# Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse

Yoshioka, Keiko

# Chapter 2. Gesture

## 2.0 Introduction

In the previous chapter, we reviewed how information is organized in L1 and L2 discourse, focusing on speech. In the present chapter, we will review studies on gesture. In 2.1, we will present an overview of gesture whose purpose is to define the phenomenon and familiarize readers with the technical terms to be used in the rest of the work. 2.2 presents views on the relationship between speech and gesture. 2.3 reviews the literature on the relationship between gesture and speakers and listeners. 2.4 presents various views on the mechanism of gesture production, followed by the review of variation in gesture in 2.5. 2.6 presents the findings on gesture in L2. 2.7 reviews the literature on gesture in narrative discourse, and finally in 2.8, we will present the research questions.

## 2.1 Defining gesture: an overview

### 2.1.1 Gesture as hand and arm movements

Although gesture has attracted scholars from ancient times[1], defining the phenomenon is not a simple task. Some consider gesture to be any bodily behaviour. Others may define gesture in a more restricted manner. In fact, how the word is used may vary considerably among studies depending on the theoretical perspective one takes. In 1960's and 70's, gesture was labelled as one form of nonverbal behaviour and studied collectively with other bodily behaviours such as gaze, facial expressions, body-posture and proximity behaviour. These studies are usually known as nonverbal communication research. As the term 'nonverbal' suggests, the investigation of this research tradition mainly focuses on what bodily behaviour expresses that is not expressed in speech or how it regulates the verbal behaviour (Müller 1999). This line of study is still active today. In contrast, in 1980's and onwards, a different line of investigation has developed which focuses more on how closely gesture is intertwined with speech in creating meaning. Unlike nonverbal communication research, gesture research, as this line of study has come to be known, focuses on the representational and functional aspects of the use of the body parts in communication, to which the present work belongs.

In the present work, the definition of gesture is restricted to the movements of hands and arms in order to achieve some communicative intent.[2] At the outset, this restriction excludes behaviours such as facial expressions, body movements, and nodding from the

---

[1] See Kendon (2004) for an excellent historical overview of gesture research from Antiquity to the present day.

[2] However, we recognize the significance of other bodily behaviour in gesture research. For instance, lip-pointing has been studied in detail (Enfield 2001, Sherzer 1972), and the relationship between gaze and gesture has been pointed out (Kendon 2004; Streeck 1993). The use of gaze is discussed briefly in the present work.

analysis. In addition, although some research includes self-touching movements such as scratching one's face or playing with nails as gestures (Goodwin 1986), they are excluded from the present discussion. However, we must go further in seeking to define gesture, since hand and arm movements can be differentiated in more detail. For instance, gestures may be produced with or without speech. Some gestures may be easily recognizable while others are not. For instance, a person may summon others by waving a hand. This hand movement is usually recognizable and can be repeated. On the other hand, people may move their hands and arms while speaking, which are frequently more idiosyncratic and less recognizable than the former example.

### 2.1.2 Hand and arm movements in continuum

For the purpose of refining the definition of gesture, as a starting point, we will review McNeill's (1992, 2000b) notion of a gesture continuum, 'Kendon's continuum', based the idea originally put forth by Kendon (1988). The strength of this notion is that it captures the range and properties of various hand and arm movements including sign languages. As mentioned in passing, in some circumstances, hand and arm movements may be used like a word, while in other cases, their movements are more picture-like. The multiple continua illustrate properties that are shared by but can be used to distinguish various hand and arm movements. Note that this notion sets language (i.e. sign languages) and gestural communication in a continuum, which is distinctively different from how gesture is viewed in nonverbal communication research.

The first version of this continuum, introduced in 1992, identifies different hand and arm movements based on two cross-cutting dimensions, the degree of speech accompaniment and conventionality (McNeill 1992). In the later version, McNeill (2000b) added two more dimensions: the extent of linguistic properties hand movements possess and their semiotic content. Four types of hand and arm movements are distinguished, namely *gesticulation, pantomime, emblem* and *sign languages*. These types of hand and arm movements are placed on each continuum as in Figure 2.1.

| **gesticulation** | **pantomime** | **emblem** | **sign languages** |
|---|---|---|---|
| | ← | → | |
| + | speech accompaniment | | — |
| — | linguistic properties | | + |
| — | conventionality | | + |
| global/synthetic | global/analytic | segmented/synthetic | segmented/analytic |

**Figure 2.1.** Hand and arm movements in continuum (based on McNeill 2000b)

The first continuum regards the degree of speech accompaniment to the movement of hand and arms. The second continuum, the degree of linguistic property, refers to how much the hand and arm movements obey any system constraints with regards to the well-formed vs. not-well-formed ways of producing the movements.

The third continuum, the degree of conventionality, is reflected in the measure of agreement about the manual form among users. The fourth continuum regards the semiotic differences of hand and arm movements.

According to the four continua, sign languages are at the one end in terms of speech accompaniment, the extent of possessing linguistic properties and conventionality. With respect to the semiotic content, they are usually segmented and analytic. Kendon distinguishes two types of sign languages, 'primary' sign languages and 'alternate' sign languages (1988b, 2004). According to Kendon, the former are sign languages used in communities of deaf people as the sole means for communication, such as American Sign Language. The primary sign languages are fully-fledged languages with their own linguistic rule systems and conventional signs (Klima & Bellugi 1979; Lidell 1980; Stokoe 1972). Like any linguistic system, sign languages can be defined in terms of grammatical units.

A slightly different case is the so-called homesigns, elaborated and complex gesture systems used by isolated deaf persons who are born into hearing parents and not exposed to any established sign language. Goldin-Meadow and her associates show that homesigns exhibit some of the features of natural languages, yet they do not develop into full ranges of linguistic features as seen in 'primary' sign languages (Butcher, Mylander & Goldin-Meadow 1991; Goldin-Meadow 2003; Goldin-Meadow & Mylander 1983; Singleton, Morford, & Goldin-Meadow 1993). Furthermore, studies show that when homesigners are gathered in a community, their gesturing develops into a widely shared system which shows some characteristics of 'primary' sign languages. The Nicaraguan sign language users seem to be the best example of such cases (Kegl, Senghas & Coppola 1999; Morford & Kegl 2000). Since 1980, when the first special education school for the deaf children opened in Managua, the sign language used by the deaf children at the school became more stabilized and became the basis for the emergence of what seems like a fully-fledged sign language (Kegl et al.).

In contrast, 'alternate' sign languages are elaborate kinesic codes used by speaker-hearers in restricted circumstances or limited occasions (Kendon 2004). Kendon compares ranges of such systems based on the degree of generalizability and complexity from technical system such as 'crane driver guider gestures', 'sawmill system', 'monastic sign languages' to 'sign languages used by the North America plain Indians and by Australian Aboriginees'. Kendon shows that the alternate sign languages may reflect different degree of relationship with spoken language, and that the generalizability of the system is related to its complexity (2004: 303). The recognition of 'alternate' sign languages is useful in understanding some findings which suggest that when hearing individuals are asked to gesture without speaking, their gesture begin to show more structures (Goldin-Meadow, McNeill, & Singleton 1996; Singleton, Goldin-Meadow, & McNeill 1995).

Next on the continua, *emblems* are highly conventionalised and possibly culturally- or group-bound behaviour whose forms and meanings do not differ from one user to another within the same culture or group and often replace speech. While some emblems may have different meanings across cultures (akin to 'false cognates'

in the lexicon), some are transcultural, or even largely global. Examples of the latter are the 'thumbs up' sign or the raised middle finger indicating sexual insult. Examples of the former are the 'fig' gesture, made by placing the thumb between the index and middle fingers. Although many cultures consider this an obscene gesture, in parts of East Europe (e.g. Hungary and Poland) and Russia, it means 'zero' or 'nothing' (Nomura 1996). Because of their crosscultural variations, *emblems* are widely studied. Surveys and dictionaries are abundant in the literature. (For a detailed analysis of emblems, see Brosnahan & Okada 1990; Calbris 1990, 2001; Morris, Collett, Marsh, & O'Shaughnessy 1979).

Some researchers subcategorise *emblems*. For instance, Kita (2002) classifies *emblems* into four different types: *performative emblems, word-like emblems, expressive emblems* and *meta-discursive emblems*. *Performative emblems* are gestures whose actions themselves perform some social function. An example is the Dutch *lekker* ('tasty') sign—performed by waving a hand sideways next to the face—to show the appreciation of food. Various types of derogatory signs are also examples of *performative* emblems. *Word-like emblems* are *emblems* that are used as words in utterances (cf. Kendon 1988). Rubbing the tip of fingers and thumb to mean 'money' in an utterance such as 'That man has a lot of (money sign)' is an example. *Expressive emblems* are *emblems* that are not necessarily directed towards the other interlocutors. For instance, many professional tennis players spontaneously clench their fist after winning an important point in a match. The action is frequently not directed at the audience or at the opponent. This type of *emblem* is considered to express the inner emotion of the individual (Kita 2002).

