Natural Language Communication with Robots

Yonatan Bisk
ISI-USC

Joint work with:

Deniz Yuret
Koç University

Daniel Marcu
ISI-USC

Components of Communication

Entity/Spatial Grounding

Understanding

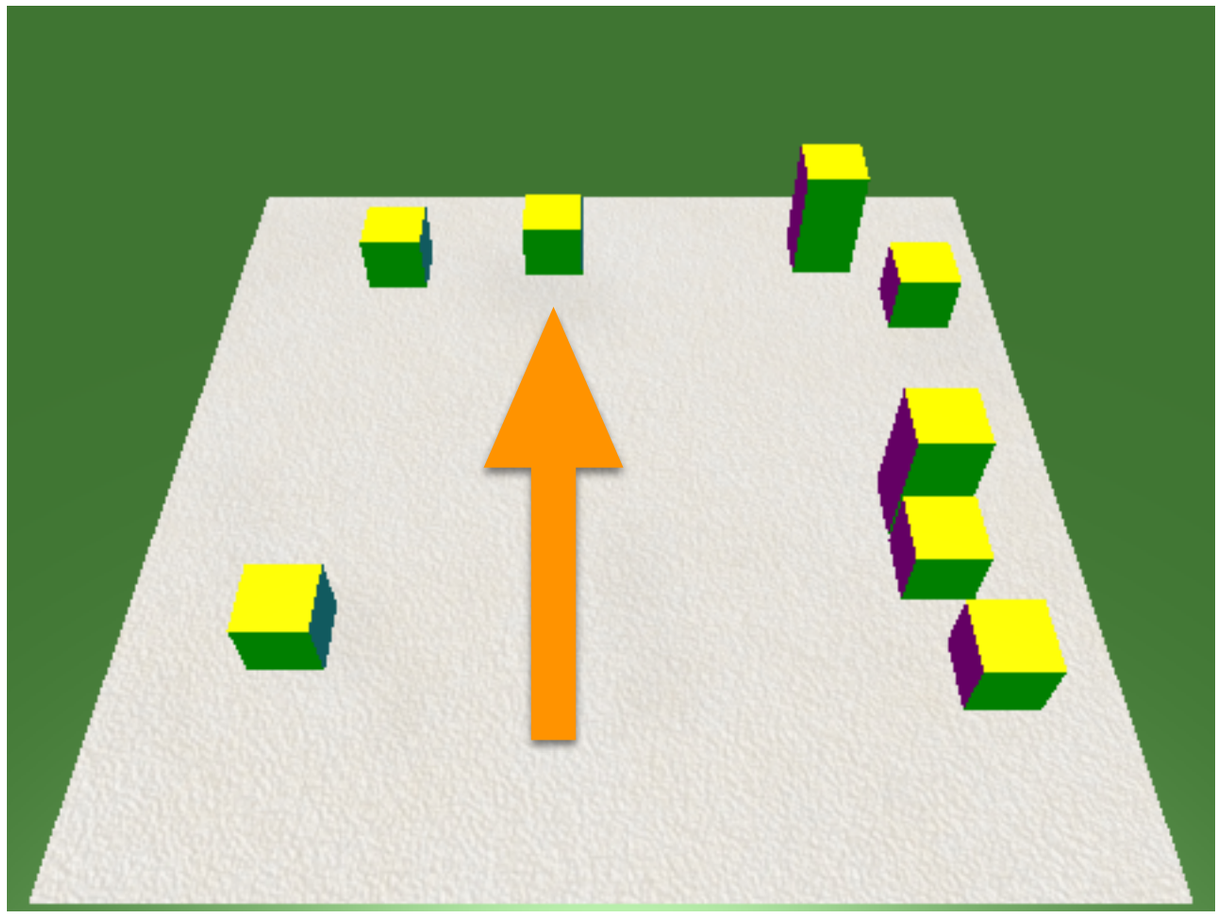
Planning and Plan Recognition

Language Generation

….

