# A Speech/Gesture Interface: Encoding Static, Locative Relationships in Verbal Discourse

Mark Tutton

University of New South Wales / Université Lille 3

*When speakers communicate, both verbal and non-verbal aspects of behaviour create and influence discourse. This paper looks at a PhD in progress, which examines how native speakers of Australian English and French use both language and gesture to describe static, locative relationships in everyday spatial scenes. The major hypotheses of the study stem from two theoretical concepts which are central to the lexical expression of spatial relationships: the degree of granularity (Narasimhan and Gullberg) in individual English and French prepositions, and the frames of reference (Levinson) adopted by speakers to encode locative relationships. A link between language and gesture is proposed by examining the ramifications these concepts may have for a speaker's gestural behaviour.*

*Keywords:* *speech, gesture, preposition, static locative relationship, English, French*

## 1 Introduction

This paper presents a framework for an experiment-based investigation into how native speakers of Australian English and French encode static, locative relationships in verbal discourse. It works from the premise that meaning is a construction of both verbal and gestural facilities (Kendon 5) and that discourse analysis should account for both verbal and gestural forms of expression. I will start by briefly examining the rise of the field of Gesture Studies, before moving on to present the theoretical concepts at the core of this study. Hypotheses are then elaborated, and the visual stimulus used in the pilot study is presented and explained. This paper constitutes an overview of the early stages of a PhD project, which aims to provide the first comparative study into how native speakers of Australian English and French use language and gesture to describe static, locative relationships in everyday spatial situations.

## 2 Background

The importance of understanding how we communicate meaning through gesture has become a topic of particular prominence in the last 20 years, so much so that the new field of Gesture Studies has come into being. Much of the interest in the human use of hand gesture has been directed towards an understanding of the relationship between co-occurring speech and hand gesture (i.e. McNeill 1992).

Within the spatial domain, studies have sought to examine how speakers of different languages fuse speech and gesture when describing dynamic motion events (cf. Kita and Özyürek), and other research exists into how semantic categories related to the spatial domain, such as size, are encoded in language and gesture (i.e. Beattie and Shovelton). Recent research (cf. Kita; Haviland) into pointing gestures has brought into focus the ways in which speakers encode the fixed location of objects in space, but apart from these works the literature remains relatively scant on the subject. This PhD seeks to respond to this void in the literature, by looking at how native speakers of Australian English and French use both verbal *and* gestural modalities to encode the semantics of static locative relationships. This represents a natural extension of prior research, and also incorporates Australian English into the English-language data: previous studies into English spatial language and gesture are heavily weighted towards North American English (cf. Kendon 345). Since gesture is not just a linguistic phenomenon but also a cultural one (Kendon 328), shifts in how and when gesture is used may conceivably vary amongst different cultural sets of same-language speakers – between American and Australian English speakers, for example. Analysing how we communicate the location of objects to other people is a logical topic of research: from describing the layout of a lounge room to explaining where a certain landmark is found on a city street, these actions are at the heart of our everyday lives. Moreover, gesture allows the speaker an easily accessible and precise means of encoding locative relationships, providing a channel which moves beyond the constraints of spatial language.

## 3 What is Gesture?

Broadly speaking, gesture may be understood as the "symbolic movements related to ongoing talk and to the expressive effort or intention (what you are trying to say)" (Gullberg 2006: 104). This use of gesture, which is in temporal synchrony with the verbalised linguistic segment, is known as *co-verbal* gesture. The union between speech and gesture in the creation of meaning is made possible since "the gestures used by speakers as they speak are partnered with speech as a part of the speaker's *final product* and are as much a part of the utterance's design as the speaker's words" (Kendon 5). Gestures do not just occur with speech, they contribute to utterance meaning: that is, utterances are the creation of language and gestural components, and not just of language alone. This bi-modal approach to speaker-constructed meaning is highlighted by McNeill, who writes that "[g]estures are meaningful and they form nonredundant combinations with the linguistic segments with which they co-occur" (2000: 44). If language and gesture fuse to form a single utterance, this has necessary import for linguistic analyses of speaker-produced discourse: both language and gesture should be understood as the twin channels of a single communicative process (McNeill

1992). While an understanding of what gesture means should not be restricted to hand movement (speakers also gesture with other body parts, very commonly with their heads), the present study focuses solely on speakers' use of hand gesture. This strategic focus allows for a fine-grained analysis of one of a speaker's most commonly-used channels of gestural communication.

