Chapter Three

Methodology

In this chapter, the task design and analytical procedure of this study will be presented and delineated. In section 3.1, we will introduce the learners participating in this investigation. In section 3.2, we will introduce the material which is used to elicit the written data. In section 3.3 we will be how the data collection is undertaken. Section 3.4 will be our analyzing procedure in treating all referential forms, including a framework for the present study, three discourse functions, and a standard measurement of distance value.

3.1 Subjects

In the task, we asked eight learners to submit written narratives serving as the database for the present study. All the learners are young adults; they are advanced EFL learners in Taiwan. Their mother tongue was Mandarin, and they learned English as a foreign language. Throughout the study, the learners will be referred to by Arabic number, from 1 to 8. More background information is included in Appendix A.

The reason for choosing advanced EFL learners as our subjects is due to the fact that productive skills in writing such as reference management plays an important role for academic purposes. But so far little has been done to investigate advanced EFL learners' referring expressions in written narratives. Therefore, the present study will

focus on the referential forms in learners' written form.

To meet the requirement of the study presented here, the advanced EFL learners

achieve one of the following English proficiency tests:

1. TOEFL (Test of English as a Foreign Language) PBT (paper-based test) over 600/

    TOEFL CBT (computer-based test) over 250/TOEFL iBT (Internet-based test)

    over 100,

2. IELTS (International English Language Testing System) 7.0, or

3. GEPT (General English Proficiency Test) High-intermediate Level.

Among the eight learners, three have taken TOEFL CBT and scored over 250, two

have taken IELTS and achieved 7.0, and three have passed GEPT High-intermediate

Level[13]. (cf. Appendix A)

All these English proficiency tests mentioned above require test takers to produce

a coherent English composition in order to demonstrate their linguistic competence in

writing. Learners crossing one of these thresholds above could be linguistically

competent enough to produce a comprehensible narrative describing the plot of a

movie. From their narratives, referential forms are expected to shed light for an

overall scope of their English writing capability.

---

[13] According to CEF (Common European Framework), the English language proficiency can be divided into seven levels, from basic to advanced: A0, A1, A2, B1, B2, C1, and C2. The English proficiency tests all over the world can correspond to different levels on the framework. With a reference to this framework, the proficiency level of all the advanced EFL learners in the present study falls on the level of C1.

3.2 Material

We first explain the reasons for choosing narrative as our source of data. The

general preference of narrative data in previous research dates back to Labov and his

colleagues' (1967) study, which suggests that in narrative the two functions of

reference and evaluation need to be performed. Without a good command of reference

management, a narrative would be difficult to understand. In addition, as Li (2000)

has argued, in narrative discourse, the events must be recapitulated in "an iconic

fashion (p. 345)," which entails that the entities or participants with different degrees

of importance or saliency would be manifested on the storyline. Lastly, in studying

the acquisition of narratives, Bamberg (1987) also reports that the development of

narrative in language learners' abilities is reflective of their communicative

competence. In terms of reference management, in narrative tasks language users are

obliged to produce unambiguous referential forms by means of various linguistic

devices so as to convey clear messages to receivers. With all its advantages, therefore,

the present study will examine narrative produced by our EFL learners.

The material adopted by the researcher to elicit advanced EFL learners' written

narrative is the American movie *Ghost* (1990), starred by Patrick Swayze (Sam),

Demi Moore (Molly), Whoopi Goldberg (Oda Mae), Tony Goldwyn (Carl), et al. This

movie was chosen because of its efficacy in helping elicit subjects' narratives. Being

selected by earlier discourse researchers in their investigation such as referring

expression (Huang, 1992), anaphora (Lin, 1992), and topic chain (Chui, 2001), this

movie has been proved to be an ideal material, because it obliges learners to produce

clear and appropriate referential forms in their narratives for receivers.

Moreover, the movie's plot is informative and easily perceptible. Five primary

participants (Sam, Molly, Oda Mae, Carl, Willie) are involved. All these major

participants need to be unambiguously presented in learners' written narrative through

appropriate use of referential forms, in order to contribute to receivers'

comprehension of the movie itself. Furthermore, the movie is long enough to allow of

learners' compulsory manipulation of different structures such as sentences or

paragraphs. In sum, this movie will undoubtedly elicit sufficient referential forms for

the purpose of the study presented here.