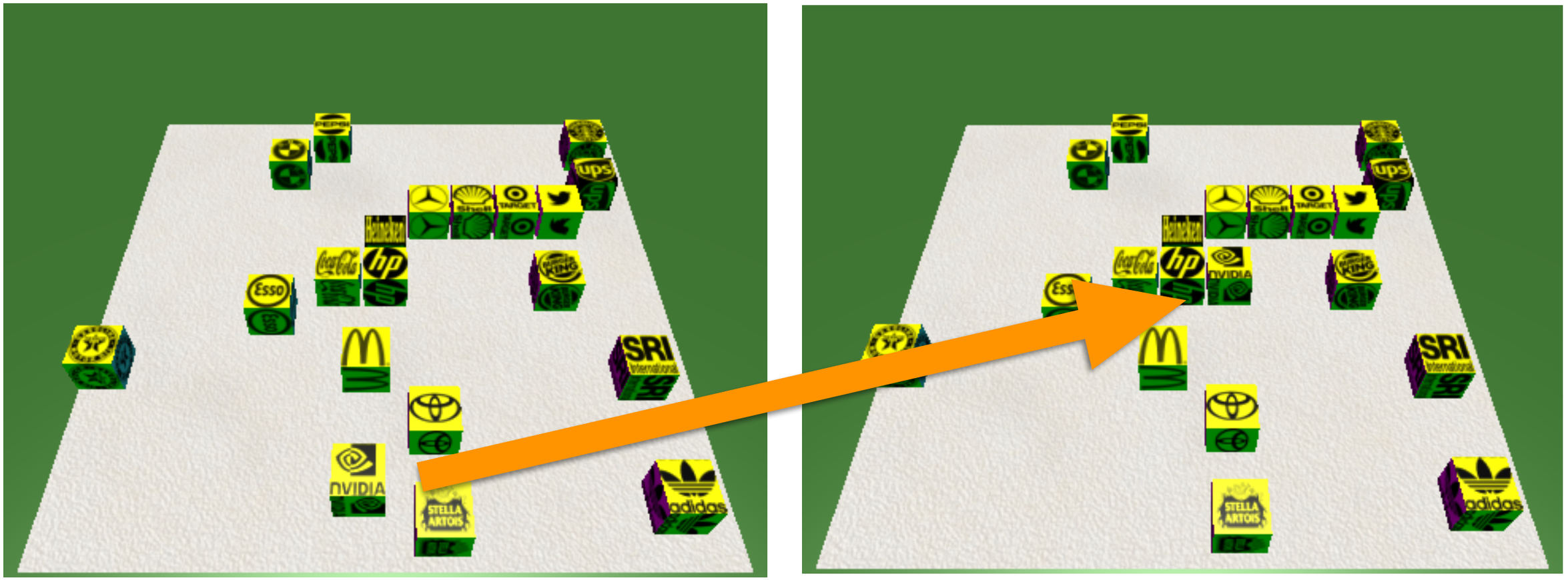
Grounding

The third block from the left

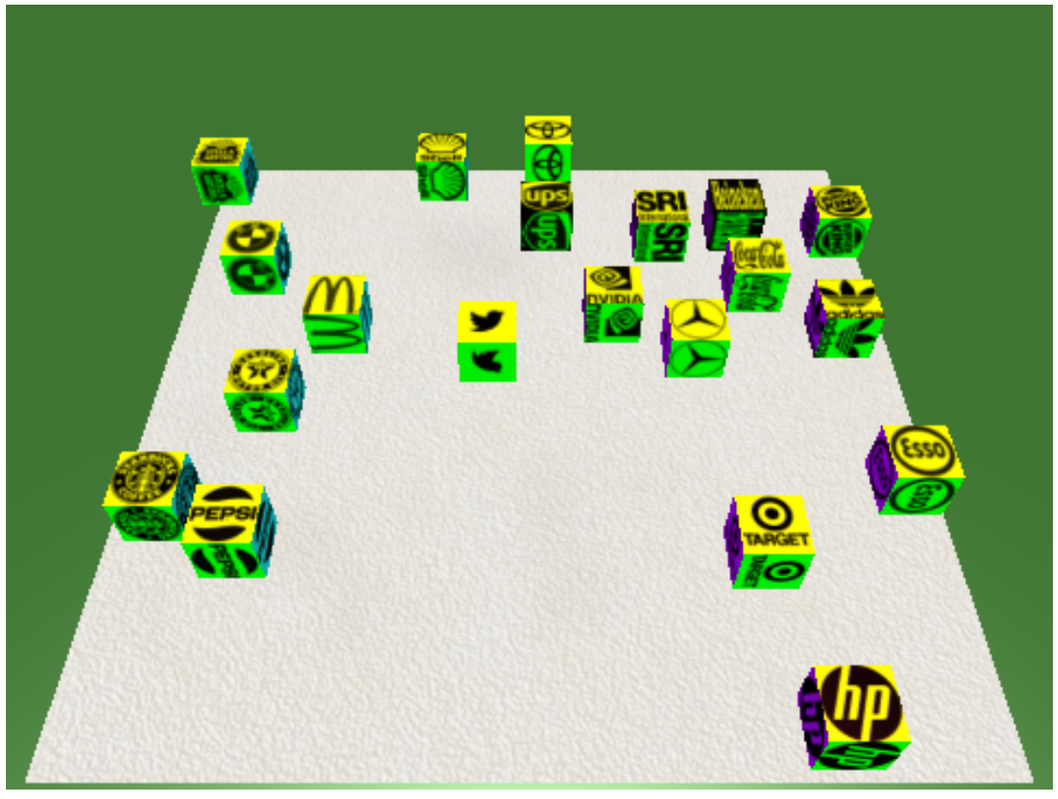


Understanding

place the **nvidia** block **east** of the **hp** block .



Plans



Draw the number six with a rigid base and a right diagonal top. Start with a line of 6 blocks in the middle of the table …

Generation

[I need to] move UPS from the left side of the board to just below Starbucks, leaving a small gap.

Goal

Introduce a dataset collection paradigm for
Human-Robot Communication:
Understanding, Learning, and Generation

1. Easily evaluated
2. Data exists in 3D space
3. Natural language utterances
4. Parallel annotation at differing levels of abstraction
5. Computer Vision can help but is not a pre-requisite

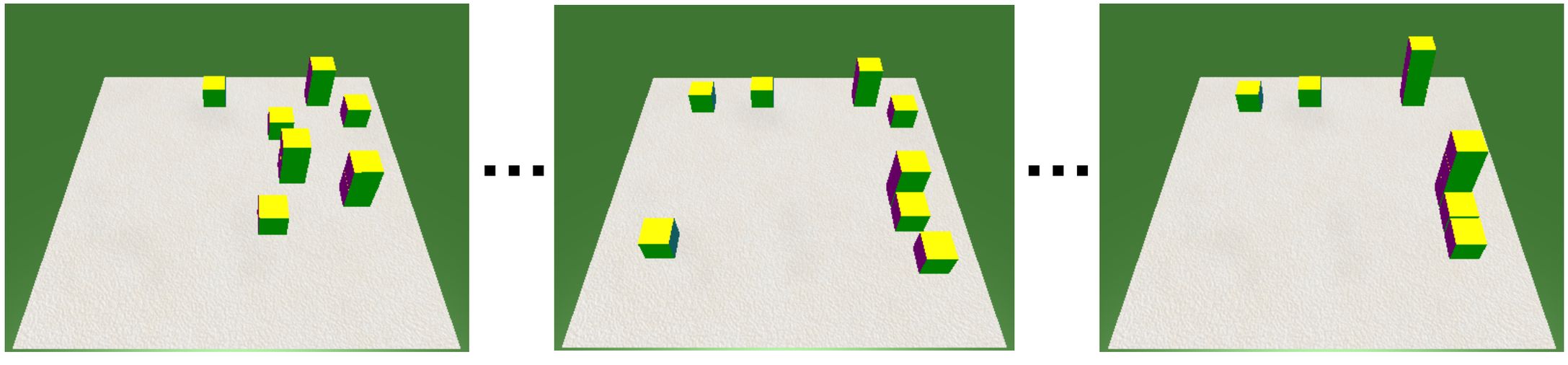
+ Models to begin addressing understanding

