Pragmatic Constraints on Gesture Use: The Effect of Downward and Non Entailing Contexts on Gesture Processing*

Gianluca Giorgolo and Stephanie Needham

Institute of Cognitive Science, Carleton University, Ottawa, Canada,
gianluca_giorgolo@carelton.ca, sneedham@connect.carleton.ca

Abstract. We report on ongoing research on the semantic and pragmatic factors that influence the interpretation of co-speech spontaneous gestures. We extend the semantic theory introduced by Giorgolo in [1] by proposing a pragmatic principle that controls the felicitousness of spontaneous gestures in different linguistic contexts. The principle is based on the idea of rationality in communication and will result in an extension of the gricean Maxim of Quantity [2]. We also present an experiment that we designed to test the predictions of the combined semantico-pragmatic principles.

Keywords: gesture, multimodality, semantics, pragmatics, semantics-pragmatics interface

1 Introduction

This paper is concerned with the identification of some of the principles that govern the interaction between language and gesture at the semantic level. The main contribution of this paper is the extension of the model for gesture semantics proposed by [1] to take into account pragmatic factors. Our extension is quite conservative as it is based on the assumption of rational behaviour in communication, a fairly standard assumption in current pragmatic theory [3]. This extensions allows us to make fairly strong predictions about the possibility of observing gestures in certain linguistic contexts. The paper presents the first results of an ongoing experiment designed to test these predictions.

In this paper, we will concentrate on the spontaneous manual gestures that typically accompany verbal language. More specifically we will focus on the class of gestures usually referred to as *iconic gestures* [4]. This class of gestures includes those hand movements that are used spontaneously by speakers and that visualize physical properties of the entities or the events referred to in the utterance. This type of gesture lacks a codified form of execution (in this

* This research is supported by an Early Researcher Award from the Ontario Ministry of Research and Innovation and NSERC Discovery Grant #371969. The authors thank Sebastien Plante for acting in the experiment stimuli and Raj Singh and Deidre Kelly for useful feedback on the experiments and the analysis.

sense they are spontaneous and free form), even though they tend to be used consistently in the same stretch of discourse [5].

In Section 2 we present in more detail the framework proposed by [1] and discuss its limitations. Section 3 introduces the simple extension we propose to the framework and discusses some of the implications of this extension. The current state of the experimental work is discussed in Section 4. We conclude with Section 5 by discussing the significance of our findings and what they mean for a theory of gesture meaning.

2 The Interpretation of Spontaneous Co-Speech Gestures

Our starting point is the *semantics* for gesture proposed by Giorgolo [1]. This analysis assumes that gestures contribute to the interpretation of an utterance by providing additional information expressed in terms of an iconic representation. According to this analysis, the iconic representation is a process that identifies the salient spatial features of the referent of a gesture (be it an entity or an event) and that encodes them as visible actions. The representation identifies an equivalence class of spatial configurations that are *indistinguishable* from the virtual space created by the hand movements at a context dependent level of description (determined by the salient features picked for the gestural representation). In this semantic model the two modalities are interfaced by an *adjunction* operation that *intersects* the semantic content of the gesture with the content of its verbal anchor point (roughly the semantic constituent connected both temporally and semantically with the gesture).

The process can be visualized as the diagram in Figure 1. The speech component σ of a multimodal utterance is interpreted in the usual way. Each verbal constituent is assigned an abstract object taken from an ontology F of entities, events and truth values. A family of (partial) mappings *Loc* connects the ontology F to a spatial ontology S by linking each entity, event, property and relation to its spatial extension (a spatial region, a spatio-temporal region, a set of regions or a set of tuples of regions). On the other side, the gesture γ is translated from a collection of motoric configurations into a virtual spatial object. This virtual space is then used to create the equivalence class of the spaces that are sufficiently similar to the represented one. The characteristic function of this class is taken to be the core meaning of the gesture. Finally the meaning contribution of the verbal component and the one provided by gesture are combined via a generalized meet operation.

This model makes already a number of strong predictions partially confirmed by experiments reported in [1]. For instance the model restricts the distribution of gestures by constraining their co-occurrence with verbal expressions. According to the model gestures can co-occur only with those linguistic constituents whose interpretation (under one of the *Loc* mappings) is of a type that can be intersected with the meaning of a gesture (basically the model restricts the verbal correlate of gestures to expressions that denote properties or relations). At the same time the model precludes the possibility of introducing discourse referents

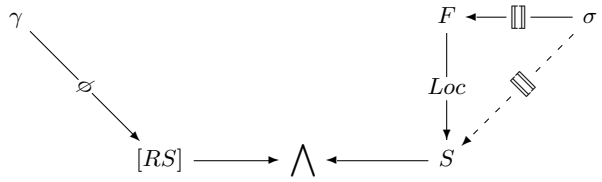


Fig. 1: Interpretation process for a multimodal utterance.

with gestures (see [6] for a different approach in which gestures have the possibility to introduce discourse referents and are in general treated as discourse segments).

