Chapter 1. Introduction

1.0 Introductory remark

The present work examines how speech and gesture together contribute to create meaning in discourse in both first and second language settings. As any observer will confirm, the movement of hands and arms frequently synchronizes with on-going speech, regardless of the language spoken and whether or not speakers are operating in their mother tongues (henceforth L1) or foreign/second/third languages (henceforth L2). The hand and arm movements may occur during exchanges between interlocutors or when speakers are engaged in monologue. Speech is linear and audible, while gesture is dynamic, spatially organised and visual. The two modes of expressions have distinctively different characteristics, yet both are utilized during communication. The present work is an attempt to examine how the two modes of expression complement each other, with a specific emphasis on how animate and inanimate referents are introduced and tracked in the production of narrative.

1.1 Organization of the study

The thesis is organized as follows: In the rest of this chapter and Chapter 2, we present the scope of the present work and survey of the relevant literature. The research questions are presented after the literature review in Chapter 2. Chapter 3 outlines the methodology adopted for the present study. Analyses of data are provided in Chapters 4, 5 and 6. Chapter 4 presents the findings on the introduction of animate referents in speech and gesture in L1 and L2 narratives. Chapter 5 presents the findings on the tracking of animate referents in these narratives. Chapter 6 presents the findings on the introduction and tracking of inanimate referents in speech and gesture in L1 and L2 narratives. Finally, Chapter 7 presents a summary and discussion of the findings and a general conclusion.

1.2 Scope of the present work

Reference introduction and tracking is a communication act in which speakers provide information about entities (animate, inanimate) in a stretch of discourse. For this to be successful, entities should be clearly identified, properly located in reference to other entities and followed through various events that may unfold in the discourse. For instance, let us imagine a witness reporting an accident as follows:

---

1 Although discourse could be both spoken and written, we will focus on the former in the rest of the discussion.
(1) I saw this person running down the street, and then he banged into a tree and fell over.

In (1), the speaker introduces a new animate entity, *this person*, and locates him in reference to an inanimate entity ‘the street’. The animate referent is then followed, or tracked, in the two subsequent events where he ‘banged into a tree’ and ‘fell over’. If we examine the statement closely, we realize that although the speaker is talking about the same person, linguistic forms used to refer to him change from ‘this person’ to ‘he’ and finally to ø through zero anaphora. Despite the changes, the listener would know that the statement is about one and the same person. As this example shows, reference making is not based on an absolute one-to-one relationship between words and persons/things. It is rather an act whereby the speaker uses available means to enable the listeners to identify the intended referent (cf. Lyons, 1977).

The type of information provided by referents varies according to their animacy. While animate referents are mostly the actors in events that constitute the core of discourse, inanimate referents often offer background and contextual/spatial information. In (1), ‘the street’ locates the animate referent in a spatial frame and ‘a tree’ represents the target of his action, ‘banging’, although the type of information provided by inanimate entities may be influenced by the nature of discourse. For instance, an enthusiastic real estate agent’s description of a house may differ from that of someone who is trying to direct others over the phone to find an item in the house. It is likely that inanimate referents in the latter discourse offer more locational information than the former, in which inanimate referents may be associated with physical and atmospheric qualities such as ‘spaciousness’ or ‘luxury’.

The data used for the present work consist of spoken narratives. It has been shown in the literature that narrative is a highly structured type of discourse (Labov 1972; Labov & Waletzky 1967), and that the organization of the entire narrative text can be viewed as an answer to an underlying question ‘What happened?’ (von Stutterheim & Klein 1989). The selection of the relevant information for the description of events is made according to the demands of this question. For instance, in the narrative of story-telling, information about the characters, the sequence of events and the locations where these events take place are considered most relevant.

In order to understand reference introduction and tracking, it is also important to note that utterances in discourse do not generally exist independently of contextual information. In other words, “the information conveyed in the wording of utterances available to the speaker and the listener at the given situation” is related to information regarding knowledge of the world, preceding information and the situational knowledge (Klein 1986: 113). To use (1) as an example again to show how contextual information is interrelated with the utterance, for the description of an accident, the speaker chooses to make reference to one animate and two inanimate entities, frames the scene at ‘the street’, but decides not to explicitly mention the relative spatial relationship between the inanimate entities. The speaker relies instead on the listener’s knowledge of the world to fill in the gap. In addition, each referring form is chosen
based on the preceding information that provides the cue for the listener to correctly identify the intended referent. Furthermore, extra-linguistic features such as voice quality and gesture may be available as situational information. The speaker in (1) may use hands to represent a person and a tree to describe the event of a person banging into a tree, with the hand movement revealing the speed and the harshness of the bang.

Although (1) is a short utterance, narrative is usually a long stretch of discourse, consisted of a number of connected utterances. During narrative, speakers usually introduce various animate and inanimate referents. Since narrative has layers of information, new and given, intertwined in a complex way, for narrative to be intelligible, it is essential that identities of referents are made clear. The question is how speakers manage this task, and how utterances interrelate with contextual information.

Because points in space are uniquely distinguishable from each other, this characteristic is incorporated in pronominal systems in signed languages. Pointing to a particular location in space is used as a way to create anaphoric linkages within discourse. Bellugi & Klima (1982) describes how pointing may be used in American Sign Language as follows:

If a referent (third person) is actually present in the discourse context between signer and addressee, specific indexical reference is made by pointing to that referent. But for non-present referents that are introduced by the speaker into the discourse context only ‘verbally’, there is another system of indexing. This consists of introducing a nominal and setting up a point in space associated with it’ pointing to that specific locus later in the discourse clearly ‘refers back’ to that nominal, even after many intervening signs.