Dataset

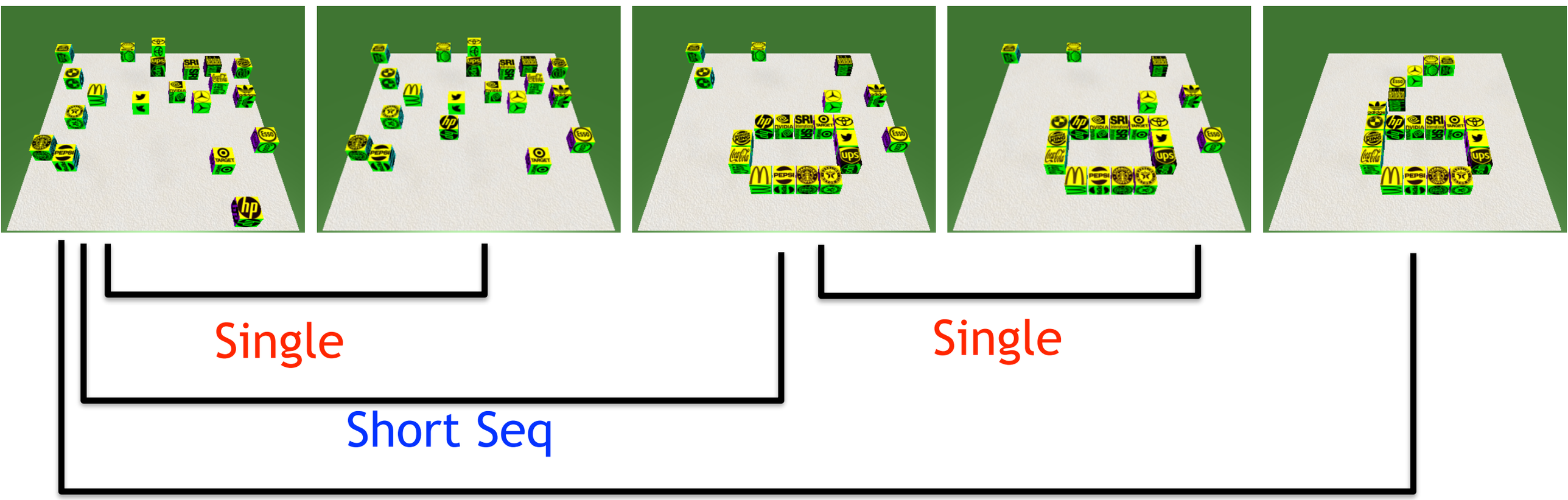
Action Sequences

Identifiable Sequences

Random Blank Sequences



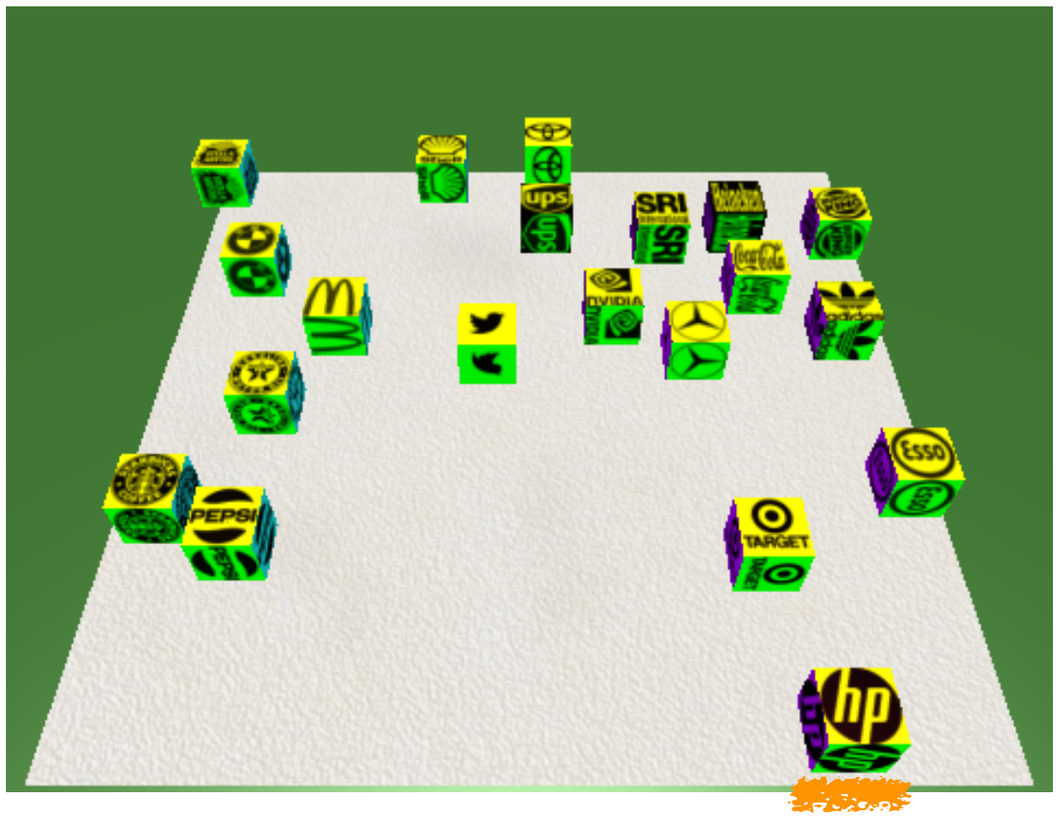
Problem Solution Sequences

