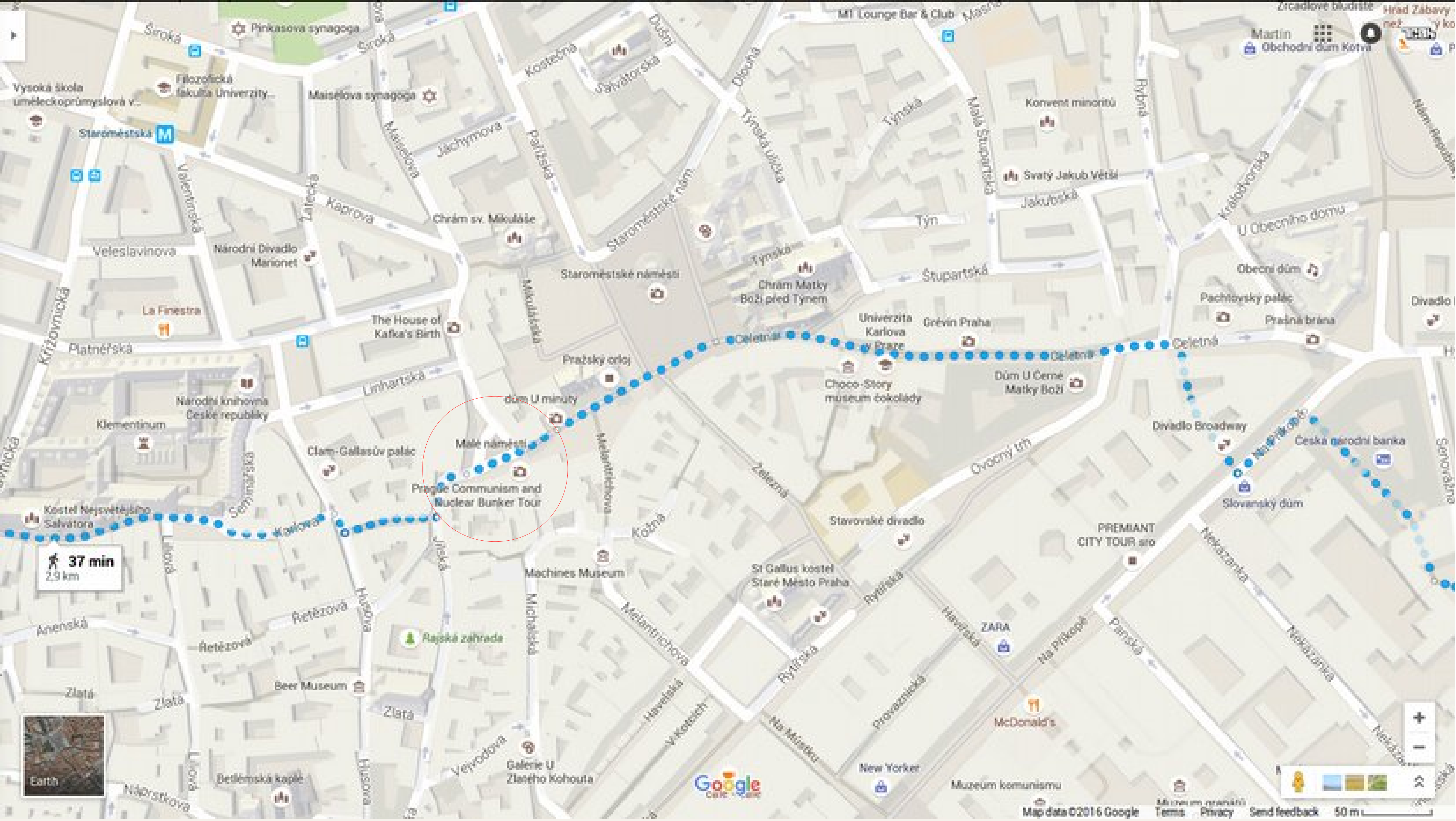
INTERACTIVE NATURAL LANGUAGE GENERATION IN VIRTUAL ENVIRONMENTS

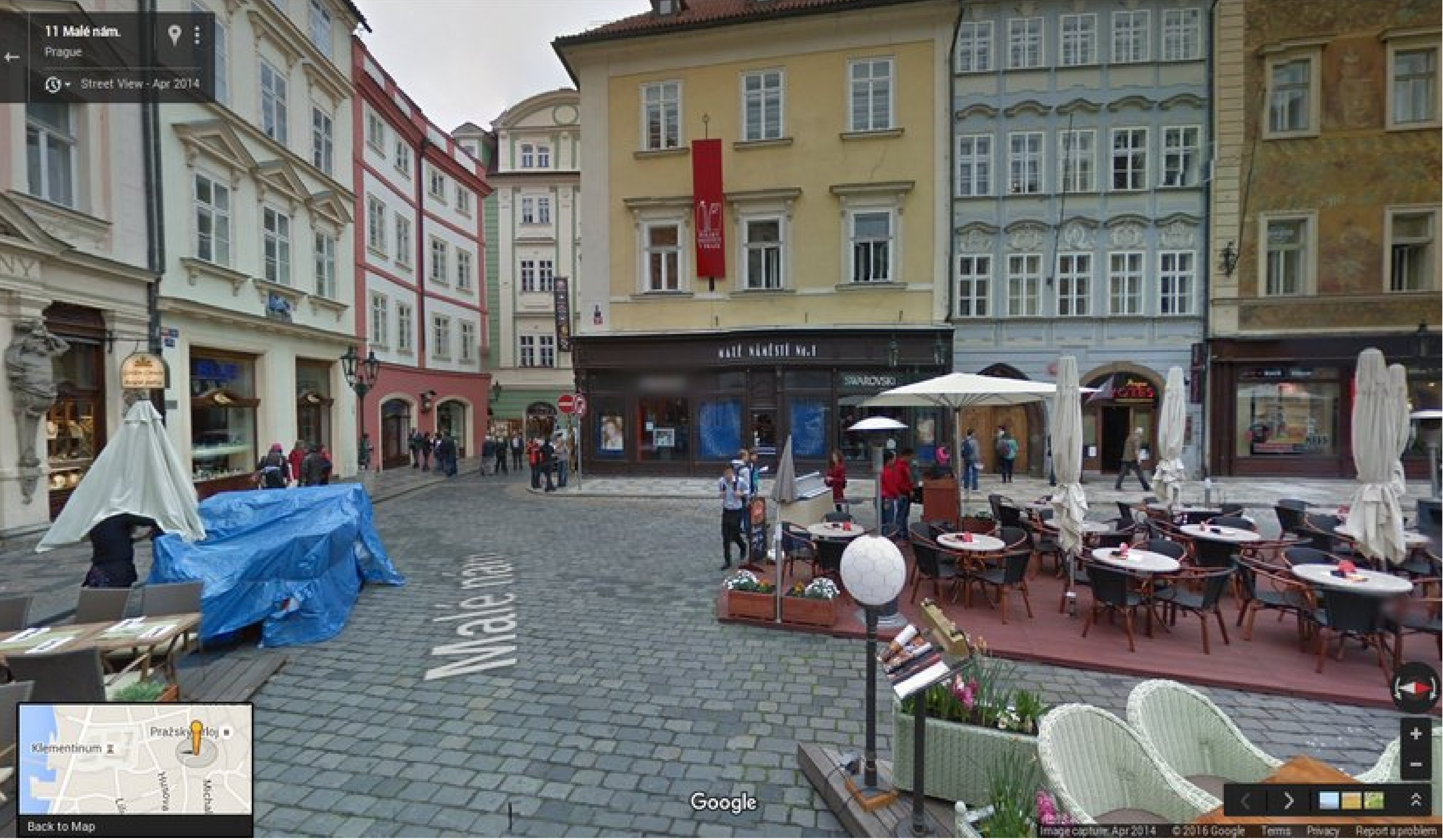
MARTIN VILLALBA
ALEXANDER KOLLER
NIKOS ENGONOPOULOS

UNIVERSITY OF POTSDAM

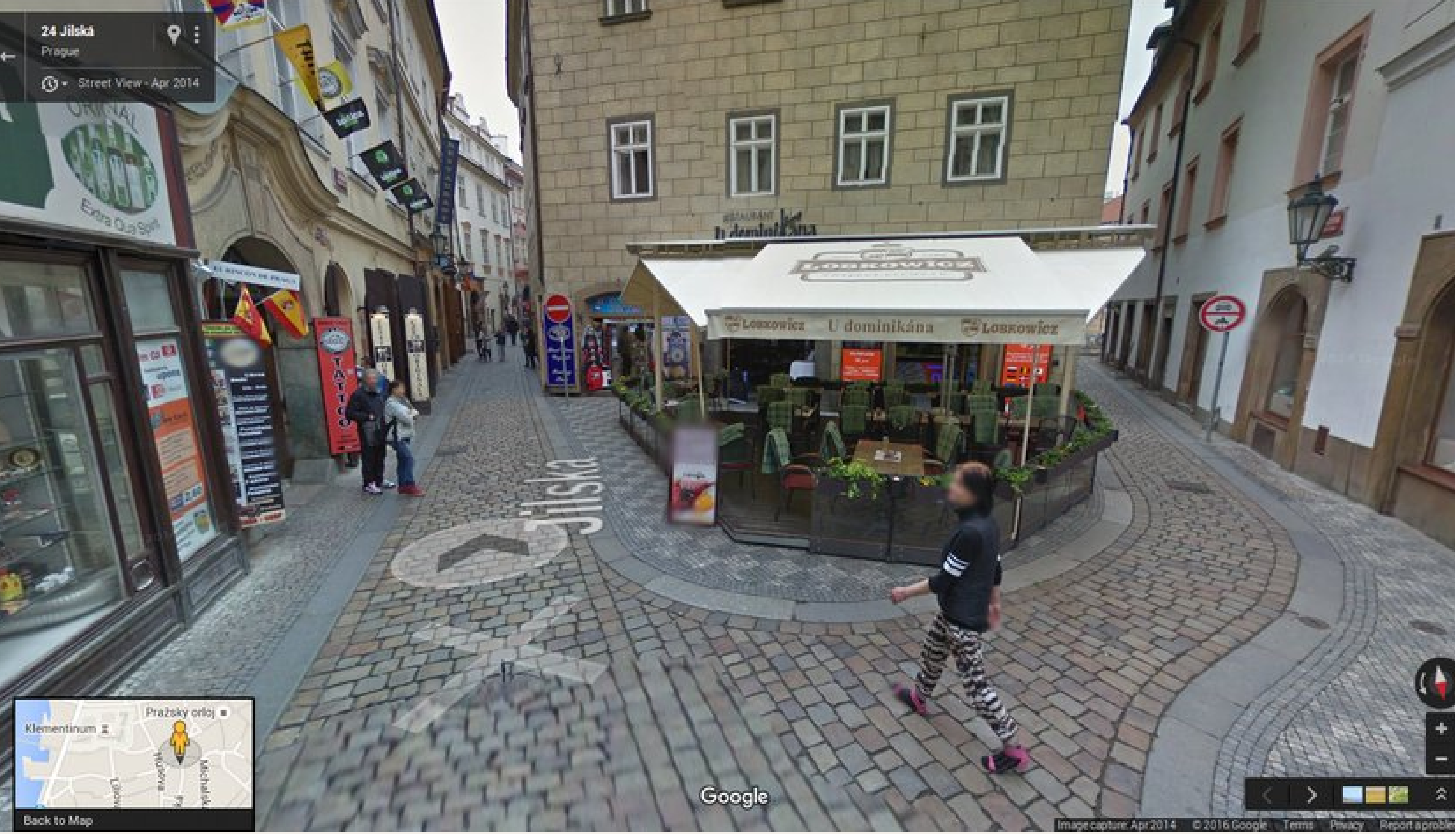
Picture by Andrei Pop, via Flickr



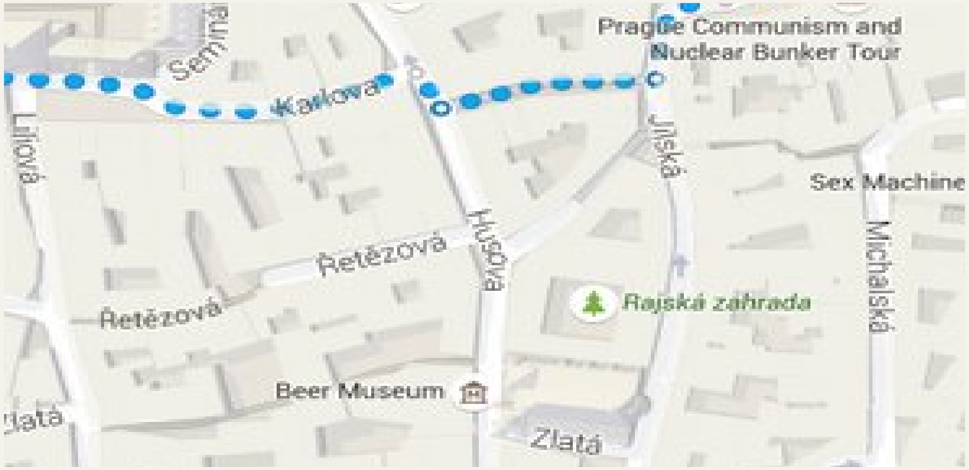
“Walk straight on Malé náměstí,
and turn left on Jilská”



"Walk straight on Malé náměstí,
and turn left on Jilská"



"Walk straight on Malé náměstí,
and turn left on Jilská"



