

Einführung in die Computerlinguistik

2: Die Rolle von Wissen in der Sprachverarbeitung

WS 2005/2006

Manfred Pinkal

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 extrem schlecht, die Förderung wurde eingestellt.
- In den späten 60er und den frühen 70er Jahren entstand die zweite Generation von computerlinguistischen Systemen:
Dialogsysteme
Die beiden vermutlich berühmtesten Systeme:
 - ELIZA, 1966, von Joseph Weizenbaum, und
 - SHRDLU, 1970, 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.
- ELIZA arbeitet mit einfachen Mustervergleichs-Techniken (**Pattern Matching**), ohne Einsatz von Wissen:
 - **Templates**: Muster mit variablen Teilen, die mit der Benutzereingabe abgeglichen werden, und
 - Template-basierten System-Äußerungen (Prompts)
- ELIZA hat in gewisser Hinsicht den **Turing-Test** absolviert (s. Lektüre), dies aber unter besonderen Rahmenbedingungen.

SHRDLU

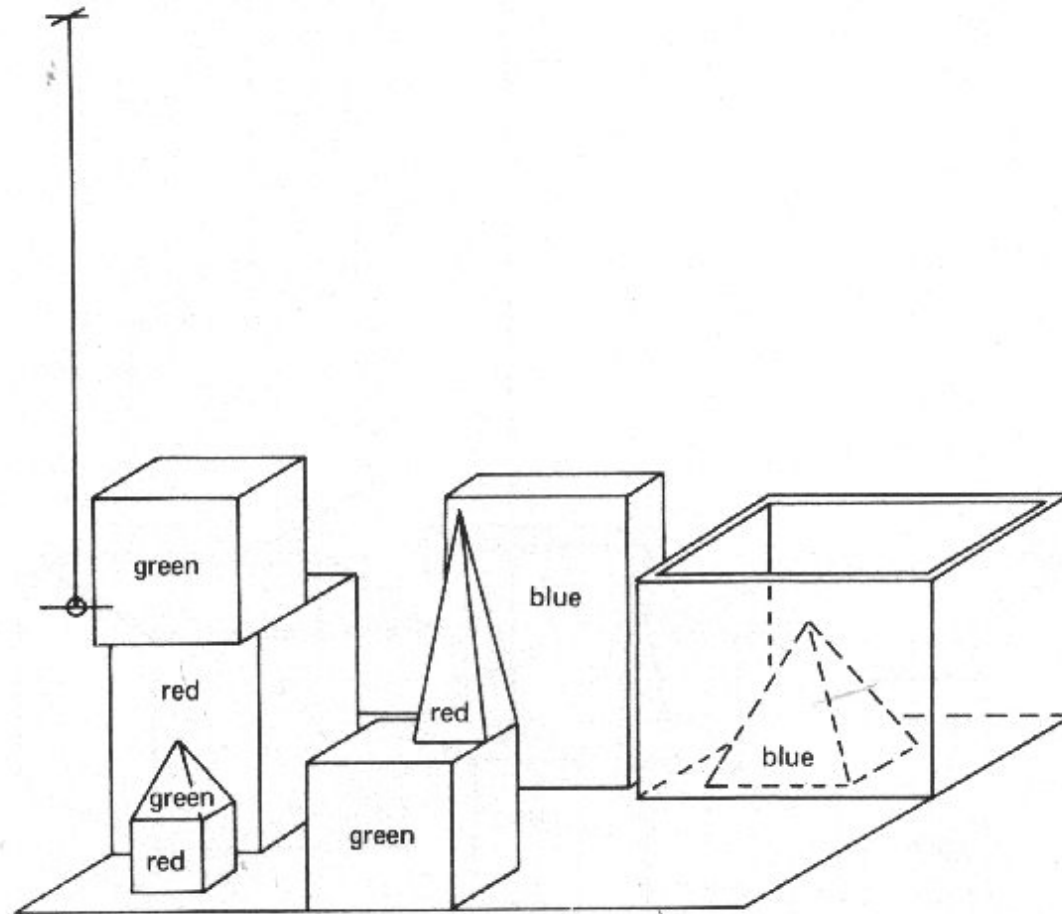
SHRDLU beantwortet Fragen, führt Anweisungen aus und lernt Begriffe.