3.3 Data collection

The narrative production task in the experiment was a recall one. For each

learner, in the first place, he or she was required to read through the instruction

provided by the researcher in order to have a general idea of how the task proceeded.

Then the learner watched the DVD of *Ghost*. After seeing the movie, the learner came

up with an English summary which was expected to be composed of five hundred

words or so, without time limit. (cf. Appendix B for the task)

A written narrative of short length (i.e. 500 words) would not be too long for our subjects to produce. By restricting the number of words, learners need not spend too much time working on the summary. In addition, although the length of the composition is limited, the summary will contain most of the major events and participants in the story. In a written summary of a movie, the learners will have to describe the plots and the correlation among participants, which would require the writer to produce a considerable number of referential forms.

The data collection was completed in one month. Each narrative, as was stipulated in the instructions, consists of 500 words or so. In each text, all the referential forms that refer to [+human] participants in the movie were counted and marked. We hereby obtained 835 tokens of referential forms for analysis. The total numbers of referring expressions in each text are exemplified in Table 3.1 below.

Table 3.1   Number of referring expressions in learners' text

	Number of words	Number of referring expressions
Learner 1	518	122
Learner 2	496	100
Learner 3	526	99
Leaner 4	522	108
Learner 5	495	93
Learner 6	507	90
Learner 7	458	102
Learner 8	501	121
Total	4032	835

In each written narrative, as Table 3.1 shows, a great number of referring expressions were produced by the advanced EFL learners to refer to different participants of the film in their written narratives. In these summaries, with a variety of referring expressions, 13 NP categories in total could be identified, including (1) Sam, (2) Molly, (3) Oda Mae, (4) Carl, (5) Willie, (6) a ghost in the subway, (7) a ghost in the hospital, (8) a ghost that possesses Oda Mae, (9) other spirits, (10) two nuns on the street, (11) Rita Miller, (12) the police, and (13) the goons. Among them, eight trivial participants were mentioned only once in the narratives. They are: (6) a ghost in the subway, (7) a ghost in the hospital, (8) a ghost that possesses Oda Mae, (9) other spirits, (10) two nuns on the street, (11) Rita Miller, (12) the police, and (13) the goons.

On the other hand, the five major participants: (1) Sam, (2) Molly, (3) Oda Mae, (4) Carl, and (5) Willie, are mentioned several times by means of various referential forms. Among them, except for the first-time introduction, all are anaphoric in nature. Namely, a majority of the referential forms can find their antecedents either within or beyond sentences. These referential forms will serve as the database for the present study.

3.4 Data analysis

In this section, there are three goals. To begin with, in view of the insufficiency

of previous frameworks, the author will devise a framework for the present study in

3.4.1. The referential forms on the model will encompass all the tokens produced by

our subjects. With a revised model, all the referential forms in our data can be

identified and categorized. Secondly, in 3.4.2, we will introduce the three discourse

functions that referential forms perform in the written narratives. With the

establishment of the three discourse functions, we can undertake a function analysis in

the next chapter. The last goal of this section will be a standard measurement which

aims at the distance value that the referential forms display. By working out the

distance values, the correlation between distance and the use of referential forms can

be explored.

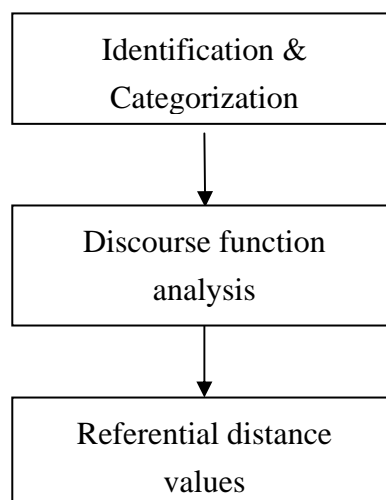The analyzing procedure of the present study can be illustrated in Figure 3.1

below.

```
┌─────────────────────┐
│   Identification &   │
│    Categorization    │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  Discourse function  │
│       analysis       │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│ Referential distance │
│       values         │
└─────────────────────┘
```

Figure 3.1 Analyzing procedure

In each of the sections below, the steps presented above in Figure 3.1 will be delineated with more details.

