

자연어처리를 위한 낱말의미의 중의성 해소에 대한 고찰

김 현 수(부산대학교)

I. 들어가는 말

만약 하나의 낱말이 단 하나의 의미로만 쓰인다면 의사소통이 지금 보다도 훨씬 더 명료할 것이라는 사실은 분명하다. 하지만 하나의 낱말이 여러 의미로 사용되어 간혹 의사소통의 장애를 일으키기는 하지만, 일반적인 세상지식을 가진 성인들에게 있어서는 그다지 큰 문제로 대두되지는 않는다. 언어적 발화 상황에서의 문맥 등이 고려되어 의사소통이 성공적으로 이끌어질 수 있기 때문이다. 그러나 인간이 아닌 기계, 즉 컴퓨터를 이용한 자연어처리에 있어서는 인간에게는 별 대수롭지 않은 낱말의 중의성 문제가 큰 장애로 등장한다.

컴퓨터가 자연언어를 분석하는데 있어서는 어휘사전(Lexikon)의 탐색 횟수를 줄이는 것과 의미의 정확한 처리가 무엇보다 중요하다. 왜냐하면 처리의 신속성과 정확성은 그 시스템의 성능과 신뢰도에 결정적인 부분을 차지하기 때문이다.

낱말의 중의성에는 일반적으로 다음의 세 가지를 들 수 있다. 첫째, 품사적 중의성이다. 이것은 영어에서는 동사와 명사, 형용사와 부사 등 다양하게 나타나지만 독일어에서는 주로 형용사와 부사에서 나타난다. 둘째, 어휘적 중의성이다. 이것은 어떤 품사의 낱말이든 한 낱말이 여러 가지 의미로 사용되고 있는 경우이다. 예를 들어 Weide는 경우에 따라 버드나무를 의미하기도 하고, 목초지를 의미하기도 하며, Tau는 이슬, 밧줄 그리고 그리스어의 19번째 문자 T라는 의미로 사용된다. 그리고 세 번째로는 구조적 중의성이다. 이것은 주로 전치사구의 문제가 많이 등장한다. 예를 들면 Hans wartet auf Tilo mit Blume.에서 전치사구 mit Blume는 Hans를 수식하는지 Tilo를 수식하는 지에 따라 문장의 의미가 달라진다. 구조적 중의성에 있어서는 그밖에도 여러 중의적 현상들을 발견할 수 있지만 자세한 논의는 다음으로 미룬다.

인간이 아닌 기계가 인간의 복잡다단한 언어를 처리하는 데 있어 어휘적 중의성을 해소하는 것이 선결과제라 생각하여 본 논문에서는 이 세 가지 가운데 어휘적

중의성을 중심으로 살펴본다.

II. 어휘적 중의성 해소를 위한 전제조건

문맥에서 어휘적으로 애매한 낱말에 대한 올바른 의미를 결정하는 문제에서는 우선 문맥인지(Kontext Rekognition)가 그 가운데 하나로 알려져 왔는데, 문맥 인지는 알려져 있거나 새로이 성립된 문맥을 표현하고 있는 프레임 혹은 스크립트에 적합한 독특한 의미로의 중의성이 제거된 어떤 낱말이 있다는 것을 아는(인지하는) 것이다. 예를 들면, 잘 알려져 있는 SAM 프로그램¹⁾에서, 각 스크립트는 그 스크립트에 적합한 일련의 낱말의미들을 그것에 관련시켰다. 즉, 레스토랑 스크립트에는 Kellner와 servieren과 같은 낱말들을 위해 주어진 독특한 의미들이 있을 수 있으며, (2-1) 즉,

(2-1) Der Kellner servierte die Suppe und die Getränke.

라는 문장이 처리될 때, 테니스경기 스크립트²⁾에서는 이 낱말이 전혀 다른 의미를 가지고 있다는 사실은 여기서는 무시된다.

그리고 다음과 같은 아주 단순한 형태의 문장에서도, 레스토랑 스크립트를 이용한 접근은 쉽게 실패한다. 즉,

(2-2) Der Anwalt gang nach Hause, das Gericht zu bereiten.

1) SAM(Script Appling Mechanism)과 PAM(Plan Appling Mechanism)은 Roger Schank가 예일대학에서 1978년에 개발한 시스템이다. 이 시스템은 문장 단어분석의 한계를 극복하여 텍스트 속에서의 단어를 분석하고자 하는 시도에서 연구되었다. 여기서 Schank는 이야기(story)를 해석하기 위하여 스크립트(Script)라는 개념을 도입했는데, 이것은 인간의 행위에 대한 전형적인(stereotypisch) 형태를 사건의 순서대로 기술한 것이다.

