Dialogue Semantics and Pragmatics

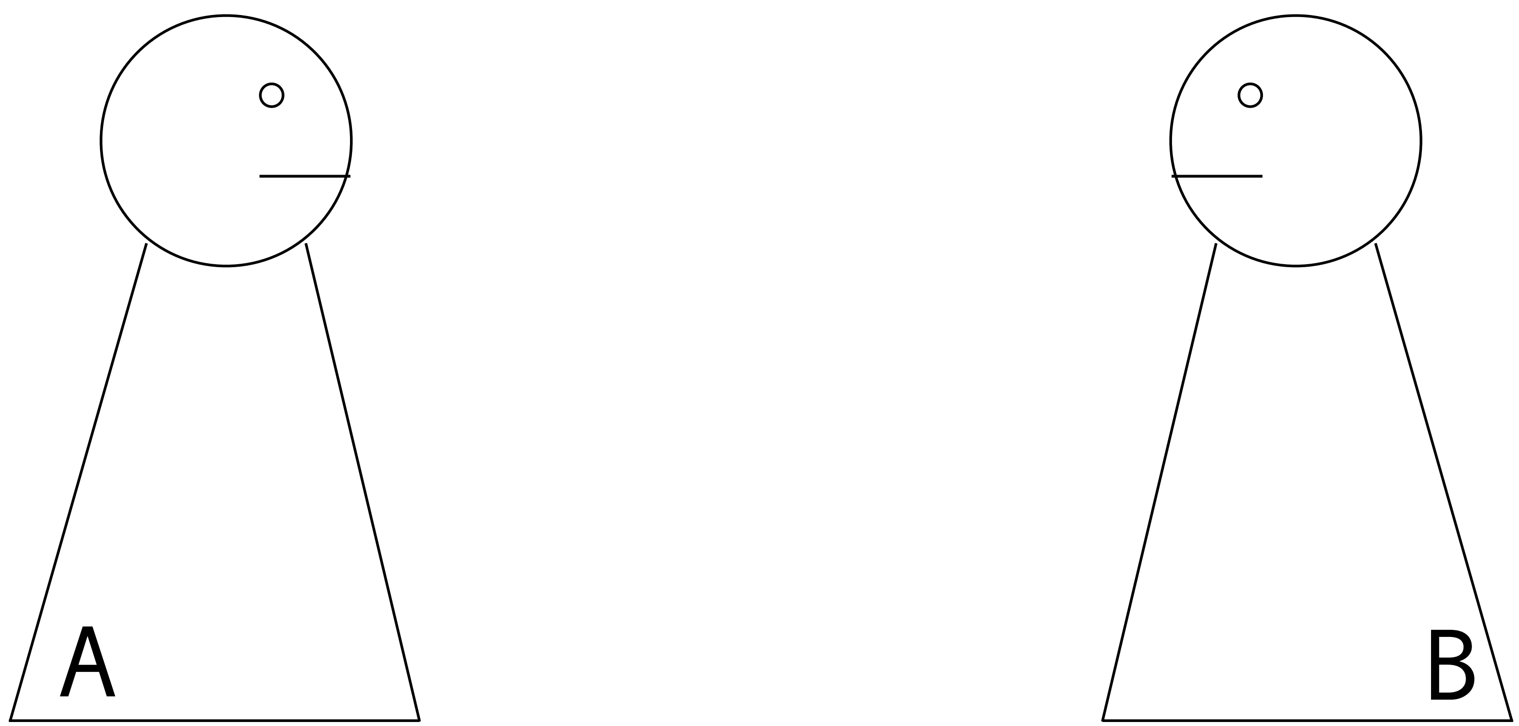
A Tutorial at the ESSENCE Fall School 2014
Part II

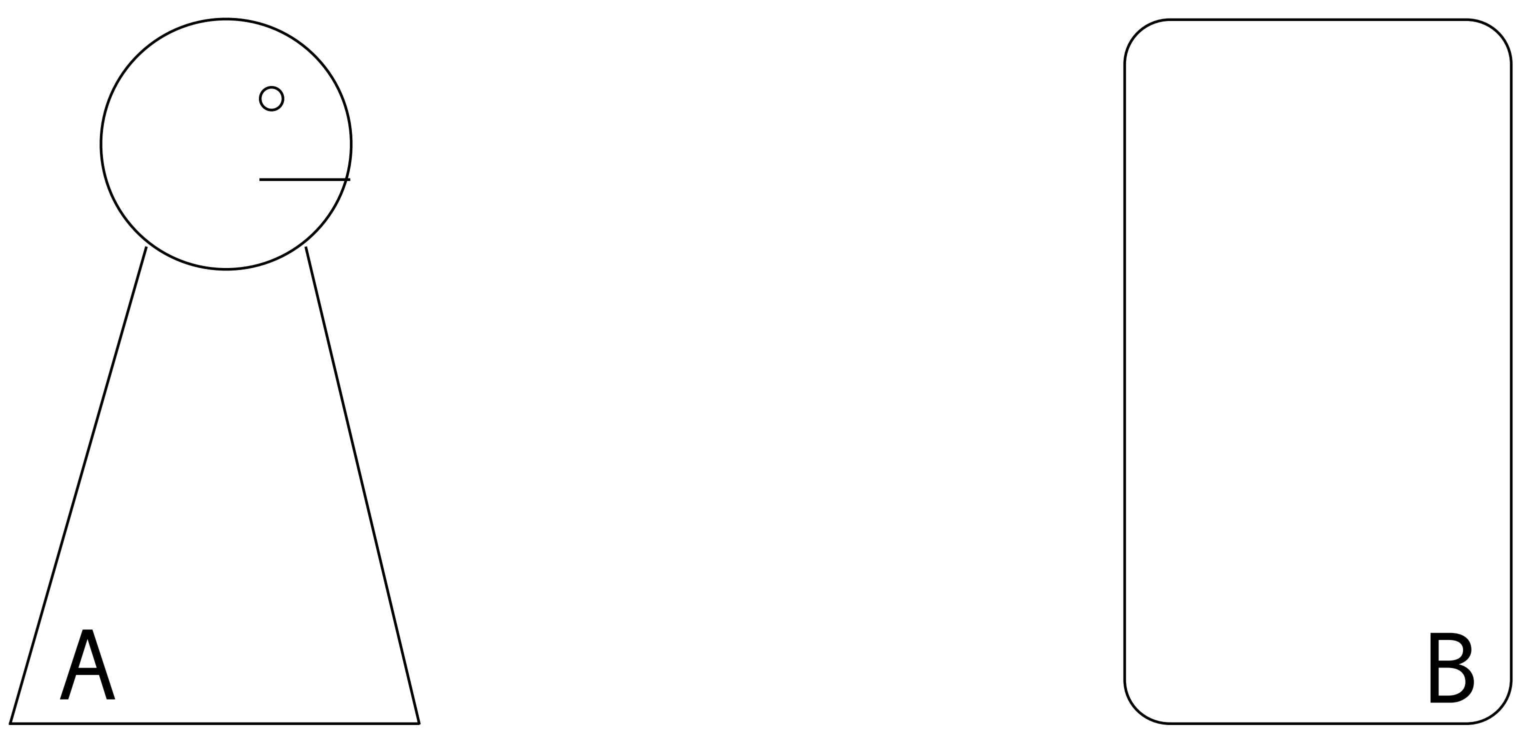
David Schlangen
Universität Bielefeld, Germany
david.schlangen@uni-bielefeld.de

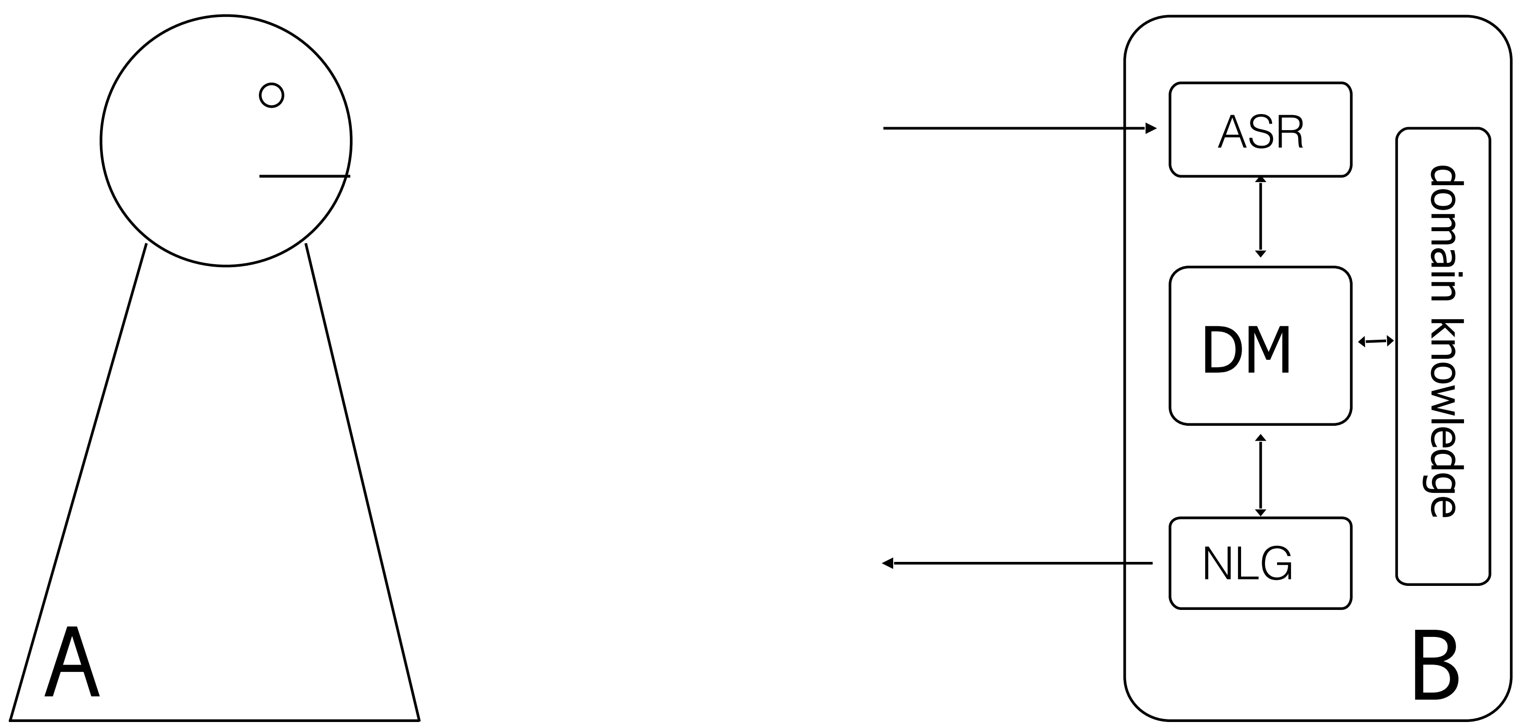
Overview

- **Part I: Foundations**
  - coordination, convention
  - communicative intentions
  - non-conventional meaning
  - grounding
  - turn-taking
  - disfluencies

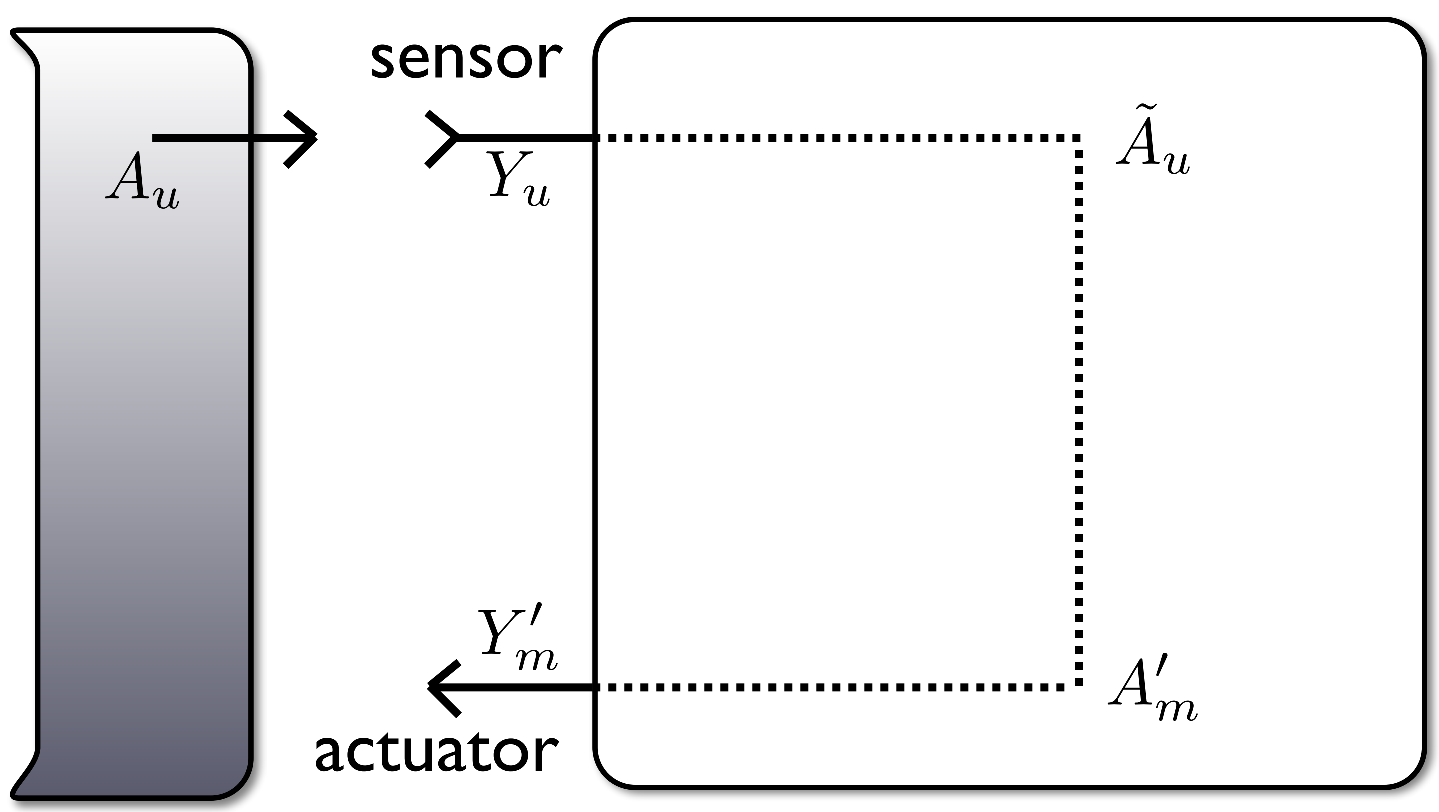
- **Part II: Computational Models**
  - approaches to dialogue modelling
  - incremental processing, turn-taking
  - an example: grounded semantics

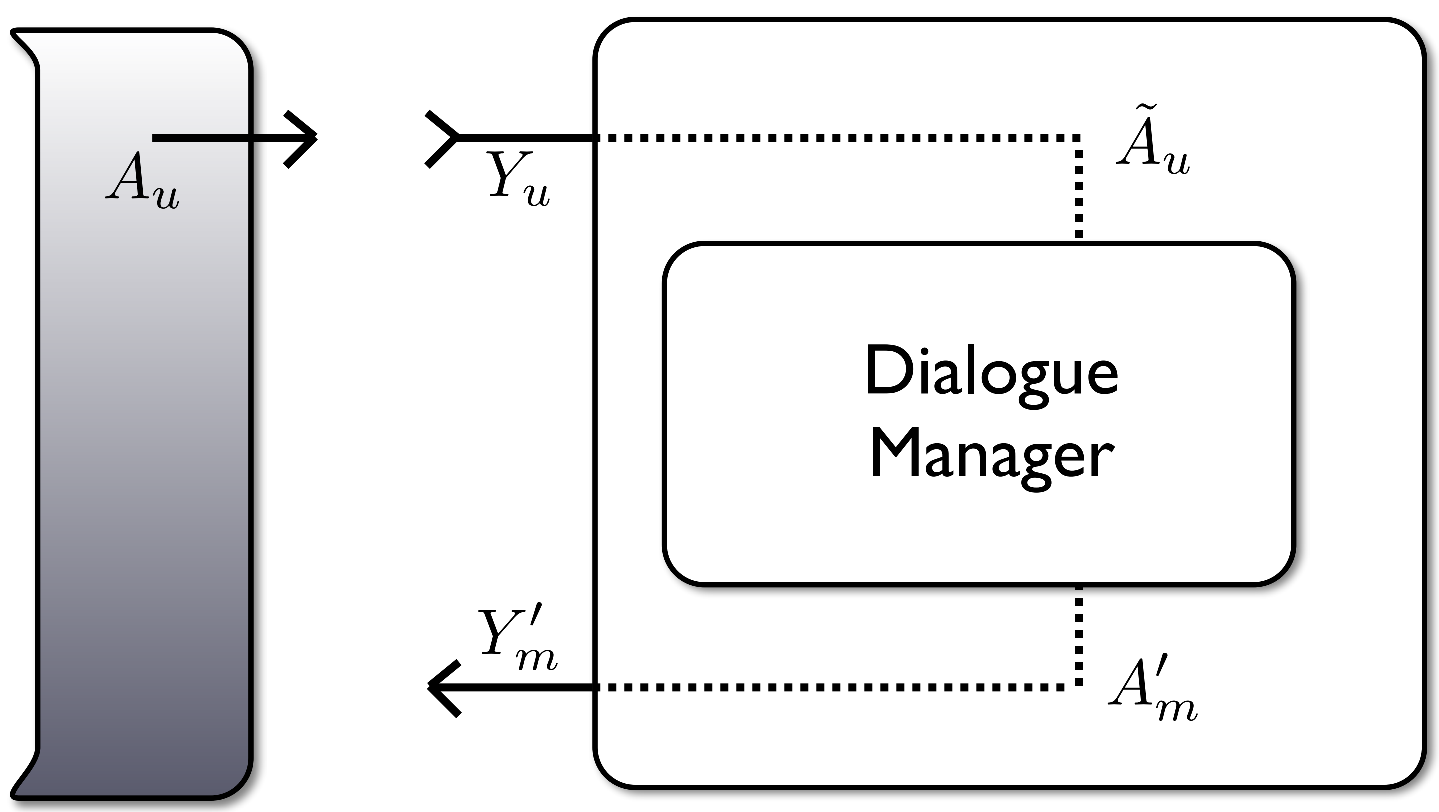

A
B


A
B

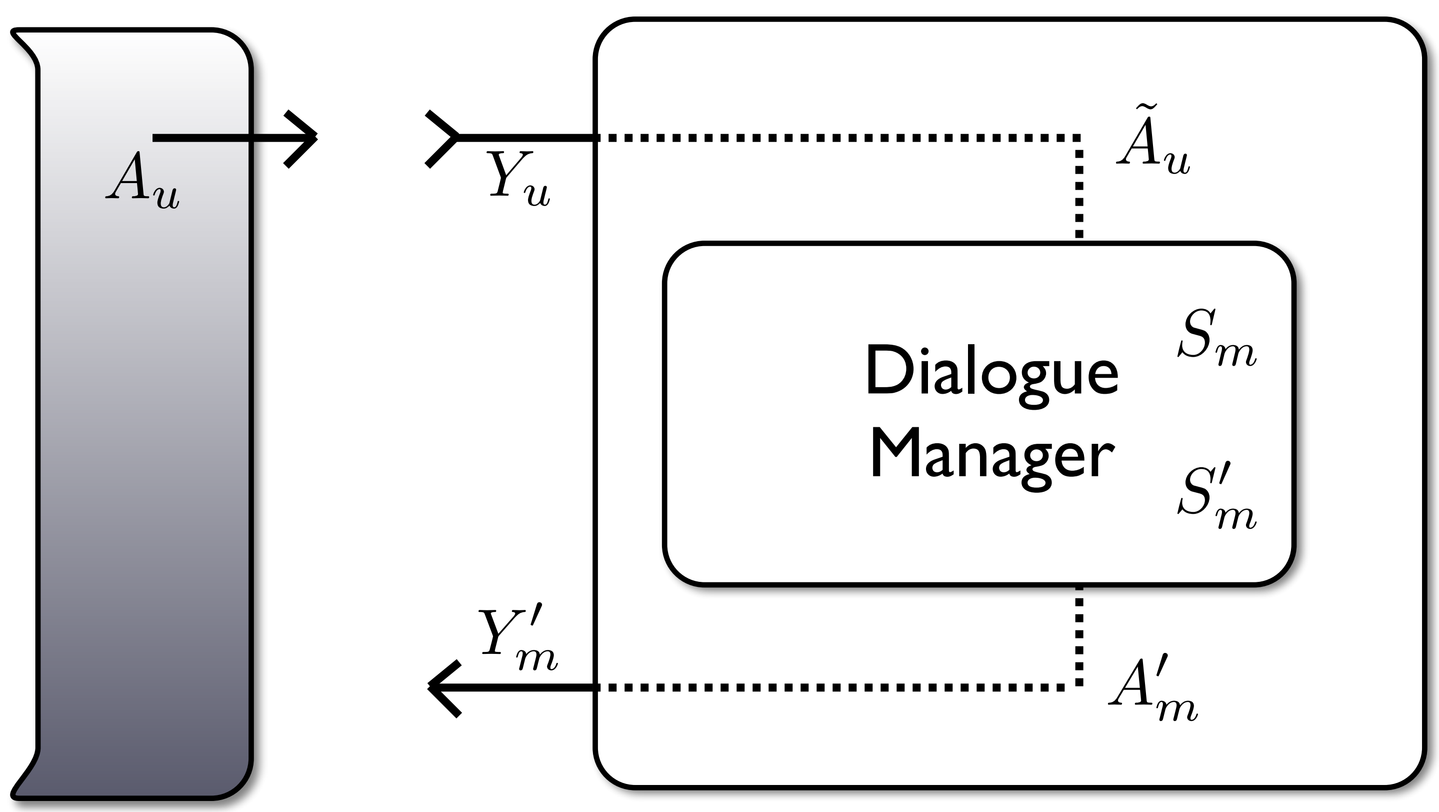


A
ASR
DM
NLG
domain knowledge
B


sensor
A_u
Y_u
$\tilde{A}_u$
Y'_m
A'_m
actuator

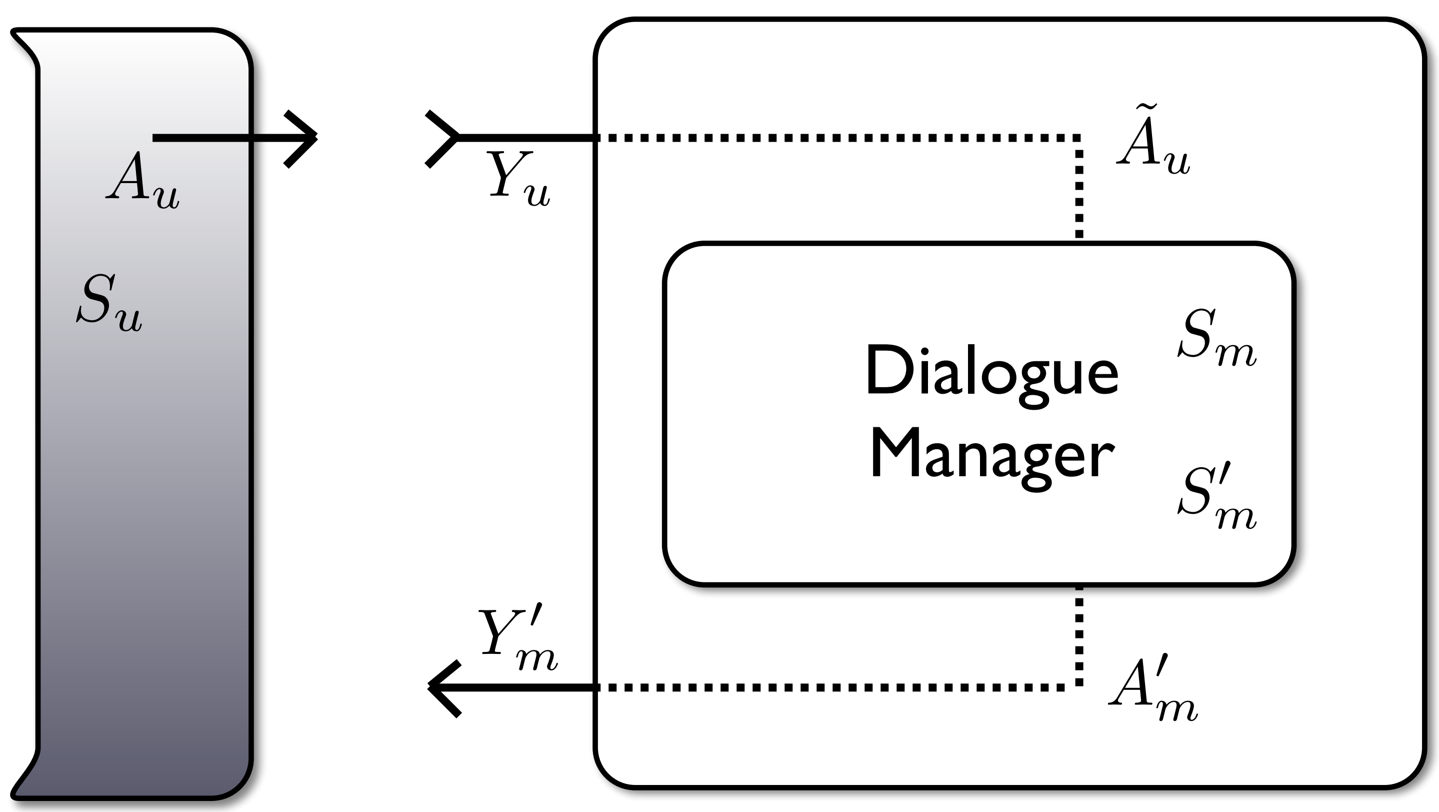


A_u
Y_u
$\tilde{A}_u$
Dialogue Manager
Y'_m
A'_m



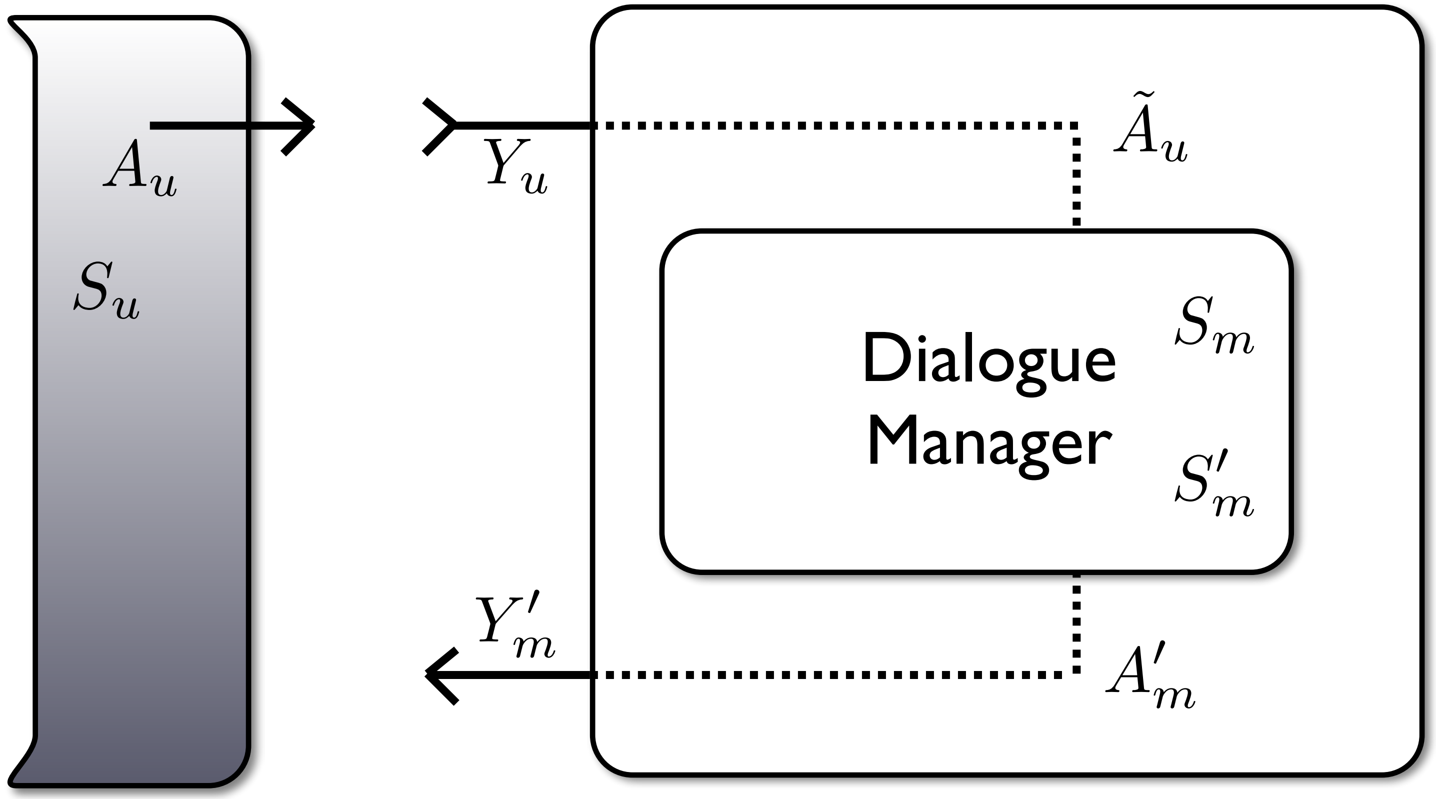
$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- -- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000) (Schlangen 2003)
- BDI (Allen 1995; Perrault & Allen 1980)



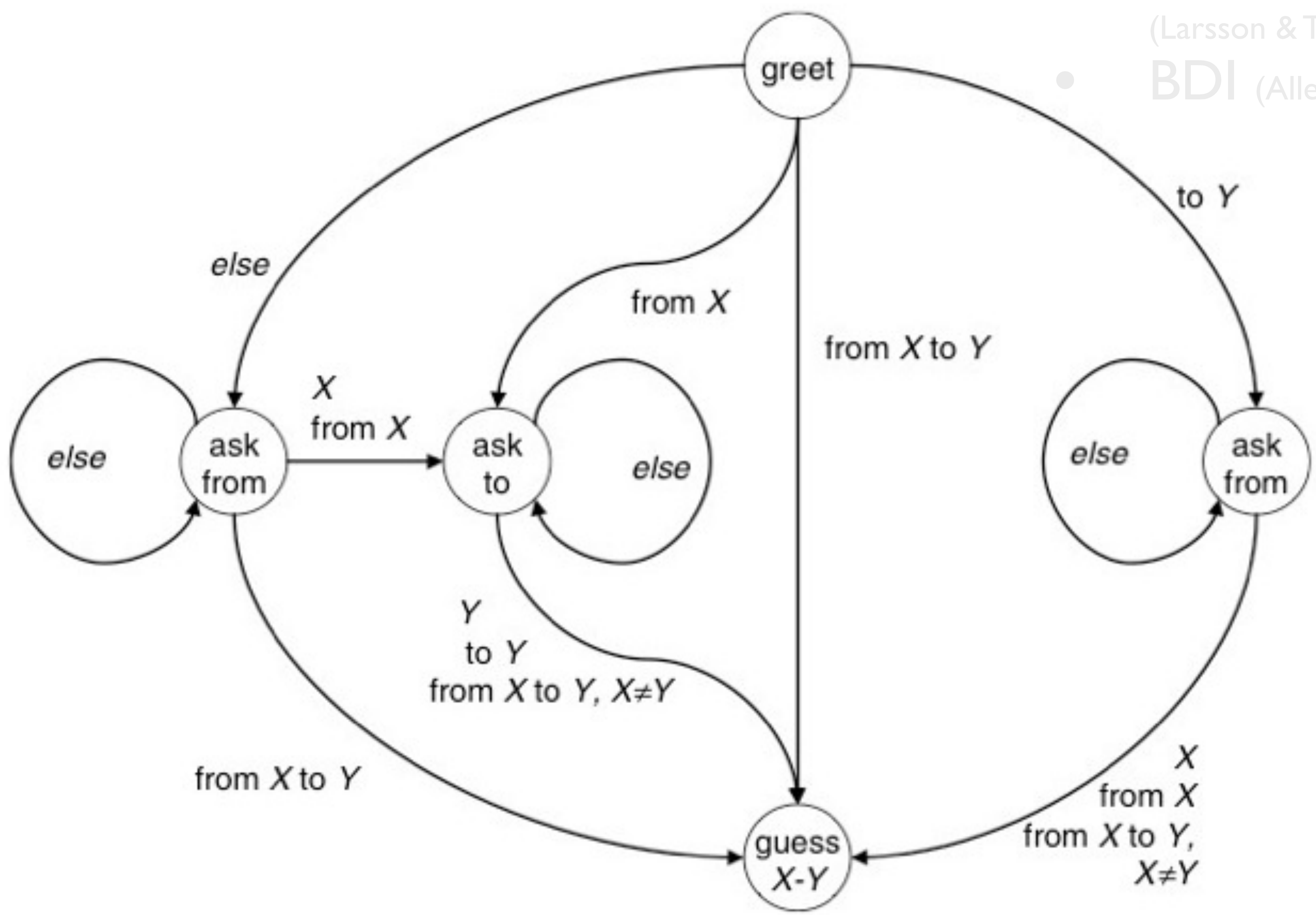
$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- -- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000)(Schlangen 2003)
- BDI (Allen 1995; Perrault & Allen 1980)