(Bellugi & Klima 1982:301)

Speakers may also use gesture to distinguish referents from each other and to create anaphoric linkage. In addition, because of its visual characteristics, gesture may offer information about physical aspects of the referents they accompany.

Although all speakers may produce utterances that fit appropriately to the entire ‘flow of information’ (Chafe 1994), languages provide different linguistic means to structure information about referents. Accordingly, with respect to the surface form, reference introduction and tracking may reflect cross-linguistic variation. Given the interrelationship between utterances and contextual information, gestural introduction and tracking of referents may reflect such cross-linguistic differences in speech. One of our aims of this study is to investigate this issue. However, there is another issue that we intend to pursue as the main purpose of the study, i.e. how second language learners cope with the task of linguistic and gestural marking of referents.

Due to the fact that producing narrative requires speakers to simultaneously attend locally to produce correct forms and globally to the flow of information, one can easily imagine that this is a cognitively demanding task for learners who are in the process of acquiring a new language. It has been suggested that producing utterances that fit the
surrounding information appropriately may pose two kinds of an ‘embedding problem’ for learners (Klein 1986). The first is judging what should be explicitly mentioned in an utterance and what can be left out given the contextual information. The next is the use of necessary linguistic elements that are available in the target language (Klein 1986: 112). Although we all perceive and talk about the world that surrounds us, how information about events is packaged and filtered into linguistic forms in a linear order is prone to cross-linguistic differences, and languages provide different means to express the organization of information. It is thus conceivable that both aspects of the embedding problem may interact with factors that learners bring to L2 discourse, learners’ knowledge of the source language (SL) and of the target language (TL). Furthermore, given the interrelationship with utterances and contextual information, the interaction between the SL and the TL may be reflected in gesture. To investigate this issue is the main aim of the study.

Thus, the scope of the present work needs to be relatively wide-ranging, and to include cross-linguistic differences in referent introduction and tracking in L1 speech and gesture, as these serve as the baseline for the examination of how gesture is integrated into L2 discourse production. In the following section, we will review the relevant literature accordingly.

1.3 Aspects of spoken discourse construction: principles at work

In order to produce understandable and meaningful discourse, speakers have to attend to various aspects of discourse construction. They first select, segment and construct information as propositional units, which are then processed for verbal production as utterances (Levelt 1989). The connected utterances form a text. For any discourse to be comprehensible, it is crucial that the text is linked in an understandable manner. Compare the previously used example (1) and (2) which is slightly different:

(1) I saw this man running down the street, and then he banged into a tree and fell down
(2) I saw him running down the street, and then this man banged into a tree and fell down

Upon hearing (2), the listener might feel confused as to the identity of the referent denoted by ‘him’. The identity of ‘this man’ is not clear, either. Thus, we intuitively know that the order ‘NP then pronoun’ does not cause problems in identifying the referent, but the use of a pronoun for a newly introduced referent does. Thus, the comparison of (1) and (2) shows the importance of the appropriate linkage in creating meaningful discourse.

The linguistic devices which help establish linkages have been variously termed anaphora or deixis etc. (Fillmore 1975; Halliday & Hasan 1976; Hanks 1992; Levinson 1983; Lyons 1977). The linking device, like pointing with a finger, indexes a referent, but in the discourse-internal context. Some researchers may distinguish between anaphora and cataphora (Bühler 1934:121 in Lyons 1977; Halliday & Hasan 1976). The distinction is made according to the direction of reference. Anaphora creates a linkage in a backward-looking manner - the attenuated form follows the
expression to which it is related as in (3). *Cataphora* creates linkage in the opposite manner to anaphora in that the attenuated form precedes the expression it is correlated with as in (4).

(3) I bought some apples yesterday. They were very sweet.
(4) I don’t think he can eat this dish. John is vegetarian.

However, anaphoric anchoring of referents—anaphoric linkage—is more prevalent in discourse, so that some researchers may use *anaphora* to refer to the both type of anchoring (Lyons 1977). We will adopt the same position whereby *anaphora* refers to both *anaphoric* and *cataphoric* expressions.

Types of anaphora vary with regard to the aspect of the speech situation that is linked between the ongoing and preceding utterances. Links can be made between animate entities, for instance. Referring forms such as third person pronouns (e.g. *he, they*), definite articles (*the frog*) or demonstratives (*that boy*) may be utilized for indicating the identities of the referents in discourse. Links may also occur in the domain of space. Spatial deictics such as *there* or an adnominal demonstrative *that*, which denotes the static scene as in *underneath that* may be used. Temporal or causal links may also be realized by various types of connectives, such as *when, while or because*.

The principle that governs anaphoric linkage between referents is the structure of information. According to this notion, speakers linguistically mark the flow of information in discourse so that meaningful messages can be decoded efficiently by listeners within their cognitive processing capacities. In order to facilitate the process, speakers indicate whether the information carried by a referent at the time of the utterance is new or given. In other words, by marking the information status of referents, speakers show the degree of presupposition expected of the listeners for the interpretation of the identities of the intended referent. Various notions and views on information structure in discourse have been proposed (e.g., Ariel 1988, 1990, 1994; Chafe 1974, 1976, 1987, 1994; Clark & Haviland 1977; Givón 1983, 1984, 1985; Gundel, Hedberg, Zacharski 1993; Halliday 1985; Halliday & Hasan 1976; Haviland & Clark 1974; Prince 1981).