*Meta-discursive emblems* are further sub-divided into two categories: *conversation regulating emblems* and *rhetorical emblems*. Examples of the former can be found in 'interactive gestures' studied by Baveles and associates (Baveles, Chovil, Lawrie, & Wade 1992; Bavelas, Chovil, Coates, & Roe 1995; Bavelas & Chovil 2000). The researchers found that certain hand shapes are performed during interactions, which help regulate the flow of conversations. For instance, putting out a hand with the palm up as if to hand over something is performed as a sign of changing the conversational floor (Baveles et al. 1992). When L2 speakers are given a correct lexical item by their listeners, the speakers may produce a similar gesture, putting out a hand with the palm up towards the listener. Furthermore, L2 speakers frequently produce gestures that seem to have meta-discursive motivations such as holding the floor or buying time. These gestures frequently accompany filled pauses during lexical search. However, such hand and arm movements are not conventionalised, thus they fail to be categorized as *emblems*. We will discuss these gestures in later sections.

*Rhetorical emblems* express the illocutionary intent of the utterance in discourse. The example mentioned by Kita (2002: 23) is the use of *mani-giunte* ('joined hands, praying hands') in Southern Italy reported by Kendon (1995). The *emblem* is formed when "two hands, each with fingers extended and adducted, are placed in contact, palms facing one another" (Kendon 1995: 258) at the level of chest in front of the speaker. The gesture is often observed when the speaker appeals for the acceptance of the utterance or the speaker's point of view.

One of the drawbacks of creating various sub-categories is that the division between the categories may become less clear. To use an above example, if the tennis player shows the clenched fist to the opponent, the gesture may simultaneously express the individual's inner emotion and perform some social function, depending on how aggressively the gesture is performed. Thus, all *emblems* have the potential to be *performative*. Nevertheless, despite the possible overlaps among the content, the categories can help us to understand the diversity of *emblems*, especially by the addition of *meta-discursive gestures*.

The next type of gesture on the continua away from sign languages is *pantomimic* gestures. They are depictions of actions which are not as conventionalised as *emblems*. The degree of speech association is attenuated, thus *mimes* may be accompanied by speech. The essential difference between *emblems* and *mimes* is the absence of standardized well-formedness required of *mimes*, leaving some space for an idiosyncratic aspect in expression. These examples of *mimes* can be found in the word-guessing games, such as charades, often performed on TV or among friends.

Various classifications agree in grouping *emblems* as one category, although they are variously termed *symbolic gestures* (Krauss, Chen, & Gottesman 2000; Ricci Bitti & Poggi 1991), and *autonomous gestures* (Kendon 1983). The most widely adopted gesture categorization, by Ekman and Friesen (1969), also distinguishes *emblems* as a separate category. In contrast, *mimes* are not considered an independent category by some researchers (Krauss et al. 2000).

At the opposite end of the continuum from sign languages are *gesticulations*, unconventionalised hand and arm movements which are almost always accompanied by speech. There are some characteristics of these hand movements that are worth noting. First, they are a visual display of meaning. The expression is not linear but dynamic with a multidimensional quality. Second, such hand and arm movements have no rules against which their 'correctness' can be measured. However, this does not exclude the possibility that their productions may exhibit certain patterns among speakers. Third, gestures are synthetic (McNeill 1992: 19). A complex hand and arm movement is not a combination of smaller units, unlike spoken utterances which can be divided into smaller units such as clauses, phrases and words. This type of hand and arm movements has also been labelled *illustrators* (Ekman & Friesen 1969*)* and *speech-focused movements* (Butterworth & Beattie 1978). Our emphasis in this thesis is on these gestures. We will not be analysing *emblems* or *mime* in the present study.

### 2.1.3 Classification of co-speech gestures

Various classification systems recognize subcategories of *gesticulation*, or speech-accompanying gestures. As an example, we will review McNeill's (1992) classification. It recognizes five subcategories of *gesticulations*: *beats, deictic gestures, iconic gestures, metaphoric gestures* and *cohesive gestures*.[3]

*Beats* are rhythmic and often repetitive flicks (vertical or horizontal) of the hand or fingers. The movements are short and quick. *Beats* tend to have the same form

---

[3] Gullberg (1998: 97) expanded 'Kendon's continuum' more in detail within speech-accompanying gestures (except for *cohesive* gesture) and within iconics based on real referential value or resemblance to reality.

irrespective of the content of the speech, and their occurrence is affected more by the supra-segmental quality of speech (Bull & Connelly 1985; McClave 1994; Nobe 1996). Other names adopted in classifications are *batons* (Efron 1972; Ekman & Friesen 1969, 1972), *speech-focused movements* (Butterworth & Beattie 1978), or *motor movements* (Krauss, Chen, & Chawla 1996; Krauss et al. 2000). Vertical *beats* can be frequently observed during political speeches.

 *Deictic* gestures are pointing gestures. The pointing may have a concrete object as a reference or not. For instance, when a speaker points to a restroom while saying, 'The restroom is right there' we have an instance of a *deictic* gesture with a concrete referent. The gesture is exophoric. On the other hand, when a speaker says, 'I used to live in Hawaii and then moved to the Netherlands' with a finger pointing two separate loci in front of the body, these deictic gestures indicate abstract loci (termed as an 'abstract deictic gesture' by McNeill 1992). An abstract deictic gesture may also capture temporality. For instance, a speaker may point backwards with a thumb while saying, 'many years ago'. Variation in hand shape may be observed among *deictic* gestures. For instance, a pointing gesture may be formed by an index finger (Figure 2.2a). An open hand may also be used to point at a concrete or abstract location (Figure 2.2b).

 *Iconic gesture*s are speech-associated hand movements that depict figural representations. For instance, when a speaker forms a round entity with two hands while saying 'There was a beehive', the gesture depicts some of the beehive's physical characteristics. Gesture may depict actions as well. A hand moving upward with wiggling fingers while someone says, 'The boy climbed up a tree', iconically captures the climbing action. The other names that are used to describe this type of gesture are *physiographic gestures* (Efron 1941), *pictographs, kinetographs & spatial movements* (Ekman & Friesen 1969), *ideational gestures* (Hadar, Wenkert-Olenik, Krauss, & Soroker 1998; Hadar & Butterworth 1997), *lexical gestures* (Krauss et al. 2000). Iconic gestures representing entities may be performed instantaneously (Figure 2.3) or by tracing the outline of an entity (cf. 'sculpting mode', Müller 1998 in Kita 2002), as in Figure 2.4.



**Figure 2.2a**. *Deictic* gesture
with an index finger



**Figure 2.2b.** *Deictic* gesture
with an open hand

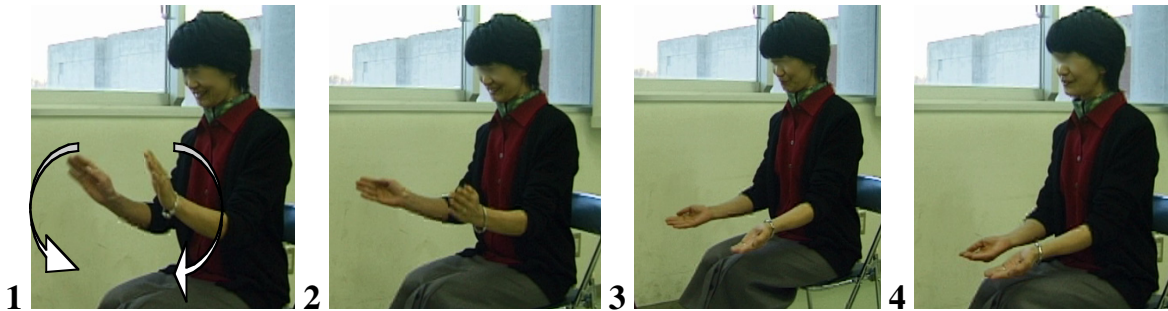**Figure 2.3.** Instantaneous *iconic* gesture



**Figure 2.4.** Gesture tracing the outline of an entity

According to McNeill (1992), *metaphoric* gestures depict abstractions. Efron (1941) and Ekman & Friesen (1969) call this group of gestures *ideographic*. However, compared to the aforementioned types of gesture, there is less uniformity in the definition of this type of gesture among the researchers. For instance, Krauss et al. (2000) questions the usefulness of the iconic/metaphoric distinction. They state, "(So) in our view it makes more sense to think of gestures as being more or less iconic rather than either iconic or metaphoric (or non-iconic)" (Krauss et al. 2000: 276). The disagreement about *metaphorics* raises the question of whether classifications of gesture should be strictly based on the form or the semantic relationship of the form to speech. Kita (2000, 2002) describes *iconic*, *metaphoric* and *deictic* gestures collectively as 'representational gestures' and categorizes *iconic* and *metaphoric* gestures as 'depicting gestures' (Kita 2002).

