Identifiability and definiteness in Chinese*

PING CHEN

Abstract

This article explores how the pragmatic notion of identifiability is encoded in Chinese. It presents a detailed analysis of the distinctive linguistic devices, including lexical, morphological, and position in sentence, which are employed in Chinese to indicate the interpretation of referents in respect of identifiability. Of the major determiners in Chinese, demonstratives are developing uses of a definite article, and yi ‘one’ + classifier has developed uses of an indefinite article, although morphologically and in some cases also functionally they have not yet been fully grammaticalized. What makes Chinese further different from languages like English is the interpretation in this regard of what are called indeterminate lexical encodings, which include bare NPs and cardinality expressions. They by themselves are neutral in respect of the interpretation of identifiability. For indeterminate expressions, there is a strong but seldom absolute correlation between the interpretation of identifiability or nonidentifiability and their occurrence in different positions in a sentence. Unlike the cases with several other languages without articles like Czech, Hindi, and Indonesian, the features of definiteness and indefiniteness cannot be obligatorily and uniquely specified for nominal expressions in Chinese. The findings in this article lead to the conclusion that definiteness as a grammatical category defined in the narrow sense has not been fully developed in Chinese.

1. Introduction

The term “identifiability” in this article denotes a pragmatic concept, and the term “definiteness” denotes a grammatical category featuring formal distinction whose core function is to mark a nominal expression as identifiable or nonidentifiable. The formal distinctions may be expressed by a variety of grammatical means in languages, including phonological,
lexical, morphological, and word order. Most typically, and also most extensively in languages, the grammatical category is encoded in terms of a contrast between a definite article like *the*, and an indefinite article like *a* in English. A definite expression with *the* differs essentially from an indefinite expression with *a* in that the former is marked as being identifiable and the latter as nonidentifiable. Whether or not a language is considered to have a grammatical category of definiteness is decided, to a large extent, based on whether there are specialized grammatical means primarily for this particular function on a par with definite and indefinite articles in languages like English. As observed by Chesterman (1991: 4), “it is via the articles that definiteness is quintessentially realized, and it is in analyses of the articles that the descriptive problems are most clearly manifested. Moreover, it is largely on the basis of the evidence of articles in article-languages that definiteness has been proposed at all as a category in other languages.”

As is the case with other grammatical categories in language like tense, number, gender, proximity, animacy, etc., the form and function do not always match. Definite expressions typically, but need not always, mark identifiability, just as a verb in past tense may be found in uses which have nothing to do with past time.

There is no fully grammaticalized definite article in Chinese. I aim to address the following issues in this article:

i. How is the pragmatic notion of identifiability encoded in Chinese?

ii. How is Chinese in this respect similar to, or different from, languages with articles like English, and languages without articles like Czech, Hindi, and Indonesian?

iii. Is it justified — and if so, to what extent and in what sense — to assert that definiteness as a grammatical category exists in Chinese?

Given the relevance of identifiability and definiteness to a wide range of linguistic phenomena, findings in this article, I would hope, would have implications for other studies involving these concepts, in particular relating to Chinese and those languages lacking *the*- and *a*-like articles, and also to other languages in general.

1.1. Definition of identifiability and definiteness

Identifiability in this article is taken as a pragmatic notion relating to the assumptions made by the speaker on the cognitive status of a referent in the mind of the addressee in the context of utterance. A referent is
considered to be identifiable if the speaker assumes that the addressee, by means of the linguistic encoding of the noun phrase and in the particular universe of discourse, is able to identify the particular entity in question among other entities of the same or different class in the context. Otherwise, it is considered to be nonidentifiable. For instance,

(1)  
   a. George finally bought a house.  
   b. George finally bought the house.

By using ‘a house’ in (1a), the speaker assumes that the addressee is not in a position to identify which particular house George bought; in uttering (1b), on the other hand, he assumes that the addressee knows which house he is talking about. The entity of house is presented as non-identifiable in (1a), and identifiable in (1b).

The terms “entity” and “referent” used in the article, it is to be noted, are shorthand for mental presentations, or mental files, of entities denoted by linguistic expressions in the universe of discourse constructed by the speaker and the addressee. I follow Lambrecht (1994: 36–37) in taking the universe of discourse as composed of two parts, the text-external world, and the text-internal world. The former comprises the participants and the spatio-temporal setting of a speech event, and the latter comprises the linguistic expressions and their meanings. Whether or not the entities exist in the “real world,” and whether they have been established in physical or linguistic terms seldom affect how the linguistics expressions are actually used to draw attention to what we are talking about, which is what linguists are really interested in, in contrast to philosophers and logicians, who are equally, if not more, interested in the ontological and epistemological aspects of the issue.1

The pragmatic notion of identifiability, or notions of a very similar nature, comes under different names in the literature, such as old vs. new (Halliday 1967), given vs. new (Clark and Clark 1977), definite vs. indefinite (Chafe 1976; Lyons 1977; Givón 1984/1990), and uniquely identifiable vs. nonidentifiable in (Gundel et al. 1993). The terminological, and in some cases substantive differences between writers need not concern us here. This article follows Chafe (1994: 93) and Lambrecht (1994: 77–79) in the usage of the terms of identifiable and nonidentifiable as defined above.

Definiteness, on the other hand, is used in this article as a grammatical concept, relating to the formal grammatical means in which identifiable and nonidentifiable referents are encoded distinctively in language. The formal grammatical means include phonological, lexical, morphological, positional, and other linguistic devices whose function it is to indicate whether a nominal expression is to be interpreted as identifiable or
nonidentifiable. Whether definiteness is a grammatical category in a particular language depends on how the notion is defined. Generally speaking, there are two senses, a broad one and a narrow one, in which definiteness is claimed to be a grammatical category. In the broad sense, it is understood as characterizing the major types of identifiable referring expressions, mainly personal pronouns, proper names, and definite noun phrases featuring one of the definite determiners in that language. Assuming that almost all languages in one way or another provide for these types of identifiable expressions, we would come to the natural conclusion that definiteness, in the broad sense of the term, is language universal. In the current literature, however, the notion of definiteness as a grammatical category is usually understood in a narrow sense: the defining criteria are whether there is a linguistic form, or forms, whose core or primary function it is to indicate identifiability, and whether the features of definiteness and indefiniteness are obligatorily and uniquely specified for nominal expressions in the language.

It is not always straightforward to decide on the basis of the above criteria whether a particular language has definiteness or not. When definiteness is marked by typically grammatical or functional morphemes in the form of affixes, clitics, or morphophonologically weak free forms (most importantly articles like the definite article the in English), which are called simple definites in C. Lyons (1999: 279), it is normally a clear case for the presence of definiteness as a grammatical category. Languages in this category include English, French, and other Germanic and Romance languages. In other languages which are genetically and geographically as diversely scattered as Chinese, Japanese, Czech, Russian, Warlpiri, Lango, Ik, Hindi, and Indonesian, identifiability is indicated primarily by forms such as proper names, demonstratives, personal pronouns, and possessives, which are called complex definites in C. Lyons (1999), or other grammatical means like word order. The major difference between simple definites and complex definites, apart from morphological autonomy and phonological weight, is that the encoding devices in the former group, whatever they may be diachronically derived from, have undergone the full process of grammaticalization and developed highly specialized uses to indicate identifiability or nonidentifiability of entities, while those in the latter group simultaneously, or even primarily, encode other grammatical features like deixis, person, saliency, topicality, and so on, in addition to identifiability. Linguists may look further into the languages that only have complex definites to determine whether definiteness is obligatorily and unambiguously encoded for nominal expressions in that language. If it is, it may be treated as a language with definiteness as a grammatical category; if it is not, it is taken as a language lacking
definiteness as a grammatical category. While identifiability, as a pragmatic concept, plays an important role in the form and function of human languages, definiteness, as a grammatical category defined in the narrow sense, may not be fully developed in some languages. It is, in the words of C. Lyons (1999: 278), the grammaticalization of identifiability. Like almost all other grammatical categories such as number, tense, or voice, it is present in some languages and absent in others.

The notion of definiteness in this article is used in its narrow sense. Applying the above criteria to Chinese, as will be elaborated in this article, I will conclude that definiteness as a grammatical category has not been fully developed in Chinese. On the other hand, the terms of definite and indefinite are also extensively used in the current literature, sometimes rather loosely, to refer to two distinctive groups of formal expressions which are normally, but not necessarily, interpreted as identifiable versus nonidentifiable in addition to other accompanying grammatical and pragmatic attributes. It is in the latter use that proper nouns, personal pronouns, demonstratives, some types of quantified NPs, etc. are referred to as definite expressions by many writers, with no commitment to any claim on whether there is definiteness as a grammatical category in the language in question. For convenience of exposition, and to ensure comparability, I also follow this practice in this article.

1.2. Definition of other related notions

It is appropriate in this connection to discuss briefly another two pairs of related notions: referential vs. nonreferential and specific vs. nonspecific.

There is a striking lack of general consensus on the definition and denotation of these terms in the current literature. Seldom do we find two writers who adopt the same definitions for these terms, or use the same term to cover the same range of linguistic phenomena. The differences are both terminological and substantive — in a way that one might ask whether or not these terms have passed their expiration date. What is presented below is a very sketchy account of my views on the relevant issues (cf. Chen 2004 for details). It is, I hope, sufficient for the purpose of this article.

I follow Payne and Huddleston (2002: 399) in defining a referential NP as one which refers to “some independently distinguishable entity, or set of entities” in the universe of discourse, where “independently distinguishable” means “distinguishable by properties other than those inherent in the meaning of the expression itself.” Nonreferential uses of nominal expressions, I propose, fall into three major groups.
First, and the “most nonreferential” are instances of what I call “non-individuated,” as the nominal expressions in this group are used primarily for the quality denoted by the expressions, rather than as individuals. They are considered to be “nonreferential in the semantic sense” by Hopper and Thompson (1984: 711) and Du Bois (1980), who have identified five types of this nonreferential use of nominals, as exemplified in the following sentences:

(2) a. incorporation of patient:
   I only wear one in my left when I’m wearing my lenses.

b. incorporation of oblique:
   We went to school yesterday.

c. noun compounding:
   pear tree, letter box

d. predicative nominal:
   He is the Prime Minister of Australia.\(^4\)

e. nominal in the scope of negation:
   Please don’t say a word.

In terms of formal encoding, the nominals in the nonreferential use can be indefinite ([2e]), definite ([2a] and [2d]), or bare nouns ([2b] and [2e]). Nonreferential expressions in this group are all characterized by the fact that they do not have continuous identity in the following discourse, and do not allow anaphoric reference.

The second group of nominals in the nonreferential use are nonspecific expressions. An expression is specific if the speaker uses it to refer to a particular entity in the universe of discourse, which may be identifiable or nonidentifiable; otherwise it is nonspecific. One of the most important defining features is that with a specific referent, the speaker may be able to provide more identifying information about it, or to use another referring expression of different linguistic encoding to refer to it. Whereas with a nonspecific referent, all the speaker knows is that it fits the description of the nominal expression. Consider the following examples:

(3) a. Everyday the chef comes to cook the dinner for us.

b. Everyday a chef comes to cook the dinner for us.

‘the chef’ in (3a) is specific, and identifiable. There are two readings for ‘a chef’ in (3b), one specific, and the other nonspecific. On the specific reading, the speaker may, though not necessarily, tell us more about his/her personal attributes such as age, appearance, etc. With ‘a chef’ on nonspecific reading, on the other hand, the speaker is unlikely to know anything beyond his/her type membership.
There are three major types of context in which a nominal expression may be subject to a specific and a nonspecific interpretation. The first type is represented by the so-called narrow scope NP, which is in the scope of another term that is quantified, as exemplified by (3b).

The second is marked by the irrealis, or non-fact modality of the proposition where the nominal is embedded, which, according to Givón (1984/1990: 393ff.), characterizes all of the following situations:

(4) a. sentences in future or habitual tense;
    b. within the scope of world-creating verbs like ‘want,’ ‘look for,’ ‘imagine,’ etc.;
    c. within the scope of complements of nonimplicative verbs and nonfactive verbs such as ‘believe,’ ‘think,’ ‘say,’ ‘claim,’ etc.;
    d. within the scope of probabilistic modal operators like ‘can,’ ‘may,’ and ‘must’;
    e. within the scope of irrealis adverbial clauses, imperative or interrogative sentences.

Consider (5), a classic example illustrating the distinction between a specific and a nonspecific interpretation of the nominal in a sentence of irrealis modality:

(5) John intends to marry a Norwegian girl.

John may already have a Norwegian girl in mind, whom he intends to marry. She may be known or unknown to the speaker. Or John simply intends to marry a Norwegian girl, whomever it may be. The noun phrase ‘a Norwegian girl’ is specific in the first case, and nonspecific in the second.

