

Frühgeschichte der Computerlinguistik

- In den späten 50er Jahren begann in den USA aus politischer Motivation die Förderung der Maschinellen Übersetzung (Russisch-Englisch) als erster computerlinguistischer Anwendungsdisziplin. Die Resultate waren katastrophal, die Förderung wurde eingestellt.
- In den späten 60er und den frühen 70er Jahren entstand die zweite Generation von computerlinguistischen Systemen: Die beiden vermutlich berühmtesten Systeme waren zwei äußerst unterschiedliche Dialogsysteme:
 - ELIZA, 1966, von Joseph Weizenbaum, und
 - SHRDLU, 1972, von T. Winograd

ELIZA: Beispieldialog [1]

(1) U: I have been very depressed lately

S: *You say you have been very depressed lately.*

(2) U: Yes, my mother beats me and my father calls me names

S: *You`re being somewhat short with me*

(3) U: What about my mother beating me?

S: *Tell me more about your family*

(4) U: My father calls me nasty names

S: *Who else in your family calls you nasty names*

(5) U: I think the dog would if he could

S: *Do you think its likely that he could*

(6) U: No, you do

S: *You don`t seen very talkative today*

ELIZA: Beispieldialog [2]

(7) U: I guess what is really
bothering me is that I don't
have any money
S: *Can you elaborate on that*

(9) U: I thought you might think
so
S: *Perhaps in your fantasy
we thought each other*

(8) U: The problem is that I
know psychiatrists are
expensive and I may not be
able to pay you
S: *Please continue, this may
be interesting*

ELIZA [3]

TEMPLATE: You _X1_

RESPONSE: You say you _X1_

TEMPLATE: _X1_ your _X2_ (*category family*) _X3_

RESPONSE: Who else in your family _X3_ ?

TEMPLATE: _X1_ if _X2_

RESPONSE: Do you think it is likely that _X2_ ?

TEMPLATE: _X1_

RESPONSE: You're being somewhat short with me.

ELIZA [4]

ELIZA ist ein Dialogsystem, das beliebig komplexe Eingaben mit beliebigem Wortschatz zu beliebigen Themen akzeptiert.

Aber: ELIZA arbeitet mit einfachen Mustervergleichs-Techniken (Pattern Matching), ohne linguistisches oder anderweitiges

Wissen: Die Datenbasis des Systems enthält Paare von

- Templates: Muster mit variablen Teilen, die mit der Benutzereingabe abgeglichen werden, und
- Systemantworten, die aus neuen Wörtern und Teilen der Benutzereingabe zusammengesetzt werden.

(Einzige „linguistische“ Aktion des Systems: Vorab werden alle Pronomen der 2. Person in Formen der 1. Person umgewandelt.)

ELIZA hat demonstriert, dass ein Computersystem ohne jede Intelligenz gute Chancen hat, den Turing-Test zu bestehen (s. Lektüre).

SHRDLU

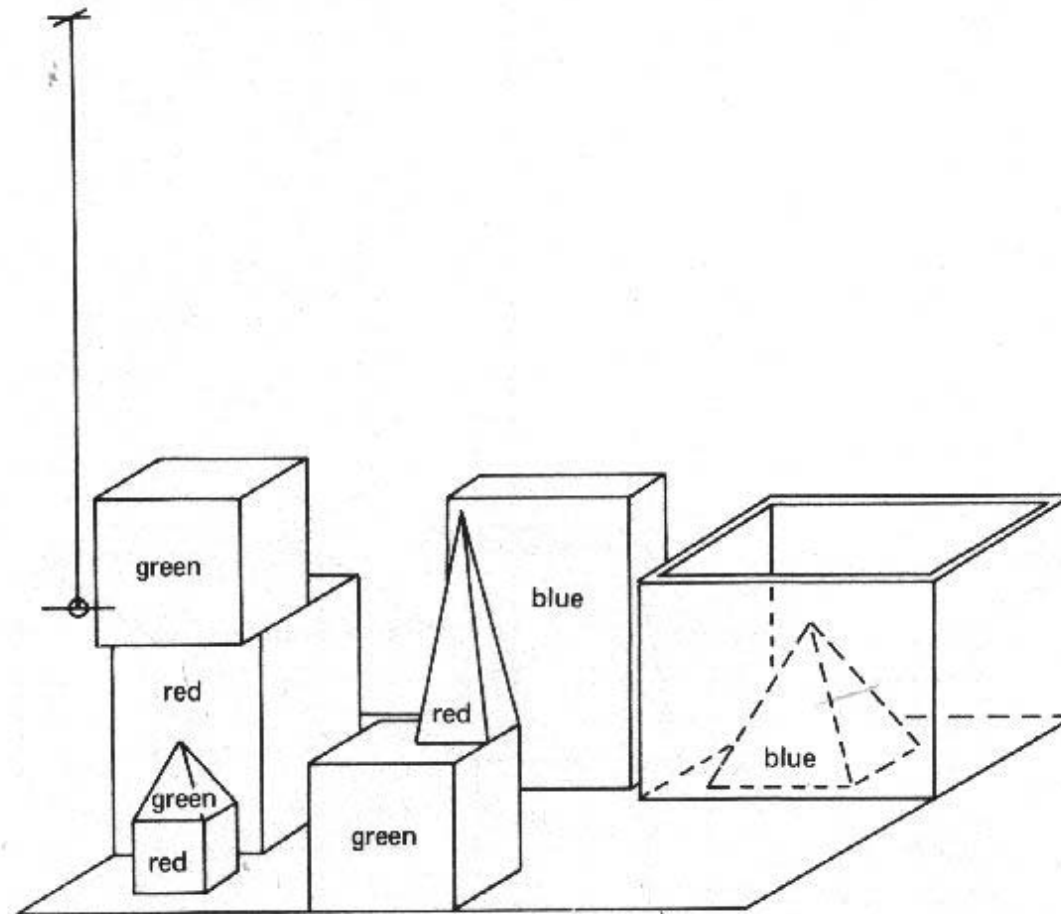
Winograds SHRDLU-System arbeitet in einer äußerst eingeschränkten Mini-Welt oder –Domäne (blocks world).