## 4 What are Static Locative Relationships?

Linguistically speaking, static locative relationships occur when a person or object is encoded as being at a fixed, unchanging point in space (cf. Hendriks et al.). The entities which enter into this relationship will be referred to as the *Figure* and the *Ground* (Talmy) respectively. The *Figure* is the entity whose location is being determined, and the *Ground* is the reference entity in relation to which this determination occurs. For example, in the sentence "the bird is in the cage", the *bird* is the entity being located and is therefore the Figure, while the *cage* is the reference entity and hence the Ground. In both English and French, static locative information is commonly encoded in the closed grammatical class of prepositions.

| | | | | |
|---|---|---|---|---|
| *1)* | *The picture*<br>NP (Figure) | *is*<br>copula | *on*<br>preposition | *the wall*<br>NP (Ground) |
| | *L'image*<br>NP (Figure) | *est*<br>copula | *au*<br>prep + article (à + le) | *mur*<br>NP (Ground) |

To investigate differences in the semantic networks of French and English locative prepositions, a theoretical approach based upon the concept of *granularity* is proposed.

## 5 Granularity

Granularity is often applied as a methodological tool to measure levels of detail. This view of the concept gives rise to modifiers such as "fine-grained" and "coarse-grained" (Gullberg, forthcoming), referring to high and low degrees of semantic precision respectively. This degree of semantic precision can be applied to utterance analysis, where a speaker's choice of a fine- or coarse-grained lexical unit may reflect shifts in speaker perspective (Narasimhan and Gullberg). It may also reflect the demands of the communicative task at hand: for example, a descriptive activity which requires the speaker to encode locative detail with as much precision as possible, may lead to much finer-grained locative expressions than may otherwise have been the case. As applied to lexical semantics, the granularity of a lexeme may conceivably be determined by assessing its entire semantic network. The problem which arises with such an approach is the difficulty of establishing an appropriate

determination of what constitutes a fine- or a coarse-grained reading. In order to overcome this hurdle, I suggest that the granularity of a lexical item should be determined from within clearly-defined contextual boundaries, and that the exact semantic criteria used to determine granularity should be specified. In the following example, English allows the locative relationship to be encoded by two different prepositions, while only one preposition is possible in French. It would be erroneous to immediately claim that the two English language prepositions are of a finer grain than their French counterpart: after all, their uses are only considered in the one *context*, and their semantic extensions in other uses are not analysed (nor indeed are they of any interest for the present study). If their semantic designations are indeed of a finer grain in the example analysed, then this holds for the context described; it is, however, by no means an indication of "overall" granular level.

| | | | | |
|---|---|---|---|---|
| *2)* | *il y a* | *une lampe* | *au-dessus de* | *la table* |
| | *there is* | *a lamp* | *above/over* | *the table* |

The sentences in example two present an important distinction between available prepositional forms in English and French. Whereas English provides speakers with two possible prepositions to encode the described situation, French allows only one, *au-dessus de*. The semantic distinction between *above* and *over* is important. *Over* implies that the figural entity is located on a point along the horizontal axis, so that the ground entity lies within its "sphere of influence" (Tyler and Evans; Dewell): that is, the lamp is seen as having a potential effect over the entire horizontal dimension of the table (i.e. through the distribution of its light).[1] Such influence on the horizontal access is not necessarily present in *above*, which encodes the superior position of the figural entity on the vertical axis. English speakers can therefore choose whether they wish to highlight the vertical dimension (*above*) or the horizontal one (*over*). In contrast to this, French has the one preposition, *au-dessus de*, which communicates the notion of superiority on the vertical axis, without particular reference to the influence of the figure on the horizontal axis of the ground entity. The question therefore becomes whether French speakers use gesture as a means of communicating this influence along the horizontal axis, if and when such an influence is salient. If gesture is used to this effect, then it may be seen as a sort of modifier, refining the locative semantics of *au-dessus de*. The construction of locative semantics would also be a joint task, realised in both verbal and gestural channels of communication. This question depends, in turn, on how speakers use gesture in the first

