An HPSG-based Approach to Synchronous Speech and Deixis

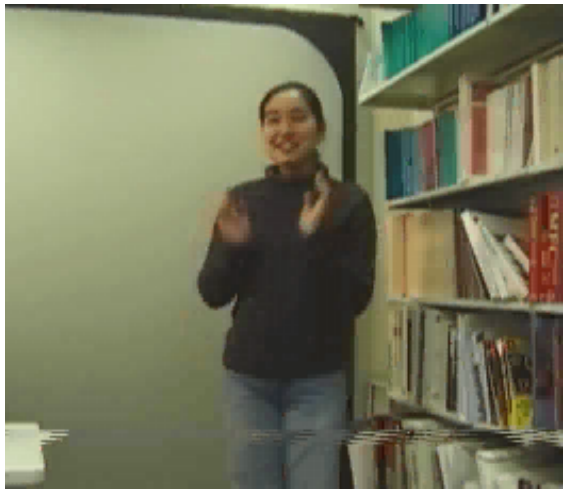
Katya Alahverdzhieva & Alex Lascarides
School of Informatics, University of Edinburgh

18th International Conference on *Head-Driven Phrase Structure Grammar*
University of Washington, Seattle

August 25, 2011

Deictic gesture: range of usage

(a) And $|_g$a as she $_N$said$_g|$, it's an environmentally friendly uh material ...



(b) I $|_g$ $_{PN}$enter my$_g|$ $_N$apartment ...

General Observation:

Distinct LFs:

1. Utterance (a)

   $\pi_1 : \exists m(material(m) \wedge environmentally\text{-}friendly(m))$

   $\pi_2 : \exists s, g(she(s) \wedge said(e_0, s, \pi_1) \wedge loc(g, x, v(\vec{p}_x)) \wedge Identity(s, x))$

   $\pi'_1 : \exists m(material(m) \wedge environmentally\text{-}friendly(m))$

   $\pi'_2 : \exists s, g(she(s) \wedge said(e_0, s, \pi'_1) \wedge classify(g, \pi'_1, v(\vec{p}_s)) \wedge Acceptance(g, \pi'_1))$

2. Utterance (b)

   $\pi_1 : \exists s, a, g(speaker(s) \wedge apartment(a) \wedge enter(e_0, s, a) \wedge$
   $loc(g, y, v(\vec{p}_y)) \wedge VirtualCounterpart(a, y))$

   $\pi'_1 : \exists s, a, g(speaker(s) \wedge apartment(a) \wedge enter(e_0, s, a) \wedge$
   $loc(g, e_1, v(\vec{p}_{e_1})) \wedge VirtualCounterpart(e_0, e_1))$

How can we get these interpretations?

Via a grammar for speech and deixis:

1. identifying the speech signal(s) semantically related with the gesture is a matter of attachment(s) in the grammar
2. relating the speech segment and the gesture is a matter of an underspecified *deictic_rel(s,d)* between the speech *s* content and the deixis *d* content
3. resolving via pragmatics *deictic_rel(s,d)* to *Identity*, *VirtualCounterpart* is logically co-dependent on (1)

Talk in a nutshell

- use the form of the speech signal, the form of the deixis signal and their relative timing to integrate speech + deixis into a single tree
- which maps to an (underspecified) meaning
- via construction rules that
  - impose temporal and linguistic constraints based on empirically extracted generalisations about well-formedness
  - introduce an underspecified relation *deictic_rel(s,d)*

Outline

Deixis Ambiguities (1)

The region pointed out is ambiguous

- What is the exact region when pointing in the direction of a book using 1-index finger?
- Ambiguity resolved via the synchronous NP, e.g., "the book", "the table", "the book cover"
- $\vec{c}$ (tip of 1-index finger) + hand shape, location and orientation determine $\vec{p}$ (the region pointed out by the gesture) [Lascarides and Stone, 2009]

Deixis Ambiguities (2)

"Attachment" ambiguous

- Which constituent does the deixis along with "...as she said" attach to:
  - "she"?
  - "as she said"?
- The grammar allows all attachments as they support distinct interpretations in context, recall:

$\pi_1 : \exists m(material(m) \wedge environmentally\text{-}friendly(m))$
$\pi_2 : \exists s, g(she(s) \wedge said(e_0, s, \pi_1) \wedge loc(g, x, v(\vec{p}_x)) \wedge Identity(s, x))$

$\pi'_1 : \exists m(material(m) \wedge environmentally\text{-}friendly(m))$
$\pi'_2 : \exists s, g(she(s) \wedge said(e_0, s, \pi'_1) \wedge classify(g, \pi'_1, v(\vec{p}_s)) \wedge Acceptance(g, \pi'_1))$

Deixis Ambiguities (3)

Ambiguous relation between speech and deixis

- in utterance (a), there's identity between the gestured space and the physical space ⇒ *Identity*
- in utterance (b), there's no identity between the gestured space and the physical space ⇒ *VirtualCounterpart*
- in the grammar, we connect speech content *s* and deixis content *d* through an underspecified *deictic_rel(s,d)*

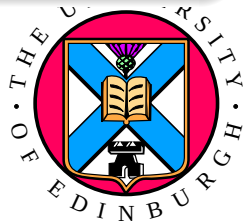
Outline

Empirical Investigation

- evidence in the literature for the interaction between speech prosody and gesture performance, e.g., [Loehr, 2004], [Ebert et al., 2011], [Giorgolo and Verstraten, 2008]
- empirical validation through a corporal study of the interaction between gesture and nuclear prominence

Hypothesis

Deixis aligns with the nuclear pitch accents (PA). In case of pre-nuclear rise, deixis aligns with the pre-nuclear PA.



Empirical Investigation, cont'd

Corpora & Data Annotation

Corpora

- corpora: a 5.53 min recording from `http://www.talkbank.org/`, and observation IS1008c, speaker C from `http://corpus.amiproject.org/`

Annotation

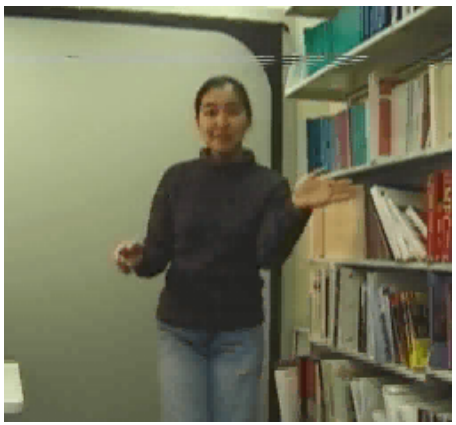
- **Prosody Annotation:** transcription, pitch accents (nuclear, non-nuclear or pre-nuclear) and prosodic phrases
- **Gesture Annotation:** hand movement (communicative vs. non-communicative), gesture phases (preparation, stroke, hold, retraction)

Empirical Investigation, cont'd

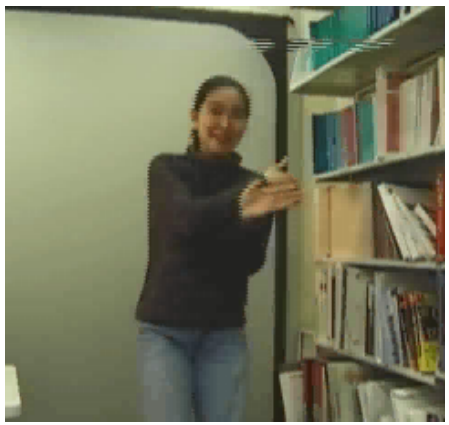
Results & Observations (1)

- temporal overlap between 86/87 deictic gestures and nuclear and/or pre-nuclear accented word
- alignment between gesture and nuclear prominence, and focussed elements

Example



(c) I keep $_{N}$going $|_{g}$ until I $_{NN}$hit Mass $_{N}$Ave $_{g}|$, I think . . .



(d) And then I $_{N}$turn . . . $|_{g}$ $_{N}$left on $_{NN}$Mass $_{g}|$ Ave . . .