Finally, following an approach initiated in Partee (1970), the distinction traditionally drawn between the referential and the attributive use of definite expressions since Donnellan (1966) is captured in this article in terms of the contrast between specific and nonspecific. Consider the nominal ‘Smith’s murderer’ in (6):

(6) Smith’s murderer is insane.

It can have a referential or specific interpretation, or an attributive or nonspecific interpretation. The expression in the specific use can be replaced with other descriptions of the same person, which is not possible with the expression in the nonspecific or attributive use.

While all of them are nonreferential, nonspecific expressions differ from the nonindividuated expressions in that they may allow anaphoric reference in subsequent discourse, although usually subject to some restrictions. As observed by Karttunen (1976: 374), for instance, it is possible for
a nominal in the nonspecific use to be followed by a short-term anaphoric pronoun or definite noun phrase, “provided the discourse continues in the same mode.” Under either interpretation, the nominal expression in (5) can be followed by an anaphoric pronoun (also cf. Heim 1988: 249ff.).

(7) a. John intends to marry a Norwegian girl. She is a linguist. (specific)
   b. John intends to marry a Norwegian girl. She must be a linguist. (nonspecific)

The third major use of nonreferential expressions is for generic reference. It refers to a kind or a genus, instead of a particular object. Since it has no direct relevance to the subject of this article, we will not discuss it here (cf. Kriska et al. 1995).

2. Cognitive basis of identifiability

As a cognitive concept, identifiability denotes a status of the referent in the mental representations of the participants of a speech event. Its social and expressive functions aside, a speech event may be taken as a process by which the speaker instructs the addressee to reconstruct a particular mental representation of events and ideas that the speaker himself has in his mind. When he chooses among a range of possible alternatives what he believes to be the most felicitous way to encode and send the message to the addressee, the speaker depends crucially on his assumptions regarding the various statuses of the entities, attributes, and their links which comprise the mental discourse model in the minds of the speech participants, particularly in the mind of the addressee — such statuses as relating to their location in memory, predictability, attention state, and so on. These assumptions, furthermore, are continuously adjusted and updated in the ongoing, dynamic process of communication.

What are the factors on the basis of which the speaker comes to assumptions one way or another on the identifiability of entities for the addressee? The status of being identifiable can be assumed by the speaker to have been established for an entity between him and the addressee by virtue of a variety of identificatory resources. Roughly speaking, they fall into two major categories. In the first category, the identifiability is directly evoked from its presence in the context of discourse, which is composed of the physical situation of utterance, and the linguistic text. In the second category, the identifiability of the entity in question is established on the basis of shared background knowledge between speaker and addressee, or inferable from other entities in discourse by virtue of the
knowledge shared by participants of the speech event about the associations between the former and the latter.

2.1. Direct physical or linguistic co-presence

The identity of an entity is considered to be contextually evoked when the entity in question is located in the spatio-temporal universe of discourse where the speaker and the addressee are co-present and can be uniquely identified by means of the linguistic expression used with or without accompanying paralinguistic expressions. For instance, catching sight of someone who has just entered the room, the speaker, most likely with accompanying bodily gesture or sight in the direction of the person, may ask the following question to the addressee without having provided any other information about the person prior to the utterance:

\[(8) \text{Do you know who he/that man/the man is?}\]

In addition to the physical situation of utterance, a more common type of the context of discourse is constructed through the use of language by the participants of the speech event. In so far as identifiability is concerned, entities that have been introduced into the text by the speaker are comparable to those that make their first appearance in the physical environment. After a referent has been introduced into the context, it can be treated as identifiable on subsequent mentions, given that enough identificatory linguistic encoding is provided. The identifiability of such a referent is taken as textually evoked, as illustrated by the following example:

\[(9) \text{There is a dog and a cat in my backyard. The dog loves to chase the cat.}\]

‘the dog’ and ‘the cat’ refer backward to the correlated entities introduced into the text by ‘a dog’ and ‘a cat.’ They represent the typical use of definite expressions in anaphoric reference.

2.2. Shared background knowledge

The identity of a referent is also established on the basis of the shared knowledge between the participants of the speech event about their physical and linguistic environments, which may vary considerably in terms of scope and nature. It may involve only the speaker and the addressee, or it may be so broad as to cover all those who live in the same social and cultural environments. In the situation of a family that has a pet dog, the
husband may ask his wife the following question when he returns from work in the evening, with the assumption that the addressee is able to establish the identity of the noun phrase ‘the dog’:

(10) Where is the dog?

In this particular context, it is normally impossible for the speaker to accompany the utterance with any deictic gestures as the dog in question is physically absent; it is not necessary for the dog to be verbally introduced into the discourse before being treated as identifiable. The identifiability of the dog derives from the background knowledge shared between the husband and the wife that they have a dog, and a dog only. It is by the same token that the identity of the referents of such noun phrases like ‘the house,’ ‘the river,’ ‘the City Council,’ ‘the Prime Minister,’ ‘the President,’ and ‘the sun,’ as well as proper nouns, is established in contexts of varying scope and nature. The knowledge involved can be very specific, as is the case with the dog in (10), or very general, as with the sun.

The identity of a referent can also be inferred from other entities or activities in the discourse through logical reasoning on the basis of the general knowledge of the interrelationship among the entities or activities involved. They are often interrelated in such a way that the mention of one will automatically bring the mental representation of others that are customarily associated with it into the consciousness of the participants of the speech event. The knowledge of such interrelationships among entities is generally shared by all the members of the group, constituting an important feature defining the membership of a certain community. Such organization of knowledge in memory is captured in terms of theoretical constructs such as frame, schema, script, scenarios, etc. in cognitive sciences. Consider the following example:

(11) David bought an old car yesterday. The horn didn’t work.

That cars have horns can be assumed, in the context of the modern society, to be part of general knowledge in the possession of ordinary language users. This enables the speaker to assume that, once the referent of a car has been introduced into discourse, the addressee is able to establish the identity of the horn as being the part of the car. In this case, the antecedent of the horn is not directly mentioned in the previous discourse; instead, it is identified as the horn of the car through what Clark (1977) calls “indirect reference by association,” which is a type of bridging cross-reference. Consider another example:

(12) Joe bought a used car yesterday, but the seller later claimed that he didn’t get the money from Joe.
Among the major stereotypical information slots that a selling frame, as presented in the first clause, characteristically has are a buyer, a seller, an object, and money that change hands. Although the seller and the money in the second clause are mentioned for the first time in the discourse, their identity has been established through the evocation of the frame as the seller and the money, which are the fillers of the slots of the transaction.

Indirect reference by association involves anaphora and shared general knowledge simultaneously. It looks backward to an entity or situation that has already been present in the universe of discourse, in a way similar to the normal anaphoric reference. Rather than in a direct reference to the correlated referent in previous discourse, it refers to one whose identifiability is inferred through association with another referent or situation in the general knowledge of the participants.

Referents which derive their identifiability through association may display varying degrees of identifiability, depending on the types of frames and referents, as well as on the extent of familiarity with the frames on the part of addressees. A distinction is drawn in the literature between two kinds of bridging cross-reference: one represented by (13) quoted from Haviland and Clark (1974: 515), and the other by (14) from Sanford and Garrod (1981: 104):

(13) a. We got some beer out of the trunk. The beer was warm.
    b. We checked the picnic supplies. The beer was warm.

(14) a. Mary put the baby’s clothes on. The clothes were made of pink wool.
    b. Mary dressed the baby. The clothes were made of pink wool.

In contrast to the direct anaphoric reference to the antecedent in the previous sentence in (13a) and (14a), the identity of the definite nominals in (13b) and (14b) is indirectly established through association in terms of the frames of ‘picnic supplies’ and ‘dressing.’ It is reported in Haviland and Clark (1974) that in psycholinguistic experiments, it takes more processing time for most subjects to establish the connection between the anaphoric definite expression in the second sentence and its “trigger” indirect antecedent in (13b) than is the case with the direct antecedent in (13a). However, no significant difference in processing time is found between (14a) and (14b) in the comprehension experimentation by Sanford and Garrod (1981). The results of the two experiments suggest that some associations are easier to establish than others. The association between ‘the clothes’ and the ‘dressing’ frame is part of the general knowledge of ordinary people so that the frame will easily or automatically activate ‘clothes’ in our mental representation. On the other hand, as noted by Brown and Yule (1983: 263), the connection between picnic supplies and
the beer is not as readily made by readers other than “a group of real ale enthusiasts who often indulge their enthusiasm on picnics at the local park.” In other words, the ‘clothes’ in (14b) is more identifiable than the ‘beer’ in (13b) with most addressees. In spite of the difference, all the referents in question are encoded in the same way as a definite NP marked by the definite article.

The identifiability of a referent may also derive from its association with information that is contained in the nominal expression itself. Consider the following example:

(15) Do you know the man that she went to dinner with last night?

It may well be that the referent of ‘the man’ appears for the first time in the universe of discourse. It is treated as an identifiable referent through the identifying function of the restrictive relative clause that follows it: she went to dinner with a man last night, and ‘the man’ refers to that particular man (cf. C. Lyons 1999: 5 for a detailed discussion of the relevant issues). A similar case called “containing inferrable” is discussed by Prince (1981: 236–237), who gives the following illustrative example:

(16) Have you heard the incredible claim that the devil speaks English backwards?

in which the identifiability of the definite referent ‘claim’ is inferenced off from the following clause that is properly contained within the inferrable NP itself.

2.3. Degrees of identifiability

It is evident from the discussion above that identifiability, as a pragmatic concept, is a matter of degree. From full identifiability to complete non-identifiability is a continuum with no clear line of demarcation anywhere along it. In languages in which identifiability is grammaticalized in terms of definiteness, speakers are usually forced to make a decision on whether to encode entities of varying degrees of identifiability in definite or indefinite terms. The cut-off line between definite vs. indefinite encoding along the continuum of identifiability is not always readily obvious in any language. It is a common phenomenon that a referent of partial identification is treated as identifiable, receiving a definite encoding in the same way as a referent of full identification. Consider the following example from Du Bois (1980: 232):

(17) The boy scribbled on the living-room wall.
As argued by Du Bois (1980: 232), it is not a necessary condition for the definite encoding of the referent of ‘the living-room wall’ for the addressee to be able to identify precisely which of the four walls of the living room is involved. Du Bois (1980: 232) maintains that the definite encoding here is justified so long as the speaker assumes that “the addressee is able to identify the particular living room in question, and to narrow down the range of possible referents to one of the four walls.” Du Bois (1980: 232) also notes that the speaker could be “violating the Gricean maxim of relevance by giving more information than people care to know,” if he specifies exactly which wall, as in (18):

(18) He scribbled on the north living-room wall.

Du Bois (1980) also points out that to present the wall as nonidentifiable as (19)

(19) He scribbled on a living room wall.

would be violating the maxim from the other direction, because it presupposes an excessive curiosity about the walls on the part of the addressee. To explain the phenomenon, Du Bois (1980) has proposed what he calls the curiosity principle:

A reference is counted as identifiable if it identifies an object close enough to satisfy the curiosity of the hearer. The identification need not to be one to satisfy a philosopher or a Sherlock Holmes, who may of course be led to demand “Which wall?” In special circumstances even an ordinary speaker might desire more precise identification. But in everyday speech such partial identification is quite common. (Du Bois 1980: 233)

The lack of full identification for referents which are encoded as definite is mostly confined to those which derive their identifiability from semantic frames discussed above. It is noted in Löbner (1985: 302) that frame-triggered referents may stand in a one-to-one relationship to the anchor, like driver to a car and president to a state, or in a one-to-many relationship, like daughter to a parent and friend to a person. Löbner argues that the identifiability of a definite expression need not be determined in an absolute sense, and a definite article can be used to mark a noun so long as the referent is one that stands in a one-to-one relationship to the anchor in spite of the fact that the overall NP may be nonidentifiable. Thus the grammaticality of (20):

(20) the mayor of a small village in Wales

A case is presented in Lambrecht (1994: 91) and C. Lyons (1999: 26) in which a referent is treated as identifiable where the conditions for
identifiability defined in Lübner’s terms do not strictly hold. In (21), for instance, there is no implication that the speaker has only one brother.

(21) I’m going to stay with my brother for a few days.

And (22) would be appropriate, as Lambrecht (1994) remarks, even if the unidentified king in question has more than one daughter.

(22) I met the daughter of a king.

As long as the information provided by the noun phrase is sufficient for the communicative purpose of the utterance, there is no need to specify it any further. Obviously, the same curiosity principle as formulated by Du Bois (1980) is at work here.

