$와 한국어 중간분류등급 Big$K$의 일차유사도를 구한다.

$BigCountSim(BigC,BigK) = \Sigma | Sim(C, K) > t2 |$,

$(\forall C \in BigC, \forall K \in BigK, \quad t2 = threshold)$

다음의 예는 중국어 중간분류등급 Ai과 한국어 중간분류등급 11,12 등과의 일차유사도 값이다:

$BigCountSim(Ai, 11) = 353$

$BigCountSim(Ai, 12) = 327$

$BigCountSim(Ai, 13) = 151$

$BigCountSim(Ai, 31) = 146$

$BigCountSim(Ai, 14) = 70$

...

4) 중국어 중간분류등급 Big$C$와 한국어 중간분류등급 Big$K$의 최종유사도 $BigSim(BigC, BigK)$를 구한다.

패턴 연쇄 (pattern chain)방법을 이용하여 단계 3)에서 얻은 일차유사도를 군집화 (Clustering)하며 유사도가 가장 큰 군집 1의 유사도는 1, 군집 i의 유사도는 "군집 i 일차유사도치/군집 1 일차유사도치"로 항상 1 보다 작다. 0.1 보다 작은 군집 유사도는 0 으로 처리한다.

아래의 예는 단계 3)에서 얻은 중국어 중간분류등급 Ai 과 한국어 중간분류등급 11, 12 등의 일차유사도에 대한 군집화를 통하여 얻은 최종 유사도 값이다:

$BigSim(Ai, 11) = 1$

$BigSim(Ai, 12) = 1$

$BigSim(Ai, 13) = 0.44$

$BigSim(Ai, 31) = 0.44$

$BigSim(Ai, 14) = 0.21$

...

5) 단계 4)에서 얻은 중간분류등급 간의 최종유사도를 단계 2)에서 구한 분류등급간 일차유사도 $CountSim(C,K)$에 반영하여 최종 유사도 $Sim(C,K)$를 구하되 분류등급 C와 K가 속하는 중간분류등급 간 유사도가 0 이면 $CountSim(C,K)$도 0 으로 되어야 하고 그렇지 않으면 원 유사도를 유지한다. 이는 정확도의 향상을 위한 것이다. 단계 4)에서 중간분류등급 간 관계에 대하여 군집화를 진행했듯이 분류등급 간 관계에서도 군집화 작업을 진행 한다. 식으로 표시하면:

$\forall C \in BigC, \forall K \in BigK,$

$If\ BigSim(BigC,BigK) > 0\ then$

$Sim(C,K) = CountSim(C,K)$

$Else$

$Sim(C,K) = 0$

3. 양국어 어휘분류망 구축을 통한 한국어 분류어휘표 평가 및 보완

분류어휘표에 등록된 한국어 어휘에는 한 개 어절로 구성된 단어도 있고 두개 이상의 어절로 구성된 구도 있다. 연구의 편의를 위하여 한 개의 어절로 구성된 어휘를 단어로, 두개 이상의 어절로 구성된 어휘를 구로 분류한다.

단어에는 순수한국어 (우리말, 예: "죽다"), 상용한자어 (대체할 우리말이 없거나 한자어를 더 많이 사용, 예: "죽순"), 기타 한자어 (대체할 우리말이 있고 한자어보다 많이 사용, 예: "죽간 ⇔ 대나무 장대", "죽간 ⇔ 글쓰기 위한 대쪽") 등이 있다.

본 연구에서는 분류어휘표와 우리말큰사전, 중국어 유의어 사전 및 금성출판사 사전 (이하 "금성사전"이라 칭함) 을 비교함으로 한국어 분류어휘표에 대한 보다 객관적인 평가분석을 도모하고자 한다.

3.1 한국어 분류어휘표 분석평가

[표1]은 등록 항목수, 어휘수, 등록항목중 단어수, 중복되지 않는 단어수 등에 대한 비교표이다. [표1]의 우리말큰사전 등록어는 분류어휘표와 비교하기 위하여 우리말큰사전에서 명사, 대명사, 의존명사, 형용사, 관형사, 타동사, 자동사, 부사 등 품사의 등록어를 추출한 것이다.