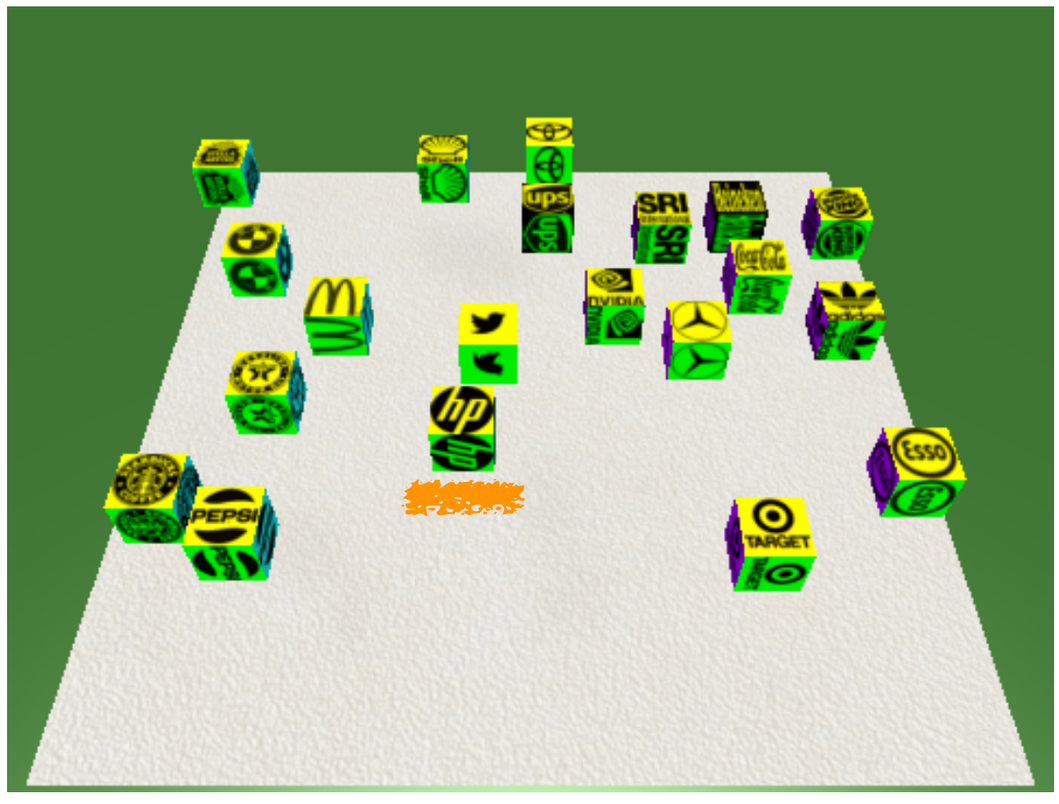


We focus on Single Actions in this work

Corpus Creation

Simple Actions

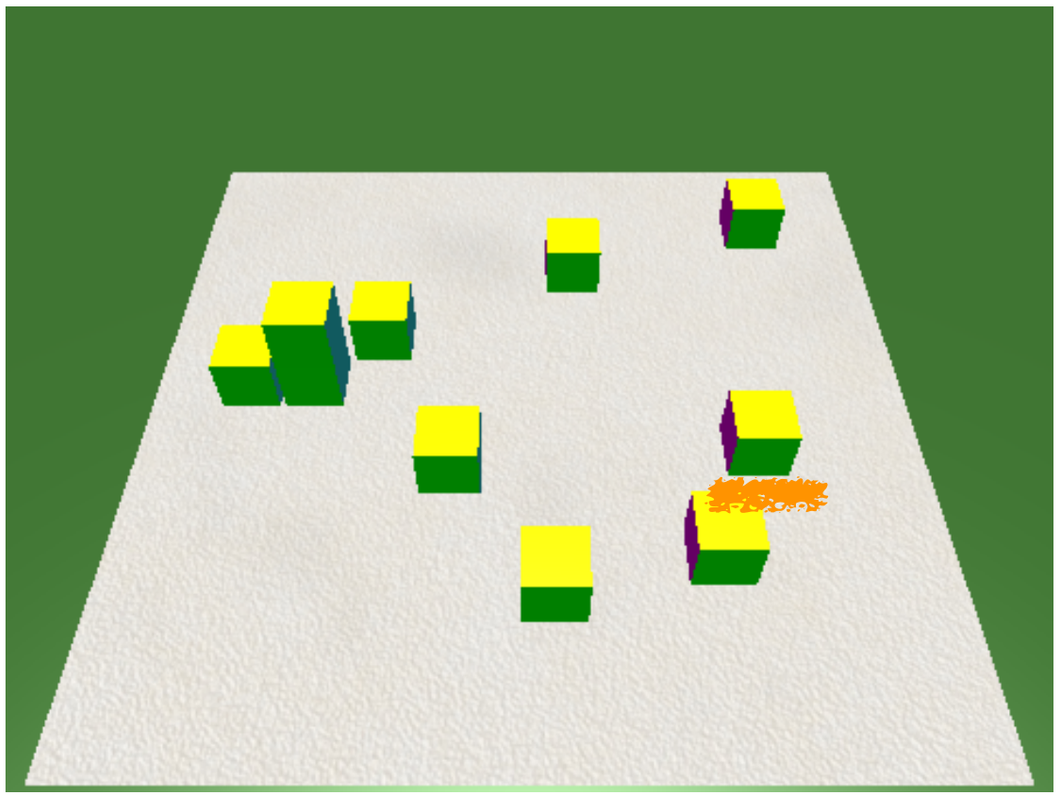


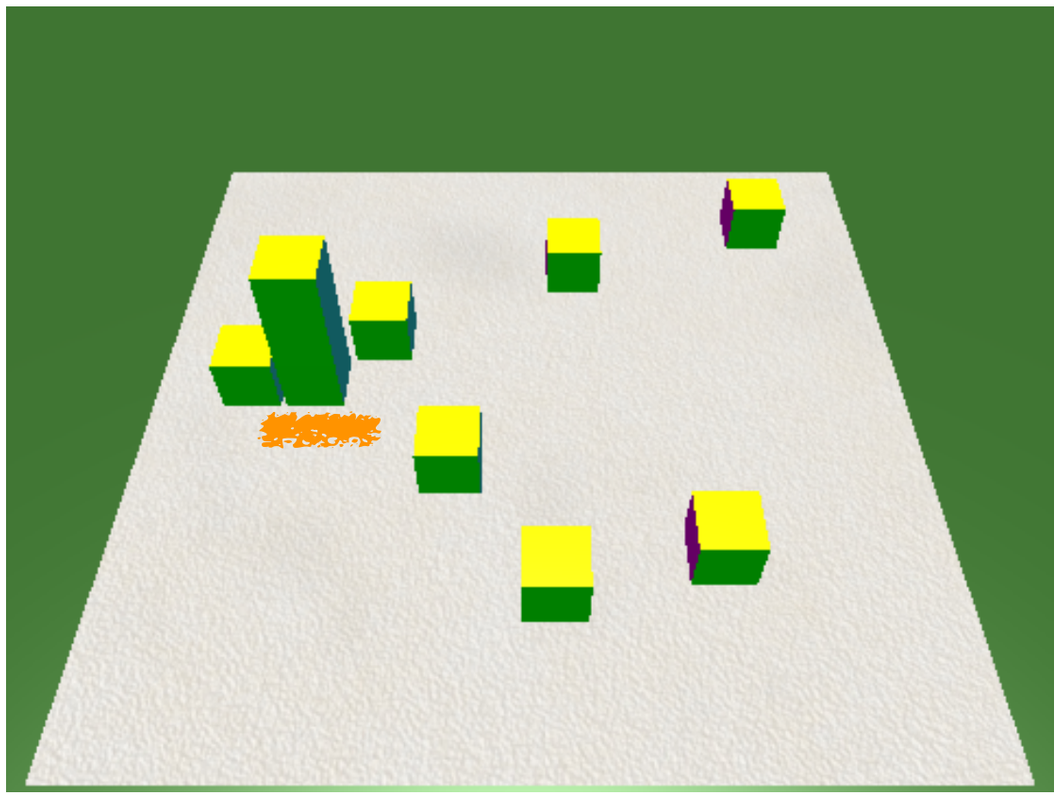