*Cohesive gestures* are a type of gesture that are thematically related but temporally separated. The recurrence of gestures shows the continuation of a particular theme. Gestural cohesion is revealed by the repeated gesture form, movement or locus in the gesture space. McNeill has recently relabelled these types of gestures 'catchments' (McNeill 2000a; McNeill, Quek, McCullough, Duncan, Furuyama, Bryll, Ma, & Ansari 2001). For instance, in the following clips, the speaker first performs a gesture with palms facing upward (Figure 2.5:1). Then the speaker is interrupted by the listener. As the speaker attempts to answer the question, his hands are placed on his lap, although the shape of the hands is retained from the first gesture (Figure 2.5:2). When the speaker resumes his narrative, he creates the same gesture in the same place (Figure 2.5:3). The recurrence of the same gesture physically shows that the on-going speech is related to the previous narrative before the interruption

occurred. This is one of the features of gesture that we focus in the present study in relation to the establishment of anaphoric linkages of reference in discourse.
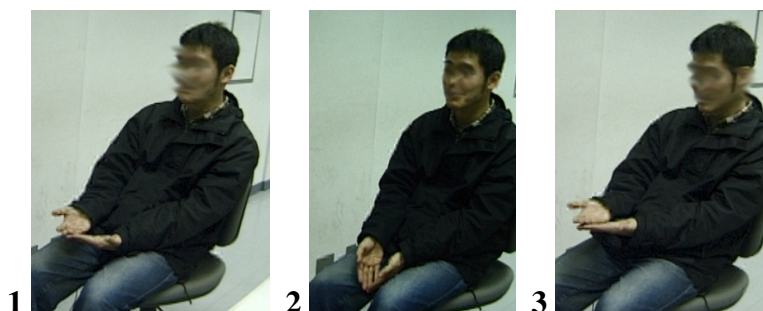


**Figure 2.5.** Cohesive gestures

Although categorized separately, it is by no means the case that the gesture types mentioned above always appear independently of each other. Gestures are typically 'densely-encoding' (Duncan 1996: 21) and it is not rare to find more than one gesture type in a single gesture. For instance, *iconic gestures* sometimes appear with superimposed *beats*, resulting in repetitive movements (vertical or horizontal) of the same iconically represented entity. Similarly, some gestures may have *iconic* and *deictic* qualities at the same time. For instance, when a hand moves across the space in front of the body from left to right when the speaker says, 'the dog runs away from the bees', the gesture iconically depicts the referent's motion and, at the same time, shows the direction of the movement (i.e., from left to right). How to categorize such gestures may vary among researchers, which may cause complications in comparing findings. We will come back to this point later.

In addition, because McNeill's categories are based on the gestures of native speakers, some of the gestures that are produced in L2 may seem difficult to categorize. For instance, gestures that occur during relatively long filled or unfilled pauses may sometimes depict an entity or a movement without any linguistic equivalent in speech. Although we will not deal with these gestures in the present work, they seem to form a specific category, peculiar to L2. Since no conversational floor change occurs during these gestures, these gestures may be motivated to keep the conversational floor or buy time (cf. 'attempt-suppressing signals' by Duncan, 1972; Duncan & Friske 1977).[4] We name these gestures 'spinning ball' gestures, just as the spinning ball on the computer screen indicates that some processing activities are taking place so that the user does not unnecessarily press keys.

Given the detailed classification of gestures, let us again review the definition of gesture in the present work. Gestures are here defined as hand and arm movements accompanying speech which are produced largely without the speaker's awareness. The types of gesture we are dealing with are, namely *beats* and *deictic*, *iconic* and

---

[4] Duncan states, '(a)n attempt-suppressing signal displayed by the speaker maintains the turn for him, regardless of the number of yielding cues concurrently being displayed. Auditors almost never attempted to take their turn when this signal was being displayed. The attempt-suppressing signal consists of one or both of the speaker's hands being engaged in gesticulation' (Duncan 1972: 287).

*cohesive* gestures. *Metaphoric* gestures are excluded from analysis, since the present work only focuses on the gestures accompanying speech relating to concrete referents.

### 2.1.4 Form, space and viewpoint

Gestures can be distinguished by various features such as palm/finger-orientation, hand-shape, the position in the gesture space and viewpoints. McNeill (1992) distinguishes palm/finger-orientation into 6 categories (e.g. palm/finger toward up, down, centre, away from the body etc.). American Sign Language hand shape descriptions are often adopted in the study of gestures. McNeill (1992) presents twenty different gesture hand shapes. Some gestures involve one hand, while others may be performed with both hands. Among two-handed gestures, a distinction could be made between those gestures where both hands jointly encode one entity and those where each hand represents different meanings. In Figure 2.6a, the two hands jointly depict one entity (e.g., a *jar*). In contrast, in Figure 2.6b, the two hands represent different entities. The gesture was performed when the speaker was describing the *dog's head* being stuck in a *jar*. The right hand depicts the movement of the *dog's head*, while the left hand represents the *jar*.
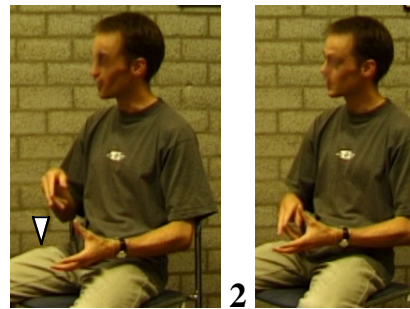


**Figure 2.6a.** Symmetrical use of hands     **Figure 2.6b.** Asymmetrical use of hands

Gesture space is where a gesture occurs. McNeill (1992) divides the space around a speaker into three major domains: *centre, periphery* and *extreme periphery*. The *periphery* and *extreme periphery* are further divided into nine sub-domains as shown in the figure below.
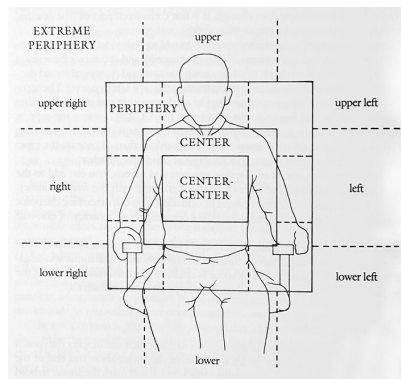


**Figure 2.7.** Gesture space (from McNeill 1992: 378)

Gestures are also distinguished by the viewpoint they represent (McNeill 1992). In some gestures, speakers act out a movement as if they themselves have become the character. These gestures are called 'character viewpoint' gestures (C-VPT gestures). On the other hand, if the gesture captures an event from the viewpoint of an observer, they are called 'objective viewpoint' gestures (O-VPT gestures). In Figures 2.8a and 2.8b, the speakers depict the same scene in the story where the character falls into the pond. However, the speaker in Figure 2.8a acts out the falling movement, with the entire body leaning forward, thus reflecting the character viewpoint. On the other hand, the gesture in Figure 2.8b depicts the trajectory of the fall, thus reflecting the objective viewpoint.
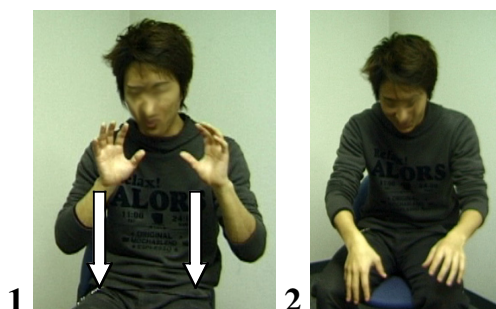


**Figure 2.8a.**
Gesture with a character viewpoint

**Figure 2.8b.**
Gesture with an object viewpoint

### 2.1.5 Gesture units, gesture phrases and phases of gesture units

It has been pointed out that gestures have units, phrases and phases (Kendon 1980, 2004; McNeill 1992). One unit of gesture has been called the G-unit. The unit "begins the moment the limb begins to move and ends when it has reached a rest position again" (McNeill 1992: 83). G-units are comprised of the following phases: *preparation, pre-stroke hold, stroke, post-stroke hold* and *retraction* (or *recovery*). The *preparation* phase is where the hand moves to the optimal position for the *stroke*. The *stroke* phase often synchronizes with the point of prosodic emphasis (McClave 1994; Kendon 1980; Nobe 2000). This is the phase of gesture where the form is semantically interpretable and the most effort is manifested. The meaning represented by this phase of gesture is usually expressed in speech. The *hold* is where the hand is held in the mid-air in the same position. A *hold* may be observed before the *stroke*. This type of *hold* is termed a *pre-stroke hold*. It has been suggested that *pre-stroke holds* are a phenomenon which reflects the synchronized timing relationship between speech and gesture (Kita 1993). The *post-stroke hold* is the phase where a hand is held for a few seconds in the same position once the gesture stroke has ended. It has been suggested that some *post-stroke holds* may have a semantic relationship with speech (Duncan 1996). However, the function of *holds* is not entirely clear. All phases but the stroke phrase, the nucleus of the movement, are optional, although the preparation phrase is rarely omitted (McNeill 1992). Kendon (2004) define the combination of the *stroke*, any *preparation* and *hold* phases as the gesture *phrase,* where the meaning of the gesture is most clearly expressed.

