

인과적 계산이론과 언어이해

공 용 현
서강대학교 철학과

Causal Computationalism and Language Understanding

Yong Hyun Kong
Department of Philosophy, Sogang University

요 약

컴퓨터의 언어이해 가능성을 반박하는 주된 근거는 형식적 기호들을 처리하는 프로그램이 의미론을 다룰 수 없다는 것이다. 그러나 인과적 계산이론에 따르면 컴퓨터 프로그램이 순전히 구문론적인 것은 아니고 컴퓨터 내부의 기호적 표상의 처리과정에서 의미론적인 지시와 해석이 일어난다고 할 수 있다.

1.

컴퓨터에 의한 자연언어처리는 인간의 심적 과정을 정보의 유한한 계산적 처리과정으로 간주하는 계산주의 모델을 토대로 하여 발전되어 왔다. 1950년대 후반의 '기계번역'을 시발로 '정보검색' '인간-컴퓨터 상면(human-computer interface)' 등 여러 갈래로 추구되어 온 이 분야의 연구는 한때 참담한 실패와 좌절을 경험하기도 했지만, 1980년대에 들어와 컴퓨터 성능의 비약적인 향상, 새로운 모델 및 이론의 등장, 현실적 요구 등 주변환경이 변함에 따라 인공지능 영역의 핵심적 자리를 차지하게 되었다..

그러나 컴퓨터에 의한 '문자인식' '음성식별' '기계번역' 등 세부 영역에서의 괄목할 만한 성공에도 불구하고 일각에서는 컴퓨터에 의한 자연언어처리가 원리적으로 불가능하다는 견해가 줄기차게 제기되어 왔다. Dreyfus와 Searle로 대

표되는 이 진영에서는 현행의 컴퓨터 이론 혹은 모델이 인간의 고유한 언어현상을 설명하기에 적합한 모델이 아니라고 주장한다. 다시말해 현재 널리 사용되고 있는 계수형 컴퓨터(digital computer)는 단지 형식적인 기호(formal symbols)만을 처리할 수 있을 뿐 자연언어를 이해하거나 생성할 수 없다는 것이다. 그들에 의하면 컴퓨터와 같은 기계는 단지 인지행위를 흉내내거나 '모의'하고 있는 것이지 인간의 내적 상태와 같은 어떤 상태를 지니거나 '복제'하고 있는 것이 아니다(Dreyfus, 1974; 1979; Searle, 1980; 1984).

'모의(simulation)'와 '복제(replication)' 혹은 '약한 의미의 인공지능(weak AI)'과 '강한 의미의 인공지능(strong AI)'의 정의와 기준에 관한 문제는 경험적인 문제라기 보다는 철학적이고 선현적인 문제이다. 이점에서 자연언어처리 분야에 종사하는 대부분의 컴퓨터과학자들은 이런 철학적 논쟁들을 그들이 당면하고 있는 과제와 전혀 무관한 것으로 여기는 경향이 있다. 그들은 이미 선제(先提)하고 있는 특정한 모델에 근거하여 실행가능한 소프트웨어를 설계하고 그것의 효과를 검증하는데 주력하고 있다. 이것은 물론 공학적 관심이며 모의의 차원에서 의 개략적 해답에 급급한 현재의 실정에서는 아마도 모델이나 기초이론의 타당성에 관한 인지과학적 혹은 인지철학적 문제를 심각하게 다룰 여유가 없을 것이다. 그러나 요구되는 모의의 수준이 아주 높아지거나 모의과정에서 예기치 않은 장벽에 부딪히게 될 때 그들은 인간 심성의 내적 구조에 주목하지 않을 수 없고, 이 관심은 결국 컴퓨터 내부의 처리과정과 인간의 심적 처리과정을 비교하는 문제로 돌려지게 될 것이다.