The above issue arises, I maintain, to a large extent as a result of the fact that the speaker is obliged, as a result of the grammatical constraints of definiteness as a grammatical category in English, to make a selection between definite vs. indefinite encoding for the referent in question. It may no longer be an issue in a language without definiteness as an obligatory grammatical category for nominal expressions, such as Chinese, where the referent may be encoded in a way that is neutral with respect to the interpretation of identifiability. The difference between English and Chinese in this respect is readily seen when (17) is translated into Chinese: the referent in question will most likely assume the form of an indeterminate expression (to be explained in Section 4) in a sentence position that does not make any clear indication or suggestion to the addressee regarding whether the expression is to be interpreted as identifiable or non-identifiable. I will return to this point later.

3. Linguistic encodings of identifiability

Irrespective of whether there is definiteness as a grammatical category, the distinction between identifiability and nonidentifiability can be encoded in one way or another in all the languages of the world (cf. Haspelmath 1997; C. Lyons 1999; inter alia). While it is typically encoded in terms of respective formal markings which can be phonological, lexical, morphological, and syntactic, languages vary considerably in the types of encodings most commonly used for the purpose and in how they are used. To bring a crosslinguistic perspective to definiteness in Chinese, let us start with a brief account, based on recent findings reported in the literature, of the linguistic encodings of identifiability in two types of languages, one with and the other without definite or indefinite articles.
The former is represented by English, and the latter by Czech, Hindi, and Indonesian.

3.1. **English**

Definite expressions in English fall into three major categories, namely, definite NPs, proper nouns, and personal pronouns. Definite NPs feature one of the following definite determiners:

1. definite article the;
2. demonstratives this/these, that/those;
3. possessives like my, our, his, and so on;
4. universal quantifiers like all, every, random any, and so on.

3.1.1. **Definite article.** As the most important definite determiner, the definite article represents an exemplar par excellence of the grammaticalization of identifiability. Its core function is to indicate that the referent, or more precisely the mental representation of the referent in the universe of discourse, that it is used with is to be interpreted as an entity that the addressee can identify from among the other members of the class in the context. Unlike other definite expressions such as demonstratives, proper nouns, or personal pronouns, the article itself does not have any descriptive content other than the ostensive function. As is the case with the overwhelming majority of the languages in the world, and certainly with all the Germanic and Romance languages, the English definite article derives diachronically from a demonstrative pronoun. As a fully grammaticalized marker of definiteness to indicate the identifiability of the noun phrase it modifies, it is neutral with regard to deixis, person, number, gender, or any other grammatical features.

Given the highly specialized role of the definite article, it is only to be expected that it stands to serve as a marker of definiteness in all the situations in which identifiability of reference is derived. The uses of the English definite article fall into four major categories, namely situational, anaphoric, shared specific or general knowledge, and associative, covering all the sources of identifiability of referents as discussed in the last section. Following are examples illustrating each of them (cf. Christopherson 1939; Hawkins 1978; C. Lyons 1999; inter alia):

(23) situational:
Get a knife for me from *the* table.

(24) anaphoric:
I saw a man pass by with a dog. *The dog* was very small and skinny, but *the man* was very large.
shared specific knowledge:
Be quiet. Do not wake up the baby (who is sleeping in the next room).

shared general knowledge:
The sun is brighter than the moon.

frame-based association:
They bought a used car. The tires were all worn out.

self-containing association:
Do you know the man who lived in this room last year?

He broke the window glass with the handle of a bike.

The uses of the as discussed above will serve as a template in the examination of the uses of other definite determiners in English and Chinese.

3.1.2. Demonstratives. Demonstratives differ from definite articles in two major aspects. First, while definite articles have adjectival uses only, demonstratives typically have adjectival, pronominal, and adverbial uses
as well. Second, the primary function of demonstratives in English is that of deixis, which has been extended to other uses as well (cf. Fillmore 1982, 1997; Himmelmann 1996; Diessel 1999). They serve to locate and identify entities with reference to their distance in relation to the speech participants in the spatio-temporal space of discourse. As determiners of definiteness, they are mainly found in deictic uses, signaling to the addressee in one way or another that the referent in question is accessible to him in relation of the position of the participants in the context of utterance. Definite articles, in contrast, are deictically neutral.

The uses of demonstratives, following Himmelmann (1996), fall into four major types: situational, discourse deictic, anaphoric, and recognitional.

(30) situational:  
Could you please give me a hand with this big box?

(31) discourse deictic:  
He did not answer our phone call as promised. This is not good.

(32) anaphoric:  
There is a zoo a couple of miles down the road. You won’t see many animals in that zoo.

(33) recognitional:  
It was filmed in California, those dusky kind of hills that they have out here by Stockton and all.

Demonstratives in situational use differ from definite articles in that the former are subject to the restriction that the referent in question must be visible to the addressee. Compare the following examples:

(34) a. Beware of the dog.
   b. Beware of that dog.

(34a), but not (34b), is felicitous if the dog is invisible, but its existence can be inferred from the context. As Hawkins (1978: 112) notes, “demonstratives are only possible in these cases if the interlocutors can actually see a dog at the time of the utterance.” The explanation lies in the deictic component in the semantics of the demonstratives, which distinguishes them from the definite article. The use of the deictics assumes that the addressee is able to locate the referent in terms of its location relative to the participants of the speech event. The assumption would be invalidated if the referent is physically absent from the immediate situation of the utterance.

Anaphora, as enunciated by J. Lyons (1977: 670), involves the transference of what are basically spatial notions to the temporal dimensions of the context of utterance and the reinterpretation of deictic location in
terms of what may be called location in the universe of discourse. With deixis underlying anaphora, the English demonstratives are found in anaphoric use as well. The anaphoric uses of the demonstratives are much less common in comparison with their deictic uses, and also in comparison with the anaphoric uses of other definite determiners like the definite article and personal pronouns. When they are used anaphorically, it is usually with a contrastive sense.  

(33) is quoted from Himmelmann (1996: 230), who characterizes it as “recognitional,” a term borrowed from Sacks and Schegloff (1979). It is first observed in Sacks and Schegloff (1979), and later developed in Schegloff (1996) that when the speaker does not know with certainty whether a referent is identifiable enough for the addressee, as happens very often in informal talks, he usually prefers a definite expression, either in the form of a proper name, or “recognitional” descriptions, which presume some familiarity on the part of the addressee with the referent rather than using an indefinite expression which treats the referent as nonidentifiable. The speaker will often try different wordings, called “try-marked” recognitionals by Sacks and Schegloff (1979), for this definite expression until he perceives recognition on the part of the addressee (cf. Ford and Fox 1996; also cf. Grice 1989 and Levinson 2000 for an explanation). Demonstratives in such recognitional uses are mainly found in situations where the speaker is not very sure whether the relevant knowledge that is crucial for the identifiability of the entity is shared by the addressee or not. As is the case when recognition on the part of the addressee is in doubt, demonstratives in such uses are typically accompanied by expressions like you know? and remember?, seeking confirmation of the information being shared by the addressee. For a detailed discussion, cf. Sacks and Schegloff (1979), Himmelmann (1996), and Schegloff (1996).

Demonstratives in recognitional use are used to refer to referents that have been previously introduced into discourse, or to introduce referents into discourse for the first time. It is, in my view, a combination of the shared knowledge and self-containing uses of definite determiners. On the one hand, the speaker appeals to the knowledge that he assumes, albeit without much certainty, to be shared by the addressee; on the other hand, he phrases the expression in a way that he hopes will provide sufficient identifying information for the addressee to identify the referent in question.

Apart from the recognitional use, which is accompanied by some restrictions, the English demonstratives are normally unacceptable for referents which derive their identifiability through shared specific or general information, as in (35) and (36), or through association, as in (37) and (38):
Be quiet. *Don’t wake up *that baby (who is sleeping in the next room).

*That sun* was covered by dark clouds.

They bought a used car. *These/those tires* are all worn out.

*He broke the window glass with *this/that handle of a bike.*

3.1.3. **Grammaticalization of demonstratives into definite articles.** It is well-attested in the languages of the world that demonstratives are the most common sources from which definite articles are derived through the process of grammaticalization. In the discussion of the cycle of the definite article, Greenberg (1978: 61) describes how a demonstrative, which he calls Stage 0, develops into a definite article, which he calls Stage 1. In a number of instances that have been studied in detail, Greenberg finds that the process of grammaticalization starts when a purely deictic element has come to identify an element as previously mentioned in discourse.

The point at which a discourse deictic becomes a definite article is where it becomes compulsory and has spread to the point at which it means ‘identified’ in general, thus including typically things known from context, general knowledge, or as with ‘the sun’ in nonscientific discourse, identified because it is the only member of its class. (Greenberg 1978: 61–62)

His view can be summarized in the diagram of (39):

(39) Stage 0 situational deictic > transitional anaphoric > Stage 1 shared knowledge association

The view is shared by Diessel (1999), who maintains that

the use of anaphoric demonstratives is usually confined to nontopical antecedents that tend to be somewhat unexpected, contrastive, or emphatic. When anaphoric demonstratives develop into definite articles their use is gradually extended from non-topical antecedents to all kinds of referents in the preceding discourse. In the course of this development, demonstratives lose their deictic function and turn into formal markers of definiteness. (Diessel 1999: 128–129)

The above will serve as our guiding criteria when we examine the emerging uses of Chinese demonstratives as definite articles. As markers of definiteness, demonstratives and definite articles differ crucially in that the former are deictic and the latter are not. The process of grammaticalization of demonstratives into definite articles is one in which the deictic force of the demonstratives is gradually bleached out, which is often accompanied by phonological reduction, loss of morphological and
grammatical autonomy, etc. As a result of which the demonstratives gradually extend their uses to situations that call for a deictically neutral determiner of definiteness. It has been attested in all the languages that did not, or do not have definite articles. As grammaticalization is by nature a gradual process, we are more likely to be concerned with transitional stages and borderline cases rather than distinct categories in the studies of the development of demonstratives into definite articles in particular languages. Demonstratives in Chinese, as we will show shortly, display some features characteristic of a transitional stage in the process.

3.1.4. Possessives. English noun phrases with possessives such as *my*, *his*, *John’s*, as premodifiers are definite expressions. It is ungrammatical to insert an indefinite determiner between the possessive and the head noun. However, as C. Lyons (1999: 24) points out, it would be wrong to assume that possessives are definite determiners crosslinguistically. In languages like Italian and Greek, possessives do not impose an interpretation of identifiability on the head noun: if the head noun is to be interpreted as identifiable, a definite article is used; and if it is nonidentifiable, an indefinite article is used, as shown by the Italian examples: *il mio libro* ‘lit. the my book’ (“my book”) and *un mio libro* ‘lit. a my book’ (“a book of mine”).

The difference between English and Italian in this regard is captured by C. Lyons (1999: 24) in terms of a typological distinction between a determiner-genitive (DG) language and an adjectival-genitive (AG) language. In DG languages, possessives appear in positions reserved for definite determiners, while in AG languages, they are in adjectival or some other position. It is also observed by C. Lyons (1999) that while a nonidentifiability reading is impossible with the basic possessive structure in DG languages, a prepositional construction is most commonly used when the head noun is nonidentifiable, with or without the co-occurrence of an indefinite marker. Examples from C. Lyons (1999) are English *a friend of mine*, French *un ami a moi* ‘a friend to me,’ German *ein Freund von mir* ‘a friend of mine,’ and Irish *cara liom* ‘friend with me’ (“a friend of mine”).

3.1.5. Indefinite markers. The most important indefinite marker in English is the indefinite article *a*. Unstressed *some* and *any* are also used as indefinite markers to indicate that the entity they modify is to be interpreted as nonidentifiable. The weakly stressed *this* also serves as indefiniteness marker, mainly in colloquial speech and typically for referents of high thematic importance with continuing presence in the ensuing discourse.
As is the case in the great majority of languages that have indefinite articles, the indefinite article in English derives from the numeral *one*. It is proposed in Givón (1981, 1984/1990) that there are a few major steps along the functional continuum along which the numeral develops gradually into an indefinite article through the process of grammaticalization:

(40) quantification > referentiality/denotation > genericity/connotation

The English article *a* serves all the three major functions, as exemplified in the following examples:

(41) quantification:
   It needs *an hour* and *a quarter*.
(42) referentiality/denotation:
   He bought *a book*.
(43) genericity/connotation:
   He is *a teacher*.

It is further proposed in Heine (1997) that the second stage of referentiality/denotation consists of three uses, namely as a presentative marker, a specific marker, and a nonspecific marker. The article at the final stage of development is called a “generalized” article by Heine.