SHRDLU beantwortet Fragen, führt Anweisungen aus und lernt Begriffe (alles eingeschränkt auf die Domäne).

Wichtige Programmkomponenten von SHRDLU sind:

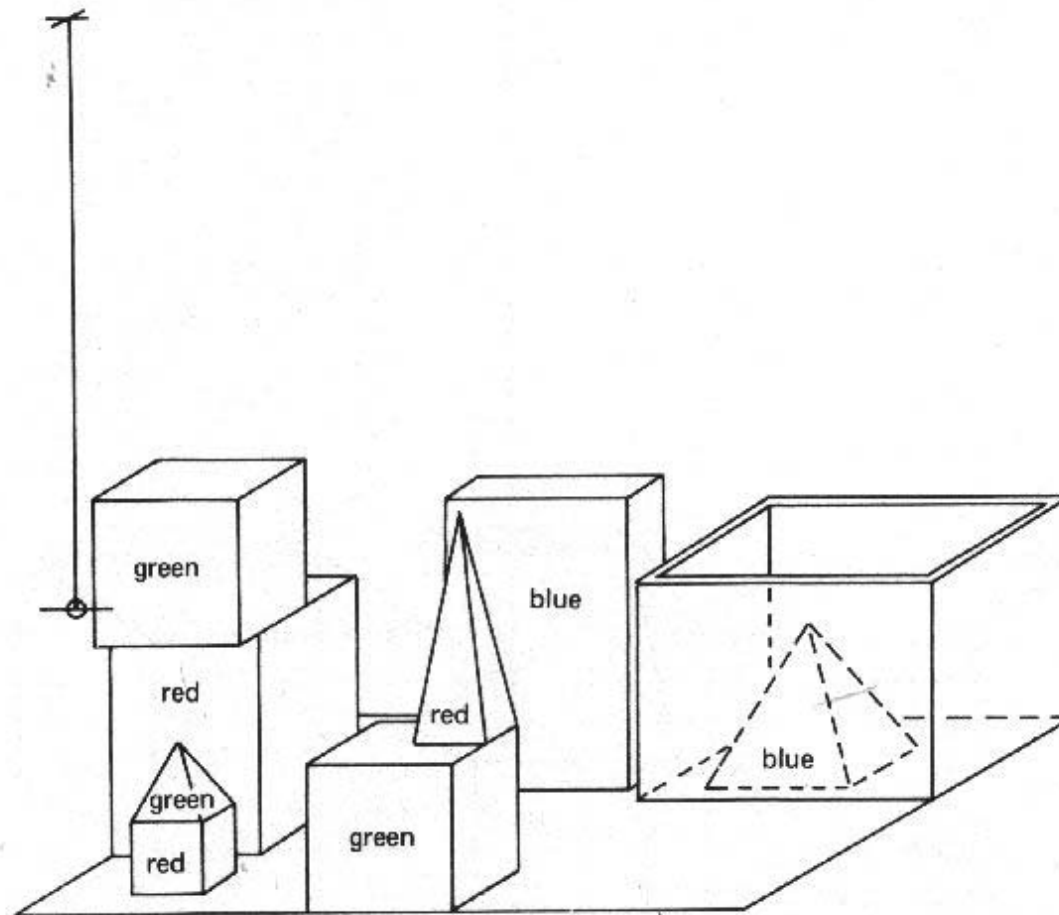
- (Linguistische) Analyse
- Generierung
- (Handlungs-)Planung
- (grafische) Visualisierung

Interessant ist die Interaktion von Analyse und Planung; die Generierungskomponente ist sehr einfach (patternbasiert); die Grafik ist computerlinguistisch nicht sehr interessant, hat aber zum durchschlagenden Erfolg des Systems beigetragen.