However, in its current form, the model cannot predict the distribution of information between the two modalities. The extension we present in the next section allows us to make simple predictions of this kind. A full model of the distribution of information between modalities would require the introduction of more theoretical constructs such as a way of quantifying the expression of a piece of information in any specific modality. In this paper we present just a first attempt in this direction that does not require the introduction of any additional theoretical assumptions besides the notion of rationality in communication which is already independently motivated by the study of verbal language.

3 Pragmatic Constraints on Gesture Interpretation

Gesture is not the primary mode of communication in most cases. This is particularly true for the spontaneous gestures we are dealing with in this paper. Although we pay attention to the hands of our interlocutor, our attention is focused primarily on the facial area with quick glimpses at what the hands are doing ([7]). We expect that communication has evolved so that both listeners and speakers take into consideration this fact when they engage in a conversation.

We can couple this observation with the general rules governing communication that have been proposed when considering language in isolation. In particular the various incarnations of the Maxim of Quantity seem to be particularly related to the situation we are dealing with. In its most general form the maxim assigns a lower and upper bound to the amount of information that a contribution to the conversation should convey. The general idea is that the contribution should be such that it presents our interlocutor with enough information to move the conversation towards the desired goal (some ideal informational state) without including irrelevant or excessive details. By observing naturally occurring gestural data, one has pretty soon the impression that gestures often play the role of providing some more details about the situation under discussion without crossing the upper bound limit imposed by the Maxim of Quantity. In a sense gestures allow us to smuggle some more information into the conversation without making it too heavy. The freedom of adding more information, however, is

combined with the lower saliency that information expressed gesturally shows. A speaker encoding a piece of information in gesture runs the risk of seeing that information disregarded by her interlocutor. We expect this fact to be built in the rules of communication and to determine how people use gestures.

We can express this rule in the following terms: *encode a piece of information as a gesture only if it provides additional information moving you closer to your goal and if it is not necessary to achieve your goal.* The idea is therefore that gestures can only monotonically increase the amount of information available and that they can do so only if their contribution is not vital for the success of the conversation. This rule allows us to preserve the validity of the Maxim of Quantity. With gesture we are apparently allowed to break it by introducing some information that often (if paraphrased in verbal terms) would make our contribution too heavy. At the same time we are not breaking the rule by marking this additional information as not particularly salient and by allowing our interlocutor to disregard it safely.[1]

If we take seriously the adjunct-like semantics for gesture we sketched above, we can find linguistic contexts in which the use of a gesture would contravene the rule just stated. The cases we focus on in this paper is the one of downward monotone and non-monotone contexts. Given the non propositional content of gestures we focus on the contexts induced by determiners and quantifiers. In those contexts deciding to use a gesture to convey information corresponding to an adjunction may result in a communicative failure. In fact in the case of a downward monotone context failing to integrate the gestural information may result on the part of the interlocutor in an interpretation that is too strict and that may even not include the state of information that is the goal of the speaker. Similarly in the case of a non monotone context the adjoined information may be crucial to determine the correct interpretation. This is never the case in an upward monotone context as any interpretation that includes the additional contribution of the gesture is in a sense a subset of the laxer interpretation without gesture. Another way of looking at this prediction is by considering the fact that only upward monotone contexts allow gesture to operate as an optional source of information that increases the overall amount of information in a monotonic way.

Our prediction is therefore that downward and non monotone contexts are bad candidates for the use of gestures. To show that this is the case we designed an experiment that takes advantage of the fact that listeners are attuned to the dispreference of speakers for using gestures in those contexts. We expect listeners to lack a strategy to interpret gestures in those contexts or to have at least a preference for not integrating them.

[1] There are linguistic expressions that play the role of moving saliency away from language and towards other media such as gesture. For example the use of deictics like “this” and “so” and similar phrases like “shaped like this” or “this big” create the effect of marking the gestural information as necessary for the communicative goals. We expect that in those cases this rule and the predictions depending on it would not apply.

