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

- ASR
- DM
- NLG

domain knowledge
Sensor

Actuator

$A_u$

$Y_u$

$Y'_m$

$A'_m$

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

- **FSA** (e.g., Oviatt et al. 1994, McTear 1998)
- **Forms** (e.g., 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 & Rudnick 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)
Dialogue Manager

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

- **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)
"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)

```
PRIVATE:
[ AGENDA : Stack(Action) ]
[ PLAN : Stack(Action) ]
[ BEL : Set(Prop) ]
[ COM : Set(Prop) ]
[ QUD : Stack(Question) ]

SHARED:
[ LU : [ SPEAKER : Participant ] ]
[ MOVE : Move ]
```

(Larsson 2002)
$U(S_m, \tilde{A}_u) = S'_m$
$G(S_m) = A_m$

- FSA (e.g., Oviatt et al. 1994, McTear 1998)
- Forms (e.g., 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
{  set(/SHARED/LU/MOVES, set([ask(?A.price(A))]))
  set(/SHARED/LU/SPEAKER, usr)
}
integrateUsrAsk
{  push(/SHARED/QUD, ?A.price(A))
  push(/PRIVATE/AGENDA, respond(?A.price(A)))
}
findPlan
{  pop(/PRIVATE/AGENDA)
  set(/PRIVATE/PLAN, stack([raise(?C.how(C)), findout(?D.dest.city(D)), ... ]))
}
selectFromPlan
{  push(/PRIVATE/AGENDA, raise(?A.how(A)))
}
selectAsk
{  add(NEXT_MOVES, ask(?A.how(A)))
  if do(fst($/PRIVATE/PLAN), raise(?A.how(A)))
}

"how do you want to travel?"
```
$U(S_m, \tilde{A}_u) = S'_m$

$G(S_m) = A_m$

- **FSA** (e.g., Oviatt et al. 1994, McTear 1998)
- **Forms** (e.g., 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)