## 2.2 Speech and gesture

The relationship between speech and gesture has intrigued many scholars over the years. Some have examined whether forcing speakers not to gesture would affect speech, and indeed found that hindering gesture leads to changes in speech (Graham & Argyle 1975; Graham & Heywood 1975; Rauscher, Krauss, & Chen 1996; Ríme & Schiaratura 1991; Ríme, Schiaratura, Hupet, & Ghysselinckx 1984). For instance, it was found that when gesturing was hindered, speakers produced more expressions describing spatial relations, fewer deictic expressions and more pauses (Graham & Heywood 1975). Rauscher et al. (1996) found that preventing speakers from gesturing adversely affected speech when the content was spatial. Speakers had more filled pauses. Ríme et al. (1984) found that preventing gestures resulted in speech with fewer imagistic expressions.

Another focus of interest has been the timing between gesture and speech. Findings suggest that gesture tend to either precede or occur simultaneously with semiotically correspondent elements in speech but rarely afterwards (Morrel-Samuels & Krauss 1992). There is general concensus in the literature that *beat* is frequently coordinated with the prosodic features of speech (Bull 1987; Condon & Ogston 1971; McClave 1994), However, there is disagreement with respect to whether the onset of other types of gestures occurs during pauses or on articulation (Butterworth & Beatti 1978; Buterworth & Hardar 1989; McNeill 1992; Nobe 2000). As one of the explanations for the discrepancy of the results, task-effect has been pointed out (Aboudan & Beatti 1996; Beatti & Aboudan 1994; Nobe 2000).

In many of the early studies, the relationship between speech and gesture in creating meaning was largely ignored, or at best, treated as peripheral. Some considered that gesturing was the reflection of speech failure, based on the findings that aphasic patients make use of gesture to overcome the difficulties they face in naming objects (Feyereisen 1986, 1987; Feyereisen & Lannoy 1991). On the other hand, Kendon (1980, 1993, 1994, 2000, 2004) suggested the complementary relationship between the two modes of expression. Furthermore, McNeill (1985, 1987, 1992) provided a theoretical framework that speech and gesture form one integrated system. Several facts are provided as evidence for this view (McNeill 1992). First, gestures mostly occur when the speaker is engaged in speech. In fact, it was found that 90% of gestures occur accompanying speech (McNeill 1992: 23). Second, gestures are semantically and pragmatically co-expressive. For instance, when the speaker produces a gesture tracing the outline of a round entity while uttering, 'The frog lived in a jar', the gesture depicts an object represented by 'a jar' in speech. Pointing and *beat* gestures may accompany the introduction of a new topic of conversation (McNeill 1992), reflecting the pragmatic structure of discourse. Third, speech and gesture tend to be synchronous. The stroke phase of a gesture frequently lines up temporally with the linguistic equivalent in the utterance. This synchronization between speech and gesture suggests integration of the two modes of expressions.

Kendon (2004) illustrates the intricate temporal coordination between speech and gesture by showing how gestural components (see 2.1.5) relate to phrasal organization

of speech. Kendon describes how the preparation phase of gesture occurs well in advance of a part in speech which coincides with the stroke phase. In other words, as the speakers get ready for the verbal component, they get ready for its gestural component, reflecting the fact that they are planned together (2004: 116). Furthermore, Kendon illustrates how the speaker inserts a pause in speech during the *preparation* phase of a gesture, as if waiting for the hands to get ready, so that the semantic coherence between speech and the *stroke* phase of gesture is achieved (2004: 117-119). The co-expressive nature of speech and gesture is also suggested when a gesture is repeated or revised. It is reported that when the same word is repeated due to the external causes such as noise, the speaker tends to repeat the gesture, although the physical features may be less prominent in the repeated version. Similarly, when the speaker revises the verbal components, gesture is also revised. According to Kendon (2004), these findings indicate the integrated nature of speech and gesture (cf. Seyfeddinipur & Kita 2001 for the association between speech and gesture in speech disfluencies).

Other findings also suggest the close relationship between speech and gesture. For instance, it has been reported that stutterers freeze their gestures when they begin to stutter. When their speech resumes its fluency, their hands simultaneously resume their actions (Mayberry & Jaques 2000). In the field of neuroimaging, human PET studies show that hand and mouth representations overlap in so-called Boca's area which is activated during both production and perception of meaningful mouth and manual movements (Nishitani & Hari 2000; Rizzolatti, Fadiga, Matelli, Bettinardi, Paulesu, Perani, & Fzio 1996).

## 2.3 Gesture in language use

### 2.3.1 Addressees and gesture

The view that speech and gesture may form an integrated system leads us to various questions about gesture. Do listeners attend to gestures? Do listeners take any information from gesture into account in comprehending the message? Do speakers direct listeners' attention to gestures? Do speakers gesture solely to benefit the listeners? Or are gestures bi-directional? These are all questions for which various research projects have been conducted. In order to clarify the nature of questions, it is helpful to separate issues concerning the addressees and the speakers. In the following, we will first review findings concerning the former.

Studies show that gestures often prompt reactions from listeners. Studies of mother-infant conversations indicate that mothers react to infant pointing by providing lexical labels (Masur 1982). Children have also been observed to attend to the gestures performed by their mothers (Schmidt 1996, 1999). Using the combination of eye-tracking and digitally manipulated computer graphics, Nobe, Hayamizu, Hasegawa, & Takahashi (2000) found that when the figure in the computer screen gazes at its gestures, the subjects' eye movements were also focused on the gesture. Similar findings have been obtained in another study (Gullberg & Holmqvist 1999). Also using eye-tracking, Gullberg & Holmqvist investigated whether listeners attended to

gestures in natural conversational narratives. Their findings show that although adults listeners fixate on the face of a speaker, they do attend to gestures as well.

Researchers in the field thus seem to agree that listeners attend to gesture, and that gesture is not just hand waving in mid-air. However, whether or not listeners take the information from the gestural channel of communication into account is a separate issue. This topic has been debated over the years and remains unsettled due to contradictory research findings. There are studies indicating that the meaning of speech-accompanying gestures is too opaque and does not help the listeners (Krauss, Morrel-Samuels, & Colasante 1991; Krauss, et al. 1996; Krauss, Dusha, Chen, & Rauscher 1995). In contrast, other studies suggest that listeners decode substantive information from gestures (Beattie & Shovelton 1999a, 1999b, 2000, 2001, 2003; Berger & Polelka 1971); Cassell, McNeill & McCullough 1999; Goldin-Meadow, Wein, & Chan 1992; Goldin-Meadow & Momeni-Sandhofer 1999; Graham & Argyle 1975; McNeill, Cassell, & McCullough 1994; McNeil, Alibali, & Godin-Meadow 2000; Kelly 2001; Kelly & Church 1997, 1998; Kelly, Barr, Church, & Lynch. 1999; Riseborough 1981; Thompson & Massaro 1994).

Krauss et al. (1991) examined whether the listeners decoded meaning from gestures by comparing the judgements of two groups of participants with respect to the semantic category of gestures. One group was asked to assign semantic categories to gestures describing a variety of subjects by viewing them with accompanying speech. The other group was asked to perform the same task but without accompanying speech. The assignment of semantic categories to gesture did not differ between the two groups. Based on this result, Krauss et al. claim that listeners do not derive meaning from gesture. In a different study, Krauss and his associates (1995) examined whether or not listeners understood the meaning of speech better when accompanied by gesture. In their series of experiments, the participants were asked to select the correct object by either watching a video where speakers describe abstract designs, novel sounds or samples of tea, or just listening to the soundtrack of the video. The results showed that there was no difference between the two groups of participants. In other words, watching gestures did not enhance correctness in choosing the object that had been described by the speakers.

On the other hand, McNeil et al. (2000) investigated whether pre-school and kindergarten children understood spoken messages better when they were accompanied by gestures. They found that gestures facilitated understanding if the message was not too complex. Kelly and his associates (Kelly 2001; Kelly & Church 1997; Kelly et al. 1999) examined whether child listeners attend to gestures from a pragmatic perspective. Their findings indicated that messages (e.g. requests) were better understood when they were accompanied by gestures.

