A Formal Analysis of Iconic Gesture

Alex Lascarides
Joint work with Matthew Stone and Katya Alahverdzhieva

School of Informatics
University of Edinburgh

Stanford 2013
Outline

1. Data and Motivation

2. Analysis using techniques from Linguistics
   - Coherence relations and dynamic semantics
   - Underspecification
   - Grammar

3. Conclusion
Iconic Gesture: An Example
So that these very low-level phonological errors tend not to get reported. . .

. . . because they are being produced continually by an iterative process below our level of awareness.
Now one thing you could do is totally audiotape hours and hours... 

...so that you get a large amount of data that you can think of as laid out on a time line.
And exhaustively go through and make sure that you really pick up all the speech errors

\[ \ldots \text{by individually analysing each unit of analysis along the timeline of your data.} \]
Allow two different coders to go through it... 

... and moreover get them to work independently and reconcile their activities.
Iconic Gesture: An Example

speech because gesture
speech so that gesture
speech by gesture
speech and moreover gesture
Speech resolves gesture meaning

(1) So that these very low-level phonological errors tend not to get reported

(2) The mouse ran on the wheel
Describing cotter pins in a lock being held in position: (Engle, 2000)

(3) They have **SPRINGS**.
Right pinched hand (as if holding a small vertical object) is just above left pinched hand (as if holding small vertical object).

- Speech only: collective vs. distributive.
- Gesture depicts a single pin and single spring:
  - ambiguous as to which hand depicts which.
  - Interpretation stems from its iconicity and an inference that it **elaborates** the speech.
- This resolves speech to a **distributive** interpretation.
The current contribution to a discourse is related to a prior contribution by:
- elaborating it, explaining it,
  drawing a contrast, continuing a narrative etc.

Relations’ semantics go beyond compositional semantics, resolving ambiguities, anaphora etc. 
John said that Bill kissed Mary. Peter did too/But Peter did.

The relations structure the context, identifying what’s salient. New contributions must connect to salient bits.
Advantages in gesture analysis

- **Uniform pragmatic theory** for communicative actions, in whatever medium.
- Supports gestures contributing distinctive content:
  - speech *because* gesture
  - Predicts **multimodal anomalies**:
    - You *walk out the doors*.
    - Linguistic analogy: You walk out the doors. Turn right. ???Push the door handle down.
Gesture interacts with prosody

Ill formed!

(4) * Your MOTHER called

Syncopation and boogie woogie in music, but not communication!
Gesture interacts with linguistic syntax

From (Kendon 2004, p.129):

(5) First of all they made everything GREASY in the whole room place.

- Exhaustiveness of greasy stuff...
- ...even if gesture temporally synchronous with made
- But not if gesture temporally overlaps only First or they.
**Construction rules** attach gesture to a phrase:

**Syntax:** Constraints on time, prosody and syntax.

**Semantics:** Introduce an *underspecified coherence relation* between the content of the speech daughter and the gesture daughter.

So timing, syntax and prosody constrain what bits of speech content a gesture can be semantically related to.

Typically have attachment ambiguity, but some readings ruled out by form.
Formalisation

Form-meaning mapping: Use Ivan’s Work!

**RMRS:** to capture the meaning of gesture that’s revealed just by the form of the hand movement(s).

**HPSG:** to articulate how multimodal form constrains meaning.

**Context and Interpretation:**

**Discourse Coherence:** helps resolve underspecified content revealed by form to a specific interpretation in context.

**Dynamic Semantics:** constrains co-reference between speech and gesture and across gestures.

Won’t talk about that here.
Form-meaning mapping: Use Ivan’s Work!

**RMRS**: to capture the meaning of gesture that’s revealed just by the form of the hand movement(s).

**HPSG**: to articulate how multimodal form constrains meaning.

**Context and Interpretation:**

**Discourse Coherence**: helps resolve underspecified content revealed by form to a specific interpretation in context.

**Dynamic Semantics**: constrains co-reference between speech and gesture and across gestures. Won’t talk about that here.