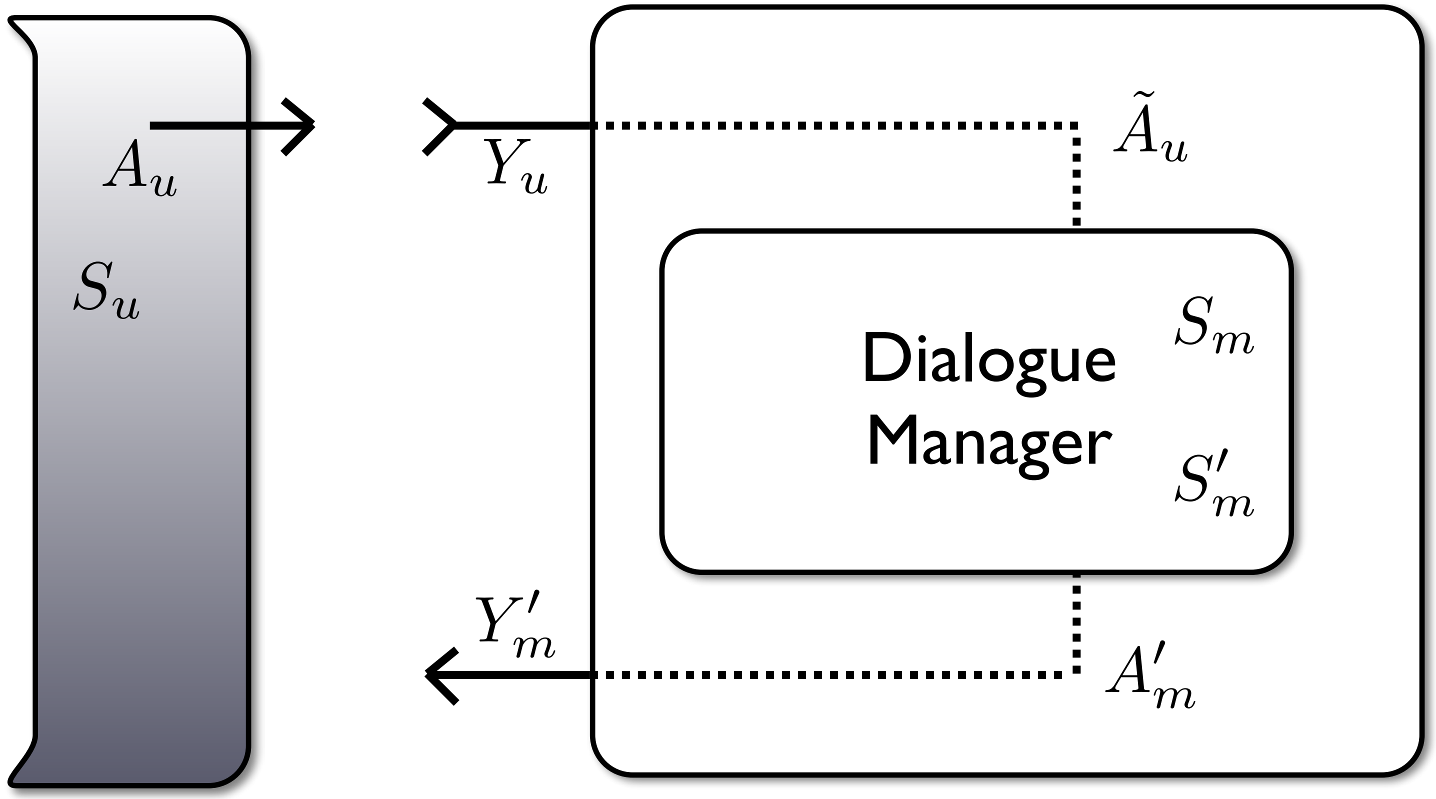


$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- -- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000)(Schlangen 2003)
- BDI (Allen 1995; Perrault & Allen 1980)



aus (Williams & Young 2007)



$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- -- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000) (Schlangen 2003)
- BDI (Allen 1995; Perrault & Allen 1980)

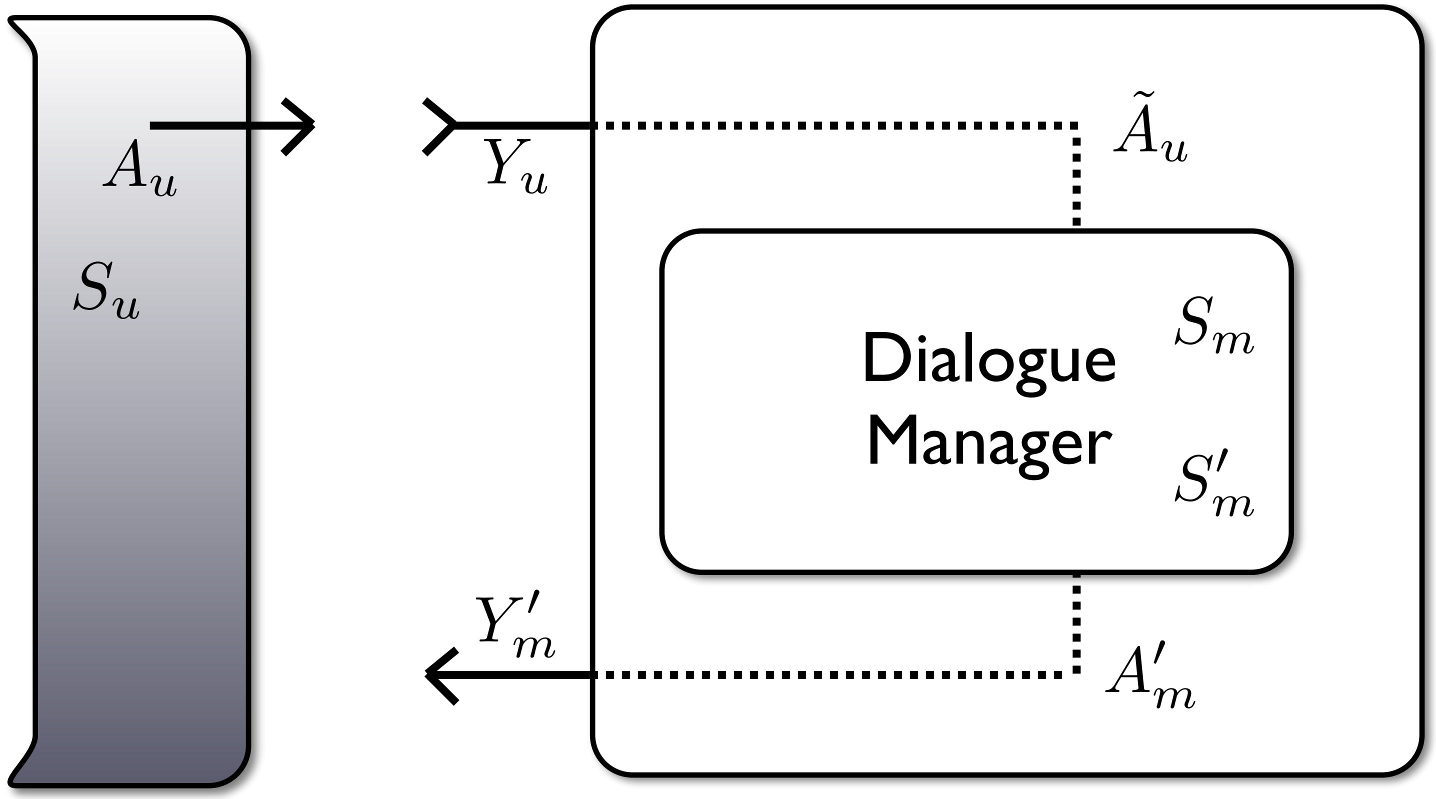
State: *Beliefs*, *Desires*, *Intentions*
Update: inference rules
Decisions: planning

"Can you give me a list of flights to Berlin?"

```
S.REQUEST(S,H,InformIf(H,S,CanDo(H,Give(H,S,List))))
B(H,W(S,InformIf(H,S,CanDo(H,Give(H,S,List)))))
B(H,W(S,KnowIf(H,S,CanDo(H,Give(H,S,List)))))
B(H,W(S,CanDo(H,Give(H,S,List))))
B(H,W(S,Give(H,S,List)))
REQUEST(H,S,Give(H,S,List))
W(H,Give(H,S,List))

W(S,Give(S,H,List))
Action:  Inform(S,H,P)
Effect:  Know(H,P)
```

"There are three flights, one at ..."

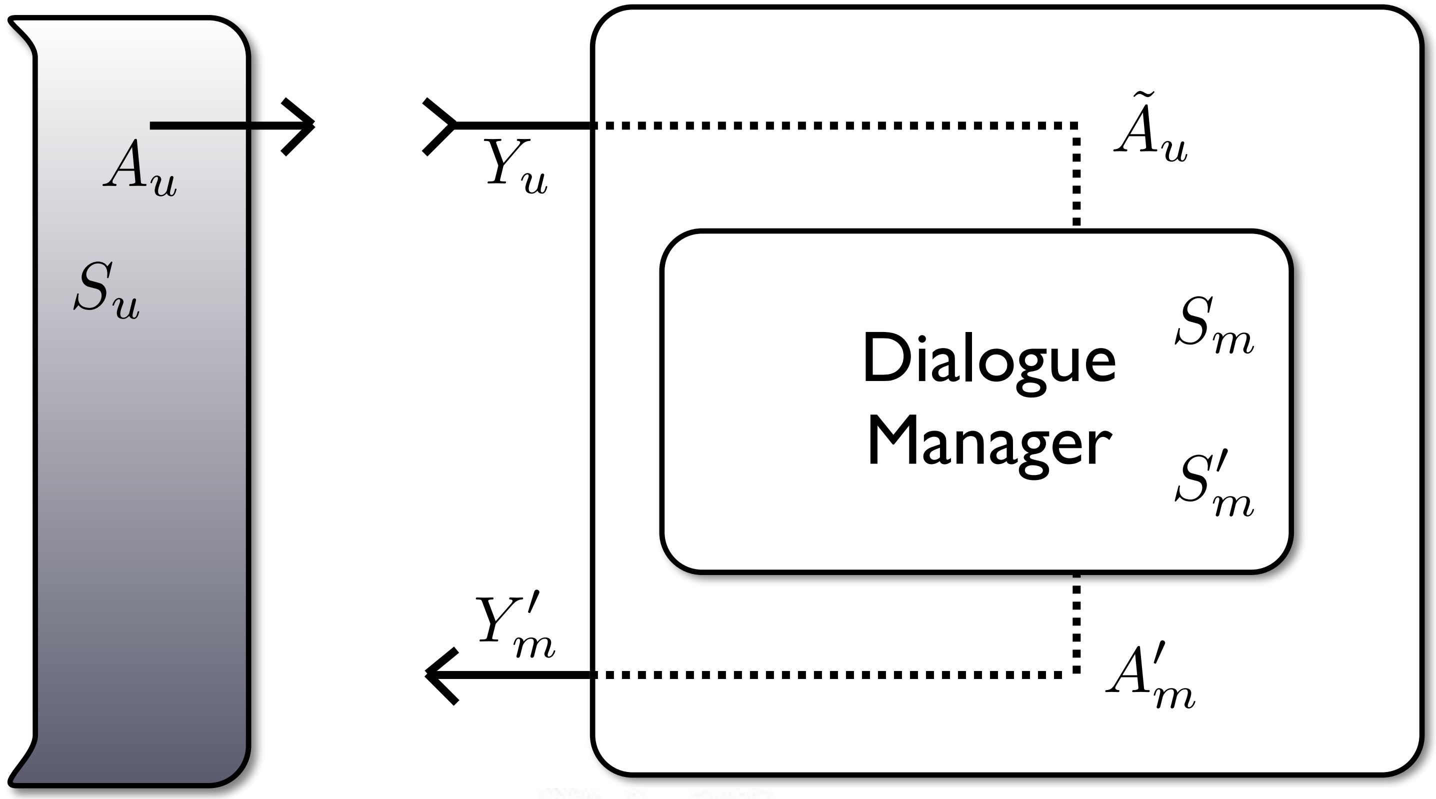


$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- -- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000)
- BDI (Allen 1995; Perrault & Allen 1980)

$$\begin{bmatrix} \text{PRIVATE} & : & \begin{bmatrix} \text{AGENDA} & : & \text{Stack(Action)} \\ \text{PLAN} & : & \text{Stack(Action)} \\ \text{BEL} & : & \text{Set(Prop)} \end{bmatrix} \\ \text{SHARED} & : & \begin{bmatrix} \text{COM} & : & \text{Set(Prop)} \\ \text{QUD} & : & \text{Stack(Question)} \\ \text{LU} & : & \begin{bmatrix} \text{SPEAKER} & : & \text{Participant} \\ \text{MOVE} & : & \text{Move} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(Larsson 2002)



$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- Conversational Scoreboard (Larsson & Traum 2000)
- BDI (Allen 1995; Perrault & Allen 1980)

"price information, please"

getLatestMove
$\left\{\begin{array}{l} \text{set}(/\textsc{shared}/\textsc{lu}/\textsc{moves},\ \mathsf{set}([\mathsf{ask}(?A.\mathsf{price}(A))])) \\ \text{set}(/\textsc{shared}/\textsc{lu}/\textsc{speaker},\ \mathsf{usr}) \end{array}\right.$

integrateUsrAsk
$\left\{\begin{array}{l} \text{push}(/\textsc{shared}/\textsc{qud},\ ?A.\mathsf{price}(A)) \\ \text{push}(/\textsc{private}/\textsc{agenda},\ \mathsf{respond}(?A.\mathsf{price}(A))) \end{array}\right.$

findPlan
$\left\{\begin{array}{l} \text{pop}(/\textsc{private}/\textsc{agenda}) \\ \text{set}(/\textsc{private}/\textsc{plan},\ \mathsf{stack}([\mathsf{raise}(?C.\mathsf{how}(C)),\ \mathsf{findout}(?D.\mathsf{dest\_city}(D)),\ \ldots])) \end{array}\right.$

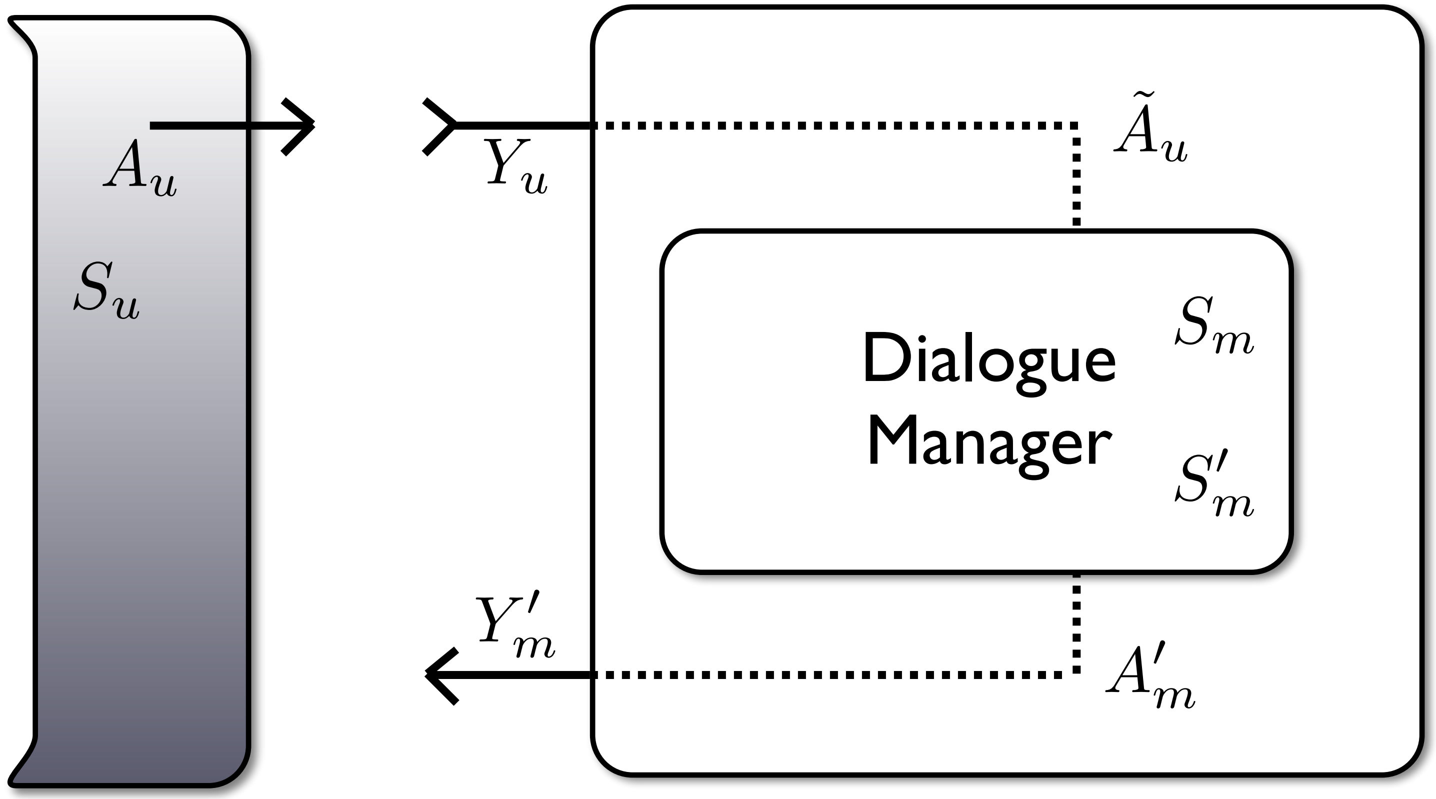
$$\left[\begin{array}{lcl} \textsc{private} & = & \left[\begin{array}{lcl} \textsc{agenda} & = & \langle\rangle \\ \textsc{plan} & = & \left\langle \begin{array}{l} \mathsf{raise}(?A.\mathsf{how}(A)) \\ \mathsf{findout}(?B.\mathsf{dest\_city}(B)) \\ \mathsf{findout}(?C.\mathsf{dept\_city}(C)) \\ \mathsf{findout}(?D.\mathsf{month}(D)) \\ \mathsf{findout}(?E.\mathsf{dept\_day}(E)) \\ \mathsf{findout}(?F.\mathsf{class}(F)) \\ \mathsf{consultDB}(?G.\mathsf{price}(G)) \end{array} \right\rangle \\ \textsc{bel} & = & \{\} \end{array}\right] \\ \textsc{shared} & = & \left[\begin{array}{lcl} \textsc{com} & = & \{\} \\ \textsc{qud} & = & \langle\ ?H.\mathsf{price}(H)\ \rangle \\ \textsc{lu} & = & \left[\begin{array}{lcl} \textsc{speaker} & = & \mathsf{usr} \\ \textsc{moves} & = & \{\ \mathsf{ask}(?H.\mathsf{price}(H))\ \} \end{array}\right] \end{array}\right] \end{array}\right]$$

selectFromPlan
$\{\ \text{push}(/\textsc{private}/\textsc{agenda},\ \mathsf{raise}(?A.\mathsf{how}(A))$

selectAsk
$\left\{\begin{array}{l} \text{add}(\textsc{next\_moves},\ \mathsf{ask}(?A.\mathsf{how}(A))) \\ \text{if\_do}(\text{fst}(\$/\textsc{private}/\textsc{plan},\ \mathsf{raise}(?A.\mathsf{how}(A \end{array}\right.$

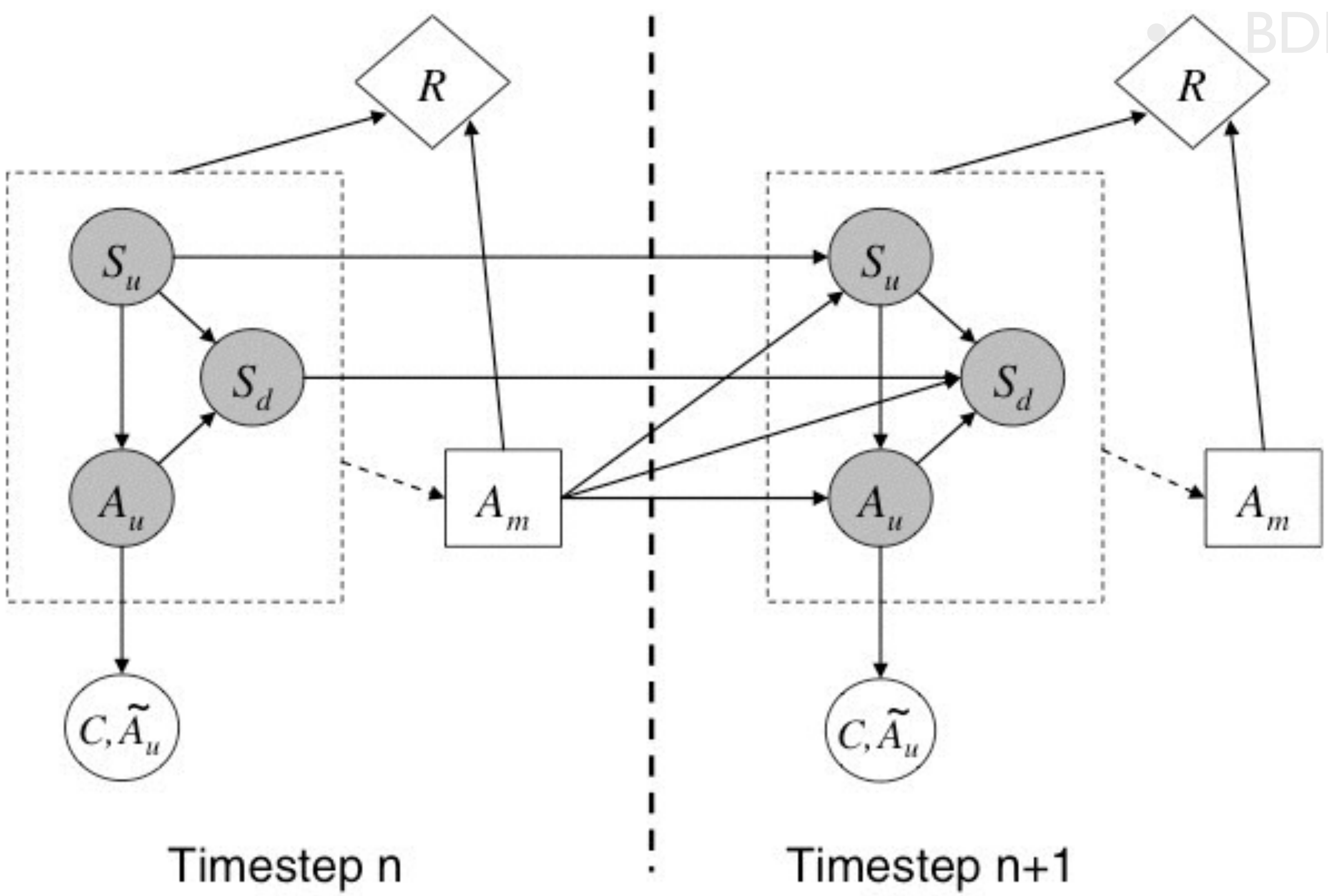
"how do you want to travel?"

(Larsson 2002)

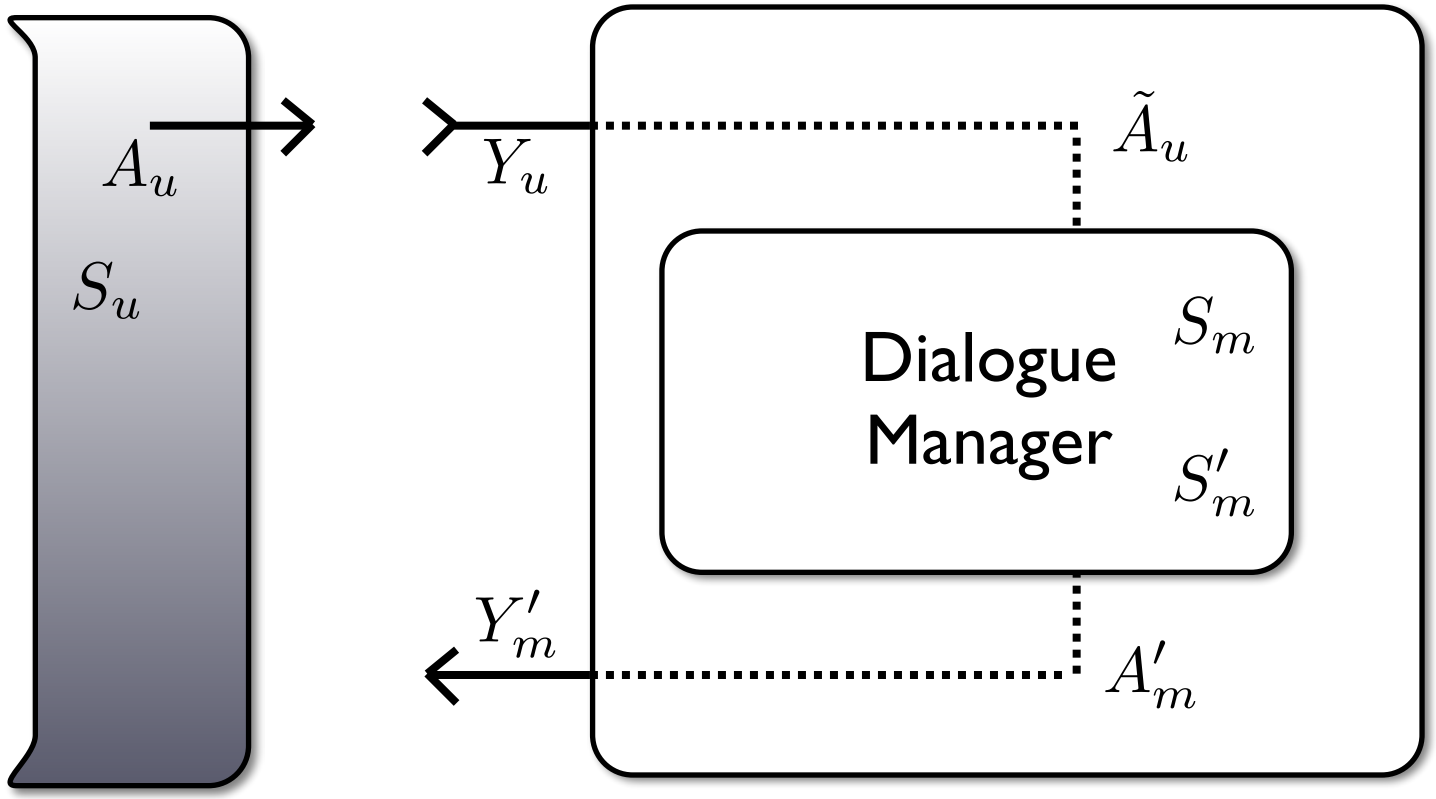


$$U(S_m, \tilde{A}_u) = S'_m$$
$$G(S_m) = A_m$$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- -- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000)(Schlangen 2003)
- BDI (Allen 1995; Perrault & Allen 1980)



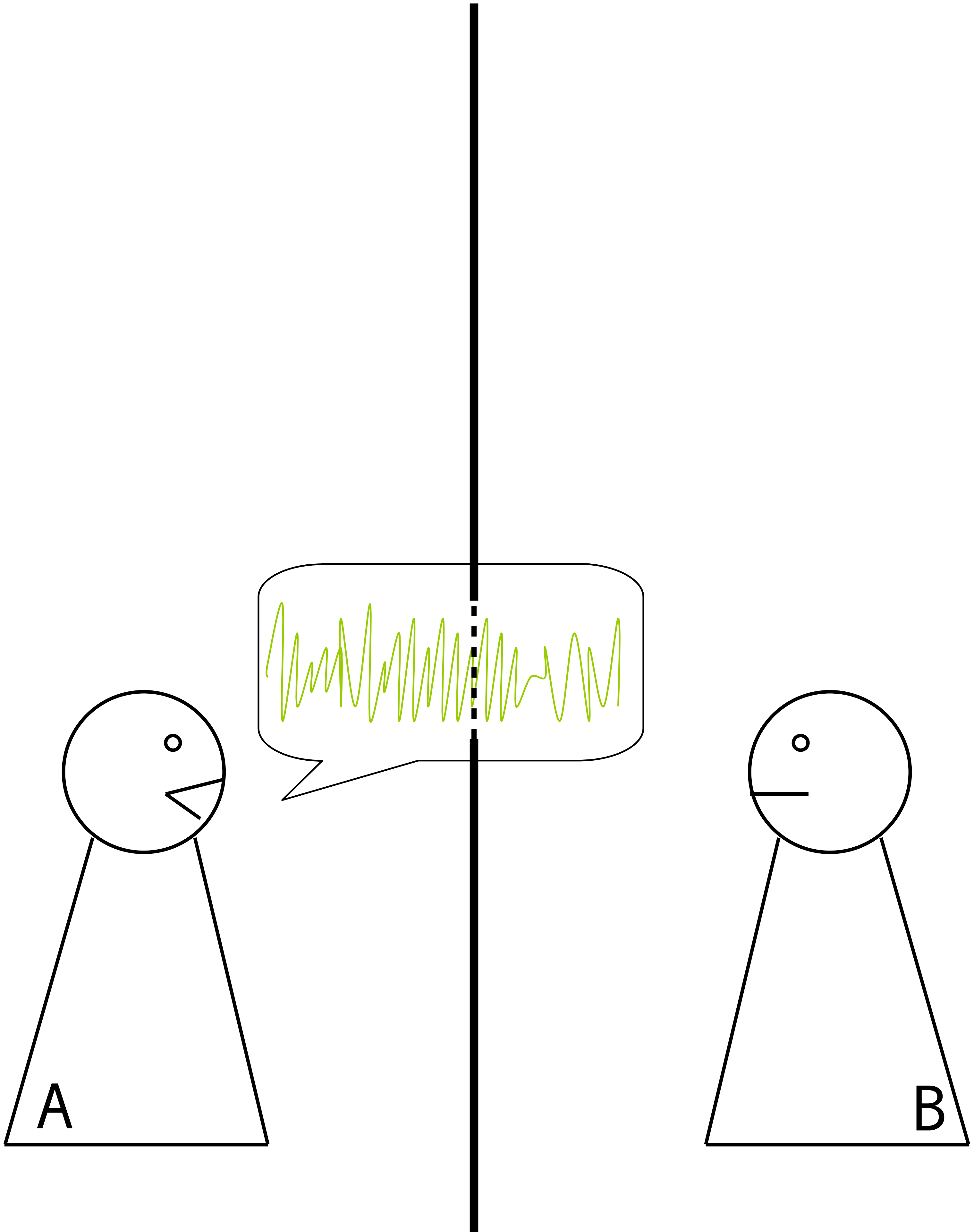
aus (Williams & Young 2007)



$U(S_m, \tilde{A}_u) = S'_m$
$G(S_m) = A_m$

- FSA (eg. Oviatt et al. 1994, McTear 1998)
- Forms (eg., VoiceXML; Goddeau et al. 1996)
- Stacks, Trees, ... (Bohus & Rudnicky 2009; Lemon et al. 2002)
- (PO)MDP (Singh et al. 2000; Williams & Young 2007)
- Conversational Scoreboard (Larsson & Traum 2000) (Schlangen 2003)
- BDI (Allen 1995; Perrault & Allen 1980)

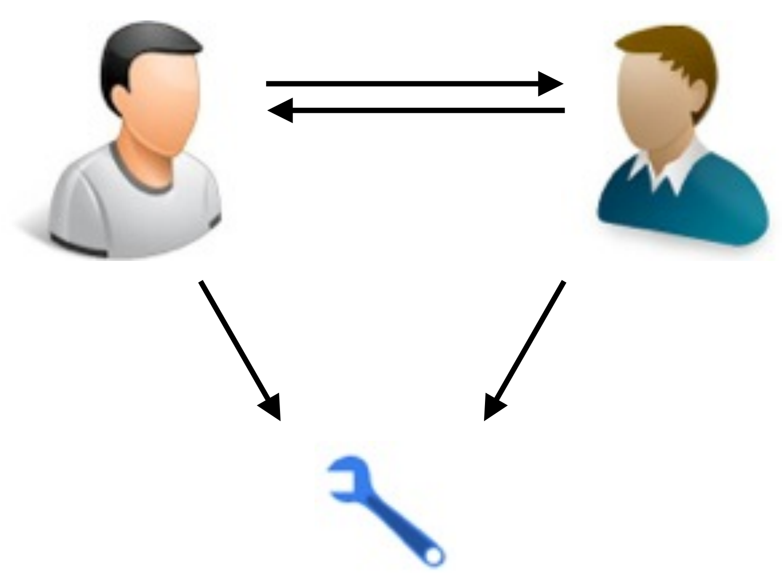
non- Incremental Dialogue Processing


A
B

situated dialogue

- participants share a common timeline:

- participants are co-located:



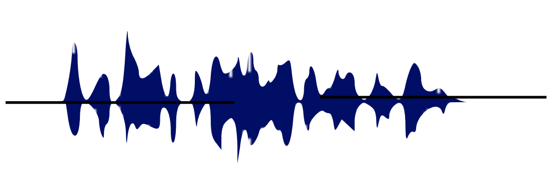
Overview

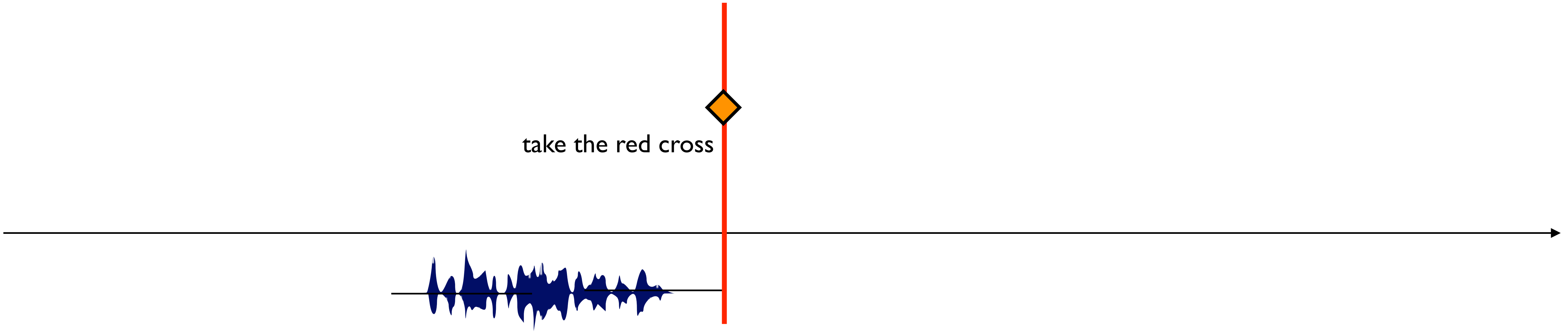
- **Part I: Foundations**
  - coordination, convention
  - communicative intentions
  - non-conventional meaning
  - grounding
  - turn-taking
  - disfluencies
- **Part II: Computational Models**
  - approaches to dialogue modelling
  - incremental processing, turn-taking
  - an example: grounded semantics

the IU model

– Motivation –

- fundamental skills of an agent:
  - to form hypotheses about the world
  - to plan and perform actions on the world
- IU model as a temporally fine-grained model of the information state of an agent, and of how it is updated




take the red cross

the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

the IU model

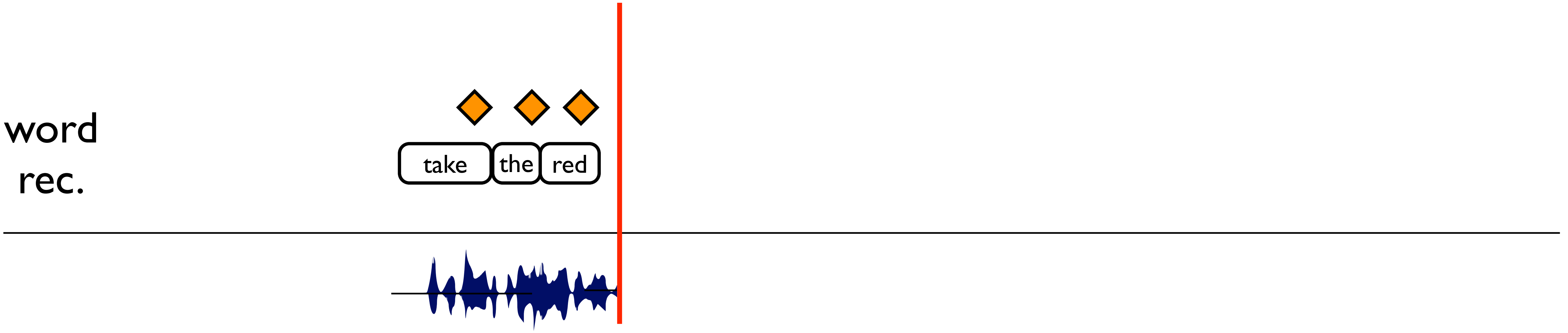
– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

the IU model

– Assumptions –

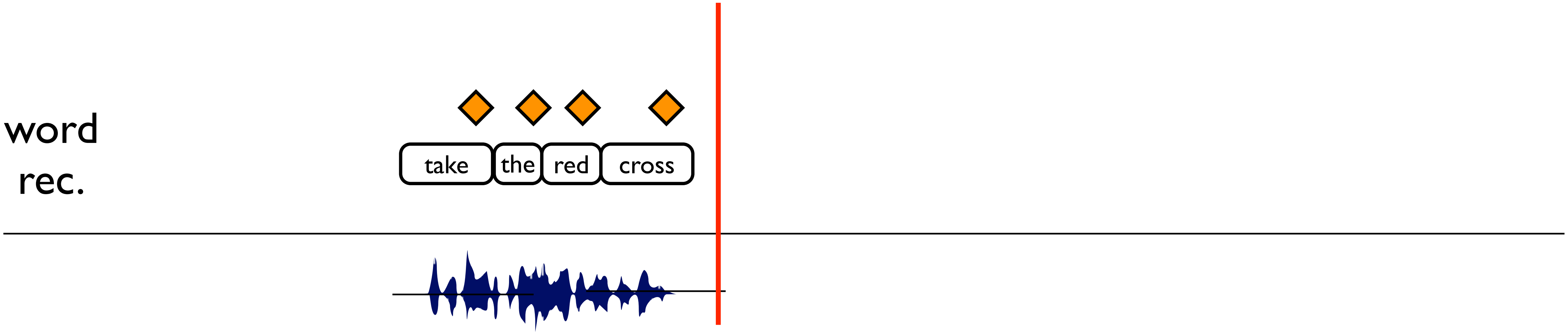
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised



the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

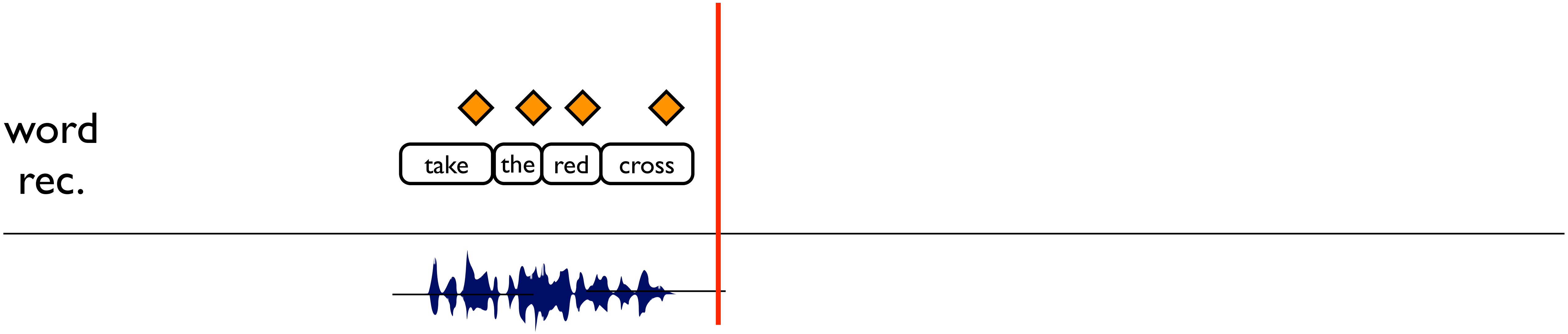


the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

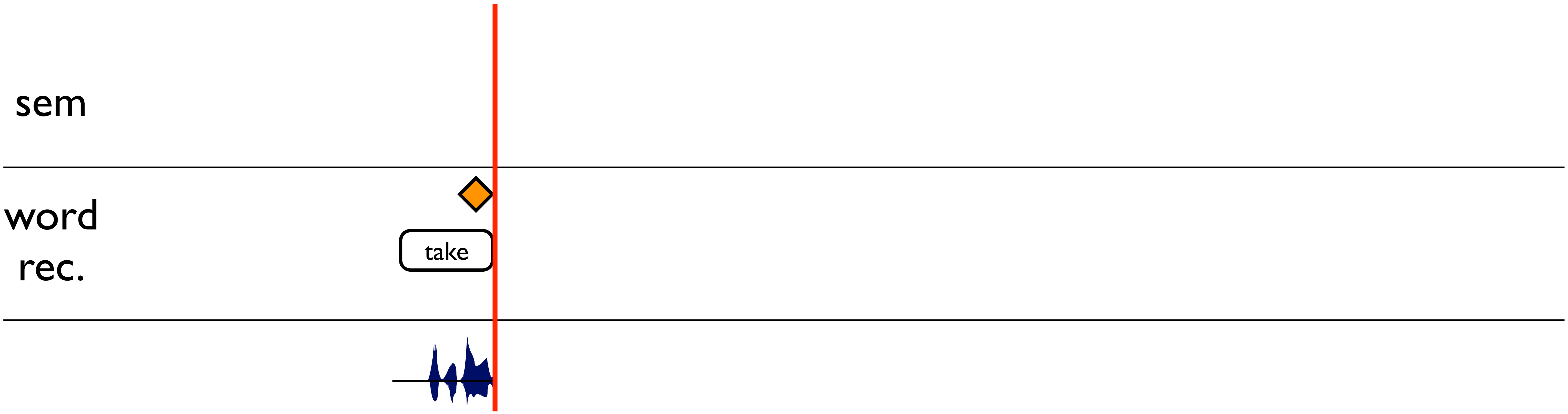
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.



the IU model
– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.

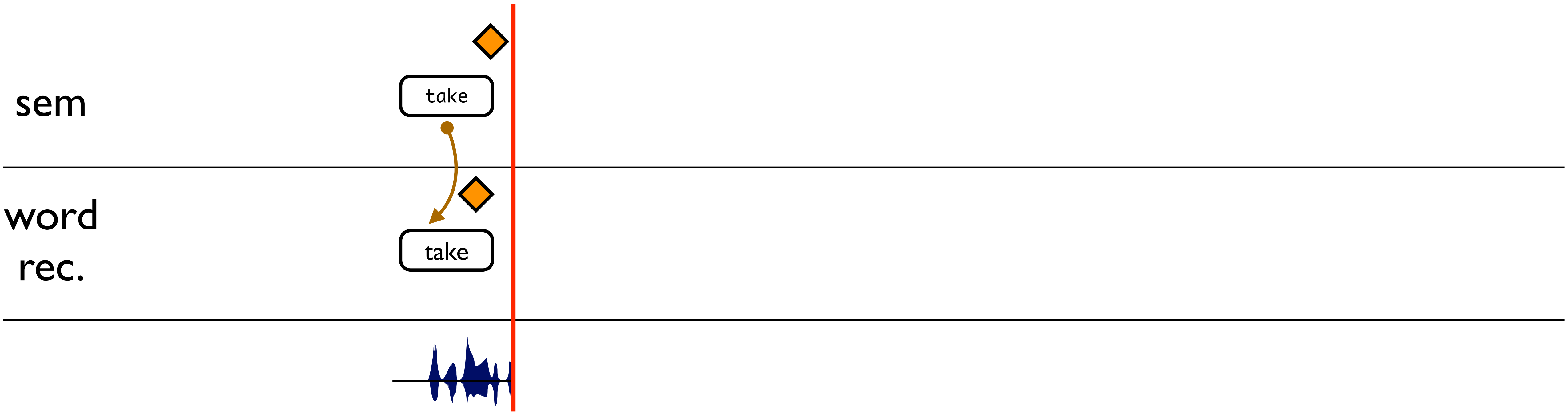


the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.

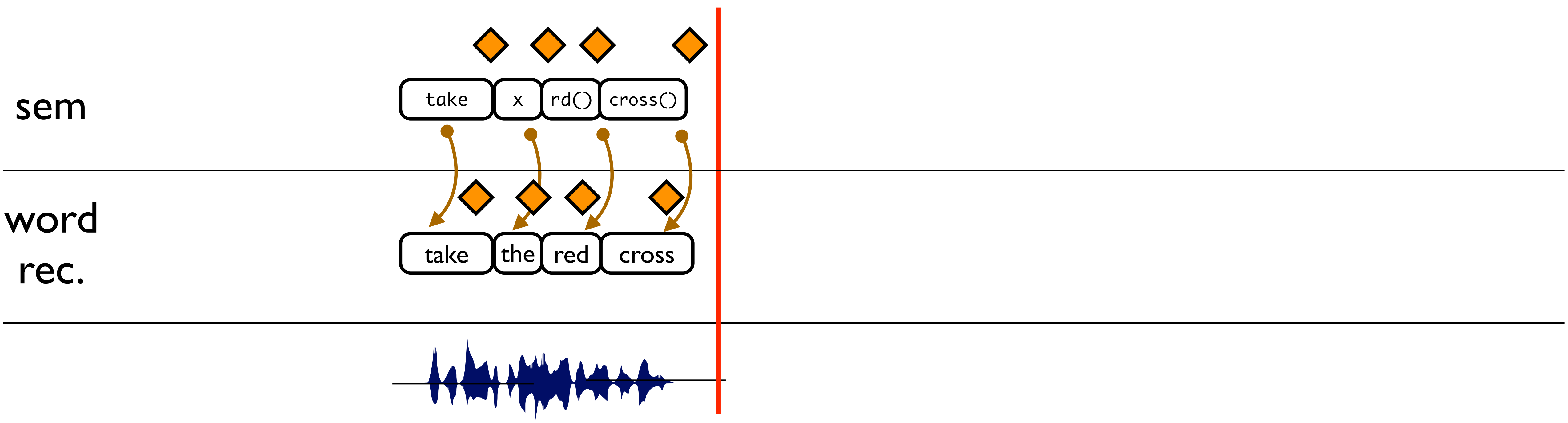


the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.



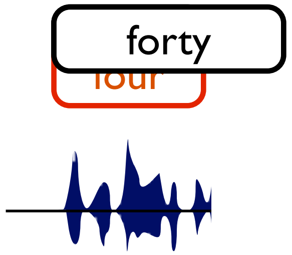
the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.

- IS may have to be revised, in light of newer information



the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised

- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.

- IS may have to be revised, in light of newer information

the IU model

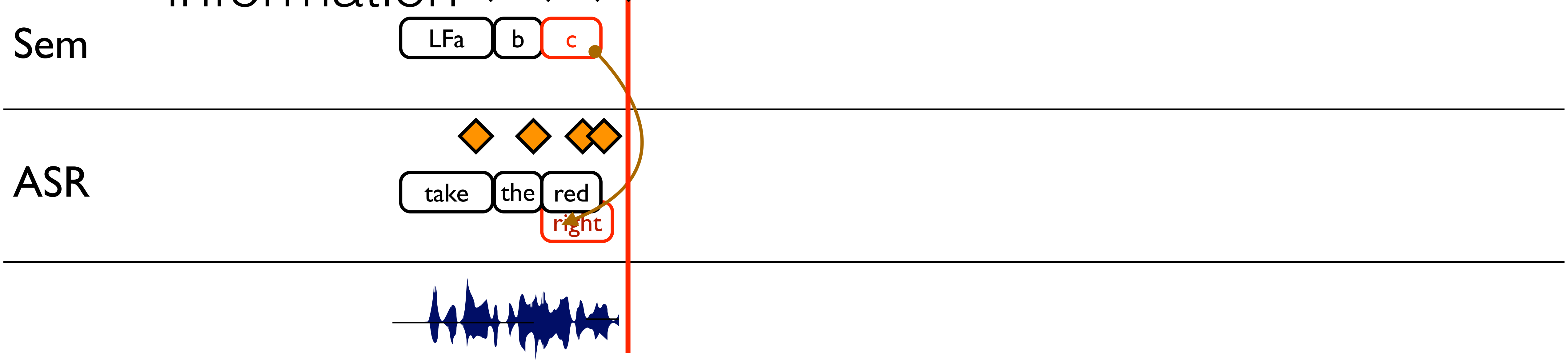
– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.
- IS may have to be revised, in light of newer information

the IU model

– Assumptions –

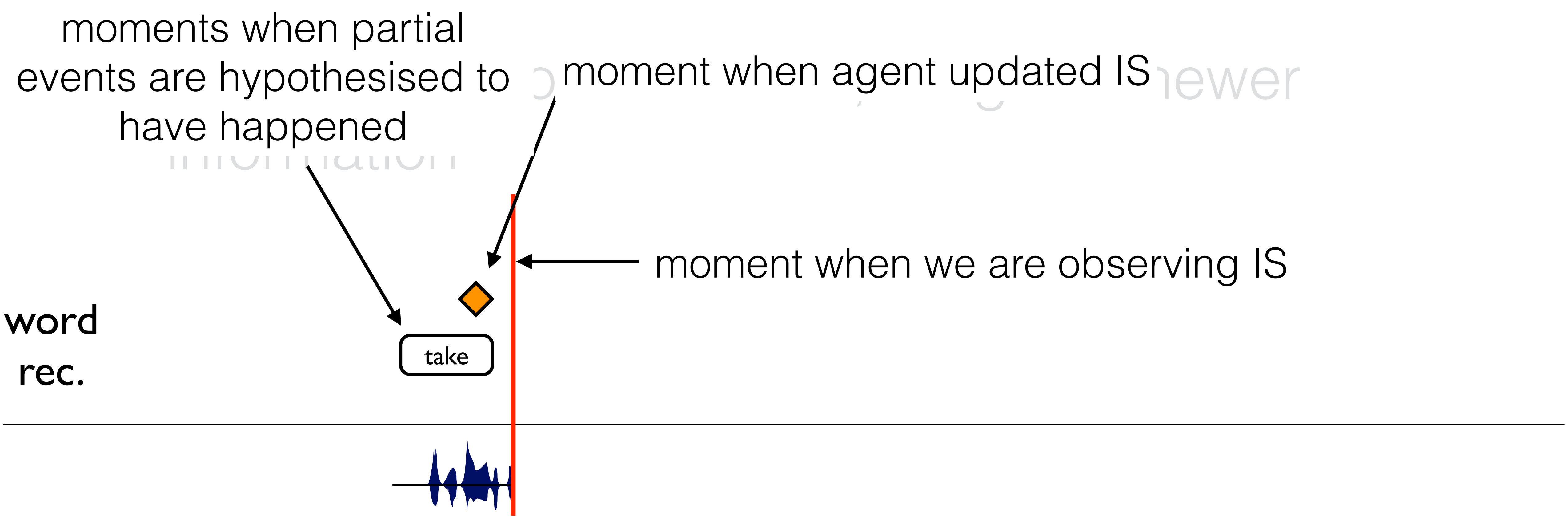
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.
- IS may have to be revised, in light of newer information



the IU model

– Assumptions –

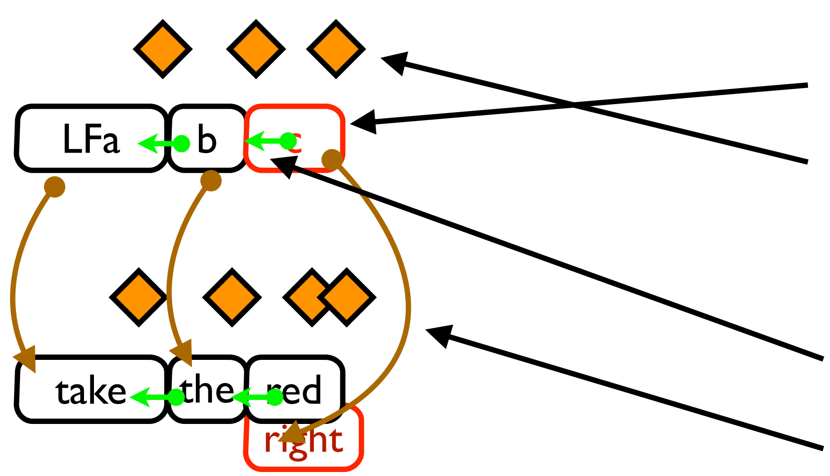
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.



the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.
- IS may have to be revised, in light of newer information

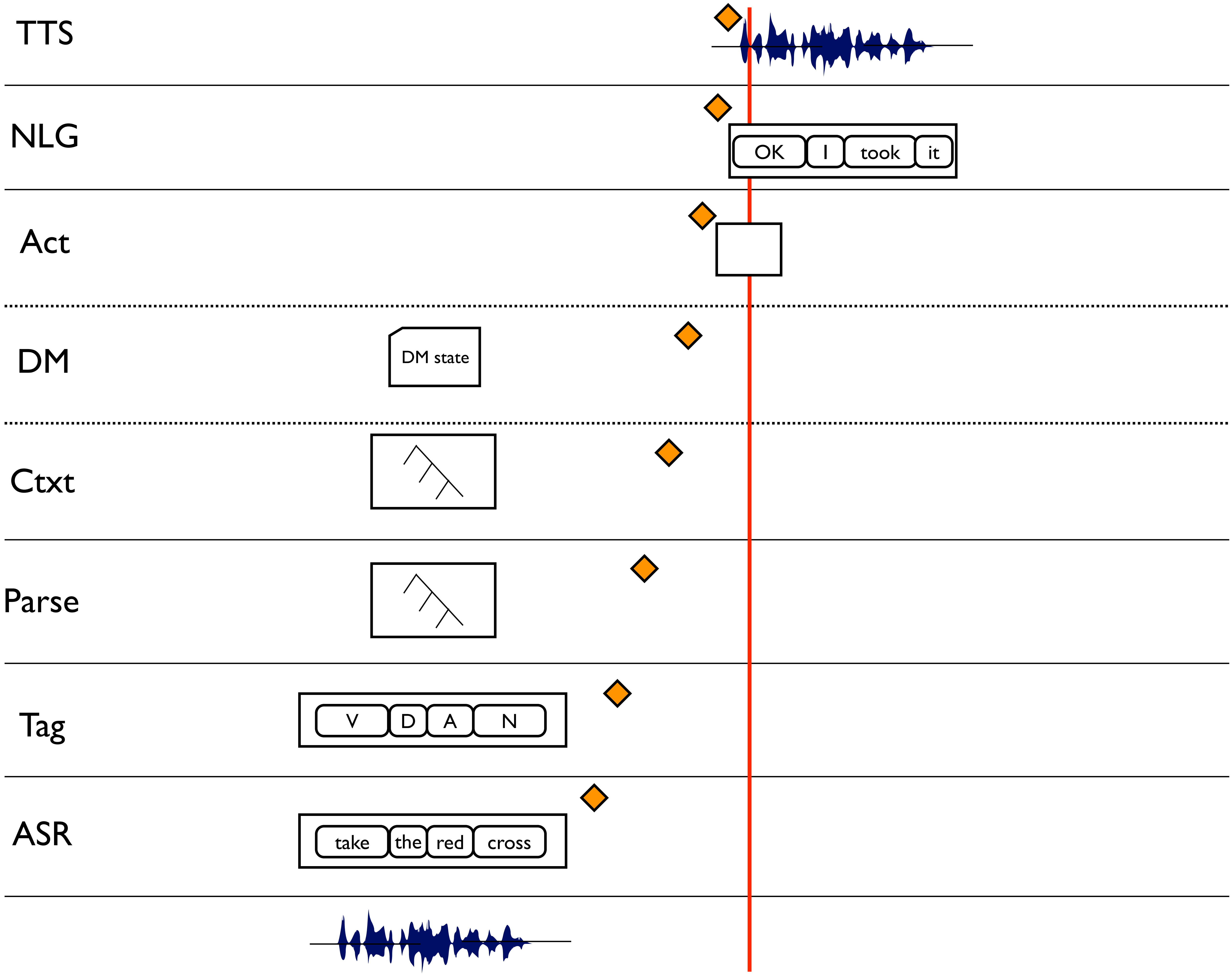


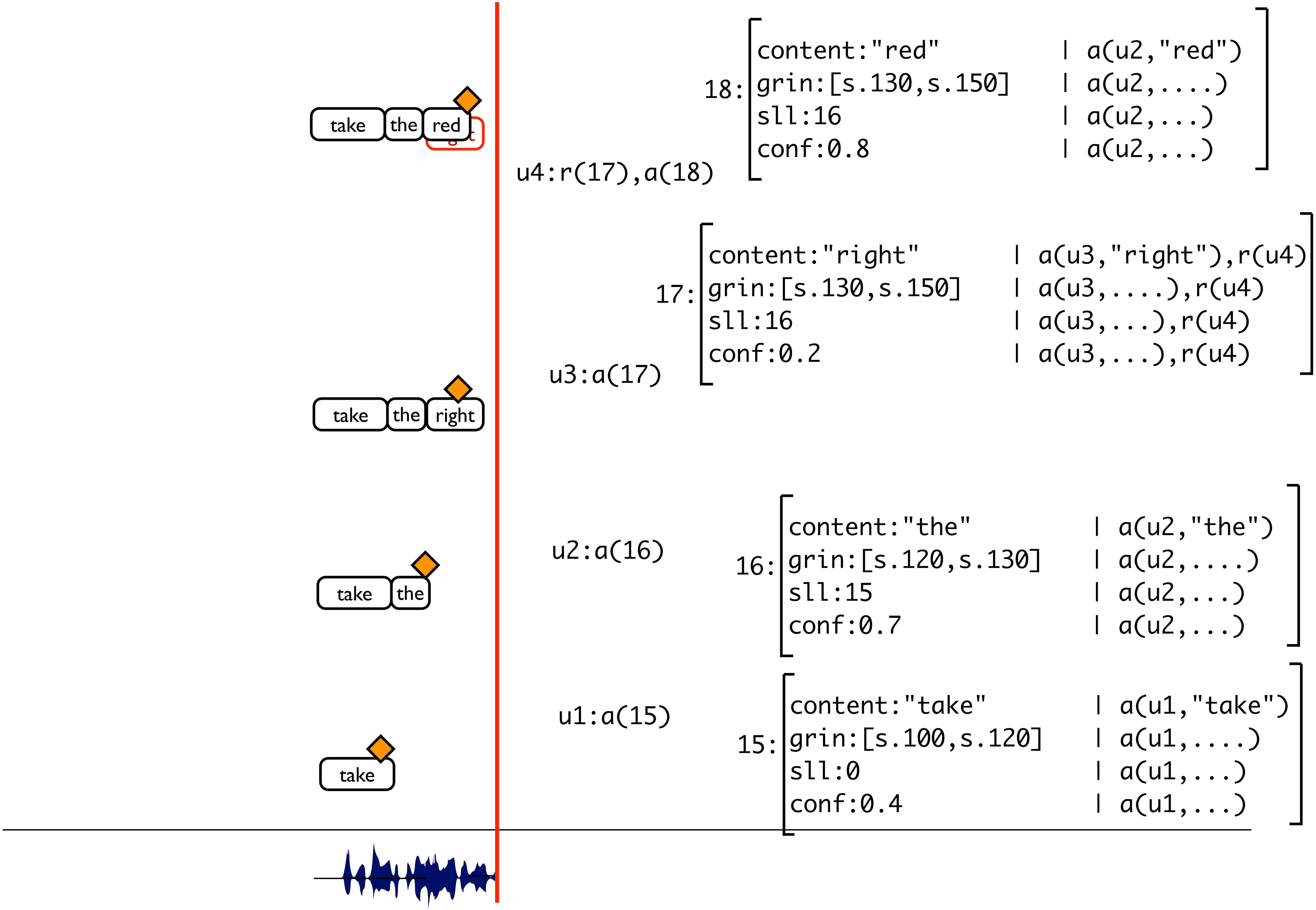
IUs
Updates
2 types of relation:
* same-level links
* grounded-in links

the IU model

– Assumptions –

- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- “Higher-level” hypotheses can be formed on the basis of “lower-level” ones.
- IS may have to be revised, in light of newer information

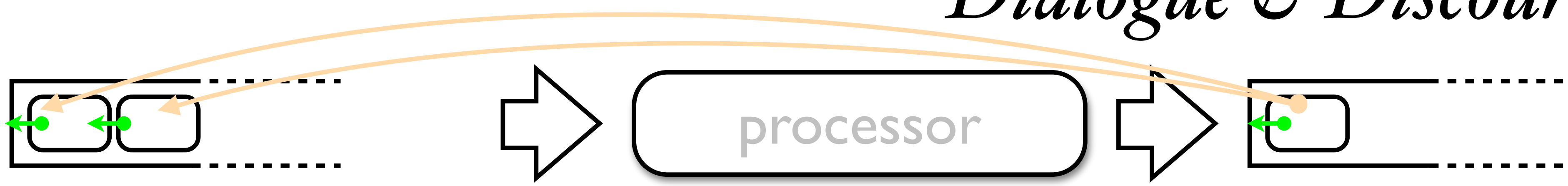

TTS
OK
I
took
it
NLG
Act
DM
DM state
Ctxt
Parse
Tag
V
D
A
N
ASR
take
the
red
cross



take
the
red
u4:r(17),a(18)
18:
content:"red" | a(u2,"red")
grin:[s.130,s.150] | a(u2,....)
sll:16 | a(u2,...)
conf:0.8 | a(u2,...)
take
the
right
u3:a(17)
17:
content:"right" | a(u3,"right"),r(u4)
grin:[s.130,s.150] | a(u3,....),r(u4)
sll:16 | a(u3,...),r(u4)
conf:0.2 | a(u3,...),r(u4)
take
the
u2:a(16)
16:
content:"the" | a(u2,"the")
grin:[s.120,s.130] | a(u2,....)
sll:15 | a(u2,...)
conf:0.7 | a(u2,...)
take
u1:a(15)
15:
content:"take" | a(u1,"take")
grin:[s.100,s.120] | a(u1,....)
sll:0 | a(u1,...)
conf:0.4 | a(u1,...)

incremental processing
the IU model

- incremental units (Schlangen & Skantze, EACL 2009 / *Dialogue & Discourse* 2011)



  - IUs: minimal units of characteristic in/output, parts of larger unit
  - part-of-relation represented as *same-level-links*
  - IUs *grounded in* IUs from other levels, which were drawn on when building them
- Implemented in InproTK (http://www.inpro.tk), Jindigo (Skantze), IPAACA (Kopp & Buschmeier)

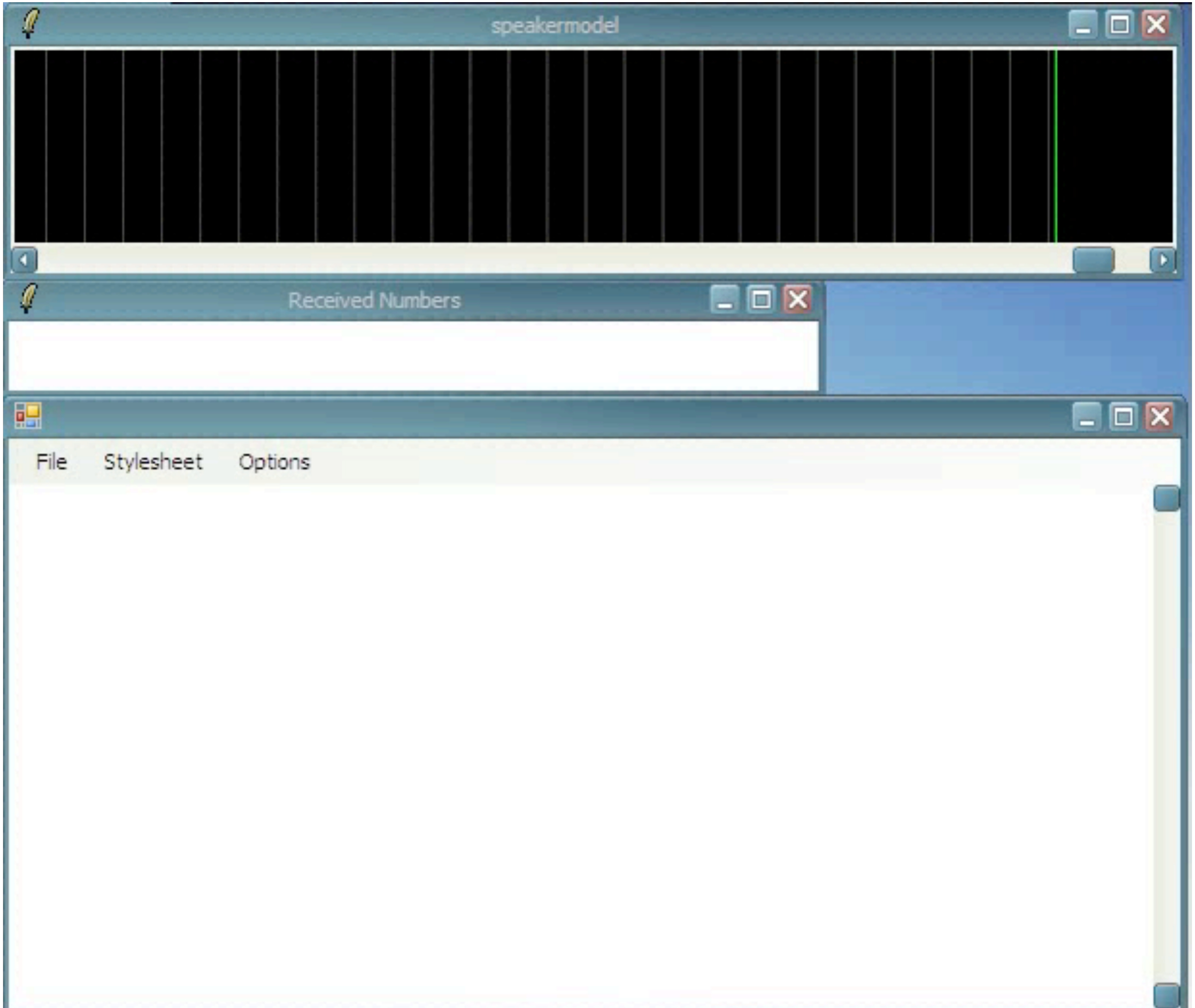
The NUMBERS systems
fast turn-taking

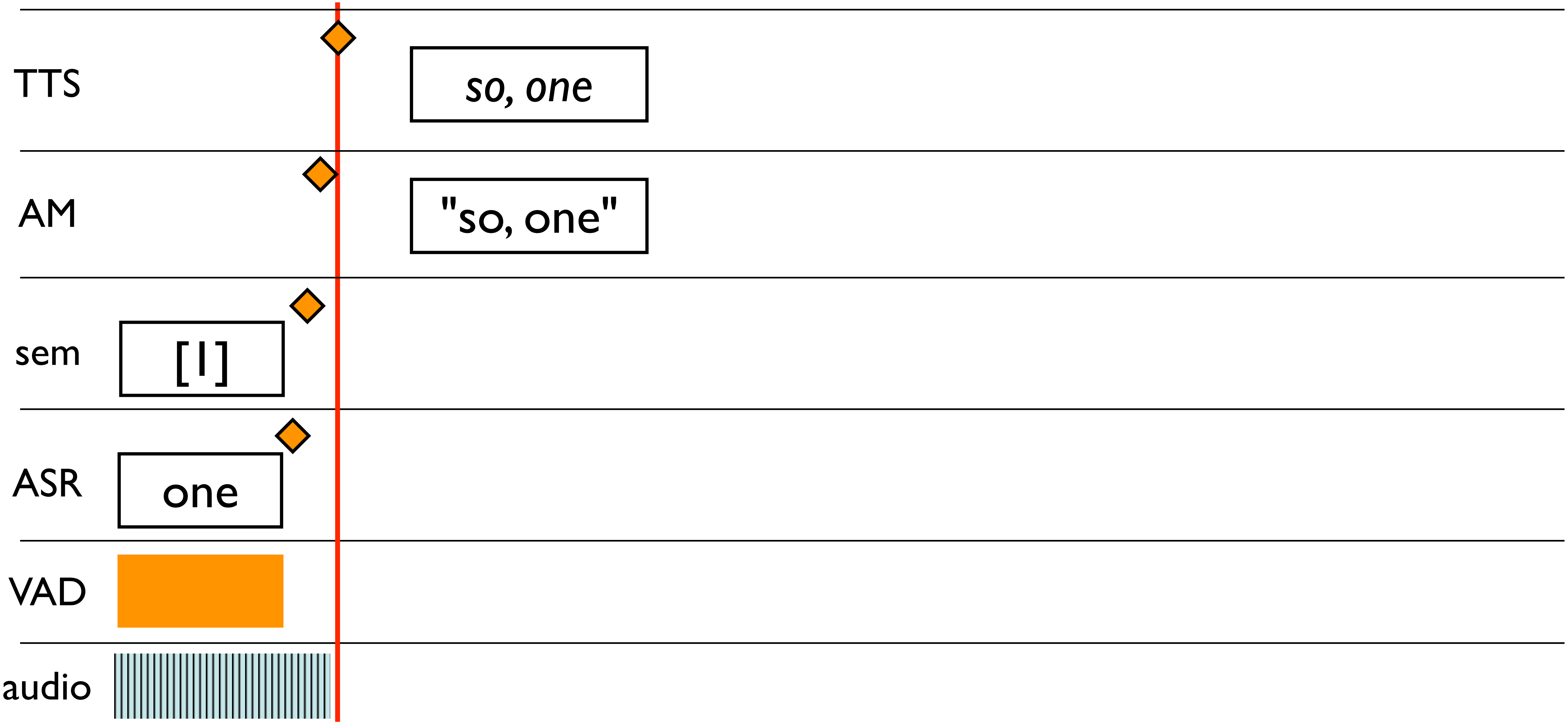
joint work with Gabriel Skantze
(Skantze & Schlangen, EACL 2009)