Wichtige Programmkomponenten von SHRDLU sind:

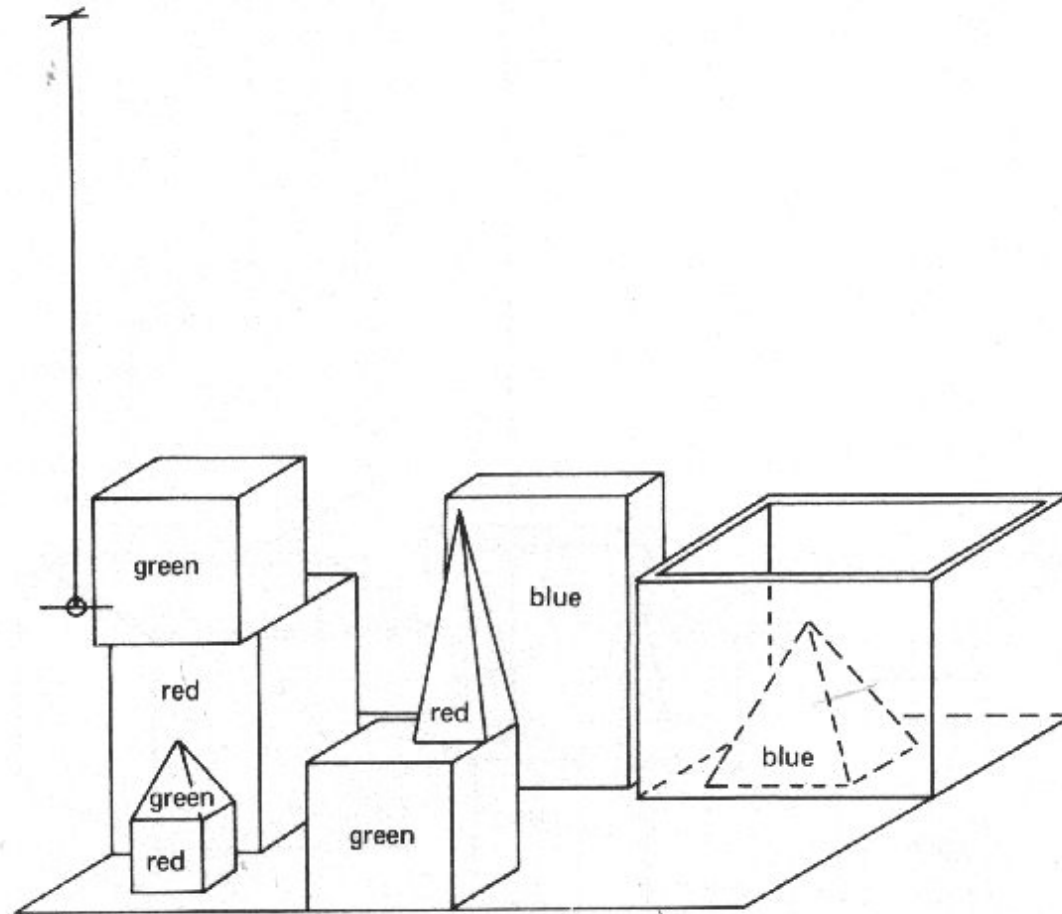
- (Linguistische) Analyse
- Generierung
- (Handlungs-)Planung
- (grafische) Visualisierung

Winograds SHRDLU-System arbeitet in einer kleinen, eingeschränkten Mini-**Welt** oder –**Domäne** ("Blocks World").