컴퓨터가 앵무새처럼 언어행위를 흉내내는데 불과한지, 아니면 참으로 언어를 이해한다고 말할 수 있는지를 판정하는 것은 쉽지 않다. 모의와 복제가 원리적으로 구별되고 있음에도 불구하고 인간 심성의 내밀성(內密性)은 이 문제를 끈질기게 사변적인 것으로 만들어버리고 만다. Turing(1950)은 '튜링검사'라고 불리우는 '모방게임(imitation game)'의 사례를 통해서, '컴퓨터가 생각할 수 있는가?'라는 질문이 본질적으로 대답될 수 없는 것이라고 주장한 바 있다. 어떤 점에서 모의와 복제에 관한 문제 그 자체는 무의미한 문제인지도 모른다. 그러나 이 문제의 핵심은 형이상학적인 것이 아니라 논리적이고 경험적인 문제이다. Dreyfus와 Searle의 반론도 '동일성'의 문제와 같은 철학적 난제를 제기하려는 데 목적이 있는 것이 아니라 실은 현행의 컴퓨터 모델의 토대가 되고 있는 계산주의 모델과 계산적 언어학 그리고 계산적 심리학의 이론을 논박하고 있는 것이다.

2.

Searle은 1980년과 1984년의 논문에서 컴퓨터의 언어이해가 불가능함을 '중국어 방(Chinese Room)의 논변'이라고 알려진 일종의 사고실험을 통해 증명을 시도하고 있다. 이 논변은 컴퓨터 모델과 인공지능의 기초에 대한 전형적인 논쟁꺼리를 던지고 있다는 점에서 매우 흥미로운 것이다. 이 논변의 요지는 다음과 같다.

밀폐된 방에 내(Searle)가 있고 긴 중국어로 된 글이 주어졌다고 가정하자. 나는 중국어를 쓸 줄도 말할 줄도 모른다. 더구나 나는 그 중국어 글이 중국어인지 일본어인지 아무런 글인지조차 구별할 수 없다고 하자. 이제 이 중국어 글의 첫번째 묶음(batch)과 2번째 묶음이 주어진 후, 이 두 묶음의 중국어 글들을 상관짓는 한자발의 규칙이 주어졌다고 하자. 그 규칙들은 영어로 되어 있고 영어는 내가 잘 이해한다. 이 규칙은 하나의 형식적 기호를 다른 형식적 기호와 관련시키는 것이고, 여기서 말하는 '형식적'이란 단지 글자의 모양에 의해서만 식별되는 것이다. 이제 3번째 중국어 묶음이 영어로 된 어떤 명령과 더불어 주어졌다고 하자. 이 명령은 세번째 묶음의 중국어 기호들에 상응하는 특정한 중국어 기호를 바깥으로 되돌려주라는 것이다. 이상과 같은 상황에서 방의 바깥으로부터 중국어로 어떤 이야기를 들려주면 그 답을 중국어로 말할 수 있다. 그렇다고 내가 중국어를 이해한다고 말할 수 있는가?

이 논변에서 Searle이 공격하는 대상은 분명히 Schank의 스크립트(scripts)를 통한 텍스트 분석 프로그램이나, Minsky의 프레임(frames)을 통한 의미 처리 프로그램이다. Searle에 의하면 이런 프로그램들에 의해서 수행되는 '질문-대답'은 실제의 질문 대답이 아니다. 그는 "내가 중국어 기호들을 조작하는 명령에 속달되고 프로그래머가 프로그램을 잘만 짤다면, 방 바깥에 있는 사람은 나와 중국어 속달자를 전혀 구별하지 못할 것이다. 그러나 나는 해석되지 않은 중국어 기호만을 사용하기 때문에 중국어를 이해한다고 말할 수 없다. 내가 방을 벗어나지 못하는 한 나는 계속 중국어에 대해서 무지한 채로 남아있을 것이다"라고 주장한다 (Searle, 1980: 357-358).

3.