```
S_u
\downarrow
A_u
\downarrow
C, \tilde{A}_u
\downarrow
```

```
S_d
\downarrow
A_m
\downarrow
C, \tilde{A}_u
\downarrow
```

```
S_u
\downarrow
A_u
\downarrow
C, \tilde{A}_u
\downarrow
```

```
S_d
\downarrow
A_m
\downarrow
C, \tilde{A}_u
\downarrow
```

Timestep n

Timestep n+1

*aus* (Williams & Young 2007)
\[ U(S_m, \tilde{A}_u) = S'_m \]
\[ G(S_m) = A_m \]

- **FSA** (e.g., Oviatt et al. 1994, McTear 1998)
- **Forms** (e.g., 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

User:

System:

750ms silence
situated dialogue

- participants share a common timeline:

A:  
B:  

- 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

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
I took it
take the red

take

take the right

take the
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
so, one
"so, one"
[l]
one

TTS
AM
sem
ASR
VAD
audio
so, one

"so, one"

[1]

one

VAD

audio
TTS

AM

sem

ASR

VAD

audio
• conditional projections
• w/ dynamic offsets

TTS

AM

sem

ASR

VAD

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

joint work with Okko Buß
(Buß et al., SIGdial 2010, semdial 2010, 2011)
delete the blue cross
which piece?
top right.
ok?
right, now take the yellow [one]...
yes?
... and turn it...
yes?
... to the left
ok.
now flip the stairs...
ok
horizontally
that's right
erm now delete the red [one]*wh-*
bottom right
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

res.

vis

sem

word rec.

\[ x = \text{red cross} \]

\[ P(I) \]
incremental statistical NLU

$$P(I|S, W)$$
incremental statistical NLU

\[ P(I \mid S, W) = \frac{P(S \mid I, W) P(I \mid W)}{P(S \mid W)} \]
incremental statistical NLU

\[ P(I \mid S, W) = \frac{P(S \mid I, W) \cdot P(I \mid W)}{P(S \mid W)} \]

\[ P(S \mid I, W) \]

res.

vis

sem

word rec.

\[ \text{take} \quad x \quad \text{rd()} \quad \text{cross()} \]

\[ \text{take} \quad \text{the} \quad \text{red} \quad \text{cross} \]
incremental statistical NLU

\[
P(I \mid S, W) = \frac{P(S \mid I, W) \cdot P(I \mid W)}{P(S \mid W)}
\]

\[
P(S \mid I, W)
\]
incremental statistical NLU

\[ P(I \mid S, W) = \frac{P(S \mid I, W) \cdot P(I \mid W)}{P(S \mid W)} \]

\[ P(S \mid I, W) \]
incremental statistical NLU

\[ P(I \mid S, W) = \frac{P(S \mid I, W) \cdot P(I \mid W)}{P(S \mid W)} \]

\[ P(U \mid I, W) \]
incremental statistical NLU

\[ P(I \mid S, W) = \frac{P(S \mid I, W) \cdot P(I \mid W)}{P(S \mid W)} \]

\[ P(U \mid I, W) = \sum_R P(U, R \mid I, W) \]
incremental statistical NLU

\[
P(I | S, W) = \frac{P(S | I, W) P(I | W)}{P(S | W)}
\]

\[
P(U | I, W) = \sum_R P(U, R | I, W) = \sum_R P(U | R) P(R | I, W)
\]
incremental statistical NLU

\[ P(I|U, W) = P(U|I, W) P(I|W) \frac{1}{P(U|W)} = \]
\[ \sum_R P(U|R) P(R|I, W) \frac{1}{P(U|W)} \]
incremental statistical NLU
– The Pento-2010 Data –
incremental statistical NLU
– The Pento-2010 Data –

\[ o_1 \text{ shape:star, colour:gray, } x\text{-pos:10, y\text{-pos:20}} \]
\[ o_2 \text{ shape:cross, colour:red, } x\text{-pos:30, y\text{-pos:10}} \]
incremental statistical NLU

\[
P(I|U, W) = \frac{1}{P(U|W)} P(I|W) \sum_R P(R|I, W) P(U|R)
\]
incremental statistical NLU

\[
P(I = o_2 | U = \text{cross}, W) = \\
\frac{1}{P(U = \text{cross} | W)} P(I = o_2 | W) \\
\sum_R P(R | I = o_2, W) P(U = \text{cross} | R)
\]
incremental statistical NLU
– Results for Pento-2010 Data –

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

<table>
<thead>
<tr>
<th></th>
<th>NB</th>
<th>ME</th>
<th>K</th>
<th>H</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>fscore</td>
<td>81.16</td>
<td>92.26</td>
<td>92.18</td>
<td>86.8</td>
<td>76.9</td>
</tr>
<tr>
<td></td>
<td>(74.5)</td>
<td>(89.4)</td>
<td>(86.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>slot</td>
<td>73.62</td>
<td>88.91</td>
<td>88.88</td>
<td>81.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(66.4)</td>
<td>(85.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frame</td>
<td>42.57</td>
<td>74.08</td>
<td>74.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(34.2)</td>
<td>(67.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>action</td>
<td>80.05</td>
<td>93.62</td>
<td>92.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>object</td>
<td>76.23</td>
<td>90.79</td>
<td>84.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>result</td>
<td>64.4</td>
<td>82.34</td>
<td>86.65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
incremental statistical NLU
– The Pento-2010 Data –

\[ o_1 \text{ shape:star, colour:gray, } \quad x\text{-pos:10, y\text{-pos:20}} \]
\[ o_2 \text{ shape:cross, colour:red, } \quad x\text{-pos:30, y\text{-pos:10}} \]
incremental statistical NLU
– The Pento-2013 Data –

\[ o_1 \text{ shape: star, colour: gray, x-pos: 10, y-pos: 20} \]
\[ o_2 \text{ shape: cross, colour: red, x-pos: 30, y-pos: 10} \]
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 –

res.

vis

sem

take x rd() cross()

take the red cross

word rec.
Properties over Time

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

yellow,T,TL,R1,C0,looked_at
yellow,T,R0,C2
blue,P,TL,R2,C2
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

During pointing gesture, all objects in grid receive \textit{pointed\_at} property

then ... take ... second ... a little right ... 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

During pointing gesture, all objects in grid receive *pointed_at* property
Properties over Time

dann ... nehmen ... zweite t bisschen rechts ist ... rüssel hobel ... ja
then ... take ... second t a little right is ... snout plane ... yes