Various other findings suggest that adults also extract meanings from gestures. In order to determine whether listeners attend to meanings conveyed by gestures, Cassell et al. (1999) showed a video clip to subjects in which people retold the story of an animated film that they had just seen. However, this video clip was actually acted out based on a script where gestures were created with a slight mismatch to the linguistic utterance. It was found that when the subjects retold what they saw, not only did they

incorporate what they had heard but also what they had seen. In some cases, the subject changed the storyline so that the content fit gestures that they saw. The results are interpreted as evidence that listeners incorporate information encoded in gestures. Riseborough (1981) found that the participants who were shown videotapes of people describing objects were better at correctly naming the objects in comparison to those who only listened to the sound of the tape. Beattie and Shovelton (1999a, 1999b, 2000) found that certain semantic features of iconic gestures, such as *position* and *size* of objects, may be taken into account during communication activities by listeners. In a follow-up study (Beattie & Shovelton 2001, 2003), they found that the understanding of speech and size increases with speech-associated gestures.

Although listeners may take information from gesture, whether gestures are made to be seen by the listener is a different issue. However, studies show that speakers direct listeners' attention to gestures. They do so by performing the gesture close to listeners (Furuyama 2000a, Streeck 1993, 1994). Furuyama (2000a) shows that during origami (paper folding) instruction, speakers performed gestures near the listeners where gestures could easily be seen. Other findings indicate that speakers can direct the attention of listeners by gazing at their own gestures in L1 (Nobe 2000; Streeck 1993; Streeck & Knapp 1992) as well as in L2 (Nash 2001). Expressions such as 'like this' have been noted to direct the attention of a listener to a gesture (Emmet 1998; Furumoto 1997; Heath 1992; Streeck 1993, 1994). Although not directing listeners' attention to gesture per se, pointing directs the listeners' attention to the relevant entity. For instance, a tour guide frequently uses a polite pointing gesture with utterances such as 'this is ~' to direct the attention of tourists.

Another evidence that speakers may produce gesture with the listener in mind is shown by the fact that speakers may change gestures depending on the location and the number of the listeners (Özyürek 2000, 2002b). Özyürek found that when the amount of size of the shared space between the speaker and the listener changed, the speakers changed their gestures accordingly. Furthermore, studies show that C-VPT gestures are more frequently observed during a spatial description task when the speaker and the listener were far (120 cm) from each other rather than close (60 cm) (Hosoma 2001). The willingness of the speaker to convey a message to the listener may also affect the frequency of gesture (Nishio 2000). These studies all suggest that speakers may, to some extent, take listeners into account when producing gestures.

In order to examine whether iconic constitute a part of the speakers' communicative intention, Melinger and Levelt (2004) examined whether speech produced with concurrent gesture is less explicit thatn speech without gestures by using a picture description task. Their results indicate that speakers who produced gestures conveying the spatial information either initially as the overview or during the descriptions omitted more required information in their speech than the speakers who did not produce gestures. Based on the results, Meliger and Level suggest that the speakers intend co-speech iconic gestures to convey information to the listeners (2004).

## 2.3.2  Speakers and gesture

We have so far examined research findings regarding the relationship between gesture and the listener. A question naturally arises with respect to whether gesture is produced solely for interactional purposes. In order to find an answer for the question, a number of studies investigated how the visibility of the listener affects the frequency of gesture, as this is one of the ways to examine whether gesture is produced for the benefit of the listener only or for both the speaker and the listener. Some findings suggest that speakers produce more gestures when listeners are visible than when the listener's view is blocked by a screen or when listeners are only available via an intercom (Cohen 1977; Cohen & Harrison 1973; Emmorey & Casey 2001; Krauss et al. 1995; Ríme 1982). Others suggest that the visibility of listeners does not affect the frequency of gesture (Lickiss & Wellens 1978). Suspecting that such conflicting results may have arisen due to the fact that the visibility of the listeners has a different affect depending on the types of gesture, Alibali, Heath, & Mayers (2001) compared the frequency of representational gesture and *beat* gestures in two conditions, with and without visibility of the listeners. They found that the frequency of representational gestures was higher in the visible condition, although the frequency of *beat* gestures showed no change between the two conditions. Furthermore, Alibali et al. noted that the speakers produced representational gestures even when the listeners were not visible, a finding which suggests that gestures may be produced to benefit not only listeners but also speakers themselves. Bavelas et al. (1992) also found that listener visibility had an affect only on certain types of gestures. Speakers produced more interactive gestures (gestures that refer to the listeners) in the condition where the listeners were visible than when they were not, but listener visibility did not affect the frequency of topic gestures (gestures that refer to the topic of conversation).

The fact that gestures may be produced to benefit both listener and speaker is most clearly demonstrated in the performance of congenitally blind individuals. They produce gestures to both blind and non-blind listeners (Goldin-Meadow 2003b; Iverson & Goldin-Meadow 1997). In some tasks (such as spatial description), the frequency of gestures by blind individuals to blind and non-blind listeners does not show any significant differences (Iverson & Goldin-Meadow 1998). Such findings seem to illustrate the robust nature of gesture, and suggest that gestures are most likely bi-directional, performed for the benefit of speaker and listener alike.

A number of studies examined how gesture benefits the speaker, with a specifi focus on the relationship between gesture and the thinking process. Goldin-Meadow and her associates have been studying gestures produced by children and adults when they are engaged in conceptually demanding tasks (i.e. mathematical tasks, Piagetian conservation tasks for children and explaining how gears which are set up in a complex manner rotate etc. for adults), where gestures convey information which is not expressed in speech (termed 'gesture-speech mismatches') (Alibali, Flevares, & Goldin-Meadow 1997; Alibali & Goldin-Meadow 1993; Church & Goldin-Meadow 1986; Garber & Goldin-Meadow 2002; Goldin-Meadow, Alibali & Church 1993; Goldin-Meadow & Singer 2003; Perry, Church, & Goldin-Meadow 1988; but see

Goldin-Meadow 2003a for reviews of various studies on the phenomenon found in both children and adults).

For instance, Church & Goldin-Meadow (1986) observed the information conveyed in speech and gesture by children who were engaged in Piagetian conservation tasks. In one of the tasks, the children watched the water in a glass poured into a dish, after which they had to explain their judgement about quantity invariance before and after the transformation. The researchers noted that during the explanation, some children conveyed information that was not expressed in speech. For instance, while saying "You poured water from the glass into the dish", a child may gesturally indicate the tallness of the glass and the shortness of the dish. Interestingly enough, among the children who did not provide clear explanations the first time, those who conveyed the complementary information by gesturing scored better in the post-task than those children whose gestures only supplied information that was expressed in speech. The researchers concluded that the complementary provision of information by speech and gesture indexes cognitive instability. In other words, it marks a transition state in the cognitive development of children who are ready for instruction for the acquisition of a new concept (Goldin-Meadow et al. 1993). Similar findings have been obtained in studies which examined the information conveyed in speech and gesture by children while they are engaged in mathematical tasks (Perry et al. 1988).

Based on the accumulated findings on 'gesture-speech mismatches', Goldin-Meadow hypothesized that gesturing may reduce the speakers' cognitive load, thereby freeing up capacity to perform other tasks. In order to examine the hypothesis, Goldin-Meadow and her associates (Goldin-Meadow, Nusbaum, Kelly, & Wagner 2001; Wagner, Nusbaum, & Godin-Meadow 2004) conducted a study which explored how gesturing on one task affected performance on another. In the study, the participants were first asked to solve a math problem on a board, after which they were given words (children) or letters (adults) to remember. Then, they were asked to explain how they solved the problem. When the explanation was over, they were asked to give either the words or letters previously assigned to them to remember. The participants were either permitted or not permitted to gesture during their explanation. The assumption was that if gesturing helped lighten the cognitive load of speakers, they should remember the words or letters better than when they were not permitted to gesture. The results were in accordance with their hypothesis. The participants who gestured during the math explanation performed the other task (remembering words/letters) better. The result suggests that at least some type of gesture may be produced to benefit the speakers.

## 2.4 Mechanism of gesture production

Some scholars have theorised about how gesture is involved in the process of speech production. One view is that gestures are mostly tied up with lexical retrieval (Butterworth & Beattie 1978; Butterworth & Hardar 1989; Krauss, et al. 2000). However, this view does not take into account gestures that encode what is not

encoded in speech. In addition, it does not offer explanations for gestures that overlap with more than one lexical item. Furthermore, the view cannot fully explain findings that gesture may constitute part of the speakers' communicative intention (Meliger & Levelt 2004). Others suggest that gestures are generated 'prelinguistically', before the propositional unit is processed for verbalization (de Ruiter 1998, 2000). According to this view, gestures are performed mainly to remember the imagery (de Ruiter 2000, and see Wesp, Hesse, Keutmann, & Wheaton 2001 for a similar view).

