COGNITIVE STATUS AND THE FORM OF REFERRING EXPRESSIONS IN DISCOURSE

eanette K. Gundel	NANCY HEDBERG
University of	Simon Fraser
Minnesota	University

RON ZACHARSKI University of Edinburgh

In this paper we propose six implicationally related cognitive statuses relevant for explicating the use of referring expressions in natural language discourse. These statuses are the conventional meanings signalled by determiners and pronouns, and interaction of the statuses with Grice's Maxim of Quantity accounts for the actual distribution and interpretation of forms when necessary conditions for the use of more than one form are met. This proposal is supported by an empirical study of the distribution of referring expressions in naturally occurring discourse in five languages-English, Japanese, Mandarin Chinese, Russian, and Spanish.*

1. INTRODUCTION. One of the more interesting facts about human language is that we can use different forms to refer to the same thing, and the same form can be used to refer to many different things. Yet people somehow manage to understand one another. A particular issue of Language, for example, can be referred to as an issue of Language, the issue, that issue, this issue, that, this, or it, and any one of these forms could be used to refer to other objects on different occasions. The question then is: what do speakers/writers know that enables them to choose an appropriate form to refer to a particular object and what do hearers/readers know that enables them to identify correctly the intended referent of a particular form?

The study of reference has a long tradition in the philosophical literature. and has been investigated from various perspectives within linguistics and psychology (see, for example, Karttunen 1976, Nunberg 1978, Hawkins 1978, 1984, 1991, Clark & Marshall 1981, Grosz 1981, Heim 1982, Maclaran 1982, Givón 1983, Ariel 1988, Kronfeld 1990, and numerous works cited therein). Although many important insights and observations have come out of this work, basic facts concerning the distribution and understanding of different forms of referring expression in natural language discourse still remain unexplained. In this paper we outline a theory whose main premise is that different determiners and pronominal forms conventionally signal different cognitive statuses (information about location in memory and attention state), thereby enabling the

* We gratefully acknowledge the University of Minnesota Graduate School for a Grant-in-Aid and Faculty Summer Research Grant to Jeanette Gundel and Cornell University for a Mellon Postdoctoral Fellowship to Nancy Hedberg. Ron Zacharski was partly supported by UK Economic and Social Research Council grant #R000 23 3460 to Edinburgh University. This paper represents a revised and expanded account of research introduced in Gundel, Hedberg. & Zacharski 1988, 1989, 1990. We are indebted to Thorstein Fretheim, Alison Marchant, Gerald Sanders, Candace Sidner, and two anonymous referees for their critical comments and suggestions. Special thanks also to Ann Mulkern, Suellen Rundquist, and Karen Schaffer for their work on various stages of this project and to Karen Frederickson, Xinghui Chen, Hui-chuan Huang, Svetlana Krylova, Sahya Luck, Yoshiko Matsumoto, Silas Oliveira, Satoko Suzuki, and Zhu Yao for help with data collection, translation, and analysis. The authors' names appear in alphabetical order.

addressee to restrict the set of possible referents. In §2 we introduce the Givenness Hierarchy, a set of implicationally related statuses which we propose are necessary for explaining the relation between referring forms and conditions for their appropriate use and interpretation across languages. In §3 we show how the Givenness Hierarchy accounts for restrictions on the distribution of forms for a particular type of reference which Prince (1981b) calls 'inferrable'. In §4 we propose correlations between statuses on the hierarchy and different forms in Mandarin Chinese, English, Japanese, Russian, and Spanish, and in §5 we present results of an empirical study of the use of referring forms in these five languages. Finally, we show in §6 how the Givenness Hierarchy interacts with the Gricean Maxim of Quantity to predict the actual distribution and interpretation of forms when necessary conditions for the use of more than one form are met.

2. THE GIVENNESS HIERARCHY. It is widely recognized that the form of referring expressions, like such other aspects of language as word order and sentence intonation, depends on the assumed cognitive status of the referent. i.e. on assumptions that a cooperative speaker can reasonably make regarding the addressee's knowledge and attention state in the particular context in which the expression is used (cf., for example, Chafe 1976, 1987, Gundel 1978, 1985, Prince 1981b, Grosz & Sidner 1986). But the nature of such cognitive statuses and the logical and empirical relations among them is still a matter of some debate. Moreover, researchers have not always distinguished the statuses themselves (e.g., whether or not an addressee already has a mental representation of a referent and whether attention is focussed on the referent) from the means by which a referent acquires a particular status (e.g. whether it has been linguistically introduced, whether it is part of general cultural knowledge, and so on).

We propose that there are six cognitive statuses relevant to the form of referring expressions in natural language discourse, and that these are related in the Givenness Hierarchy shown in 1.