[1] This use of *over* to encode location on the horizontal axis is commonly revealed in everyday situations, through expressions like "over on your right", "over to the side" or just "move over". Note that it is not possible to use *over* to refer to the vertical axis in a similar way: *over to the top, *over to the bottom.

place. McNeill (1992: 27) argues that co-occurring speech and gesture "present the same meanings at the same time". Within the framework of such "semantic synchrony" however, McNeill adds that the meanings encoded in both speech and gesture may "complement" each other (1992). It therefore seems possible that gesture may be used to refine the locative semantics of *au-dessus de* by encoding complementary spatial information – such as the salience of the horizontal axis in the Figure/Ground relationship. However, it is also possible that a French speaker's gesture may just reflect the coarse semantic grain of *au-dessus de*. Considered under the umbrella of the semantic synchrony rule, does gesture reflect the semantic granularity of prepositions, or does it refine it?

## 6 Frames of Reference

The second theoretical framework used in this research project is Levinson's frames of reference. These frames of reference address the different ways languages encode a speaker's perception of spatial relationships. The three frames of reference proposed by Levinson are as follows:

> Absolute frame of reference: Spatial coordinates are cardinal directions, used to locate the figural entity in relation to the ground entity. For example, "The station is north of the park".
>
> Intrinsic frame of reference: The spatial coordinates used are intrinsic spatial properties of the ground object. For example, "The ball is in front of me" (37). The Figure entity (the ball) is located in relation to the Ground entity's intrinsic front – the front of the speaker, lexicalised by "me".
>
> Relative frame of reference: This frame of reference involves the mapping of an observer's point of view onto the Figure and Ground entities. It is this point of view which furnishes the spatial co-ordinates in the relationship. For example, "John is in front of the tree" (44). The speaker imbues the Ground entity (the tree), with a front. This process of projecting a spatial property from observer to Ground entity therefore enables a particular locative relationship to be conceived and lexicalised.

Given that neither French nor English requires speakers to use absolute frames of reference for the small-scale spatial relationships being investigated in the present study (the stimulus pictures being of a lounge room and of a street), it is intrinsic and relative frames of reference which are of particular interest. These two frames of reference differ interestingly (and substantially) in terms of how an entity is seen to possess spatial properties. With an intrinsic frame of reference, the entity has its own spatial attributes (for example, an intrinsic front, back, right hand side and left hand side), and it is precisely these intrinsic attributes which are referenced in the locative expression. In contrast to this, a relative frame of reference incorporates the observer's point of view into the locative relationship: that is, it is the observer who projects her own spatial properties onto the ground entity, thereby allowing a spatial relationship between the Figure and Ground to be encoded. For example, if a speaker were to say "there is a ball to the left of the rug", this would necessarily mean that she has

used a relative frame of reference: rugs do not have intrinsic left or right hand sides, and so it is the speaker who creates the spatial relationship between the Figure and Ground by projecting her own left hand side onto the rug. Since the differences between these frames of reference have necessary import for spatial reasoning – a speaker must incorporate the additional step of considering the spatial properties of a third entity when a relative frame of reference is used – I hypothesise that a speaker's use of co-verbal hand gesture will also differ depending on whether an intrinsic or a relative frame of reference is used. I explain this in more detail in section 9.

## 7 Pilot Testing

In order to test the proposed experimental set-up for the main experiments, a pilot study was undertaken at the University of New South Wales, Sydney in second semester, 2006. Eight native speakers of Australian English and eight native speakers of French participated in this study. Speakers were placed into pairs, to enable the filming of four dyads for each language. One speaker was the designated "describer", who had the task of describing a picture to their partner, known as the "receiver". Both speakers were seated opposite each other, and each had a copy of the picture on a stand next to them (see Figure 1, below). On the receiver's copy of the picture however, three items were missing (the cat, the bone and the ball). The describer was instructed to describe the lounge room to the receiver, focusing on the location of 14 items which were listed next to the picture. The describer was also told to describe the location of these items as clearly as possible. Based on the description provided by the describer, the receiver had to work out which three items were missing. Once the describer had completed her description, the receiver was allowed to ask questions. No time limit was imposed for this task. The picture used was specifically devised for this experiment, in order to encourage the use of both intrinsic and relative frames of references, and a wide range of locative prepositions and prepositional expressions (for example, *on, on top of, next to, to the right of, in front of,* etc). Therefore, the location of the dog just *in front of* the table entails the use of a relative frame of reference (since tables don't have intrinsic fronts), although the rug is located in front of the fireplace's intrinsic front.