Move HP in front of Twitter and slightly to the left

Corpus Creation

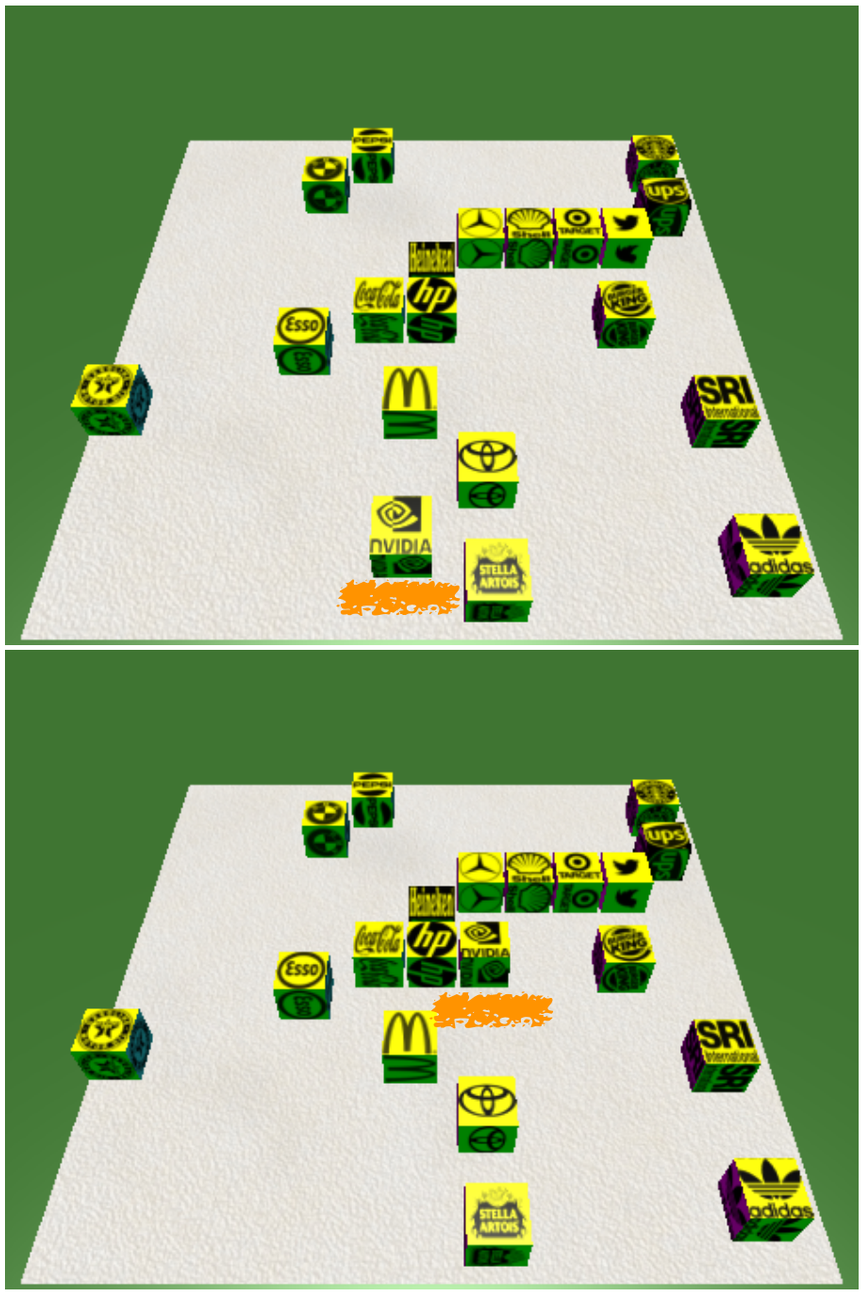
Difficult Actions





Remove the block above the right bottom block and place it on top of the left stack of blocks.