3.4.1 Analytical framework

To devise our own scale, we shall begin by reexamining Ariel's (1990) and Brown's (1983) models first. The overlapping referential constructions that can refer to [+human] NPs in both Ariel's and Brown's frameworks are: zero anaphora, pronoun, definite NP, demonstrative, name (full NP/ proper noun), and NP after possessive (name +modifier). All these referring expressions except for demonstrative will be included in the framework of this study.

The reason for excluding demonstrative from the present study is because in our data, no subjects produced demonstrative to refer to the participants. In addition to demonstrative, there are also topic-marking structures which are excluded. They are topic-marking constructions in Brown's but not in Ariel's framework: left dislocation, right dislocation, passivization, Y-movement, existential/ presentative, generics, and cleft/ focus construction. All these constructions were never found in our data, and will not be considered in our framework, either.
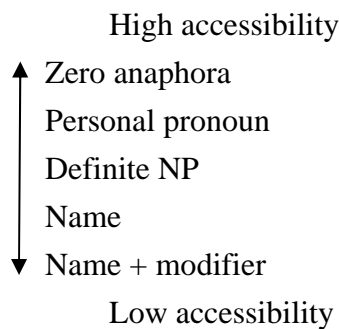
After we reconsider all the models in previous studies, the referential realizations in the written data presented here can be classified into five categories: (1) Zero, (2) Pronoun, (3) Definite NP, (4) Name, and (5) Name + modifier. The arrangement order

of the present framework primarily follows Ariel's (1990) criterion of *Informativity*

and Givon's (1983) *Iconicity* principle. The Informativity criterion indicates that each

referential form contains various amount of information load. The more informative

forms tend to be more easily identified and therefore more rigid than the less

informative ones. For instance, a name (e.g. Sam) would be more informative and

more rigid than a pronominal form (e.g. he). The Iconicity principle, according to

Givon (1983), suggests that the amount of coding material for a certain referring

expression should suffice for receiver's successful interpretation of the intended

referent. As the topic or referent becomes more continuous and thus more predictable,

a referential form with less coding material serves as a better retriever. But if the topic

or referent becomes less continuous and thus less predictable, a more informative

form will be required. In other words, the choice of anaphoric expressions is based on

receiver's interpretation need, no more or less. For example, if a pronominal form (e.g.

she) is sufficient to recover the referent in receiver's consciousness, then there is no

need to use a definite NP (e.g. his daughter).

Based on Ariel's criterion and Givon's principle, the analytical framework of

referential forms adopted in this study can be arranged as the sequence illustrated

below:

Most accessible/predictable/ continuous topic
(highly implicit, least informative, and least rigid marker)

       High accessibility
Zero anaphora
Personal pronoun
Definite NP
Name
Name + modifier
       Low accessibility

Least accessible/predictable/continuous topic
(highly explicit, most informative, and most rigid marker)

The scale we presented above ranks the five referential forms based on the

criteria of topic continuity, accessibility, predictability, and semantic content. On the

upper end are the referential forms that represent more continuous, accessible, and

predictable topics. These forms carry little semantic content and thematic information,

and they mark highly accessible topics. On the lower end are referential forms that

represent topic with lower topic continuity, accessibility, predictability. These forms

carry richer semantic content and thematic information, and they are markers for

low-accessibility topics. In what follows, each referential form is elaborated and

illustrated with some instances:

1) ***Zero anaphora (Zero):*** When a series of events are conducted by the same

participant, the participant becomes a continuous topic in subject position, and Zero

marker would occur. Specifically, a Zero occurs when a known NP is abridged in

subject position. Zero anaphora has two variants: deleted NPs following conjunction

and elliptic nominal forms in detached participial construction. Below are two

examples:

(1) He follows his body to the hospital and ___ϕ___ meets another ghost.

(2) Along with the robber Lopez, Carl rushed to Oda Mae's house, ___ϕ___ trying
     to get the money back.

In example (1) mentioned above, the second unexpressed subject that follows the

conjunction *and* is already activated in the previous clause *He follows his body* and is

spared, and would be labeled as a Zero. In example (2) above, the subject topic is

activated in the main clause *Carl rushed to Oda Mae's house*, and its second

occurrence is realized as a Zero marker which heads a participial clause *trying to get*

the money back.

   2) ***Personal pronoun (Pronoun):*** This form includes all the personal pronouns,

both singular and plural ones. In the data of the present study, our subjects did not

produce any first or second personal pronouns[14], we focus on third personal pronouns.