Vgl. R. C. Schrank and C. K. Riesbeck(ed.)(1981): Inside Computer Understanding -Five Programs Plus Miniatures, Erlbaum, Hillsdale, NJ, S. 75-119.

2) 테니스 스크립트에서 servieren은 다음과 같은 의미를 가진다. 즉 Tilo servierte den Ball zu Beginn eines Ballwechsels.

이 문장에서 만약 레스토랑 스크립트만이 유효하다면, 법정이나 재판울 의미하는 Gericht는 이용할 수 없는 것이 될 것이며, 이 낱말의 법률적인 의미는 잘못 선택될 것이다. 만약 의미를 결정하는 일이 낱말 essen을 보다 더 적절한 스크립트속에 적용시킬 기회를 가질 때까지 미루어진다면, 의미형태를 선택하는데 있어서는 어떤 스크립트를 결정하는가하는 문제가 있을 수 있다. 그러나 Charniak (1984)³⁾이 지적하는 것처럼, 실제로 50개 혹은 그 이상의 스크립트가 동시에 유효한 것으로 여겨질 수 있고, 그 각각의 스크립트는 고유어휘사전을 가지고 있으며, 선택된 스크립트 가운데서는 광범위한 검색과 선택이 필요하다. 따라서 문맥 인지를 하지 않고 중의성을 해소한다는 것은 이러한 스크립트 접근이라는 주요 장점을 소멸시켜버리는 결과를 초래한다.

나아가, 심지어 하나의 스크립트에서조차도, 다의적인 낱말은 여전히 애매한 것으로 남는 경우가 있다. 변호업무 스크립트에서, 낱말 Gericht는 경우에 따라서는 법관을 그리고 다른 경우에 있어서는 법원 건물 혹은 재판울 의미할 수도 있다. 더욱이 제일 먼저 어떤 스크립트를 불러 올 것인가를 결정하는 것은 다음의 문장 (2-3), (2-4), (2-5)에 있는 spielen과 같이 여러 낱말들과 결합할 수 있는 가능성을 지닌 애매한 낱말에 대하여 특정한 의미를 결정하도록 요구한다.

(2-3) Tilo spielte mit einer Puppe. (Script = Spiel)

(2-4) Tilo spielte Gitarre. (Script = Musik-machen)

(2-5) Tilo spielte den Hamlet gut. (Script = Theater)

하지만 아래의 문장(2-6)을 보면 낱말 Gitarre는 음악을 연주하는 스크립트를 연상시키기에는 불충분하다는 것을 주목해야 한다. 즉,

(2-6) Das Baby spielte mit einer Gitarre. (Script = Spiel)

3) Vgl. Charniak, Eugene(1984): "Cognitive Science is methodologically fine". In: Kintsch, Walter, Miller, James R. and Polson, Peter G.(ed.)(1984): Methods and tactics in cognitive science, Hillsdale, NJ: Lawrence Erlbaum Associates, S. 263-274.

일반적으로, 낱말 의미는 단지 포괄적인 문맥에만 의존하는 것이 아니라, 옆에 있는 낱말들의 의미와 같은 국부적인 단서(local cue)에도 의존한다. 문장(2-7)에 서는, 즉

(2-7) Als Tilo nach Hause kam, war sie blau.

이 문장에서는 술이 취한 상태를 의미하는 blau는 전체 문맥을 고려하지 않고서도 단지 '색깔이 푸르다'는 의미는 거부될 수 있다. 그러나 흔히 중의성을 해소하기 위한 한 방법으로 중의적 낱말의 한 가지 의미와 그 주변에 있는 낱말들 사이의 "의미 연합"(semantische Assoziation)이 필요할 때도 있다. 다음 문장(2-8)을 보자.

(2-8) Der Sekretär weckte dich aus dem Schläfe auf.

위 문장에서 동사 aufwecken과 주어로 쓰인 명사 Sekretär가 의미 연합함으로써 무생물인 책상의 의미가 아니라, 생물을 나타내는 의미, 즉 비서의 의미가 선택될 수 있음을 알 수 있다.