It is, of course, difficult to predict which prepositions speakers may or may not use: the way in which a particular locative relationship is construed is a function of the way in which the relationship is visually considered by a speaker. This, in turn, depends on the way in which the speaker creates a visual trajectory through the image. For example, one speaker may process the picture along the lateral axis, leading to the observation that *next to the fireplace is a cat*, whereas an approach which considers locative relationships along the vertical axis might lead to an utterance like *beneath the picture of the young boy is a cat*. A speaker's visual trajectory thus becomes a

determining factor of lexical and gestural encoding. Analysis of this pilot-test data is currently in progress.

Figure 1. Pilot-Test Stimulus Picture[2]

## 8 Preliminary Results: Pilot Test

In this section I present several examples from the pilot test data, and argue that the granular level at which a locative relationship is encoded is the function of both language and gesture. The examples which follow are taken from the discourse of the first English-language describer, known hereafter as Paul.

Paul makes frequent use of gesture in his description of the visual stimulus. I will highlight several key instances which show how the granularity of a locative expression is a function of both lexical and gestural components.

In his description of the ball in the picture (see Figure 1), Paul states the following:

[2] Drawing by Peta Dzubiel

*3)* *it's on <u>the bottom of the the picture</u> um [..] it [..] <u>around the the centre</u>*[3]
Gesture 1 Gesture 2

Gesture 2 is of particular interest as far as the concept of granularity is concerned. Paul's right hand moves out in front of the right side of his torso, its palm facing slightly left. His hand then moves up and down, and back and forth across the lateral axis, making a motion as though washing the space with his hand. The gesture ends with a direct, downward movement of the hand slightly to the right of the space he has created with this "washing" movement; this coincides with the lexeme "centre". This gestural reference to the slightly right-of-centre location of the figure is unlexicalised. It does, however, perfectly reflect the location of the ball in the picture, which, as the speaker's gesture shows, is not in the centre of the picture, but just to the right of the centre. The preposition *around* does not encode direction on the lateral axis, giving instead a general indication of location. Paul's gesture refines the general idea of proximity encoded by *around*: his verbal and gestural package shows that the ball is not only *around* the centre, but that it is slightly *to the right* of it. This gesture therefore refines the locative semantics of the prepositional phrase *around the centre*. Different strands of information are presented in language and gesture, and these combine to form an overall package of locative information.

Later in his description, Paul describes the location of the book in the following way:

*4)* *the book is at <u>the woman's feet</u>*
Gesture 3

Gesture 3 starts with Paul's right hand reaching out in front of him, palm facing down and fingers slightly curved as though placing an item: this suggests the general idea of location encoded by the preposition *at*. What happens next is a particularly salient example of how gesture communicates locative information which refines the co-occurring prepositional expression. Paul's right hand moves down and over towards the left, palm still facing down and fingers curved slightly inward, before moving across to the right-hand side in a single, straight motion. This left-to-right movement of the hand co-occurs with the nominal *feet*, and reveals several pieces of information about the book and its location. Firstly, the lateral direction of the gesture in frontal space shows that the book is *in front* of the woman's feet, as opposed to being at a particular side; note that *at* remains unspecific as to any precise location on the frontal or lateral axes. Secondly, the emphasis on the length of the book, as encoded by Paul's left-to-right gesture along the lateral axis, suggests that it is

[3] Brief transcription code: <u>GESTURE</u>: underlining indicates co-occurring gesture
[..]: pause (up to 3 seconds)

positioned lengthways at the woman's feet. This focus on length, performed in temporal synchrony to the nominal *feet*, also suggests that the length of the area occupied by the book is at least equal to that occupied by the woman's feet. The combined lexical/gestural package which Paul produces here offers much more information than the verbal component alone would suggest: the gesture encodes locative information which refines the locative semantics of the preposition, while also providing spatial information pertaining to the entity's size and direction (cf. Gullberg forthcoming).