In the following, we will briefly review Givón’s position (Givón 1983, 1984, 1985), as it has been widely adopted in studies of referential introduction and tracking in both L1 and L2. Framing the information structure with the notion of *topic continuity*, Givón (1983) argues that a topic continuing from the preceding clause will be more predictable, and as a result, may be easier for the listener to process than new or re-introduced topics. He illustrates the relationship between the various indexical markings and the new and given informational status of a referent as in Figure 1.1.

According to Givón’s hierarchy of forms, zero anaphora shows the strongest relationship to continuous topics, while full NPs are most likely related to discontinuous topics. Givón (1984) also presents the principle of *the quantity universal* which states that ‘more continuous, predictable, non-disruptive topics will be marked by *less marking material*; while less continuous, unpredictable/surprising, or
disruptive topic will be marked by *more marking material*’ (Givón 1984: 126, italics in the original).

*more continuous/accessible topics*

*Zero anaphora*

Unstressed/bound pronouns

Stressed/independent pronouns

*Full NPs*

*more discontinuous/inaccessible topics*

<table>
<thead>
<tr>
<th>NEW (discontinuous)</th>
<th>GIVEN (continuous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>een kikker</td>
<td>het kikker</td>
</tr>
<tr>
<td><em>a frog</em></td>
<td><em>the frog</em></td>
</tr>
<tr>
<td>kaeru</td>
<td>kaeru</td>
</tr>
<tr>
<td><em>frog</em></td>
<td><em>frog</em></td>
</tr>
<tr>
<td>hij</td>
<td><em>he</em></td>
</tr>
<tr>
<td><em>ø</em></td>
<td><em>ø</em></td>
</tr>
</tbody>
</table>

**Figure 1.1.** Topic continuity and referential forms (based on Givón 1983)

For instance, Figure 1.2 shows how differences in the informational status of referents may be encoded in the choice of referential forms in Dutch and Japanese according to Givón’s view. Thus, in Dutch, the form related to the most discontinuous topics is a full NP with an indefinite article *een*. The next in line is an NP with a definite article *het*, followed by a pronoun, and lastly by zero anaphora (*ø*). In Japanese, there are basically two forms, NP and zero anaphora (*ø*). Although numeral classifiers are used optionally, there is no obligatory marker to be used with NPs.

**Figure 1.2.** Informational status and referential forms in Dutch and Japanese

The Givón’s principle of *the quantity universal*, has been investigated with various languages, and there is general consensus about the view in the literature (Clancy 1980, 1992; Clancy & Downing 1987; Givón 1983, 1984; Hinds 1983; Marslen-Wilson, Levy, & Tyler 1982).

While Givón’s view is based on the horizontal distance between the present and the last mention of a referent in discourse, the choice of referential forms may also be influenced by differences in the referential importance among characters. For instance, Chafe (1994) shows that protagonists in stories are more likely to be introduced with proper names than peripheral characters. In the developmental studies, children after a certain age treat the protagonists and peripheral characters differently like adult speakers with respect to their choice of referential forms (Karmiloff-Smith 1981, Wigglesworth 1990). Thus, reference to animate entities in narrative discourse is governed basically by two factors. One regards the horizontal distance between mentions of referents, and the other regards the vertical differences between the importance of characters. However, how these two factors, in particular the former, are realized by available linguistic means show variation among languages.
1.4 Reference introduction and tracking cross-linguistically

1.4.1 Reference to animate entities

As mentioned in passing, various studies have verified the claims by researchers such as Chafe (1994) and Givón (1983, 1984) that referents being introduced for the first time are linguistically indicated differently from their subsequent mentions by the choice of referential forms. The marking of newness occurs both locally (i.e. at the level of noun) or globally (i.e. at the level of clause).

With respect to local newness marking, in most of the Indo-European languages, nominal determiners, such as definite and definite articles, are obligatory and are used to distinguish the informational status of a referent. Observe the following example:

(5) A boy had a dog and a frog. The frog was kept in a jar.2

In (5), when *frog* is newly introduced, it is marked by an indefinite article as in *a frog*. Upon the second mention, the same referent is marked with a definite article as *the frog*. The articles clearly distinguish the two NPs as new vs. given information. However, such local newness marking with article systems is not universal. In languages such as Polish, Chinese, Japanese and Finnish, these devices are not available. Accordingly, other means of marking the informational status of referents are adopted. For instance, in Japanese, numeral classifiers such as *h(p)iki* (a counter for animals) may be utilized in the place of an indefinite article, although its use is not obligatory and speakers may prefer to use other structures to mark the informational status of a referent. In addition, in Japanese, the informational status of the referent may be reflected in the choice of the post-positional particles. Studies indicate that the use of the nominal particle *ga* in Japanese is associated with the introduction of a referent (Hinds 1983; Nakahama 2003b; Ono, Thomson, & Suzuki 2000). Similarly, in Korean, another language with post-positional particles, the nominal particle *ka/i* is clearly preferred to other particles when referents are introduced (Nakahama 2003b).

With respect to global marking, various studies have shown how *new* and *given* differences in information status may relate to the syntactic/semantic roles assumed by the referents (Du Bois 1985, 1987; Du Bois, Kumpf & Ashby 2003). Preferred argument structure, as it is termed, presents views that are relevant to the present work. It states that speakers irrespective of the language spoken tend to avoid assigning agent role (A) to the newly introduced referent. Instead, new referents assume either subject role of an intransitive verb (S) or become the transitive object (O). Let use give an example. An English fable traditionally starts with a line like the following:

(6) One upon a time, there was a king who had a beautiful wife.

---

2 All the examples in this section are taken from the ‘Frog story’ narratives.
In (6), two new referents are introduced, a king and his wife. If we examine their introductions, the former does not assume agentive role in its introduction, and is introduced as the subject of an intransitive verb. The second referent is introduced as an object, again without assuming agentive role. Du Bois’ principle has been studied for various languages, and its universality is very largely confirmed (Du Bois 1987; see Du Bois et al. 2003 for a compilation of work on the principle).