Hayes는 "중의적 낱말의 의미와 그 낱말의 사용을 둘러싸고 있는 문맥 사이의 연합은 그 낱말의 해석이 주변 낱말의 의미를 통해서 나와야 한다는 강한 증거다"⁴⁾라고 말한다. 그러나 중의성을 제거하는 주변 낱말들 자체도 또한 중의적 낱말인 경우도 있다. 이것에 대한 예는 schwerer Kohl을 들 수 있다. 낱말 schwer는 경우에 따라 '심한' 또는 '소화하기 어려운'을 의미할 수 있고 Kohl은 '양배추'나 '불합리함'(Unsinn)의 뜻일 수도 있다. 그러나 각각의 의미 중에서 단지 하나의 의미만이 다른 낱말의 의미와 일치하기 때문에 그것들은 서로 중의성을 해소하고 있는 것이다.

또한 통사적인 측면을 살펴본다면, 격삽입구(case-slot) 표지기에 대한 지식과 격삽입구 삽입물(filler)에 대한 제한은 또한 중의성을 제거하는 정보의 중요한 원천이다. 예를 들어 문장(2-3)과 (2-4)는, 음악을 연주하는 스크립트에서 피동자

4) Hayes, Philip J.(1977a): Some association-based techniques for lexical disambiguation by machine, Technical report 25, Department of Computer Science, University of Rochester, June 1977, S. 43.

(patient)가 OBJ에 의해 표지되기를 요구하고, 반면에 놀이 스크립트에서는 mit 를 이용하여 그것을 표지한다는 것을 알고 있다면 쉽게 중의성이 해소될 수 있다. 그리고 놀이 스크립트도 또한 그 피동자를 OBJ로 표지하는데, 그러나 이것은 음악연주 스크립트와는 구별된다. 왜냐하면 전자(놀이 스크립트)는 피동자가 놀이에 이용되기를 요구하는 반면에, 후자는 음악연주를 위한 악기를 요구하고 있기 때문이다.

동사 spielen이 나타난 예문들에서 동사와 격 표시기들은, schwerer Kohl의 예에 있는 낱말들과 마찬가지로 서로의 의미를 참조함으로써 중의성이 해소될 수 있다. 그러나 각 낱말들이 개별적으로는 많은 의미를 가질지라도, 단지 하나만의 서로 적합한 (만약 문장이 구조적인 측면에서도 심한 중의성을 띠고 있지 않다면) 의미결합이 존재한다.

그러나 때로는 고차원의 추론과 비교적 고도의 작업이 필요할 경우도 있다. 다음 문장들을 비교해 보자. 즉,

(2-2) Der Anwalt gang nach Hause, das Gericht zu bereiten.

(2-10) Der Anwalt hielt vor dem Gericht an, und sah die Leute.

문장 (2-10)은 몇 가지 민감한 의미를 가지고 있다는 것에 주목할 필요가 있다. 즉, Gericht는 법정에 모인 사람들을 의미하는 낱말 Leute와 함께 법정을 의미하거나, 또는 레스토랑 안에서 음식을 먹고 있는 사람들을 나타내는 Leute와 더불어 음식을 나타낼 수도 있다. 그리고 특정한 사례에 어떤 것이 더 나은가를 결정하는 것은 선행하는 문맥에 대한 추론을 요구한다. 다른 예에서 즉,

(2-11) Tilo schlug einen Nagel mit dem Hammer ein und der Kopf flog ab.

위의 문장 (2-11)에서 낱말 Kopf는 못의 머리 혹은 Tilo의 머리가 아니라 '망치의 머리'로 가장 자연스럽게 해석되는데, 그러나 이 의미를 해결하기 위하여서는, 망치의 머리라고 생각할 수 있는 원심력에 대한 지식 그리고 Tilo의 머리 혹은 못

의 머리는 어디론가 사라져 버릴 수 있는가에 대한 원인에 관한 추론이 요구된다.

따라서 낱말의미의 중의성 해소를 위하여 필수적인 것으로 다음을 제안할 수 있다5).

- 문맥에 대한 지식
- 인접하는 낱말들 사이의 의미연합을 찾아내는 메카니즘
- 통사적 중의성해소 단서를 다루는 메카니즘
- 애매한 낱말들 사이에 있는 선택제약 조정을 다루는 메카니즘
- 마지막 단계로서의 추론

III. 인공지능에서의 어휘적 중의성해소 연구

이 장에서는 국부적인 단서를 고려하는 어휘적 중의성 해소에 대한 자연어처리 연구 가운데 세 가지 상이한 접근 방식들을 살펴본다. 즉, Yorick Wilks의 선호 의미 시스템, Philip Hayes의 의미적 판단 그리고 Branimir Boguraev의 프레임과 의미연합을 통한 방식이 그것이다.

1. 선호 의미

Wilks의 선호의미 시스템6)은 아마도 어휘적 중의성 해소에 대한 필요에 의해

5) Hirst, Graeme(1987): *Semantic interpretation and the resolution of ambiguity* (Studies in national language processing), Cambridge University Press, S. 80.