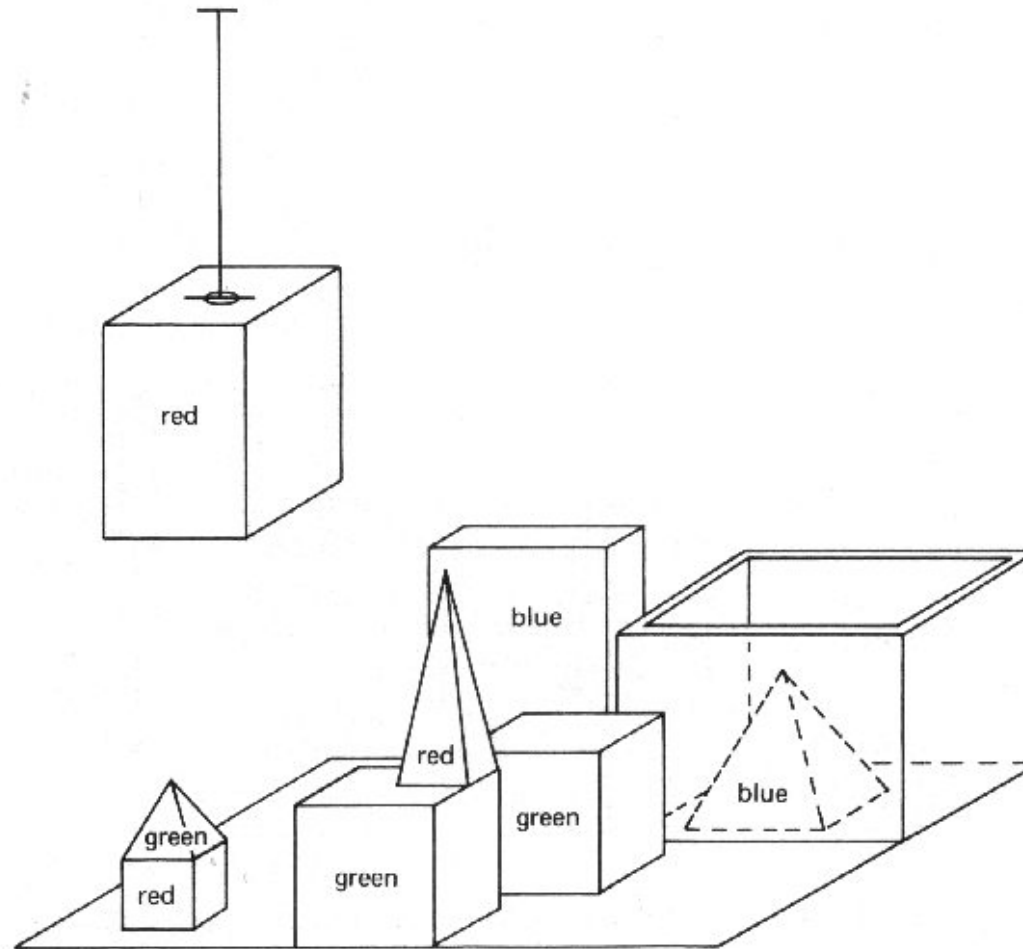
Interessant ist die Interaktion von Analyse und Planung; die Generierungskomponente ist sehr einfach (patternbasiert); die Grafik ist computerlinguistisch nicht sehr interessant, hat aber zum durchschlagenden Erfolg des Systems beigetragen.