Searle의 사고실험이 과연 컴퓨터의 언어이해 가능성을 부정하는 확고한 증거가 될 수 있을까? 이 논변에 대한 가능한 몇몇 반론들을 살펴보자. 첫째 반론은 무엇보다 이 사고실험의 일관성에 관한 것으로서, 이 속에 실현불가능한 상황이 예시되어 있음을 지적할 수 있다(Moor, 1988: 39-44).

Searle은 '중국어 방 내부에 있는 Searle'이 주어진 기호들을 식별하는 능력에 있어서 아무런 제한이 없고, 기호들을 연결짓는 규칙 또한 간단한 것처럼 생각하고 있지만 실제의 상황은 그렇지 않다. 먼저 첫번째와 두번째로 주어지는 중국어 묶음의 양이 너무 방대하다. 이 양은 중국어를 모국어로 사용하는 자가 이 세상에서 알아야 할 대부분의 지식을 망라한 엄청난 양이 될 것이다. 따라서 그것들을 조작하는 규칙도 매우 복잡한 것이 될 것이다. '방 안에 있는 Searle'이 이것을 감당할 수 있을까?

이 반박은 우리들 중 어느 누구도 초인적인 능력을 갖지 못한다면 사실상 '방 안의 Searle'이 될 수 없다는 점에서 일리가 있다. 그러나 '방 안의 Searle'이 자연인이 아니라 프로그램이 예화된 것이라는 점에서, 그리고 원리상 프로그램의 처리능력을 극도로 높일 수 있다는 점에서, 이 반박은 결정적이지 못하다.

Searle의 논지는 순수한 형식체계로서의 프로그램이 바깥 세계와 인과관계와 같은 실질적인 관계를 맺을 수 없기 때문에 언어를 이해할 수 없다는 것이다.

그렇다면 중국어 방의 변형으로서, 내부에 컴퓨터 두뇌가 장착된 로봇트가 있고 그것이 외부 사물을 지각하고 그에 상응하는 적절한 행동을 취할 수 있다고 가정하면, 그 로봇트가 언어를 이해하고 또 심적 상태를 갖는다고 말할 수 있지 않을까(Robot Reply)? '로봇트 반론'에 대한 Searle의 응답은 다음과 같다.

언어에 대한 이해란 근본적으로 그것이 지향성(intentionality)을 지닌다는 것이다. 로봇트의 지각작용과 운동능력은 지향성이나 이해에 아무런 역할을 못한다. 그 예로서 Searle은 '로봇트 속의 Searle'을 가정한다. 여기서는 중국어 문자가 중국사람의 손에 의해서 주어지는 것이 아니라, 로봇트의 눈과 귀를 통해서 들어온다. 하지만 이 경우에도 Searle은 역시 중국어를 전혀 이해하지 못한다. 왜냐하면 로봇트의 감각기관에서 처리되고 있는 것들 역시 형식적 기호에 지나지 않기 때문이다.

이러한 주장에 대해 Boden(1990)은 Searle이 잘못된 유추를 사용하고 있다고 논박한다. 그녀에 의하면 Searle이 생각하고 있는 예제는 계산주의 모델이 말하고 있는 바와는 다른 것이다. Searle이 가정한 '로봇트 속의 Searle'은 인간 내부의 두뇌 처럼 그 자체가 완전한 지향성을 지니고 있는 어떤 것이다. 그러나 대부분의 계산주의자들은 지향성을 뇌에 귀속시키지 않는다. 계산주의자들은 뇌가 무엇을 보고, 이해한다고 여기지 않는다. 지향적 상태란 '인간'이 지니는 특성이 지 '두뇌'가 지니는 특성이 아닌 것이다. 비록 표상과 같은 심적 과정이 두뇌 속에서 일어나는 것이라고 가정되더라도, 지각의 능력이나 명제적 태도는 전체로서의 인간에 귀속되는 것이다. '로봇트 반론'에 대한 Searle의 응답은 로봇트의 의사두뇌(pseudo-brain)를 가정하고 있다는 점에서 범주오류를 일으키고 있는 것이다. 즉 두뇌를 인과적 연쇄의 기반이 아니라 지능의 전적인 담지자로 보고 있는 것이다.