6) Wilks의 선호의미론 시스템은 1970년대 미국에서 Schank의 MARGIE 시스템과 더불어 인공지능에 있어서의 중요한 연구 가운데 하나였다. 이 시스템에는 영어 입력문장과 불어 출력문장 사이에 (트리와 망구조에 80개의 의미소(semantic primitive)를 기반으로 하는) 국제어(interlingua, 과학자용 인공 국제어)가 있었다. 그러나 여기서는 많은 의미 가운데서 적절한 표현을 선택하는데 있어서 표현에 대한 본질을 특유의 결속성 알고리즘보다도 덜 강조했다. 실제로 이 시스템에 존재하는 낱말의미와 이 시스템에 존재하고 있지 않는 낱말의미간의 의미선택문제 그리고 구조적 중의성은 이 시스템이 안고 있는 난점이었다.

Vgl. Wilks, Yorick Alexander.(1977): *Good and Bad Arguments for Semantic*

서 명쾌하게 기획된 최초의 자연언어이해 시스템일 것이다. 이 시스템의 전략은 의미가 일치해야했던 주형(鑄型, template) 형태로 표현된 선택제약에 기반을 두었다. 그러나 그 제약은 절대적인 것은 아니었으며, 선호만을 그 처리의 중점에 두었다. 선호에 일치된 낱말 의미는 언제나 우선권이 주어졌으나, 만약 그 어느 것도 이용할 수 없을 경우에는 이 시스템은 그 의미들 가운데서 취할 수 있는 것 가운데 하나를 선택했다. 또한 이 시스템은 (3-1)과 같은 문장에서 은유(Metapher)에 대하여 다루는 것도 허용한다⁷⁾.

(3-1) Mein Auto trinkt Gasolin.

동사 trinken이 생물 주어를 선호하지만, 여기서는 자동차도 주어로 취하고 있으며, 이 경우에는 동사가 아니라 술어전체가 이 주어를 선호하는 것으로 분류되었다. 예를 들면, 형용사 groß는 물질적 대상을 수식하도록 기대된다. 선호는 (3-2)에 있는 낱말 ihn에 대한 것과 같이, 심지어 대용어(Anaphora) 해결에도 이용되었는데, 이 대용어는 바위를 나타내고 있는 것은 아니다. 왜냐하면 trinken은 액체로 된 대상을 선호하기 때문이다.

(3-2) Ich kaufte einen Wein, setzte mich auf einen Stein, und trank ihn.

이러한 접근은 Riesbeck에 의한 구문분석(parsing)과 의미해석에 대한 예견적(predictive) 방법과 일치하고 있다. 즉, Riesbeck의 분석기에서는, trinken과 같은 동사는 그 문장에서 다음에 오는 것이 액체여야 한다는 명확한 예건을 미리 준비하고 있었다.

이러한 접근방식의 결점은 선호 또는 선택제약만으로는, 전체적인 문맥, 인접한 개념들의 연합이나 이러한 개념연합을 위한 추론을 필요로 하는 것들처럼 모든 어

Primitives, Communication and Cognition 10(3/4), S. 181-221.

7) Vgl. Wilks, Yorick Alexander(1982): "Does anyone really still believe this kind of thing?", In: Sparck Jones, Karen and Wilks, Yorick Alexander(ed.) (1983): Automatic natural language processing, Chichester : Ellis Horwood/John Wiley, S. 182-189.

휘적인 중의적 사례들을 다루는데 불충분하다는 사실이다. Boguraev⁸⁾는 Wilks의 접근이 다음의 ziehen의 각각 다른 의미에서와 같은 다의적인 동사들을 충분하게 다룰 수 없다는 것을 지적한다.

(3-3) Tilo zog Tanja ins andere Zimmer. (이끌다, 잡아 끌다)

(3-4) Tilo zog ein Hemd unter den Pullover. (껴입다, 받쳐 입다)

(3-5) Das Institut zog gestern in ein neues Gebäude. (이사하다, 옮기다)

(3-6) Tilo zog den Rauch tief in die Lungen. (들이 마시다)

Boguraev는 Wilks 시스템에 있는 주형이 입력문장을 의미표현으로 번역하는 데 있어서의 일반적인 기능과 낱말의 중의성을 해소하는 데 있어서의 특별한 기능 둘 다를 동시에 갖추어야 한다는 것이 문제라고 지적했다⁹⁾.