Nine Annotations



1. **coca cola** , **hp** , **nvidia** .
2. **nvidia** , to the **right** of **hp**
3. place the **nvidia** block **east** of the **hp** block .
4. move the **nvidia** block to the **right** of the **hp** block
5. place the **nvidia** block to the **east** of the **hp** block .
6. move the **nvidia** block directly to the **right** of the **hp** block .
7. move the **nvidia** block just to the **right** of the **hp** block **in line with** the **mercedes** block .
8. put the **nvidia** block on the **right** end of **the row of blocks that includes** the **coca cola** and **hp** blocks .
9. put the **nvidia** block on the **same row as** the **coca cola** block, in the **first open space to the right** of the **coca cola** block .

Corpus Statistics[V1]

	Actions	Types	Tokens	Ave Len
MNIST	11,870	1,359	~257K	15 tokens
Random	2,492	1,172	~84K	23.5 tokens

Natural Language Understanding

Action Understanding

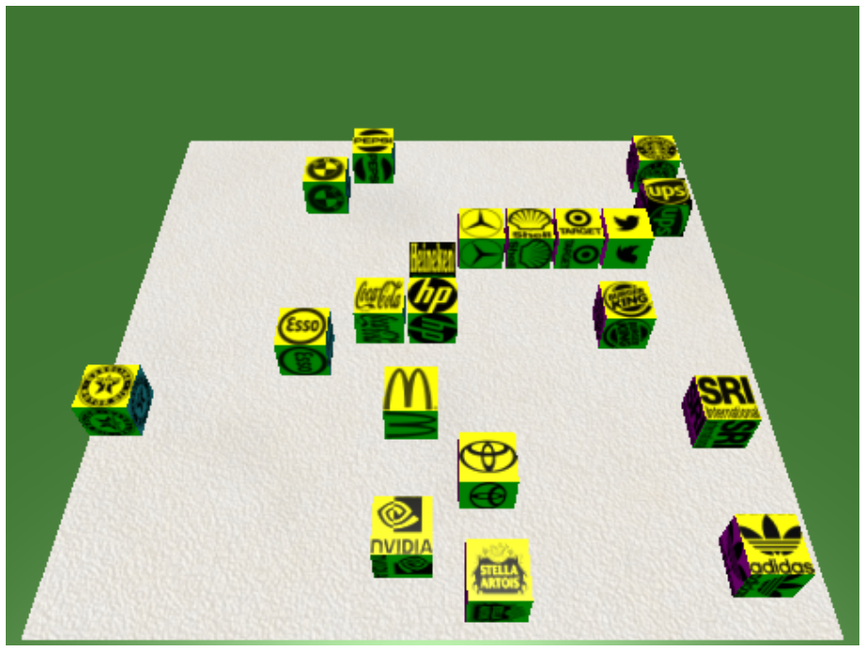
Given:

World

Utterance

Goal:

Execute a command



place the **nvidia** block **east** of the **hp** block .

Block to Move

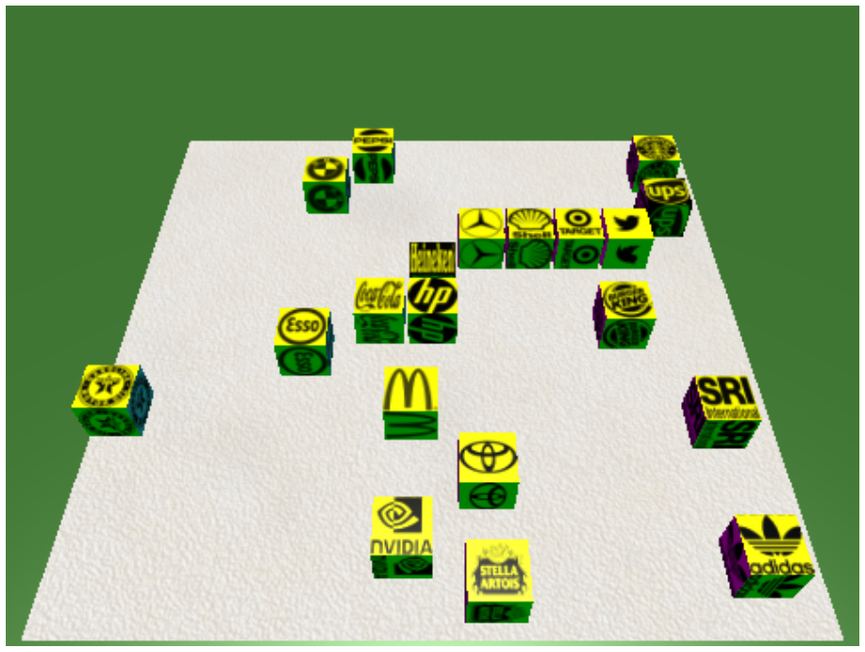
$(x, y, z)_S$

Where to Move

$(x, y, z)_T$

World Representation

Images (w/ Occlusion)



Exact Locations

Adidas	0.8	0.1	0.76
BMW	-0.3	0.1	-0.4
Burger King	0.5	0.1	0.14
Coke	-0.07	0.1	0.00

…

This Work
20 x 3 Matrix

Evaluation: Euclidean Distance

Block to Move

$$||(x, y, z)_{SPred} - (x, y, z)_{SGold}||_2$$

Where to Move

$$||(x, y, z)_{TPred} - (x, y, z)_{TGold}||_2$$

Baseline Models

Output: Block to Move $(x, y, z)_S$ Where to Move $(x, y, z)_T$

Random

Random Block to move
Random Block to place it next to

We also Perform Human Evaluation

Center

Perfect knowledge of which block to move
Always place it in the center of the board

Simple Semantics

Model 1: A Discrete world (Source, Direction, Reference)