then ... take ... second ... a little right is ... snout plane ... yes
incremental statistical NLU
– Results for Pento-2013 Data –

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

<table>
<thead>
<tr>
<th>Version</th>
<th>Acc</th>
<th>Top 2</th>
<th>Top 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze (baseline) NLU</td>
<td>18%</td>
<td>59%</td>
<td>77%</td>
</tr>
<tr>
<td>NLU + Gaze</td>
<td>53%</td>
<td>62%</td>
<td>80%</td>
</tr>
<tr>
<td>NLU + Point</td>
<td>52%</td>
<td>65%</td>
<td>90%</td>
</tr>
<tr>
<td>NLU + Gaze + Point</td>
<td>53%</td>
<td>70%</td>
<td>91%</td>
</tr>
<tr>
<td>NLU + Gaze-F</td>
<td>53%</td>
<td>65%</td>
<td>78%</td>
</tr>
<tr>
<td>NLU + Point-F</td>
<td>57%</td>
<td>68%</td>
<td>88%</td>
</tr>
<tr>
<td>NLU+Gaze-F+Point-F</td>
<td>56%</td>
<td>69%</td>
<td>85%</td>
</tr>
</tbody>
</table>
incremental statistical NLU
– The Pento-2013 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, via CV –

joint work with Liva Dia
incremental statistical NLU
– The Pento-2013 Data, via CV –

\[ o_1 \text{ shape:star, colour:gray, } x\text{-pos:10, } y\text{-pos:20} \]
\[ o_2 \text{ shape:cross, colour:red, } x\text{-pos:30, } y\text{-pos:10,} \]
\[ \ldots \]
incremental statistical NLU
– The Pento-2013 Data, via CV –
incremental statistical NLU
– The Pento-2013 Data, via CV –

\[ o_1 \text{ shape:star, colour:gray, x-pos:10, y-pos:20} \]
\[ o_2 \text{ shape:cross, colour:red, x-pos:30, y-pos:10, \ldots} \]
incremental statistical NLU
– The Pento-2013 Data, via CV –

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>cross</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>gray</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>red</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>cross</td>
<td>0.2</td>
<td>0.9</td>
</tr>
<tr>
<td>gray</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>red</td>
<td>0.1</td>
<td>0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>0.4</td>
<td>0.05</td>
</tr>
<tr>
<td>cross</td>
<td>1</td>
<td>0.45</td>
</tr>
<tr>
<td>gray</td>
<td>0.45</td>
<td>0.2</td>
</tr>
<tr>
<td>red</td>
<td>0.05</td>
<td>0.3</td>
</tr>
</tbody>
</table>

...
incremental statistical NLU
– The Pento-2013 Data, via CV –

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>cross</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>gray</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>red</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>cross</td>
<td>0.2</td>
<td>0.9</td>
</tr>
<tr>
<td>gray</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>red</td>
<td>0.1</td>
<td>0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>0.4</td>
<td>0.05</td>
</tr>
<tr>
<td>cross</td>
<td>1</td>
<td>0.45</td>
</tr>
<tr>
<td>gray</td>
<td>0.45</td>
<td>0.2</td>
</tr>
<tr>
<td>red</td>
<td>0.05</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>0.78</td>
</tr>
<tr>
<td>CV orig, argmax</td>
<td>0.77</td>
</tr>
<tr>
<td>CV orig, distr</td>
<td>0.75</td>
</tr>
</tbody>
</table>
incremental statistical NLU
– The Pento-2013 Data, via CV –

<table>
<thead>
<tr>
<th></th>
<th>star</th>
<th>cross</th>
<th>gray</th>
<th>red</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1,0</td>
<td>0,1</td>
<td>1,0</td>
<td>0,1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>star</th>
<th>cross</th>
<th>gray</th>
<th>red</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.8</td>
<td>0.2</td>
<td>0.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>star</th>
<th>cross</th>
<th>gray</th>
<th>red</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4</td>
<td>1</td>
<td>0.45</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>0.78</td>
</tr>
<tr>
<td>CV orig, argmax</td>
<td>0.77</td>
</tr>
<tr>
<td>CV orig, distr</td>
<td>0.75</td>
</tr>
<tr>
<td>CV filt, argmax</td>
<td>0.59</td>
</tr>
<tr>
<td>CV filt, distr</td>
<td>0.60</td>
</tr>
</tbody>
</table>
incremental statistical NLU
– generative vs. discriminative –

\[ P(I \mid S, W) = \frac{P(S \mid I, W) \cdot P(I \mid W)}{P(S \mid W)} \]

\[ P(I \mid S, W) = P_W(I | S) \]
incremental statistical NLU
– discriminative model –

\[ P(I \mid S, W) = P_W(I \mid \star^+) \]

Training:
* for each occurrence 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 –

\[ P(I \mid S, W) = P_W(I \mid \star^+) \]

**Training:**
* for each occurrence 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

**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
incremental statistical NLU
– discriminative model –

\[ P(I \mid S, W) = P_W(I \mid \star^+) \]

**Training:**
* for each occurrence 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

<table>
<thead>
<tr>
<th>Version</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>0.78</td>
</tr>
<tr>
<td>CV orig, argmax</td>
<td>0.77</td>
</tr>
<tr>
<td>CV orig, distr</td>
<td>0.75</td>
</tr>
<tr>
<td>CV filt, argmax</td>
<td>0.59</td>
</tr>
<tr>
<td>CV filt, distr</td>
<td>0.60</td>
</tr>
<tr>
<td>discr, CV filt &amp; distr</td>
<td>0.66</td>
</tr>
</tbody>
</table>
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
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
**Incremental Processing and Projection in Dialogue (InPro)**

2006-2012

SFB (collaborative research center) 673, alignment in communication

**Center of Excellence, Cognitive Interaction Technology (CITEC)**

2010-2014

**DFG**

2010-2016

Disfluencies, Exclamations, & Laughter in Dialogue

2014-2016

University of Potsdam

Bielefeld University
• **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!