1

Effective and Spurious Ambiguities due to some Co-verbal Gestures in Multimodal Dialogue

Frédéric Landragin

Gesture Workshop

February 27th, 2009, Bielefeld

2

Content

1. Objectives
2. Context: co-verbal gesture in human-machine dialogue systems
   - vocal systems and multimodal systems
   - a preliminary Wizard of Oz study
3. Analysis of a classical example: "*put that there*" + 1 curve-like gesture
   - step 1: analyzing the visual context (= support for the gesture)
   - step 2: analyzing the gesture
   - step 3: analyzing the linguistic utterance
   - step 4: confronting analyses for references to objects resolution
   - step 5: confronting analyses for references to actions resolution
   - step 6: confronting analyses for speech acts processing
4. Conclusion
   - multimodality is essential at each level of interpretation
   - from spurious to effective ambiguity: importance of situational and task contexts

3

Objectives

1. Going back over one of the most famous examples of gesture processing in human-machine dialogue, "*put that there*" (Bolt, 1980)
2. Clarifying the steps of the interpretation of co-verbal gestures, with a parallel with the steps of natural language processing
3. Showing the nature of the intervention of task
4. Providing recommendations for the design of future multimodal human-machine dialogue systems

4

Multimodal dialogue

1. Towards a **spontaneous interaction** between the user and the machine
2. Situations that emphasize the objects that are displayed on the screen
   - a lot of references to objects
   - interest for co-verbal referential gestures
3. A particular use of gesture
   - communication with a machine → restriction of the range of possible gestures whatever the capture device (*set of cameras, touch screen*)
   - choice of a touch screen → no expressive nor paraverbal gestures
   - speech in input of the system → presence of a "push-to-talk" mechanism that regulates the use of synchronization gestures
   - still possible: **quasi-linguistics** and **co-verbal gestures like illustrative and deictic ones**

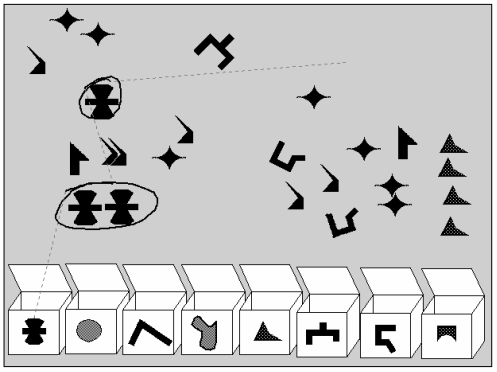
A “Wizard of Oz” study

1. The subject thinks he communicates with a machine but it is a human (the wizard) that simulates the system’s reactions
2. The task is dedicated to actions like “*put that there*”

Example from Magnét’Oz corpus (Wolff, 1999)

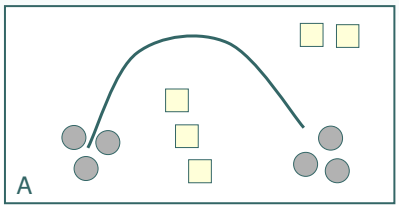
« *range cet objet et ces deux-là dans la première boîte* »

“*put this object and these two ones in the first box*”

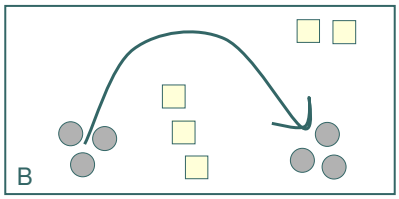


Examples of objects move

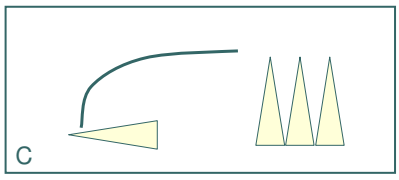
Linguistic utterance: “*put that there*” or “*move that there*”



A. Co-verbal gesture
   - refers to “*that*” and to “*there*”
   - possibly illustrates the way to “*put*”



B. Quasi-linguistic or co-verbal gesture
   - refers to “*that*” and to “*there*”
   - works like a quasi-linguistic action, or illustrates the move trajectory