4 Experiment

The experiment is designed to measure whether subjects integrate or not the gestural information in their mental representation depending on the monotonicity of the linguistic contexts in which the gesture appears. The prediction is that preference for integration is dependent on the context and that the upward monotone contexts will show a strong preference for integrating gesture (around 75% according to previous measurements of [1]). On the other hand, downward and non monotone contexts will show a strong dispreference for an integrated interpretation.

4.1 Experimental Setup

We designed a set of stimuli to test our hypotheses. The experiment consisted of 12 stimuli grouped in three conditions corresponding to the three possible monotonic behaviours of determiners and quantifiers (upward entailing, downward entailing and non-entailing) and additionally subdivided according to the position occupied by the linguistic anchor point for the gesture (restrictor or predicate position of a determiner).

Each stimulus is a short audio and video recording of a confederate engaged in a seemingly "natural" conversation. The part of the conversation shown to the subject is a short monologue introducing a context ended with a quantified statement about the main topic of the monologue. The monologues are all constructed according to a single template. A class of entities or (possible) events is introduced together with a number of relevant subcategories. Each subcategory is identified by some physical property and an iconic gesture is associated with that property. The final statement contains a reference to the general category accompanied by a gesture associated with one of the subcategories. The clip is cut in such a way that the statement gives the impression of being the first part of a longer utterance. Subjects were asked to pick the most likely continuation of the monologue from a list of sentences. To clarify the setup we present here one of the stimuli used, the "spiders" stimulus. The monologue is reported below. The speech segments accompanied by a gesture are annotated with square brackets ([]) and with a tag identifying the gesture. The gestures are shown in Figure 2.

> *"Spiders" transcript*: Last week I booked a flight down to South America 'cause I'm going there for vacation this summer and I was talking to some friends about —you know— what it's like there and one of my friends was telling me about the spiders there especially in the jungle and he was saying that there's these [LITTLE : little spiders] that are everywhere all the time and these [BIG : big fat spiders] that pretty much only come out at night and fortunately none of the [BIG : spiders] are actually poisonous...

The possible continuations for this stimulus are listed below:

1. but the small ones are deadly poisonous
2. in fact some people get bitten on purpose by the small ones because they think it helps prevent heart diseases
3. so I definitely need to buy some serum in case I get bitten
4. the spiders living there are among the deadliest on the planet

(a) LITTLE (b) BIG

Fig. 2: Still frames of the gestures used in the "spiders" stimulus.

The continuations are constructed in a way that allows us to use them to infer the interpretation that subjects assign to the stimuli. We assume that there are always two available interpretations: one that includes the gestural information and one constructed without it. To detect which interpretation is associated with each stimulus the continuations are designed to be compatible with only one of the two available interpretations and in contradiction with the other. In other words, given two interpretations Γ and Δ each continuation p is constructed in such a way that only one of $\Gamma \wedge p$ and $\Delta \wedge p$ is satisfiable.

However notice that, given the semantics of gesture, it is possible to construct this type of continuation only in the case of non-entailing contexts. In the case of upward and downward entailing contexts this is not possible. The reason in both cases is the fact that one interpretation entails the other. In the case of an upward entailing context $\Gamma[\cdot]^{\uparrow}$ we have that the common ground including the gestural information is entailed by the one that does not include it, in symbols $\Gamma[G(A)]^{\uparrow} \leq_t \Gamma[A]^{\uparrow}$, where G is the denotation of a gesture and A the denotation of the linguistic constituent that the gesture modifies. This follows from the definition of upward entailing function and from the intesective semantics of gestures (we have in general that $G(A) \leq A$, where A is an object of some boolean type and G a gesture modifying it). This means that whenever the addition of a proposition p to a common ground $\Gamma[G(A)]^{\uparrow}$ is satisfiable so it is its addition to the weaker common ground $\Gamma[A]^{\uparrow}$ (this is a consequence of the fact that conjunction with a fixed proposition is an upward monotone operation). It is however possible to construct a proposition whose addition to $\Gamma[A]^{\uparrow}$ is satisfiable, while its addition to $\Gamma[G(A)]^{\uparrow}$ leads to a contradiction. In the case of downward entailing contexts we have the dual situation: given that $\Gamma[A]^{\downarrow} \leq_t \Gamma[G(A)]^{\downarrow}$ we can

only find a proposition that is compatible with $\Gamma[G(A)]^{\downarrow}$ without being compatible with $\Gamma[A]^{\downarrow}$, as any proposition compatible with $\Gamma[A]^{\downarrow}$ will be compatible also with $\Gamma[G(A)]^{\downarrow}$ (where with compatible we mean that its conjunction with a proposition does not lead to a contradiction). Table 1 summarizes the types of continuations used in the different conditions to measure whether the gesture had been integrated in the interpretation or not. The types are codified as pairs that define whether the continuation is compatible (expressed by +) or not (expressed by -) with the interpretation that includes the information conveyed by gesture (first component of the pair) and the interpretation that does not include the gestural contribution (second component of the pair). Notice that this does not prevent us from distinguishing between the two interpretations as in both cases we expect to see a dispreference for the continuations compatible with a single interpretation (those that are not $\langle +, + \rangle$).