기본등록어는 우리말큰사전에서 위의 품사에 한하여 추출된 어휘중 금성사전에도 등록되어 있는 어휘이다. 본 연구에서는 기본등록어를 한국어 분류어휘표가 마땅히 수록하여야 할 기본 어휘로 본다.

		등록 항목수	어휘수	단어 항목수	단어수
분류어휘표	한국어	33,980	27,206	33,662	26,910
	중국어	60,196	38,703	51,200	32,063
우리말큰사전	한국어	395,244	322,801	394,216	322,003
	한자어	248,739	219,956	243,259	214,920
기본등록어	한국어	--	--	--	114,388
[중]유의어사전	중국어	61,150	50,335	61,150	50,335
분류어휘표vs 우리말큰사전	한국어		25,528		25,528
	중국어		39,650		17,010
분류어휘표 vs [중]유의어사전	중국어	--	46,278	--	13,985
우리말큰사전 vs [중]유의어사전	중국어	--	36,087	--	17,090

[표1] 분류어휘표 vs 우리말큰사전 vs [중]유의어사전[1, 2]

[표 1]을 통하여 등록어휘중 구 아닌 단어의 비례, 등록어휘중 사전등록어의 비례, 등록어휘가 기본등록어를 반영한 정도 등 측면으로 부터 분류어휘표가 수록한 한국어 등록어의 상황을 살펴보고자 한다.

- 한국어 등록어 분석

(1) 등록어휘중 단어의 비중

27,206 개의 등록 어휘중 1 개의 어절로 구성된 어휘를 단어로 보고 2 개 이상의 어절로 구성된 어휘를 구로 간주한다. 이 중 단어는 26,910 으로서 전체 등록어휘의 98.91%를 차지하는 바 분류어휘표에서의 한국어 부분이 일본어에 대한 번역으로 이루어 진 것을 감안하면 매우 높은 비례라고 할 수 있다.

---

1 등록 항목수: 사전에 등록된 모든 Item의 수
어휘수: 사전에 등록된 모든 중복되지 않는 항목의 수
단어 항목수: 사전에 등록된 모든 단어의 수 (등록 항목 - 등록 구)
단어수: 사전에 등록된 모든 중복되지 않는 단어의 수 (어휘 수 - 구의 수)
분류어휘표vs우리말큰사전: 두 사전에서 일대일 매칭이 되는 한국어거나 중국어 어휘 및 단어

2 우리말큰사전 등록어는 명사, 대명사, 의존명사, 형용사, 관형사, 타동사, 자동사, 부사에 한하여 추출하였다.

등록 어휘에서 구가 적고 단어가 많을수록 등록어의 가용도(可用度)는 높아진다.

(2) 사전등록어의 비중

등록된 한국어 단어에서 우리말큰사전에 등록되어 있는 단어를 사전등록어라고 지칭하였다. 분류어휘표의 총 26,910 개 단어중 사전등록어 (우리말큰사전에 등록되어 있는 단어)는 25,528 개로서 전체 등록 단어의 94.86%를 차지한다.

(3) 기본등록어에 대한 반영률 (적용률) : 등록어 vs 기본등록어

반영률은 기본등록어 중 분류어휘표에 실제로 등록된 사전등록어의 비중을 말하는 것으로 반영률을 통하여 분류어휘표가 한국어 중심어 (의미있는 어휘)를 어느 정도 반영하고 있는가를 살펴 볼 수 있다.

본 연구에서는 우리말큰사전의 등록어 중에서 명사, 동사 (타동사, 자동사), 형용사, 관형사, 부사 등 품사를 가진 사전등록어 322,803 중 금성사전에도 등록되어 있는 단어를 기본등록어로 규정하고 이런 기준에 따라 114,388 개의 기본등록어를 추출하였다. 여기에서 금성사전을 참조한 것은 우리말큰사전에는 非상용 역사적 어휘를 대량 포함하고 있는 반면 금성사전은 일반적으로 현대용어와 일상용어에 치중하고 있기 때문에 이를 참조한 등록어는 보다 실용성이 높기 때문이다.