For example,

(3) Sam could not believe what <u>he</u> had just realized. <u>He</u> knew <u>he</u> must take
     actions as soon as possible to protect Molly from harm.

(4) Together, <u>they</u> make Molly believe there was ghost and that Sam's ghost did
     exist, and <u>they</u> also successfully withdraw the illicit money and donate it to
     the charity.

---

[14] It is reasonable that first- and second-person pronouns do not appear in our data, because the data
are narratives of a film, in which advanced EFL are not required to provide any evaluation on it.

As is illustrated in the above examples, Pronoun in the present study includes both singular and plural third personal pronouns.

3) **Definite NP[15]**: Definite NP here encompasses a variety of referring expressions. It refers to NPs which are preceded by a definite article *the*, with or without an adjective[16]. It also includes NPs within genitive construction, i.e. preceded by possessive pronoun. Furthermore, definite NPs qualified by relative clauses are also counted as Definite NP, since a relative clause usually serves as a further supplementary remark in modifying its head noun. Here are some examples:

(5) As <u>the couple</u> is about to get married; however, one night after Sam and Molly leaves a theater, Sam is shot to death by a robber.

(6) So in the end, as <u>his beloved woman</u> is safe and sound, Sam is left with no regret and vanishes from the world.

(7) One day <u>the man who killed him</u> breaks into the apartment.

In example (5) mentioned above, the Definite NP comprises the noun *couple* and the preceding definite article *the*. In example (6), the Definite NP refers to the noun *woman* in genitive construction, the preceding possessive *his*, and an adjective *beloved*. In example (7) above, the Definite NP consists of the head noun *the man* and the following relative clause *who killed him*.

---

[15] The present study will focus on Definite NPs. Learners also produced indefinite NPs such as *a ghost*, but the frequency of occurrence is relatively low. Only 13 tokens were identified in all texts. They will be confined to the discussion of the discourse function of Introduction.

[16] An adjective which precedes the noun in the category of Definite NP would not be called 'modifier,' lest it should be confused with the fifth referential category: Name + modifier.

4) *Name:* Name in the present study is unlikely to cause ambiguity in learners'

narratives, whether it is a first name (e.g. Sam) or a full name (e.g. Sam Wheat), since

each participant in the movie is unique and has their own identity. Therefore, both

first name and full name are equally predictable, and both can be counted as Names.

Examples of Name are:

(8) At first, <u>Molly</u> doesn't believe it but after <u>Oda Mae</u> tells her things that only
       she and <u>Sam</u> know, she finally trusts her.

(9) <u>Oda Mae Brown</u> was persuaded and pressured into helping him.

As the above examples have suggested, full name and first name are equally

rigid in anchoring the participants in the film, and can both be regarded as Name in

this study.

5) *Name + modifier:* This is the combination of name plus some additional

information. In this category, there are several varieties, including names which are

modified by relative clause (restrictive and nonrestrictive) and those that occur as

apposition. In this category, 'modifier' refers to the additional information (a relative

clause or a definite NP) which follows or precedes its adjacent proper name.

Similar to Definite NP, this category also contains relative clause. The major

difference between the two categories consists in the use of proper name. Under this

category of Name + modifier, there would be a proper name.

According to Ariel and Halmari (1996), the expression of Name + modifier is

used to specify the entities with lowest degree of accessibility .Here are some

examples containing Name + modifier:

> (10) One day, as Molly and <u>Carl, who is Sam's colleague as well as friend</u>, go
> out for a walk, <u>Willie, the one who kills Sam</u>, slips into the house, trying to
> find out something.

> (11) <u>Carl, his best friend</u>, showed up and ordered the murderer to find the
> account password as soon as possible.

> (12) At Oda Mae's work place, <u>the robber Lopez</u> attempted to kill her.

Example (10) mentioned above contains two nonrestrictive relative clauses with

which the head nouns *Carl* and *Willie* are further modified. In example (11) and (12)

presented above, we find appositive constructions that contain proper nouns *Lopez*

and *Carl* and some complementary remarks *the robber* and *his best friend* to modify

the participants. The use of Name + modifier is not common in English text and is

confined to formal written form. But in our data, learners did produce such form. The

function of Name + modifier will be further probed in Chapter 4.