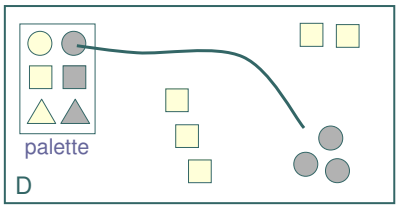


C. Co-verbal gesture
   - refers to “*that*” and to “*there*”
   - illustrates the rotating and the moving

A more complicated example

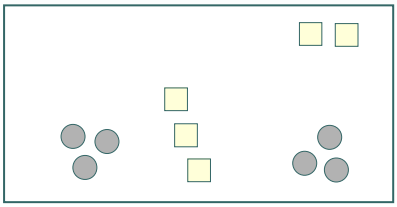
D. Not a move but a duplication

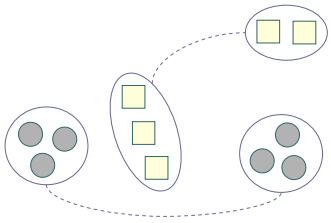
- linguistic utterance: "*put that there*"
- the presence of a palette is a strong argument for the duplication interpretation
- the gesture points out "*that*" and "*there*"
- possibly, the gesture illustrates the move trajectory (if we suppose that the duplicated object appears from the early beginning of the gesture)



(1) Analysis of the visual context

1. Application of Gestalt principles in order to identify perceptual groups
   - spatial proximity
   - similarity of form, de colour…
   - continuity
   - … (Wertheimer, 1923)
2. Detection of "affordances"
   - does the visual context predict any action?
3. Towards a logical formalization
   - group ({sim, prox}, {●,●,●})
   - group ({sim, prox}, {□,□,□})
   - group ({sim}, {□,□,□,□,□})
   - …



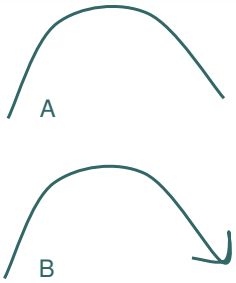


2 perceptual groups for similarity

4 groups for proximity + similarity

9

(2) Analysis of the gesture trajectory



1. Form recognition
   - analysis of the whole form: is it a quasi-linguistic gesture known in the lexicon?
   - analysis of partial elements: identification of the arrow in example B
2. Modelling the trajectory
   - departure point (x_1, y_1); arrival point (x_2, y_2)
   - **prosodic analysis**: regular curve, with no significant stop point
   - **syntactic analysis**: modelling the curve as a succession of curved portions with constant radius, and of significant lexical elements like turning-back points, intersections, superimpositions, etc. (Bellalem, 1995)
   - **semantic analysis**: adding some semantic features that correspond for instance to the inside of the curve (is it a circling gesture?), to the orientation that is specified by the arrow, etc.
   - **pragmatic analysis**: to B corresponds the order to execute an action

10

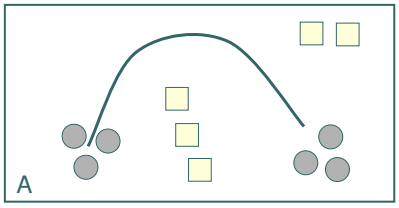
(3) Analysis of the verbal utterance

1. Modelling the sentence "*put / move that there*"
   - **prosodic analysis**: the intonation corresponds to a command
   - **syntactic analysis**: verb with imperative mood with two arguments
   - **semantic analysis**: "*that*" corresponds to the object; "*there*" to the place;
     the semantics of "*move*" brings the notion of route;
     the semantics of "*put*" is more vague (moving action, duplication action?)
   - **pragmatic analysis**: speech act = command (order)
2. More precisely on "*that*"
   - very underdetermined (no gender, no number, no category)
   - demonstrative but not anaphoric (because of the dialogue history), then: needs to be saturated by the extra-linguistic context
   - precondition: "*that*" = something that can be moved or duplicated
3. More precisely on "*there*"
   - deictic that needs to be saturated by the extra-linguistic context
   - it is a positioning action, whose nature (in particular the precision) depends on the nature of "*that*" (cf. "*put a carpet here*" versus "*put a nail here*")

(4) Confronting analyses for references to objects resolution

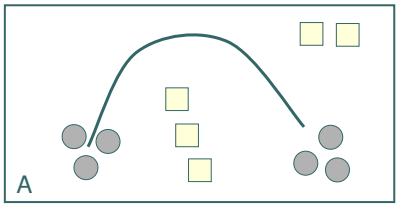
1. Multimodal synchronization
   - 2 linguistic elements that are not saturated, but only 1 gesture
   - then: exploitation of the gesture extremities as designations, with one constraint (same **order of appearance** for words and designations), and one additional argument (good **temporal synchronization**) :
     (x_1, y_1) = departure point for the interpretation of "*that*"; (x_2, y_2) for "*there*"