Furthermore, studies have shown that speakers may mark the newness of referents by using particular constructions. For instance, in English, the use of an existential construction as in (6) above is a typical way of introducing referents in a fable. The use of existential constructions for reference introduction has been observed in Indonesian (Cumming 1994), Spanish and French (Bentivoglio 1993; Lambrecht 1988; Ocampo 1993) and Japanese (Ono et al. 2000).

The newness of referents may be marked by their position in the clause in respect of the verb (Hickmann & Liang 1990; Hickmann, Kail, & Roland 1995). Studies show that the post-verbal position is often preferred for the introduction of new referents. For instance, with the use of so-called dummy subjects, the newly introduced referents in (7) and (8) are placed in post-verbal position. Inversions as in (9) may also be used to introduce referents post-verbally.

(7) There is an owl living in a tree.
(8) Er zit een kikker in een pot.

*There is a frog in a jar*

(9) The boy looked into a hole, and out came an owl.

In short, findings suggest that speakers indicate the newness of referents in many ways, although cross-linguistic variation may be observed due to the grammatical characteristics of the language spoken.

With respect to reference tracking, speakers irrespective of the language spoken, make use of explicit referential forms and attenuated forms. However, the availability and the active use of them may vary among languages. Some languages have pronominal systems, while others may not have active use of such system. In addition, the degree of complexity of the pronominal systems may vary from language to language. For instance, the German pronominal system is more complex than its English counterpart as the former incorporates a complex gender/-case marking system (Hendriks 2003). On the other hand, pure personal pronouns do not exist in some languages (e.g. Japanese), although demonstratives may be utilized as a productive resource for tracking referents. In some languages, pronouns may be dropped (e.g., languages such as Spanish or Italian) because of the system of inflectional morphology attached to the verb stem, which can indicate the intended referent.

Although speakers use zero anaphora, cross-linguistic variation exists as to the syntactic and semantic rules governing its occurrence (Fox 1987). For instance, some languages provide looser constraints than others. In topic-comment languages such as Chinese Japanese and Korean, the use of zero anaphora is less regulated than in most
of the European languages. Contextually retrievable information is often expressed by zero anaphora (Furuyama 2001; Huang 2000b; Iwasaki 2002; Jung 2004; Li & Thompson 1976, 1979; Watanabe 1994).

It has been suggested that the choice of referential forms used in tracking referents is influenced by the distance between the present and the last mentions of a referent, and the presence of an intervening subject (Clancy 1980; Gullberg 2003; Hendriks 2003; Hickmann & Hendriks 1999; Nakahama 2003a; Yanagimachi 1997). These findings suggest that if a referent is a continuous topic, the least attenuated forms is usually used. In contrast, when a referent is re-introduced into a narrative, speakers tend to use an explicit form. In addition, some researchers provide a cognitive view of how the choice of referential forms is made (Tomlin & Pu 1991). Thus, speakers of various languages may, in some cases, use the same referential forms to track referents, although the preferred forms may show cross-linguistic variation.

1.4.2 Reference to inanimate entities (space)
In narrative, accounts of events are frequently provided with spatial information. Although some inanimate referents may be introduced into discourse without any relationship to motion verbs as in (10), most of the information provided by inanimate referents is either about a static location of animate referents involving one position, or changes of location of animate referents involving two positions (the source and the target) (Klein & Perdue 1997) as in (11) and (12) respectively.

(10) The boy found a beehive.
(11) A frog family was behind a tree trunk.
(12) The boy was tipped off the cliff into the water.

Talmy (1983) shows that spatial scenes involve one object, the Figure, which is located or moving, in relation to another object, the Ground. In (11) and (12), the frog family and the boy represent the Figure, and a tree trunk and the cliff represent the Ground. In order to locate the Figure in an event description, it is necessary to relate it to an inanimate referent. It has been suggested that how the relation between Figure and Ground is expressed in discourse may show cross-linguistic variation. We will come back to this point later.

With respect to reference to static locations in narratives, researchers have recently found that, depending on the language spoken, speakers may use different approaches in describing the spatial relationship of inanimate entities. Carroll, Murcia-Serra, Watorek, & Bediscioli (2000) conducted a crosslinguistic study of the discourse of the description of spatial relationships (e.g., description of the layout of a house or a room), and found that spatial descriptions may reflect different perspectives. Two perspectives were recognized, object-based and location-based. Carroll et al. found that, depending on the perspective assumed, speakers may introduce and track inanimate entities differently. For instance, when an inanimate referent is introduced into the narrative, the object-based perspective can be realized linguistically in two separate ways. One is to provide a statement asserting the existence of an object (13a),
while the second is to assert its existence via an attribute of another object (13b). The location-based perspective is expressed whereby the introduction of a referent is performed with locative expressions (13c).

(13) a) There is a desk
b) There is a house with a chimney
c) Beside the chair is a desk

Note that while the object-based perspective is realized by existentials whereby the existence of the object is clearly expressed linguistically (by the noun), the location-based perspective is realized by locational (adverbial) means where the existence of the inanimate entity (e.g., the chair in example c) is implied but not explicitly stated. When inanimate referents are tracked, the two perspectives are again differently realized in the linguistic context. For instance, speakers taking the object-based perspective will focus on the entities (e.g. There is a café; next to it is a taxi stand). The pronoun it relates the place and the object closely. While speakers taking the location-based perspective will not focus on the object (e.g. There is a square at the end of the street; they sell balloons there). The adverbial there does not relate the space to a specific object.