The NUMBERS systems
fast turn-taking

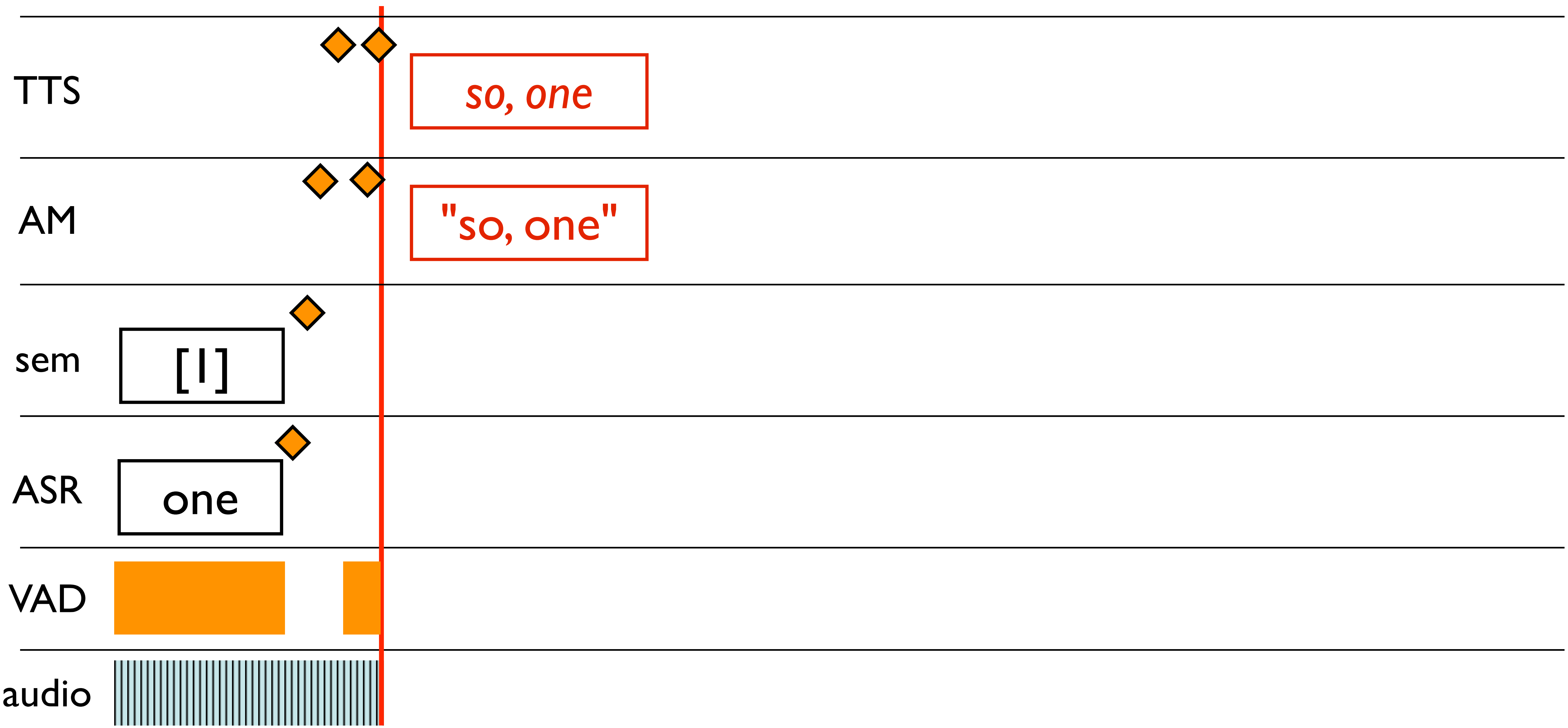
- user dictates a string of digits to system
- system tries to ground its understanding, as quickly as possible
- processing based on IU-model:
    - minimal units trigger updates
    - processors implement update functions

the *numbers* system

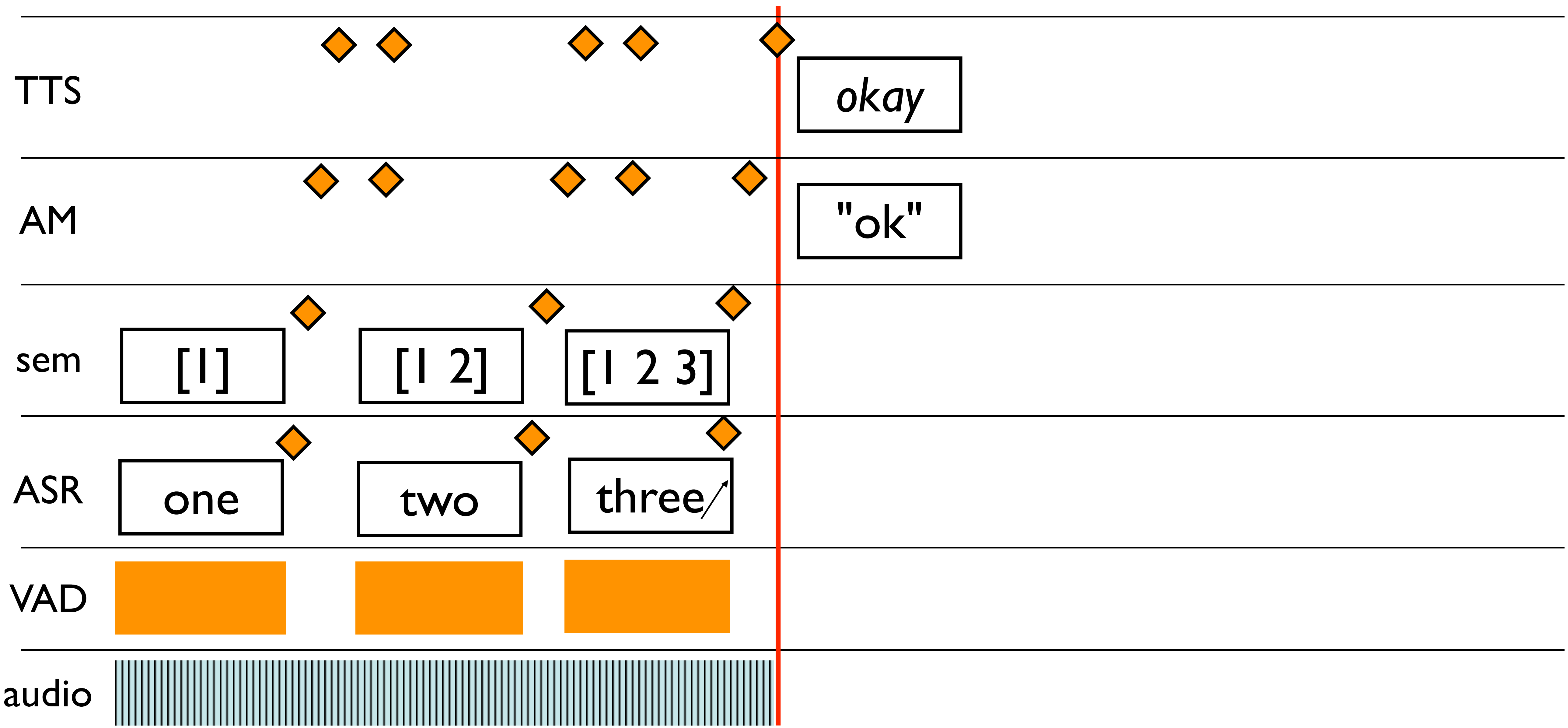


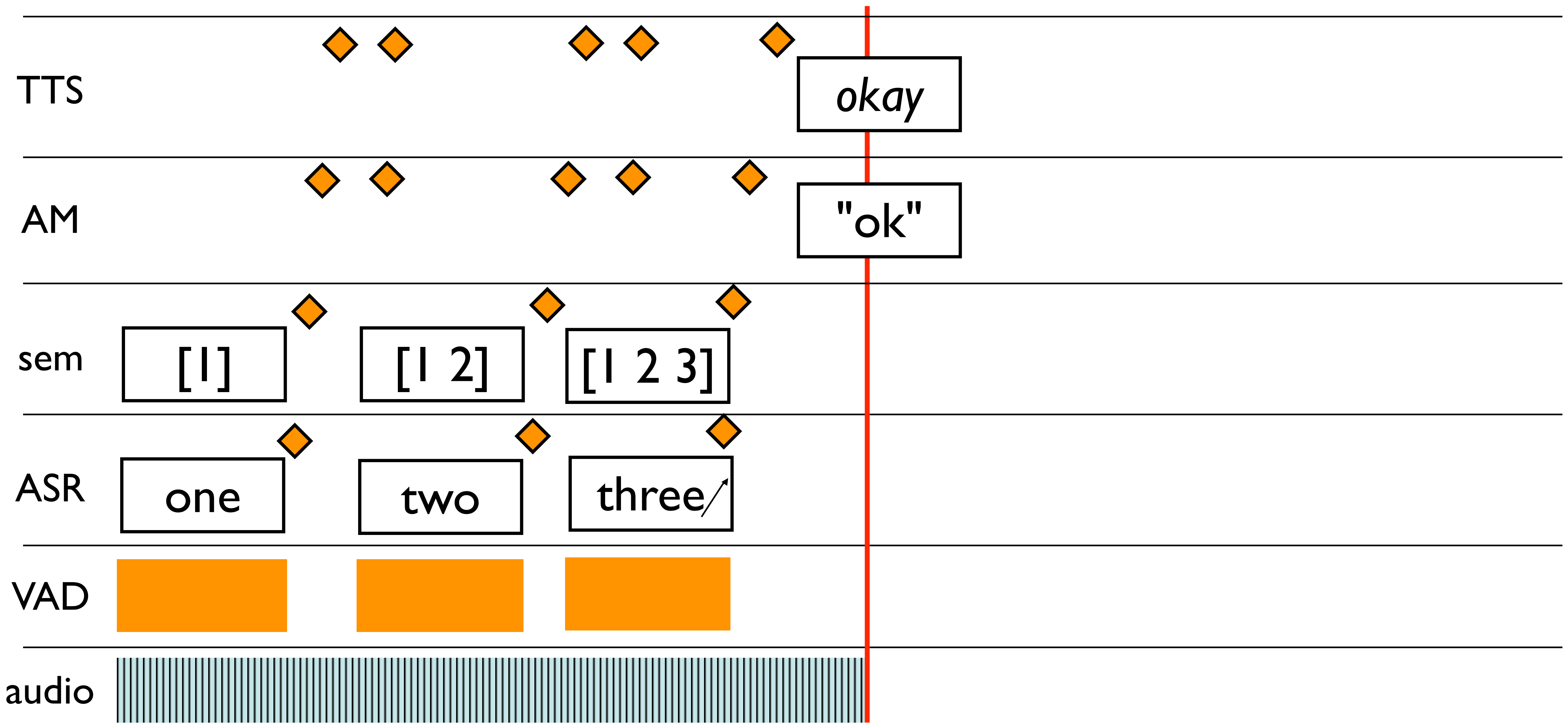


TTS
so, one
AM
"so, one"
sem
[1]
ASR
one
VAD
audio



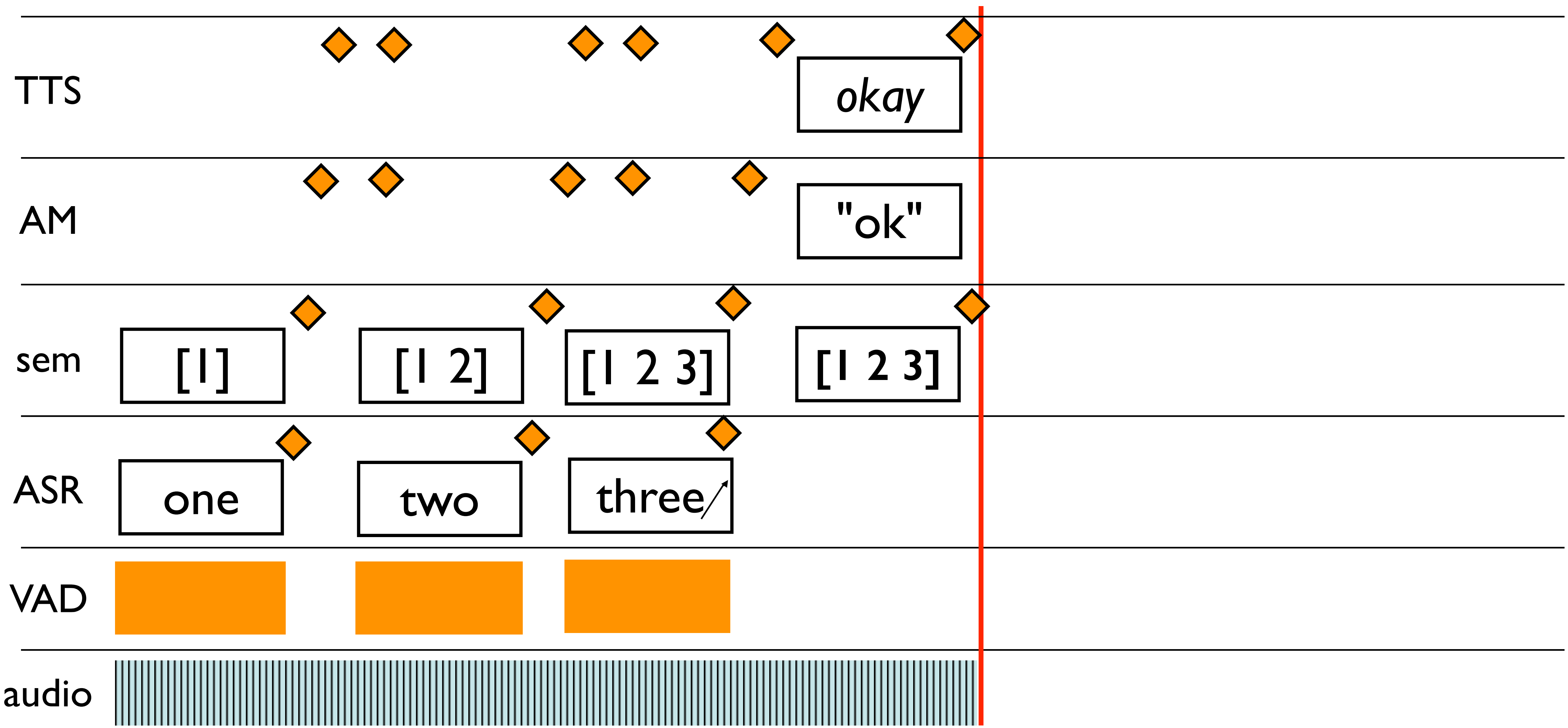
TTS
so, one
AM
"so, one"
sem
[1]
ASR
one
VAD
audio


TTS
okay
AM
"ok"
sem
[1]
[1 2]
[1 2 3]
ASR
one
two
three
VAD
audio



TTS
okay
AM
"ok"
sem
[1]
[1 2]
[1 2 3]
ASR
one
two
three
VAD
audio

- conditional projections
- w/ dynamic offsets

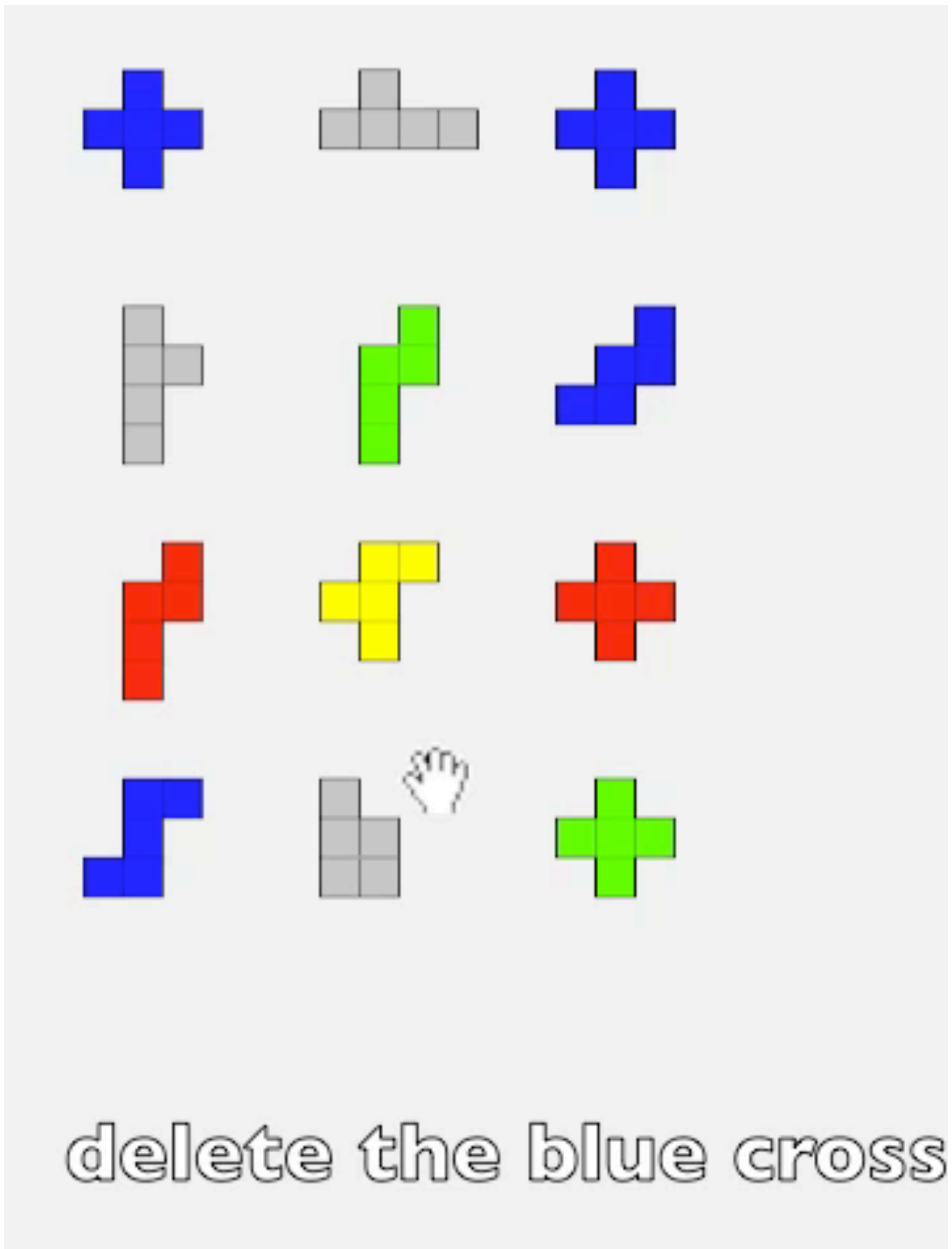


The PENTO-10 system
fast turn-taking, immediate exec

joint work with Okko Buß

(Buß *et al.*, SIGdial 2010, semdial 2010, 2011)

Pentomino System



U: *delete the blue cross*

S: which piece?

U: *top right.*

S: ok?

U: *right, now take the yellow [one]...*

S: yes?

U: *... and turn it...*

S: yes?

U: *... to the left*

S: ok.

U: *now flip the stairs...*

S: ok

U: *horizontally*

U: *that's right*

U: *erm now delete the red [one]*

S: *wh-*

U: *bottom right*

U: *correct.*

Evaluation

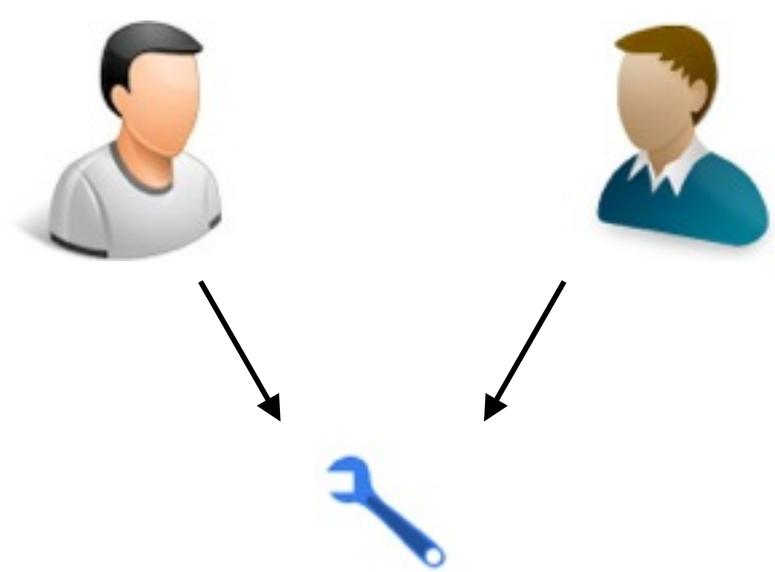
- Faster task completion compared to non-incremental versions of the systems
- Higher subjective ratings („would use again“, „behaves as expected“)
- *Not* higher task success rate

Overview

- **Part I: Foundations**
  - coordination, convention
  - communicative intentions
  - non-conventional meaning
  - grounding
  - turn-taking
  - disfluencies
- **Part II: Computational Models**
  - approaches to dialogue modelling
  - incremental processing, turn-taking
  - an example: grounded semantics

overview

- the *incremental units* model of incremental dialogue processing
- realising fast turn-taking
- a simple model of incremental reference resolution

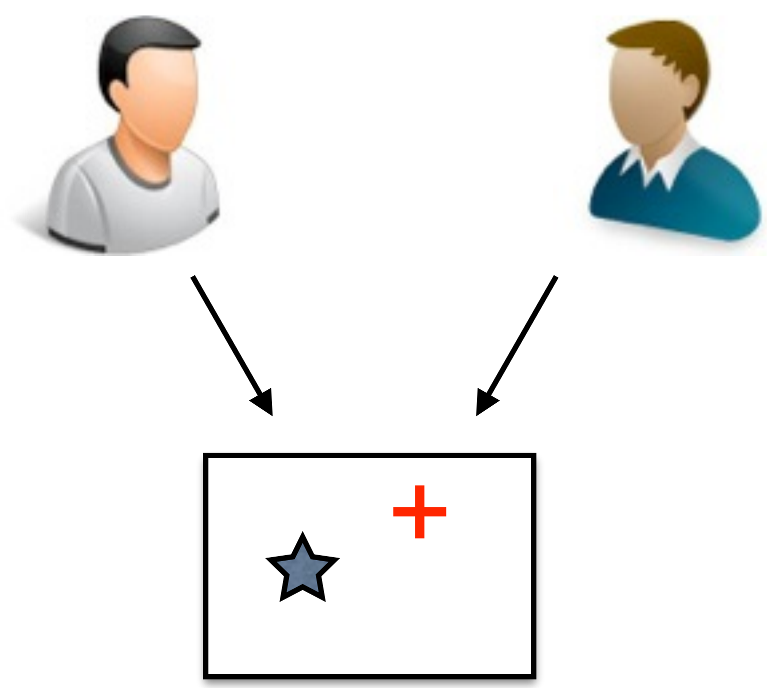


incremental statistical NLU

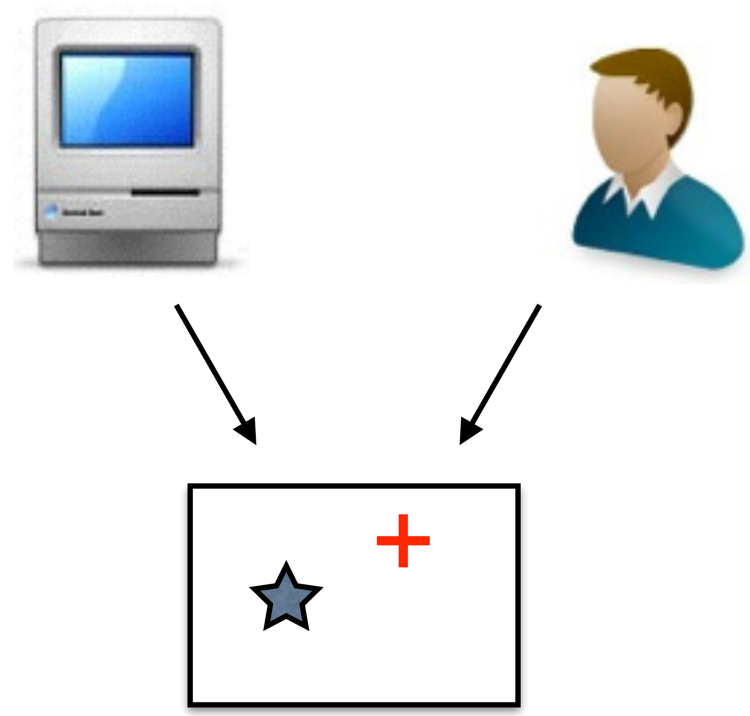
joint work with Casey Kennington

(Kennington *et al.*, SIGdial 2012, 2013; Coling 2014; Computer Speech & Language 2014)