Problem I

The real world is complicated to deal with

Problem II

We need to refer to individual objects

Problem III

Sometimes there are misunderstandings

REFERRING EXPRESSIONS

A NOUN PHRASE THAT IDENTIFIES UNIQUELY A CERTAIN OBJECT WITHIN A SCENE

Part I

Instructions in a virtual environment

Part II

A model of listener's understanding

Part III

Generating the best RE

Part IV

Dealing with misunderstandings

Future work

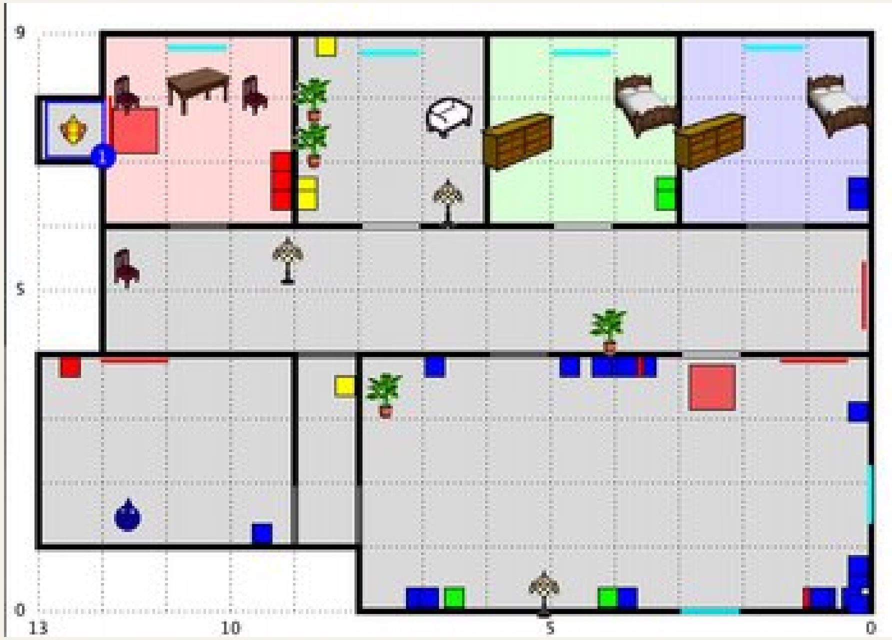
PART I: GIVING INSTRUCTIONS

INSTRUCTIONS IN A VIRTUAL ENVIRONMENT

Picture by Andrei Pop, via Flickr

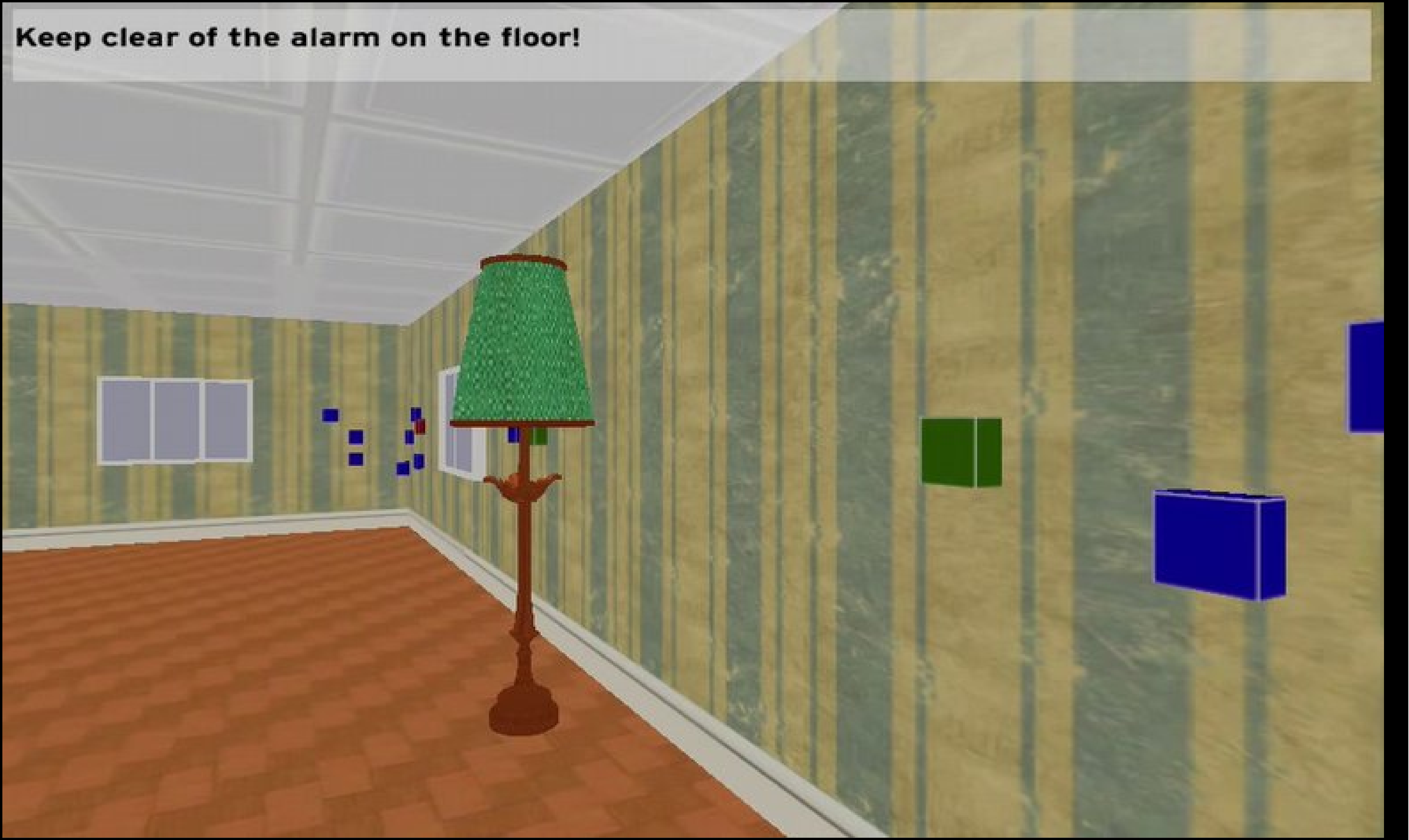
METHODOLOGY: The GIVE Challenge

GENERATING INSTRUCTIONS IN VIRTUAL ENVIRONMENTS

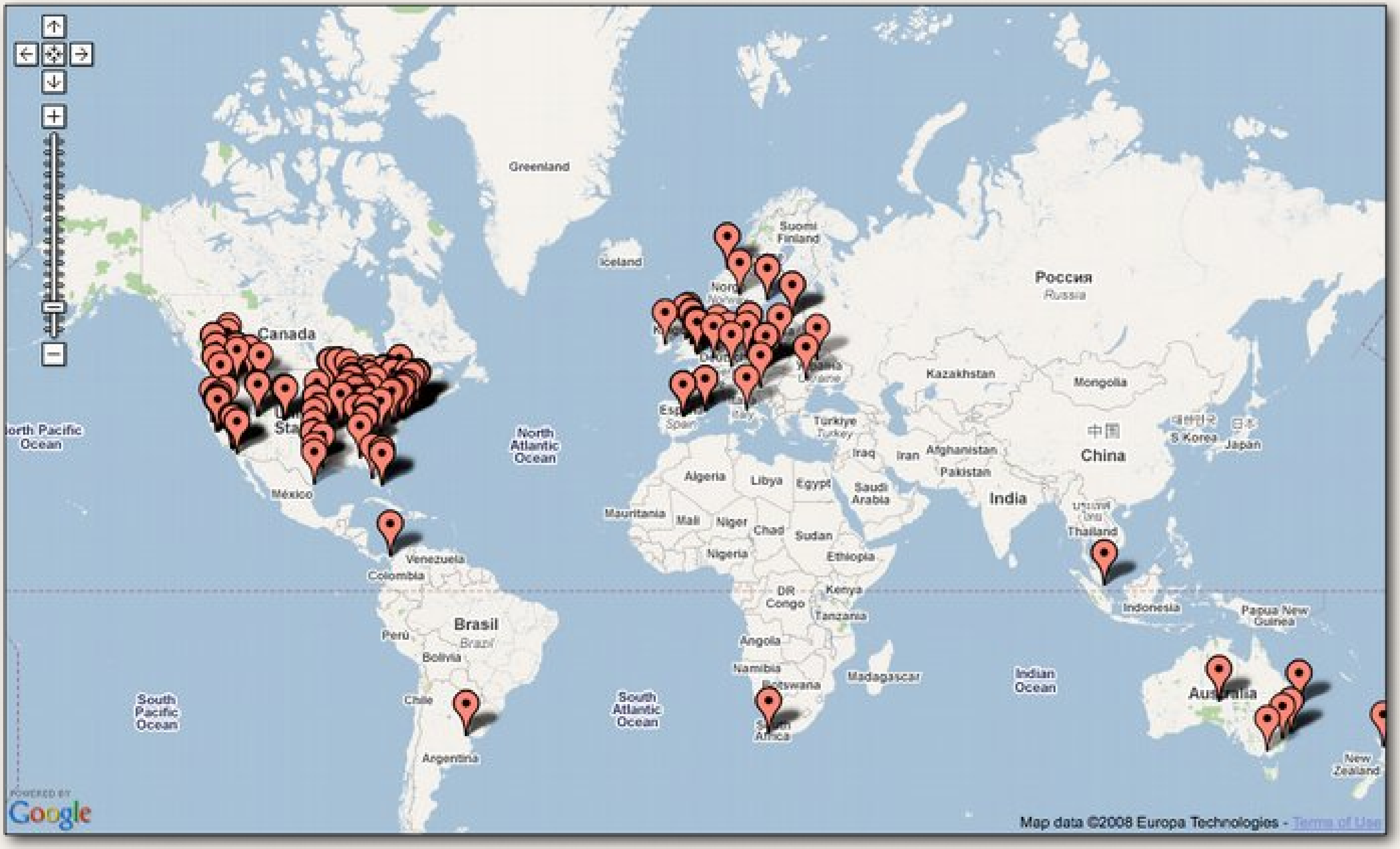


Help a human player solve a puzzle through automatically generated, real-time instructions

Report on the Second NLG Challenge on
Generating Instructions in Virtual Environments (Koller et al, 2010)



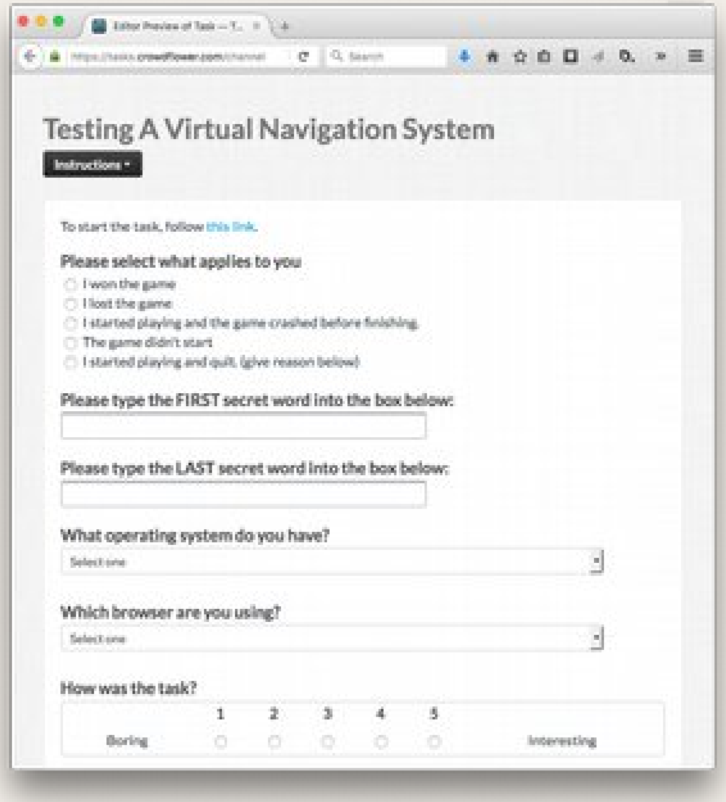
Keep clear of the alarm on the floor!



Data collected for GIVE-1 on 20/11/2008

	Year	Systems	Games
GIVE-1	2008/09	5	1143
GIVE-2	2009/10	7	1825
GIVE-2.5	2011	8	661

CROWDSOURCING
OUR EXPERIENCE



Testing A Virtual Navigation System

Instructions ▾

To start the task, follow this link.

Please select what applies to you

○ I won the game
○ I lost the game
○ I started playing and the game crashed before finishing.
○ The game didn't start
○ I started playing and quit. (give reason below)

Please type the FIRST secret word into the box below:

Please type the LAST secret word into the box below:

What operating system do you have?

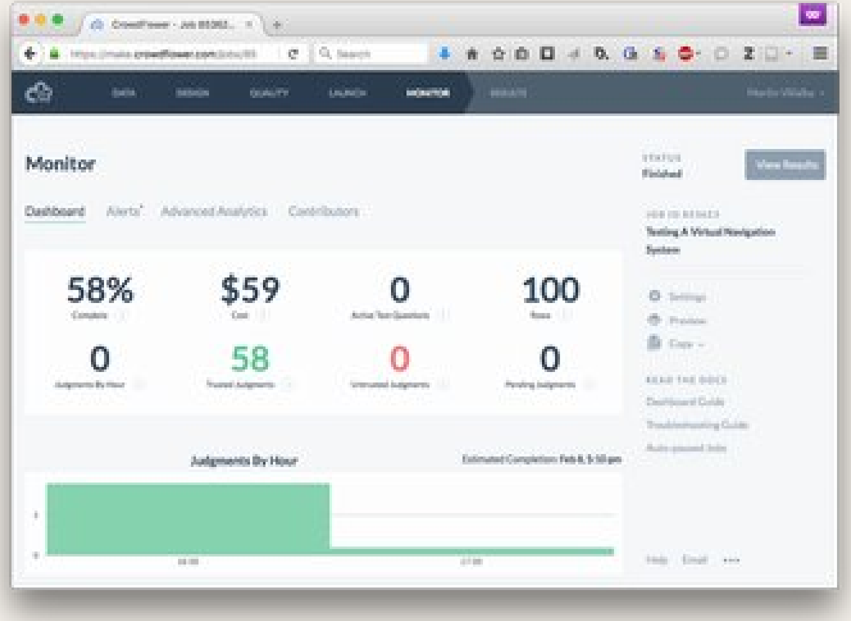
Select one

Which browser are you using?

Select one