2. Contextual interpretation of "*that*" and "*there*"
   - "that": ambiguity on the object(s)
   i. the task may impose one object
   ii. the task does not impose anything and other arguments lead to the group: (x_1, y_1) near the centre of the group, similarity of the group's objects…
   - "*there*": considering the nature of "*that*", determination of the precise place



(5a) Confronting analyses for references to actions resolution

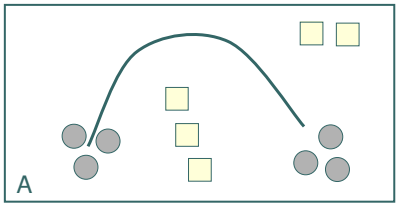
1. With "*move that there*"
   - if the application has the only primitive `move(object,place,route)`, then the gesture is **deictic** and **illustrative**
   - if the application has the only primitive `move(object,place)`, then we can consider a spurious ambiguity between a **purely deictic gesture** (the object disappears and appears at its new place, and the form of the gesture corresponds to the mandatory transition between the two designations) and a **deictic and illustrative gesture** (the illustrative aspect being modelled by `N` calls of the primitive: `move(object,place`$_1$`)`, `move(object,place`$_2$`)`, etc.)
   - if the application has both primitives, another spurious ambiguity appears
   - with the route hypothesis, the main argument to decide between spurious and effective ambiguity is the role of the route considering the visual context: avoiding some objects…



13

(5b) Confronting analyses for references to actions resolution

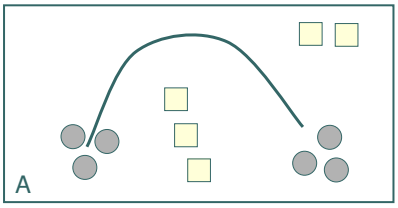
2. With "*put that there*"
   - if the application has one or several primitives **`duplicate()`** besides primitives **`move()`**, then additional ambiguities may appear
   - the task decides: in example A, the duplication hypothesis is not very probable
3. With "*put a circle here*"
   - if the application has one or several primitives **`create()`** besides primitives **`duplicate()`** and **`move()`**, then additional ambiguities may appear
   - in the case of a pointing gesture, there is an ambiguity between "*a circle*" as "*a particular circle*" (existing object, so a moving or duplicating action) and "*a new circle*" with no referent (so a creating action)

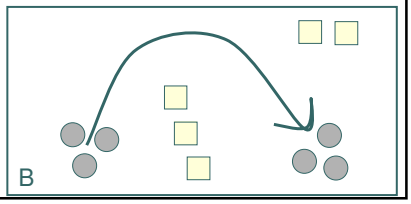


14

(6) Confronting analyses for speech acts processing

1. With "*put that there*" and A gesture
   - prosodic act = **command** (neutral, i.e. with no constraint on the semantic content)
   - linguistic act = **command** ("*put*", so a moving or duplicating semantic content)
   - gestural act = **none** (co-verbal gesture, linked to the linguistic part of the utterance)
   - resulting multimodal dialogue act = the **linguistic act**, with no ambiguity
2. With "*put that there*" and B gesture
   - prosodic act = **command** (neutral)
   - linguistic act = **command** ("*put*")
   - gestural act = **command** ("*move*" or "*duplicate*" semantic content considering the arrow of the quasi-linguistic form)
   - resulting dialogue act = when unifying the linguistic and the gestural act, the semantic contents must be compatible (spurious ambiguities are solved here)





Conclusion

1. Six steps for the interpretation process
   - the task, the context and the interactions between words, gestures and visual elements are essential for a good comprehension from the system
   - *as a recommendation for the design of multimodal systems:* there is a **first multimodal fusion** for the resolution of references to objects, a **second one** for the resolution of references to actions, and a **third one** for the identification of the dialogue acts
2. About co-verbal gestures
   - "*put that there*" is not so simple…
   - deictic and illustrative properties of a gesture are complementary, even in a constrained context like touch-screen based human-machine interaction
3. About ambiguities
   - ambiguities due to the vagueness of language and gesture
   - ambiguities due to the confrontation of the modalities
   - ambiguities due to the multiplicity of task primitives