McNeill claims that speech and gesture arise from a single process of language production and that the two modes of expression are considered as two elements of one process (McNeill 1985, 1992; McNeill & Duncan 2000). In order to illustrate the process of gesture-speech integration, McNeill (McNeill 1992, 2000) has presented the notion of 'growth point' which is the smallest essential unit which combines 'imagery and linguistic categorical content' (McNeill & Duncan 2000: 144). One evidence that gesture and speech create meanings together comes from the fact that speakers sometimes encode information in gesture that is not expressed in speech. McNeill (2000) mentions a case where a reference to 'ascending' the pipe by a Spanish speaker was accompanied by a hand that moves upward while fingers were wiggling left and right. Thus, the manner of ascent is expressed by hands. Furthermore, the horizontal direction of hand movement also conveys information about which direction the referent moved (i.e., from left to right). It has been reported that such information is rarely conveyed in speech (McCullough 1993).

McNeill (2000) further argues that the confluence of gesture and speech suggests that the speaker is thinking in terms of a combination of imagery and linguistic categories, with the salient information underlining such thinking concentrated at the point of the gesture stroke. Put differently, by examining gestures and how they synchronize with speech, one can infer the on-line thinking of the speaker. Thus, it is claimed that gestures are 'window(s)' into mental processes (McNeill 1992, 2000).

There is yet another view that gesture plays a role in conceptualising a message to be verbalized (Alibali, Kita, Bigelow, Wolfman, & Kelin 2001; Alibali, Kita, & Young 2000; Kita 2000, Kita & Özyürek 2003). It is assumed that information about action and spatial information interacts with linguistic representation during the production of gesture. All these views (except for the one by McNeill) are reflected in the models relating speech to gesture based on Levelt's (1989) processing model of speech production.

## 2.5 Variation in gesture

### 2.5.1 Culture and gesture
Casual associations between gesture (or other bodily behaviour) and culture abound. For instance, it has been suggested that Japanese speakers learn not to gesticulate (von Raffler-Engel 1975, in Neu 1990). The Chinese, it is claimed, consider the use of many gestures impolite (Chen 1990), while for the Japanese, the ideal is the 'inscrutable' controlled expressionless face (Argyle 1975). Despite a lack of evidence, such views seem to prevail both within and outside the countries concerned. Similarly,

general views about the gestural behaviour of Europeans also exist. The idea that southern Europeans gesticulate more often than northern Europeans is more likely to be met with agreement than denial. One of the reasons why such ideas persistently exist is that they appeal to generally held cultural stereotypes. However, there has hardly been any research suggesting that the frequency of gesture is culturally bound. It is worth noting that Gullberg (1998) found no difference in frequency of gesture during narratives produced by native speakers of Swedish and French, despite the cultural stereotype that northern Europeans are not overtly expressive.

An interesting example of the use of gesture that is indeed culturally motivated is observed in Ghana where pointing with left hand is considered a taboo. It is observed that speakers suppress left-hand pointing, and consequently, right-hand pointing is produced even at the expense of physically strenuous position when speakers point leftward across their body (Kita & Essegbey 2001, and see Kita 2003 for various findings about pointing).

## 2.5.2 Cross-linguistic differences and gesture

Research has shown that speakers of languages such as Guugu Yimithirr in Aboriginal Australia (Haviland 1993, 1996), some of the American Indian (Brown & Levinson 1993) and Mayan Indian languages (Haviland 2000) use pointing that tracks cardinal directions. Research on such use of pointing has lead to the re-consideration of 'linguistic relativity', an idea variously attributable to Humboldt, Boas, Sapir and Whorf, with a focus on how the idea may still be functioning but in a slightly different way from originally conceived (cf. Gumperz & Levinson 1996).

The view that speech and gesture may be interrelated has recently motivated researchers to examine possible cross-linguistic or typological variations in gesture performance (Kita & Özürek 2003; McNeill 2000; McNeill & Duncan 2000; Özyürek & Kita 1999). Kita & Özyürek (Kita & Özyürek 2003; Özyürek & Kita 1999) have examined gestures accompanying the description of motion events by speakers of English, Turkish and Japanese. Their analyses of how native speakers from three language groups describe a particular motion event (ie. 'swing') have shown that speakers of Turkish and Japanese, languages which lack the equivalent of an English verb 'swing', gesturally describe the motion event by mixing straight and curving trajectories, while speakers of English unilaterally depict an arc. In other words, these findings indicate that gesture performance may be influenced by aspects of the language spoken.

McNeill and his associates have attempted to apply 'thinking for speaking' (see 1.4.2), a notion advocated by Slobin (1987, 1996a, 1996b), to the analysis of gesture (Duncan 1996; McNeill 2000; McNeill & Duncan 2000). They investigated whether speakers of typologically different languages exhibit differences in their gestural patterns when describing motion events. When gesture production by native speakers of a path-verb language (i.e. English) and a manner-verb language (i.e. Spanish) is compared, the findings show cross-linguistic variation (McNeill 2000; McNeill & Duncan 2000). As we have stated above, Spanish speakers often only mention the path of the motion and leave out the manner in speech, which is often encoded in gesture.

However, such gesture patterns are not observed in the gestures made by English native speakers (Slobin 1996b). Given that English has more manner-of-motion verbs than Spanish, it was suggested that the typological differences in the language spoken are reflected in gesture.

Duncan (1996) has compared the gestures of native speakers of Chinese and English and noted some differences between the two groups with respect to the establishment and maintenance of TOPIC reference. For instance, a high frequency of repeated gestures among speakers of Chinese, as compared to English speakers was observed. The repeated gestures were sometimes used as a means to index the chain between given information and the focus of the subsequent discourse. In other words, the gesture accompanying the COMMENT in a TOPIC-COMMENT clause was repeated in synchrony with the COMMENT as it became the TOPIC in the immediately succeeding utterance (1996: 183-4). These repeated gestures thus appeared to be related to topic establishment or maintenance (1996: 185). Since the above findings were obtained at the level of the sentence, Duncan suggests that how speakers of typologically different languages construct discourse and perform gestures at a level beyond single utterances remains an issue that warrants further investigation (1996: 204).

## 2.6 Gestures by adult second language learners

### 2.6.1 Descriptive findings of gesture in L2
Despite the omnipresent characteristics of gesture, systematic studies of gesture in L2 are scarce. The majority of studies have been descriptive, and with a few exceptions, they are conducted with relatively little attention paid to theoretical issues in second language acquisition (SLA). Similar to studies of gesture in L1, gestures in L2 are examined with various foci from pedagogical, interactional to psycho-linguistic. As in L1 studies, gestures are typically categorized as a non-verbal or kinesic behaviour, and the definition of gesture is rarely provided. Furthermore, except for a few recent studies, the focus of investigation has mostly been on what gesture adds to the meaning conveyed by the speech channel rather than how speech and gesture as one system create meaning.

Research has suggested some pedagogical implications of gesture for language learning (Harris 2003; Kellerman 1992), teaching (Antes 1996; Hurley 1992) and assessment of learners' competence in communication (Jungheim 1994, 1995; Neu 1990). The importance of gesture by teachers as input has also been noted (Lazaraton 2004). However, there has so far been no systematic study conducted to examine whether or not learners indeed take account of the information provided in the gestural channel. In addition, questions remain with respect to whether gestures (other than 'emblems') are teachable or whether the teaching of gestures enhances language learning (but see Allen 1995 which suggests that presenting foreign expressions with emblematic gestures help retain meanings). Furthermore, the relationship between the use of gesture and proficiency is the area that requires more investigation (see below). In the rest of the section, we will focus on the gestures produced by learners.

A number of studies have examined the quantitative differences in gesture production between in their L1 and L2 (Chen 1990; Dushay 1991; Gullberg 1998, 2003; Hadar, Dar, & Teitelman 2001; Kita 1993; Marcos 1979; Nobe 1993; Sainbury & Wood 1977; Seto 2000; Yanagimachi 1999). Several studies have indicated that speakers produce more gestures in L2 than in their L1. For instance, it has been noted that English-Spanish and Spanish-English bilinguals use more gestures of all kinds in the weaker language (Marcos 1979). Japanese learners of English and English learners of Japanese have been noted to produce a larger number of gestures in L2 than in L1 (Jungheim 1995; Kita 1993; Nobe 1993). In addition, French learners of English have been observed to produce more gestures in L2 than in L1 during descriptive and emotional oral tasks (Sainbury & Wood 1977). French learners of Swedish and Swedish learners of French also produce a larger number of gestures in L2 (Gullberg 1998). Hadar, et al. (2001) found similar results in Israeli learners of English.