How was the task?

	1	2	3	4	5	
Boring	○	○	○	○	○	Interesting



Monitor

Dashboard | Alerts | Advanced Analytics | Contributors

58% Complete	$59 Cost	0 Active Test Questions	100 Rows
0 Judgments By Hour	58 Trusted Judgments	0 Untrusted Judgments	0 Pending Judgments

STATUS
Finished

View Results

Testing A Virtual Navigation System

Settings
Preview
Copy

READ THE DOCS
Dashboard Guide
Troubleshooting Guide
Auto-paused Jobs

Judgments By Hour

Estimated Completion: Feb 8, 5:10 pm

CrowdFlower

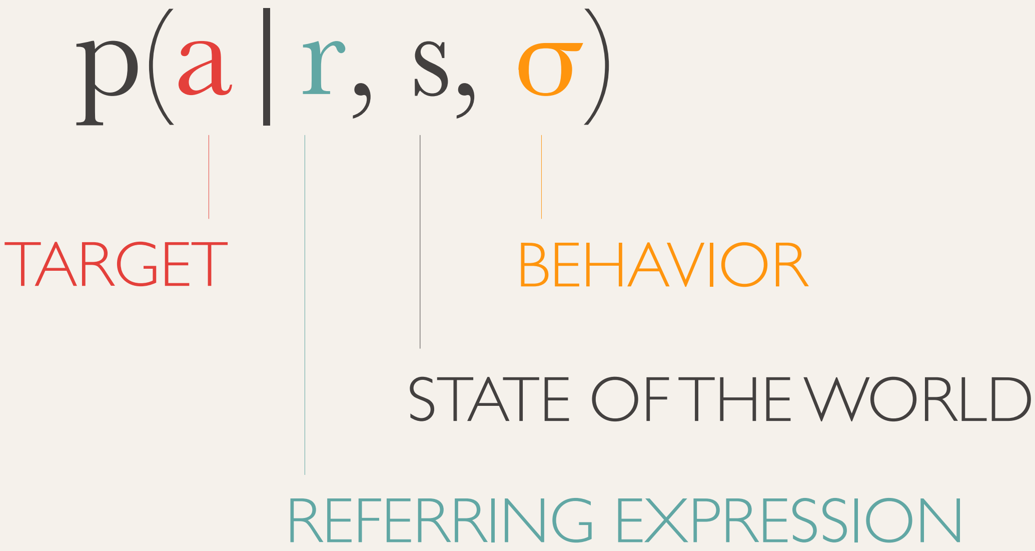
Available in Europe
Waived fee for educational purposes

PART II: LISTENER'S UNDERSTANDING

A MODEL OF LISTENER'S UNDERSTANDING

PROBABILISTIC FRAMEWORK

We want our instructions to have a high degree of success.
For that, we need to maximize this probability



$$p(a \mid r, s, \sigma)$$

TARGET

BEHAVIOR

STATE OF THE WORLD

REFERRING EXPRESSION

PROBABILISTIC FRAMEWORK

We'll split this into two models:

$$p(a \mid r, s, \sigma) \propto \underbrace{p(a \mid r, s)}_{\text{SEMANTIC MODEL (Psem)}} \; \underbrace{p(a \mid \sigma)}_{\text{OBSERVATIONAL MODEL (Pobs)}}$$

The Psem model tells us which RE has a higher chance of success

The Pobs model tells us when we need to give you a new RE

LOG-LINEAR MODELS

Both models are log-linear,
because they are written in this form:

$$p(a \mid r, s) \propto \exp(w_1 f_1(a, r, s) + \ldots + w_n f_n(a, r, s))$$

f_i are called FEATURE FUNCTIONS

w_i are the associated WEIGHTS

We select the features, but the weights
are learned from the training data

SEMANTIC MODEL

EXAMPLE FEATURES FOR Psem

SEMANTIC FEATURES

Is the color of the object mentioned in the RE?
Is the relative position of an object mentioned in the RE?