2. 의미적 판단

Boguraev의 중의성 해소 시스템은 의미론에 대한 Wilks의 접근에 기초를 두었지만, 선호가 아니라 의미적 판단을 사용함으로써 단순한 선호방법에 있는 많은 결점들을 제거하고자 시도했다. 이 시스템은 어휘적 중의성 해소와 구조적 중의성 해소 둘 다를 의미해석으로 통합했다.

이 시스템은 ATN 분석기¹⁰⁾로 구성되어 있었는데, 이 분석기에는 의미적 절차

8) Vgl. Boguraev, Branimir Konstantonov(1979): Automatic resolution of linguistic ambiguities, Doctoral Dissertation, Computer Laboratory, University of Cambridge, August 1979, 3.23-3.25.

9) Hirst, Graeme(1987): a.a.O., S. 81.

10) ATN이란 augmented transition network를 축약한 것으로써 확장전이망이라고 한다. 우선 전이망(transition network)이란 유한상태 자동기계(FSA: finite state automata)를 문법 표현에 이용한 것으로, 상태(state)라고 하는 노드(node)와 이름 붙은 아크(labeled arcs)로 분석이나 생성과정 상의 어떤 상태를 표현하면서, 이름 붙은 아크로 상태간의 전이방식을 규정하는 것이다. 분석의 경우에는 아크에 규정된 전이방

가 명사구와 절의 완성에 수행되는 행위로서 부가되어 있었다. 이러한 절차는 위의 두 가지 중의성 해소와 Wilks와 비슷한 형식으로 의미적 표현을 구축하도록 설계되었다. 각 절차들은 먼저 선택제약, 선호, 동사 격구조 그리고 낱말의 중의성을 해소하기 위한 시도 등을 적용했다. 그리고 더 나아가 어휘적 중의성 해소는 구조적 중의성 해소와 통합되어 처리되었다. 즉, 판단은 잠재적 해석에 대한 의미적 결속성(coherence)에 따라 행해졌는데, 이 결속성이 만약 받아들이기 어려운 것으로 판명되었을 경우에는 거부되었다. 만약 모든 해석이 받아들이기 어려운 것으로 판명될 경우, 이것은 분석기에 신호를 보냈고, 되돌아가서 다른 경로를 찾는다. 의미적 판단은 단지 어휘적 지식만을 토대로 해서 이루어 졌으며, 이 시스템은 일반적인 세상지식베이스는 포함하고 있지 않았다.

3. 프레임과 의미연합

CSAW 프로그램에 대한 Hayes의 연구¹¹⁾는, 낱말들 사이의 의미적 연합을 찾는 것을 강조하는, 아마도 어휘적 중의성 해소에서의 지식의 다중적 원천을 고려한 최초의 연구였을 것이다¹²⁾. 또한 이 프로그램은 격구조와 선택제약을 이용했는데,

식에 따라 입력문장을 검사해 나가고, 생성의 경우에는 역시 아크에 규정된 전이 방식에 따라 출력문장을 생성해 나간다. 그러나 전이망 문법은 자연어의 중요한 특징인 순환성(recursion)(예를 들어, VP←VP PP는 순환성을 갖는다)을 표현하기 어렵기 때문에 자연어 문법의 순환적 특성을 아주 단순하면서도 효과적으로 표현하도록 설계한 것이 순환 전이망(RTN: recursive transition network) 문법이다. 이것은 하나의 매우 복잡한 전이망 대신에 여러 개의 분리된 전이망의 집합으로 문법을 정의하고, 필요에 따라 아크에 다른 전이망의 이름을 지정하여 전이망들 사이의 호출을 허용함으로써 표현력을 크게 높여준다. 이 같은 순환전이망 문법의 기능을 확장하여 구문분석작업(parsing)이 진행되는 동안 수집가능한 문법적 정보를 수집, 보관하는 동시에 문법적 조건들을 검사할 수 있도록 기능을 보강한 것이 확장 전이망(augmented transition network)이다. 이것에 보강된 주 기능은 구문분석작업 결과로서 문법구조를 생성하는 것과 구문분석작업 과정에서 문법적 조건들을 검사할 수 있다는 것이다.

Vgl. 김영택 등(공저)(1994): 자연언어처리, 서울, 교학사, S. 112-139.

Shapiro, Stuart C.(ed.)(1987): Encyclopedia of artificial intelligence, John Wiley & Sons, Inc, S. 323-332.

11) Vgl. Hayes, Philip J.(1977b): "On semantic nets, frames and associations", Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, MA, August 1977, S. 99-107.