Empirical Investigation, cont'd

Results & Observations (2)

- the semantically related speech element is not prosodically prominent
- if the gestured space is identical to the physical space

Example

And $|_g$ a as she $_N$said$_g|$, it's an environmentally friendly uh material

Outline

Modelling Form

Typed Feature Structures

$$\begin{bmatrix} \textit{communicative\_gesture\_deictic} & \\ \textsc{Hand-shape}: & \text{open-flat} \\ \textsc{Palm-orientation}: & \text{vertical} \\ \textsc{Finger-orientation}: & \text{forward} \\ \textsc{Hand-movement}: & \text{downwards} \\ \textsc{Hand-location}: & \vec{c} \end{bmatrix}$$

Modelling Meaning

Robust Minimal Recursion Semantics

$$
\begin{aligned}
& h_0 \\
& l_1 : a_1 : \mathit{deictic\_q}(i)\ \mathit{RSTR}(a_1, h_1)\ \mathit{BODY}(a_1, h_2) \\
& l_2 : a_2 : \mathit{sp\_ref}(g)\ \mathit{ARG1}(a_2, i)\ \mathit{ARG2}(a_2, v(\vec{p})) \\
& l_2 : a_3 : \mathit{hand\_shape\_open\_flat}(e_0)\ \mathit{ARG1}(a_3, i) \\
& l_2 : a_4 : \mathit{palm\_orient\_vertical}(e_1)\ \mathit{ARG1}(a_4, i) \\
& l_2 : a_5 : \mathit{finger\_orient\_forward}(e_3)\ \mathit{ARG1}(a_5, i) \\
& l_2 : a_6 : \mathit{hand\_move\_downwards}(e_5)\ \mathit{ARG1}(a_6, i) \\
& h_1 =_q l_2
\end{aligned}
$$