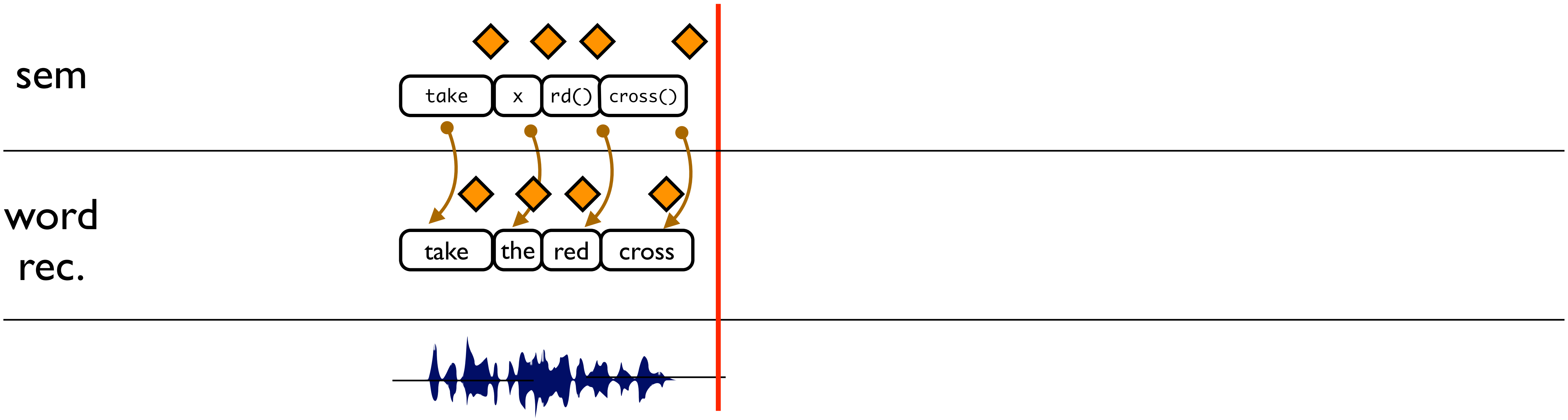
incremental statistical NLU



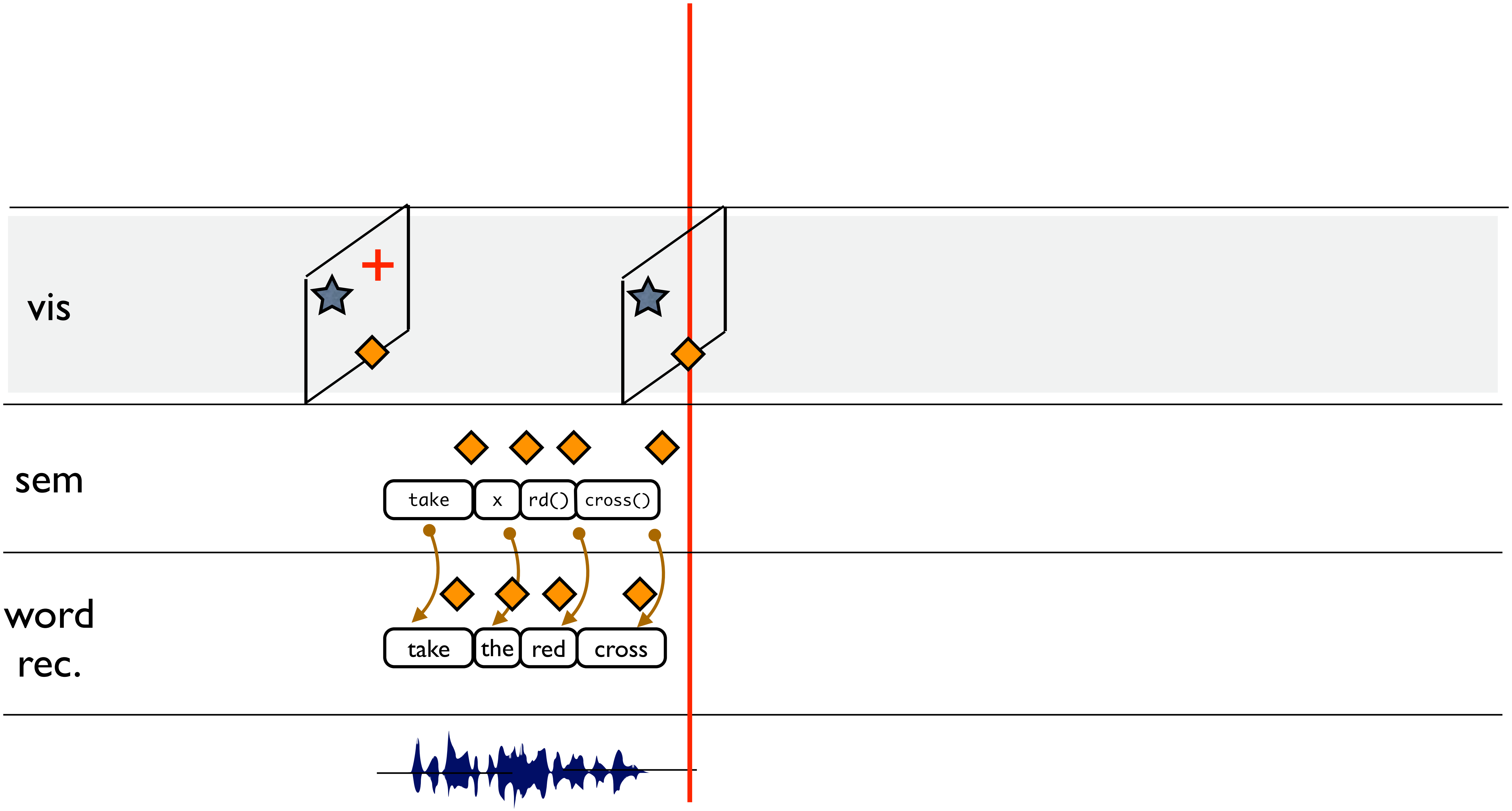
incremental statistical NLU



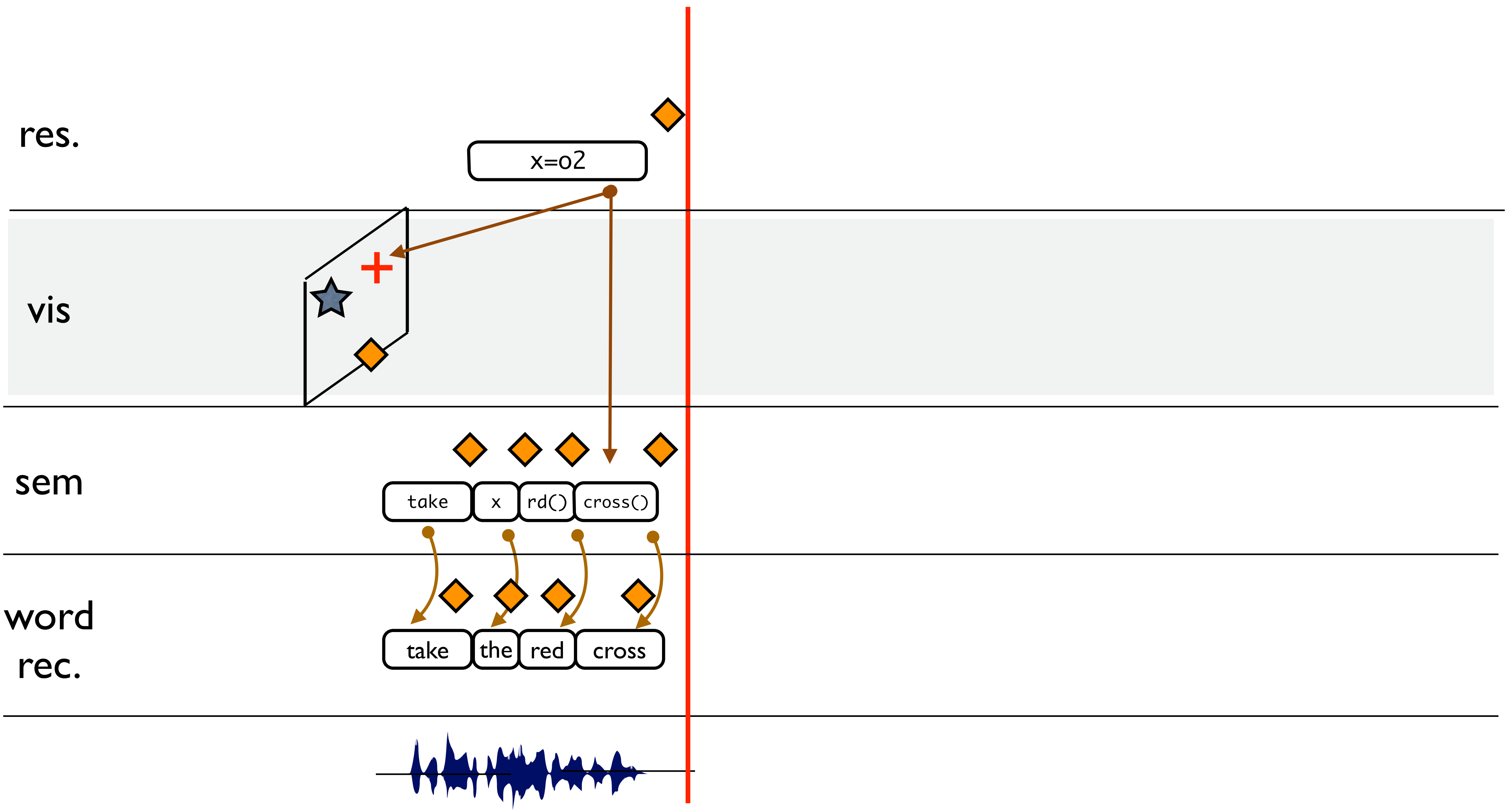
incremental statistical NLU



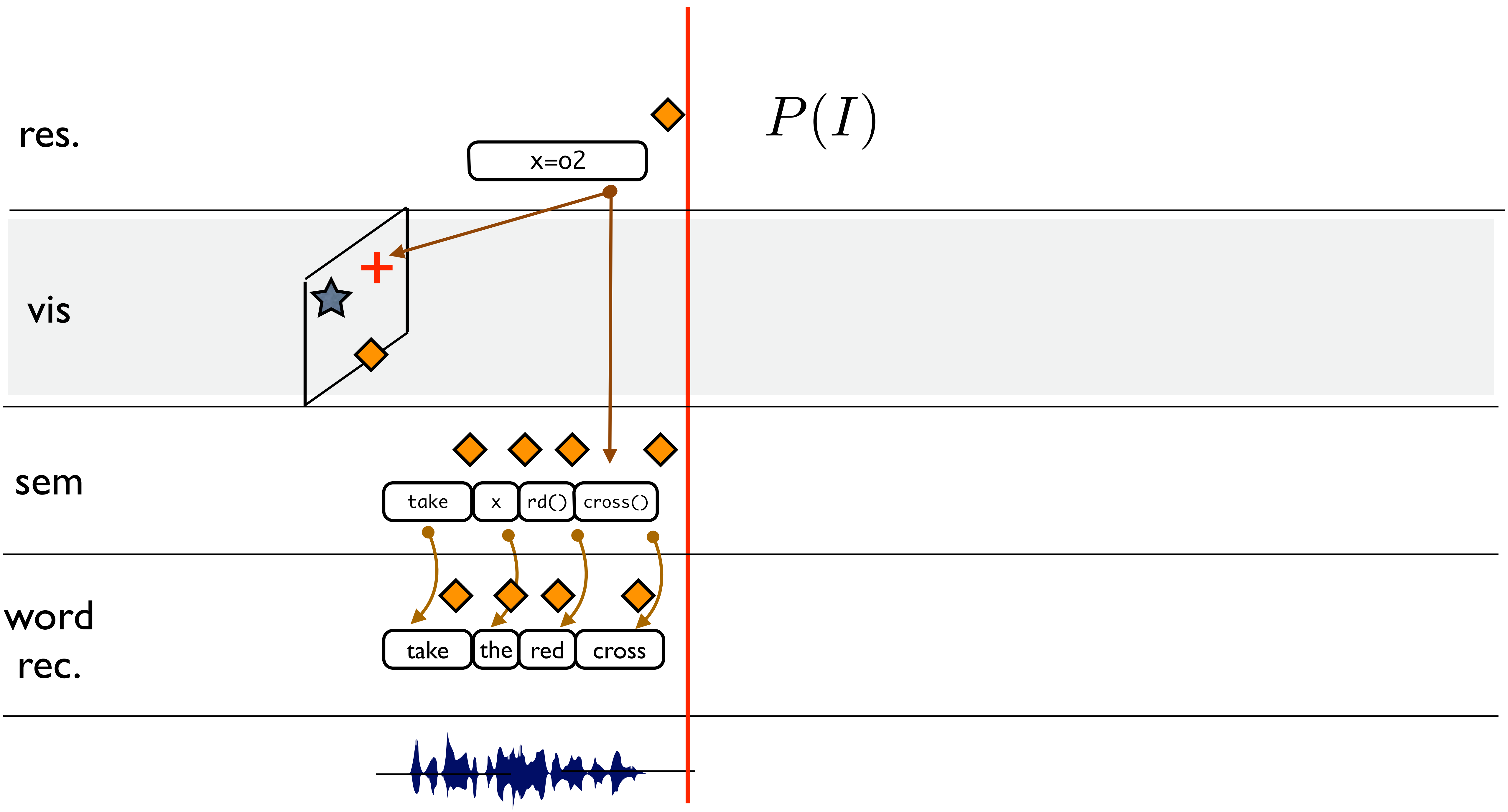
incremental statistical NLU



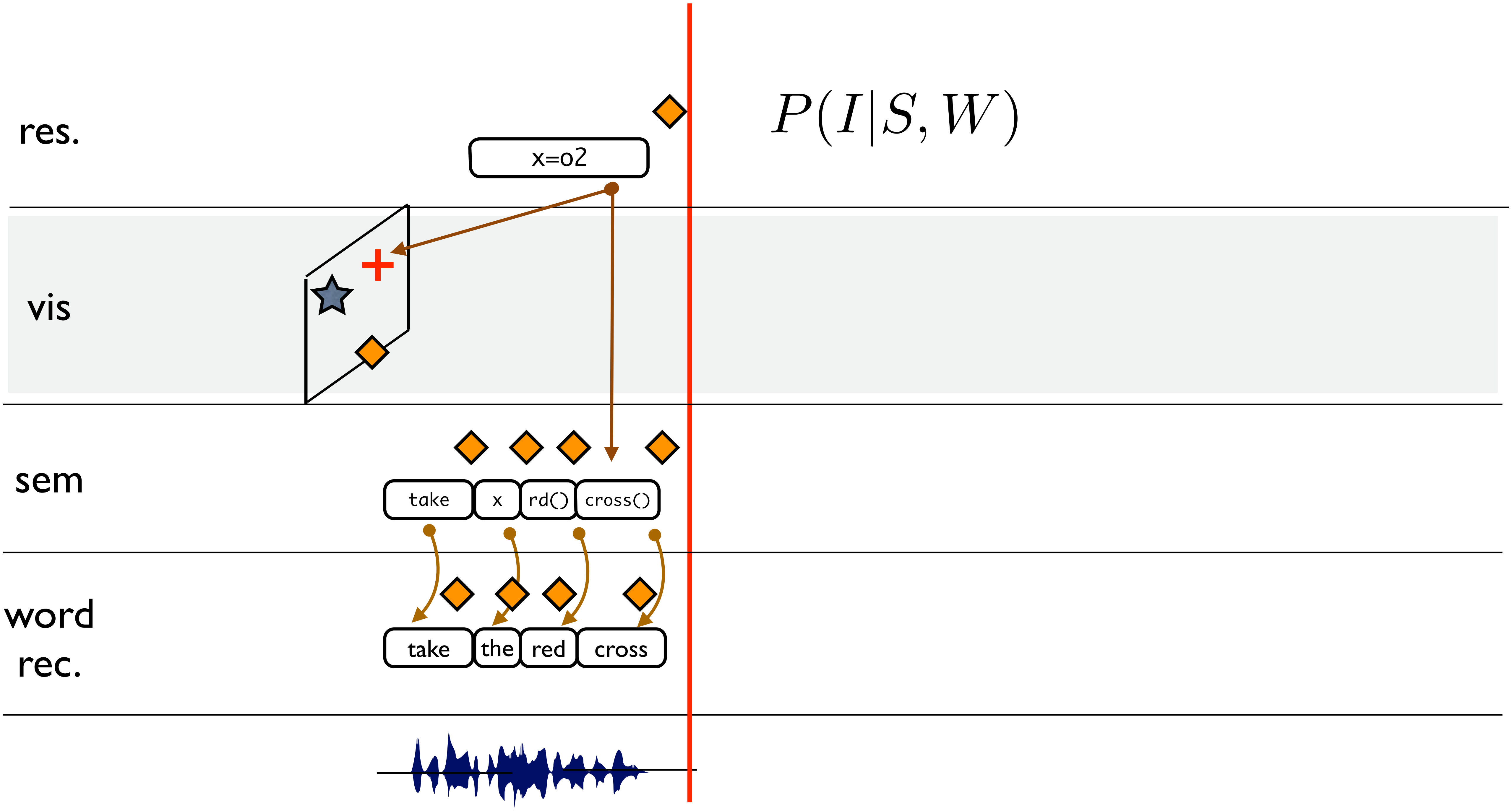
incremental statistical NLU



incremental statistical NLU

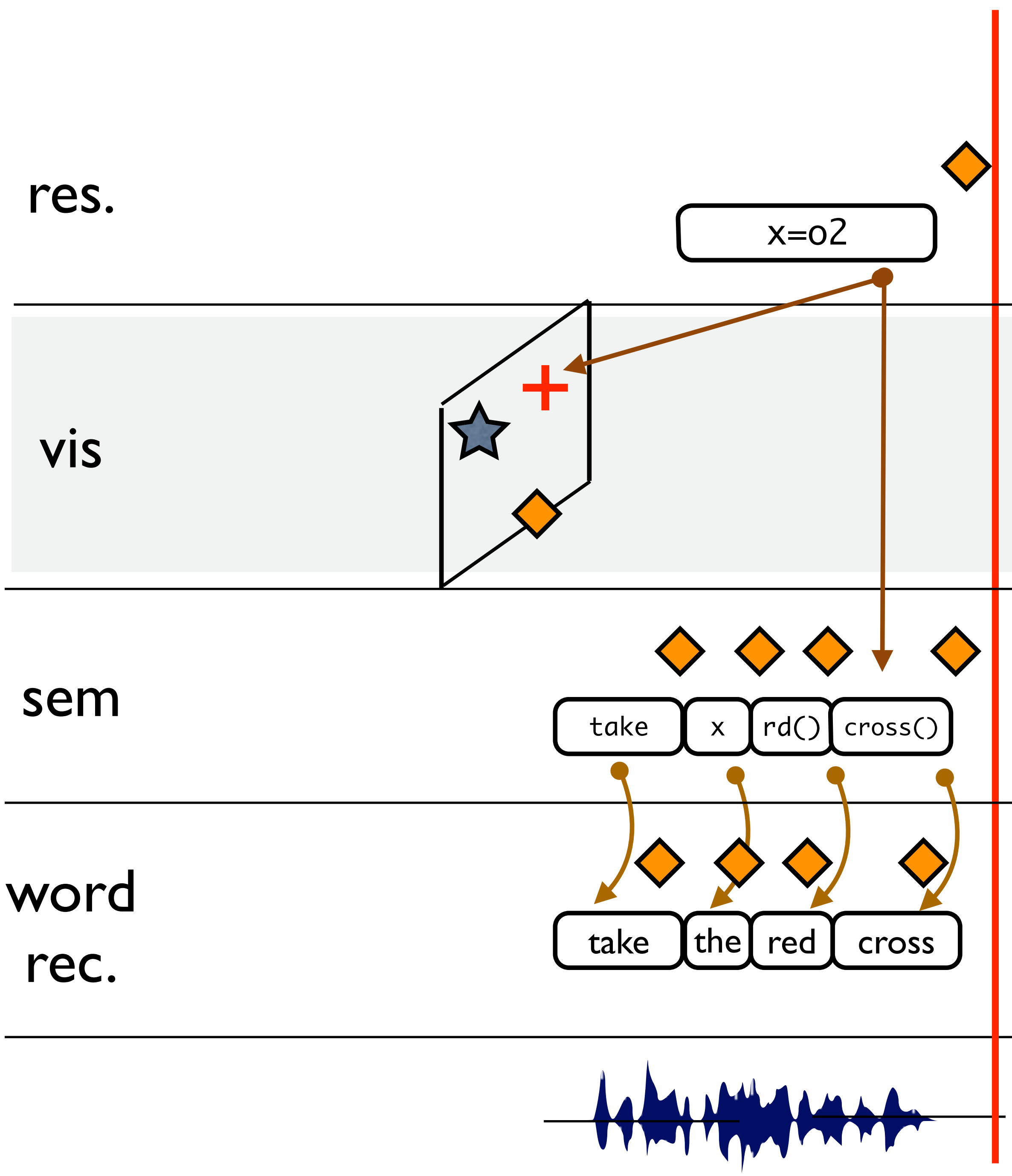


incremental statistical NLU



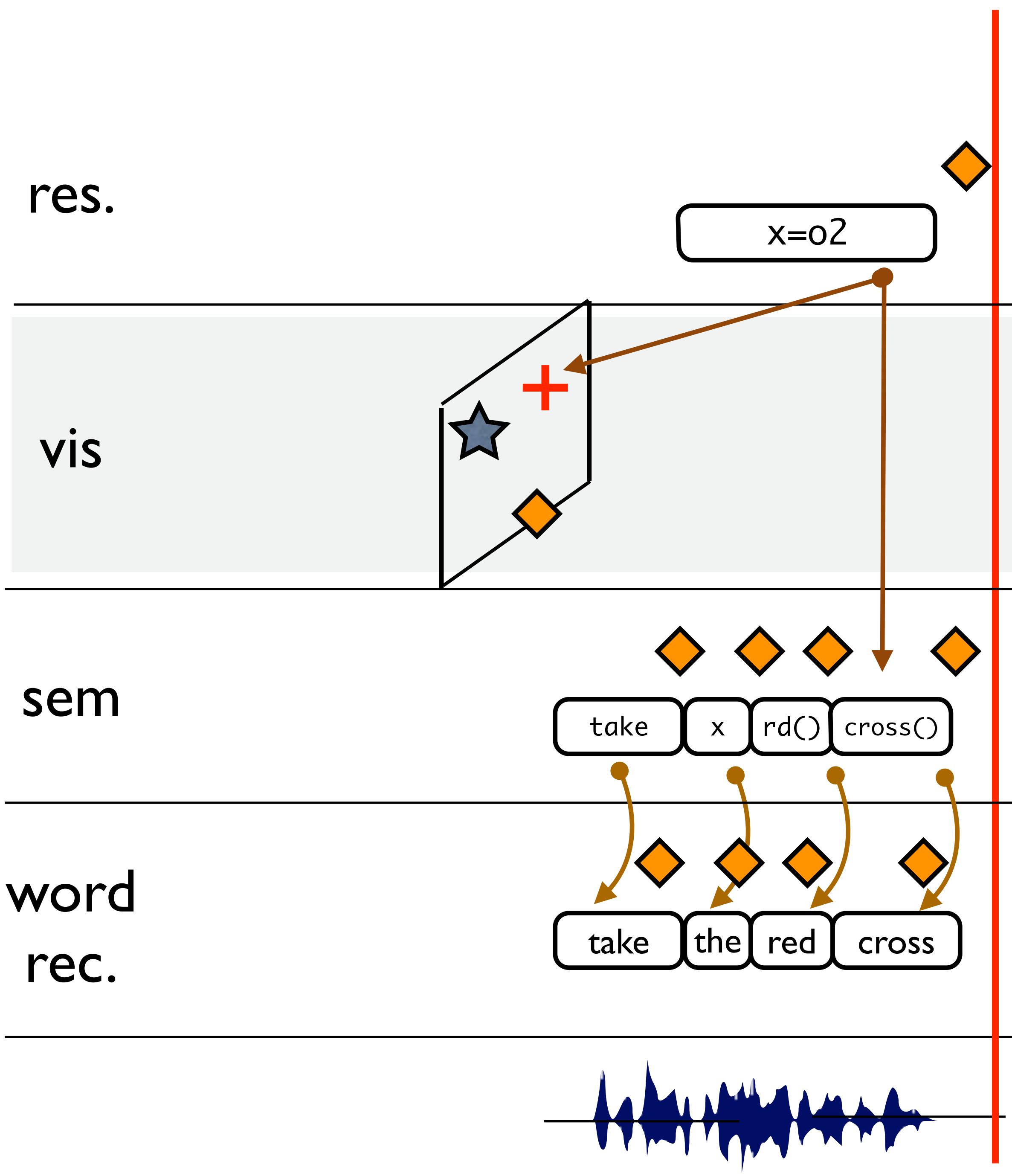
$$P(I|S,W)$$

incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\; P(I|W)}{P(S|W)}$$

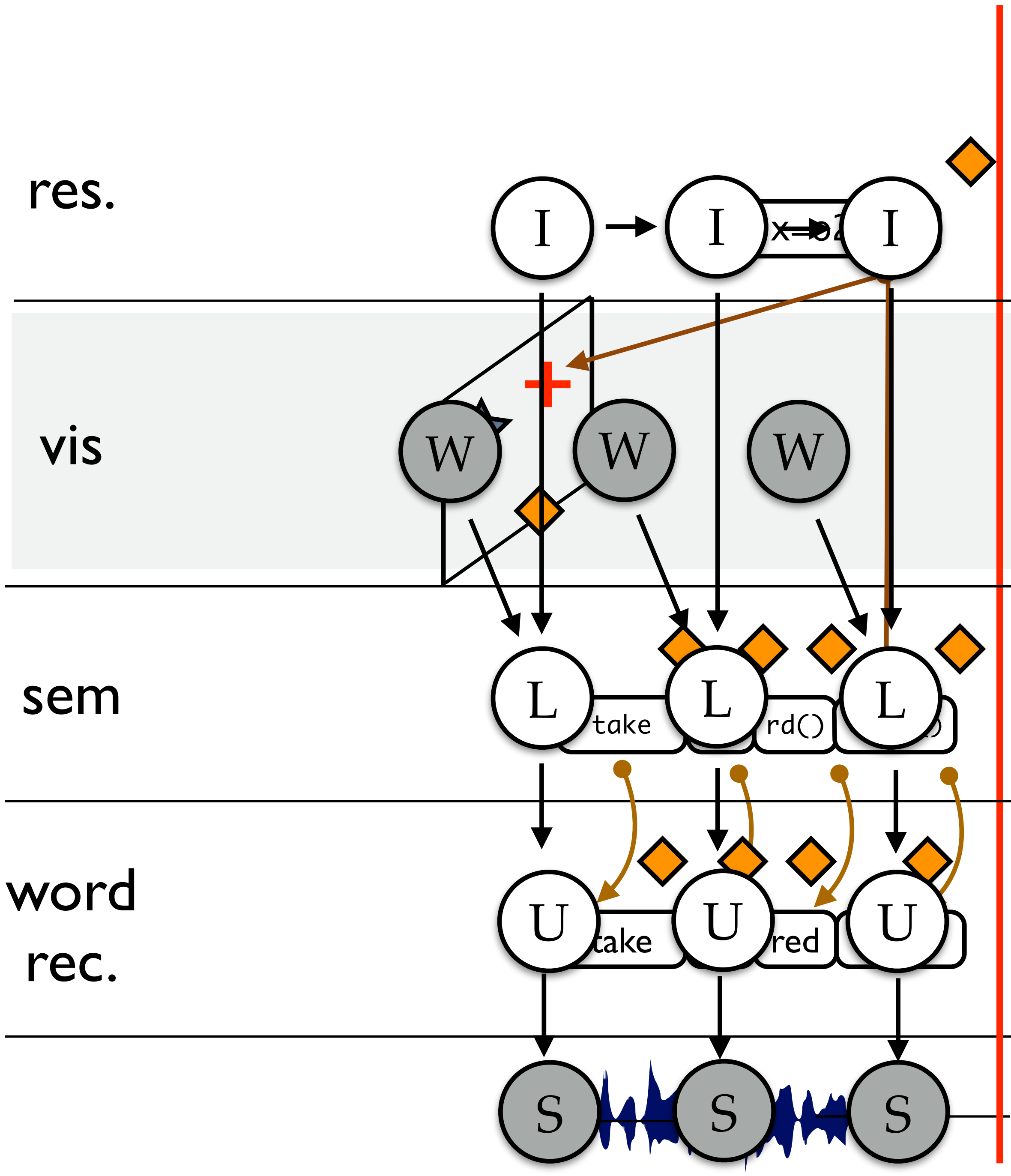
incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\; P(I|W)}{P(S|W)}$$

$$P(S \mid I, W)$$

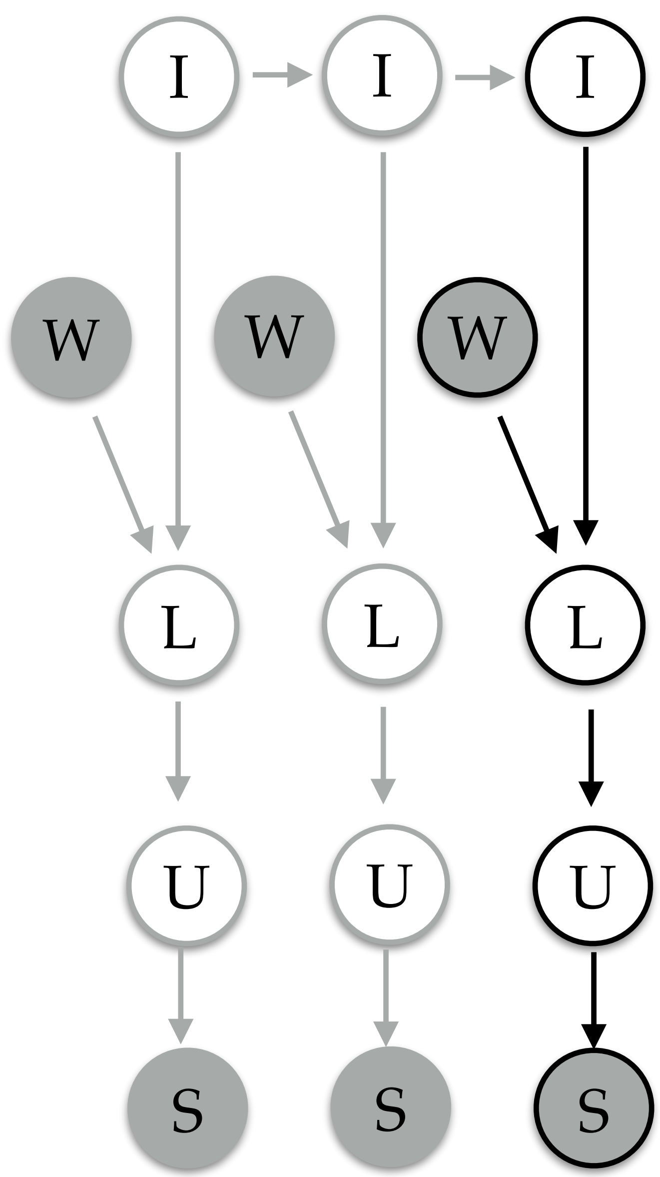
incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\, P(I|W)}{P(S|W)}$$

$$P(S \mid I, W)$$

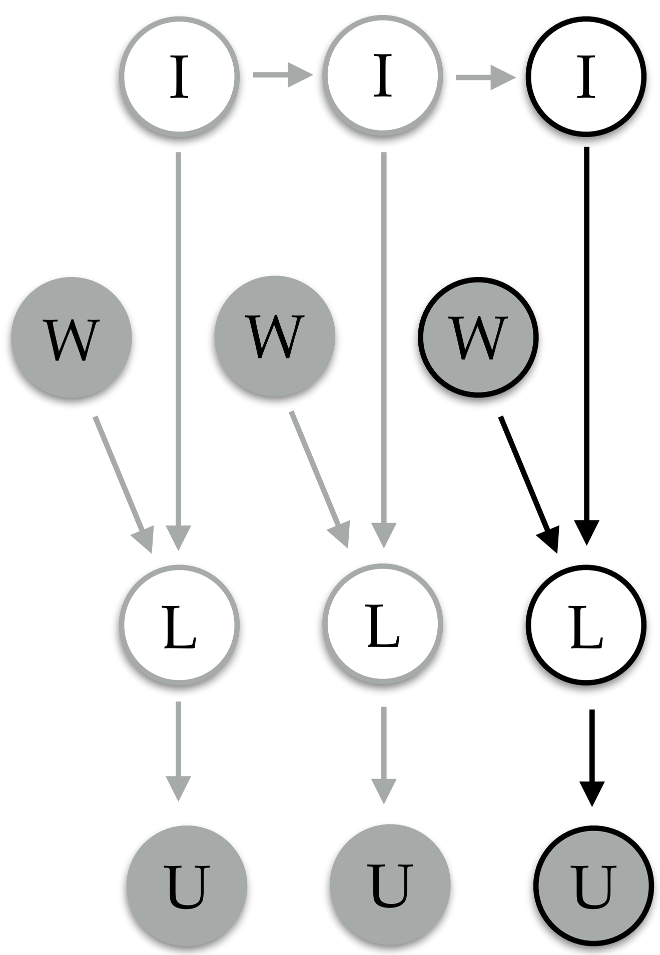
incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\ P(I|W)}{P(S|W)}$$

$$P(S \mid I, W)$$

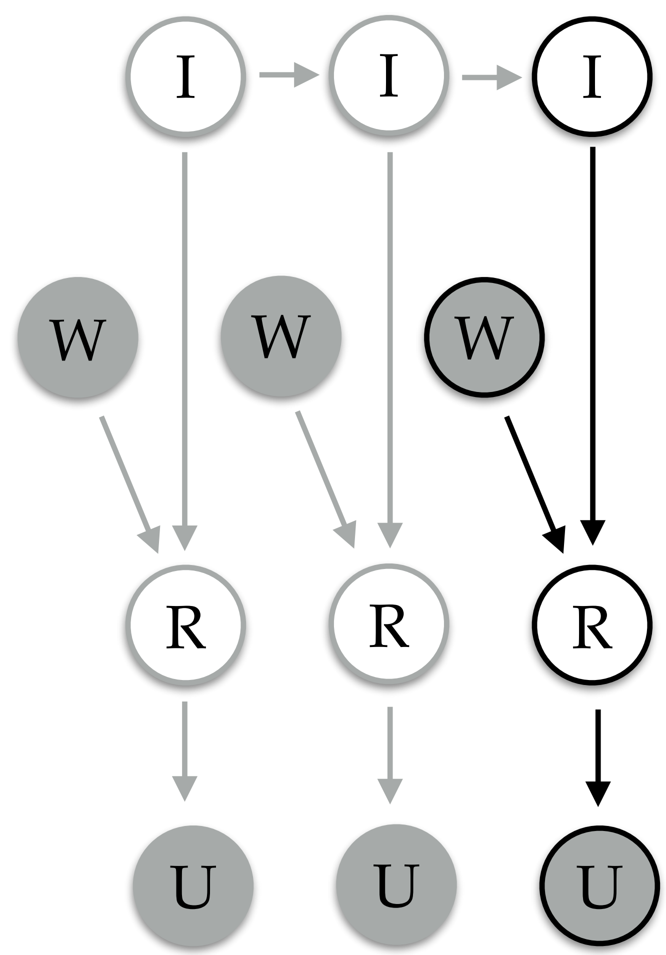
incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\ P(I|W)}{P(S|W)}$$

$$P(\ U \mid I,\ W\ )$$

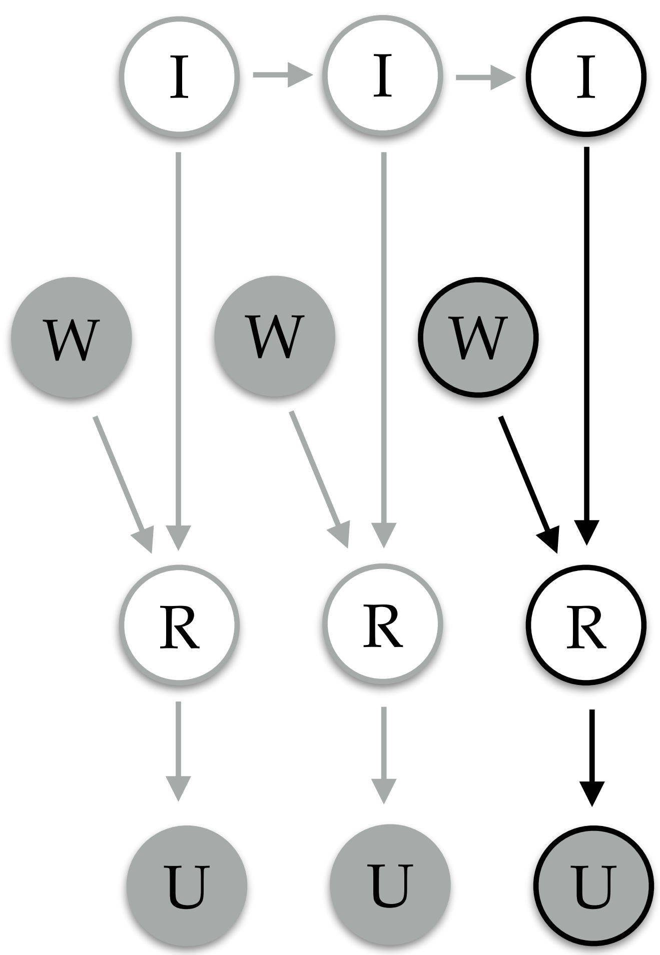
incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\; P(I|W)}{P(S|W)}$$

$$P(U|I, W) =$$
$$\sum_R P(U, R|I, W)$$

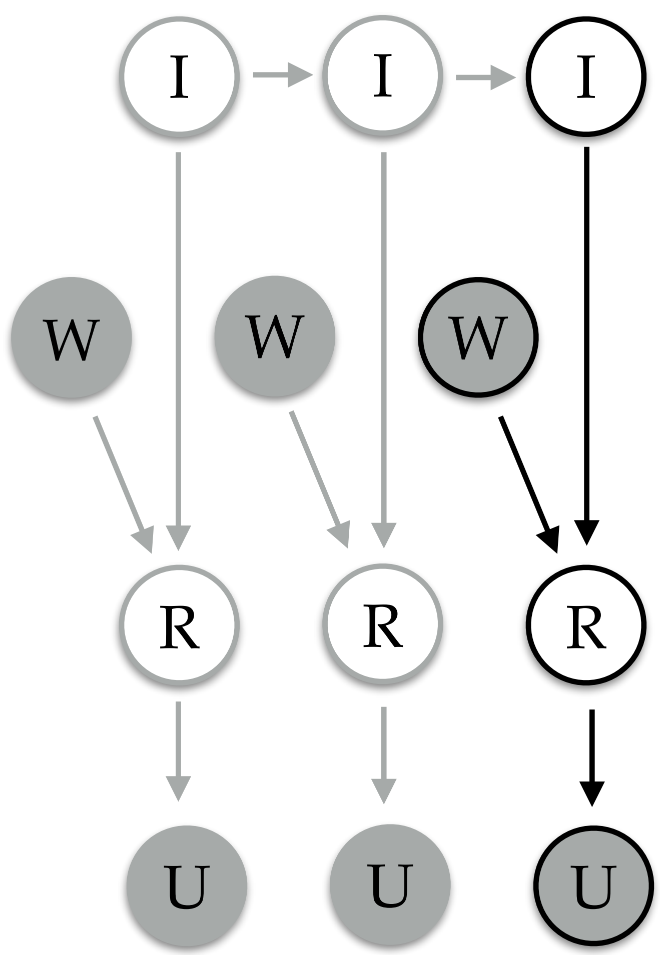
incremental statistical NLU



$$P(I \mid S, W) = \frac{P(S \mid I, W)\ P(I|W)}{P(S|W)}$$

$$P(U|I,W) = \\ \sum_R P(U,R|I,W) = \\ \sum_R P(U|R)P(R|I,W)$$

incremental statistical NLU



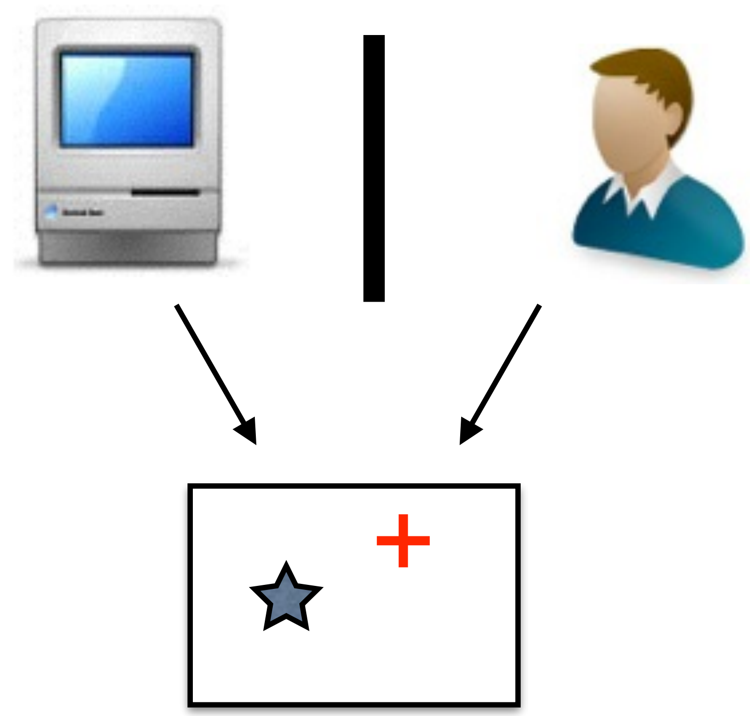
$$P(I|U,W) =$$

$$P(U|I,W)P(I|W)\frac{1}{P(U|W)} =$$

$$\sum_R P(U|R)P(R|I,W) \\ P(I|W)\frac{1}{P(U|W)}$$

incremental statistical NLU

– The Pento-2010 Data –

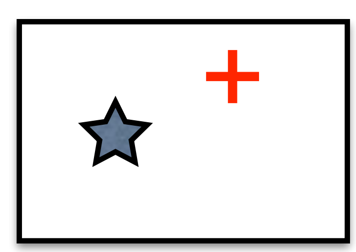


incremental statistical NLU

– The Pento-2010 Data –

o_1 shape:star, colour:gray,
x-pos:10, y-pos:20

o_2 shape:cross, colour:red,
x-pos:30, y-pos:10

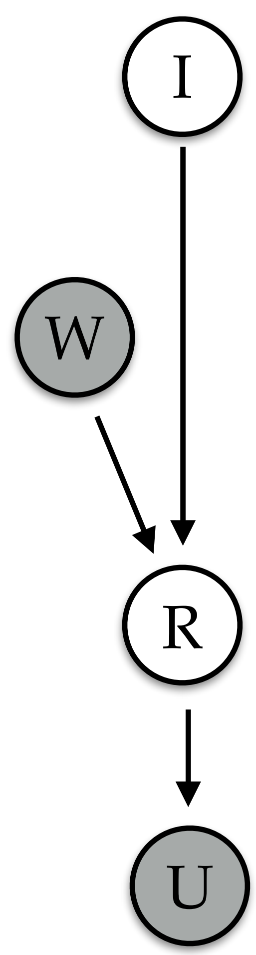


	o	o
star	1	0
cross	0	1
gray	1	0
red	0	1

incremental statistical NLU

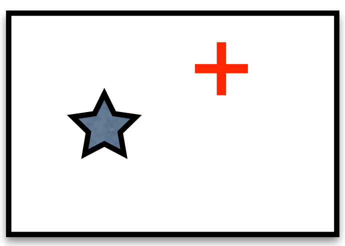
	o	o
star	1	0
cross	0	1
gray	1	0
red	0	1