Carroll and her associates (Carroll 1997; Carroll et al. 2000) examined the descriptive narratives of native speakers of English and Romance languages (French, Italian and Spanish) and German. The results show a clear contrast. While the speakers of English and Romance languages showed a preference for an object-based perspective, native speakers of German preferred a location-based perspective. Although no findings are available for speakers of Dutch and Japanese, cross-linguistic differences found in spatial descriptions among speakers in Carroll et al. (2000) provide insights for the present work.

Interestingly, research on acquisition of spatial expressions by children suggest that the speed of acquisition and the development of their ability to contrast different spatial relations with linguistic means is largely influenced by the number of terms available in the language and variety of uses the terms are put to use (Johnston & Slobin 1979). Bowerman and her associate (Bowerman 1996: Choi & Bowerman 1991) maintain that the language-specific semantic categorization of spatial domain may be formulated very early. They show that children construct such categorization with the help of linguistic input.

Recent research has illustrated that the way speakers refer to inanimate entities in narratives may be influenced by the language spoken. For instance, Slobin (1996a, 1996b, 2004) has suggested that how speakers talk about motion events is filtered by the language spoken. According to this mildly Whorfian view, there is a language-specific knowledge mechanism – thinking for speaking – activated during speaking. Slobin’s theoretical underpinning comes from Talmy’s (1985, 1991, 2000) typological distinction between languages, according to which speakers of languages such as English, Dutch and Chinese (henceforth: manner-verb languages) conflate events differently from speakers of languages such as Spanish, French and Japanese.
In short, speakers of ‘manner-verb’ languages depict path of movements in path-phrases, where speakers of ‘path-verb’ languages express path of movements by verbs. The different ways of conflating events have been empirically confirmed by various studies (Naigles & Terrazas 1998; Naigles, Eisenberg, Kako, Highter, & McGraw. 1998; Özcaliskan & Slobin 1999, 2000).

With reference to the introduction of *inanimate* entities, Slobin (1996) has shown that speakers of two types of languages may adopt different rhetorical styles in depicting the locations of motion events. Slobin claims that the speakers of manner-verb languages attend more to the trajectory of the motion than to information about locations, while speakers of path-verb languages provide less information on the trajectory of the motion and attend more to static scene setting. The following examples (from Slobin 1996b) show the difference between the two narrative styles of speakers of a manner-verb language (English) and a path-verb language (Spanish).

(14) a) He (=deer) threw him (boy) over a cliff into a pond.

   b) Los tiró a un precipicio donde había harta agua. Entonces se cayeron.

   *(The deer) threw them at a cliff where there was lots of water. Then they fell.*

In the English example, (14a), information about the location is kept to a minimum. Instead, the trajectory of the motion is provided in detail, as in ‘over a cliff into a pond’. On the other hand, in the Spanish example, (14b), the information on the trajectory of the motion is kept to a minimum. Instead, the speaker focuses more on a description of the location. The findings of Carroll et al. (2000) and Slobin (1996a, 1996b, 2003) suggest that although the selection and segmentation of information into propositional units is a universal prerequisite for discourse production, language-specific structural constraints may interact with the language production process at the conceptual level of discourse production.

With reference to inanimate reference tracking, it has been suggested that inanimate entities are rarely tracked anaphorically in story-telling narratives (Hofling 2003). The infrequency of inanimate tracking may be due to the permanent nature of referents. As we mentioned in passing, once inanimate entities are introduced, speakers may take them for granted and omit any following reference.

### 1.5 The characteristics of the source and the target languages

The source and the target languages used for the present study are Dutch and Japanese, respectively. The choice was based on the fact that the two languages are typologically different in two respects. With respect to *animate* reference introduction and tracking, the two languages vary in the linguistic resources available to mark the flow of information. With respect to *inanimate* reference introduction and tracking, Dutch and Japanese differ in the way their speakers attend to spatial information. In the

---

3 Talmy (1985, 2000) distinguishes the two groups as ‘satellite-framed’ and ‘verb-framed’ languages. However, due to the difficulty in defining ‘satellite’, we will avoid using the terms.
following, a brief review of some of the linguistic characteristics of the two languages relevant to the present study is provided.

### 1.5.1 Dutch

Dutch is one of the Indo-European languages with various finite verbs, rich verb inflections, an article system and prepositions. Like other Indo-European languages, Dutch is a subject-prominent language (Li & Thomson 1976). It is generally assumed that the language has the underlying SOV word order with an obligatory verb second rule for the main clause (Koster 1975). Because finite verbs and infinitives can be separated, the possible word order of the Dutch language is complex. The word order patterns in Figure 1.3 are all possible for Dutch.

![Figure 1.3](image)

As with other Germanic languages, the use of articles is obligatory in Dutch and speakers use the two types of articles to distinguish the two types of informational status of referents.

(15) Een jongetje heeft een kikker, en de kikker woont in een potje.

_A boy has a frog, and the frog lives in a jar._

In Dutch, the pronominal system, gender, number and to some extent formality are grammaticalized. In addition, two distinctions of proximity are lexicalised in Dutch demonstratives. Tables 1.1a and 1.1b show the matrix. Thus, Dutch speakers have three types of referential forms, namely NP, pronouns and zero-reference.