Move the BMW block in front of the Adidas block

Move the Source block Direction the Reference block

∈ [1,20]

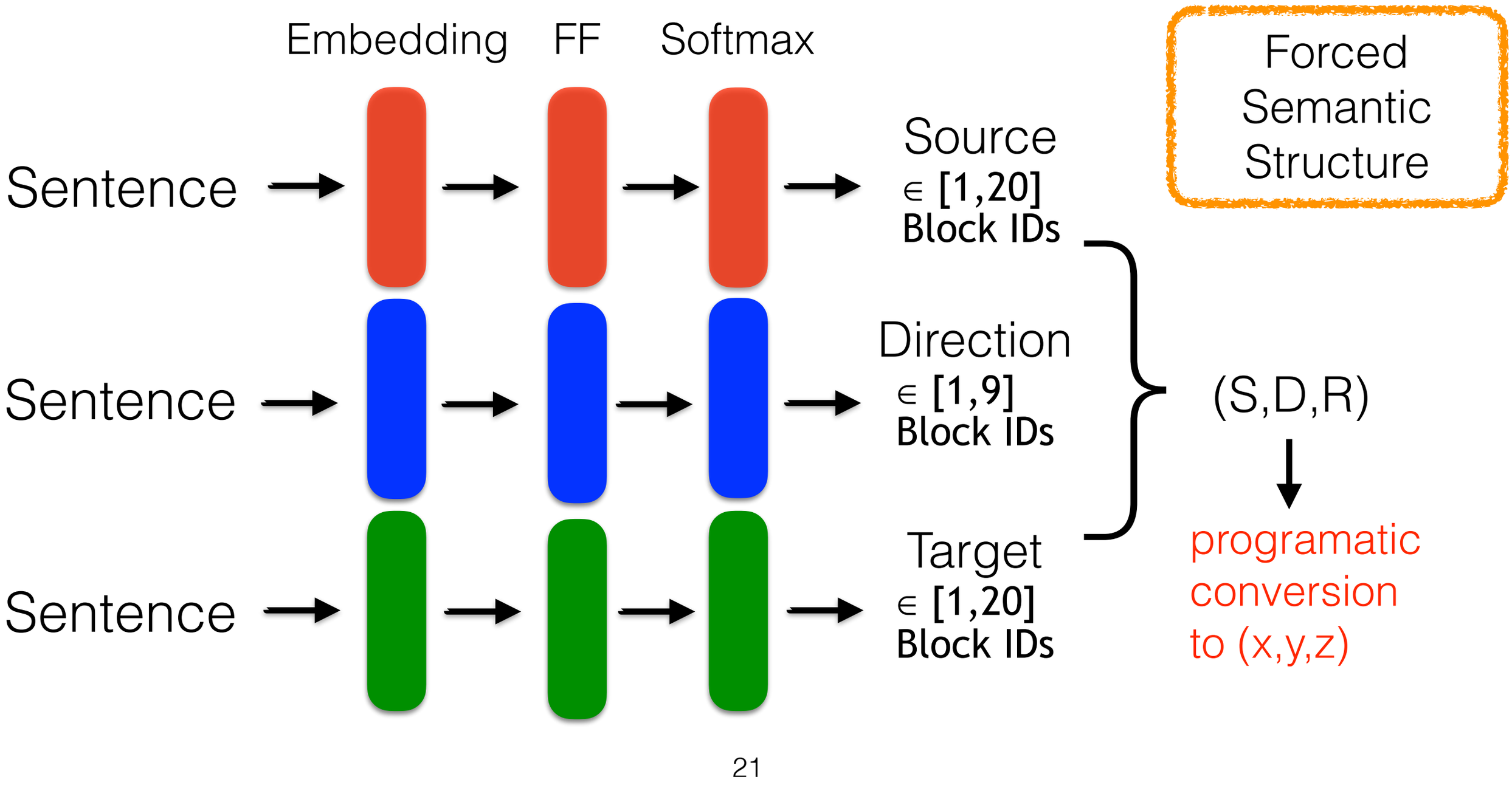
∈ [1,9]

NW	N	NE
W	TOP	E
SW	S	SE

∈ [1,20]

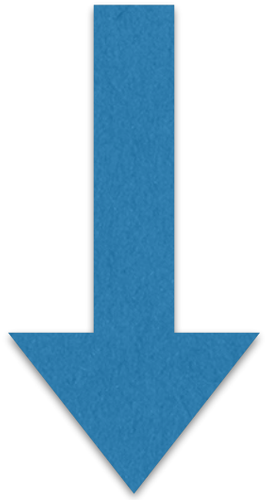
Simple Semantics

Model 1: A Discrete world (Source, Direction, Reference)