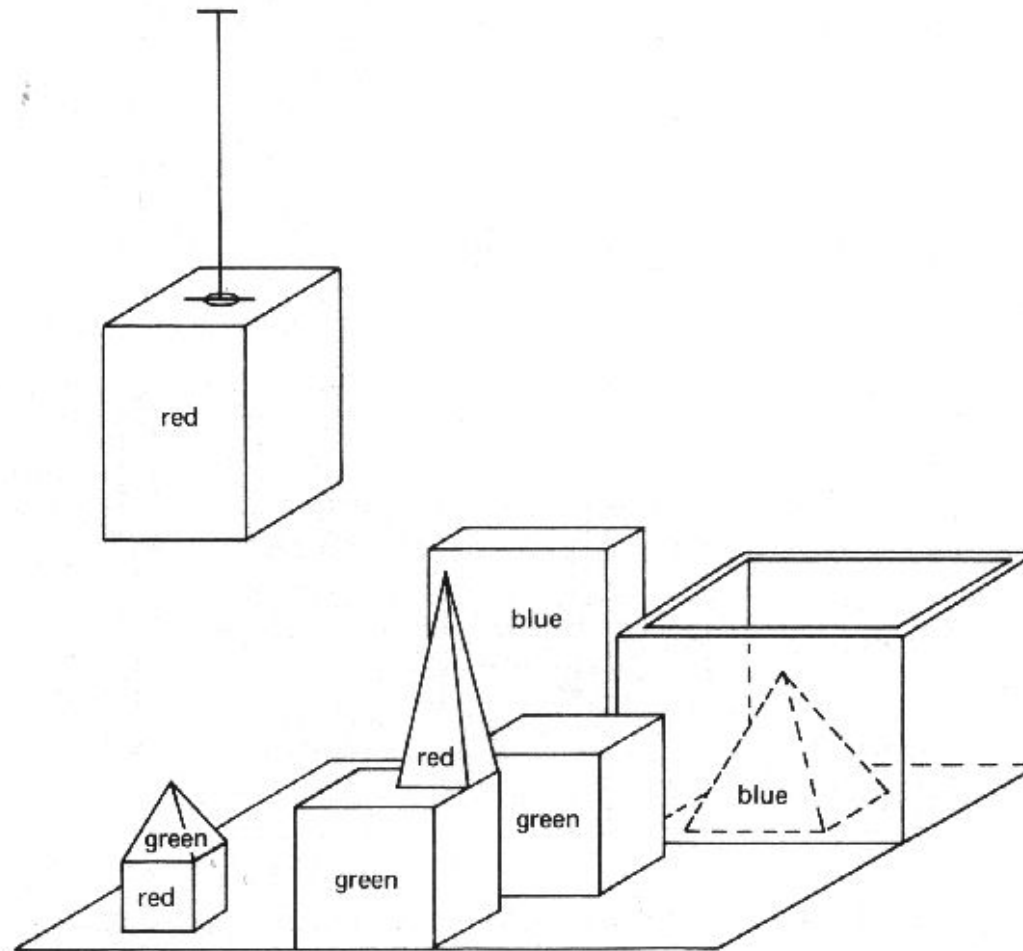


SHRDLU: Die “Blocks World”