CONFUSION FEATURES

Is the color of another object mentioned in the instruction?

SALIENCE FEATURES

Is an object visible? Is it in the room?
How visually salient is it?

OBSERVATIONAL MODEL

EXAMPLE FEATURES FOR Pobs

How much closer has the player moved towards an object? Has he entered the same room?

How has the visual salience of an object evolved?
(might indicate a loss of interest)

How much has the angle to an object changed?
(might indicate (dis)interest)

Has the user remained still in the last seconds?
(might indicate confusion)

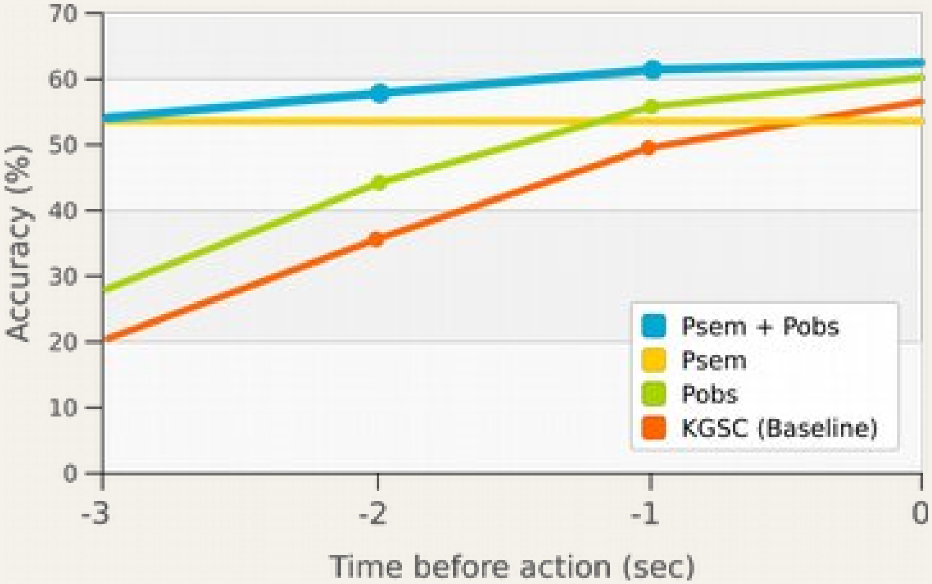
RESULTS

COMBINED MODEL

The combined model outperforms both individual models

The Psem model outperforms Pobs and the baseline early on

The Pobs model improves late accuracy



Predicting the resolution of referring expressions from user behavior
(Engonopoulos, Villalba, Titov & Koller, 2013)

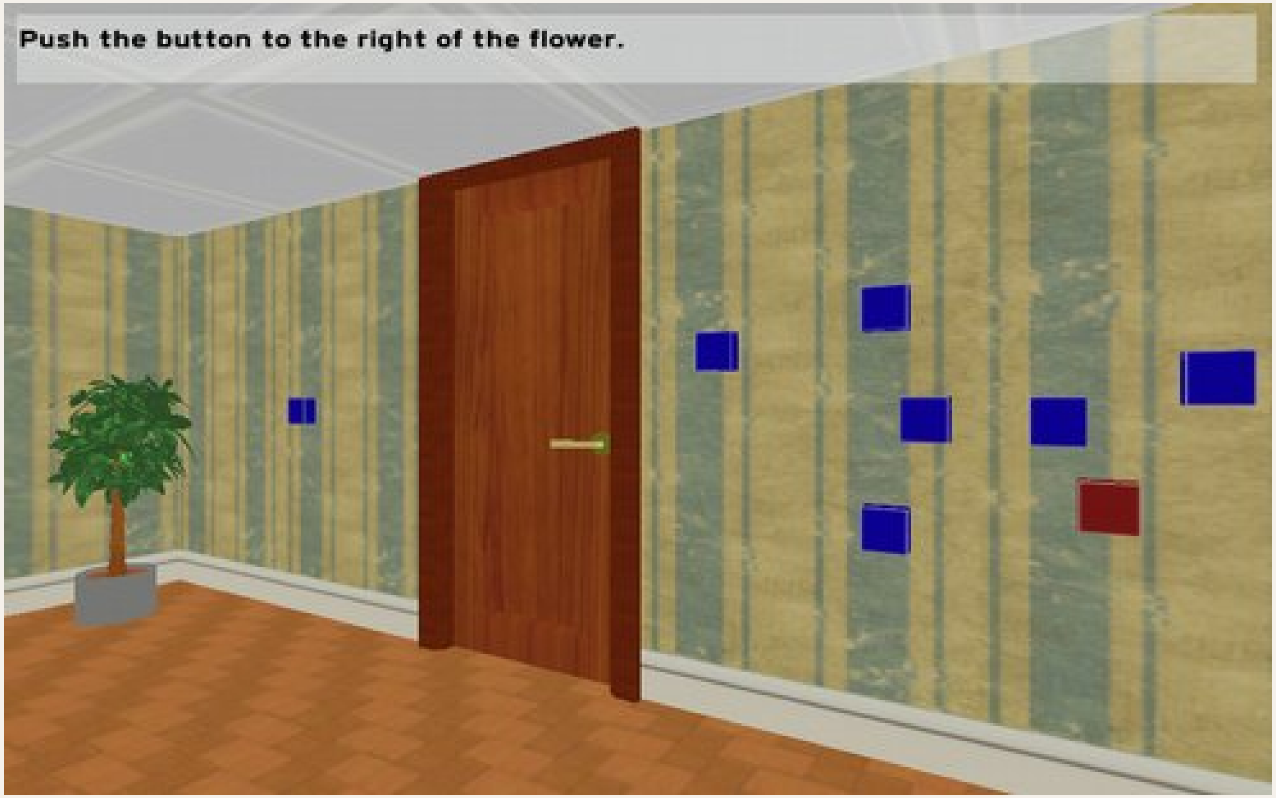
Additional corpora containing eye-tracking recordings collected in 2012

Over 8hs of recorded interactions

Using listener gaze to augment speech generation
in a virtual 3D environment (Staudte, Koller, Garoufi & Crocker, 2012)

EYE-TRACKING MODEL

EXAMPLE FEATURES

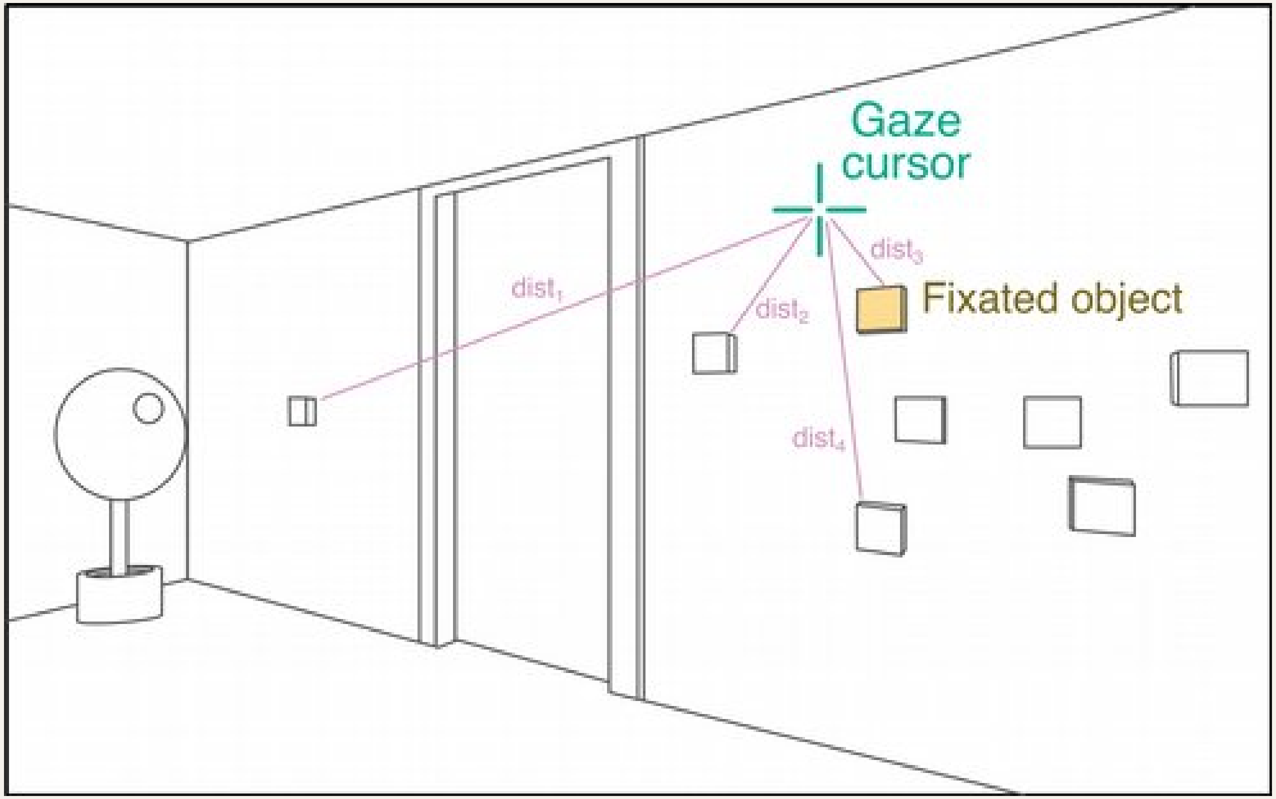


Push the button to the right of the flower.

Has the user seen the object? For how long?
Is the user's gaze fixated in the object?
How close is the user's gaze to the object?

EYE-TRACKING MODEL

EXAMPLE FEATURES



Has the user seen the object? For how long?