Under the referential scale presented above, the less accessible and predictable

the referred entity is, the more rigid and informative its referential form will be. For

example, a Definition NP (e.g. the robber) is more informative and semantically richer

than a Pronoun (e.g. he). That is to say, compared with pronominal form, a Definite

NP would be a better retriever for receiver's successful activation of the referred

entity. Each referential form produced by our learners will be assigned one of the five

categories on the scale that we have just designed.

In Section 3.3, we have collected 835 referential forms across the written data provided by our EFL learners. All these tokens represent [+human] entities, and they refer to certain participants in the movie. Each token will be assigned into the five categories on the scale that we devised here: Zero, Pronoun, Definite NP, Name, and Name + modifier. Then the frequency and percentage of each category will be figured out, and then we will investigate the distributional pattern of each form, including how and why a certain category displays such a pattern.

All the written data were reviewed by two well-educated English native speakers (college undergraduates). They provided their intuitive judgments over the use of all the referential forms. If there was any inappropriateness for a certain referential form, the native speakers would suggest the form they thought fit in the context. For example, a pronominal form occurs at a certain clause may cause ambiguity and consequently impede readers' understanding of the text. When the hindrance occurred, the native speakers could provide a more acceptable form such as a full NP to replace the pronoun.

Inevitably, there would be inconsistency and disagreement between the two native speakers' judgments. In order to solve the discrepancy, after all the written data were reviewed, the researcher served as a mediator and led a discussion with the two

native speakers. For each inconsistent modification, the two native speakers would

exchange their opinions about the choices until an agreement is reached by both of

them.

The total number of identified referential forms that were considered

inappropriate is 35. After the native reviewers identified these problematic referential

forms, we will manage to explain the possible causes for their inappropriateness. In

each section of the next Chapter, the appropriateness issue will be further discussed.

In the next section, we will move to the discourse function of each referential

form.

3.4.2 Discourse functions

The reason for defining different discourse functions is because that the input

and output of referring expressions are determined by different discourse contexts

(Chaudron & Parker, 1990; Clancy, 1992; Givon, 1983; Yule, 1981). As has been

pointed out in previous studies, when introducing or reintroducing a referent, a full

NP tends to be used. When a referent is already activated and needs to be maintained,

a pronominal form is preferred. In other words, for the receiver's successful encoding

and decoding of referents, and for the receiver's mental processing to go smoothly, the

correlation between discourse function and referential forms must be considered by

the narrators.

Ever since Clancy's (1980, 1992) pioneering investigation on referential strategies, the three function categories of Introductions, Same Subjects, and Switch Subjects have been widely accepted by subsequent researchers. In this study, for the purpose of analytical consideration, the functions are: 1) *NP Introduction*, 2) *NP Maintenance* (i.e. Same Subjects), and 3) *NP Reintroduction* (i.e. Switch Subjects). When discussing the Introduction function, we inspect all NPs regardless of their positions in sentences. But when discussing the discourse functions of Maintenance and Reintroduction, we only consider the subject NPs, since only a subject topic can be maintained and reintroduced. Each type of discourse function is characterized in the following parts.

Firstly, Introduction refers to the context in which a completely new NP is introduced in the text for the first time. Since a brand-new referent has not entered the receiver's focus of attention, it requires more efforts of delineation from the writer. For example,

(13) Sam's friend, Carl, came to Molly to comfort her and even wanted to kiss her.

In example (13) above, the Name + modifier *Sam's friend, Carl* is produced to introduce the new participant to the story.

Secondly, Maintenance function refers to reservation of the same NP in subject position at its second or ensuing occurrence, i.e. in the adjacent following clause(s).

Below is an example of subject NP Maintenance.

(14) Just when he was annoyed about not being able to contact Molly, <u>he</u> met a spiritual fixer, Oda Mae. <u>He</u> asked Oda Mae to help him deliver messages to Molly.

In the above example, both of the two Pronouns *he* refer to the major participant *Sam*. After *Sam* is introduced and presumed to have already been established in receiver's state of consciousness, the subject of the same entity *Sam* is said to maintain the topic continuity.

Lastly, if another subject referent intervenes, the earlier subject referent has to be reintroduced into the text. Reintroduction function refers to a previously mentioned subject NP that is reintroduced into the immediate context after the interference of other different NP category in subject position. See the example below.

(15) *Sam* is shot to death by a robber. Molly is too sad to carry on her life for some time. On the other hand, <u>Sam</u> does not leave the world as he dies.