12) 그러나 Quillian의 교육용 언어이해기(Teachable Language Comprehender)(1969)

이 들은 중의성을 해소하는데 절대적이며 우선적인 것이었다. CSAW 프로그램은 우선 가장 제한적 방법 즉, 격상입구, 연합 그리고 마지막으로 선택제약을 이용하고자 했다. 이 프로그램은 우선 입력문장을 ATN을 가지고 분석했으며, 그런 다음 다중 의미를 가지고 있었던 낱말에 대해 중의성을 제거하기 시작했는데, 이 모든 것들은 하나의 해석으로 집약될 때까지 함께 고려되었다. 만약 어떤 선택이 한 부분에서 이루어지면, 이 선택에 대한 다른 가능성들이 병렬적으로 처리되었다. 즉, 다중 처리가 가능했던 방식이었다.

그리고 의미연합 추출은 Hayes가 사용했던 지식표현에 의해 촉진되었다. 즉, 의미망(semantic network)이 프레임 시스템에 첨가되었던 것이다. 일반적으로 말하면, 프레임 시스템과 의미망은 같은 종류의 구조를 표시하는 단순한 변이형들이며, Hayes의 지식표현에서는 두 표시방법 모두가 사용되었다. 이때 경우에 따라서 지식베이스를 프레임 시스템으로 생각할 수 있으며, 또 각 프레임은 의미망의 각 영역이라고 생각할 수 있을 것이다. 또한 지식베이스를 각 영역들로 나누는 망으로 생각할 수 있으며, 각 영역은 하나의 프레임이라고 생각할 수 있다.

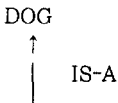
프레임들은 두 가지 위계구조로 이루어져 있었다. 하나는 ISA 관계¹³⁾를 위한

는 의미연합에 의한 중의성해소의 초보적 형태를 포함하고 있었다. 주어진 (i)에서 즉, (i) Tilo schießt den Lehrer.

만약 Tilo가 사친사이거나 또는 폭력단원으로 알려져 있었다면, 낱말 연합은 schießen이 '사친적이다'의 의미 또는 '권총을 쏴다'라는 의미인지를 결정하는데 이용될 수 있었을 것이다.

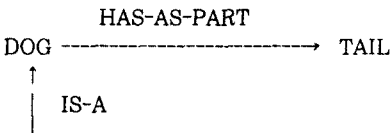
Quillian은 1961년에 처음으로 이것을 제안했다.

- 13) ISA 관계는 영어의 "is a"의 관계를 말한다. 즉, "무엇은 무엇이다"는 것을 나타내고 있는 구조이다. 예를 들면,



FIDO

이것을 문장으로 표현하면 "Fido is a dog"이다. 그리고 이와 함께 소유관계를 나타내는 HAS-AS-PART 관계도 있다. 예를 들어,



것이며, 다른 하나는 부분-전체(part-of)¹⁴⁾의 관계를 위한 것이다. 명사 의미들은 망에 있는 교점으로 표현되었으며, 동사들은 프레임으로 표현되었다.¹⁵⁾

그러므로 연합은 이러한 표현들을 살펴봄으로써 알 수 있다. 예를 들면, 하나의 교점은 이것이 속하는 어떤 프레임과 관련되어 있다. 즉, 인간의 손을 표현하고 있는 교점은 사람을 표현하고 있는 프레임 속에 존재한다. 아래의 문장 (3-7)에서 Tilo는 사람에게 대한 사례이기 때문에 중의성을 제거하는 의미연합이 나타난다.

(3-7) Tilos Hand ist groß.

또한 하나의 교점은 이것이 존재했던 ISA 위계구조 위에 있는 것 또는 아래에 있는 것에 관련되는 프레임과 연합되어 있으며, 다른 연합들은 ISA와 부분-전체 위계구조에서 작용하는 규칙에 의해 명기되었다¹⁶⁾. 이와 같은 의미연합은 교점과 프레임 사이의 지식베이스에서 연결 연쇄(Kette der Verbindung)를 찾는데 달려 있었다.

Hayes는 의미연합이 대부분 의미들 사이에 큰 차이가 있는 동음이의어를 다루는데 유용하다는 것을 알았다. 그러나 하나 이상의 의미를 위한 연합이 발견되기 쉬운 다의적 낱말¹⁷⁾을 다루는 데는 비교적 성공적이지 못하다. 비록 명사와 마찬가지로 동사도 의미연합에 의해 중의성이 제거될 수 있지만, Hayes의 처리방식에 있어서는 처리의 복잡함 때문에 동사를 포함하지 않았다¹⁸⁾.