Boden은 이해과정을 포함한 모든 인지과정의 기본적 수준을 고수준의 프로그램이 아니라 기계-코드로 표현될 수 있는 신경과학적 수준으로 보고 있다. 이 견해는 연결론적 접근 방식으로서 최근 신경망 모델을 근간으로 한 문자인식이나 음성인식 프로그램을 상기할 때 상당히 설득력 있게 느껴진다. 그러나 신경망을 모델로 한 프로그램의 작동과정이 실제의 두뇌 속에서 처리되는 생리적 과정과 꼭 같은 것이라고 단언할 수는 없다. 연결론(connectionism)에서의 처리과정이 비록 명제적 기호들을 엄격한 형식논리에 의해 조작하는 기호론(symbolism)에서의 처리과정과 다른 것이기는 해도 이 과정이 넓은 의미의 계산과정의 범주를 벗어나지는 않는다. 문제는 계산과정 중 언제 어디에서 어떻게 지향작용 또는 이해가 일어나는 것인지를 설명하는 일이다. 기호들이 복잡하게 결합하다 보면 저절로 이런 속성들이 창발(emergence)된다고 말할 수 있을까?

이점에서 Searle은 연결론의 모델 역시 컴퓨터의 언어이해 가능성을 뒷받침하는 적절한 근거가 될 수 없다고 주장한다. 연결론자들은 중국어 방 속에 있는 '한 사람의 Searle'이 아니라 '수 많은 사람들'이 서로 연결된 연결망 속에서 언어의 이해가 일어난다고 주장하지만, 이 주장은 인간의 두뇌 내에서는 성립되지 모르지만 컴퓨터 내에서는 불가능하다.

4.

Searle의 논변이 여러 형태의 반론에도 불구하고 유지될 수 있는 이유는, 그의 논변 밑바탕에 좀처럼 반박하기 힘든 두가지의 근본적인 가정이 깔려있기 때문이다. 그 가정이란 첫째, 컴퓨터의 하드웨어는 두뇌와 달리 지향성을 일으키는 인과적 힘을 지니지 못한다는 것이고 둘째, 컴퓨터의 프로그램은 순수하게 형식적이기 때문에 의미론을 다룰 수 없다는 것이다. 이 두 가정은 사실 한 가정의 두 가지 다른 표현일 수 있다.

먼저 첫번째 가정을 검토해 보자. Searle에 의하면, 의미있는 기호들이란 지향성을 일으키는 진정한 인과적 힘을 가진 어떤 것 속에서만 구체화 되는 것이다. 인간의 두뇌는 의미있는 기호들을 지닌다. 그러나 컴퓨터는 그런 것들을 지닐 수 없다. 더 정확히 말해서 신경단백질은 가능하지만 금속이나 실리콘 조각은 불가능하다는 것이다. 이 말은 두뇌의 생화학적 속성을 강조함과 동시에 Newell(1980)의 '물리 기호 체계'를 거부하는 것이다. Searle에 의하면 컴퓨터는 사실상 기호를 조작하지도 못하고, 어떤 대상을 지시하거나 해석하지도 못한다. 그에 있어서 지향작용은 마치 광합성과 같은 생물학적 현상이다. 신경단백질은 탄소동화작용을 수행하는 엽록체처럼 심적 표상들을 산출하지만 금속이나 실리콘은 그런 것들을 산출하지 못한다. 다시 말해서 비유기체로서의 컴퓨터 하드웨어는 본질적으로 심적 기능을 지닐 수 없다는 것이다.