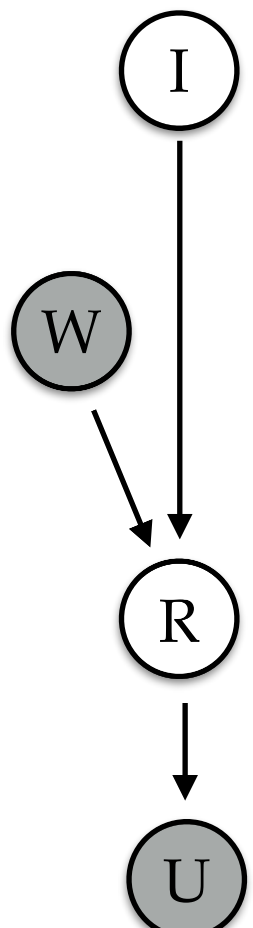
	o	o
star	.5	0
cross	0	.5
gray	.5	0
red	0	.5



$$P(I|U,W) = \frac{1}{P(U|W)} P(I|W) \sum_R P(R|I,W)P(U|R)$$

incremental statistical NLU



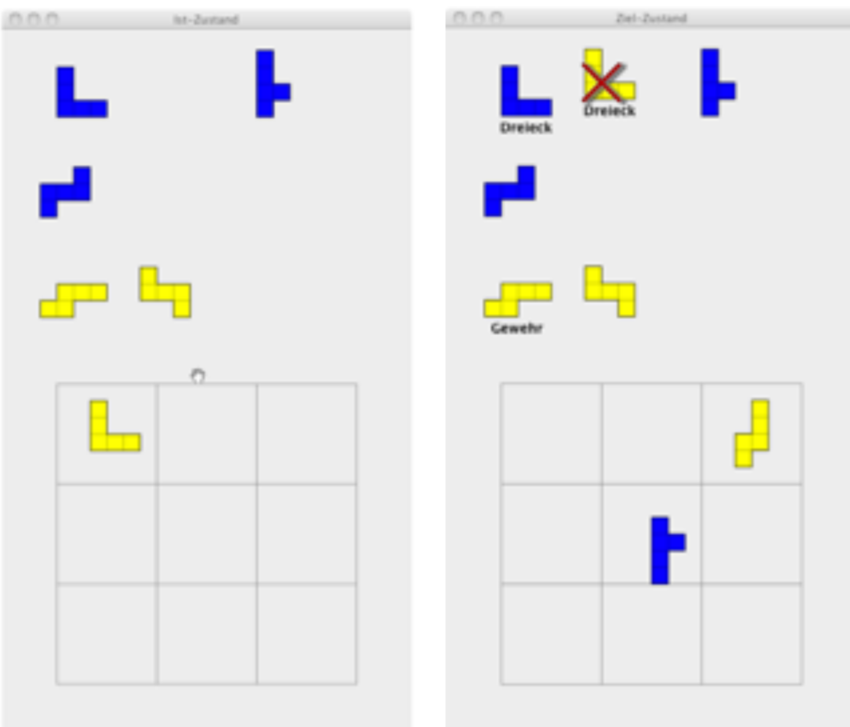


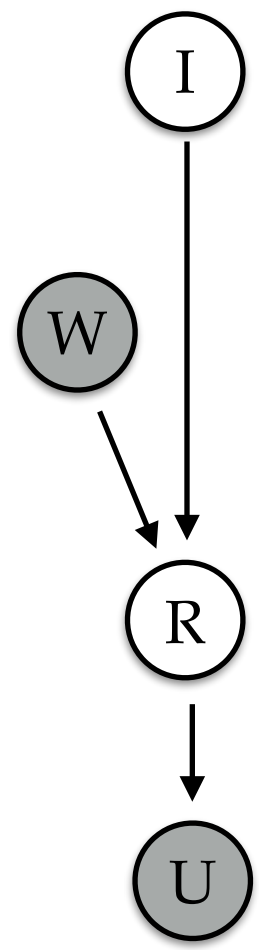
$$P(I = o_2 | U = \text{cross}, W) =$$
$$\frac{1}{P(U{=}\text{cross}|W)} P(I = o_2 | W)$$
$$\sum_R P(R | I = o_2, W) P(U = \text{cross} | R)$$

	o	o
star	.5	0
cross	0	.5
gray	.5	0
red	0	.5

incremental statistical NLU

– Results for Pento-2010 Data –



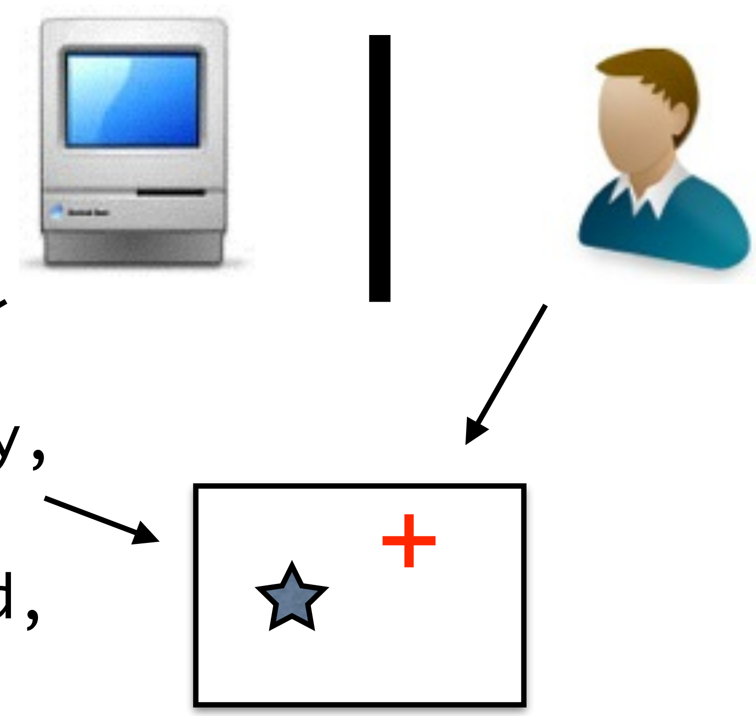


$$P(I|U,W) = \frac{1}{P(U|W)} P(I|W) \sum_R P(R|I,W) P(U|R)$$

	NB	ME	K	H	P
fscore	81.16	92.26	92.18	76.9	
	(74.5)	(89.4)	(86.8)		
slot	73.62	88.91	88.88		
	(66.4)	(85.1)	(81.6)		
frame	42.57	74.08	74.76		
	(34.2)	(67.2)	(61.2)		
action	80.05	93.62	92.62		
object	76.27	$\boxed{90.79}$	84.71		64.3
result	64.4	82.54	86.65		

incremental statistical NLU

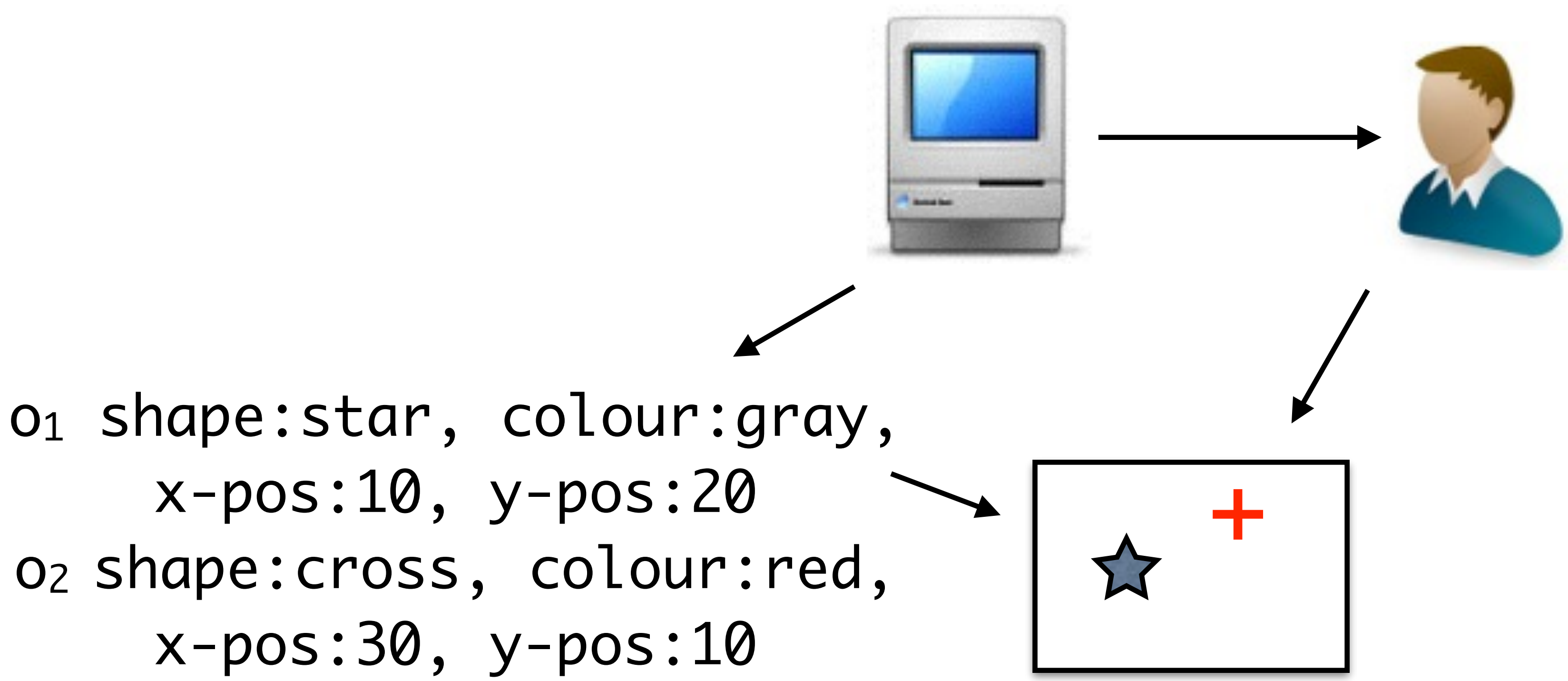
– The Pento-2010 Data –



o_1 `shape:star, colour:gray,`
`x-pos:10, y-pos:20`
o_2 `shape:cross, colour:red,`
`x-pos:30, y-pos:10`

incremental statistical NLU

– The Pento-2013 Data –



multimodal processing

– mint.tools / venice –

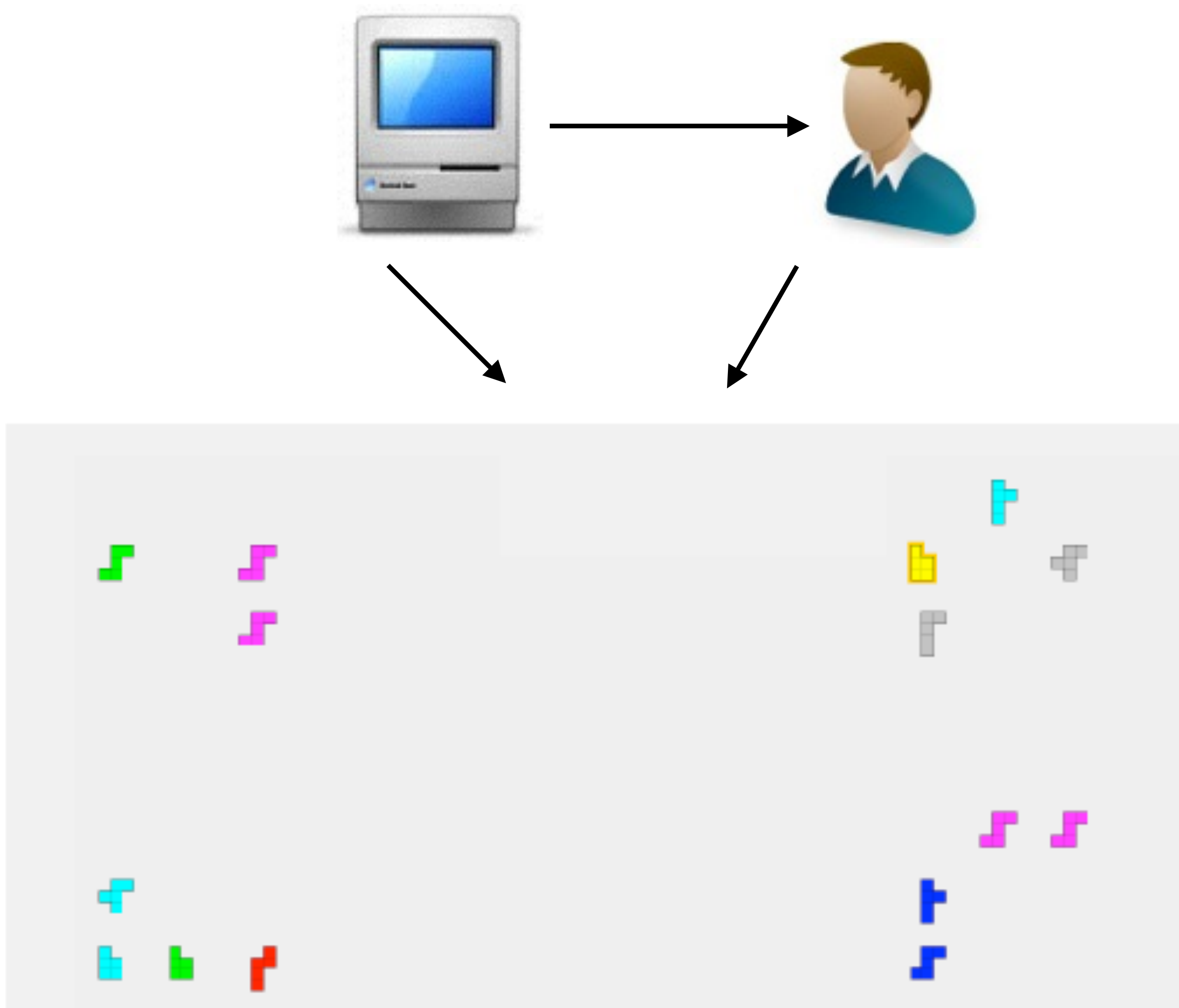
joint work with Spyros Kousidis

(Kousidis *et al.*, SIGdial 2013, 2014; ICMI 2014)

technical infrastructure for combining various sensors (motion capture, eye tracking) and various actuators (NAO robot) with InproTK

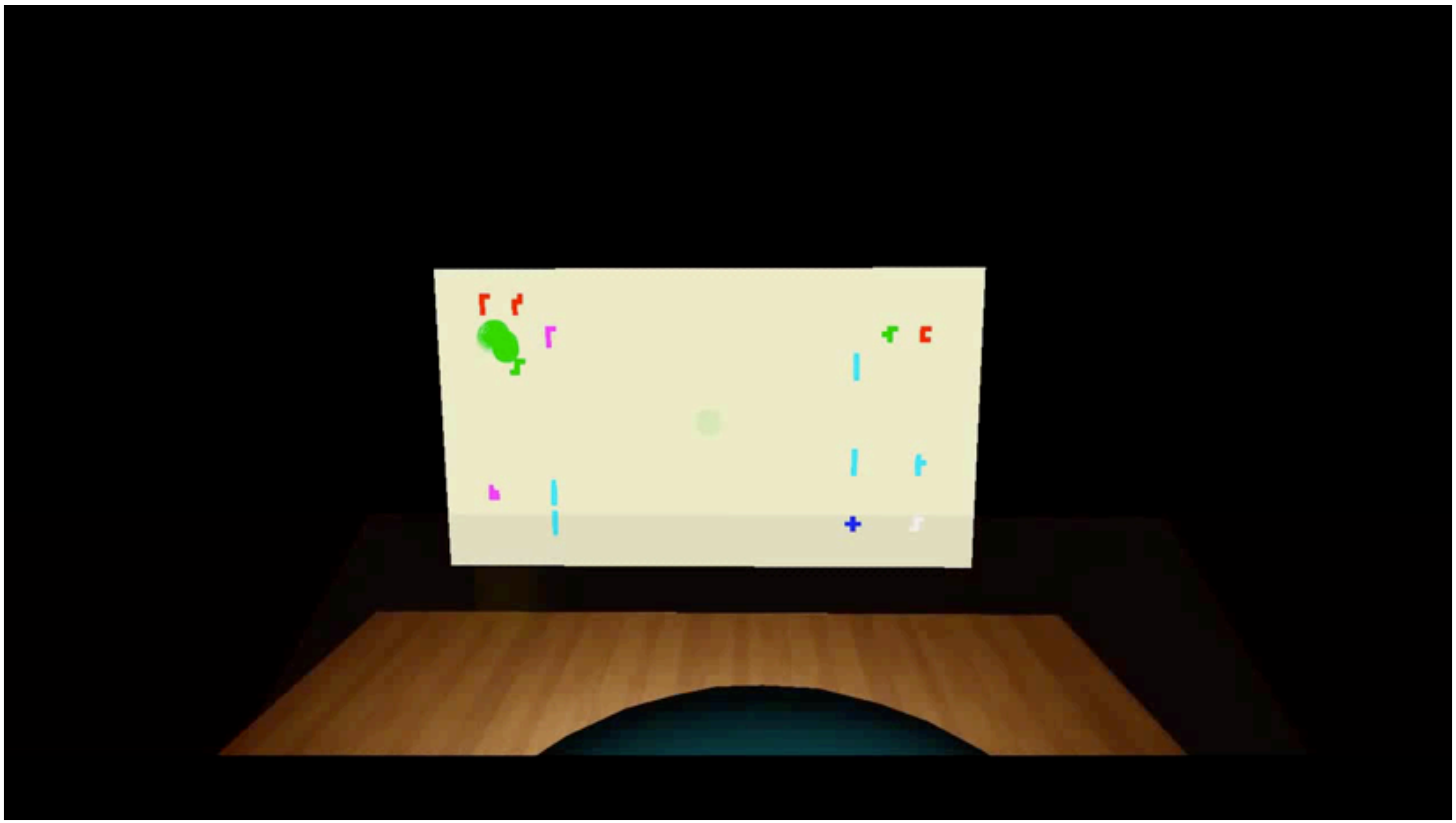
incremental statistical NLU

– The Pento-2013 Data –



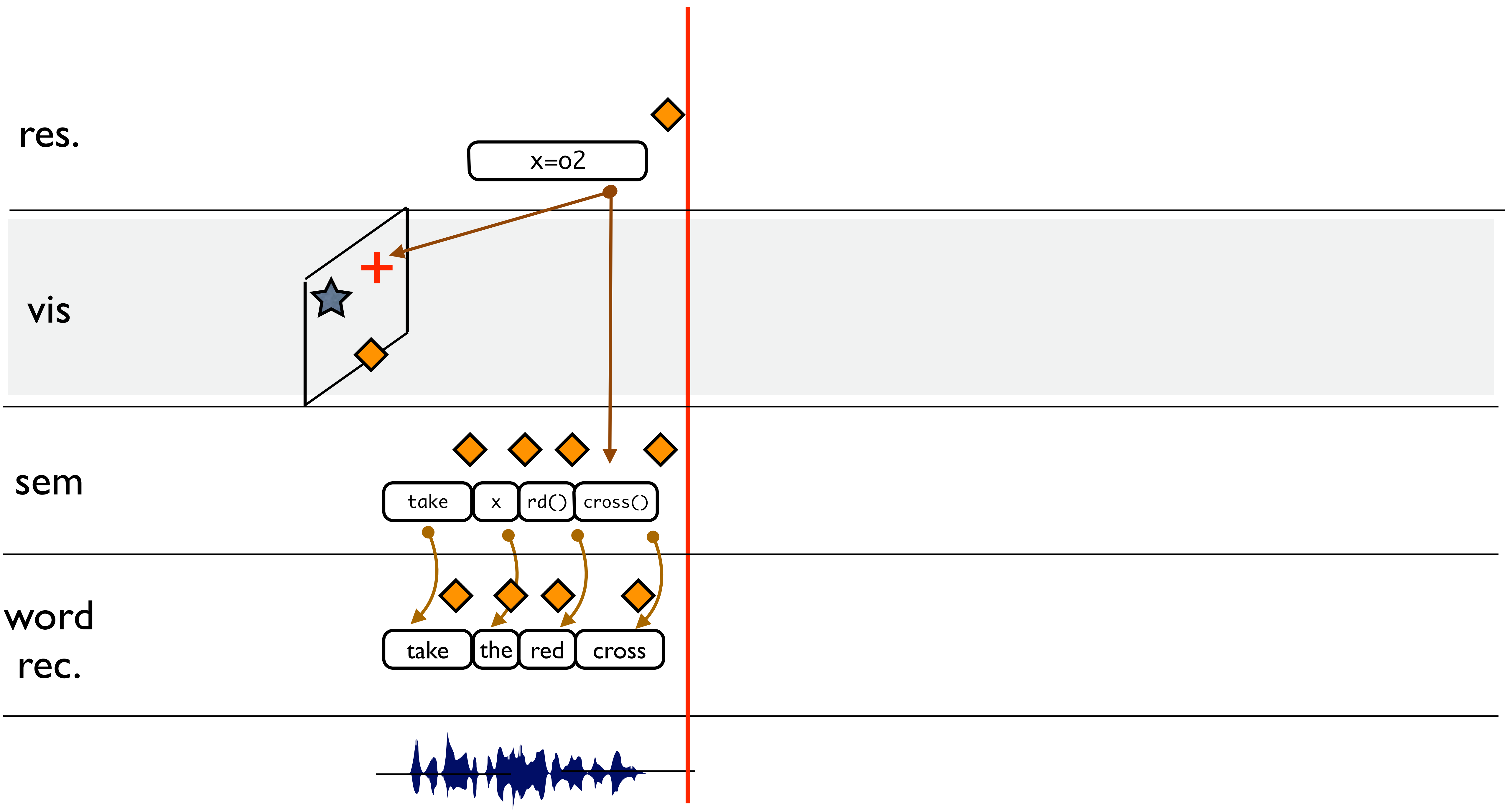
incremental statistical NLU

– The Pento-2013 Data –



incremental statistical NLU

– Adding Gaze & Gestures –



Properties over Time

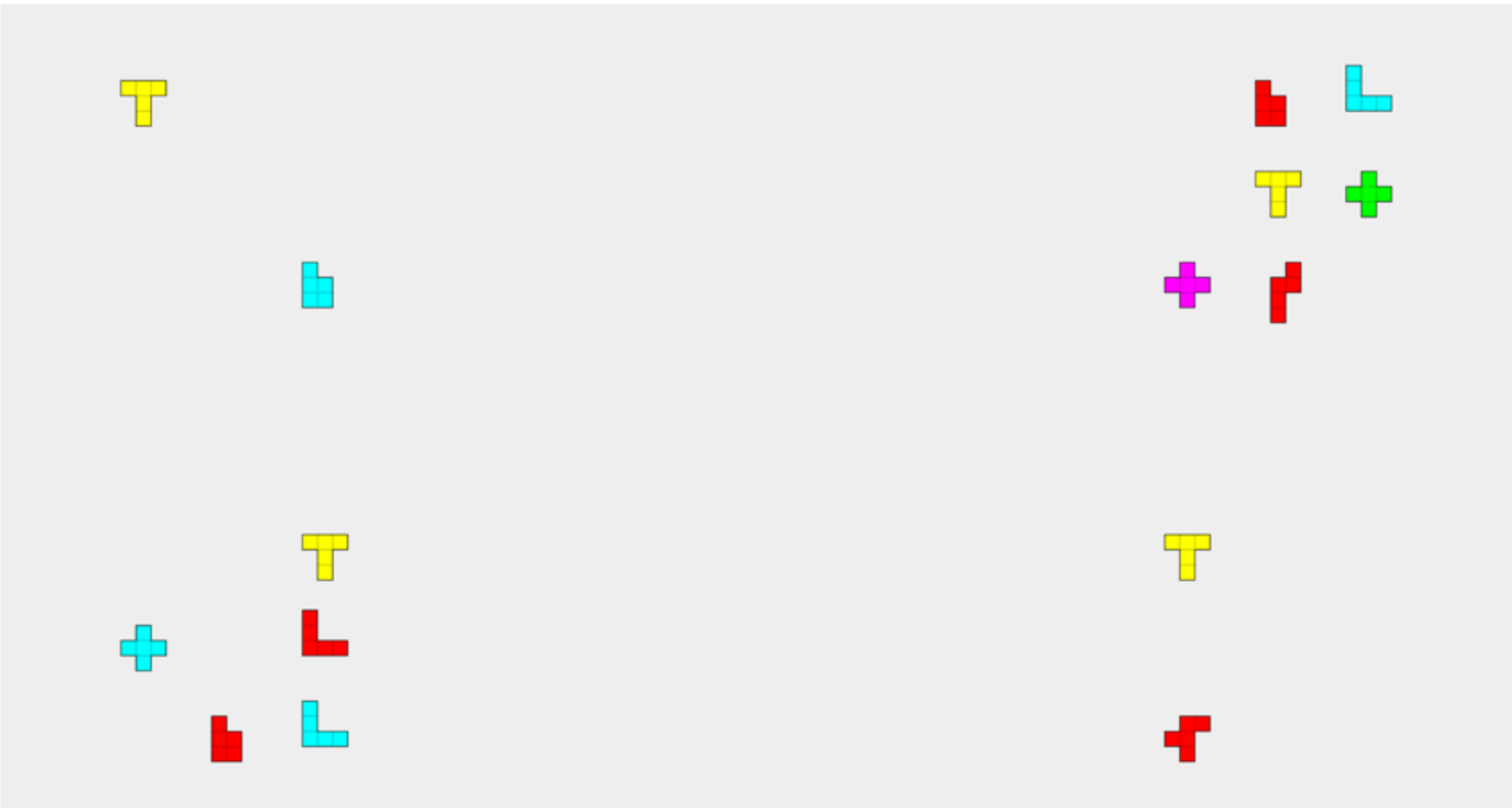
dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes

Properties over Time

dann	...	**nehmen**	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	**take**	...	second	t	a little	right	is ...	snout	plane	...	yes