It is important to note that, as argued in C. Lyons (1999: 33–34), a noun phrase in English is indefinite if it has no definite determiner, whether or not it has an indefinite determiner. Thus, count nouns in the plural and mass nouns, as in (44) and (45), are indefinite where they are not interpreted generically:

(44) John has gone out to buy *milk*.
(45) I have already put *spoons* on the table.

Cardinality terms like *two*, *three*, and *many* are neutral with respect to definiteness. They can be preceded by a definite determiner, as in:

(46) Pass me those *three books*.
(47) I’ve only read *a few of the many books she’s written*.

When they occur without any determiner, they are indefinite in a way similar to count nouns in the plural and mass nouns.

3.2. *Languages without definite articles*

Two recent studies, Cummins (1998) and Porterfield and Srivastav (1988), have explored the correlation between identifiability and definiteness in
Czech, Hindi, and Indonesian, which, like Chinese, are languages that lack simple definiteness markers such as the English definite article. For lack of space, what is presented here is a very brief summary of their main points. Readers are referred to the two papers for a detailed account of their findings.

Identifiability in Czech, according to Cummins (1998), is marked in terms of various linguistic devices including word order and intonation. Bare NPs and cardinality NPs in sentence-initial positions are definite expressions. Their default reading in sentence-final positions is indefinite. Bare NPs in final positions are definite only when the entity in question is “at hand in context or in memory” (Cummins 1998: 578). A demonstrative *ten* is often used as a definiteness marker in the language, serving as the most close approximate to an definite article like the English *the*. Both bare NPs and *ten*-modified NPs are found in anaphoric uses.

Bare singular NPs in Hindi, as concluded by Porterfield and Srivastav (1988), are always to be interpreted as definite expressions. In situations where they are apparently indefinite, they are in kind-level generic use. In Indonesian, bare NPs are always in generic use (cf. Porterfield and Srivastav 1988).

What the three languages have in common is that the bare NPs, cardinality NPs, and demonstrative modified NPs only have one reading in the sentence, which is assigned to them according to their lexical and morphological encodings and word order. Generic uses aside, they are to be interpreted as identifiable in certain positions or uses, and as non-identifiable in other positions or uses. It is claimed in Cummins (1998: 567) that the underlying notions of definite and indefinite appear to be as unitary and as central to the semantics and pragmatics of NPs in Czech as they are in languages with articles like English and German. Based on his studies of Czech as an articleless language, Cummins (1998) proposes that as a grammaticalized subcategory of nominal determination, definiteness is a linguistic universal. It is further claimed in Porterfield and Srivastav (1988) that the feature of definiteness vs. indefiniteness should be obligatorily and uniquely specified in all languages.

While apparently belonging to the same group of languages without articles, Chinese, as I will demonstrate below, represents a case that is different from what has been established for the three languages.

## 4. Definiteness in Chinese

The identifiable vs. nonidentifiable contrast is encoded in a more complex manner in Chinese. Three major types of linguistic devices — namely,
lexical, morphological, and positional — are employed to indicate or suggest to the addressee whether the nominal expressions should be interpreted as being of identifiable or nonidentifiable reference.

4.1. **Lexical**

There is no *the*-like definite article in Chinese. In terms of lexical encoding, other than proper names and personal pronouns, three major groups of definite determiners serve the function of marking a referent as identifiable in Chinese. They are demonstratives, possessives, and universal quantifiers.

4.1.1. **Demonstratives.** The most important definite determiners in Chinese are demonstratives *zhe* ‘this’ and *na* ‘that’ and their plural forms *zhexie* ‘these’ and *naxie* ‘those.’ In contemporary Chinese, particularly in the Beijing dialect, *zhe* and *na* can also take the forms *zhei* and *nei*, respectively, when the deictic distinction is highlighted as in situational uses (cf. Lü 1990 [1956], 1990 [1985]; Zhang and Fang 1996; Fang 2002).9 They are used both as pronouns and adjectives. It is to be noted that the singular vs. plural distinction is not strictly observed in the use of Chinese demonstratives, or for that matter, Chinese personal pronouns and nouns. More often than not, a singular form is used where semantics would dictate a plural form. As discussed earlier, demonstratives are the most common source from which definite articles or similar determiners of definiteness derive. Chinese is no exception in this regard. With phonological reduction and the deictic component in their meanings weakened in varying degrees, *zhe* and *na* in Chinese are the closest to definite articles in other languages, as has often been noted in the literature (cf. Lü 1990 [1956], 1990 [1985], 1990; S. Huang 1999; Tao 1999; Fang 2002).

As demonstratives, *zhe* and *na* serve all the major functions of demonstratives as discussed in the last section. Consider the following examples:

(48) situational:
Qing ba *zhe/na* zhang yizi ban dao *na* jian fangjian qu.
Please BA this/that CL chair move to that CL room go
‘Please move this/that chair to that room.’

(49) discourse deictic:
Ta xiang huiqu? *Zhe* ni ke buneng daying.
He want return this you surely cannot agree
‘He wants to go back? You surely cannot give your permission to that.’
anaphoric (contrastive):

Yi ge xiaohaizi... lu zhong de shihou ne, pengdao le one CL kid road middle DE time SFP run:into PFV yi ge nühaizi, zhe ge shihou, zhe ge xiaohaizi you kan one CL girl this CL time this CL kid again look le na ge nühaizi, kan le yi yan. PFV that CL girl look PFV one eye

‘A kid in the middle of the road ran into a girl,... this time the kid had another look at the girl.’

When used anaphorically, zhe is preferred for a referent that has just been introduced into discourse. When the referent is referred to later in the discourse, particularly after several other intervening referents, na is more often used than zhe. Consider the following example from a story in Zhongguo Yuyan Gushi:10

Zai tong xiang cheng li de yi tiao da lu pang, zhang zhe on lead to town in DE one CL big road side grow DUR yi ke da shu. Zhe ke shu de shugan shang you yi ge hen one CL big tree this CL tree DE trunk up have one CL very da de dong. Yi tian, yi ge dayude ren xiang jin big DE cavity one day one CL fisherman person intend enter cheng mai yu ... zhe wei yufu jimang pao dao na ke town sell fish this CL fisherman hurriedly run to that CL you da shudong de shu xia duo yu ... Yi tian, na wei have big cavity DE tree down dodge rain one day that CL wang shudong li fang yu de yufu you jinguo zheli. to cavity in put fish DE fisherman again pass here

‘On the side of a main road leading to the town, there was a big tree. On the trunk of the tree there was a very big cavity. One day, a fishermen wanted to go to the town to sell fish ... The fisherman rushed to the tree with a big cavity to take shelter from the rain ... One day, the fisherman who had put a fish in the cavity passed by again.’

The proximal zhe is used for the tree and the fisherman right after they have just been introduced into the discourse. When the referents appear again later in the story, both of them are introduced by the distal na. This is the normal pattern for the selection of demonstratives in such anaphoric uses.

recognitional:

Ta jiu gei na .... xiaoahazi ...., get na ge huan ta he then give that kid give that CL return he
maozi de xiaohaizi san ge bale.
cap DE kid three CL pear
‘He then gave the . . . . . . boy, . . . . . . the boy who returned the cap to him three pears.’

na is the determiner that is used to introduce such try-marked recognitionals in Chinese, as illustrated in (52). The speaker of (52) is uncertain from the beginning whether the addressee is able to identify which boy he is referring to, as evidenced by his pauses of hesitation; after the first try with na . . . xiaohaizi, he offers another more detailed description with a relative restrictive clause. na is used with both try-marked recognitionals in the utterance.

Compared with the English demonstratives, the Chinese demonstratives are subject to less restriction when used as markers of definiteness. In fact, they are found in some of the contexts where in English the definite article is regularly used as a marker of definiteness and the demonstratives are not generally allowed. Such uses of the Chinese demonstratives are illustrated in the following examples:

(53) anaphoric (noncontrastive):
You yi ge lieren . . . yang zhe yi zhi gou. Zhe zhi gou hen have one CL hunter keep DUR one CL dog this CL dog very dongshi.
intelligent
‘There was a hunter who had a dog. The dog was very intelligent.’

(54) shared general knowledge:
Zhe tianqi zhen guai, shi-er yue le, ke yidian bu leng. This weather really strange twelve month CRS but bit not cold
‘The weather is really strange. It is December now, but it is not cold at all.’

(55) frame-based association:
Ta mai le yi liang jiu che, na luntai dou mo ping le. he buy PFV one CL old car that tire even wear flat CRS
‘He bought an old car. All the tires are worn out.’

(56) self-containing association with accompanying restrictive relative clause:
Shang ge yue lai kan ni de na ge ren, wo jintian last CL month come see you DE that CL person I today you jian dao ta le.
again see to he CRS
‘The person who came to see you last month, I saw him again today.’
Unlike the situation in English, the use of the Chinese demonstratives in noncontrastive anaphoric reference as in (53), and with restrictive relative clauses as in (56) is very common in Chinese of all styles. On the other hand, uses of the Chinese demonstratives with referents whose identifiability stems from shared general knowledge or frame-based association as exemplified in (54) and (55), are only found in texts of the vernacular style.

Demonstratives in Chinese, as is obvious from the discussion above, have extended their use to definite articles, serving some of the functions that are characteristic of the definite article like *the* in English. It is by virtue of the functions beyond those of pure deictics that Lü (1990), S. Huang (1999), Tao (1999), and Fang (2002) rightly claim that *zhe* and *na* are used, in some situations, as definite articles.

While it is clear that *zhe* and *na* have started on the path of grammaticalization into definite articles, there is evidence which suggests that they are still far from reaching the endpoint yet.

First, most of the instances of *zhe* and *na* which apparently have much weakened or no deictic force are found in anaphoric and recognitional uses, as reported in studies based on text counts in S. Huang (1999) and Tao (1999). These uses, as argued in Greenberg (1978) and Himmelmann (1996), are characteristic of the beginning or transitional stage of the grammaticalization of demonstratives toward definite articles, rather than representing typical uses of fully grammaticalized definite articles.

There is evidence that *zhe* and *na* have preserved their deictic force, to a certain extent, in these uses which are considered to be transitional. When *zhe* and *na* are found in contrastive anaphoric use, as discussed above, *zhe* is preferred over *na* as the anaphoric device for an antecedent that is recently introduced into discourse. In noncontrastive anaphoric use, *zhe* outnumbers *na* by approximately three to one in my corpus. In textual deictic uses, where no contrast between referents is normally involved, *zhe* outnumbers *na* by approximately six to one (Tao 1999: 82). This makes best sense if we assume that the proximity of *zhe* makes it a better anaphoric device than the distal *na* in referring to an antecedent recently introduced into discourse.

The deictic distinction may also play a role in a more delicate manner in those circumstances which are apparently neutral with regard to proximity. No deictic element appears to be involved when referents depend on accompanying relative clauses for their identifiability: the referents become identifiable through the qualitative information provided by the relative clause and the noun itself, and no locative information is at play in the process. Table 1 presents the distribution of *zhe* and *na* as the definite determiner of an NP modified by a relative clause.
Table 1 shows that while both *zhe* and *na* are found in such uses, *na* is by far the preferred determiner with a head noun modified by a restrictive clause. A total of 36 such instances are found in my corpus of data. *na* is used in 33 instances and *zhe* only in three. When the referent is used with a nonrestrictive relative clause, no preference is evident. A total of eight such instances are found in the corpus, with *zhe* used in four instances, and *na* in the other four. One explanation for the differential is that of the two demonstratives, *na* is the more grammaticalized, or unmarked, determiner of definiteness for referents which are neutral with respect to the deictically based distinctions, and is thus more appropriate for uses with nouns modified by restrictive relative clauses. It may also be explained, I propose, in terms of the inherent deictic distinctions between the two demonstratives. When a referent depends crucially on the accompanying restrictive relative clause for identifying information, it is very likely that it is further away from the speaker, both physically and cognitively, than a referent that has established its identifiability through other means. Given the circumstances, the distal demonstrative is more appropriate for that referent than the proximal demonstrative.

Second, *zhe* and *na* are either not allowed, or very rarely found for the uses which are considered to be prototypical of definite articles, namely, for uses with referents which derive their identification through shared specific and general knowledge or frame-based association. Consider (57):

(57) Anjing dianr, bie ba *na* haizi chaoxing le.
    quiet bit don’t BA that baby wake:up CRS
    ‘Be quiet. Don’t wake up that baby.’

In a way similar to the behaviors of the English demonstratives in (34) and (35), the Chinese demonstrative would be infelicitous if the baby is not visible to the addressee, and the addressee has no previous knowledge that there is a baby in the house.