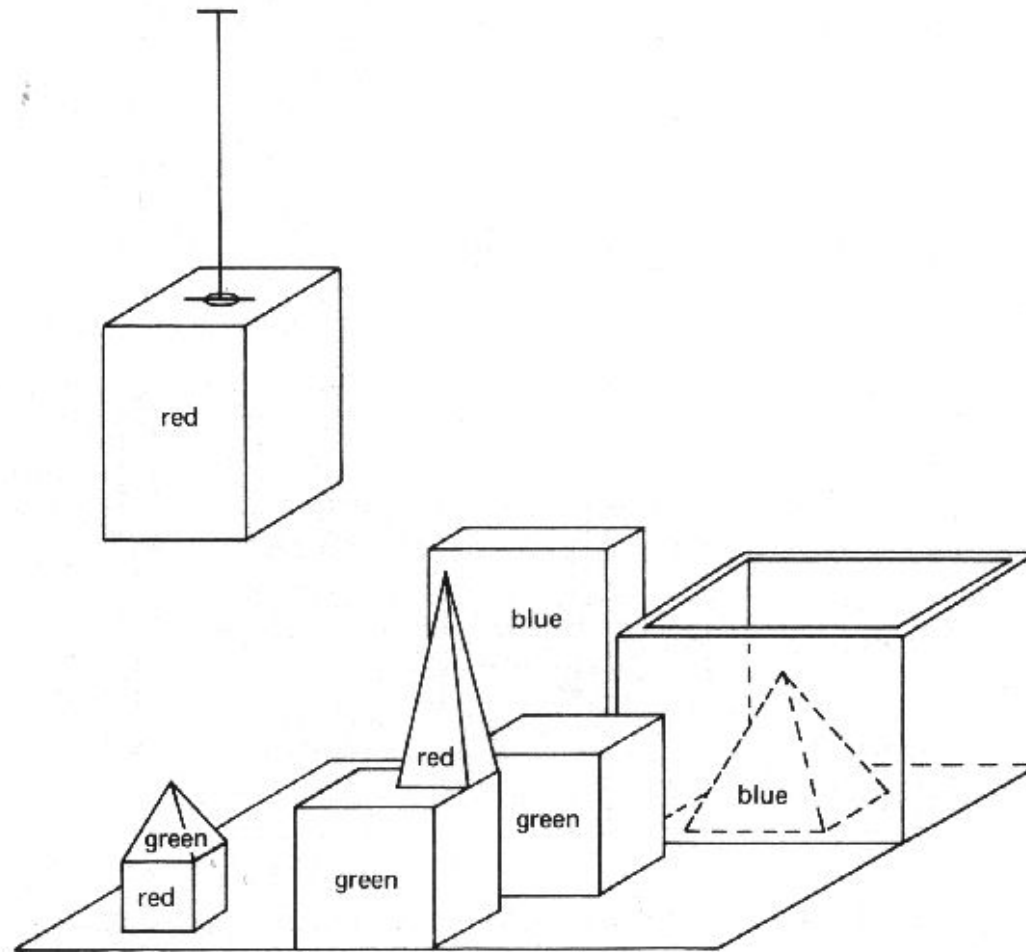
Vorlesung Semantik 2002/2003 © M. Pinkal UdS Computerlinguistik