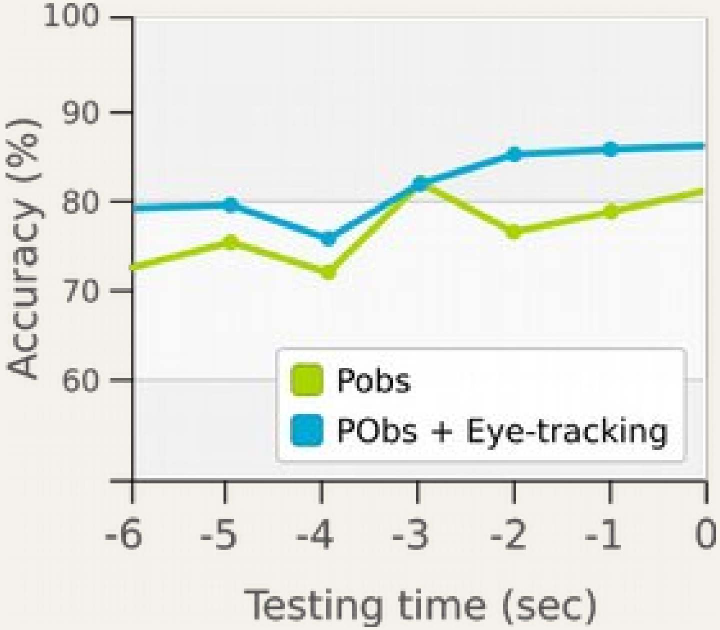
Is the user's gaze fixated in the object?

How close is the user's gaze to the object?

RESULTS

EYE-TRACKING

Adding eye-tracking features improves prediction accuracy on hard scenes



The impact of listener gaze on predicting reference Resolution (Koleva, Villaba, Staudte & Koller, 2015)

PART III: GENERATION

HOW TO CREATE THE PERFECT R.E.

Picture by Wired, via Flickr

REFERRING EXPRESSIONS

We'll define the BEST Referring Expression
as the one with the highest probability
of being correctly understood

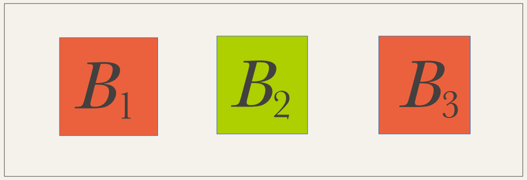
SEMANTICALLY INTERPRETED GRAMMAR

A Semantically Interpreted Grammar (SIG) provides translations between strings and sets via an intermediate grammar

SIG

SEMANTICALLY INTERPRETED GRAMMAR

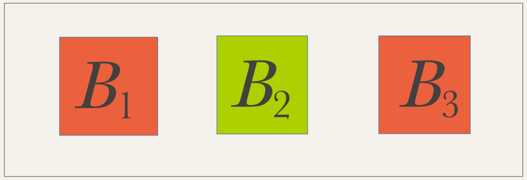
GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · *w1*	$\text{uniq}(R_1) = \text{if } (\mathrm{R}_1 \text{ is singleton}) \text{ then } \mathrm{R}_1 \text{ else } \varnothing$
N → leftof(N, NP)	*w1* · to the left of · *w2*	$\{\text{a} \in R_1 \mid \textit{exists}\ \text{b} \in R_2 \text{ s.t. } (\text{a,b}) \in \lvert\text{left\_of}\rvert\}$
N → green(N)	green · *w1*	$\lvert\text{green}\rvert \cap R_1$
N → red(N)	red · *w1*	$\lvert\text{red}\rvert \cap R_1$
N → button	button	$\lvert\text{button}\rvert$

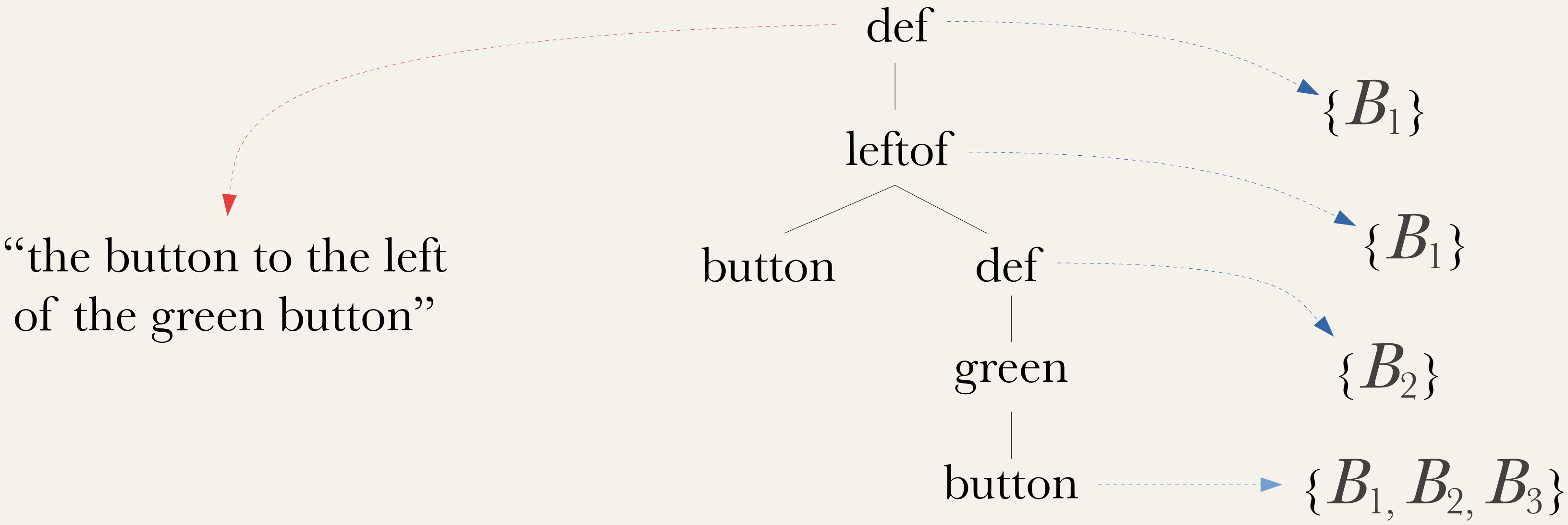


SIG

SEMANTICALLY INTERPRETED GRAMMAR

GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · *w1*	uniq(R_1)= if (R_1 is singleton) then R_1 else ∅
N → leftof(N, NP)	*w1* · to the left of · *w2*	$\{a \in R_1 \mid \textit{exists}\ b \in R_2 \text{ s.t. } (a,b) \in \lvert\text{left\_of}\rvert\}$
N → green(N)	green · *w1*	$\lvert\text{green}\rvert \cap R_1$
N → red(N)	red · *w1*	$\lvert\text{red}\rvert \cap R_1$
N → button	button	$\lvert\text{button}\rvert$





"the button to the left of the green button"

def → $\{B_1\}$

leftof → $\{B_1\}$

button

def → $\{B_2\}$

green

button → $\{B_1, B_2, B_3\}$

SIG

CHART BASED GENERATION

All possible REs are stored in a Chart, eliminating backtracking and preventing a combinatorial explosion

Each possible RE can be then scored, and we pick the best one

Generation effective referring expressions using charts (Engonopoulos & Koller, 2014)

SIG

CHART BASED GENERATION

We'll judge each RE based on our probabilistic model

$$p(a \mid r, s, \sigma) \propto \underbrace{p(a \mid r, s)}_{\text{Psem}}\, p(a \mid \sigma)$$

TARGET

REFERRING EXPRESSION

STATE OF THE WORLD

BEHAVIOR

Generation effective referring expressions
using charts (Engonopoulos & Koller, 2014)

SIG

CHART BASED GENERATION

We'll judge each RE based on our probabilistic model

$$p(a \mid r, s, \sigma) \propto \underbrace{p(a \mid r, s)}_{\text{Psem}} \, p(a \mid \sigma)$$

Generation effective referring expressions
using charts (Engonopoulos & Koller, 2014)

PART IV: MISUNDERSTANDINGS

HOW TO DETECT AND CORRECT MISTAKES

Picture by Filippo Monteforte/AFP/Getty Images, via The Atlantic

DETECTING MISUNDERSTANDINGS

Our Pobs model gives us a good approximation of which object has captured the user's interest.

$$p(a \mid r, s, \sigma) \propto p(a \mid r, s)\, \underbrace{p(a \mid \sigma)}_{\text{Pobs}}$$

TARGET

BEHAVIOR

STATE OF THE WORLD

REFERRING EXPRESSION

DETECTING MISUNDERSTANDINGS

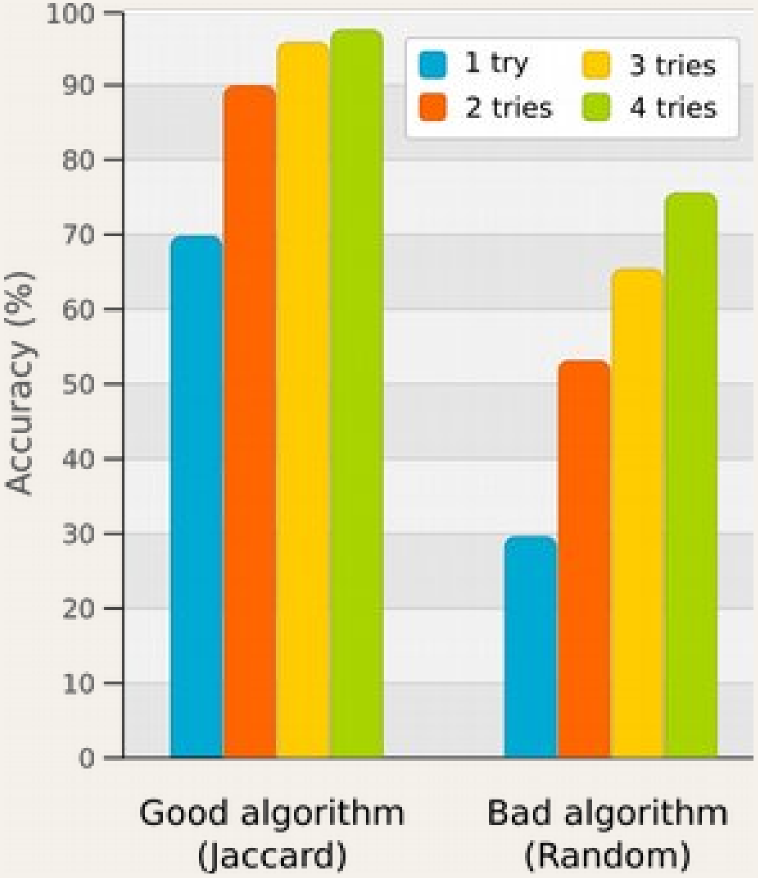
Our Pobs model gives us a good approximation of which object has captured the user's interest.

$$p(a \mid r, s, \sigma) \propto p(a \mid r, s) \underbrace{p(a \mid \sigma)}_{\text{Pobs}}$$

If the object a with the highest probability is different from *our* intended target, the user misunderstood our RE!