However, other studies present different findings (Chen 1990; Dushay 1991; Seto 2000). Dushay (1991) found that as speakers switched from L1 to L2, the frequency of gestures decreased. Dushay studied gestures made by English learners of Spanish in two referential tasks, the description of novel abstract figures and novel synthesized sounds. The results showed that when the learners described a novel sound, there was a smaller number of 'speech-focused gestures' in comparison to native production. Seto (2000) compared the frequency of gesture in eight Australian speakers of Japanese and eight Japanese speakers of English who were simultaneously engaged in discussion. The findings showed that, for both groups, the frequency of gesture was higher when the speakers were engaged in conversation in English. However, there are number of methodological weaknesses, thus it is difficult to judge the results.

Variations among studies are found with respect to which types of gesture are frequently made in L2. Gullberg (1998) noted a significant increase in *metaphorics* and *deictics* in L2, while speakers used more *iconics* in L1 than in L2. An increase of *deictics* in L2 was also noted by Sherman & Nicoladis (2004). Yanagimachi (1999) also reported a higher frequency of *beat* and *deictic* gestures by learners during a story-retelling task than their native counterparts, for whom *iconics* predominated. However, no baseline data is presented in this study, so it is difficult to compare its results with other findings. A significant increase in *beats* (Marcos 1979; Nobe 1993) and *representational* gestures (i.e. *iconic*, *metaphoric* and *deictic* gestures) in L2 has been noted (Nobe 1993). On the other hand, Sherman & Nicoladis (2004) found no difference between L1 and L2 with respect to the frequency of the use of *iconics*.

There seem to be a number of reasons why quantitative findings about L2 gesture are contradictory and difficult to interpret. First, what is meant by 'gesture' may vary among studies. Since some studies do not provide a clear definition of the term, it is difficult to judge whether 'gesture' includes only co-speech gestures or any hand movements regardless of speech accompaniment. Second, how gesture is counted and classified may vary among studies. As we have mentioned in passing, gestures are frequently 'densely-encoding' (Duncan 1996: 21). For instance, *iconics* may sometimes be superimposed by *beats*. Such gestures can be counted only as *iconics* or the combination of *iconics* and *beats*. Moving a hand across the gesture space to show

the movement of a figure may be interpreted as *iconic*, *deictic* or both[5]. Without a clear coding system, validity and reliability in analysis may not be achieved. However, the information of how gesture is classified and counted is frequently missing from the description of analyses, making comparison of the results difficult. In fact, it has been recently noted, 'there has been so little empirical work done on gesture in applied linguistics that there is really no "standard" available for transcription, and each researcher is left to herself to develop or employ a system that suits her purpose' (Lazaraton 2004: 93). However, as we have shown, gesture studies in L1 have demonstrated the importance of the semantic/temporal synchronization between speech and phases of gesture. In addition, handedness, hand-orientation, position in the gesture space and viewpoints of gesture are all useful information in analyzing gesture behaviour. Accordingly, a system for coding L2 gestures should be guided by these notions, not by each researcher's purpose for a study. Third, task variations may affect the results (Gullberg 1998: 159). Fourth, many studies do not offer SL baseline data. As with any comparative linguistic analysis of L1 and L2, baseline data are crucial in order to properly assess gestural trends in group performance.

There are relatively few studies that address the issue of the interrelationship between linguistic proficiency and gesture (Chen 1990; Gullberg 2003; Kita 1993). Kita (1993) found that the number of gestures decreased as the oral proficiency of learners increased. Similarly, Gullberg (2003) found that the number of abstract deictic gestures decrease as the proficiency of the learner increases. On the other hand, Chen (1990) found no significant difference in the frequency of gestures between Chinese learners of English with high and low proficiency. However, as has been rightly suggested (Gullberg 1998), without baseline data, it is difficult to judge Chen's explanation that the learners' low frequency of gesture may be due to the fact that the use of gestures is reputedly considered impolite in Chinese culture.

### 2.6.2 Gesture and SLA studies
L2 gestures have been systematically framed in taxonomies of communication strategies (CS) (Bialystok 1990, 1994; Chen 1990; Dörnyei & Scott 1997; Faerch & Kasper 1983; Gullberg 1998; Kellerman 1991; Kellerman, Bongaerts & Poulisse 1987; Paribakht 1985; Poulisse 1987, 1990, 1993; Tarone 1977, 1983). However, except for Gullberg (1998), they are not about gesture *per se* and the treatment of gesture is generally monolithic. Traditionallyt, L2 gesture are seen as a non-verbal strategy performed primarily to replace missing lexical items (Tarone 1977), and accordingly *mime* has been the focus of attention. More recent accounts describe gesture beyond this lexical substitute view (cf. 'replacing' and 'accompanying' gestures in Paribakht 1985), although *mime* remains the dominant focus. *Mime* may sometimes be categorized as a 'paralinguistic strategy' (Chen 1990; Dörnyei & Scott 1997) without detailed specifications as to what this strategy entails. In many of the taxonomies, the word 'gesture' is often used interchangeably with *mime,* reflecting a loose definition.

---

[5] For instance, Gullberg coded gestures indicating movement of referents across discourse space as deictic gestures (1998:94), while these gestures were coded as iconic gestures by Sherman & Nicoladis (2004:151).

In addition, the taxonomies are commonly based on the surface phenomena where gesture forms an independent category separate from oral strategies.

On the other hand, gesture is treated together with oral strategies in cognitively oriented frameworks of CS as manifestations of shared processes (Bialystok 1990; Kellerman 1991; Kellerman et al. 1987; Poulisse 1990, 1993). According to the Nijmegen Group, gestural and oral strategies share two underlying processes, 'concept-based' and 'code-based' ('co-co' processes) (Kellerman 1991; Kellerman et al. 1987; Poulisse 1990). It is claimed that the former entails the manipulation of the concept, while the latter involves the manipulation of the means of encoding. With respect to gesture, the former process is manifested as *iconics* and *mime* which express semantic properties of referents. In contrast, ostensive gestures reflect the latter process. Although gesturing by definition involves 'code-based' processes as there is a change of means of expressions, the distinction between *iconics* and pointing gestures is based on the idea that the former may involve a process operating at the conceptual level where characteristics of the referent are exploited visually. A code gesture merely point to the property (colour, item) for which the learner has no name (Kellerman 1991).

Gullberg (1998) systematically examined L2 compensatory gestures by applying the aforementioned Nijmegen taxonomy along with three other categories of her own, namely, 'hedging', 'avoidance' and 'overt appeal' (1998: 100). Her findings show that learners produce a high frequency of abstract deictic gestures which function differently from the 'code' strategies based on concrete deictic gestures or from oral 'code' strategies. Gullberg maintains:

> 'Code strategies based on abstract deictics … direct attention not only to another mode of expression, but also to a different level, discourse. The principal strategic function at discourse level is to create redundancy, a very different function from that oral Code strategies.'
>
> (Gullberg 1998: 219)

Thus, Gullberg's (1998) findings provide counter-evidence to a commonly held view that learners produce more gestures in their L2 than in L1 in order to compensate for information (mostly lexical) that is not provided in speech. Gullberg (2003) examines this issue of redundancy further in a separate study (see 2.7 below). In addition to the high frequency of abstract deictics, her findings also indicated the occurrence of gestural hedging. Given that gesture could only function as a strategy in the presence of interlocutors, a functional approach to the studies of gestural strategies was called for (cf. Kida & Faraco in press).

In addition to studying gesture in L2 with the framework of communication strategies, some researchers have framed gesture with the perspective of inner speech (McCafferty 1998; McCafferty & Ahmed 2000; Negueruela, Lantolf, Jordan, & Gelabert 2004).

### 2.6.3 Interaction between SL and TL in L2 gesture

Some researchers have investigated whether L1-based gesture patterns transfer to L2 (Kellerman & van Hoof 2003; Özyürek 2002a; Stam 1999). Stam (1999) compared the gesture patterns of intermediate and advanced Spanish learners of English with Spanish and English base-line data. Theoretically framed within Talmy's typology of motion events, the study investigated gestures that coexpressed motions. The focus of the analysis was the expressions of path elements in the retelling of a cartoon story. The typical Spanish pattern is to place a gesture on the verb which expresses the element of path in a motion event, which differs from the typical English pattern where the path gesture occurs on the satellite or ground NP. The question was whether the L1 gestural pattern could be observed in L2 gesture production. The results showed that this was indeed the case. The results for the manner gestures were inconclusive partly due to the small number of instances.