그러나 뇌의 지향작용이 과연 광합성에 비견될 수 있을까? 광합성은 탄소복합물에 관한 것이지만 단백질에 관한 것은 아니다. 우리는 엽록체가 광합성을 일으키고 있는 과정에 대해서 알고 있지만 거기에서 지향성을 설명할 어떤 단서를 찾아낼 수는 없다. 한편, 신경단백질이 지향성을 유지한다고 볼 수 있지만 어떻게 해서 그것이 가능하게 되는지에 대해서는 거의 모르고 있다.

지향성에 관한 Searle의 이런 생물학적 태도는 일부 신경생리학적 연결주의자들의 신념과 부합되는 점이 있고, 신경칩(neural chip) 또는 생체칩(bio chip)의 개발을 촉진시키는 한 이유가 될 수 있다. 그러나 이런 소자의 개발이 가능한지 의문이지만, 개발이 된다 해도 이것이 두뇌의 세포에 대한 복제가 아닌 이상 과연 Searle이 말하는 바와 같은 지향성을 일으킬 수 있는지도 의문이다. 한편 지향성이 생물학적인 토대에서만 설명될 수 있다는 가정 자체도 상당한 근거가 없다. Searle과 달리 Brentano는 지향성을 순수히 심리적인 것으로 보고 있고, Chisholm은 논리적 개념으로, Dennett는 이것이 도구적 기능을 하는 것으로 파악하고 있다(Boden, 1990: 92-94; Dennett, 1987). 어쩌면 지향성의 문제가 선험적 문제가 아니라 과학의 발전에 따라 해명될 수 있는 경험적 문제일지 모른다. 그러나 현재의 시점에서 지향성에 대해 물질적 토대로 말하는 것은 - 그것이 물리적 차원이든 생화학적 차원이든 아니면 더 고 수준의 차원이든 - Searle의 주장을 포함하여 한갓 추측에 불과하고, 컴퓨터의 언어이해 가능성을 다루는데 거의 도움이 되지 않는다.

5.

이제 두 번째 가정 - 컴퓨터 프로그램은 순수하게 형식적이다 - 을 검토해 보자. Searle은 '컴퓨터 프로그램'이라는 용어를 아주 피상적이고 모호하게 다루고 있다. 프로그램이란 '튜링기계' '알고리즘' '명령어 군' '테이프나 디스크에 수록된 물리적 대상' '루틴(routine) 실행 절차' 등 여러가지를 지칭할 수 있다 (Moor, 1988: 42). Searle은 이 중에서 특히 프로그램의 추상적인 면을 강조하고 있다.

컴퓨터 프로그램이란 무엇보다도 컴퓨터를 '위한' 프로그램이다. 프로그램이 하드웨어 상에서 처리될 때 기계는 반드시 어떤 결과를 산출한다. 이것은 단순한 기호논리의 체계가 아니다. 프로그램과 컴퓨터 계산의 본질적 기능은 컴퓨터 내에서 어떤 일이 일어나게끔 하는 것인 반면, 기호논리의 영역에서 일어나는 일은 그야말로 해석되지 않은 기호들의 조작에 지나지 않는다. 어떤 점에서 컴퓨터의 계산논리가 튜링 기계와 같은 수학적, 가상적 기계 속에서 일어나는 시공을 초월한 추상적인 관계들을 추구하는 것으로 보일 수 있다. 그러나 컴퓨터의 계산이론을 결코 이런 식으로 묘사할 수는 없다. 컴퓨터의 계산과정은 입력에서 출력까지 실제의 시간과 장소에서 인과적으로 산출되는 기호적 표상들의 처리과정이다. 이 점에서 컴퓨터의 계산이론은 추상적 계산이론이 아니라 인과적 계산이론(causal computationalism)이라고 할 수 있다(Moor, 1988: 47-48).