<table>
<thead>
<tr>
<th>Table 1.1a. Dutch pronouns in subject role</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>1st person</strong></td>
</tr>
<tr>
<td><strong>2nd person informal</strong></td>
</tr>
<tr>
<td><strong>formal</strong></td>
</tr>
<tr>
<td><strong>3rd person</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1.1b. Dutch demonstrative pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>proximate</strong></td>
</tr>
<tr>
<td><strong>distal</strong></td>
</tr>
</tbody>
</table>
Dutch allows existentials to be used with verbs of posture as in (16), which are more common for introducing new entities than the combination of *er/daar* with a copula as in ‘Er is’ (There is) (Hendriks 1993). In other words, in Dutch, manner verbs are normally used with locative expressions.

(16) En daar zat een kikker
    *And over there there was a frog (sitting)*

Certain combinations of adverbial locatives and prepositions are grammaticized in Dutch. For instance, the adverbial locatives shown in (17) can be combined with postpositions such as *binnen* (‘inside’), *achter* (‘behind’), *in* (‘inside’), etc., as in (18).

(17) er, hier, daar
    **there, here, there**

(18) erachter, hieronder, daarin
    **there after, here under, there inside**

According to Talmy’s categorisation (1985, 1987), Dutch is a ‘manner-verb’ language where the path of motion is depicted by a *path* phrase as in (20).

(19) De kikker sprong uit z’n potje
    *The frog jumped out of his pot*

In (19), the path of the frog’s motion is expressed not by the verb but by a path adjunct, *uit* (‘out’). Previous studies show that Dutch speakers share such rhetorical styles with other speakers of ‘manner-verb’ language (Slobin 1996a, 1996b).

### 1.5.2 Japanese

Japanese shares characteristics with Altaic languages, such as a rich agglutinating verb morphology, modifier head constituent order, a lack of an article system, a lack of grammatical gender and the use of postpositions rather than prepositions (Iwasaki 2002). Japanese subjects lack person and number morphology. The basic word order of Japanese is SOV, and although variations do occur by scrambling, items being right-dislocated to the position after the verb are very rare. The roles of NPs are marked by post-positional particles. In addition, Japanese is classified as a language that is both topic-prominent and subject-prominent (Li & Thompson 1976). As one of the characteristics of topic-prominent languages, a so-called ‘dummy subject’ is absent from Japanese. When existential constructions are used, the introduced referent assumes subject role as in (20).

(20) Aru tokoro ni ojiisan to obaasan ga sundeimashita
    **a certain place DAT old man and old woman NOM live-ASP:PAST**
    *In a certain place, there lived an old man and an old woman*
In (20), the newly introduced referents, an old man and an old woman, marked by a nominal marker *ga*, assume subject role.

Unlike Dutch which offers various pronouns (*hij, zij, ze, die* etc), in narrative discourse, authentic third-person pronouns, the equivalent of ‘he’, ‘she’ and ‘they’ are absent in Japanese (Kuno 1973). Translations of these pronouns do exist in Japanese as *kare* (‘he’), *kanojo* (‘she’), and *karera* (‘they’), but they are originally derived from demonstratives and their use is highly limited (Hinds 1983). Japanese is also a clause-chaining language (Iwasaki 2002). Several medial clauses may be chained before the sentence is concluded by the final clause with a verb expressing aspect and tense. The use of zero anaphora is extremely common and largely pragmatic. Generally, contextually retrievable information is often expressed with zero anaphora (see Chapter 4 for more discussion).

Another characteristic of Japanese referential forms is the frequent use of demonstratives. The demonstratives used in Japanese discourse such as *sore* (‘that’) or *kore* (‘this’) may refer to object, person, location, event etc. When demonstratives are used to refer to persons, an adnominal *sono* (‘that’ as in ‘that person’) is used together with an NP. Reference to objects can be expressed by either a nominal demonstrative *sore* (‘that’) or an adnominal *sono* (‘that’) plus an NP. Figure 1.4 shows that demonstratives can also refer to textual information larger than a single lexical item in Japanese.

There are three series of demonstratives in Japanese: speaker proximate, addressee-proximate and distal. The three types of demonstratives differ according to the first syllable. The speaker-proximate demonstrative has *ko-* , addressee-proximate has *so-* and distal has *a-* at the beginning of the word. The demonstratives are ‘nominal’, ‘adnominal’ and ‘adverbial’.

![Figure 1.4. Japanese demonstratives (adopted from Iwasaki (2002))](image)

Table 1.2 shows some of the demonstratives. Like anaphora in discourse, *so-* types of demonstratives are most common. However, *ko-* and *a-* type demonstratives may also occur under specific circumstances. For instance, the speaker-proximate adverbial demonstrative *koo* (‘in this way’), is frequently used when the narrator demonstrates the manner of actions.
Unlike Dutch, Japanese is a path-verb language where the path of motion is depicted through the verb as in (21).

(21) Fukuro ga detekimashita
Owl NOM exit-come:PAST
An owl came out

In (21), the path of the motion is expressed by the verb *detekuru* (‘exit and come’).

### 1.6 Reference introduction and tracking in L2 narrative

#### 1.6.1 Reference to animate entities

Findings suggest difficulties learners face in marking the introduction and tracking of referents in a TL-like manner. At a very low level of proficiency, learners may be unable to provide the information explicitly in their narratives and rely instead on implicit information (Klein & Perdue 1997). For instance, in tracking referents, it is often the case that the identity of the referent is not linguistically supplied in situations where it is clear from the context (Klein & Perdue 1992).

However, with learners of somewhat higher proficiency, a more complex picture emerges. With respect to reference introduction, studies indicate that the availability of newness markers in the SL influences the marking of new referents in the TL. For instance, Jarvis (2002) compared how Swedish and Finnish learners of English use articles to mark newly introduced referents. Swedish has an article system while Finnish does not. The results show that Swedish learners are more successful in using English articles than their Finnish counterparts. The interaction between SL and TL in newness marking is also suggested by Nakahama (2003a), who compared the acquisition of post-positional *ga* as a newness marker by Korean and English learners of Japanese. Both Korean and Japanese utilize post-positional particles as optional newness markers. Nakahama found that although the acquisition of *ga* interacts with proficiency in both Korean and English learners, Korean learners use *ga* more frequently than English learners, who tend to omit post-positional particles altogether from their utterances.