Sentences like (54) and (55), while acceptable in colloquial speech, are statistically very rare. Only three instances in my corpus are found in
which a demonstrative is used as a marker of definiteness for referents that derive their identification from shared general knowledge as in (54), and two instances for uses of frame-based association as in (55). All of the five instances occur in the vernacular pear stories, and not a single instance is attested in the written Chinese fables. Even these five instances are open to another interpretation, not as adnominal determiner, but as markers of hesitation and false start, which abound with some speakers.

Granted that *zhe* and *na* are sometimes used in deictically neutral contexts, neither of the Chinese demonstratives has developed in the direction of a definite article to such an extent that their primary function is to serve as a deictically neutral marker of definiteness like the English *the*. In his studies of the historical development of definite articles in languages, Greenberg (1978: 61) observes that a definite article “develops from a purely deictic element which has come to identify an element as previously mentioned in discourse. Such a use is often an additional function of an element which is also a pure deictic, but sometimes there is a particular demonstrative which has assumed this as its basic function.” From the above discussion, it is clear that the anaphoric use, as well as the recognitional use, are better taken as additional functions, rather than basic functions, of *zhe* and *na*. More importantly, uses characteristic of fully grammaticalized definite articles, like those marking shared specific and general knowledge, and frame-based association, are exceptional rather than the norm with either *zhe* or *na*. It is even unclear, based on the text frequency studies of the uses of *zhe* and *na* as definite determiners, which of the two has developed more deictically neutral uses characteristic of a definite article. Lü (1990: 592) and S. Huang (1999) maintain that *na* has developed further than *zhe* toward a definite article, while Tao (1999) and Fang (2002) suggest *zhe* as the most likely candidate for definite article in Chinese. It is an issue open to further investigation.

Finally, the Chinese demonstratives, at the current stage of development, fail to fulfill what Greenberg (1978: 61) takes to be a crucial criterion for deciding when a discourse deictic has become a definite article: the criterion of being “compulsory” when definiteness is marked for a referent in discourse. It is argued in Xu (1987), for instance, that a zero NP is the unmarked form for noncontrastive anaphoric uses that is functionally equivalent to the English definite article, and *zhe* and *na* are marked forms for such uses.

In sum, while demonstratives in Chinese have developed some functions which are typically served by definite articles in languages like English, they are, generally speaking, still much closer to demonstratives than definite articles on the path of grammaticalization.
4.1.2. **Possessives.** Chinese possessives, which are formed by suffixing the possessive marker *de* to personal pronouns or nouns, do not have the same distribution and interpretation as those in English. In English, as we know, an indefinite article cannot come between a possessive and the head noun, and the possessive imposes a reading of definiteness on the nominal expression. A Chinese possessive, on the other hand, can be separated from the head noun by an indefinite marker, with an interpretation of nonidentifiability for the nominal expression, as illustrated in (58):

(58) Zhe shi wo-de yi ge pengyou gaosu wo de.

'This be my one CL friend tell I DE
A friend of mine told me this.'

In terms of the typological distinction between DG languages and AG languages drawn by C. Lyons (1999), as discussed earlier in this article, it appears that while English is a DG language, Chinese is an AG language.

Chinese possessives, however, cannot be counted as those of a typical AG language like Italian. In Italian, as discussed earlier, whether the NP with a possessive is definite or indefinite depends on the presence of a definite or an indefinite article. In Chinese, the possessive does impose a definiteness interpretation on the nominal expression when the head of the expression is a bare NP, which in itself is unmarked for identifiability or nonidentifiability. Consider the following example:

(59) Wo-de qianbi zenme zhao bu dao le?

'my pencil how find not arrive CRS
'How come I cannot find my pencil/*a pencil of mine?'

In other words, whether the nominal expression with a possessive and a bare noun is to be interpreted as identifiable or nonidentifiable in Chinese depends on the presence or absence of an indefinite determiner between the possessive and the bare noun. Since an indefinite determiner is perfectly grammatical in that position in Chinese, its absence strongly implicates an interpretation of identifiability on the nominal expression in a way that is theoretically captured in terms of Gricean conversational implicature (cf. Grice 1989; Levinson 2000).

4.1.3. **Universal quantifiers.** Also used as markers of definiteness in Chinese are universal quantifiers, which include collective universal quantifiers *suoyou* ‘all,’ *yiqie* ‘all,’ and distributive universal quantifiers *mei* ‘each, every,’ and *ge* ‘each, every.’ Consider the following examples:

(60) Yiqie yinsu dou dei kaolu jinqu.

'all factor all must consider in
'All factors have to be taken into consideration.'
Every student has his own room.

Interrogative words in Chinese like shenme ‘what,’ shei ‘who,’ and nei ‘which,’ in their stressed form, are used as what are called free-choice quantifiers, referring to any arbitrary member of a whole class. They are treated by some linguists as indefinite determiners (cf. Lü 1990 [1956]; Chao 1968: 324). There are close semantic connections, as noted by Haspelmath (1997: 12, 154), between the distributive universal quantifiers and the free-choice quantifiers, and free-choice quantifiers may diachronically evolve into universal quantifiers. The free-choice quantifiers in Chinese have the same semantics as the universal quantifiers in terms of scope behaviors. Consider the following sentences:

(62) Mei ge xuesheng dou yinsu kaolu jinqu.
    Every CL student all have to take all factors into consideration.

(63) Nei ge xuesheng dou yinsu kaolu jinqu.
    Every CL student all have to take all factors into consideration.

the NP liang men waiyu ‘two foreign languages’ in the two sentences is subject to the same scope ambiguity in terms of specificity.

Furthermore, yiqie in (60) and mei in (61) can be replaced by free-choice quantifiers, with no change in the meaning of the sentence:

(64) Shenme yinsu dou yinsu kaolu jinqu.
    ‘All factors have to be taken into consideration.’

(65) Nei ge xuesheng dou you ziji de fangjian.
    ‘Every student has his own room.’

Finally, the free-choice quantifiers in Chinese behave much more like universal quantifiers than indefinite determiners in that they usually occupy a preverbal position in sentence, and are not allowed in some indefiniteness-inclined sentential positions (more detail on this term later). Consider the following sentence:

(66) a. Ta shouyou de cai dou chang le yidianr.
    He all DE dish all PFV taste bit
    ‘He had a taste of all the dishes.’

b. Ta ge yang cai dou chang le yidianr.
    He every CL dish all PFV taste bit
    ‘He had a taste of every dish.’

c. Ta shenme cai dou cang le yidianr.
    He any dish all PFV taste bit
    ‘He had a taste of all the dishes/every dish.’
(67)  a. */?Ta chang le yidianr *shouyou de cai.
    b. */??Ta chang le yidianr *ge yang cai.
    c. *Ta cang le yidianr *shenme cai.

No such restriction is applicable to words like *shenme in stressed form as an interrogative, as in (68), or in unstressed form as an indefinite determiner as in (69):

(68) interrogative *shenme in stressed form
    Ta chang le yidianr *shenme cai?
    he taste PFV bit what dish
    ‘What dish/dishes did he have a taste of?’

(69) indefinite *shenme in unstressed form
    Rang ta chang yidianr *shenme cai  ba.
    let he taste bit sm dish SFP
    ‘Let him have a taste of some dish/dishes.’

It is thus justified both on the grounds of semantics and syntactical behaviors to treat the freedom-choice quantifiers in Chinese as definite determiners on a par with collective and distributive universal quantifiers.

4.1.4. **Indefinite determiners.** The most important indefinite determiner in Chinese is *yi ‘one’ + classifier. *yi on its own is a numeral, and can still be used in the same way as all the other numerals in Chinese. Unlike the other numerals, however, *yi + classifier has undergone the process of grammaticalization toward a marker of indefiniteness in much the same way as the English indefinite article was derived from the numeral ‘one.’ Other than the fact that *yi + classifier can be used both as a pronominal and as a determiner, it serves all the major functions of a regular indefinite article as the English *a, and moreover extends to other uses that have not been reported for indefinite determiners in English or other languages. It is found in uses characteristic of each of the five stages of development from a numeral to a grammatical indefiniteness marker as proposed by Givón (1981) and Heine (1997), namely, as a numeral, a presentative marker, a marker of nonidentifiable specific reference, a marker of nonidentifiable nonspecific reference, and what Heine (1997) calls a generalized article. They are illustrated respectively in the following examples:

(70) numeral:
    Wo zhi yao *yi zhi pingguo jiu gou le.
    I only want one CL apple then enough CRS
    ‘I only want one apple.’
(71) presentative marker:
Yì zhì xiǎo qī′e yaoyabaibai zou le shanglai.
one CL little penguin swaying walk PFV up
‘A little penguin was waddling up.’

(72) nonidentifiable specific reference:
Zhe jiàn shì wò zuòtiān qǐng le (yì) ge rén lái.
this CL issue I yesterday invite PFV one CL person come
‘For this issue I invited a person here yesterday.’

(73) nonidentifiable nonspecific reference:
Gāngkuān qu zhǎo (yì) ge rén lái, shènmé rén dōu xǐng.
Hurriedly go find one CL person come any person all fine
‘Hurry up and get somebody; anybody will be just fine.’

(74) generalized article:
Tā kàn shāngqu xiāng (yì) ge fāguòrén.
he look up like one CL Frenchman
‘He looks like a Frenchman.’

It is proposed in Heine (1997: 76) that the more stages an item has passed through on its way from numeral to indefinite article, the more it is affected by grammaticalization processes such as bleaching, cliticization, and phonetic erosion. The behaviors of yi + classifier in Chinese have provided another piece of evidence in support of the claim. While yi in (70) must be stressed, it is unstressed in other uses, and is in fact commonly omitted, leaving as it were the function of definiteness marking served by the classifier alone. In his seminal analysis of the uses of yi + classifier, Lü (1990 [1944]) has discussed in detail the conditions under which yi can be omitted. It is observed in his paper that, apart from prosodic constraints, yi is more likely to be omitted when used as a maker of nonidentifiable nonspecific reference, as in (73), than as a marker of nonidentifiable specific reference, as in (72) (cf. Lü 1990 [1944]: 167). A more inclusive, and theoretically more revealing account, I would propose, is that, other prosodic conditions as elaborated in Lü (1990 [1944]) being equal, the further down the grammaticalization continuum of the uses of the indefiniteness marker, the more weakened is the morphological and phonological weight of yi in yi + classifier. yi in (70) always appears in stressed form; yi in (71) cannot be omitted, but is usually unstressed; naturalness of the sentences with yi omitted increases from (72) through (73) to (74): while it is more natural to have yi in (72), (74) sounds much better without yi. The extent of the phonological reduction of the Chinese numeral yi correlates perfectly with the order of its development through the five stages along the continuum of grammaticalization.
What is more interesting is the fact that the Chinese indefinite determiner has developed uses that are not reported for indefinite articles in other languages. Heine (1997: 73) observes that at the final stage of development, the generalized article is no longer restricted to singular nouns, but is extended to plural and mass nouns as is the case in Spanish. At the same time, he asserts that it still may not be used with a noun marked for definiteness or a proper noun. This restriction is relaxed in Chinese.

In addition to the uses in (70) through (74), yi + classifier can also be used with identifiable referents such as proper names. Consider the following examples ([75] and [76] are quoted from Lü 1990 [1944]):

(75) An na meizi... taiju de ge zhangfu jun shang tian jun.
I that sister praise CSC CL husband smart up plus smart
‘My sister praised her husband in such a way that he looked even smarter.’

(76) Dangxia ba ge Zhang San Li Si xia de mudengkoudai.
instantly BA CL Zhang San Li Si scare CSC dumbstruck
‘Zhang San and Li Si were instantly struck dumb with fear.’

(77) Lao Zhang bei pengyou ba ge taitai gei pian zou le.
Old Zhang BEI friend BA CL wife by cheat go CRS
‘Lao Zhang was cheated by his friend and lost his wife.’

(78) Ta ba ge pibao gei diu le.
he BA CL hat lose CRS
‘He lost his bag.’

Unlike the cases of (70) through (74), the nominal expressions modified by the indefinite determiner in (75) through (78) are all of identifiable reference.