Looking further at the concept of granularity, the pilot test brings to light a phenomenon which warrants further investigation: this is a linguistic/gestural occurrence in which a speaker uses both language and gesture to encode a locative relationship, and in which the gesture is seemingly incongruent with the locative semantics of the co-occurring linguistic unit. While such occurrences initially appear to be contradictory associations of language and gesture, they in fact allow the speaker to encode the locative relationship to a finer degree of granularity. This phenomenon was particularly salient in Paul's discourse. In the following example, Paul describes the location of the DVD player to his partner, using gesture to encode locative information beyond the semantics of the preposition *on*.

*5)* *and the DVD is <u>just on the compartment [..]</u> of the [..] ah TV stand [..]*
Gesture 4

In gesture 4 Paul's right hand slides forward five times in a dipping, downward motion. The left hand moves underneath the right hand on its third motion back and forth, thereby encoding the idea of physical support concurrently evoked by the preposition "on". The particularly salient feature of this gesture is how it pictorially encodes information which is absent from the co-occurring linguistic segment: the right hand's downward, sliding motion back and forth on the frontal axis makes a clear reference to the DVD player being *underneath* another surface, and therefore *inside* the compartment itself. That the gesture highlights this spatial property as being salient is important: it shows that speakers use gesture to encode unlexicalised locative information. The synchrony of the preposition *on* with the gesture encoding *underneath* initially appears to be an incongruent language/gesture combination. Such adjudication is incorrect however, since the gesture works *with* (and not *against*) the linguistic segment to encode the idea that the DVD player is *on* the compartment, and that this compartment is *underneath* the upper surface of the TV stand itself. Together, speech and gesture achieve a finer degree of semantic precision than would otherwise be afforded by the preposition alone.

## 9 Hypotheses

Preliminary hypotheses, based upon the theoretical material addressed in this paper, are proposed below.

> a. Gesture will be used to provide additional spatial information about referents (for example, physical size). This will occur more frequently when an **intrinsic** frame of reference is used.

Gesture can be used to encode additional, unlexicalised spatial information (Emmorey and Casey), including the size and shape of a referent (Gullberg, forthcoming). Given that intrinsic frames of reference are used for ground entities which possess understood spatial properties, speakers will not need to specify these through gesture. They will therefore have more freedom to use the gesture to encode salient, unlexicalised spatial information.

> b. When a **relative** frame of reference is applied, gesture will instead be used to encode details concerning the location of the figural entity to the ground entity.

Just as the ground entity's spatial properties do not require gestural elaboration when an intrinsic frame of reference is used, I propose that the opposite is true as far as relative frames of reference are concerned. Since the speaker is required to project spatial properties onto the ground entity, the speaker needs to clearly represent what these spatial properties are, to avoid ambiguous interpretations of the utterance. Gesture provides speakers with a modality to clearly (and economically) represent spatial relationships, thus providing an ideal solution to communicate the attribution of new spatial properties through relative frames of reference.

> c. The granularity of spatial prepositions will play a defining role in the information encoded by concomitant gesture.

It is hypothesised that locative semantic granularity, as lexicalised by prepositions, will play a role in gesture production: this stems from an understanding of utterances as a partnership between gesture and speech (Kendon 5). However, whether gesture more commonly reflects the locative semantics of the preposition or adds additional semantic precision remains to be discovered. Both are plausible possibilities. If speakers in fact do both, the question then becomes why a speaker may choose to refine the locative semantics of a spatial preposition at a certain time, but not at another. Examples from the pilot test data show that gesture can indeed refine the locative semantics of a spatial preposition (see examples 3) and 4)). With this in mind, French speakers, who use the preposition *au-dessus de* to lexically encode the semantics of both *above* and *over*, may use gesture as a way of refining this comparatively coarse-grained preposition. This would entail representing, in gesture, the salient spatial axis depending on the desired interpretation: that is, the vertical axis for *above* or the horizontal axis for *over* (cf. example 2), section 5).