그러나 이런 기호적 표상들이 두뇌 속에서의 심적 표상들처럼 능동적 기능 즉 지시와 해석의 기능을 지닐 수 있을까? Newell은(1976; 1980) 컴퓨터의 기호체계를 물리적 기호체계로 보고, 컴퓨터 내의 어떤 표상이 대상을 지시할 뿐만 아니라 다른 표상들을 해석할 수 있다고 주장한다. 컴퓨터 프로그램의 수행 과정을 기계-코드의 수준에서 본다면 하나의 연산이 단일한 연산을 이끌어 낸다는 점에서 형식적이고 선언적이다. 그러나 고급언어, 예를들어 LISP나 PROLOG의 수준에서 본다면 프로그램된 명령어들은 단순한 형식적 패턴도 아니고 선언적 진술도 아니다. 이것들은 적절한 하드웨어의 맥락에 따라 주어진 절차적 특성이고 수행될 세부적 절차를 만들어내는 것이다. 이 점에서 프로그램은 컴퓨터 내에 띄여진 구조로서의 표상을 표현하는 매개체일 뿐 만 아니라 표상적 활동의 매개체이기도 하다.

이런 접근은 컴퓨터의 언어이해를 위한 새로운 의미론의 발판을 마련한다. 그러나 의미론은 지시적인 것이 아니라 인과적인 의미론이다(Smith, 1982). 인과적 의미론에서는 전통적으로 구분해 오던 구문론과 의미론의 차이를 부정하지는 않는다. 그러나 하위 수준의 구문론으로부터 상위 수준의 의미론을 구성할 수 있다는 점에서 그 둘을 전혀 별개의 것으로 보지는 않는다. '중국어 방'의 변형으로서 카드섹션이 진행되는 운동장을 예로 들어 보자. 카드를 펼치는 개개의 사람들은 어떤 규칙에 따라 특정한 색깔의 카드를 특정한 순간 들었다 내렸다 할 것이다. 이것은 완전히 구문론적 원칙에 따르는 것이다. 여기서 사람들은 자기 쪽에서 어떤 일이 일어나고 있는지 모를 수 있다. 그러나 운동장 건너편에서 어떤 글자가 펼쳐지고 있는지는 알 수 있다. 이 예는 운동장 내부에서(즉 '중국어 방'

안에서) 모종의 이해가 일어나고 있음을 보이고 있다. 아마도 Searle은 이런 식의 언어이해가 본질적인 언어 이해가 아니라고 주장할 것이다. 그러나 컴퓨터 내에서의 처리과정이 내적으로나 외적으로나 어떤 것을 지시하고 인과적으로 산출한다는 점에서 이 과정은 분명히 의미론적이다.

6.

인과적 계산이론은 계산주의라는 기본 모델을 토대로 하고 있다. 여기서 말하는 계산주의란 형식적 기호처리(formal symbol manipulating) 체계로서의 기호론과 병행적 분산 처리(parallel distributed processing) 체계로서의 연결론을 포괄하는 넓은 의미의 계산주의이다. 인과적 계산이론은 자연언어의 컴퓨터 처리를 위한 이론이지만 이것이 반드시 현재의 계수형 컴퓨터에만 적용되는 것은 아니다. 이것은 계산주의 모델이 그런 것처럼 물리주의(physicalism), 심리주의(mentalism), 하드웨어(hardware), 웨트웨어(wetware) 등 상반된 입장 사이에서 중립적이다. 이 이론의 요지는 언어이해와 같은 과정이 물리적 과정임을 주장하는 것이 아니라 '어떤 매체를 통하든' 그 과정이 계산적 과정임을 주장하는 것이다.