Gesture interpretation	Upward entailing context	Downward entailing context	Non-entailing context
Integrated	$\langle +, + \rangle$	$\langle +, - \rangle$	$\langle +, - \rangle$
Non integrated	$\langle -, + \rangle$	$\langle +, + \rangle$	$\langle -, + \rangle$

Table 1: Compatibility of continuations with respect to integration of gesture in the interpretation. The pairs indicate whether the continuation used in the different condition is compatible (+) or not (-) with the integrated (first component) and the non integrated (second component) gesture interpretation.

Returning to the "spiders" example, continuation 1 and 3 can be added to the common ground only if the gesture has been integrated in the interpretation (an effect similar to the one obtained by substituting the last sentence of the monologue with 'none of the big spiders are poisonous'). Continuation 2 is the only continuation compatible with the non-integration of the gesture, but at the same time it is also compatible with an integrated interpretation. The fact that none of the big spiders are poisonous does not entail that the small ones are, even though pragmatic factors may favor this interpretation (a fact that would make our result stronger).

To each list of continuations we added a distractor in the form of a sentence contradicting both the integrated and the non integrated interpretation (see continuation 4 in the "spiders" example).[2]

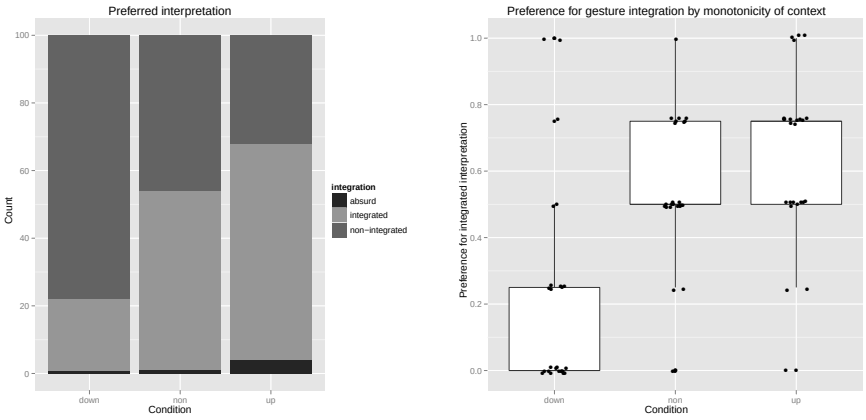
The experiment was run online with subjects recruited mainly among the undergraduate population of Carleton University. The instructions made no mention of gestures to avoid an unnatural focus of subjects on the manual modality. The sequence of stimuli was randomized as was the order in which the proposed

[2] The distractor was added mainly to control for low quality entries in the results, given we could not control for particularly poor experimental conditions.

continuations were presented. Subjects were allowed to watch each clip more than one time.

4.2 Results

We ran a first version of the experiment with 25 subjects. The results are summarized in Figure 3. As we can see the prediction is only partially confirmed by these results. The effect of the monotonicity of the linguistic context is clearly evident but it does not go in the direction predicted by our hypothesis. In upward entailing contexts we observe a clear preference for integrated interpretations, while in the case of downward entailing contexts the preference is for non-integrated interpretations. However non-entailing contexts seem to also trigger a preference for integrated interpretations. This is confirmed by a Tukey's HSD test: the mean preference of the upward monotone context condition is significantly different from the one of downward monotone contexts ($p = 0.0000008$), but it is not different from the mean preference of non monotone contexts ($p = 0.317$), and at the same time the mean preferences of downward monotone contexts are significantly different from non monotonic ones ($p = 0.0019$) while we expected these two condition two show a similar distribution of preferences.