FIDO

위의 그림은 "Fido has [a] tail"을 나타내고 있으며, 여기서 HAS-AS-PART관계는 Tail이라는 값을 나타내는 지시사 역할을 하고 있다.

Vgl. Haugeland, John(ed.)(1982): Mind Design - Philosophy, Psychology, Artificial Intelligence, The MIT Press, S. 146-148.

14) 각주(13)에서 말하고 있는 HAS-AS-PART관계를 달리 표현한 것에 불과하다.

15) Hayes, Philip J.(1976): "A process to implement some word-sense disambiguations", Technical report 6, Department of Computer Science, University of Rochester, S. 30.

16) Ebd., S. 126.

17) Ebd., S. 30.

18) Hayes, Philip J.(1977a): a.a.O., S. 46.

V. 의미 해석기의 구비요건

여기서는 앞에서 제시한 방안들을 구현하는 의미해석기에 있어서, 이것들이 갖 추어야 될 요건들을 몇 가지 제시하고자 한다. 이 요건들은 아래에서 언급되는 것처럼 다음의 다섯 가지 정도로 요약할 수 있을 것이다. 이러한 요건을 갖춤으로써 의미해석기는 자연어처리에 있어서 보다 명확한 의미를 사용자에게 제시할 수 있을 것이다.

1. 복합성(Kompositionalität).

복합성은 의미해석기에 있어서 분명하고 절실하게 요구되는 것이다. 통사적으로 완전한 각 구성요소들 만이 의미적 대상이 되어야 하며, 이러한 작은 의미적 대상이 보다 더 큰 의미적 대상의 한 부분으로 구성될 경우에도 동일성이 유지되어야 하기 때문이다.

2. 의미적 대상(semantisches Objekt).

의미해석기가 무엇을 의미적 대상으로 삼는가는 중요하다. 문장 내의 낱말들이 적합한 의미를 가지기 위하여 의미적 대상은 지식구조를 표현하는 프레임(frame)과 삽입소(slot) 그리고 진술 등이 되어야 한다.

3. 비임시성(Nicht-ad hoc).

자연어처리에 있어서 중요한 것 가운데 하나가 의미해석에 있어서의 임시성을 줄이고 시스템이 세련되어, 산만하고 불필요한 결과물을 생성해내는 것을 차단하는 것이다. 그렇게 함으로써 의미규칙이나 지식표현 형식은 의미적 대상을 올바르게 처리할 수 있고 새로운 의미적 대상을 구성할 수 있다. 통사규칙과 의미규칙은 복합성을 유지하며 의미적 대상들을 구성하기에 적합해야 하기 때문에 각 규칙은 보편적이며 서로 연관성을 지닌 것이어야 한다. 이러한 규칙은 또한 문장의 의미를 파악하기 위하여 문장의 통사적 구조가 기여하는 바도 고려될 수 있어야 하는 것이다.

4. 분석기의 피드백.

구조적 애매성제거를 위해서는 피드백이 필요하며 이러한 피드백이 해석기에도 적용될 수 있기 위해서는 해석기를 분석기와 동시에 운용하는 것이 필수적이다. 그

리고 문장의 해석된 일부분은 언제나 피드백을 적용한 방법으로 이용되기 위해서는 이러한 표현의 일부분이 적합한 의미적 대상이어야 한다.

5. 의미적 복잡성(Komplexität).

의미해석기는 몬테규와 그 밖의 학자들이 다루었던 의미의 모든 복잡성을 다룰 수 있어야 한다. 의미적 복잡성에는 내포, 불명확한 문맥, 통칭적 개념, 복잡한 수량화(Quantifikation) 등과 같은 것이 포함된다.

IV. 마치는 말

일반적으로 다양한 의미를 지닌 낱말들에 나타나는 세 가지 중의성 즉, 품사적 중의성, 어휘적 중의성 그리고 구조적 중의성 가운데 본 논문에서는 어휘적 중의성을 중심으로 중의성을 어떻게 해소할 수 있는가에 대해 살펴보았다. 그 접근 방법으로는 선호의미를 이용한 방식과 의미적 판단 그리고 프레임과 의미연합을 이용한 것이었다. 이 들 세 가지 접근 방법들은 각각 장점과 단점을 포함하고 있어 어떤 접근 방식이 가장 효율적인 것인가에 대해 논하는 것은 그리 쉬운 일은 아니다. 선호의미를 통한 접근방법에서는 동사가 선호하는 의미를 찾다보면 대응어도 손쉽게 의미해석을 할 수 있는 장점은 지니고 있으나 두 개 이상의 개념이 결합된 경우와 동사 자체가 다양한 의미로 사용될 경우에는 문맥에 적합한 의미를 취하지 못한다는 결점이 있었다. 그리고 의미적 판단을 통한 접근방법은 어휘적인 것만이 아니라 구조적 중의성까지 포함하는 의미해석 방식으로 중의성 해소를 시도했으나 단지 의미적 판단을 어휘지식만을 토대로 행했기 때문에 많은 오류를 범할 수밖에 없는 한계를 지니고 있었다. 마지막으로 프레임과 의미연합을 통한 접근방법은 동음이의어를 처리하는데 있어서는 유용한 반면 다의적 낱말의 처리에 있어서는 이 낱말들이 하나 이상의 의미연합을 할 수 있기 때문에 어려움이 있었다.