Re-using linguistic formalisms yields uniform approach to interpreting communicative acts, whatever their modality.
Gesture’s form has components:
- Hand shape, finger direction, palm direction, position (relative to torso), path of movement...

and each of these potentially reveals stuff about meaning.

Gesture for (1):

\[
\begin{array}{|l|}
\hline
rh-depict \\
\hline
\hline
\text{HAND-SHAPE} & \text{asl-s} \\
\text{FINGER-DIRECTION} & \text{down} \\
\text{PALM-DIRECTION} & \text{left} \\
\text{TRAJECTORY} & \text{sagittal-circle} \\
\text{MOVEMENT-DIRECTION} & \{ \text{iterative, clockwise} \} \\
\text{LOCATION} & \text{central-right} \\
\hline
\end{array}
\]
Underspecifying Linguistic Meaning: MRS

- Semantic ambiguity without syntactic ambiguity:
  - semantic scope, word senses...
- Underspecified LF is a **partial description** of logical form.

(6) a. Every french bank has some money.

b. \( \text{every}(x, \text{french}(x) \land \text{bank}_{s1}(x), \text{some}(y, \text{money}(y), \text{have}_{s2}(e, x, y)) \text{some}(y, \text{money}(y), \text{every}(x, \text{french}(x) \land \text{bank}_{s2}(x), \text{have}_{s1}(e, x, y)) \text{...}

c. \( l_1 : \text{every}(x, h_2, h_3), l_4 : \text{french}(x), l_4 : \text{bank}(x), l_5 : \text{some}(y, h_6, h_7), l_8 : \text{money}(y), l_9 : \text{have}(e, x, y), h_2 \geq l_4, h_6 \geq l_8 \)
Underspecifying Iconic Meaning: RMRS

Factorisation of Elementary Predications

\( l_9 : \text{have}(e, x, y) \) becomes
\( l_9 : a : \text{have}(e), \text{ARG1}(a, x), \text{ARG2}(a, y) \)

RMRS can underspecify more stuff:
- what arity predicates have (cf. subcat info)  
- what sort and value of arguments they take  
- the argument position of a variable  
- dependencies

All needed for mapping gesture form to content.
Underspecified semantics of gesture

- Each element in gestural form conveys an analogous bit of descriptive content.
- Convention yields the underspecified predicates from the feature structure:

  \[ l_1 : a_1 : \text{hand\_shape\_asl-s}(i_1) \]

- No ARGs or variable equalities
Hierarchy for Resolving Underspecified Predicates

\[ l : a : \text{hand_shape_asl-a(i)} \]

\[ l : a : \text{something_held(x)} \]
\[ l : a : \text{marker_point(x)} \]
\[ l : a : \text{event_of_holding(e)} \]
\[ l : a : \text{literal_holding(e)} \]
\[ l : a : \text{metaphorical_holding(e)} \]
\[ l : a : \text{carry*(e)} \]
\[ l : a : \text{sustain(e)} \]

**marker_point**: 1-place predicate  \hspace{1cm} **sustain**: 3-place predicate

Alex Lascarides  \hspace{1cm} Semantics of Gesture
Situated Spoken Phrase Constraint

A gesture can attach to a temporally overlapping constituent and any of its higher projections.

They made everything GREASY in the whole room.
- greasy ↑ whole clause

They made everything greasy in the whole room.
- they

The cable unexpectedly and abruptly snapped
- unexpectedly ↑ whole clause
Example construction rule (simplified)
Gestural meaning that’s derivable from its form is highly underspecified. RMRS flexible enough to formalise this.

Speech and co-speech gesture should be integrated in the grammar. HPSG flexible enough to formalise this.

Coherence relations are needed to model gesture because:
- Underspecified content is resolved via reasoning about the coherence of the gesture performance;
- Discourse structure constrains what can be gestured now.

Dynamic semantics constrains gesture reference.