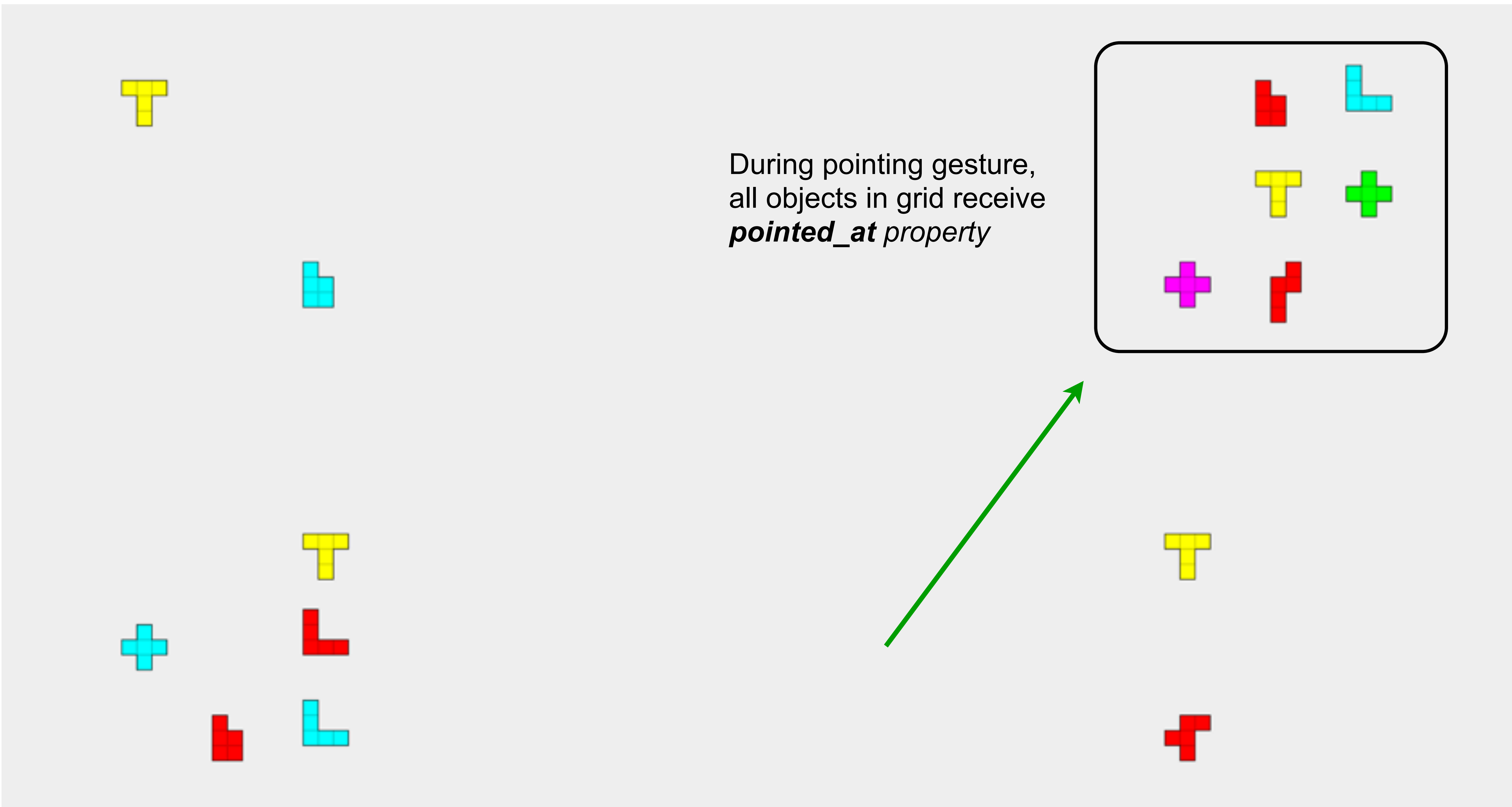
Properties over Time

dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes



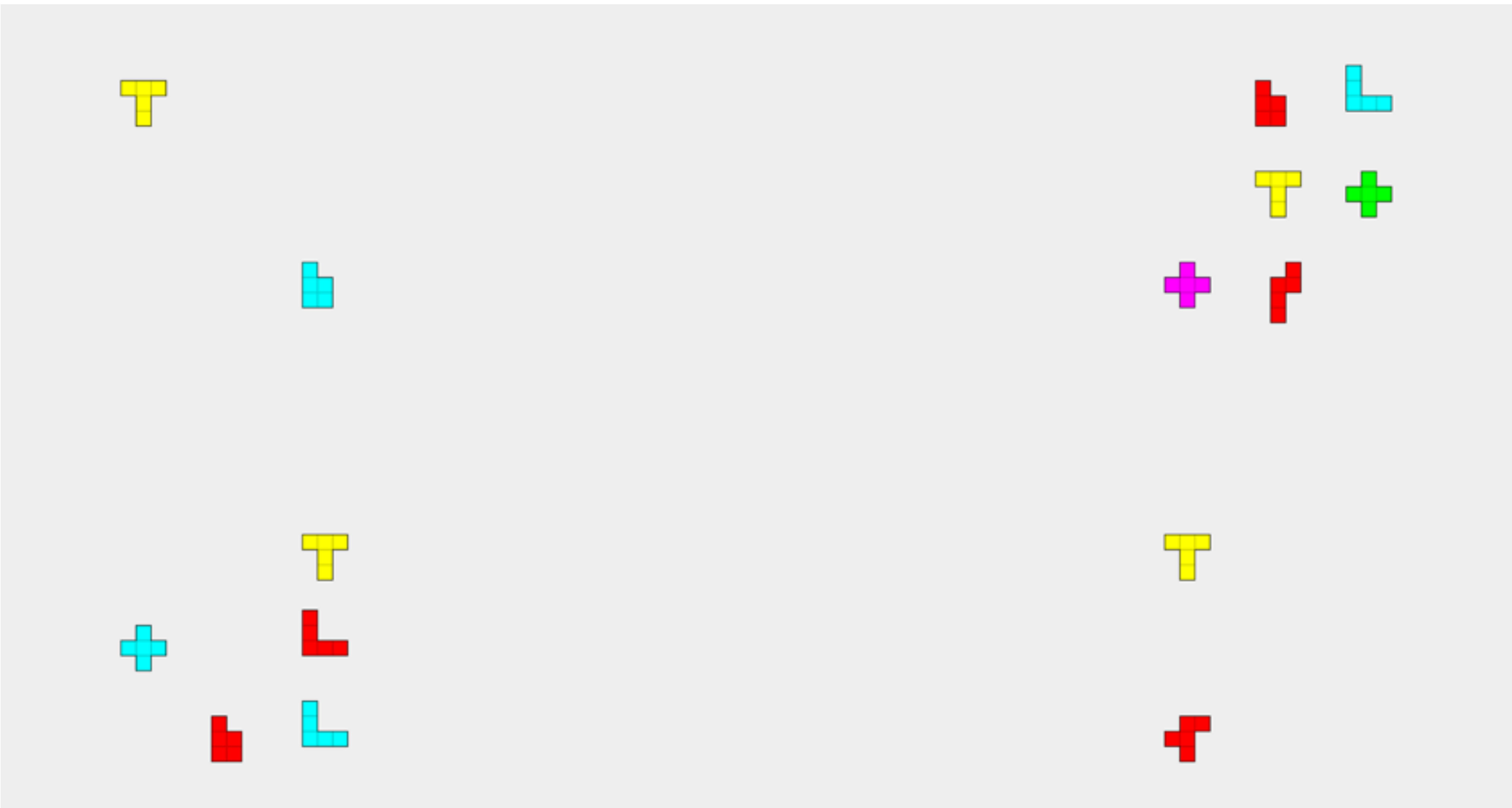
Properties over Time

dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes



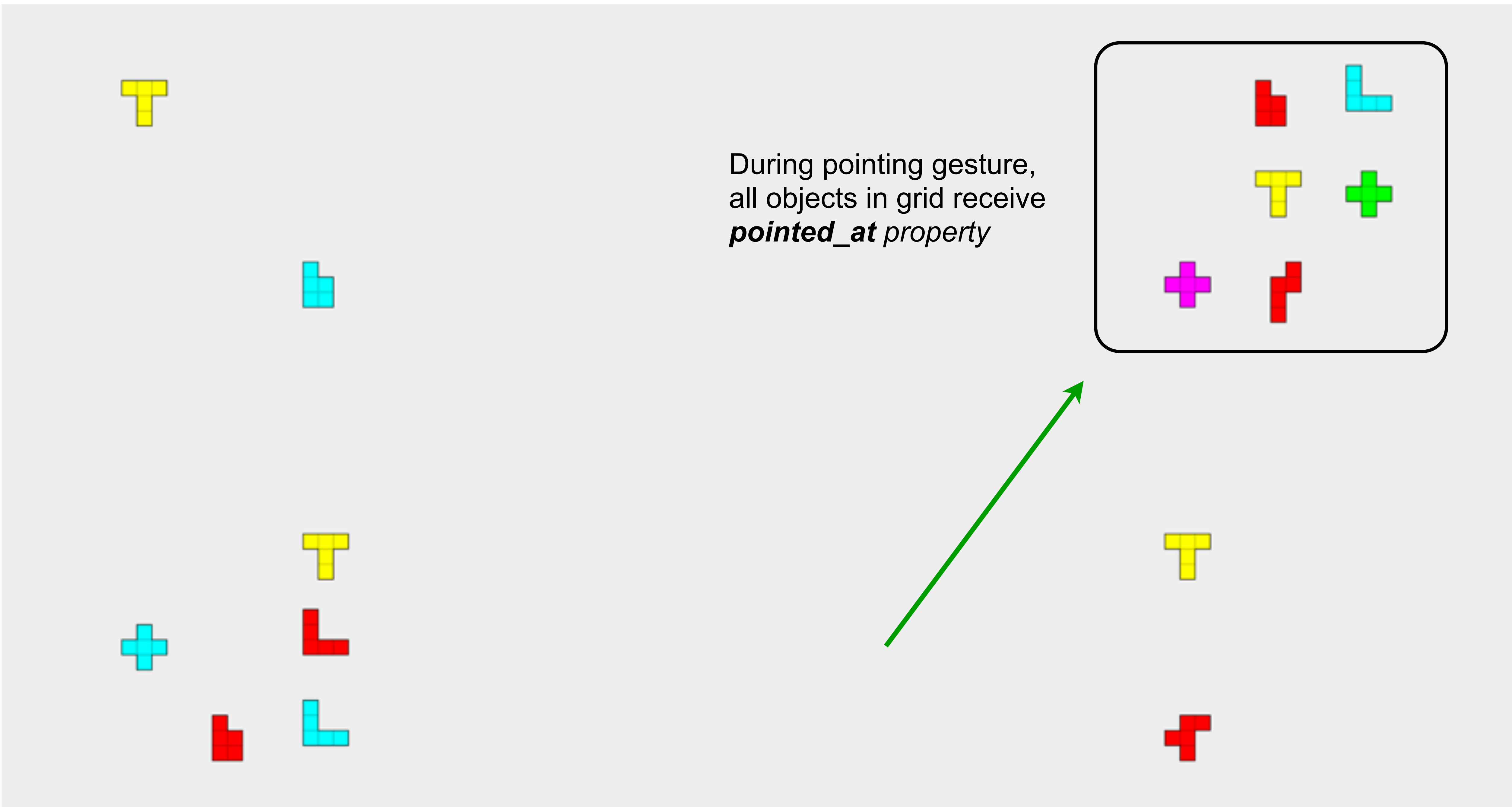
Properties over Time

dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes



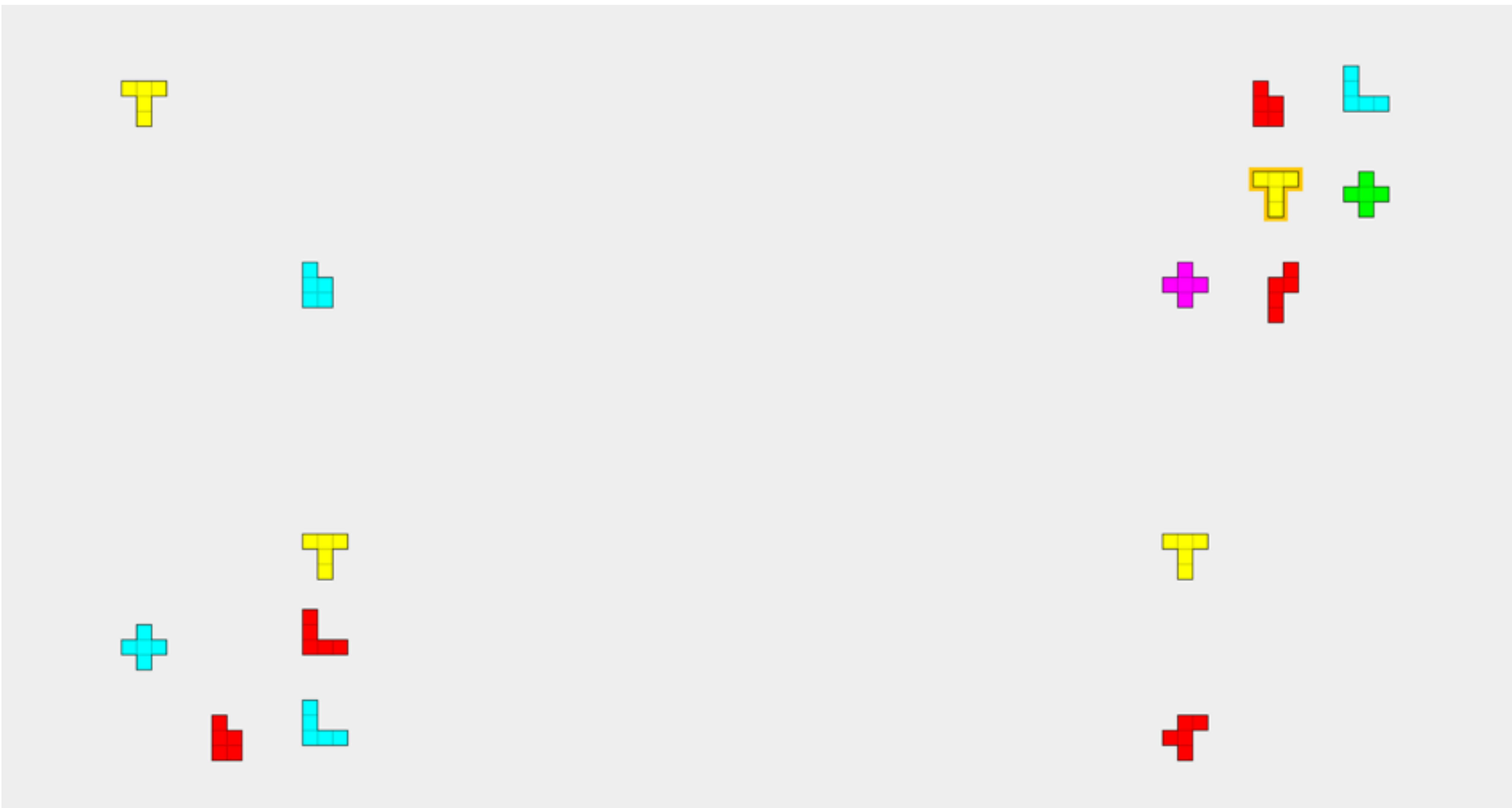
Properties over Time

dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes



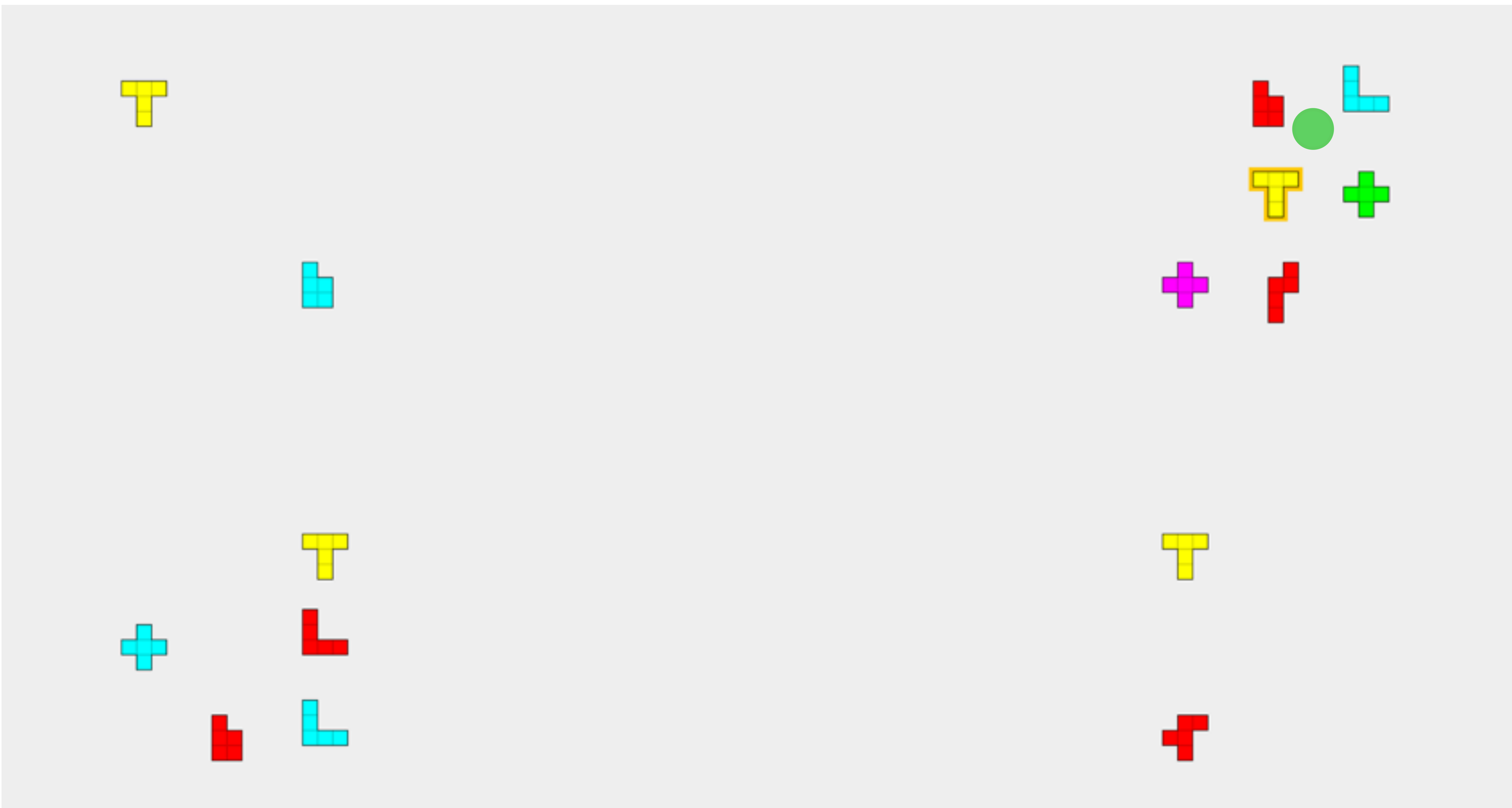
Properties over Time

dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes



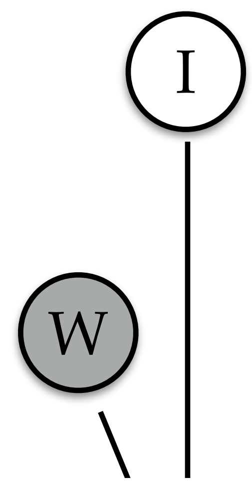
Properties over Time

dann	...	nehmen	...	zweite	t	bisschen	rechts	ist ...	rüssel	hobel	...	ja
then	...	take	...	second	t	a little	right	is ...	snout	plane	...	yes



incremental statistical NLU

– Results for Pento-2013 Data –



$$P(I|U,W) = \frac{1}{P(U|W)} P(I|W) \sum_R P(R|I,W) P(U|R)$$

Version	Acc	Top 2	Top 4
Gaze	18%		
(baseline) NLU	50%	59%	77%
NLU + Gaze	53%	62%	80%
NLU + Point	52%	65%	90%
NLU + Gaze + Point	53%	**70%**	**91%**
NLU + Gaze-F	53%	65%	78%
NLU + Point-F	**57%**	68%	88%
NLU+Gaze-F+Point-F	56%	69%	85%

incremental statistical NLU

– The Pento-2013 Data –

```
o1 shape:star, colour:gray,
     x-pos:10, y-pos:20
o2 shape:cross, colour:red,
     x-pos:30, y-pos:10
```

incremental statistical NLU

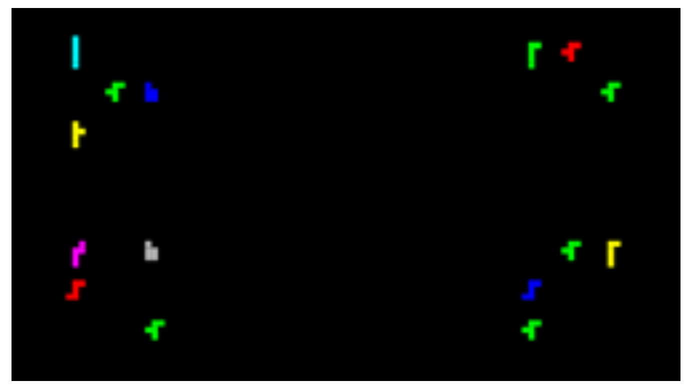
– The Pento-2013 Data, via CV –

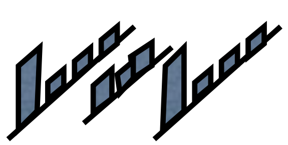
joint work with Liva Dia

incremental statistical NLU

– The Pento-2013 Data, via CV –

o_1 shape:star, colour:gray,
x-pos:10, y-pos:20
o_2 shape:cross, colour:red,
x-pos:30, y-pos:10,
. . .



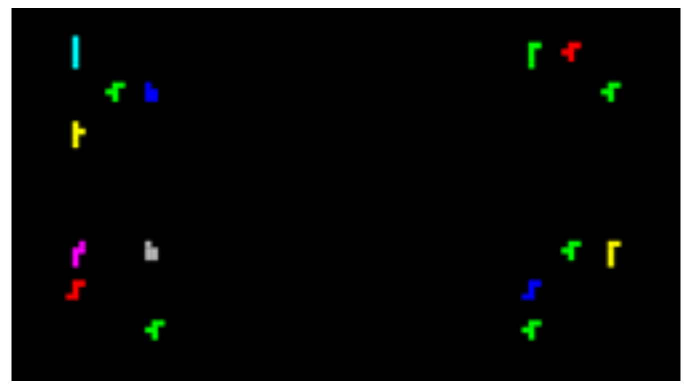


. . .

incremental statistical NLU

– The Pento-2013 Data, via CV –

o_1 shape:star, colour:gray,
x-pos:10, y-pos:20
o_2 shape:cross, colour:red,
x-pos:30, y-pos:10,
. . .



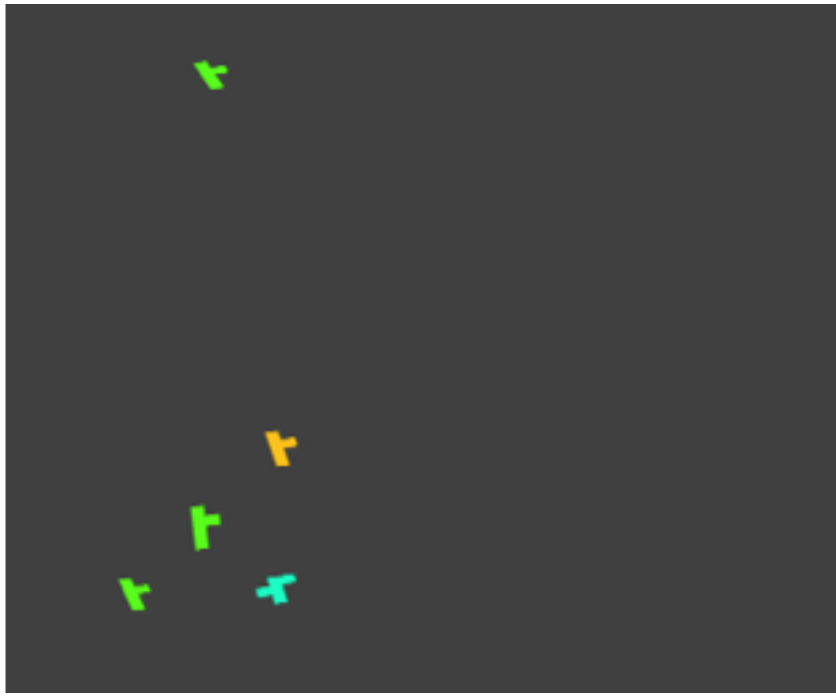
. . .

o_1 shape:star, colour:gray,
x-pos:10, y-pos:20
o_2 shape:cross, colour:red,
x-pos:30, y-pos:10,
. . .

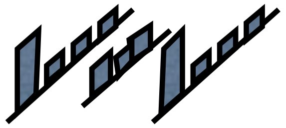
incremental statistical NLU

– The Pento-2013 Data, via CV –

```
o1 shape:star, colour:gray,
     x-pos:10, y-pos:20
o2 shape:cross, colour:red,
     x-pos:30, y-pos:10,
         . . .
```



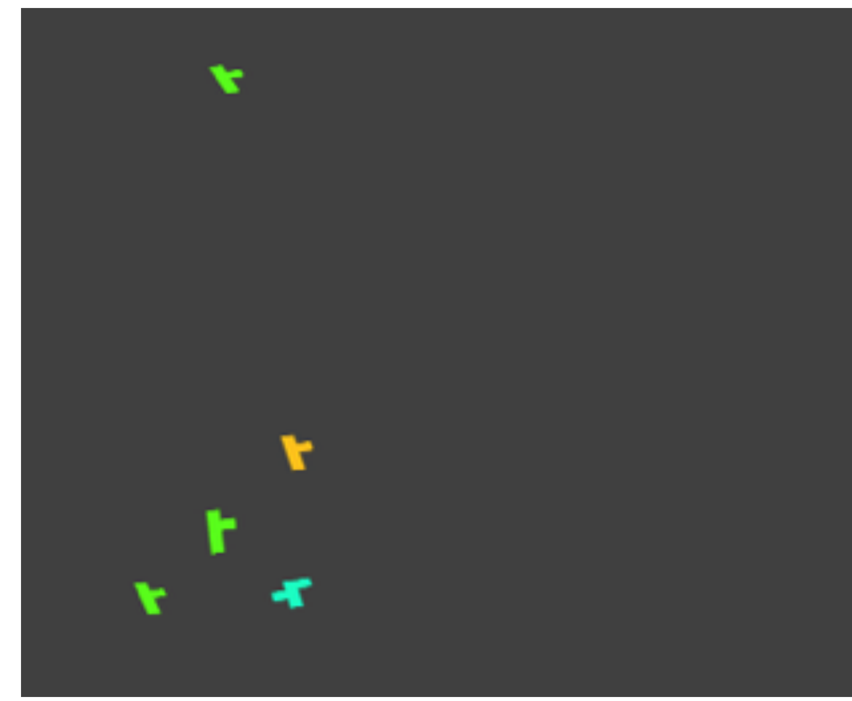
. . .



incremental statistical NLU

– The Pento-2013 Data, via CV –

	o	o
star	1	0
cross	0	1
gray	1	0
red	0	1



	o	o
star	0.8	0.1
cross	0.2	0.9
gray	0.9	0.4
red	0.1	0.6

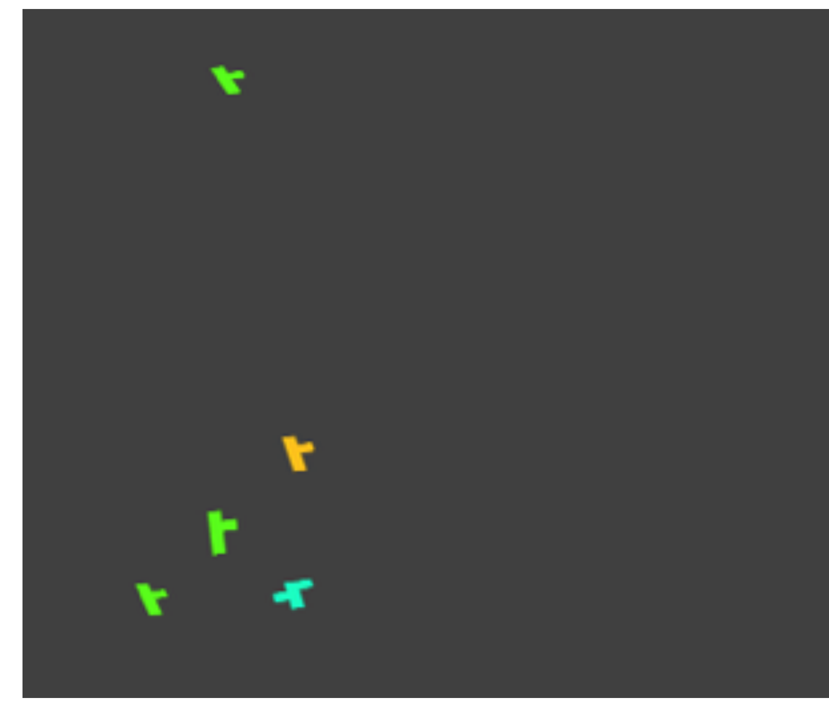
...

	o	o
star	0.4	0.05
cross	1	0.45
gray	0.45	0.2
red	0.05	0.3

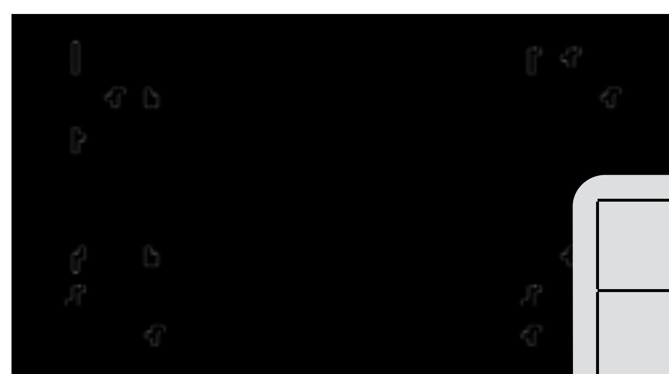
incremental statistical NLU

– The Pento-2013 Data, via CV –

	o	o
star	1	0
cross	0	1
gray	1	0
red	0	1



	o	o
star	0.8	0.1
cross	0.2	0.9
gray	0.9	0.4
red	0.1	0.6



Version	Acc
Base	0.78
CV orig, argmax	0.77
CV orig, distr	0.75

	o	o
star	0.4	0.05
cross	1	0.45
gray	0.45	0.2
red	0.05	0.3

. . .

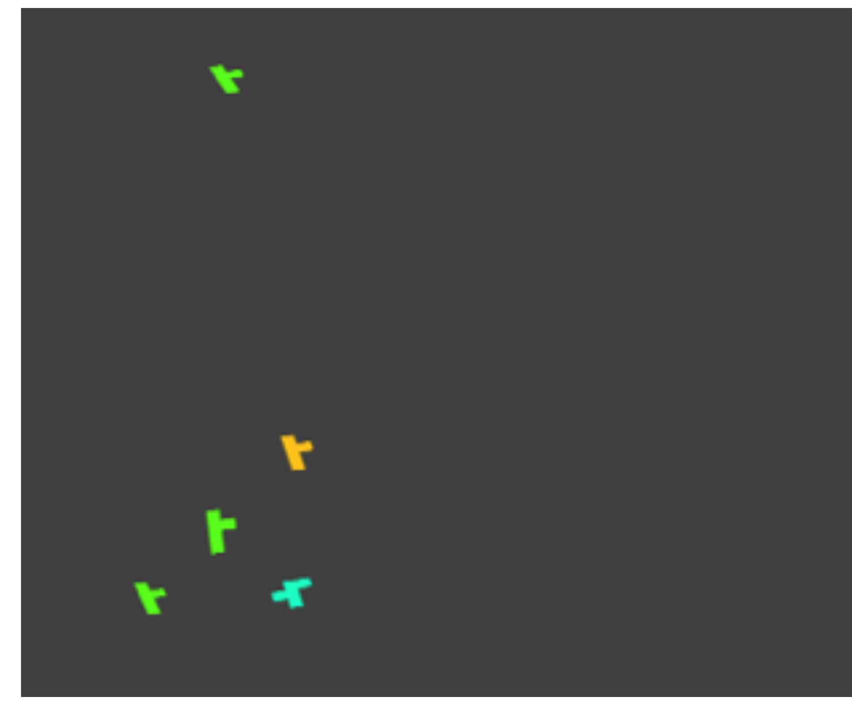
incremental statistical NLU

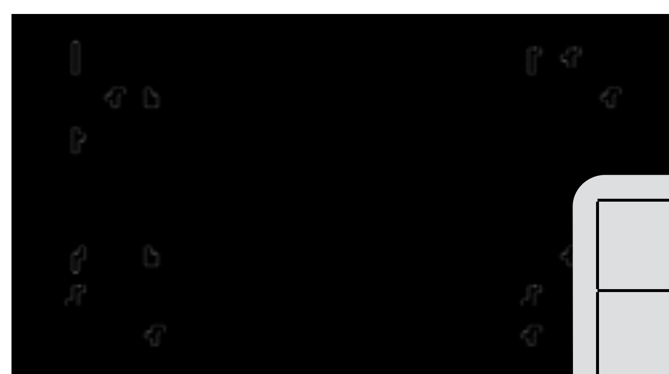
– The Pento-2013 Data, via CV –

	o	o
star	1	0
cross	0	1
gray	1	0
red	0	1

	o	o
star	0.8	0.1
cross	0.2	0.9
gray	0.9	0.4
red	0.1	0.6

	o	o
star	0.4	0.05
cross	1	0.45
gray	0.45	0.2
red	0.05	0.3



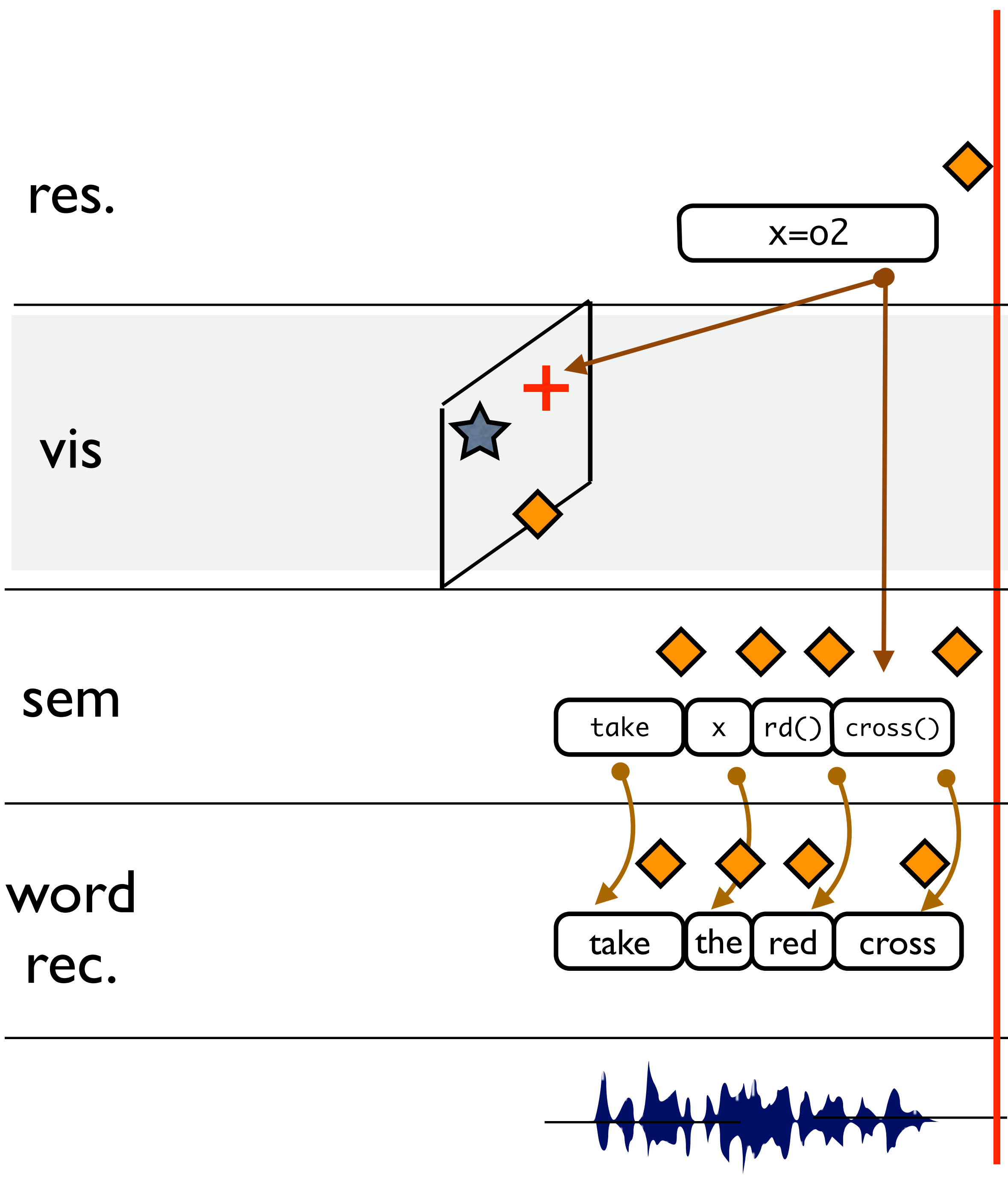


...