[표 1]에는 표시되어 있지 않지만 분류어휘표의 등록어휘 중 기본등록어에 속하는 어휘는 21,786 이기에 (참고: 사전등록어는 25,528 개) 현재 분류어휘표가 한국어 기본등록어에 대한 반영률은 19.05%이다.

분류어휘표 내 중국어 등록어 분석

38,703 개의 등록 어휘 중 32,063 개는 단어로 등록어휘에서의 단어 비례는 82.84%로 한국어 단어 비례보다 낮다.

중국어 유의어사전 등록어를 중국어 기본등록어로 간주할 경우 분류어휘표에는 13,985 개의 기본등록어를 포함하고 있으며 이는 분류어휘표 내 중국어 등록어 (32,063 개)의 43.72%를 차지하며 기본등록어에 대한 반영률은 27.78% (13,985/50335)로서 한국어보다 (19.05%) 조금 높은 편이다.

분류어휘표에서는 32,063 개의 단어가 51,200 번 출현함으로써 평균 출현빈도가 1.60 로 나타난다. 분류어휘표 등록어휘 중 13,985 개의 기본등록어는 총 43,278 번 출현하여 그 출현빈도는 3.09 번까지 올라가는 것을 볼 수 있다. 중국어 유의어 사전에서 단어의 출현 빈도가 1.21 (61,150/50,335)에 그치는 것을 볼 때 분류어휘표에서의 중국어는 한정된 단어에 대하여 집중적으로 사용하는 양상을 나타내고 있다. 이러한 문제의 원인이 일본어의 특성때문인지 아니면 번역과정에서 불가피하게 상용어휘를 집중적으로 사용한 때문인지는 일본어 등록어에 대한 연구를 추가로 더 진행하여야 알 수 있을 것이다.

3.2 어휘분류망을 이용한 한국어 분류어휘표 보완

한국어 분류어휘표에 대한 보완에는 어휘 자원의 보충 및 분류체계의 보완 두 가지 측면이 있는데 본 연구에서는 어휘 보완에 대해서만 살펴보기로 한다.

3.1 절에서의 비교 분석에서도 볼 수 있듯이 분류어휘표가 한국어 기본등록어에 대한 반영률은 19.05%, 중국어 등록어에 대한 반영률은 27.78%로 분류어휘표에 대한 어휘 보완이 필요하다. 이를 위해 우선 기본등록어의 추출에 대하여 살펴보고 다음 각 분류 클래스로의 어휘 보완 및 어휘의 등록 형태에 대하여 생각해 보기로 한다.

- 기본등록어 분석/추출

  기존 분류어휘표에 대한 평가와 이의 어휘 보완을 위해서는 우선 기본등록어를 추출하여야 한다. 우선 기본등록어를 한 개 어절로 구성된 단어로 제한하며 본 연구에서는 실제 의미를 가진 한국어 중심어를 명사 (명사, 대명사, 의존명사), 동사 (타동사, 자동사), 형용사 (관형사), 부사로 한정한 후 우리말큰사전에서 해당 등록단어 322,003 를 추출하였다.

  우리말큰사전의 등록어에는 상용/비상용 한자어, 현대/고대 한자어를 모두 포함하고 있는데 아래의 등록어 예처럼 현대 중국어와 현대 한국어에서 사용하는 한자어 (사자 - [獅子], [死者], [使者]), 현대 중국어 어휘이지만 현대 한국어에서는 다른 어휘를 사용하는 한자어 (사자 - [私資], [師資], [寫字]), 그리고 고대 중국어에서는 사용했지만 현대 중국어와 한국어에서는 사용하지 않는 한자어 (사자 - [私子], [師子]) 등도 많이 포함하고 있다.