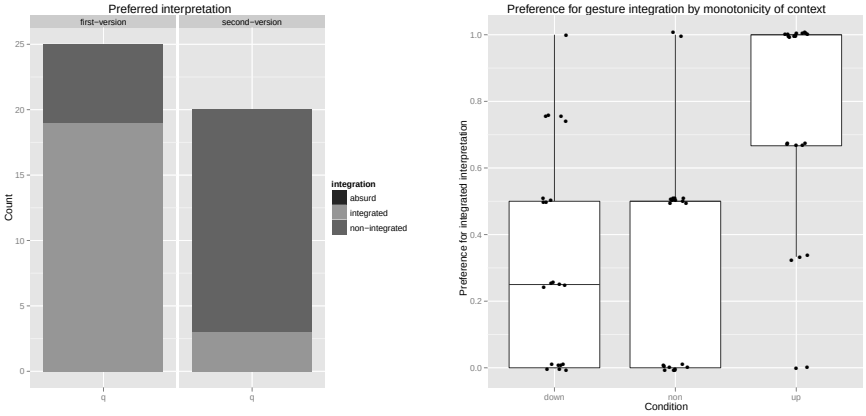
(a) Absolute count of chosen interpretation.

(b) Distribution of integrated interpretation by condition.

Fig. 3: Preferences for integrated and non-integrated interpretations for the first version of the experiment.

However by going through the stimuli again we realized that some of the non monotone tokens were not constructed according to the rules presented above. In particular the favoured answers in those cases were compatible with both an integrated and non integrated interpretation. To test this hypothesis and to

justify the exclusion of the wrong test items, we ran another session of the experiment with 20 more subjects and corrected the list of continuations in one of the affected stimuli. Figure 4(a) shows the change in response with respect to the modified item. The switch to the non integrated interpretation is pretty clear. Figure 4(b) shows instead the distribution of the preferences for the integrated interpretation in the second experiment after removing the wrongly constructed stimuli. The distributions now show more clearly a behaviour close to our predictions. Performing a Tukey's HSD test shows the mean preference for the integrated interpretation in upward monotone contexts is significantly different from the one of downward and non monotone contexts (p-values respectively of 0.00013 and 0.00008), while the distribution of preference for integrated interpretations in downward and non-monotone contexts is to all purposes the same ($p = 0.991$), as our hypothesis predicted.



(a) Count of preferences for different interpretation in the two version of the incorrect stimulus.

(b) Distribution of integrated interpretation by condition.

Fig. 4: Figure (a) shows the effect of the correction in the continuation for the selected non monotone context. Figure (b) shows the distribution of the preference for the integrated interpretation by condition in the second version of the experiment after removal of the incorrect stimuli.

5 Conclusion

We extended the semantics presented in [1] by adding a rule that deals with the pragmatic constraints of gesture use. The extension is based on the notion of rational behaviour in communication and more particularly on the Maxim of Quantity. This new framework makes strong and unexpected predictions about

the possibility of encoding information as gesture in certain linguistic contexts. To test the validity of these predictions we designed an experiment that confirmed our expectations. The results are possibly compatible with other theories of gesture meaning. For instance it is possible to adapt the semantics proposed by Lascarides and Stone in [6] to deal with the data we collected. However such an adaptation would require very strong assumptions as it would be based on a theory of anaphoric-like connections between gesture and language. Our theory does not require any such stipulation as it is based on the simplest possible semantics we can associate with gesture and a general principle about human communication.

We are in the process of extending our empirical explorations to further confirm our hypotheses. In particular we want to use a different experimental setting that does not force our subjects to focus too much on logical operations. At the same time we want to confirm our expectations about the possibility of breaking the pragmatic rule we proposed by means of specific linguistic expressions.

References

1. Giorgolo, G.: Space and Time in Our Hands. PhD thesis, Utrecht Institute for Linguistics OTS, Utrecht: LOT publications 262 (2010).
2. Grice, H. P.: Logic and conversation, in P. Cole and J. Morgan (eds.), Syntax and Semantics, 3: Speech Acts, pp. 41–58, New York: Academic Press (1975).
3. Horn, L. R.: Implicature, in L. R. Horn and G. Ward (eds.), The Handbook of Pragmatics, Oxford: Blackwell Publishing (2004).
4. McNeill, D.: Hand and Mind. Chicago: University of Chicago Press (1992).
5. McNeill, D.: Gesture and Thought. Chicago: University of Chicago Press (2005).
6. Lascarides, A., M. Stone: A Formal Semantics Analysis of Gesture, Journal of Semantics (2009).
7. M. Mancas, F. Pirri and M. Pizzoli: Human-motion Saliency in Multi-motion Scenes and in Close Interaction, Proceedings of Gesture Workshop, Athens (2011).