그러나 인과적 계산이론이 컴퓨터의 언어이해 가능성에 대한 선형적 반박을 해소할 수 있다 해도, 컴퓨터가 실제로 언어를 이해하고 있다고 주장하기 위해서는 각 분야에서 정당하게 제기되고 있는 많은 난점들을 극복해야만 한다. 예를 들어 '문자 식별'에서의 현상학적 문제, '기계번역'에서의 맥락의 문제, '텍스트 이해'에서의 해석학적 배경의 문제를 해결하기 위해 인과적 계산이론은 고전적 의미론의 '명제'나 '술어'보다 더 원초적인 단위를 계산해야 될지 모르고, 구문 및 의미 처리를 위한 더 정교한 규칙을 고안해야 될지 모른다. 명제 단위가 아닌 표상체계로 지식기반을 구축할 수 있는가? 어떤 검색 방식이 가장 효과적인가? 인지양상 전반을 망라하는 단원화(modularization)와 이들을 묶는 메타인지적 구조를 설계할 수 있을까? 이것들은 아마도 수많은 사람의 노력이 요구되는 복잡한 과제일 것이다.

그러나 이런 문제가 계산주의자들에게 큰 부담이 되고 있는 것은 사실이지만, 자연언어를 능숙히 처리하는 컴퓨터를 설계할 수 있다는 낙관론을 원리적으로 배척할 수 있는 근거는 아무데도 없다. 문제의 본질은 철학적인 것이 아니라 경험적인 것이며, 기계의 본성과 인간 심성의 내부 구조에 대한 집요한 탐구에 의해 해결될 수 있는 것이다. 간혹 "크레이(Cray) 컴퓨터 수십대를 연결하면 이 문제가 해결될 수 있다"는 식의 성급한 발상이 나올 수 있지만, 이는 채택하고 있는 모델이나 이론의 부적절함을 노출시키고 있을 뿐 문제해결을 위한 전향적인 태도가 아닌 것으로 여겨진다. 인간의 두뇌는 크레이 컴퓨터 만큼 빠르게 작동하지 않는다. 우리가 지금 모르고 있는 것은 두뇌의 처리 속도가 아니라 처리 방식이다.

< 참고 문헌 >

- Boden, Margaret E. (1990). 'Escaping from the Chinese Room,' Boden ed. *The Philosophy of Artificial Intelligence*. Oxford: Oxford Univ. Press, pp. 89-124.
- Cole, D., Fetzer, J. & Rankin, T. (1990). *Philosophy, Mind and Cognitive Inquiry*. Dordrecht: Kluwer Academic Pub.
- Dennett, Daniel C. (1987). *The Intentional Stance*. Cambridge: MIT Press.
- Dreyfus, H & Haugeland, J. (1974). 'The Computer as a Mistaken Model of Mind,' S. C. Brown ed. *Philosophy of Psychology*. New York: Harper & Row, pp. 247-258.
- Dreyfus Hubert L. (1979). *What Computers Can't Do*. New York: Harper & Row.
- Grishman, Ralph. (1986). *Computational Linguistics*. Cambridge: Cambridge Univ. Press.
- Moor, James H. (1988). 'The Pseudo Realization Fallacy and the Chinese Room Argument,' J. Fetzer ed. *Aspects of Artificial Intelligence*. Dordrecht: Kluwer Academic Pub.
- Minsky, Marvin. ed. (1967). *Semantic Information Processing*. MIT Press.
- Newell, A. & Simon, H. (1976). 'Computer Science as Empirical Inquiry,' J. Haugeland ed. (1982). *Mind Design*. Cambridge: MIT Press.
- Newell, A. (1980). 'Physical Symbol Systems,' *Cognitive Science* 4, pp. 135-83
- Schank, R. & Abelson, R. (1977). *Scripts Plans Goals and Understanding*. Hillsdale: Lawrence Erlbaum.
- Schank, Roger C. (1984). *The Cognitive Computer*. Reading: Addison-Wesley.
- Searle, John. (1980). 'Minds, Brains, and Programs,' D. Hotstadter & D. C. Dennett eds. *The Mind's I*. New York: Basic Books, pp. 353-373.
- Searle, John. (1984). *Minds, Brains and Science*. Cambridge: Harvard Univ.
- Smith, George W. (1991). *Computers and Human Language*. Oxford: Oxford Univ.
- Winograd, T. & Flores, F. (1986). *Understanding Computers and Cognition*. New York: Addison-Wesley.