(1) THE GIVENNESS HIERARCHY:

in focus >	activated	> familiar >	uniquely identifiable	> referential	type > identifiable
{ <i>it</i> }	$ \begin{cases} that \\ this \\ this \\ N \end{cases} $	{that N}	{the N}	{indefinite this N}	{ <i>a</i> N}

Each status on the hierarchy is a necessary and sufficient condition for the appropriate use of a different form or forms.1 The relevant English forms are given in 1. In using a particular form, a speaker thus signals that she assumes the associated cognitive status is met and, since each status entails all lower statuses, she also signals that all lower statuses (statuses to the right) have

¹ We assume here that demonstrative this and indefinite this are two distinct forms rather than just different uses of the same form. This assumption predicts correctly that a (proximal) demonstrative whose referent must be activated will not necessarily have an indefinite use in other languages.

Computoringuistik

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

I've come to the conclusion that I will not extend my (24) a. CONTRACT AT THE BANK. [personal letter]

b. #I've come to THE CONCLUSION.

LA SOMBRA LARGA Y NEGRA DE LOS HOMBRES (25) a the shadow large and black of the men siguió.

followed

- 'The long and black shadow of the men followed.' [Juan Rulfo, No oves ladrar los perros]
- b. #LA SOMBRA siguió.

the shadow followed

'The shadow followed.'

While pronominals and the definite article appear to require the same statuses across languages, the situation is more variable for demonstratives and indefinite articles.

4.2.1 DEMONSTRATIVES. As noted in §2, the referent of a noun phrase introduced by a demonstrative has to be at least familiar in English. A demonstrative determiner is inappropriate when the referent is uniquely identifiable but not familiar to the addressee. In 24, for example, the conclusion cannot be replaced by that conclusion or this conclusion. This is also the case for demonstrative determiners in Japanese, Russian, and Spanish. Thus, la sombra in 25 could not be replaced by esta sombra, esa sombra, or aquella sombra in the particular context in which 25 occurs. And the Spanish, Japanese, and Russian examples in 26-28 are appropriate only if the addressee can be assumed to know that the speaker's neighbor has a dog.

(26) Spanish:

Ese No pudo dormir anoche. perro Aquel not could sleep last.night dog

de al lado no me dejó dormir.

of to.the next.door not me let sleep

'I couldn't sleep last night. That dog next door kept me awake.' (27) Japanese:

wa hitobanjuu nemurenakatta. Tonari Kinoo vesterday TOP all night couldn't sleep neighbor no ie no ano inu no sei da. GEN house GEN that dog GEN reason is

'I couldn't sleep last night. That neighbor's dog is the reason.' (28) Russian:

Èta sobaka u soseda mne vsju noč' Ta at neighbor me all night dog ne davala spat'. not allow to.sleep 'That dog next door kept me awake all night.' However, as we noted in §4.1, the Chinese distal demonstrative determiner nèi in a sentence like 23 only requires the referent to be uniquely identifiable. but not necessarily familiar.

The languages also differ as to which demonstrative determiners, if any, require the referent to be activated. In Spanish, which has a three-way distinction in demonstratives, only the proximal determiner este requires activation. Thus, both the medial ese and the distal aquel are possible in 26 above. even if the dog has not been activated by the immediate linguistic or extralinguistic context.17 But in Japanese, which also has three demonstratives, both proximal and medial determiners, like pronouns, require activation. Although the facts are not entirely clear, it appears that the medial demonstrative sono would be appropriate in 27 only if the addressee were currently aware of the dog, i.e. if it could be assumed that a representation of the dog is already activated.18

While Japanese and Spanish differ as to whether or not the medial demonstrative determiner requires activation, the proximal demonstrative determiner requires activation in both languages, as it does in Chinese and English. Russian, however, differs from the other languages in lacking the activation condition on the proximal determiner. Thus the proximal èta is possible (and in fact preferred) in 28 above, even if the dog has not been recently mentioned and is not present in the immediate discourse context, i.e. if the dog is familiar, but not activated.

Finally, the languages appear to differ in whether one or more demonstrative forms require not only that the referent be activated, but that it be speakeractivated. As we noted in §2, the referent of English that can be used to comment on the remarks of another speaker (activated, but not speaker-activated). as in 7 above, or it can be used to comment on the speaker's own remarks (speaker-activated), as in 29.

(29) John, this speech was a magnificent triumph for the President. He showed he could stay awake for twelve whole minutes. He showed that he could speak every word off of his teleprompter, even the long ones. But the speech doesn't have any chance of putting the scandal behind him, because the scandal is not about mistakes, as he said, and it's not about mismanagement, as the Tower Commission said. It is about a betrayal of principles, it's about lying,

¹⁷ Although both forms require only familiarity, they are not equally appropriate in all contexts. In languages which have two demonstratives that require only familiarity, the more distal form is generally restricted to contexts where distance (either spatial or cognitive) is being emphasized. Thus, Spanish speakers find the medial ese to be more natural than the distal aquel in an example like 26.