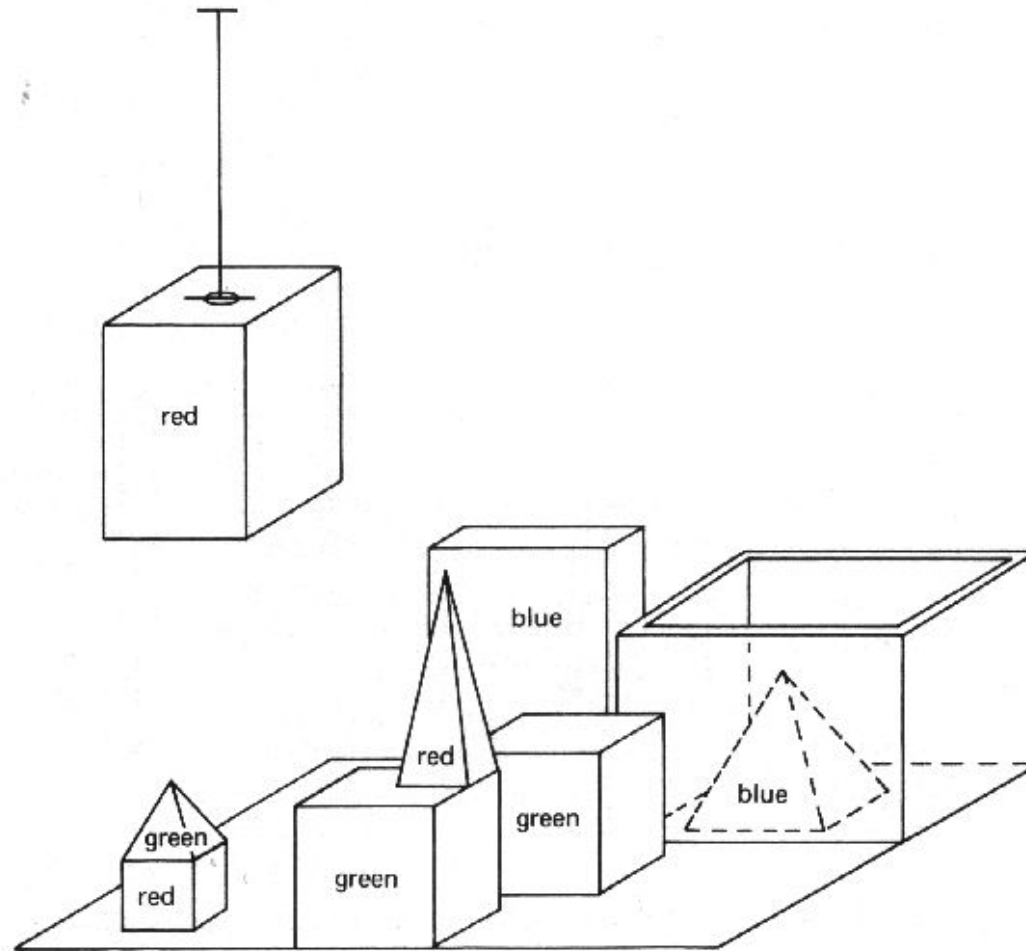
U: Pick up a big red block
S: *OK.*



(Pick up a big red block)

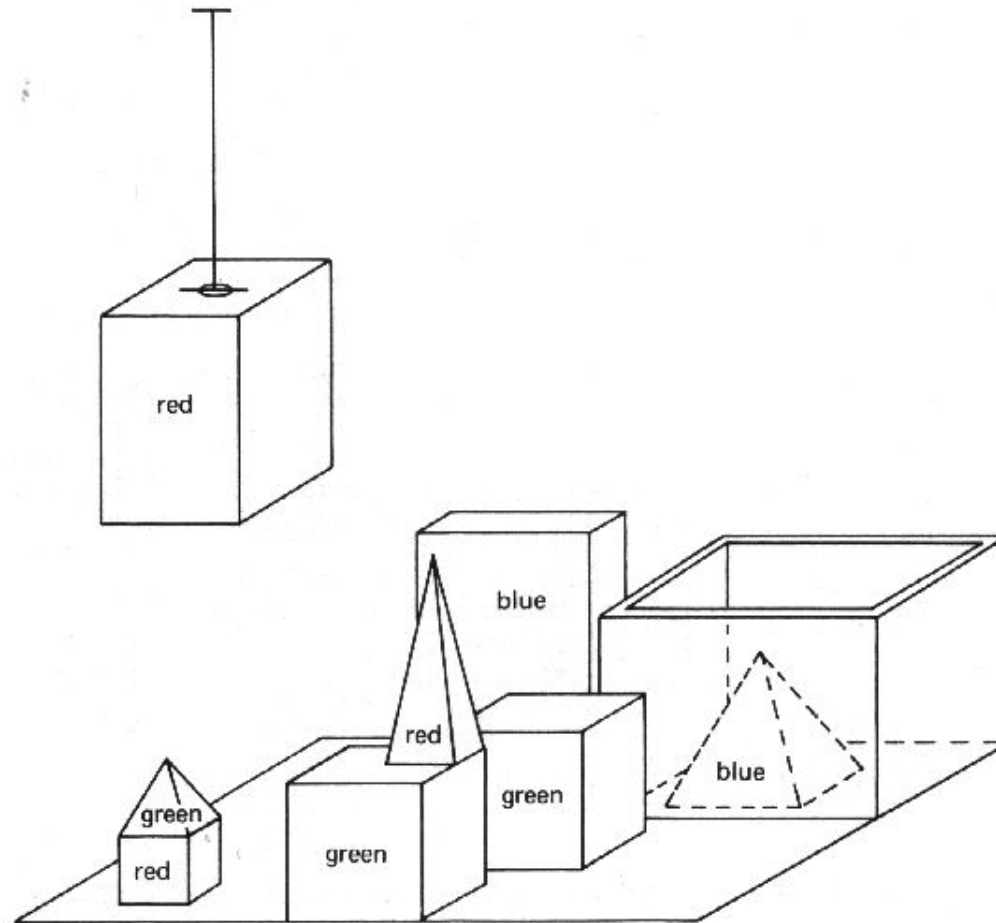


U: Grasp the pyramid.