Rules for Integrating Speech and Deixis in the Grammar

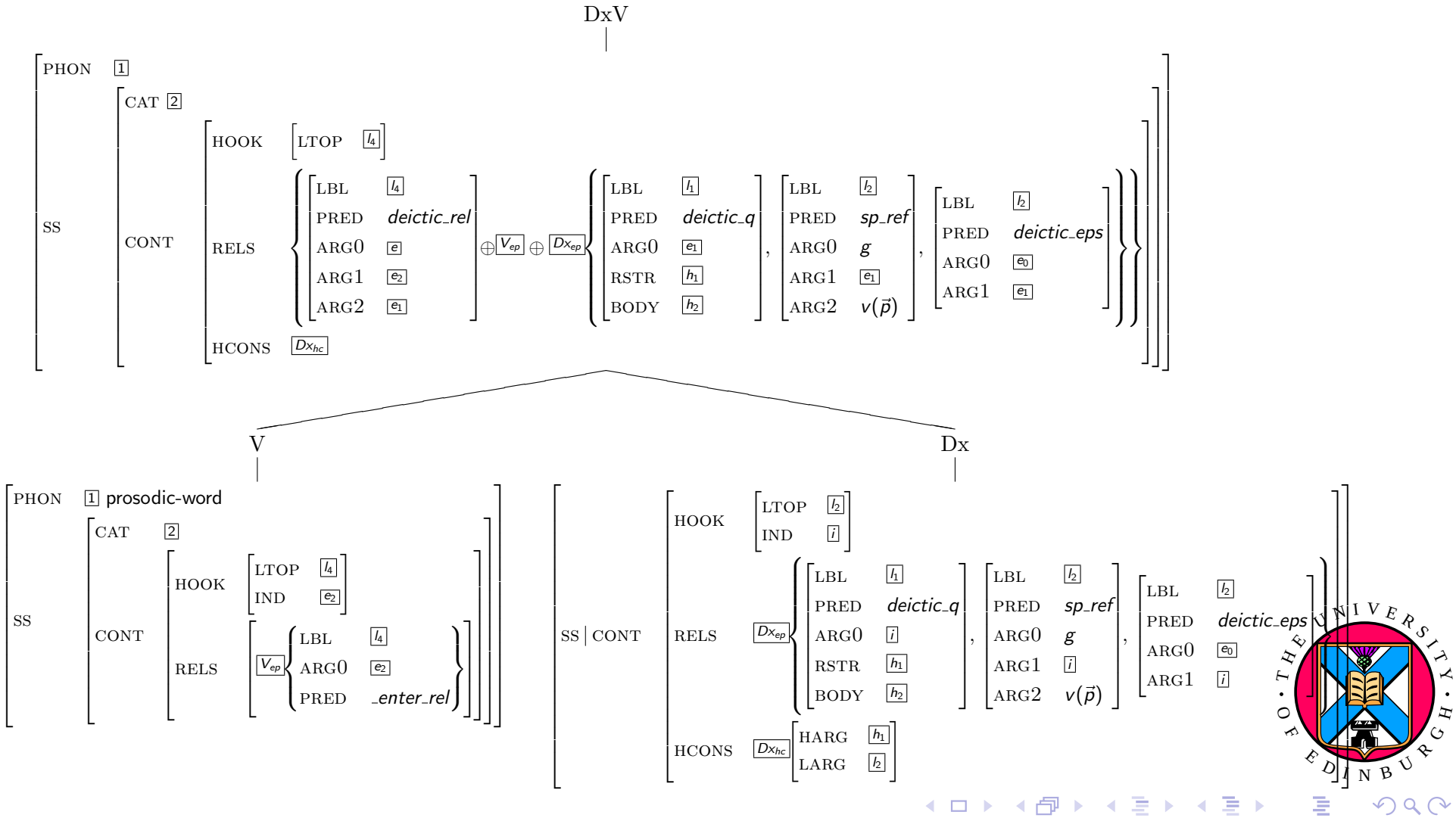
Deictic Prosodic Word Constraint

Deictic gesture *d* attaches to the (pre-)nuclear accented word *w* if there's an overlap between the timing of *w* and the timing of *d*.

- The rule applied to "I enter my apartment" + deixis:
  - "enter" + deixis

Deictic Prosodic Word Constraint in Practice

"enter" + deixis



Rules for Integrating Speech and Deixis in the Grammar

Deictic Head Argument Constraint

Deictic gesture *d* attaches to a (pre-)nuclear prominent head *w* saturated with its arguments if there is an overlap between the timing of *d* and the timing of *w*.

- The rule applied to "I enter my apartment":
  - "enter my apartment" + deixis
  - "I enter my apartment" + deixis

Rules for Integrating Speech and Deixis in the Grammar

Deictic Prosodic Word with Defeasible Constraint

Deictic gesture attaches to a non-prosodically marked element if the mapping v from gestured space $\vec{p}$ to space in denotation $v(\vec{p})$ is identity.

- The rule applied to "And a as she said . . .":
  - "she" + deixis

Outline

Conclusions and Future Work

- well-established methods can be used for the form-meaning mapping of multimodal actions
- . . . achieved by a multimodal grammar which captures constraints from the speech signal, the deictic signal and their relative timing
- the grammar rules are formalised in HPSG since it can interface prosody–syntax–semantics
- implementation of the theoretical findings into a grammar engineering platform, e.g., LKB/PET

Thank you!

Thanks to Nicholas Asher, Sasha Calhoun, Jean Carletta, Jonathan Kilgour, Ewan Klein and Mark Steedman for helpful discussions.

References I

Ebert, Cornelia and Evert, Stefan and Wilmes, Katharina.
2011.
Focus Marking via Gestures.
In I. Reich et al. (eds.) *Proceedings of Sinn & Bedeutung 15*.
Universaar - Saarland University Press: Saarbrücken, Germany.

Giorgolo, Gianluca and Frans Verstraten.
2008.
Perception of speech-and-gesture integration.
In *Proceedings of the International Conference on Auditory-Visual Speech Processing 2008*. pages 31–36.

References II

Loehr, Daniel.
2004.
Gesture and Intonation.
Georgetown University, Washington DC.
Doctoral Dissertation.

Lascarides, Alex and Matthew Stone.
A Formal Semantic Analysis of Gesture.
Journal of Semantics, 2009.