DETECTING MISUNDERSTANDINGS

A single correction can drastically improve accuracy. Giving just one new RE might be all we need



Interpreting NL Instructions using language, vision, and behavior (Benotti, Lau & Villalba, 2014)

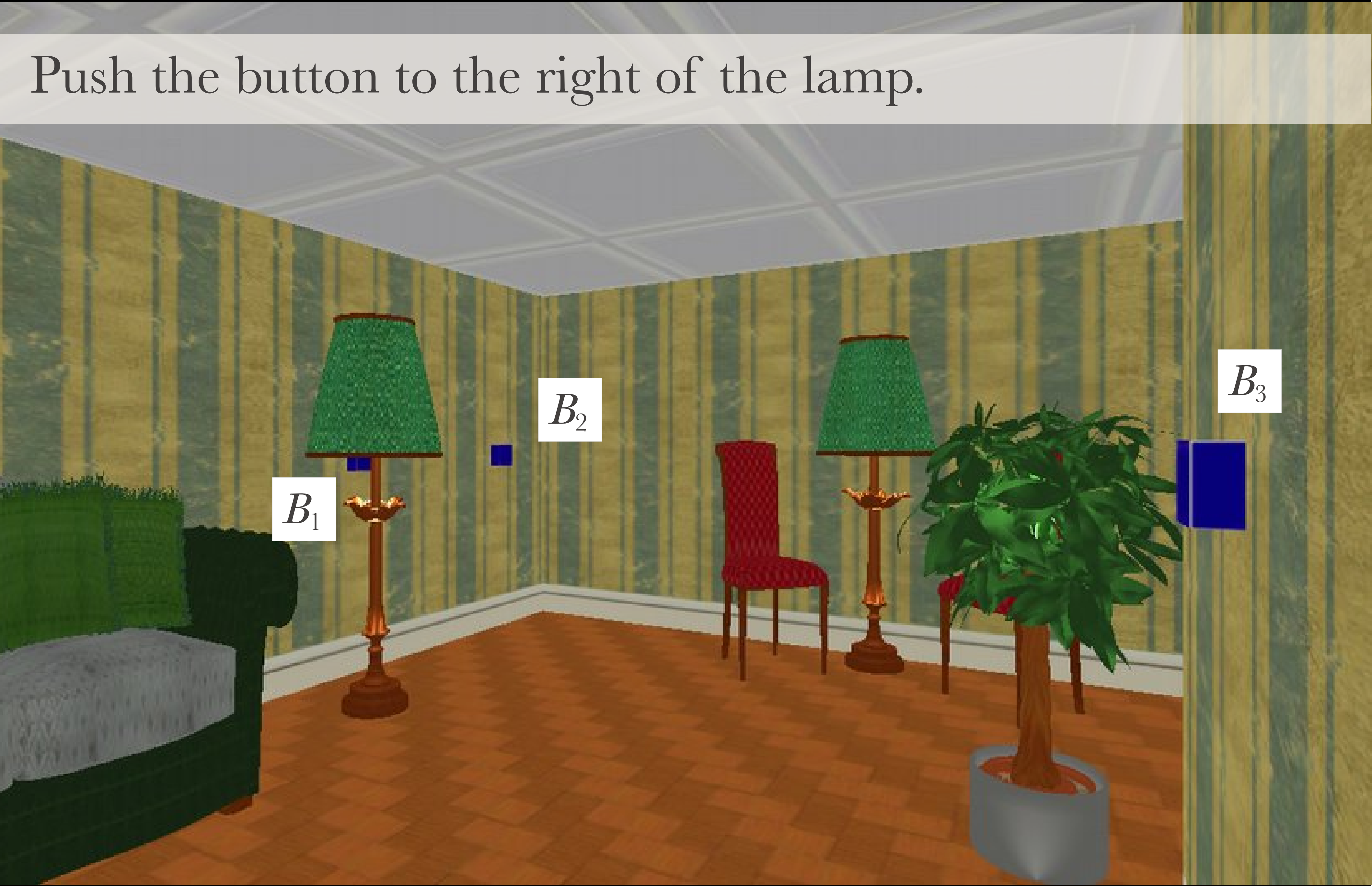
DETECTING MISUNDERSTANDINGS

We defined a referring expression as

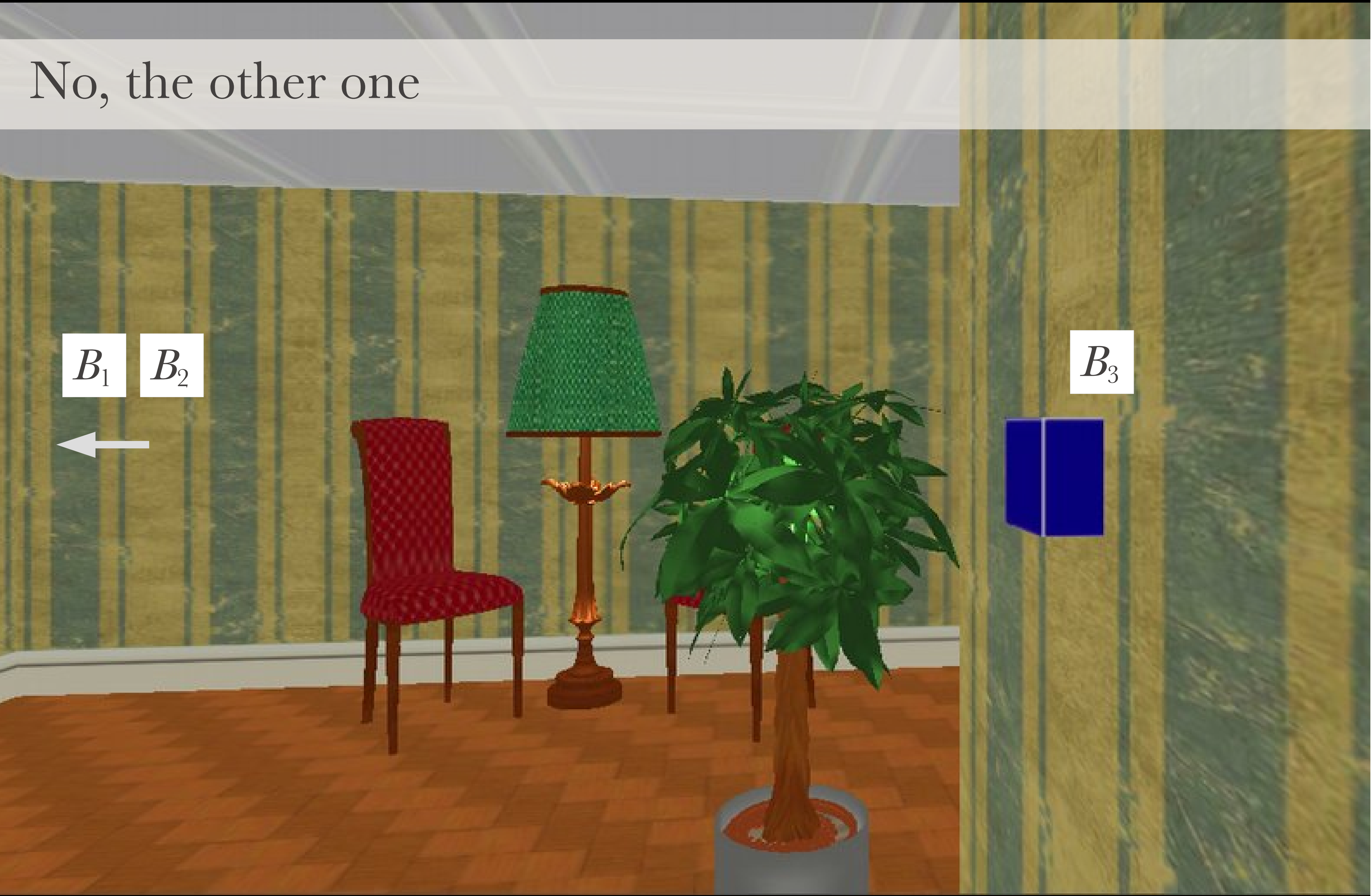
A NOUN PHRASE THAT IDENTIFIES UNIQUELY A CERTAIN OBJECT WITHIN A SCENE

We rarely make those

Push the button to the right of the lamp.



No, the other one



CORRECTING MISUNDERSTANDINGS

CONTEXT SET

Given an intended target a_{int},
the set of objects $\{a_1 \ldots a_n\}$ such that

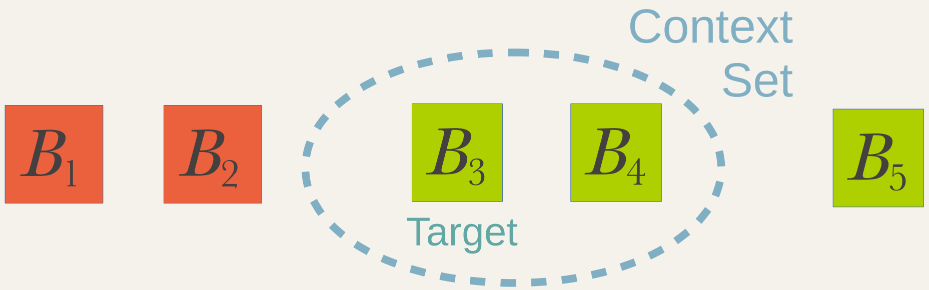
$$p(a_i \mid r,s,\sigma) \geq p(a_{int} \mid r,s,\sigma)$$

will be defined as the CONTEXT SET

CORRECTING MISUNDERSTANDINGS

GENERATION WITH CONTEXT SET

Strategy 1: globally unique REs



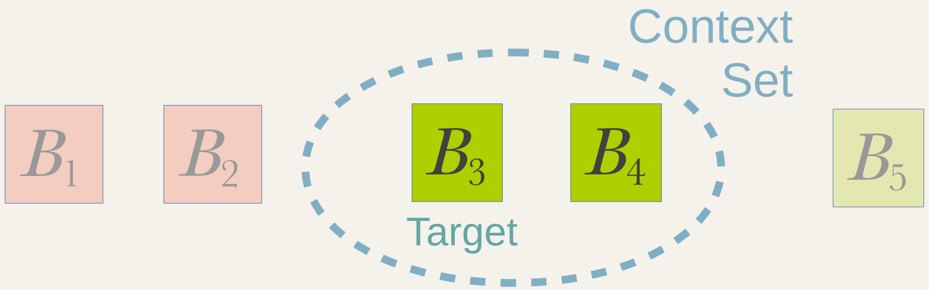
The button to the right of the red button to the right of the red button