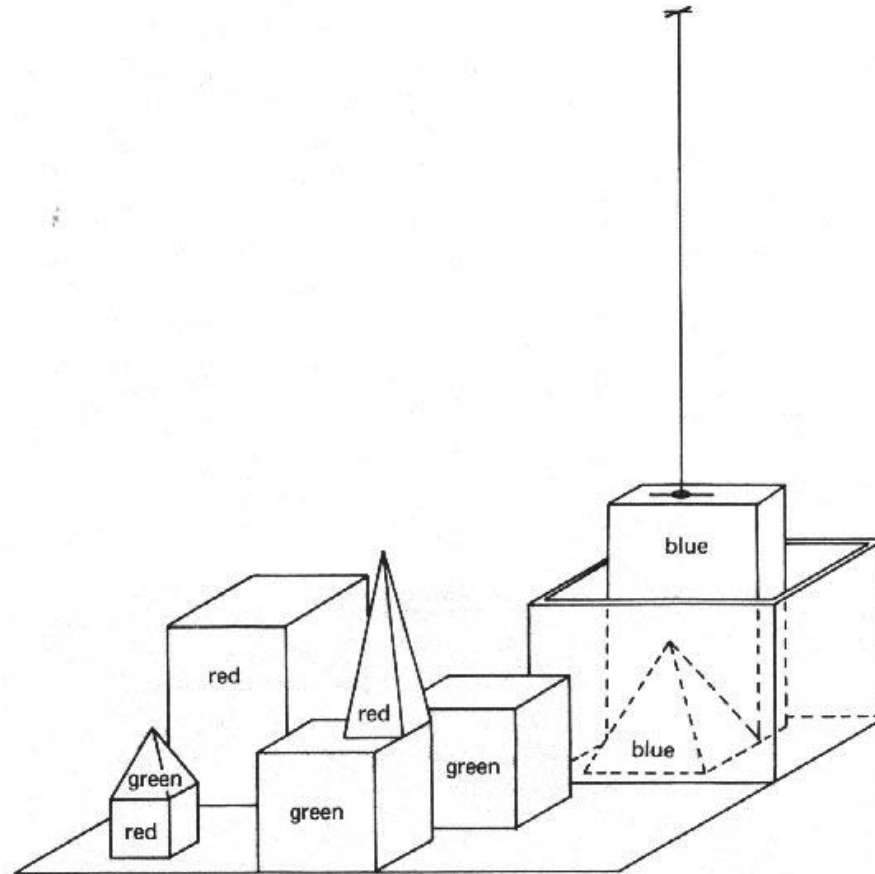
U: Grasp the pyramid.

S: I don't understand which pyramid you mean.

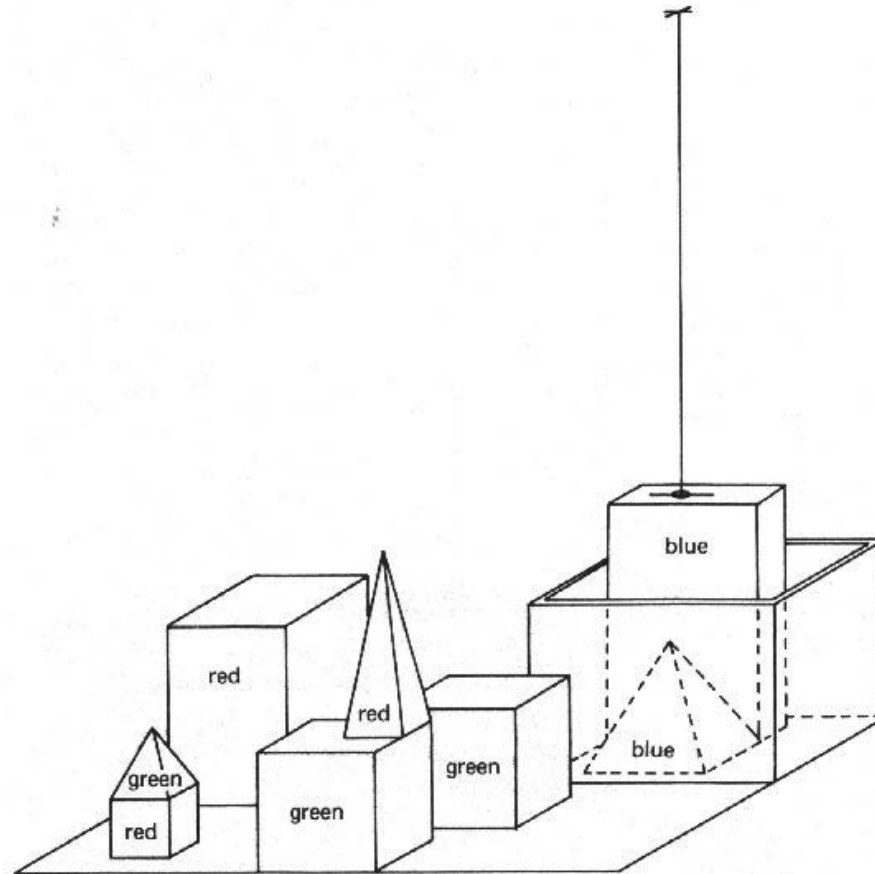


U: Find a block which is taller than the one you are holding and put it into the box.