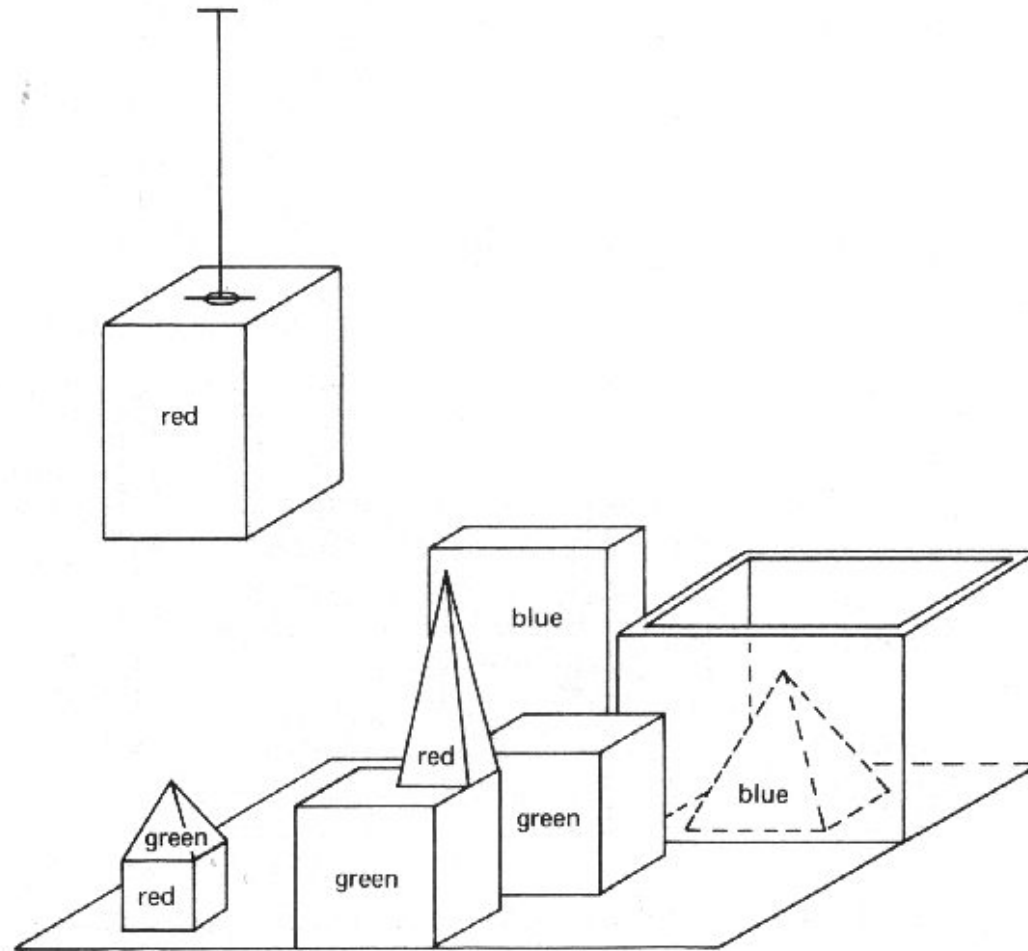
SHRDLU: Die “Blocks World”



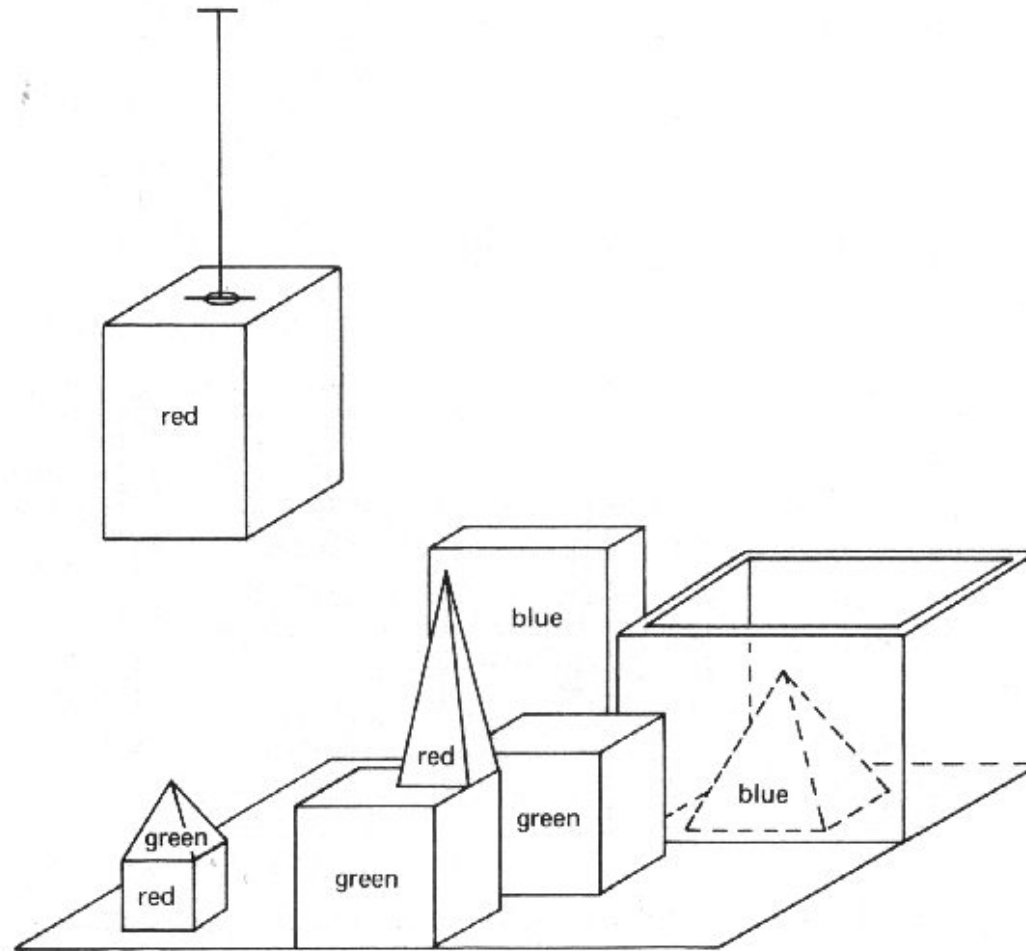
U: Pick up a big red block
S: OK.