Furthermore, studies show that when the form used for marking newness of referents in the SL is not available in the TL, learners may look for other linguistic means to perform the same function (Jin 1994). In Jin’s study of English learners of Chinese, a language without articles, the learners overused demonstratives in

---

**Table 1.2. Some demonstrative expressions (based on Iwasaki (2002))**

<table>
<thead>
<tr>
<th>Nominal:</th>
<th>Speaker proximate</th>
<th>Addressee proximate</th>
<th>Distal</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>kore (this one)</td>
<td>sore (that one)</td>
<td>are (that one)</td>
</tr>
<tr>
<td>place</td>
<td>koko (here)</td>
<td>soko (there)</td>
<td>asoko (over there)</td>
</tr>
<tr>
<td>Adnominal</td>
<td>kono +NP (this NP)</td>
<td>sono +NP (that NP)</td>
<td>ano + NP (that NP)</td>
</tr>
<tr>
<td>Adverbial</td>
<td>koo (in this way)</td>
<td>soo (in that way)</td>
<td>aa (in that way)</td>
</tr>
</tbody>
</table>

- Unlike Dutch, Japanese is a path-verb language where the path of motion is depicted through the verb as in (21).
- (21) Fukuro ga detekimashita
  Owl NOM exit-come:PAST
  An owl came out

In (21), the path of the motion is expressed by the verb *detekuru* (‘exit and come’).
combinations with NPs to denote reference maintenance. Jin interprets this as an influence of the SL in that the English learners may have relied on demonstratives in Chinese to perform the function of definite articles. The findings reviewed above indicate that the way newness of referents is marked in the SL may influence newness marking in the TL.

In contrast, research has indicated that L2 speakers do not differ from L1 speakers with respect to their preference in assigning syntactic/semantic roles to new referents. Kumpf (1992) shows that the learners avoid assigning agentive role (A) to referents upon their introduction. Similarly, Sasaki (1997) found out that learners introduced a new topic in object position as a full NP.

As for tracking animate referents, studies find that learners are generally over-explicit in their use of referential forms, irrespective of the source-target language pairing. Learners often repeat NPs even when the explicit mention of referents is not necessary, as the example from Kumpf (1992) shows in (22).

(22) last time I got a cigarette lighter, my cigarette lighter broke after that I tried to find a cigarette lighter from a supermarket they are selling cigarette lighter like five bucks or something like that so I bought cigarette lighter there, and I thought that it was gonna be fit but when I plug into cigarette lighter fuse keep blowing so I decided to buy a part from dealer and they charged me twenty-seven dollars? for cigarette lighter! I couldn’t believe it.

(from Kumpf 1992)

In (23), the speaker refers to the same referent, cigarette lighter, explicitly with a full NP over a stretch of discourse without switching to attenuated forms. The phenomenon of over-explicitness in L2 has been noted in studies with Japanese learners of English (Chaudron & Parker 1990), learners of English with various L1s (Mandarin, Cantonese, Malay, Korean, Russian, Mende, Kikongo and Vietnamese) (Williams 1989), Spanish learners of English (Muñoz 2000), Japanese and Spanish learners of English (Kumpf 1992), English learners of French (Fakhri 1989), French learners of Swedish and Swedish learners of French (Gullberg 1998), English and Japanese learners of Chinese (Polio 1995), English learners of Japanese (Yanagimachi 1997, 2000) and English learners of Korean (Jung 2004). Some studies show that over-explicitness is actually a transfer from L1 where it is used as a rhetorical device to show emotional involvement (Bartelt, 1983) or a strategic device used only for nouns such as ‘man’ or ‘child’ (Nistov 2001).

Among them, the findings most relevant to the present work are those by Yanagimachi (1997, 2000) and Polio (1995). Yanagimachi (1997, 2000) examined the
interaction between proficiency and the choice of referential forms by American learners of Japanese. His findings show that learners’ use of zero anaphora in reference to animate entities in their story-retellings is much less frequent in comparison to their native speaker counterparts. However, it is difficult to judge whether the infrequent use of zero anaphora in L2 is a general characteristics of learners or an effect of SL. On the other hand, Polio (1995) examined the use of zero anaphora by Japanese and English learners of Chinese. She found that both groups of learners were reluctant to produce zero anaphora in the conditions where it is not syntactically or semantically constrained, despite the fact that zero anaphora was not as strongly constrained in Japanese as in Chinese. Based on her results, Polio concludes that ‘there is something that is preventing the NNSs from using zero pronouns as often as the NSs’ (Polio 1995: 370), although she does not provide further explanation.

Recent studies show that learners’ linguistic awareness of the source and target languages may in fact interact with the choice of referential forms. Some of the evidence comes from Chinese learners of English, French and German in Hendriks (2003). Despite the fact that all the learners share the same source language, differences were observed among the groups as to the choice of referential forms in the target languages. Hendriks found that when the target language referential process (left dislocation in French) licenses the source-like referential process (topic-raising in Chinese), learners incorporate the form in their L2 discourse and use it with the frequency preferred by native speakers of French (cf. Hendriks 2000). With regard to over-explicitness, comparative analyses with native narratives show that Chinese learners of German were the only learner group that showed this tendency. Hendriks maintains that the over-explicitness of Chinese learners is most likely due to the avoidance of the German pronominal system because of its complex nature in comparison to its English and French counterparts.