## 10 Conclusion

This paper proposed a methodology for investigating the relationship between language and gesture in spontaneous verbal discourse. The investigation of this relationship was justified on the basis that meaning is the joint construction of both language and gesture (Kendon 5): researchers analysing verbal discourse should consider the relationship between verbal and non-verbal phenomena in order to arrive at a fuller understanding of speaker-constructed meaning. The semantics of static locative relationships are lexically encoded by English and French locative prepositions, yet how a speaker determines the overall granular level of a locative relationship can be a function of both verbal *and* gestural components. Within the context of the everyday, local spatial situations investigated in this study, French and English speakers use either intrinsic or relative frames of reference. It is hypothesised that each of these will result in a different use of hand gesture: relative frames of reference will warrant gestural explication of the spatial properties attributed to the ground entity, whereas intrinsic frames of reference will leave the speaker freer to encode additional, unlexicalised spatial information. Examples from the pilot test data show that gesture is used to encode salient, unlexicalised spatial information, and that the granular level of a locative expression can be a function of both verbally and gesturally encoded information.

## Works Cited

Beattie, Geoffrey, and Heather Shovelton. "When Size Really Matters: How a Single Semantic Feature is Represented in the Speech and Gesture Modalities." *Gesture* 6.1 (2006): 63-84.

Dewell, Robert. "*Over* Again: Image-schema Transformations in Semantic Analysis." *Cognitive Linguistics* 5 (1994): 351-80.

Emmorey, Karen, and Shannon Casey. "Gesture, Thought and Spatial Language." *Gesture* 1.1 (2001): 35-50.

Gullberg, Marianne. "Some Reasons for Studying Gesture and Second Language Acquisition (Hommage à Adam Kendon)." [Special Issue on Gestures and SLA, M. Gullberg, Ed.] *International Review of Applied Linguistics* 44.2 (2006): 103-24.

---. "Language-specific Encoding of Placement Events in Gestures." *Events Representations in Language and Cognition.* Eds. E. Pederson, R. Tomlin, and J. Bohnemeyer Cambridge: Cambridge UP, forthcoming.

Haviland, John B. “Pointing, Gesture Spaces, and Mental Maps.” *Language and Gesture.* Ed. David McNeill. Cambridge: Cambridge UP, 2000. 13-46.

Hendriks, Henriëtte, Marzena Watorek, and Patrizia Giuliano. “L’Expression de la Localisation et du Mouvement dans les Descriptions et les Récits en L1 et L2.” *Langages* 155 (2004): 106-26.

Kendon, Adam. *Gesture: Visible Action as Utterance.* Cambridge: Cambridge UP, 2004.

Kita, Sotaro. *Pointing: Where Language, Culture and Cognition Meet.* New Jersey: Lawrence Erlbaum, 2003.

Kita, Sotaro, and Asli Özyürek. “What Does Cross-linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking.” *Journal of Memory and Language* 48 (2003): 16-32.

Levinson, Stephen C. *Space in Language and Cognition – Explorations in Cognitive Diversity.* Cambridge: Cambridge UP, 2003.

McNeill, David. *Hand and mind: What Gestures Reveal about Thought.* Chicago: The University of Chicago Press, 1992.

---. “Analogic/Analytic Representations and Cross-linguistic Differences in Thinking for Speaking*.” *Cognitive Linguistics* 11.1/2 (2000): 43-60.

Narasimhan, Bhuvana, and Marianne Gullberg. “Perspective-shifts in Event Descriptions in Tamil Child Language.” *Journal of Child Language* 33 (2006): 99-124.

Talmy, Leonard. *Toward a Cognitive Semantics.* Cambridge, MA: MIT-press, 2000.

Tyler, Andrea, and Vyvyan Evans. “Reconsidering Prepositional Polysemy Networks: The Case of *Over*.” *Language* 77.4 (2001): 724-65.

ADDRESS FOR CORRESPONDENCE:
mark.tutton@student.unsw.edu.au