The function of (yi) + classifier in (75) through (78), as I have argued in detail in Chen (2003), is to de-individuate the referent it is used with to make it less likely to serve as a topic of continuing reference in ensuing discourse. It serves as a backgrounding device in a way that is opposite to the indefinite use of this in English. Originally and still primarily a definite demonstrative, this in the new usage as an indefinite determiner to introduce new referents into discourse serves to mark a nonidentifiable referent as of high thematic importance. yi + classifier in Chinese, on the other hand, operates in the opposite direction. As an indefiniteness determiner, it has developed uses with an identifiable referent, otherwise of high inherent saliency as is the case with the proper names in (76) and with “wife” in (77), to mark it as of low thematic importance.
In the final stage of the development, as discussed earlier, the indefinite determiner $yi +$ classifier is used with nominal expressions whose referentiality is completely bleached out, leaving us with only genericity and connotation, as is the case with the predicative nominal in (74). Other nominal expressions used with this generalized indefinite determiner, through the process of analogy, may acquire some of the features that characterize the nonreferential nominal expressions that are commonly used with the determiner at this stage of development, with their referentiality in effect bleached out to a certain extent. In other words, the use of the indefinite determiner at its final stage of development is further extended to definite expressions to induce sense of nonreferentiality to identifiable entities, which is in turn suggestive of lower thematic importance (cf. Chen 2003 for a more detailed discussion).

In terms of the typical functions representing the five major stages of grammaticalization of the numeral ‘one’ into an indefinite article, we conclude that the numeral $yi$ in Chinese has undergone the full process of grammaticalization, and has developed into an $a$-like indefinite article. The conclusion, however, has to be qualified. Unlike the indefinite article in English, as noted earlier, $yi$ has both an attributive and a pronominal use, which means it has retained a higher degree of morphological autonomy than a fully grammaticalized indefinite article. If we follow the assumption in Heine (1997: 71) that an indefinite article cannot be used as a pronoun (e.g. *I want $a$ vs. I want $one$), we cannot accept without any reservations the claim that there is an $a$-like indefinite article in Chinese. As will be discussed later in the article, moreover, $yi$ differs from $a$ in another more important respect: while it is compulsory for a nonidentifiable referent in English to be marked by $a$ or another device of indefiniteness, it is not the case in Chinese.

Other markers of indefiniteness in Chinese include interrogative-turned indefinite determiners like *shenme ‘some/any,’ and *shei ‘someone/anyone,’ which are always in weakened phonological form (cf. Chao 1968: 651–657). Consider the following examples:

(79) Ta meiyou shuo shenme yaojin de shir.
    He not:have say any important DE thing
    ‘He did not say anything important.’

(80) Na zhi beizi bei shei da po le.
    that CL cup BEI somebody hit break PFV
    ‘That cup was broken by somebody.’

What are sometimes called mid-scalar quantifiers like $few$, $several$, and $many$ are often grouped with indefinite pronouns in many languages (cf. Haspelmath 1997: 11–12). They do not constitute a single category in
Chinese. While yixie ‘several’ is an indefinite determiner, ji ‘several,’ as will be discussed shortly, should be treated as a modifier of cardinality expression on a par with numerals like liang ‘two’ and san ‘three,’ which may have an identifiability reading in certain sentential positions when not accompanied by other definite or indefinite determiners.

4.2. Morphological

Monosyllabic classifiers in Chinese, and occasionally monosyllabic nouns as well, may undergo the morphological process of reduplication to gain the same meaning as that of distributive universal quantifiers, as illustrated in the following examples:

(81) Zhong-zhong yinsu dou dei kaolu jinqu.
CL factors all must consider in
‘Every factor has to be taken into consideration.’

(82) Ren-ren dou shuo ta shi ge hao ren.
person all say he be CL nice person
‘Everybody says he is a nice person.’

The reduplicated classifier or noun gives rise to the same scope ambiguity with regard to quantified NPs in sentences as universal quantifiers, which is illustrated in (62) and (63); it also displays the same syntactic behaviors as universal quantifiers in that they are not allowed in the indefiniteness inclined positions like that in (67).

(83) *Ta chang le yidianr yang-yang cai.
he taste PFV bit CL dish
‘He had a taste of every dish.’

4.3. Indeterminate expressions

The lexical and morphological markings of NPs in Chinese we have discussed so far fall into two groups, definite and indefinite. NPs which are lexically or morphologically encoded as definite or indefinite, on the assumption that they are used referentially, are always interpreted as of either identifiable or nonidentifiable reference in utterance, no matter what position they occupy in sentences. The encodings are determinate in relation to the interpretation of identifiability.

There are other types of NPs which, so far as their lexical or morphological encodings are concerned, are neutral with respect to the interpretation of identifiability. One is the bare NP, by which I mean an NP that
is not marked by any of the definite or indefinite determiners we have discussed above, and has not undergone the morphological process of reduplication either. The other is the cardinality expression, a term borrowed from C. Lyons (1999), which is an NP modified by a cardinal numeral or a quantifier like ji. They constitute what I call the indeterminate encodings of NP with respect to the interpretation of identifiability.

In most, but not all, instances, whether the indeterminate expressions are to be interpreted as identifiable or nonidentifiable is indicated, or suggested, by the position of the noun phrase in sentences. It has been a well-known statement in studies of Chinese grammar that a bare noun in subject position, like keren ‘guest’ in the classic example (84a), is identifiable, and one in postverbal position, in keren in (84b), is nonidentifiable.

(84) a. Keren lai le.
   guest come PFV
   ‘The guest has arrived.’

b. Lai keren le.
   ‘A guest/guests has/have arrived.’

The statement, as I will show shortly, needs qualification and modification to account for the full range of relevant phenomena in Chinese.

It has been elaborated in this article that definite expressions derive their identifiability from a variety of sources, and there is some correlation, and overlapping, between the particular types of definite determiners and the ways in which the identification of the referents is established. For instance, while both the definite article and the demonstratives in English are used for simple anaphoric reference, the former, but not the latter, is used for referents of frame-based association. When interpreted as of identifiable reference, indeterminate expressions in Chinese are used for referents which derive identifiability from the whole range of sources that is covered by the English definite article and demonstratives, as illustrated in the following sentences:

(85) bare noun: situational
   Xiaoxin, bie ba haizi chaoxing le.
   careful don’t BA baby wake CRS
   ‘Be careful. Don’t wake up the baby.’

(86) cardinality expression: situational
   Xiaoxin, bie ba liang ge haizi chaoxing le.
   careful don’t BA two CL baby wake CRS
   ‘Be careful. Don’t wake up the two babies.’
(87) bare noun and cardinality expressions: anaphoric
Ta yang le yi zhi maor he ji tiao gou, maor you fei
He raise PFV one CL cat and several CL dog cat both fat
you da, ji tiao gou que shou de pi bao gu.
and big several CL dog but thin CSC skin cover bone
‘He has a cat and several dogs. The cat is big and fat, but the
dogs are very skinny.’

(88) bare noun: shared specific knowledge
Ganggang jie dao tongzhi, laoban xiawu lai jiancha
just receive arrive notice boss afternoon come inspect
gongzuo.
work
‘The notice has just come: the boss is coming for an inspection in
the afternoon.’

(89) cardinality expression: shared specific knowledge
Ba wu shangchuangzi dou guan shang.
BA five CL window all shut up
‘Shut all the five windows.’

(90) bare noun: shared general knowledge
Yueliang sheng shang lai le.
moon rise up come CRS
‘The moon has risen.’

(91) cardinality expression: shared general knowledge
Zai nar, si ge jijie dou you gezi de tezheng.
in there four CL season all have self DE characteristics
‘Each of the four seasons there has its own charateristics.’

(92) bare noun and cardinality expression: frame-based association
Ta mai le yi liang jiu che, (si zhi) luntai dou mo ping
he buy PFV one CL old car four CL tire all wear flat
le.
CRS
‘He bought a used car. (The four tires) the tires are all worn out.

(93) bare noun and cardinality expression: self-containing association
Ni renshe zuotian lai de (san ge) ren ma?
you know yesterday come DE three CL person Q
‘Do you know the (three) people who came yesterday?’

Indeterminate expressions on nonidentifiable reading also serve the major
functions that are served by indefinite determiners like (yi) + classifier
exemplified in (72) through (74). Following are some examples:
(94) bare noun and cardinality expression: nonidentifiable specific
Wo jia li lai le (ji wei) keren.
I home in come PFV several CL guest
'I have had (several visitors) a visitor/visitors at home.'

(95) bare noun: nonidentifiable nonspecific
Women dei zhao ren bangmang, shenme ren dou xing.
we must find person help any person all fine
'We must get somebody to help us; anybody will be just fine.'

(96) cardinality expression: nonidentifiable nonspecific
Women dei zhao san ge ren bangmang, shenme ren
dou xing.
all fine
'We must get three people to help us; anybody will be just fine.'

(97) bare noun and cardinality expression: generic/connotation
Li Ming he Zhang Hong kan shangqu xiang (liang ge)
Li Ming and Zhang Hong look up like two CL
faguoren.
Frenchman
'Li Ming and Zhang Hong look like Frenchmen.'

Factors which govern the choice between the determinate and the inde-
terminate encoding for identifiable or nonidentifiable referents will be
addressed in a separate paper.

4.4. Positional

In addition to the lexical and morphological encodings, position in sen-
tences also serves as an important device to indicate identifiability or
nonidentifiability of nominal expressions in Chinese. It is extensively at-
tested in both languages with definite articles, like English and languages
without articles, like Czech and Russian, that most sentential positions,
or semantic roles which typically occupy particular syntactic positions in
sentence, differ in their preference for nominal expressions characterized
by various degrees of identifiability. Subject as well as other preverbal
positions in an SVO language usually display a strong inclination for
definite expressions; some postverbal positions display a inclination for
indefinite expressions (cf. Keenan and Comrie 1977; Chafe 1980; Givón
Thompson 1997; inter alia). I call the former definiteness-inclined posi-
tions, and the latter definiteness-inclined positions. There are also
positions which show relatively weak inclination one way or the other. The inclination may be grammaticalized in a way that expressions of the opposite encodings in terms of the definite versus indefinite dichotomy are simply ungrammatical in the syntactic slot. In the overwhelming majority of cases, however, instead of being in the form of grammatical restrictions, the inclination is manifested in terms of higher text frequencies of the nominal expressions of definite versus indefinite encodings occupying the particular position in sentences.

The major types of definite or indefinite expressions discussed earlier also differ with regard to the degree of identifiability of reference, which may relate to their grammaticality or naturalness in some positions of sentences. Of proper names, personal pronouns, and NPs modified by definite determiners, for example, the referent of a personal pronoun normally registers a higher degree of identifiability than that of the other two types (cf. Ariel 1990). The strength of inclination of sentential position for definite or indefinite expressions correlates positively with the degree of identifiability of the referent. I will come to this point shortly.

The discourse pragmatic notion of topic, as demonstrated in the studies by Hopper and Thompson (1980, 1984), Givón (1983, 1984/1990), Tomlin (1983), Van Oosten (1986), Du Bois (1987), and Thompson (1997), stands as a most important factor underlying the correlation between the definiteness inclination of syntactic positions (and semantic roles that typically assume the respective positions) and the identifiability, as well as referentiality and activation status, of nominal expressions that occupy those positions. I am adopting the notion of topic as defined by Lambrecht (1994):

(98) A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee’s knowledge of this referent. (Lambrecht 1994: 131)

The topic expression of a sentence, furthermore, typically contains information that bears close relevance to the preceding discourse. It is more likely, although by no means an absolute certainty, for the subject expression of the sentence than other sentential components to serve at the same time as the topic of the utterance. A highly identifiable referent, which is by definition referential, is more likely to serve as the topic of the utterance than a referent with a lower degree of identifiability. Hence the high correlation between subject and identifiability of reference.

It has been established in the literature that the following sentential positions display an inclination for definiteness or indefiniteness in

(99) Definiteness-inclined positions in Chinese:

- subject
- object
- preverbal object
- first object of ditransitive sentence

(100) Indefiniteness-inclined positions in Chinese:

- object of the presentative verb you
- postverbal NP in presentative sentences
- postverbal NP in existential sentences
- second object of ditransitive sentences

After subject, object is the next most likely topic expression in a Chinese sentence, which is characterized as “secondary topic” by some writers. Nominal expressions in the other two positions in (99) also display features of high topicality (cf. Tsao 1990; Zhang and Fang 1996; inter alia). The primary function of the four positions in (100), on the other hand, is to introduce new and nonidentifiable referents into discourse, which are highly unlikely to serve as the topic of utterance.

Now let us consider how the sentential positions correlate with definite expressions, indefinite expressions, and expressions of indeterminate encodings in Chinese with regard to the interpretation of identifiability of reference.

4.4.1. Positions and determinate expressions. Nominal expressions which are lexically or morphologically encoded as definite or indefinite may be subject to severe restrictions in their eligibility to occur in some of the positions listed in (99) and (100). It is ungrammatical to have a reduplicated classifier or noun in any of the definiteness-inclined positions:

(101) a. *You ren-ren lai le.  
    have person come CRS

b. *Qianmain lai le ren-ren.  
    front come PFV person

c. *Qianmian zuo ge-ge daibiao.  
    front sit CL representative

d. *Women gei le ta yang-yang liwu.  
    we give PFV he CL present

Indefinite expressions can hardly occur as subject with stative predicate, as illustrated by (102a). Explanation for this phenomenon in Chinese and
other languages is captured in terms of the thetic vs. categorical distinction in Sasse (1987).