In excerpt (15) presented above, the full NP *Sam* is produced to reintroduce the activated entity in the movie. The intervening NP between two mentions of the same entity *Sam* is *Molly*.

The major difference between the discourse functions of Maintenance and Reintroduction is the existence of intervening subject topics. While entities in Maintenance function comprise successive subject NPs referring to the same participant in the film, entities in Reintroduction function denotes subject NPs that

cannot retrieve their antecedents in the immediately preceding clause, but in two or

more clauses away.

The number of subjects that perform discourse function is 470. See Table 3.2

below for their distribution in each text.

Table 3.2   Number of referring expressions at subject position in learners' text

	All referring expressions	Referring expressions at subject position
Learner 1	122	72
Learner 2	100	49
Learner 3	99	62
Learner 4	108	59
Learner 5	93	51
Learner 6	90	53
Learner 7	102	52
Learner 8	121	72
Total	835	470

As Table 3.2 above shows, among all the referential forms, more than half of the

referential forms are identified in subject position. Out of the total number of 835

referential forms, 470 can be identified in subject position. Each referential form that

occurs in subject position fulfills one of the discourse functions: Introduction,

Maintenance, or Reintroduction. The correlation between referential forms and the

discourse function they fulfill will then be further discussed.

3.4.3 Referential distance value

To learn more about the role that distance plays in referential forms, we find it

necessary to propose a linear measurement in determining the distance values. In the

present study, we adopt Givon's (1983) standard measurement. In Givon and his

associates' research, the measurement which is consistently followed is *Referential*

distance (i.e. look-back).In their studies, they divided the texts into the basic units of

clauses. The distance value is figured out by counting the number of clauses to the left

from a given referent to its immediately preceding occurrence. Both of the two

mentions representing the same participant occur at subject position, because the

focus will be subject NPs.

A minimum integral of distance value will be 1 (clause), which denotes that the

subject topic in discussion is highly continuous and refers back to its antecedent in the

previous clause. Let's inspect excerpt (14) in 3.4.2 again as an illustration.

(14) Just when he was annoyed about not being able to contact Molly, <u>he</u> met a
spiritual fixer, Oda Mae. <u>He</u> asked Oda Mae to help him deliver messages to
Molly.

The underlined referential form *He* in *He asked Oda Mae to help him*… displays

a referential distance value of 1, for its antecedent *he* can be identified in the

immediately preceding clause *he met a spiritual fixer, Oda Mae*.

On the other hand, a maximum distance value will be 20, which means that the

subject topic (e.g. a referential form which represents a new participant) under

discussion is least continuous and can hardly trace back to its antecedent, if any. Let's

inspect excerpt (11) in 3.4.1 again as an illustration.

(11) <u>Carl, his best friend</u>, showed up and ordered the murderer to find the
account password as soon as possible.

In the above excerpt, the referential form of Name + modifier *Carl, his best*
friend is meant for the introduction of a new participant into the story, and cannot find
any antecedent. In dealing with such form, we assign it a maximum distance value of
20. The correlation between the distributional proportion of each referential forms and
the average referential distance value will be quantified and figured in Chapter 4.

Other than the linear perspective of distance value, we also consider the
hierarchical approach of episode boundary and use it as a complementary tool to
explain the distribution of the referential forms. According to hierarchical approach, it
is the episodic organization that influences referential behavior (Fox, 1989a, 1987b;
Hinds, 1977, 1979; van Dijk, 1982). They propose that the demarcation of narrative
units accounts for the alternation of full NPs and pronominal forms in the written
narratives. In the present study, the demarcation can be identified by means of
different episodes of the film. Each episode involves some individual scenic features
in terms of a) time, b) place, c) event, and d) character, all of which can help signal
the demarcation of episodic units. The hierarchical approach, in addition to linear
analysis, may shed some light on the reference management.

3.4.4 Summary

In this section, we have devised a scale of referential forms which is intended to

describe the referring expressions produced by the advanced EFL learners. We have

also elaborated on the three major discourse functions that each referential form may

perform. The distributional pattern and frequency of each referring expression's

discourse function will be further examined in the next Chapter. On top of that, we

have also introduced a linear standard measurement in order to figure out the distance

value that each referential form displays. Lastly, we presented the hierarchical

perspective in dealing with the alternation of full NPs and pronominal forms.