Similarly, Yanagimachi (1997) explains that the over-explicitness of referential forms may be the direct result of avoidance by learners of complex target-language-specific linguistic constructions such as passives and the use of ‘giving’ and ‘receiving’ verbs which are necessary to create coherent discourse. It has been suggested that languages may differ in the way that events are interpreted and introduced by the choice of voice and focus (Nakahama 2003b; Kim 2001). For instance, of the two Asian languages Korean and Japanese, speakers of the latter fix the focus of narratives by assigning the subject roles to a limited number of main referents. In order to fix these viewpoints, various complex structures such as passives and give-receive constructions may be utilized in L1 discourse. However, learners seem to be keenly aware of the ambiguity that erroneous use of such structures may cause, consequently may avoid their use altogether. The result is the phenomenon of ‘view-hopping’ (Yanagimachi 1997), where NPs are used in every clause to refer to different referents in subject role. In (23), the viewpoint is not fixed, and the subjects of utterances switch between the dog and the two human referents. As a result, the American learner of Japanese could not produce the optimal conditions for using zero anaphora.
CHAPTER 1

(23) …uni no, inu wa uhhm, akachan o miru to, uhhm, ø (=inu) kawashi [kanashi]soo, ni narimassu. Uhh, obaasan to o, ojisan wa, m, tsk, akachan to, uhh, tsk asonde, anoo inu wa, sooji suru, anoo, uhhm, m, sooji nado shiteimasu. Uhm, ojiisan to, obaasan wa, m, inu, o mimasen, anoo, inu wa uchi o, demasu

‘…when the dog in the house, uhhm, sees the baby, hum, ø (=dog) feels lonely. Uhh, the grandma and granpa, m, tsk, play, uhh, tsk, with the baby, and uhm, the dog, vacuums, uhhm m, is vacuuming and so on. Uhm, the grandpa, and grandma, m, do not see the dog, uhm, the dog leaves, the house.’

With respect to the interaction between referential importance and the choice of referential forms, findings suggest that, unlike L1 learners, L2 adult learners, even at an early phase, may be able to use a global structuring in narrative (Broeder 1991; Strömqvist & Day 1990)

In summary, the findings above indicate that learners may be aware, at least to some extent, of the pragmatic principle governing the introduction and tracking of animate referents. When SL-like referential forms are available in the TL, learners transfer their SL-based knowledge to the TL, in accordance with Andersen’s (1983) ‘transfer to somewhere’ principle. However, at the same time, learners show a keen awareness of the problems that erroneous use of forms or structures may cause. Thus, they may avoid using error-prone forms or constructions in their target languages (cf. Schachter’s notion of ‘avoidance’, Schachter, 1974) even if the use of other forms may be non-TL like. As one explanation for the over-explicit nature of referencing in L2, Williams (1988) states that learners face difficulties in striking the right balance between redundancy and ambiguity. However, recent studies such as Hendriks (2003) mentioned above show that these opposing forces seem to interact with learners’ linguistic knowledge of the SL and the TL.

1.6.2 Reference to inanimate entities (space)

It should be stated first that studies of the introduction and tracking of inanimate referents in L2 narrative production are extremely rare. The majority of the research involves either learners at the very beginning or at the advanced stages of proficiency. The vast amount of work compiled during the “European Science Foundation Project” (Perdue 1993) shows that learners with very limited proficiency use a fixed order to show the location of an entity. In short, the entity which is located (theme) is uttered before mention of the entity in relation to which it is located (relatum). In (24), the learner is directing the listener to move an object to an appropriate location.

(24) bag in the table
   ‘put the bag on the table’ (Carroll 1990, in Klein & Purdue 1997: 329)
In (24), the learner first draws attention to the entity, ‘bag’, and subsequently mentions the place where it needs to be located.

On the other hand, studies of L2 event construal and spatial descriptions indicate that recognizing the full range of principles that govern the construction of discourse in the spatial and temporal domains is difficult even for highly advanced learners (Carroll et al. 2000; Carroll & Lambert 2003; Carroll & von Stutterheim 2003; Harley 1989; von Stutterheim & Nüse 2003; von Stutterheim, Nüse, & Murcia-Serra 2002). For instance, Harley (1989) studied the acquisition of spatial expressions by English-speaking immersion students. Her study was based on Clark’s (1985) observation that, unlike in English where the notion of motion is expressed in prepositions, in Romance languages, verbs of motion typically combine notions of motion with those of directions (Clark 1985: 746). Harley examined how English speakers learn to use verbs to express directional information in French using the data consisted of students’ compositions. Her results show that English-speaking students rely more on prepositions than French native speakers to express direction of motions.

Carroll et al. (2000) studied the aforementioned two perspectives that distinguish the descriptions of inanimate entities in discourse by advanced English and Spanish learners of German. Close examination of the descriptive discourse revealed that the learners adhered to their SL-based perspective while introducing and tracking inanimate entities. The results are in line with the finding that TL-like perspectives in event narrating are acquired very late, if ever (Carroll et al. 2000; Carroll & von Stutterheim 2003). Thus, although limited in number, previous studies on inanimate reference introduction and tracking in L2 show that learners tend to organize information about inanimate referents in the way that conforms to the norm of their SL.

While speakers in both L1 and L2 use various linguistic devices in creating anaphoric linkage in narrative discourse, it has been suggested that linkage between referents during discourse may also be indicated by the use of the hand and space in front of the speaker. Therefore, in the next section, we will review the literature on gesture to see how referent introduction and tracking is performed manually.