Van Hoof (Kellerman & van Hoof 2003; van Hoof 2000) also examined the path gesture patterns of native speakers of English, Dutch and Spanish, and Dutch and Spanish learners of English. Similar to the study by Stam (1999), the focus was on the synchronization patterns between speech and gesture with respect to information about the path element in motion events. The results indicate cross-linguistic variations in gestures. The gestures produced by Spanish native speakers showed the same pattern as in Stam (1999). On the other hand, the speakers of Dutch and English, while belonging to the same typological group, revealed intra-typological differences in their gesture patterns. The patterns found in the Dutch L1 and Spanish L1 speakers were also apparent in L2 production. The research findings indicate the likelihood of the cross-linguistic influence of the source languages on L2 production. Finally, Özyürek (2002a) investigated gestures made by Turkish learners of English. Özyürek found that the pattern of acquisition noted in speech was mirrored in gesture. In other words, in comparison to the beginning/intermediate learners, the advanced Turkish learners of English in her study not only produced a higher proportion of expressions conflating manner and path of motion in speech but also of gestures encoding both manner and path of motion.

With a few exceptions, the studies reviewed so far have examined gestures in isolation, extracted from surrounding utterances. As a consequence, the association between gesture and the ongoing discourse context is generally missing from the analyses, which have tended to focus on intrasentential elements. In the following, we will review studies of L1 and L2 which have focused on gesture at the level of discourse.

### 2.7 Gesture in L1 and L2 narrative discourse

Studies have indicated that some gesture performance may be related to the structure of discourse (Cassell & McNeill 1991; Furuyama 2000b, 2001; Gullberg 1998; 2003; Levy & Fowler 2000; Levy & McNeill 1992; Marslen-Wilson, et al. 1982; McNeill 1992; McNeill, Cassell, & Levy 1993; McNeill & Levy 1982, 1993; McNeill, Levy, & Pedelty 1990; O'Neill & Holmes 2002). It has been suggested that gestures may mark

the transition between narrative levels. McNeill (1992) distinguishes three levels of narrative: narrative proper, meta-narrative and para-narrative. Narrative proper is where reference to events is made. Meta-narrative is the level where reference is made to the storyline or the structure of the narrative. For instance, the speaker's comment such as 'The boy was probably sad' is considered a meta-narrative comment. Para-narrative is the level where the reference is made to the physical setting and interlocutors. A question by a speaker to a listener such as "Have you ever seen a comic called ~?" is an example of a para-narrative statement.

Studies have shown that the transition from one narrative level to another is often accompanied by *beat* gestures (McNeill 1992; McNeill, et al. 1990), and that gestures mark the introduction of new topics or unexpected lexical information. It has been pointed out that at the start of a new episode in a narrative, referent(s) are marked by gesture (Marslen-Wilson et al. 1982; McNeill 1992; Levy & McNeill 1992). Adopting Givón's scale of *linguistic quantity* which states that 'the less predictable/accessible/continuous a topic is, the more coding material is used to represent it in language' (Givón 1985), McNeill (1992) presents a similar progression of gestures as in Figure 2.9. The arrow indicates an increase in 'communicative dynamism' (Firbas 1964 in McNeill 1992), which is defined as the extent to which the message at the point "pushes the communication forward" (McNeill 1992: 207). The basic assumption underlying the McNeill scale is the same as Givón's: when the information is predictable (given), there is less need for coding material. However, when the information is unpredictable (new), the speaker will provide more complex material in both speech and gesture (Levy & McNeill 1992; McNeill 1992; McNeill & Levy 1993).
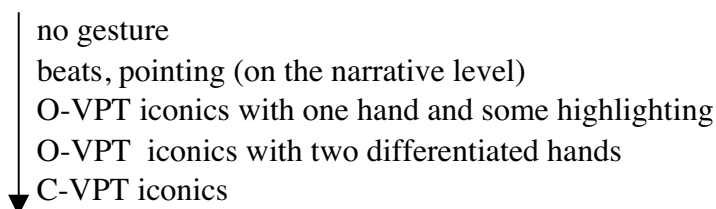
no gesture
beats, pointing (on the narrative level)
O-VPT iconics with one hand and some highlighting
O-VPT iconics with two differentiated hands
C-VPT iconics

**Figure 2.9.** Progression of gestures (based on McNeill (1992)).

As in speech, referents are manually tracked throughout the development of a storyline. Unlike gestural marking of the introduction of referents, which is generally performed by *iconics* and *deictics*, gestural tracking of referents is often performed with abstract *deictics* (Levy & McNeill 1992; McNeill 1992; McNeill et al. 1990). Thus, gesture accompanying the introduction of a referent performs two tasks. The first is to highlight its newness, and the second is to assign a certain locus in the gesture space for the particular referent. In other words, gesture creates an association between the referent and the space in front of the speaker. When the referent is mentioned again later in the discourse, the hand may point to the same locus as if to reactivate the association between the referent and its assigned position in space. Thus, gestural anaphoric linkages are performed visibly by pointing to a designated location

in the gesture space (Gullberg 1998). It is worth noting that such gestures create redundancy between speech and gesture which is characteristically different from other gestures such as *iconics* which often complement the information conveyed in speech.

In addition to the repetitive use of the same location in the gesture space, the recurrence of some physical gesture features may help establish cohesion in discourse. The notion of 'catchment' has been proposed to capture the aspect of gesturally cohesive linkages (McNeill 2000). Any partially or fully recurring features of shape, space, movement, orientation, dynamics etc are recognized as such and are considered to provide a 'window into discourse cohesion' (McNeill 2000: 316). Furuyama (2000b) shows how repetition of an *iconic* gesture helps clarify the identity of the intended referent of zero anaphora when there is more than one possible candidate for the animate entity in subject position. Furthermore, it has been suggested that hands register a referent's movements (McNeill et al. 1993). Thus, when a referent moves from one point to another during the development of the storyline, the hand traces its move. As a consequence, the gesture space in front of the speaker is charged with multi-layered discursive meaning.

The gestural introduction and tracking of referents in discourse is not a phenomenon limited to native adult speakers. Research shows that children make use of gesture to refer to story characters (O'Neill & Holmes 2002). With respect to L2 speakers, Gullberg (2003) has examined gestures accompanying introduction and tracking of animate referents in the narrative discourse of adult learners. Three informational types of referents were distinguished, *introduced, maintained* and *re-introduced*. Gullberg found that French learners of Swedish and Swedish learners of French perform similarly to natives with respect to the frequency of gestural marking of *introduced* or *re-introduced* referents. However, the learners produced gestures accompanying *maintained* referents significantly more than the native speakers of either TL. Given that the learners who had trouble with the complex pronominal systems of French and Swedish often used NPs for *maintained* referents in speech, Gullberg concludes that over-explicitness in speech is mirrored in the gestural performance of the learners (2003). While Gullberg's insightful study reveals the complex inter-relationship between speech and gesture in L2 discourse, the target languages of her study share cross-linguistic similarities. What is not known yet is how learners with typologically different SL and TL approach the TL conventions for linguistic and gestural referential management.

## 2.8 Research questions

The present work takes the line of investigation on linguistic and gestural (i.e. bi-modal) reference introduction and tracking further by examining the gestures of native speakers of typologically different languages and adult language learners. The study is exploratory in nature and does not aim to test any hypothesis. Thus the central aims of the present study are the descriptions and explanations of speakers' (both L1 and L2) preference for bi-modal means in encoding reference to entities in narrative discourse.

Based on a thorough survey of the literature, the following three general research questions were formulated:

1)  Do speakers of typologically different languages (i.e. Dutch and Japanese) show cross-linguistic variation in their bi-modal introduction and tracking of referents in narratives?
2)  What are the similarities and differences of bi-modal reference introduction and tracking between those by Dutch learners of Japanese and by native speakers of Dutch and Japanese?
3)  How can learner preferences for linguistic devices and gestural markings be explained?

With respect to Question 1), the following subquestions were formulated:

●  Does the availability of various linguistic means to mark the information status of referents affect the quantitative and qualitative aspects of gestures accompanying the introduction and tracking of animate referents in discourse? Would having various linguistic means to mark the differences in the information status of referents result in less reliance on gestural resources for marking and distinguishing referents? Alternatively, do referential gestures show quantitative and qualitative similarities among native speakers regardless of the language spoken? Moreover, do differences in organizing discourse to describe inanimate referents affect the quantitative and qualitative aspects of gestures accompanying the introduction and tracking of inanimate referents in discourse?

With respect to Question 2), the following subquestions were formulated:

●  Which linguistic devices do learners use to introduce and track referents? Do learners' choices of linguistic forms reflect any principles of organizing information in narrative discourse? Do learners' gestures show quantitative and qualitative similarities with native speakers of SL or TL?

The sub-question for Question 3) is the following:

●  How does the SL and the TL interact in learners' gestural encoding of reference to entities in L2 narrative discourse?

Thus, the present study focuses on L2 learners' linguistic and gestural referential management in the light of SL and TL conventions. The following chapter presents the methodology adopted for the study.