S: *By „it“, I assume you mean the block which is taller than the one I am holding. OK*

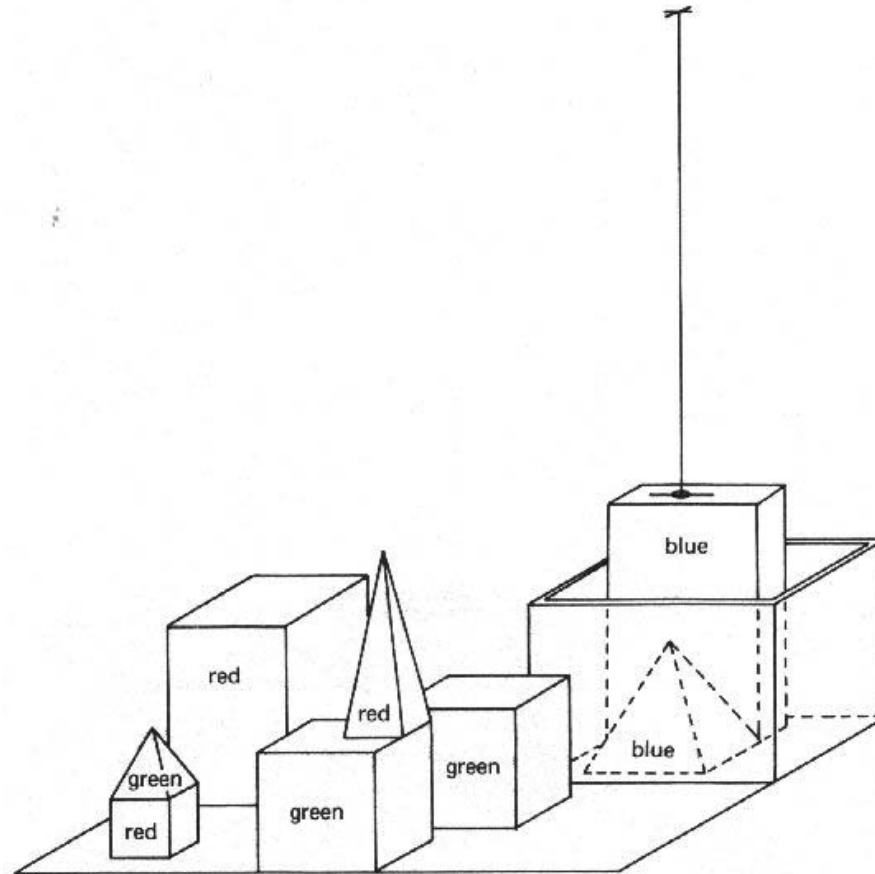


(U: Find a block which is taller than the one you are holding and put it into the box.)

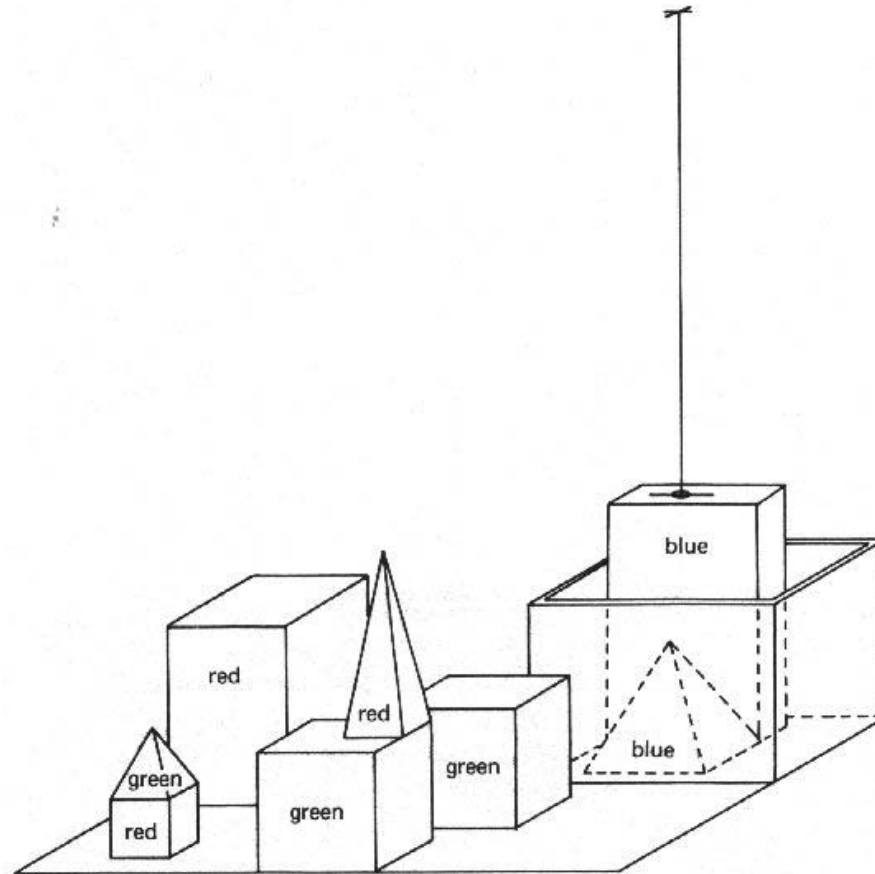


U: What does the box contain?