> 예: "사자: 1. [獅子] 2.[死者] 3.[使者] 4. [師子] 5. [私資] 6.[師資] 7. [私子] 8.[寫字]..."
>
> 현대 중국어, 한국어에서 한자어로 사용: [獅子], [死者], [使者]
>
> 현대 중국어, 한국어에서 다른 어휘 사용: [私資], [師資], [寫字]
>
> 고대 중국어, 한국어에서 다른 어휘 사용: [私子], [師子]
>
> (고대 중국어, 한국어에서만 사용:...)

본 연구에서는 한국어 기본등록어를 순수 한국어 (우리말)와 현대중국어, 현대한국어에서 모두 사용하는 한자어로 한정시키고자 하였고 비상용 및 역사적 한자어를 기본등록어에서 제외하기 위하여 금성사전을 참조하였다. 중심어에 한하여 우리말큰사전에서 추출한 단어 중 금성사전에도 동시에 등록되어 있는 단어를 현대어휘 및 상용어휘로 간주하여 이를 최종 기본등록어로 확정하여 114,388 의 기본등록어를 추출하였다.

- ClassNet 을 이용한 분류어휘표 어휘 보완

  우선 중국어이거나 한국어 단어의 한자어 있는 한국어에 대하여 해당 단어의 중국어 대응어를 찾는다. 예를 들면, 한국어와 중국어 대응으로서 다음과 같다. 사자[2] -> [死者]

  다음 해당 중국어 단어의 중국어 유의어 사전에서의 가장 가까운 거리의 분류등급을 판단한다. 유의어 사전에 해당 등록어가 있으면 바로 결정할 수 있고, 없으면 해당 단어와 유의어사전의 분류등급 간의 유사도를 이용하여 가장 가까운 분류등급을 찾고, 해당 분류등급이 분류어휘표에서의 관련 분류등급을 어휘분류망을 통해 얻는다. 이 때

단어 유사도 계산은 다이스 계수 (Dice-Coefficient) 공식에 의해 진행되어 지며 단어와 분류등급 간 거리는 분류등급 내 일정한 유사도 관계를 갖는 단어 개수의 합으로 간단히 대체할 수 있다.

분류어휘표에 넣어야 할 어휘가 순수한 한국어 (우리말)인 경우 우선 사전 (우리말 큰사전, 금성출판사사전)에서 한자어로 된 동의어 혹은 반의어를 찾아보고 이러한 한자어를 찾을 수 없다면 한국어 단어 유사도 계산을 진행하되 다이스 계수 공식에 문자별 가중치를 도입한다, 즉 한국어에서 단어의 의미를 결정하는데 많이 쓰이는 어미, 접사에 더 많은 가중치를 준다[3].

- 분류어휘표에서의 등록어 형태

분류어휘표에서 한자어의 정확한 의미 전달을 위하여 한자어 대신 대응되는 중국어 단어를 직접 기입하거나 동시 기입하는 방식을 권장한다.

예: "죽간: 1.[竹竿]대나무 장대 2.[竹簡]글 쓰기 위한 대쪽"

"사자: [獅子]", "사자: [死者]"....

4. 실험 및 토론[4]

작업의 편리를 위하여 중국어 유의어사전에서 각기 7 개의 중간분류등급에 속하는 임의의 54 개 분류등급을 선택하여 실험하였다 (참고로 중국어 유의어사전의 1/4 부분에 대하여 단어 대 단어 관계를 추출하였는데 560 만개 이상의 레코드가 생성되었다). 위에서 언급하였던 중간 분류등급에 대한 군집화 부분은 미완성 상태이어서 본 실험에서는 반자동으로 구축하였다.

7 개의 중간분류등급에서 중간분류등급 간 유사도가 0.19 이상인 관계 46 쌍을 얻었고 여기에서 군집 1 에 속하는 중간분류등급 간 관계는(1 의 유사도를 가진, 유사도가 가장 큰 분류등급쌍)는 100% 의 정확도를 나타냈으며 그 이하의 군집 관계는 유의어사전과 분류어휘표의 분류등급 간 유사도를 일일이 단어를 체크해 가며 확인할 수 없었지만 하위분류 등급 노드에 대해 살펴본 결과 중간분류등급 간 유사도를 역시 적절하게 나타내고 있다는 결론을 내릴 수 있었다.