CORRECTING MISUNDERSTANDINGS

GENERATION WITH CONTEXT SET

Strategy 1: globally unique REs

Strategy 2: objects outside the CS are irrelevant



The leftmost button

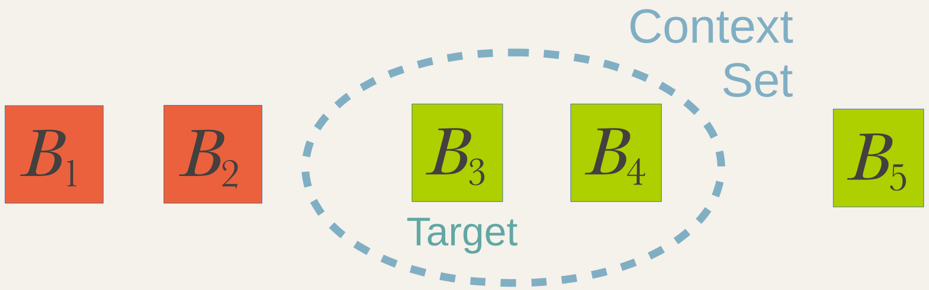
CORRECTING MISUNDERSTANDINGS

GENERATION WITH CONTEXT SET

Strategy 1: globally unique REs

Strategy 2: objects outside the CS are irrelevant

Strategy 3: We only refer to the intended target in relation to other objects in the CS



The button to the left of the green button
to the left of the green button

CORRECTING MISUNDERSTANDINGS

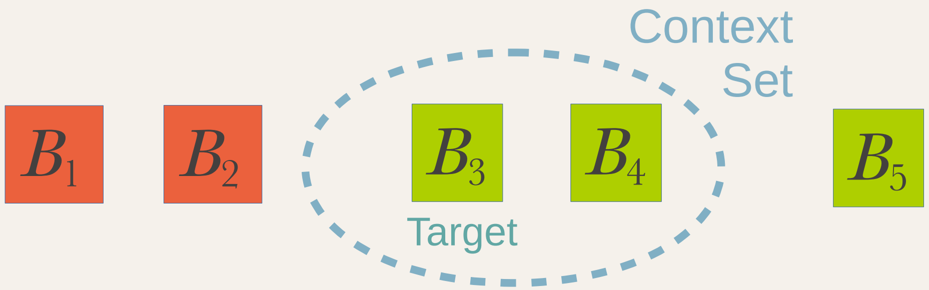
GENERATION WITH CONTEXT SET

Strategy 1: globally unique REs

Strategy 2: objects outside the CS are irrelevant

Strategy 3: We only refer to the intended target in relation to other objects in the CS

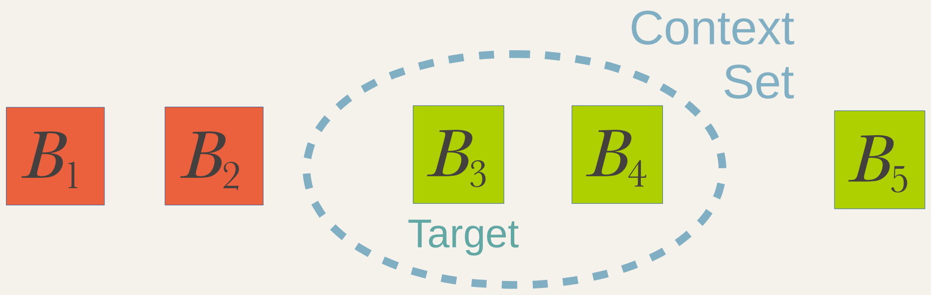
Strategy 4: The RE must be unique within the CS

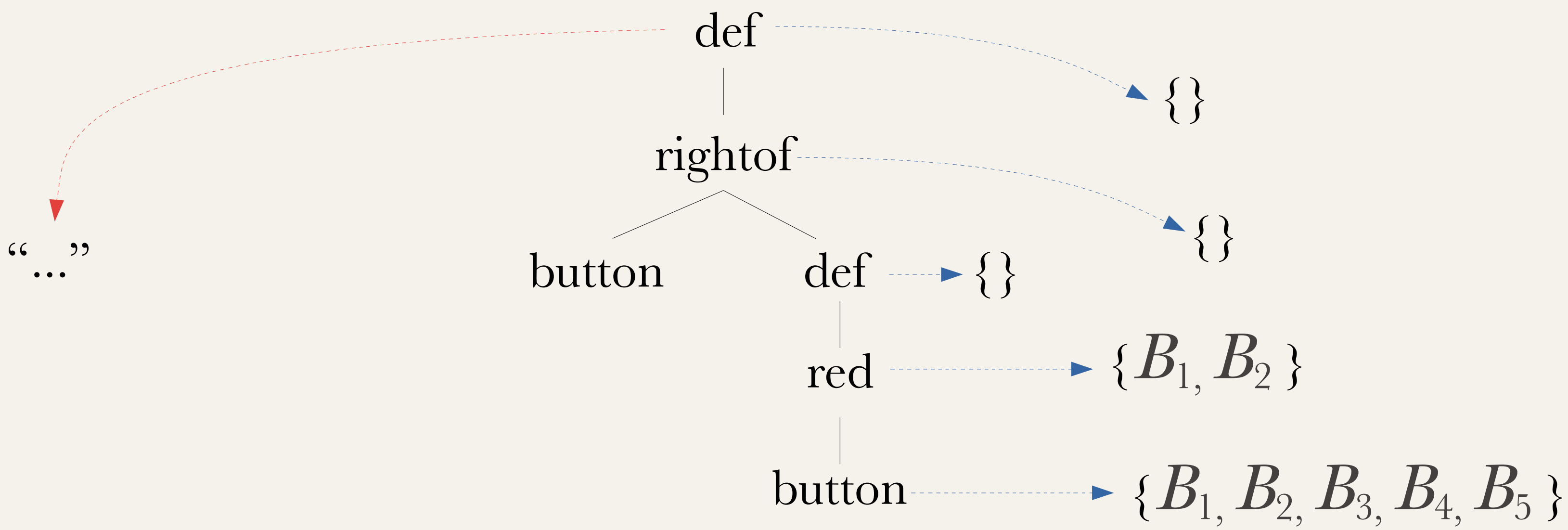


The button to the right of the red button

CORRECTING MISUNDERSTANDINGS

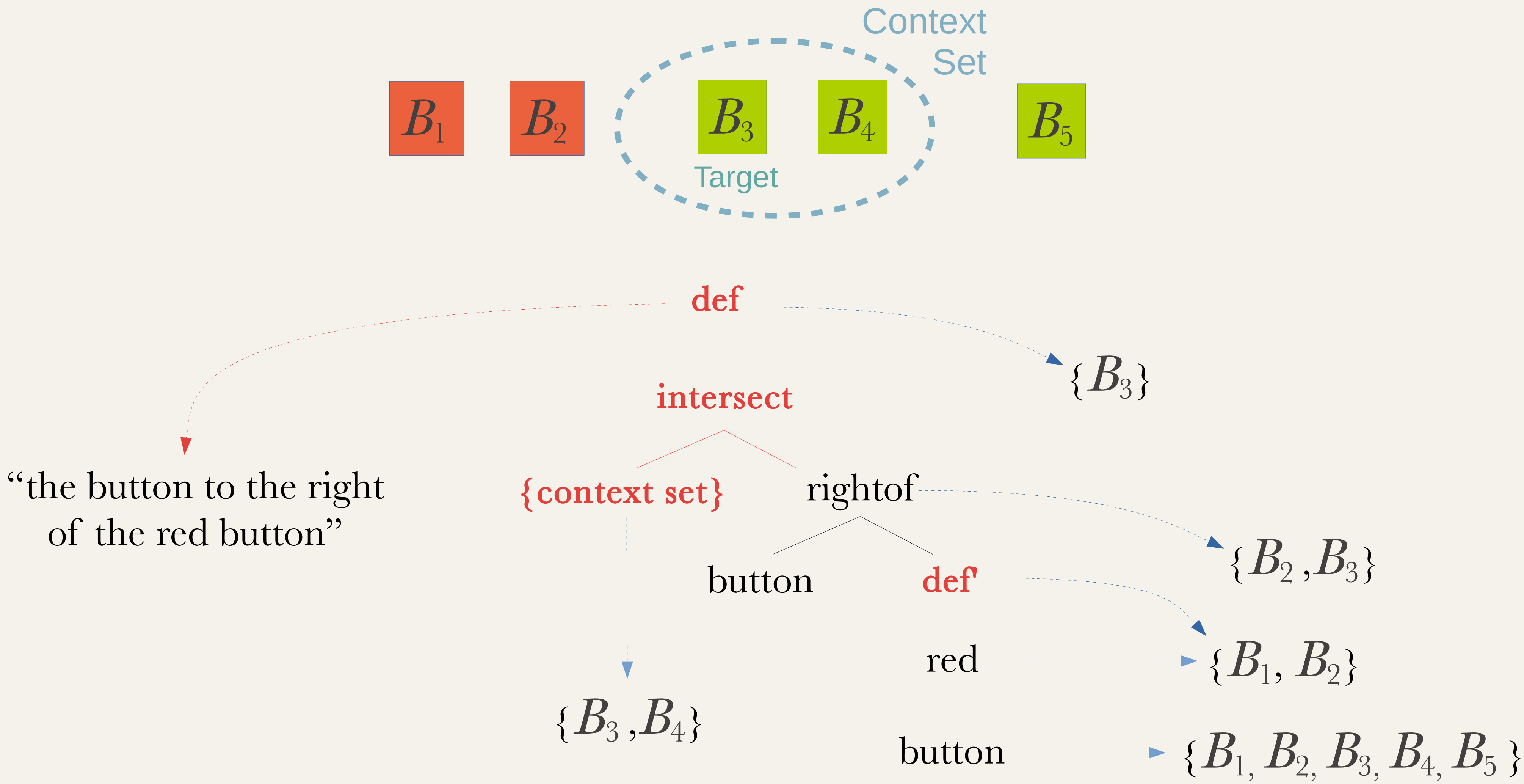
GENERATION WITH CONTEXT SET





CORRECTING MISUNDERSTANDINGS

GENERATION WITH CONTEXT SET



Context Set

B_1 B_2 B_3 B_4 B_5

Target

def

intersect

{context set}

rightof

button

def'

red

button

$\{B_3\}$

$\{B_2, B_3\}$

$\{B_1, B_2\}$

$\{B_1, B_2, B_3, B_4, B_5\}$

$\{B_3, B_4\}$

"the button to the right of the red button"