자연어처리에 있어서의 낱말의미 해석은 중요하게 다루어져야 할 영역이다. 결국 특정한 한 자연어를 다른 언어로 번역하고자 하는 기계번역 시스템에서 이러한 낱말의 중의성이 해결되지 않는다면 번역의 정확도를 기대할 수 없다. 따라서 이 부분의 연구는 지속적으로 이루어져야 할 것이다.

Literaturverzeichnis

- Boguraev, Branimir Konstantonov (1979): Automatic resolution of linguistic ambiguities, Doctoral Dissertation, Computer Laboratory, University of Cambridge, August 1979.
- Charniak, Eugene (1984): "Cognitive Science is methodologically fine", In: Kintsch, Walter, Miller, James R. and Polson, Peter G. (ed.)(1984): Methods and tactics in cognitive science, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Haugeland, John(ed.)(1982): Mind Design - Philosophy, Psychology, Artificial Intelligence, The MIT Press.
- Hayes, Philip J.(1976): "A process to implement some word-sense disambiguations", Technical report 6, Department of Computer Science, University of Rochester.
- Hayes, Philip J.(1977a): Some association-based techniques for lexical disambiguation by machine, Technical report 25, Department of Computer Science, University of Rochester, June 1977.
- Hayes, Philip J.(1977b): "On semantic nets, frames and associations", Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, MA, August 1977.
- Hirst, Graeme (1987): Semantic interpretation and the resolution of ambiguity(Studies in natural language processing), Cambridge University Press.
- Schank, Roger C. and Riesbeck, Christopher K.(ed.)(1981): Inside Computer Understanding -Five Programs Plus Miniatures, Erlbaum, Hillsdale, NJ.
- Shapiro, Stuart C.(ed.)(1987): Encyclopedia of artificial intelligence, John Wiley & Sons, Inc.
- Wilks, Yorick Alexander.(1977): Good and Bad Arguments for Semantic Primitives, Communication and Cognition 10(3/4).
- Wilks, Yorick Alexander (1982): "Does anyone really still believe this

kind of thing?". In: Sparck Jones, Karen and Wilks, Yorick Alexander(ed.)(1983): Automatic natural language processing, Chichester: Ellis Horwood/John Wiley.

김영택 등(공저)(1994): 자연언어처리, 서울, 교학사.

Zusammenfassung

Die lexikalische Disambiguierung im computergestützten sprachlichen Verfahren

Kim, Hyun-Shu (Pusan Universität)

Im computergestützten sprachlichen Verfahren (eng. Natural Language Processing, NLP) gibt es drei Ambiguitäten der Wortbedeutung: Ambiguität der Wortart, lexikalische Ambiguität und strukturelle Ambiguität. In dieser Arbeit wurde von diesen die lexikalische Ambiguität betrachtet.

Für die lexikalische Disambiguierung wird als Vorbedingungen im folgenden gefordert:

- Kenntnis des Kontexts
- ein Mechanismus, der die semantische Assoziation zwischen den benachbarten Wörtern findet.
- ein Mechanismus, der einen Anhaltspunkt zur syntaktischen Disambiguierung gibt.
- ein Mechanismus, der die Beschränkungen der Auswahl zwischen den ambigen Wörtern kontrolliert.
- ein Schlussfolgerung als letzten Schritt

Zur lexikalischen Disambiguierung wurden drei Verfahren, d.h. Vorzugsbedeutung, semantische Entscheidung und Frame und semantische Verbindung in dieser Arbeit betrachtet. Zwar haben diese Verfahren

Effektivität bei der lexikalischen Disambiguierung bewiesen, aber es war ungenügend. Deswegen sollte das Studium der semantischen Interpretation in dem computergestützten sprachlichen Verfahren weiter verfolgt werden.