End-to-End Model

Move the BMW block in front of the Adidas block



$(x, y, z)_{SPred}$

or

$(x, y, z)_{TPred}$

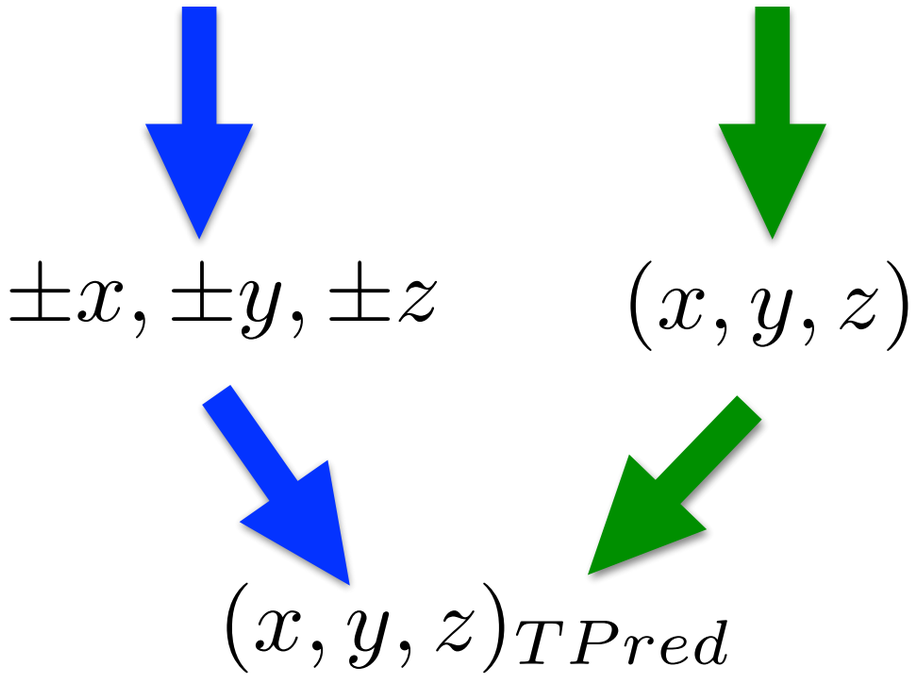
End-to-End Model

Move the BMW block in front of the Adidas block

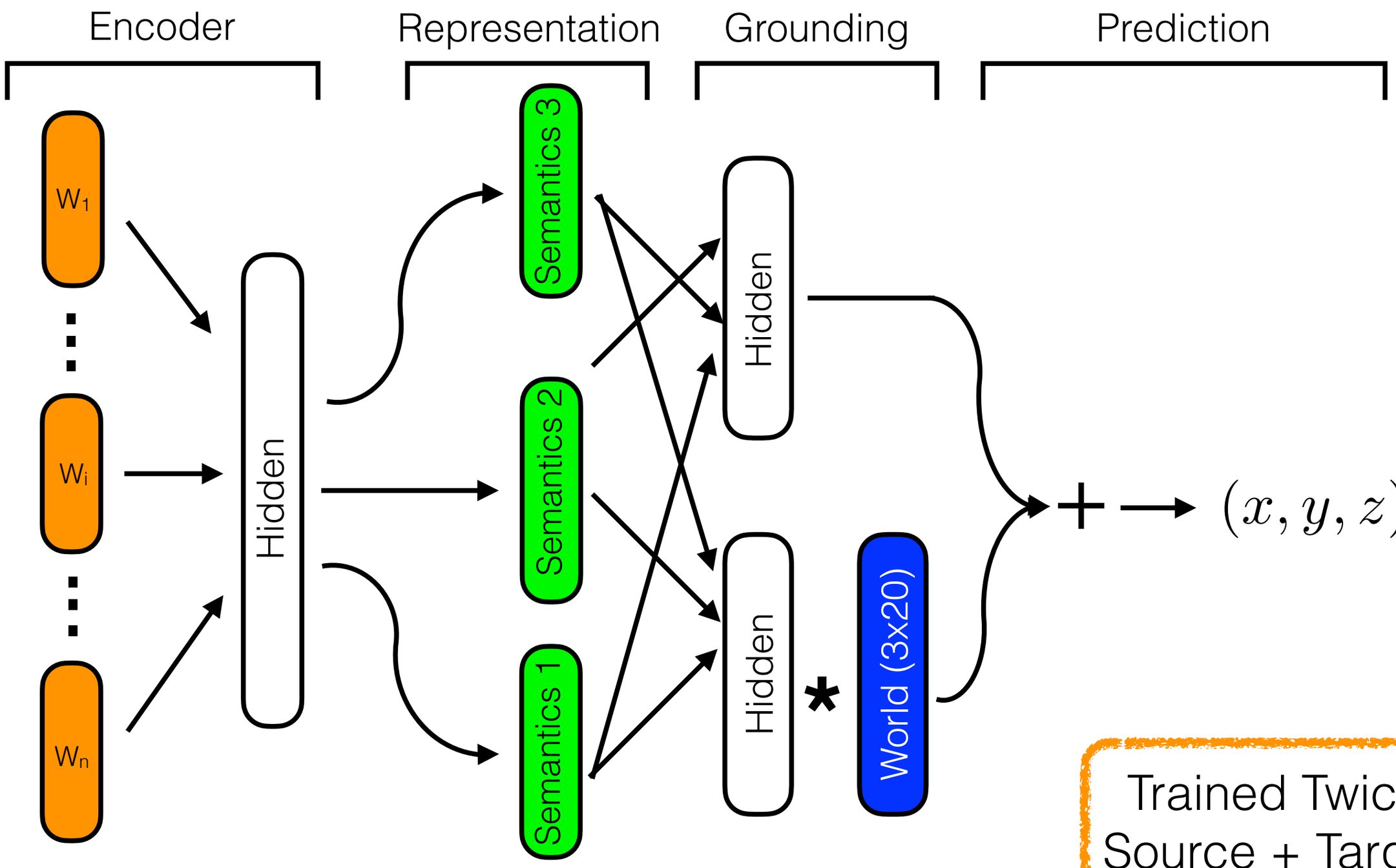
Direction

Reference

Assumed Logic:
Can we encode this?



End-to-End Model



MNIST Performance

	Source	Target
	Mean	Mean
Human	0.00	0.53
Simple Semantics	**0.14**	**0.98**
End-To-End	0.19	1.05
Center Baseline		3.43
Random Baseline	6.49	6.21

Blank Block Performance

	Source	Target
	Mean	Mean
Human	0.30	1.39
Simple Semantics	5.00	5.57
End-To-End	**3.47**	**3.70**
Center Baseline		4.06
Random Baseline	4.97	5.44

Common Errors

Multi-relation actions

Place block 20 parallel with the 8 block and slightly to the right of the 6 block.

Geometric Understanding

Continue the diagonal row of 20, 19 and 15 downward with 13.

Grammatical Ambiguity

19 moved from behind the 8 to under the 18th block.

Summary

This Work:

- Initial Models for Language Understanding
- An environment for exploring grounded phenomena

Moving Forward:

- Language Generation, Planning, …
- Increased task difficulty.

Thanks!

http://nlg.isi.edu/language-grounding/