CORRECTING MISUNDERSTANDINGS

GENERATION WITH CONTEXT SET

GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · *w1*	uniq(R_1)= if (R$_1$ is singleton) then R$_1$ else ∅
N → leftof(N, NP)	*w1* · to the left of · *w2*	$\{a \in R_1 \mid \textit{exists}\ b \in R_2 \text{ s.t. } (a,b) \in \lvert \text{left\_of} \rvert\}$
N → green(N)	green · *w1*	$\lvert \text{green} \rvert \cap R_1$
N → red(N)	red · *w1*	$\lvert \text{red} \rvert \cap R_1$
N → button	button	$\lvert \text{button} \rvert$

CORRECTING MISUNDERSTANDINGS

GENERATION WITH CONTEXT SET

GRAMMAR RULE	STRING	DENOTATION
NP → def'(N)	the · *w1*	member (R_1)= R_1
N → leftof(N, NP)	*w1* · to the left of · *w2*	$\{a \in R_1 \mid \textit{exists}\ b \in R_2 \text{ s.t. } (a,b) \in \lvert\text{left\_of}\rvert \}$
N → green(N)	green · *w1*	$\lvert\text{green}\rvert \cap R_1$
N → red(N)	red · *w1*	$\lvert\text{red}\rvert \cap R_1$
N → button	button	$\lvert\text{button}\rvert$
NPCS → def(N)	the · *w1*	$\text{uniq}(\lvert\text{context set}\rvert \cap R_1)$

FUTURE WORK

WHERE DO WE GO FROM HERE?

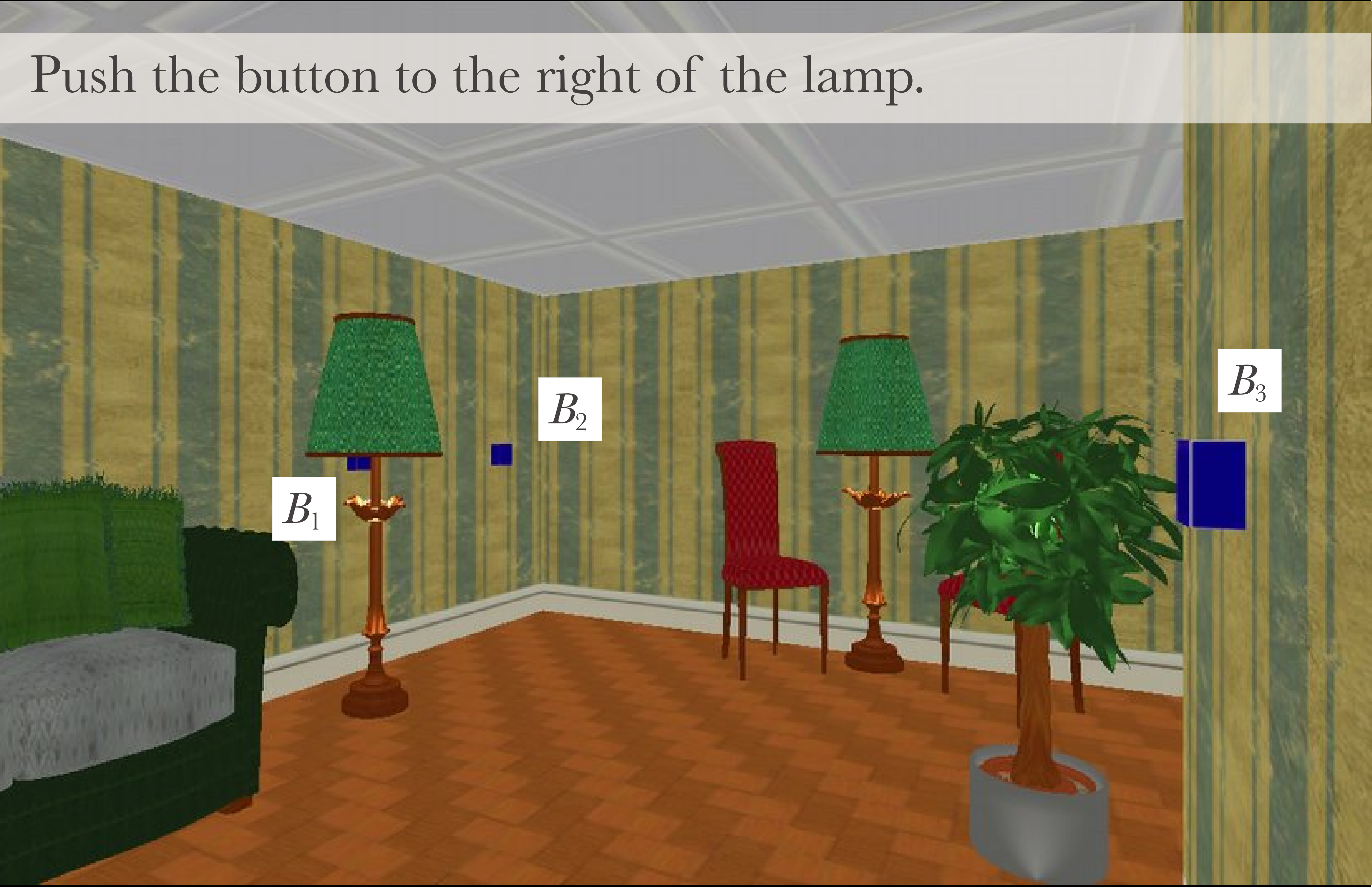
Picture by Jen Scheer, via Flickr

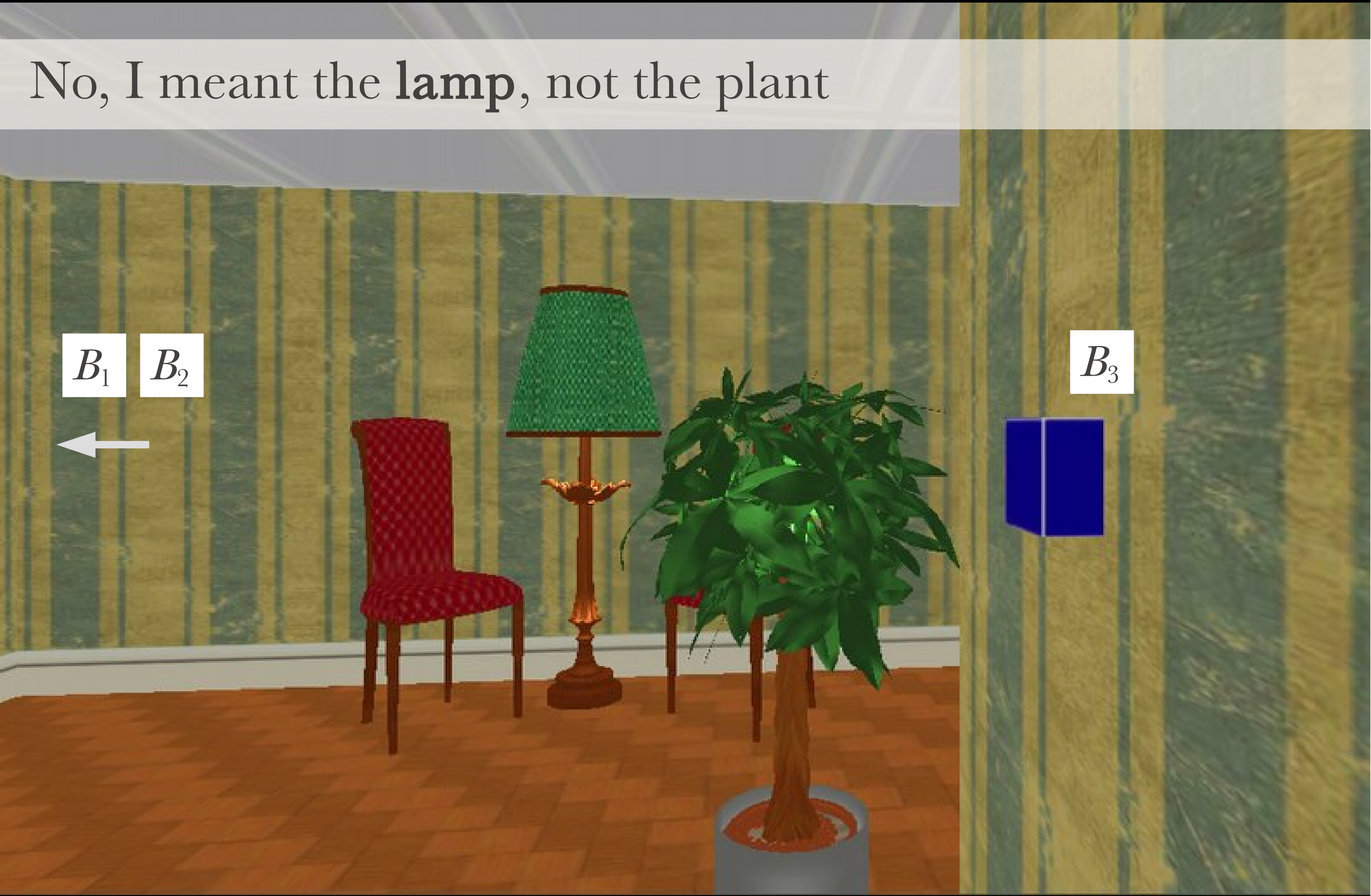
FUTURE WORK

CONTRASTIVE REs

Contrastive REs are vital to keep users from making (possibly costly) mistakes

Push the button to the right of the lamp.





No, I meant the **lamp**, not the plant

Part I

Instructions in a virtual environment

Part II

A model of listener's understanding

Part III

Generating the best RE

Part IV

Dealing with misunderstandings

Future work

QUESTIONS?

THANK YOU FOR YOUR ATTENTION