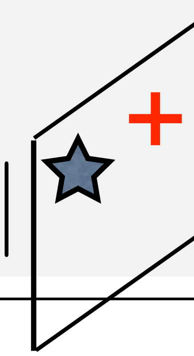
Version	Acc
Base	0.78
CV orig, argmax	0.77
CV orig, distr	0.75
CV filt, argmax	0.59
CV filt, distr	0.60

incremental statistical NLU

– generative vs. discriminative –

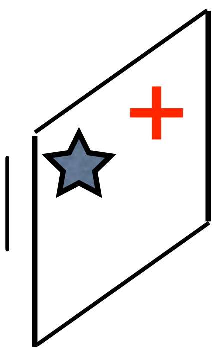


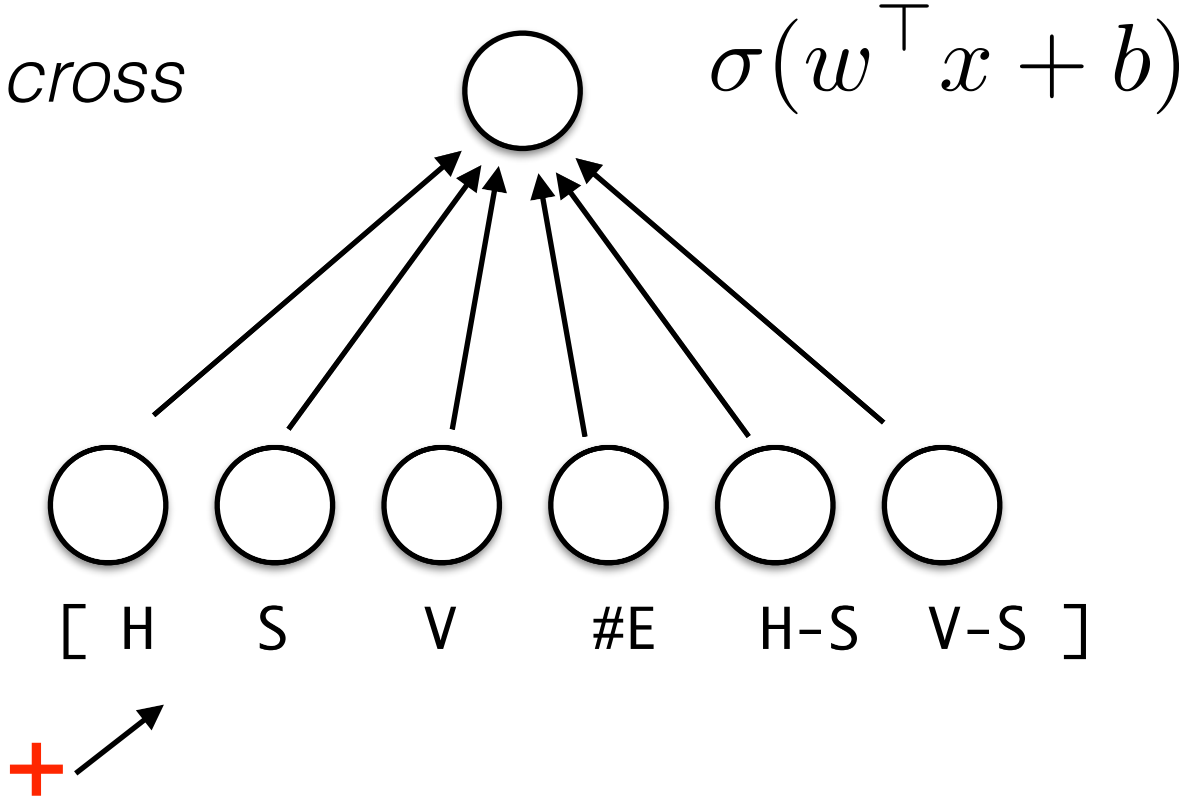
$$P(I \mid S, W) = \frac{P(S \mid I, W)\ P(I|W)}{P(S|W)}$$

$$P(I \mid S, W) = P_W(I \mid$$  $$)$$

incremental statistical NLU

– discriminative model –

$P(I \mid S, W) = P_W(I \mid$  $)$

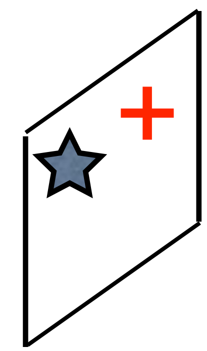


Training:

* for each occurence of word in corpus:
  * pair it with (features of) object utterance referred to [positive example]
  * pair with (features of) *n* objects in scene it didn't refer to [negative examples]
  * train (binary) logistic regression classifier for word

incremental statistical NLU

– discriminative model –

$P(I \mid S, W) = P_W(I \mid$  $)$

Testing:

* for each word of utterance:
  * test each object in scene against classifier of this word
  * update distribution from previous step with distribution from this step

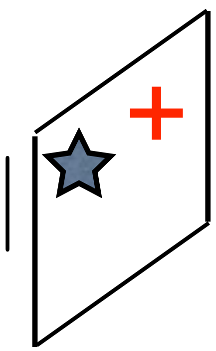
Training:

* for each occurence of word in corpus:
  * pair it with (features of) object utterance referred to [positive example]
  * pair with (features of) *n* objects in scene it didn't refer to [negative examples]
  * train (binary) logistic regression classifier for word

incremental statistical NLU

– discriminative model –

$$P(I \mid S, W) = P_W(I \mid \quad)$$



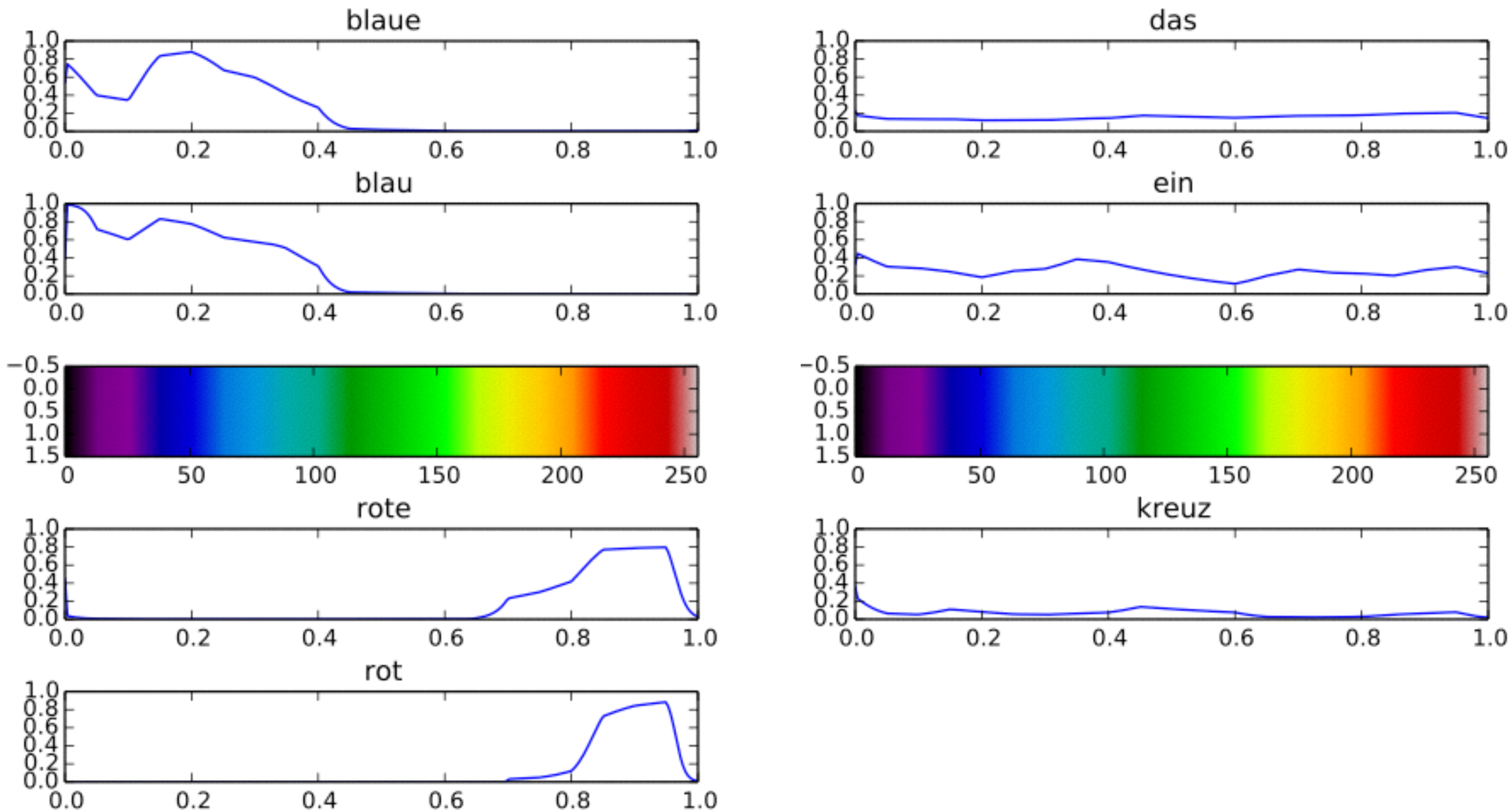
Version	Acc
Base	0.78
CV orig, argmax	0.77
CV orig, distr	0.75
CV filt, argmax	0.59
CV filt, distr	0.60
discr, CV filt & distr.	**0.66**

Training:

* for each occurence of word in corpus:
  * pair it with (features of) object utterance referred to [positive example]
  * pair with (features of) *n* objects in scene it didn't refer to [negative examples]
  * train (binary) logistic regression classifier for word

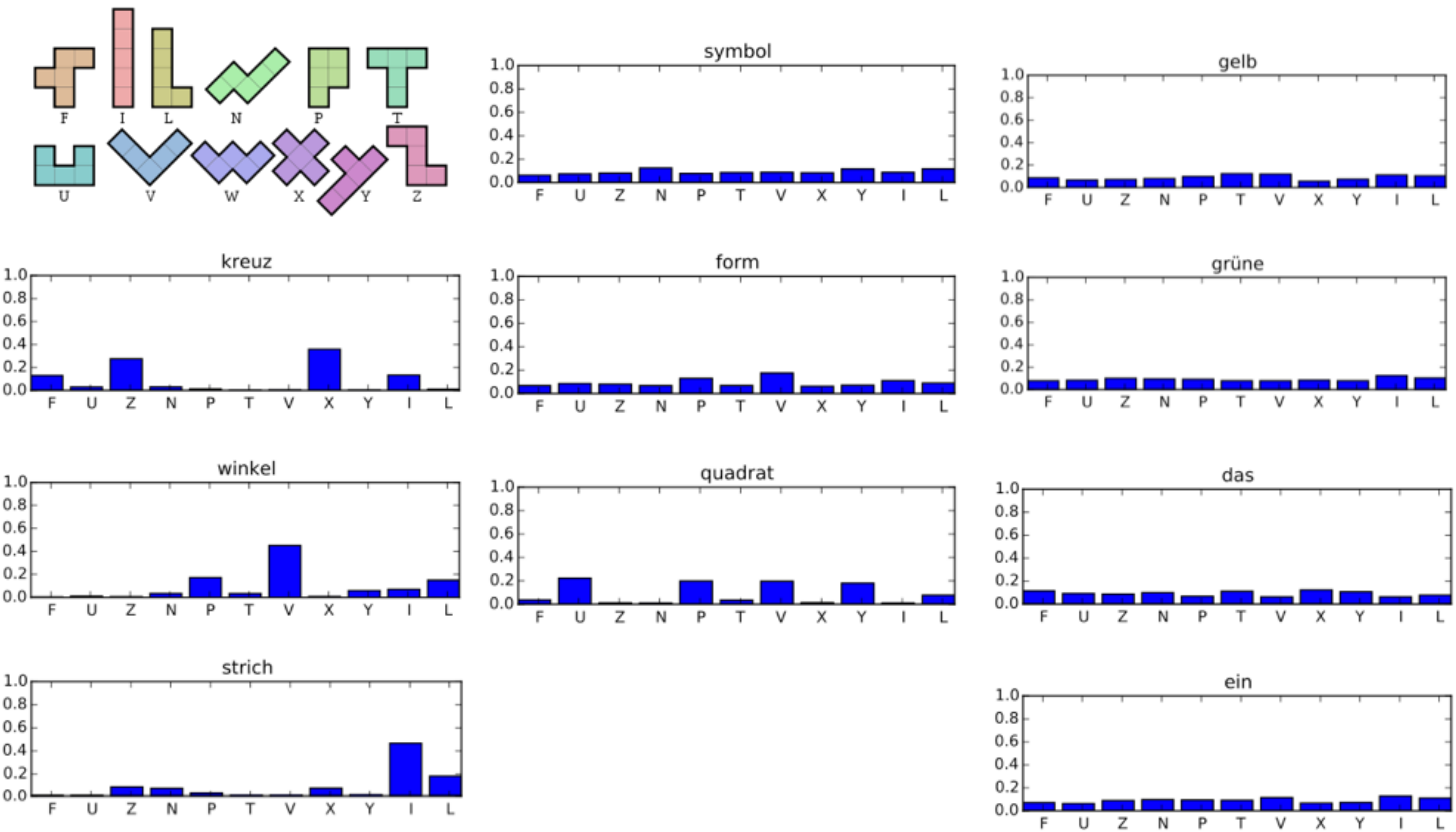
incremental statistical NLU

– discriminative model.. word semantics? –



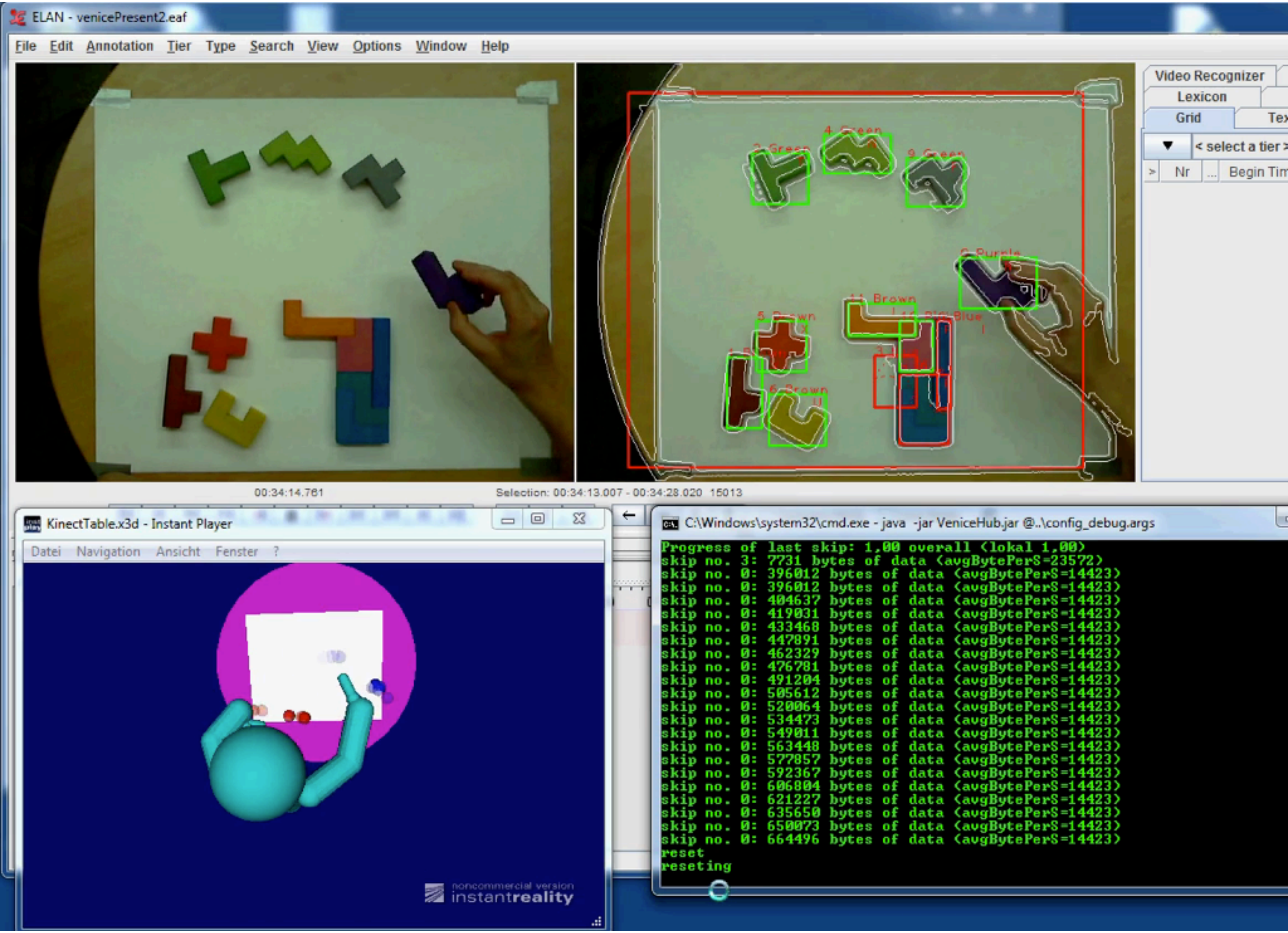
incremental statistical NLU

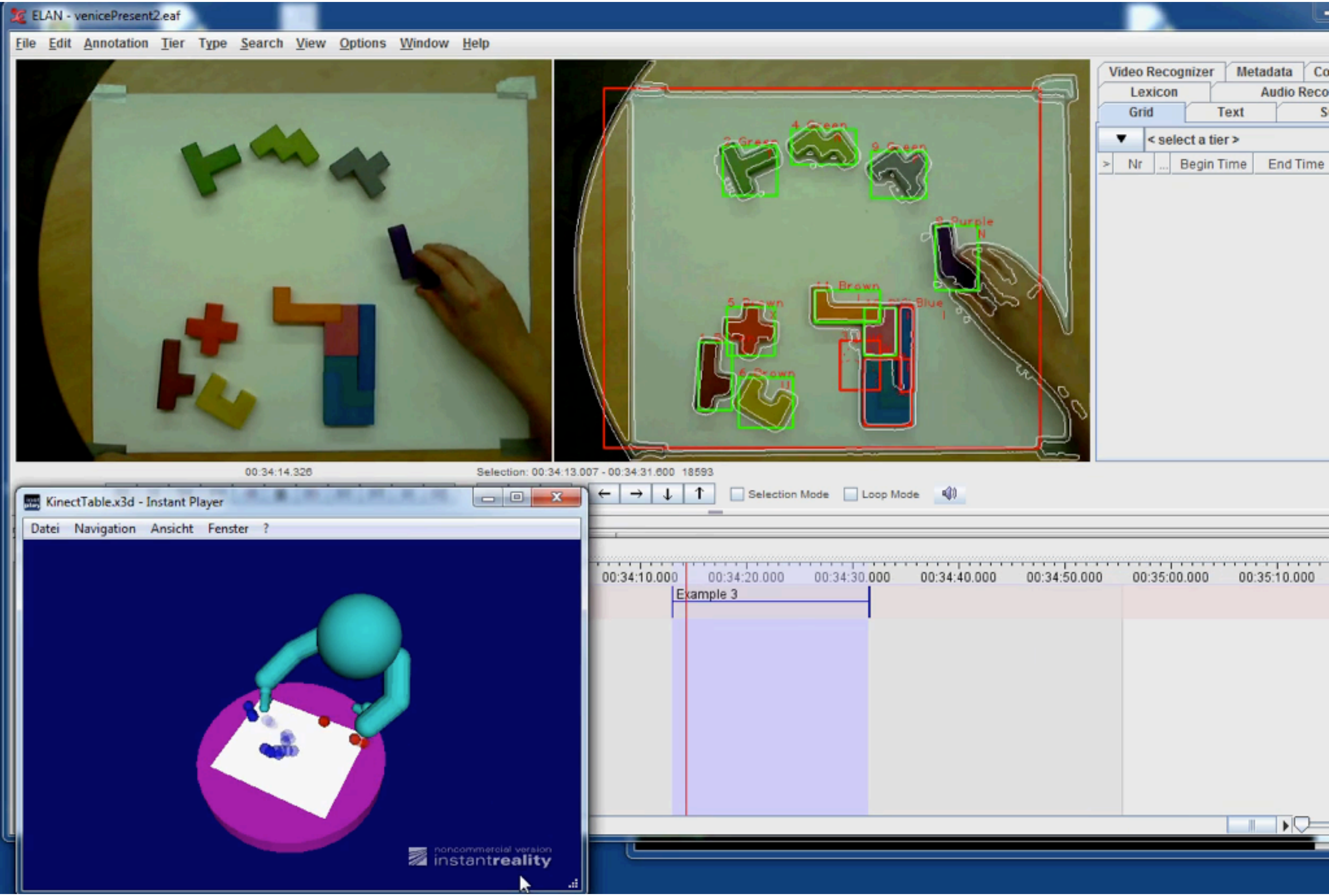
– discriminative model.. word semantics? –

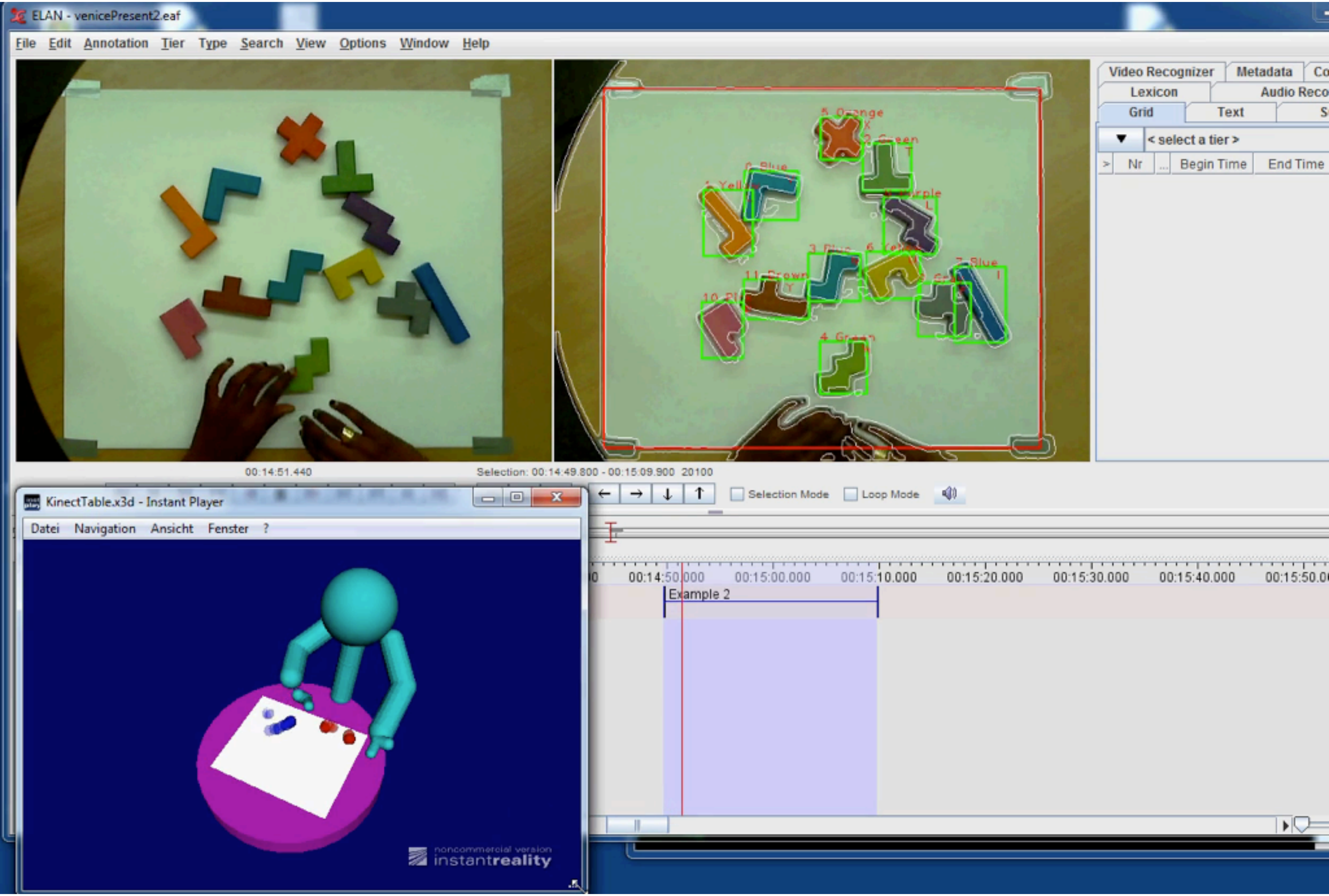


incremental statistical NLU

- models that learn from annotated data (instances of reference), and that
- can be applied incrementally, on-line, and continuously produce hypotheses,
- can make use of additional information sources such as gestures, eye gaze
- can work with non-symbolic input (images)
- perform relatively well


ELAN - venicePresent2.eaf
File Edit Annotation Tier Type Search View Options Window Help
Video Recognizer
Lexicon
Grid
< select a tier
Nr
Begin Tim
3 Green
4 Green
9 Green
8 Purple
11 Brown
5 Brown
6 Brown
Big Blue
00:34:14.761
Selection: 00:34:13.007 - 00:34:28.020 15013
KinectTable.x3d - Instant Player
Datei Navigation Ansicht Fenster ?
noncommercial version
instantreality
C:\Windows\system32\cmd.exe - java -jar VeniceHub.jar @..\config_debug.args
Progress of last skip: 1,00 overall (lokal 1,00)
skip no. 3: 7731 bytes of data (avgBytePerS=23572)
skip no. 0: 396012 bytes of data (avgBytePerS=14423)
skip no. 0: 396012 bytes of data (avgBytePerS=14423)
skip no. 0: 404637 bytes of data (avgBytePerS=14423)
skip no. 0: 419031 bytes of data (avgBytePerS=14423)
skip no. 0: 433468 bytes of data (avgBytePerS=14423)
skip no. 0: 447891 bytes of data (avgBytePerS=14423)
skip no. 0: 462329 bytes of data (avgBytePerS=14423)
skip no. 0: 476781 bytes of data (avgBytePerS=14423)
skip no. 0: 491204 bytes of data (avgBytePerS=14423)
skip no. 0: 505612 bytes of data (avgBytePerS=14423)
skip no. 0: 520064 bytes of data (avgBytePerS=14423)
skip no. 0: 534473 bytes of data (avgBytePerS=14423)
skip no. 0: 549011 bytes of data (avgBytePerS=14423)
skip no. 0: 563448 bytes of data (avgBytePerS=14423)
skip no. 0: 577857 bytes of data (avgBytePerS=14423)
skip no. 0: 592367 bytes of data (avgBytePerS=14423)
skip no. 0: 606804 bytes of data (avgBytePerS=14423)
skip no. 0: 621227 bytes of data (avgBytePerS=14423)
skip no. 0: 635650 bytes of data (avgBytePerS=14423)
skip no. 0: 650073 bytes of data (avgBytePerS=14423)
skip no. 0: 664496 bytes of data (avgBytePerS=14423)
reset
reseting


ELAN - venicePresent2.eaf
File Edit Annotation Tier Type Search View Options Window Help
Video Recognizer
Metadata
Lexicon
Grid
Text
< select a tier >
Nr
Begin Time
End Time
00:34:14.326
Selection: 00:34:13.007 - 00:34:31.600 18593
Selection Mode
Loop Mode
KinectTable.x3d - Instant Player
Datei Navigation Ansicht Fenster ?
00:34:10.000
00:34:20.000
00:34:30.000
00:34:40.000
00:34:50.000
00:35:00.000
00:35:10.000
Example 3
noncommercial version
instantreality



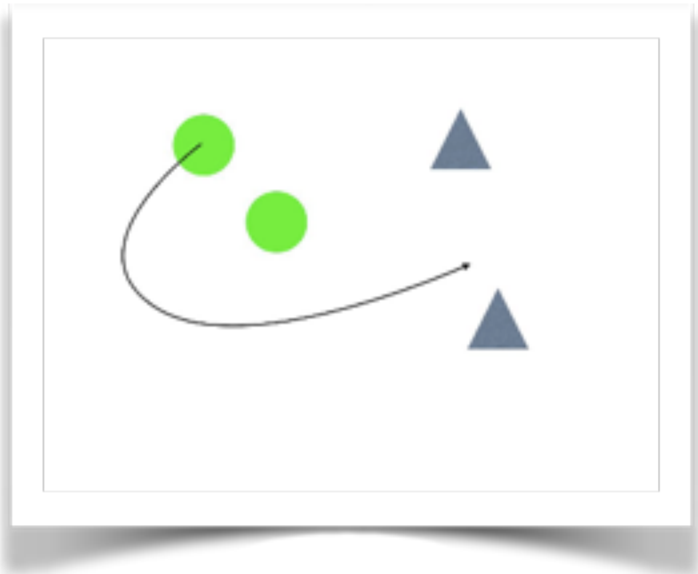
ELAN - venicePresent2.eaf
File Edit Annotation Tier Type Search View Options Window Help
Video Recognizer
Metadata
Lexicon
Grid
Text
< select a tier >
Nr
Begin Time
End Time
00:14:51.440
Selection: 00:14:49.800 - 00:15:09.900 20100
Selection Mode
Loop Mode
KinectTable.x3d - Instant Player
Datei Navigation Ansicht Fenster ?
00:14:50.000
00:15:00.000
00:15:10.000
00:15:20.000
00:15:30.000
00:15:40.000
Example 2
noncommercial version
instantreality

incremental statistical NLU

- models that learn from annotated data (instances of reference), and that
- can be applied incrementally, on-line, and continuously produce hypotheses,
- can make use of additional information sources such as gestures, eye gaze
- can work with non-symbolic input (images)
- perform relatively well

towards a model of sit. com.
– future work –

- learning through interaction / active supervision
- compositionality
- more types of gestures
- managing uncertainty; clarification sequences; active grounding
- embodiment



hier ist ein graues Dreieck
here is a gray triangle

und hier ist ein grüner Kreis
and here is a green circle

hier ist noch ein grüner Kreis
here is another green circle

und hier ist noch ein graues Dreieck
and here is another gray triangle

und von dem oberen grünen Kreis [...]
and from the top green cross [...]

towards a model of sit. com.
– future work –

- learning through (active) interaction / supervision
- compositionality
- more types of gestures
- managing uncertainty; clarification sequences; active grounding
- embodiment

Overview

- **Part I: Foundations**
  - coordination, convention
  - communicative intentions
  - non-conventional meaning
  - grounding
  - turn-taking
  - disfluencies
- **Part II: Computational Models**
  - approaches to dialogue modelling
  - incremental processing, turn-taking
  - an example: grounded semantics

2006-2012 Incremental Processing and Projection in Dialogue (InPro)

Emmy Noether-Programm
Deutsche Forschungsgemeinschaft DFG


University of Potsdam

2010-2014 SFB (collaborative research center) 673, alignment in communication

SFB 673 Alignment in Communication

2010-2018 Center of Excellence, Cognitive Interaction Technology (CIT-EC)

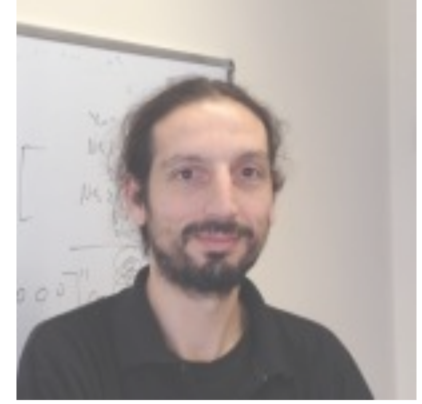
CITEC

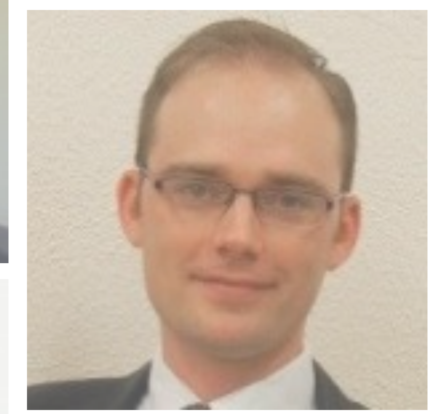

Bielefeld University

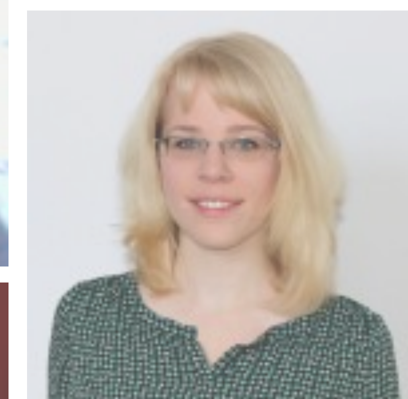
2014-2016 Disfluencies, Exclamations, & Laughter in Dialogue

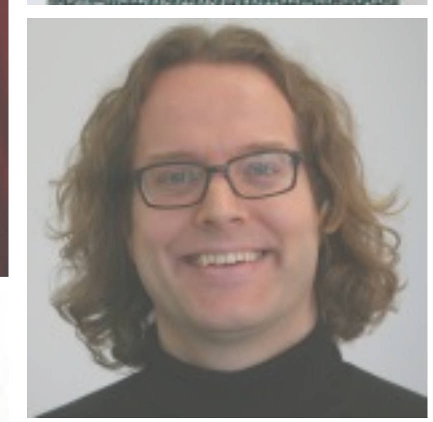
DFG Deutsche Forschungsgemeinschaft

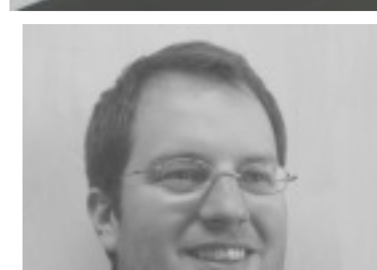
- *Post-Docs*
  - Spyros Kousidis (PhD Dublin)
  - Iwan de Kok (PhD U Twente)
  - Julian Hough (PhD Queen Mary, U London)
- *PhD Students*
  - Casey Kennington
  - Ting Han
  - Birte Carlmeyer
  - Simon Betz
- *Alumni*
  - Timo Baumann (PhD)
  - Gabriel Skantze (Post-Doc)
  - Okko Buß (PhD)
  - Michaela Atterer (Post-Doc)









thank you!