54 개의 중국어 하위 분류등급에는 총 7,319 개의 관계가 이루어졌으며 이에 대한 군집화 작업을 아직 진행하지 못하였기에 정확도 분석을 할 수 없다. 그러나 중간분류등급에 대한 분석에서 보여주듯이 관계의 출현빈도 (관계가 성립하는 차수)는 분류등급 간의 유사관계를 적절하게 표현하고 있었고 이 점은 하위 분류등급에서도 마찬가지라고 생각한다.

단어간 유사도 계산에서 단순히 다이스계수 공식을 이용했기에 0.5 이하의 유사성은 정확도가 낮았다. 본 실험에서는 임계치 t1 를 0.6 으로 설정하여 정확한 분류등급 간 관계를 얻었지만 만약 단어간 유사도에 의미결정자와 단어 내부구조 정보를 도입하면 단어간 유사도 계산에서도 정확도의 향상을 기대할 수 있을 것으로 생각한다.

향후 유의어사전과 분류어휘표의 분류체계에 대한 비교 (비교방법론)에 대한 연구가 추가로 더해져야 할 것이다.

분류어휘표에 대한 분석에서 보여주듯이 단어 대 단어 번역을 통한 한국어 분류어휘표 구축은 단어의 질 (등록어 중 사전등록어 비중)이 좋았다.

---

[3] 이런 어미, 접사는 이운재 등의 관련 연구 (이운재, 김선배, 1999) 에 의해 이미 어느 정도 추출된 상태이다

[4] 본 연구는 다양한 데이터에 대한 대량의 미세 조절과 어휘분류망의 구축 방법에 대한 여러 가지 테스트가 필요했기에 많은 프로그래밍을 하지 않고도 데이터간 관계를 쉽게 얻을 수 있는 마이크로소프트 Access 를 사용하였다.

그러나 이러한 방법을 통한 자체 분류체계 구축은 번역담당자에 의하여 그 결과가 좌우될 수 있다는 점이 큰 부담으로 된다. 또한 단어의 반영률 (등록어가 기본등록어에 대비한 비율)이 낮은 것도 (한국어 19.05%, 중국어 27.78%) 하나의 문제로 된다.

하나의 새로운 언어에서의 분류체계 확립은 컴퓨터 보조 하에서 반자동으로 진행되어야 한다고 생각한다. 향후의 연구에서는 사람의 번역을 통하여 얻은 분류체계에서 어휘의 자동/반자동 보완에 대한 연구를 계속 진행하여야 한다.

5. 결론

본 연구에서는 기존의 양국어 단일 언어 어휘 분류체계를 이용하여 양국어 어휘 분류등급 간의 개념유사도에 의한 양국어 분류체계 간의 연관 관계를 구축하는데 대한 실험을 진행하였다. 중국어 유의어사전과 한국어 분류어휘표를 이용하여 양국어 어휘 분류체계계에서의 분류등급 간의 개념유사성 및 양국어간의 어휘 유사성에 의하여 어휘분류망 을 자동 구축하였다.

한국어 분류어휘표의 어휘 구성에 대한 定性, 定量분석을 진행하였고 이의 어휘적 보완 방법에 대한 토론을 시도하였다. 본 연구가 한국어 자체 어휘 분류체계의 구축 방법론의 연구에 도움되기를 기대한다.

6. 참고문헌

1. 梅家局등, *同義詞詞林*, 중국상해사서출판사발행소
2. 국립국어연구소, *分類語彙表*, 수영출판사
3. 兪士汶 등, *현대한어어법정보사전*, 청화대학출판사
4. Oi Yee Kwong, *Aligning WN with Additional Lexical Resources*, COLING-ACL'98
5. Michael McHale, *A Comparison of WordNet and Roget's Taxonomy for Measuring Semantic Similarity*, COLING-ACL'98
6. 이운재, 김선배 (1999), "*어휘 제약 특성을 이용한 모호성 해소*", 비공식 세미나 발표자료