(102) a. *Yi ge ren hen shanchang dalie.
   one CL person very good:at hunt
b. You yi ge ren hen shanchang dalie.
   have one CL person very good:at hunt
   ‘There was someone who was a very good hunter.’

In the great majority of cases, as said earlier, the inclination is manifested, not in terms of grammatical restriction, but in terms of higher frequency of expressions of one category in contrast with the other in texts. As amply demonstrated in statistical studies of Chinese sentences and discourse, subject, ba object, and the other definiteness-inclined positions are overwhelmingly definite, and the indefiniteness-inclined positions are overwhelmingly indefinite. Similar findings are reported in other languages as well (cf. Givón 1983, 1984/1990).

The text frequency findings of my studies, which will be presented below, demonstrate that nominal expressions of definite or indefinite encodings may also occur in some sentential positions of the opposite inclination, with frequency in texts varying from relatively low to extremely low. It is reported by Li (1986) that definite expressions as the postverbal NP in presentative and existential sentences are attested in his corpus of data, although such occurrences are very rare in text counts. (103), (105), and (106) are quoted from Li (1986):

(103) You zuo men zou jin le Hu Si he Gu Ba Nainai. from left door walk in PFV Hu Si and Gu Ba Nainai
   ‘From the left door walk in Hu Si and Gu Ba Nainai.’
(104) Pingmo shang jianjian chuxian le Wang Gang de miankong. screen on gradually appear PFV Wang Gang DE face
   ‘Wang Gang’s face gradually took shape on the screen.’
(105) Ta . . . sheng pang jiu tang zhe Chen Changjie. he body side then lie DUR Chen Changjie
   ‘Beside him lies Chen Changjie.’
(106) Ta huiguo sheng qu, yan qian jiu zhan zhe ta.¹³ she turn body go eye front then stand DUR he
   ‘She turned around, and he stood right in front of her.’

Furthermore, indefinite expressions are found in use as ba object, as in (107), although much less frequently than definite expressions (cf. Lü (1990 [1948]; Zhang 2000).
On the other hand, while statistically many more definite and indeterminate expressions occupy subject position, indefinite expressions serving as subject with dynamic predicate as in (71), repeated below as (108), are actually quite common in Chinese, so long as certain conditions are met (cf. Fan 1985; Sasse 1987).14

(108) Yi\(\text{ CL}\)\(\text{ zhi}\)\(\text{ CL}\)\(\text{ xiao}\)\(\text{ CL}\)\(\text{ qi’e}\)\(\text{ CL}\)\(\text{ yaoyaobaibai}\)\(\text{ CL}\)\(\text{ zou}\)\(\text{ PFV}\)\(\text{ le}\)\(\text{ CL}\)\(\text{ shanglai}\).

‘A little penguin was waddling up.’

The subject in (108) presents a nonidentifiable referent that is introduced for the first time into discourse. The grammaticality, and fairly common occurrence of sentences like (108) clearly demonstrate that, as expected, there is no one-to-one correlation between the discourse pragmatic role of topic and the syntactic position of subject, just as there is no one-to-one correlation between semantic roles and syntactic roles. There are many competing pragmatic, semantic, and syntactic factors at work which may result in the separation of subject and topic in utterances.

Other positions in sentences, like postverbal object in transitive sentences and oblique object, normally admit determinate expressions of either category, as well as indeterminate expressions. Text counts may reveal a reference for one category of expressions rather than another, but not as strong and obvious as is the case with the positions in (99) and (100).

4.4.2. Positions and indeterminate expressions. Expressions which are lexically and morphologically indeterminate with regard to identifiability are not subject to the same kind of restrictions as expressions of determinate encodings. Generally speaking, they occur freely in positions of either inclination, as well as in other positions. At the same time, they display a strong inclination to be interpreted as identifiable in definiteness-inclined positions, and as nonidentifiable in indefiniteness-inclined positions. The strength of the inclination varies with the types of indeterminate encoding and the sentential positions that are occupied by the expressions.

I have examined 24 narrative stories in Chinese Fables in my corpus to ascertain how the indeterminate encodings and the interpretation in respect of identifiability correlate. I have chosen for the purpose two most
prominent definiteness-inclined positions, subject and ba object, and two most prominent indefiniteness-inclined positions, which are object of the presentative verb you and the postverbal NP in presentative sentences. Postverbal object in transitive sentences is also included as representing positions not belonging to either category. The outcomes are presented in Tables 2 and 3.

It is evident from Table 2 and Table 3 that the great majority of the expressions of indeterminate encodings in the position of subject and as ba object are interpreted as identifiable. The percentage of identifiability reading is even higher for those in the subject position than as ba object. In this limited corpus of data, all the indeterminate expressions in the two indefiniteness-inclined positions, namely, object of presentative verb you and postverbal NP in presentative sentences, are of nonidentifiable reference, although it is attested in Li (1986), as discussed above, that they can be of identifiable reference. Identifiable and nonidentifiable referents encoded as bare NP are fairly equally divided in the position of postverbal object in transitive sentences, while the ratio is about one to three for cardinality expressions in the same position. The typical uses and interpretations of indeterminate expressions in the relevant positions are illustrated by the following examples:

(109) Qianyuan da shu xia ji tian qian pa man le mayi, front:yard big tree down several day ago creep full PFV ant dao zuotian, mayi dou pa zou le, di shang zhi jian to yesterday ant all creep away CRS ground on only see
yi gen niao mao.
several CL bird feather
‘Several days ago there were ants all over the bottom of the big
tree in the front yard. Yesterday, the ants all disappeared, and
there were only a few bird feathers on the ground.’

(110) Zhangfu zhang lai le ji wei niwangong. Ji wei
husband find come PFV several CL tiler several CL
niwangong yong le yi tian shijian ba fangding xiu hao le
tiler use PFV one day time BA roof fix fine CRS
‘The husband hired several tilers. The tilers took a day to fix
the roof.’

The referent mayi ‘ant’ in (109) assumes the same lexical form, as a bare
NP, on its two occurrences in the sentence; and so does the referent
niwangong ‘tiler’ in (110), as a cardinality expression. In each case, the
nominal expression on the two occurrences differs in its position in the
sentence. The bare NP mayi in (109) appears as postverbal NP in the pre-
sentative sentence on the first occurrence, interpreted as a nonidentifiable
referent; and, as subject on the second occurrence, the expression is in
anaphoric use, referring to an identifiable referent. The cardinality ex-
pression ji wei niwangong on the first occurrence as the object of the tran-
sitive verb refers to a new referent that is introduced into the discourse for
the first time, and on the second occurrence, in subject position it is in-
terpreted as a fully identifiable referent coreferential with the previously
introduced referent.

While the identifiability or nonidentifiability of reference of determi-
nate expressions is unambiguously indicated by the lexical and mor-
phological encodings of expressions irrespective of their positioning in
sentences, the interpretation of indeterminate expressions in respect of
identifiability is inferred by the addressee, or is expected by the speaker
to be inferred by the addressee, based on his assessment of how likely the
expression in question is meant to serve as the topic of the utterance as
well as on his whole range of knowledge of the utterance and its context
of use. When he hears an utterance like keren lai le ‘The guest/guests
has/have arrived,’ he will start searching in the universe of discourse for a
referent that meets the description of the NP keren, and which is sup-
posed to be already identifiable for him: he knows that it is highly likely
that the expression, because of its position as subject, is meant to be the
topic of the utterance, and as such, normally has to be assumed by the
speaker to be identifiable for him. In an overwhelming number of cases,
he will easily locate that referent, in spite of the fact that the encoding
itself is less informative in comparison with encodings marked by definite
determiners like demonstratives. On the other hand, with an utterance like
\textit{lai keren le} ‘A guest/guests has/have arrived; here arrives/arrive a guest/
some guests,’ instead of searching for an identifiable referent fitting the
description, most likely he will simply establish a new file in his mental
representation of the event for a referent that is not supposed to have
been known to him: he knows that position in sentence is normally used
for introducing new referents into discourse.

Since identifiability or nonidentifiability is encoded unequivocally and
irrevocably in the case of determinate encoding, communication breaks
down when the addressee cannot find an identifiable referent in the uni-
verson of discourse that meets the description of the definite expression.
There is more flexibility with indeterminate expressions in definiteness-
inclined positions which are normally expected to be interpreted as iden-
tifiable. As we have discussed earlier, the possibility is great that the sub-
ject expression is also the topic of the utterance, but it is not an absolute
certainty. Subject can also serve as the position in which new referents,
encoded in indefinite terms, are introduced into discourse, as illustrated
by (108). Given that possibility, when the addressee cannot find a suitable
identifiable referent satisfying the description of the indeterminate ex-
pression in the subject position, he will most likely read it in the same way
as he interprets the indefinite subject in (108), treating it as a new referent
making its first appearance in discourse, particularly when some other
conditions are met (cf. Sasse 1987; Lambrecht 1994).\textsuperscript{15}

In comparison, cardinality expressions in definiteness-inclined positions
are even more likely to receive a nonidentifiability reading than bare NPs,
presumably due to the fact that the descriptive content of numerals and
quantifiers makes the expression more compatible than bare NP with a
presentative use. Consider the following sentences:

\begin{enumerate}
  \item (111) \textit{Wuyun piao le guolai, jianjian ba yueliang zhezhu le.}
dark:cloud float PFV come gradually BA moon blot:out CRS
  ‘Dark clouds float over, and gradually blot out the moon.’
  \item (112) \textit{Da qun de lu tingdao tonglei de zhaohuan, beng zhe
big group DE deer hear same:kind DE call leap DUR
tiao zhe lai dao lieren shengbian.}
lump DUR come to hunter side
  ‘Deer in large numbers ran vivaciously to the hunter when they
heard the calls from other deer.’
  \item (113) \textit{Liang ge qiangdao cong zheli zou guo, kaodao zhouwei
two CL bandit from here walk by see around
meiyou ren, zhiyou Ou Ji zhenme yi ge xiaohaizi, jiu
have:not person only Ou Ji this one CL kid then
\end{enumerate}
‘Two bandits passed by, saw nobody around except the kid Ou Ji, and then had an evil idea . . . . . . The two bandits tied Ou Ji’s hands behind his back.’

Both (111) and (112) are felicitous when the referents of wuyun ‘dark clouds’ and da qun de lu ‘deer in large numbers’ appear for the first time in discourse. (113), which is quoted from a Chinese fable in my corpus of data, offers another illustrative example. The cardinality expression liang ge qiangdao ‘two bandits’ in the two clauses has exactly the same lexical encoding, and occupies the same syntactic position; it is nonidentifiable on the first occurrence, and identifiable on the second.

Finally, there are situations where the same indeterminate expression in an utterance can be interpreted either as an identifiable referent or a nonidentifiable referent. Consider the following as the opening sentence for a conversation:

(114) Zuowan lianyouchang da huo, tie men dou shao hua le. last:night refinery big huo iron gate even burn melt CRS ‘There was a big fire at the refinery last night. The/an iron gate/gates melted in the fire.’

(115) Xiao Li de bing hen zhong, yiyuan de san wei xinzhangbing zhuanjia shuo de shoushu. Xiao Li DE illness very serious hospital DE three CL xinzhangbing zhuanjia heard:ailment expert say immediately move operation ‘Xiao Li is seriously ill. (The) three cardiologists at the hospital said that he needed to undergo an operation immediately.’

The bare noun tie men ‘iron gate’ in (114) can refer to an identifiable referent, in the situation where there is only one iron gate at the refinery that is supposed to be known to the addressee as part of their shared background knowledge or as a frame triggered entity. The sentence is also appropriate in the context in which the addressee is not assumed to know, and probably the speaker himself does not know either, how many iron gates there were to the refinery and which one or ones were melted in the fire. All the speaker intends to tell the addressee is that the fire was so destructive that it melted one or more iron gates. In the latter case, the expression obviously refers to a nonidentifiable referent. The same applies to (115). The referent of san wei xinzhangbing zhuanjia ‘three cardiologists’ can be three out of many cardiologists at the hospital, or it can be
the three cardiologists that the hospital has, or the three doctors that the 
speaker and the addressee have mentioned earlier in discourse. It is out of 
the question to interpret the indeterminate expressions in (114) and (115) 
as in generic use.