(Pick up a big red block)

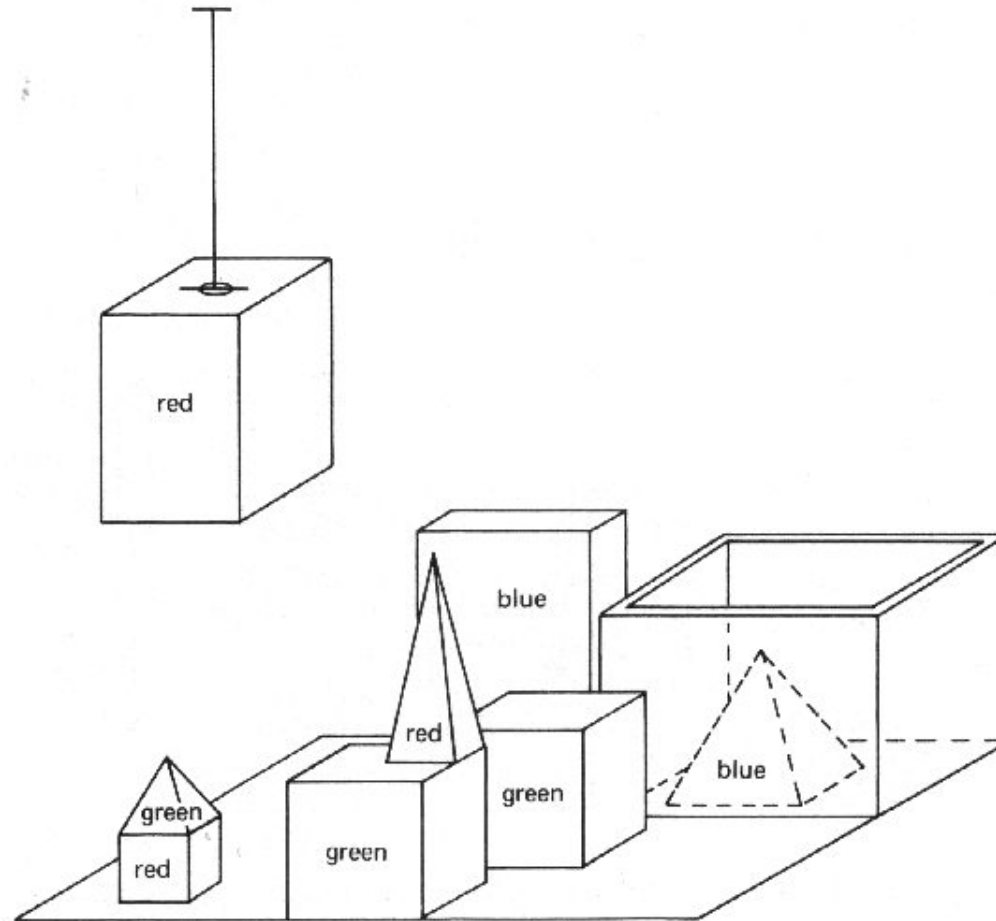


U: Grasp the pyramid.