S: *The blue pyramid and the blue block*



U: What is the pyramid supported by?
S: *The box*



U: How many blocks are not in the box?

S: *Four of them*

SHRDLU [2]

SHRDLU ist ein wissensbasiertes Dialogsystem.

Im Unterschied z.B. zu ELIZA verwendet es bei der Dialogführung unterschiedliche Arten von Wissen, und zwar

- Sprachliches/linguistisches Wissen
- Kontextwissen
- Weltwissen

Sprachliches Wissen in SHRDLU: Beispiele

Morphologisches Wissen:

regelmäßige Verben bilden *grasp* ist regelmäßiges Verb
Präteritum auf -ed *put* ist unregelm. Verb mit Prät. *put*

Syntaktisches Wissen:

In Imperativen steht das *grasp* ist transitives Verb
Verb an erster Stelle *stop* ist intransitives Verb

Semantisches Wissen:

A+N in attributiven *red* bezeichnet rote Dinge (?)
Konstruktionen bezeichnet *pyramid* ...
Dinge, die unter A und unter *grasp* ...
N fallen

Sprachliches Wissen in SHRDLU: Beispiele

Grammatik

Lexikon

Morphologisches Wissen:

regelmäßige Verben bilden	<i>grasp</i> ist regelmäßiges Verb
Präteritum auf -ed	<i>put</i> ist unregelm. Verb mit Prät. <i>put</i>

Syntaktisches Wissen:

In Imperativen steht das	<i>grasp</i> ist transitives Verb
Verb an erster Stelle	<i>stop</i> ist intransitives Verb

Semantisches Wissen:

A+N in attributiven	<i>red</i> bezeichnet rote Dinge (?)
Konstruktionen bezeichnet	<i>pyramid</i> ...
Dinge, die unter A und unter	<i>grasp</i> ...
N fallen	

Grammatisches und lexikalisches Wissen

- Morphologische, syntaktische, semantische Regularitäten sind tendenziell in der Grammatik kodiert
- Spezielle morphologische, syntaktische, semantische Information über Einzelwörter sind im Lexikon kodiert.

Die Grenze zwischen Grammatik und Lexikon kann unterschiedlich gezogen werden:

- Vollformenlexikon / Stammlexikon + Flexionsmorphologie (Übung dazu kommt)
- Funktionswörter (Konjunktionen, Artikel): Grammatik oder Lexikon?

Außersprachliches Wissen

- Kontextwissen: Dialoggeschichte, Äußerungssituation

Welches Objekt wurden zuletzt erwähnt ? Welches Objekt kommt im Blickfeld vor?

(Put it into the box. What is the pyramid supported by?.)

- Weltwissen: Episodisches Wissen

Es gibt zwei rote Klötze. Eine Pyramide ist in der Kiste.

- Weltwissen: Regelwissen

Zwei Objekte können nicht den gleichen Platz einnehmen. Um ein Objekt auf ein anderes zu stellen, muss das zweite eine gerade Auflagefläche besitzen.

Wissen in der Sprachverarbeitung:

- sprachliches/ „linguistisches“ Wissen
 - phonetisches
 - morphologisches
 - syntaktisches
 - semantisches
 - pragmatisches Wissen
- außersprachliches Wissen
 - Kontextwissen
 - episodisches Wissen
 - Regelwissen

Wozu wird Wissen eingesetzt?

Wissen wird in der – menschlichen und maschinellen Sprachverarbeitung eingesetzt, um linguistische Struktur unterschiedlicher Arten und Ebenen aufeinander abzubilden:

- Speech → Text
- Text → Speech
- Wortkette → Bedeutungsinformation
- Bedeutungsinformation → Handlungsplan
- Bedeutungsinformation → Wortkette
- deutscher Satz → englischer Satz

Das zentrale Problem ist die Mehrdeutigkeit (Ambiguität) auf allen Ebenen: Wie kommen wir zu einer eindeutigen Abbildung?

Woher kommt das Wissen?

Alternativen:

- Manuelle Grammatik- und Lexikon-Entwicklung, Erstellung von extralinguistischen Wissensbasen
- Automatische Extraktion von Wissen aus Sprachdaten (Korpora) mit statistischen Techniken und Lernverfahren (seit den neunziger Jahren zunehmend wichtig)