By putting indeterminate expressions in sentence positions of two distin-
tinctive categories in respect of inclination for definite or indefinite ex-
pressions, the speaker indicates, or suggests, to the addressee whether the 
expressions are meant to be interpreted as identifiable or nonidentifiable. 
While it works in the majority of cases, it is to be stressed that, unlike 
the lexical and morphological encodings, position in sentences does not 
function as an unambiguous, watertight encoding device for marking 
identifiability or nonidentifiability of reference for nominal expressions in 
Chinese, except in one or two syntactic slots, such as the postverbal NP in 
presentative sentences. As far as determinate expressions are concerned, 
most sentential positions in Chinese, as in many other languages, favor 
definite or indefinite expressions, as the case may be, but seldom reject 
categorically expressions of the other category. With regard to indeter-
minate expressions, most sentential positions invite, some more strongly 
than others, but seldom impose categorically, a reading of identifiability 
or nonidentifiability. There is simply no sentential position, or specific 
word order in Chinese whose core or primary function is to mark the 
distinction of the constituent in terms of identifiability.

It is appropriate in this connection to make a comparison between 
Chinese and the other four languages discussed earlier with respect to the 
use of zero NP as a coding device for identifiability. Count nouns in the 
plural and mass nouns with zero determiner in English, as observed ear-
lier in this article, are always interpreted as nonidentifiable, standing in a 
paradigmatic contrast with the same nouns preceded by the definite arti-
cle. Bare NPs in Czech are interpreted as definite or indefinite, depending 
on their position in sentences. Those in Hindi and Indonesian are either 
definite or for generic use. The four languages all share the feature that 
the interpretation of zero NPs with regard to identifiability of the refer-
ents is unambiguously determined. In contrast, the interpretation of zero 
NPs in Chinese with respect to identifiability correlates closely, but im-
perfectly, with their grammatical position in sentences; it is sometimes 
ambiguous or indeterminate.

5. Concluding remarks

In conclusion, we now summarize our answer to the first two ques-
tions raised at the beginning of this article: (i) how is identifiability
encoded in Chinese? and (ii) how is Chinese in this respect similar with, or different from, the other languages that are briefly discussed in this article?

The pragmatic distinction between identifiability and nonidentifiability is expressed in Chinese in terms of distinctive lexical and morphological encodings and in terms of the positioning of nominal expressions in sentences. Of the major definite determiners in Chinese, demonstratives are developing functions, more advanced in spoken Chinese than in written Chinese, which are served by the definite article in English in marking referents whose identifiability is established through shared general knowledge, and in anaphoric and associative uses, although they have still preserved their deictic force to a considerable extent in these situations. The most important marker of indefiniteness in Chinese is yi + classifier, which has developed all the major functions, and more, that are served by indefinite articles in languages like English. In terms of morphology, and also in terms of function in the case of demonstratives, the definite and indefinite determiners in Chinese have not yet acquired the full status of specialized grammatical marker of definiteness and indefiniteness like the articles in English. Possessives in Chinese display features characteristic of AD languages like Italian, rather than DG languages like English, although they at the same time have their own features not shared by those of typical AD languages. Monosyllabic classifiers and nouns in Chinese undergo reduplication to gain the additional meaning as that of universal quantifiers, which in Chinese are taken as definiteness markers on the grounds of their semantics and syntactic behaviors.

What makes Chinese further different from languages like English is the existence of what I call indeterminate expressions, which consist of bare NPs and cardinality expressions, and the correlation between the interpretation of the expressions in respect of identifiability and their positioning in sentences. As far as their lexical and morphological encodings are concerned, the indeterminate expressions are neutral with respect to the interpretation of identifiability. Whether the speaker intends them to be interpreted by the addressee as identifiable or nonidentifiable is indicated, or suggested, by their positioning in sentences. Some sentential positions in Chinese display a strong inclination for definite expressions, and some for indefinite expressions. They are called definiteness-inclined and indefiniteness-inclined positions, respectively. With regard to indeterminate expressions, there is a strong, but not absolute, correlation between the interpretation in respect of identifiability of reference and the inclination of the position occupied by the expressions in sentences. While the pragmatic status of identifiability is unequivocally and irrevocably
expressed for nominal expressions in referential use when they are en-
coded in determinate lexical and morphological terms, the interpretation
of indeterminate expressions with respect to identifiability is expected by
the speaker to be inferred by the addressee, on the basis of the topicality
of the position occupied by the expression, the availability of an identifi-
able referent in the context that meets the descriptive content of the ex-
pression, as well as other relevant information of the utterance in the
universe of discourse.

Finally, we turn to the last question: is it justified to assert that defi-
niteness as a grammatical category, in the narrow sense of the term,
exists in Chinese? English presents an exemplary case for the existence of
definiteness as a grammatical category, as the language displays a para-
digmatic contrast between two simple, fully grammaticalized articles, one
definite and the other indefinite. A language which only has complex
definite markers or other grammatical devices to encode the distinction
between identifiability and nonidentifiability may also be treated, as dis-
cussed earlier, as having definiteness as a grammatical category if definite-
ness is obligatorily and unambiguously marked in that language. Czech,
Hindi, and Indonesian, based on the findings by Porterfield and Srivastav
(1988) and Cummins (1998), meet this criterion, since the features of
identifiability and nonidentifiability are obligatorily and unambiguously
encoded in those languages in terms of a variety of linguistic devices in-
cluding word order.

Chinese, however, represents a case that does not accord very well with
what has been established for these languages in the current literature. To
begin with, there is no simple, fully grammaticalized marker of definite-
ness in Chinese, like the definite article in English. In spite of the fact that
demonstratives in Chinese have developed some uses that are normally
served by definite articles in other languages, their basic or primary func-
tions are still far from being those of deictically neutral determiners of
definiteness like English the. While the Chinese numeral yi ‘one’ has ar-
guably reached the endpoint of grammaticalization into an indefinite
article, there is no paradigmatic contrast between it and a highly gram-
maticalized marker of definiteness. Furthermore, it is not obligatory to
mark a nominal expression as either definite or indefinite in Chinese, as is
the case in English. To the extent that situations are abundant in Chinese
in which the interpretation of bare NPs and cardinality expressions with
respect to identifiability cannot be determined solely in terms of their
position in sentences, and may even be ambiguous or indeterminate with
regard to identifiability, the features of definiteness and indefiniteness
cannot be uniquely and unambiguously specified for nominal expressions
in Chinese. This leads to the conclusion that definiteness as a grammatical
category, as defined in the narrow sense of the term, has not been fully developed in Chinese.

Received 13 August 2002
Revised version received 5 June 2003

Notes

* I am grateful to Sandra A. Thompson and two anonymous reviewers of Linguistics for their valuable comments, criticisms, and suggestions on an earlier draft of this article.

Following are the abbreviations used in this article. For a detailed explanation of the terms, cf. Li and Thompson (1981).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>preverbal object marker (ba)</td>
</tr>
<tr>
<td>DE</td>
<td>nominalizer (de)</td>
</tr>
<tr>
<td>CL</td>
<td>classifier</td>
</tr>
<tr>
<td>CRS</td>
<td>Current Relevant State (le)</td>
</tr>
<tr>
<td>CSC</td>
<td>complex stative construction (de)</td>
</tr>
<tr>
<td>DUR</td>
<td>durative aspect marker (zhe)</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective aspect marker (le)</td>
</tr>
<tr>
<td>SFP</td>
<td>sentence-final particle</td>
</tr>
</tbody>
</table>

Correspondence address: Department of Asian Languages and Studies, University of Queensland, Brisbane, QLD 4072, Australia. E-mail: p.chen@mailbox.uq.edu.au.

1. To illustrate this point, let us consider the following sentences ([i] and [ii] are from Givón 1984: 120):

   (i) There was once a unicorn. The unicorn loved lettuce.
   (ii) There was once a rabbit. The rabbit loved lettuce.
   (iii) There was once a rabbit under this tree. The rabbit loved lettuce.

   Unicorns do not exist in the real world; the existence of ‘rabbit’ in (ii) is established in linguistic terms, and in (iii), in both physical and linguistic terms. Yet, the unicorn in (i) and the rabbit in (ii) and (iii) display exactly the same set of features in terms of both form and function.

2. It has been argued that there are semantic and pragmatic components other than identifiability to the grammatical notion of definiteness, the most prominent of which is inclusiveness (cf. Hawkins 1978). C. Lyons (1999: 278) notes that in languages like Chinese and Korean which do not have an explicit definiteness marker such as the English definite article, what is customarily referred to as definiteness is usually an element of discourse organization, relating to whether the referents are familiar or already established in the discourse. It is obviously identifiability rather than inclusiveness that is involved here. Furthermore, C. Lyons argues that demonstratives, when used as markers of definiteness, indicate identifiability, not inclusiveness. It is also suggested by Epstein (1993) on the basis of data from Old French, that some definite articles may also serve the function of marking referents that are nonidentifiable but prominent. I leave the issue open whether and how the notion of inclusiveness, or other relevant notions, also plays a role here.

3. To give an example illustrating the diversity of the usage of the terms in question, let us consider the following sentence:

   (i) A friend has just sent me a lovely Valentine card.
J. Lyons (1977: 189) maintains that the subject NP is subject to both specific and non-specific interpretation. However, it is argued in Krifka et al. (1995: 16) that this NP, as the subject of a particular predication, can only make specific reference.

4. The nominal expression in question is in predicative use. It can also be interpreted as being in equative use in this sentence, in which case it is referential and identifiable.

5. The relativity of the identifiability of referents often manifests itself in the grammatical structure of language. It is illustrated by Prince (1981: 476) in the following sentences:

   (i) A friend of yours bought a Toyota.
   (ii) A friend of Steve’s bought a Toyota.
   (iii) A friend of my neighbors bought a Toyota.
   (iv) A friend of a guy I know bought a Toyota.
   (v) ?A friend of a guy’s bought a Toyota.

There is a scale of identifiability for the referent ‘a friend,’ which correlates with the identifiability of the anchor in the postnominal phrase. The referent becomes less and less identifiable as the identifiability of the anchor decreases. As shown in the sentences above, the subject position of the English sentence, while admitting indefinite expressions, displays a clear inclination to have a NP at the higher end of the scale of identifiability. For similar observations, cf. Givón (1984/1990: 431) and Lambrecht (1994: 85–86).

6. Note in this connection the following quote from J. Lyons (1977: 188): “It is a characteristic feature of the grammar of English that common nouns in the singular (except when they are used as mass nouns) must be introduced with an article (whether definite or indefinite), a demonstrative adjective, or some other determiner. Not all languages that have what might be described a definite or indefinite article are like English in this respect.” It would be interesting to investigate how referents like ‘the living-room wall’ in (13) are encoded in those languages with respect to definiteness.

7. Diessel (1999: 5) has identified a fourth use, called identificational, as exemplified in the following sentence:

   (i) C’est Pascal.
       This/it is Pascal
       ‘It/this is Pascal.’

8. Note the observation by Clark and Marshall (1981: 44) that the use of that in the sentence I met a woman yesterday; that woman was a doctor “attracts contrastive stress and implies that there is a contrasting set of women.”


11. The uses to be discussed here have little to do with the use of a in A Mr. Smith came to see you this morning, where a means ‘a certain,’ indicating that the referent is non-identifiable for the addressee. yi + classifier can be used in a similar way in Chinese, which need not concern us here.

12. Chinese linguists often disagree over the definition and application of the terms for sentential components as listed in (99) and (100). In this article I am using these terms
in a way that is, to the extent possible, fairly uncontroversial and compatible with different theoretical frameworks. For want of space to do full justice to the issues involved, I have left aside in this article the notion of ‘topic’ as a syntactic component (cf. Chen 1994, 1996).

13. While the postverbal NP in this existential sentence is a personal pronoun, no case has been reported in Li (1986) where a personal pronoun occurs in the postverbal slot in a presentative sentence.

14. A distinction is drawn in the studies of Aktionsart between stative and dynamic sentences. While it occurs frequently in a dynamic sentence, as illustrated by (108), an indefinite expression is normally unacceptable as subject of a stative sentence in Chinese, as in other languages, as shown in the following example:

(i) *Yi zhi xiao qi’e hen e.
   one CL little penguin very hungry
   ‘A little penguin was very hungry.’

15. I myself recently heard the following utterance from the manager to a waiter in a Chinese restaurant:

(i) Keren lai le, ling libian qu.
   customer come CRS lead inside go
   ‘There are some customers coming. Lead them to the inside.’

when it was obvious from the body language of the restaurant staff and the customers that the customers had not patronized the restaurant before.

References


Identifiability and definiteness in Chinese


Li, Lingding (1986). *Xiandai Hanyu Juxing [Sentence Patterns in Modern Chinese]*. Beijing: Shangwu Yinshuguan.