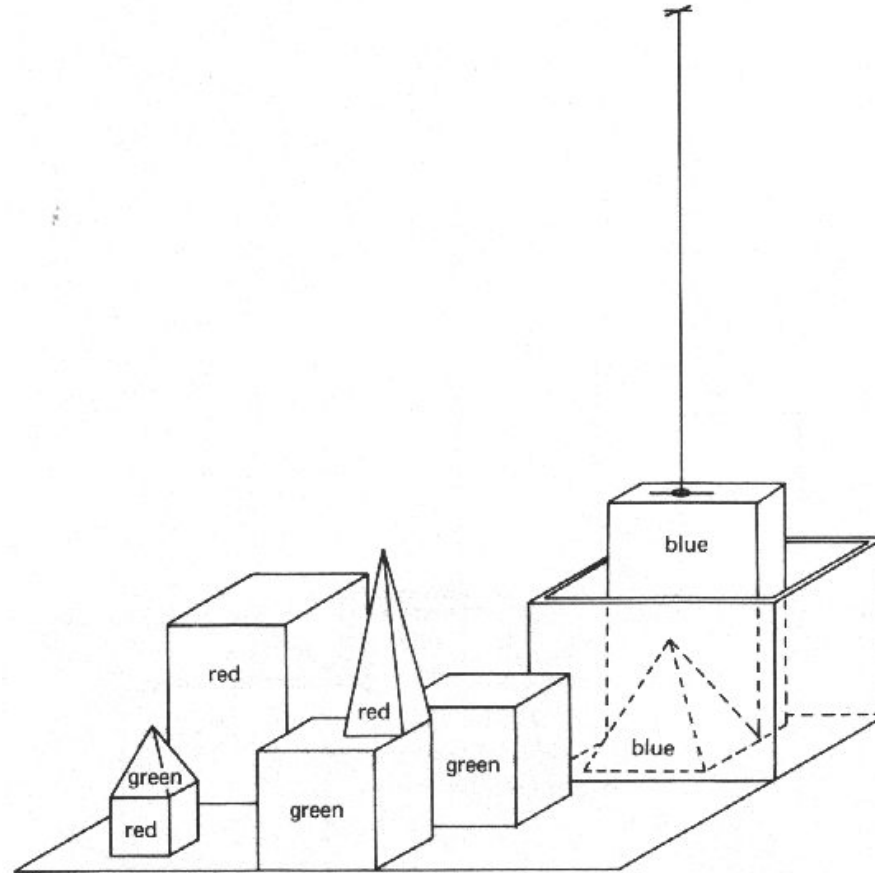
U: Grasp the pyramid.

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

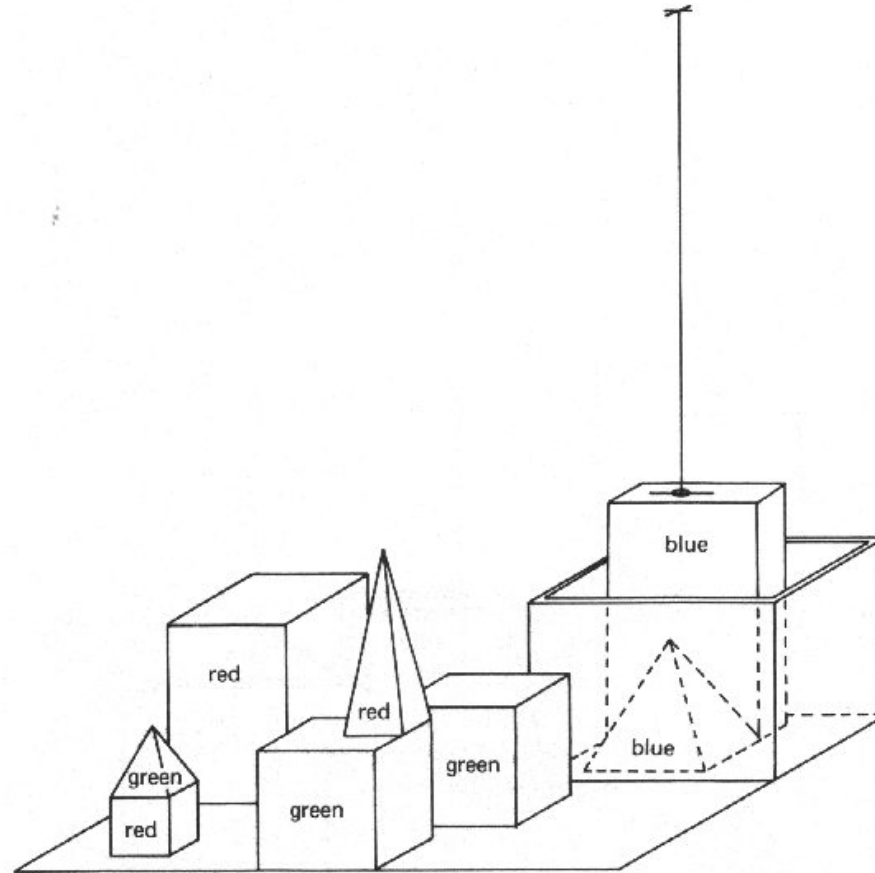


U: Find a block which is taller than the one you are holding and put it into the box.

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

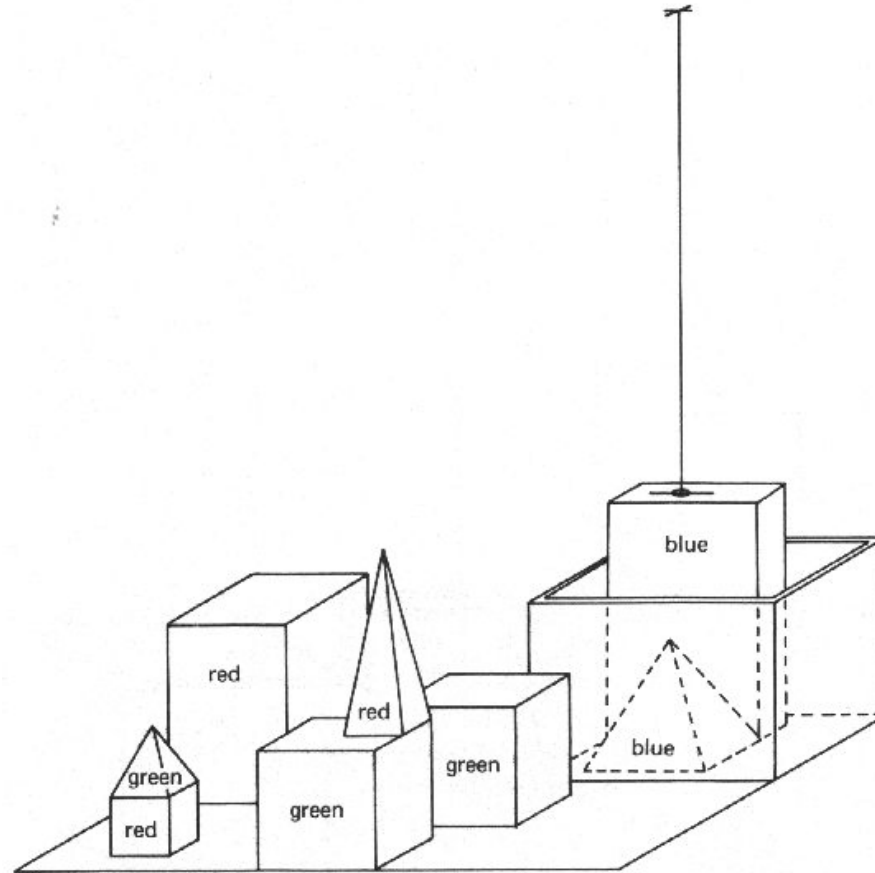


(U: Find a block which is taller than the one you are holding and put it into the box.)

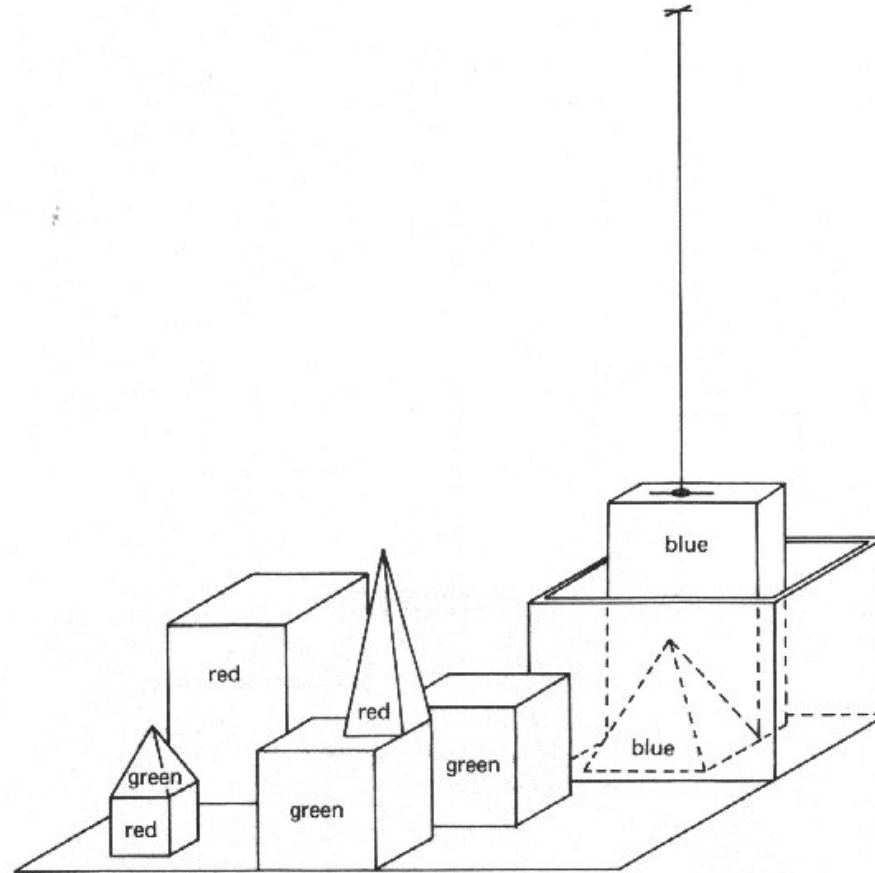


U: What does the box contain?

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

Morphologisches Wissen:

regelmäßige Verben bilden
Präteritum auf -ed

Syntaktisches Wissen:

In Imperativen steht das
Verb an erster Stelle

Semantisches Wissen:

A+N in attributiven
Konstruktionen bezeichnet
Dinge, die unter A und unter
N fallen

Lexikon

grasp ist regelmäßiges Verb
put ist unregelm. Verb mit Prät. *put*

grasp ist transitives Verb
stop ist intransitives Verb

red bezeichnet rote Dinge (?)
pyramid ...
grasp ...

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.

Außersprachliches Wissen

- Kontextwissen:
 - Sprachlicher Kontext / Dialoggeschichte: Welches Objekt wurden zuletzt erwähnt? (*Put it into the box.*)
 - Situationskontext: Welche Objekte kommen in der Äußerungssituation vor? (*What is the pyramid supported by?*)
- Weltwissen:
 - Episodisches Wissen: Wissen über Einzelfakten
 - "Es gibt zwei rote Klötze."*
 - "Die Kiste enthält eine Pyramide"*
 - Regelwissen: Wissen über mathematische, naturwissenschaftliche, gesellschaftliche Regularitäten
 - "Zwei Objekte können nicht den gleichen Platz einnehmen."*
 - "Ein Objekt muss eine ebene Auflagefläche besitzen, damit ein zweites stabil darauf stehen kann"*

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**? Hierzu ist typischerweise Kontext- und Weltwissen nötig.

Woher kommt das Wissen?

Alternativen:

- Manuelle Grammatik- und Lexikon-Entwicklung, Erstellung von extralinguistischen Wissensbasen
- Automatische Extraktion von Wissen aus Sprachdaten (Korpora) mit statistischen Techniken/ maschinellen Lernverfahren