¹⁸ There has been considerable debate regarding conditions on the appropriate use of sono and ano (cf. Kuroda 1965, Hinds 1973, Kuno 1973, and Kitagawa 1979). Our claim that sono requires activation while ano requires only familiarity appears to be at least consistent with all positions. But it is not clear whether this distinction can account for all differences in the distribution of these. two forms.

and it's about breaking the law. And THOSE ISSUES remain. [PBS. The McLaughlin Group. 3/6/87]

However, pronominal and determiner *this* REQUIRE speaker activation. Thus, while *those issues* can be replaced by *these issues* in 29, *this* is unnatural across speaker boundaries in 30.

(30) A: I think that my novels are better than his.

B: I agree with that (statement)/??this (statement).

Speaker activation also appears to be required for proximal demonstrative pronouns and determiners in Japanese (*kore*, *kono*) and Spanish (*ésta*, *esta*). Thus *kore* in 31 is inappropriate in B's utterance, because the referent has been activated by A.

(31) Japanese:

A: Watashi no hon wa Mishima no yori ii. I GEN book TOP Mishima GEN than good 'My book is better than Mishima's.'

B: $\begin{cases} Sore \\ \#Kore \end{cases}$ wa honto da. TOP true be 'That/this is true.'

And, while the facts are less clear than in Japanese, most Spanish speakers that we consulted find the proximal *esto* to be less natural than the medial *eso* in 32.

(32) Spanish:

A: Mis cuentos son mejores que los de Ortega. my stories are better than those of Ortega 'My stories are better than Ortega's.'

B: Estoy de acuerdo con {eso am in agreement with 'l agree with that/this.'

Finally, as illustrated in 33 and 34, proximal demonstratives in Chinese and Russian do not require speaker activation. (As noted above, the proximal determiner in Russian does not require activation at all.) In fact, speakers prefer the proximal ($zh\dot{e}$, $\dot{e}to$) to the distal ($n\dot{e}i$, to) forms in these examples.

(33) Chinese:

A: Wõ juéde wõde xiǎoshuõ bǐ lǚxūn xiẽ-de hǎo. I think my novel than Luxun write-NOM good 'I think my novels are better than Luxun's.'

B: $\begin{cases} Zh \dot{e} ige \\ ?N \dot{e} ige \end{cases}$ wo chéngrèn. I admit

'I agree with this/that.

(34) Russian:

A: Moi rasskazy lučše čem ego rasskazy.
 My stories better than his stories
 'My stories are better than his stories.'

B: $Ja \ s$ I with $\begin{cases} etim \\ ?tem \end{cases}$ soglasen. 'I agree with this/that.'

4.2.2. THE INDEFINITE ARTICLE. Of the languages we examined, only Chinese, English, and Spanish have an indefinite article, and these languages are alike in that the indefinite article appears to require only type identifiability. But distribution of the indefinite article differs somewhat in the three languages.

As noted in §2, a noun phrase introduced by English a may be referential or merely type identifiable. Thus, a sentence like 35 can have an interpretation where the speaker intends to refer to a particular car and wants the addressee to construct a representation of that car (the referential reading), or it can have an interpretation where the speaker intends only to assert that Ellen bought at least one car, i.e., she is now a car-owner (the nonreferential, merely typeidentifiable reading). The two interpretations are distinguished in 35' and 35", respectively.

- (35) Ellen bought a car.
- (35') We went to Southtown Toyota yesterday. Ellen bought a car. It's a Regatta Blue Corolla with a sunroof.
- (35") A: You'll never guess what happened today.

B: Don't tell me. Ellen bought a car.

In contrast to a, which requires only type identifiability, the referent of indefinite *this* must be referential. Thus, 36 is appropriate in the context of 36', but not in 36''.

(36) Ellen bought this car.

- (36') We went to Southtown Toyota Yesterday. Ellen bought this car. It's a Regatta Blue Corolla with a sunroof.
- (36") A: You'll never guess what happened today.

B: #Don't tell me. Ellen bought this car.

The Spanish indefinite article *un* also requires only type identifiability, as illustrated by examples like 37 and 38, where the referents of the phrases *una gaseosa* and *una celosa fanatica* are type identifiable, but not referential.

(37) El quiere (una) gaseosa pero no encuentre.

he wants a soft.drink but not find

'He wants a soft.drink, but can't find (one).'

(38) No sos (una) celosa fanática.

not be a jealous fanatic

'Don't be a fanatic jealous (woman).'

However, unlike the English indefinite article, the Spanish indefinite article is

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

optional in clearly nonreferential contexts like 37 and 38, and in some nonreferential contexts it is not used at all.

Chinese does not have an obligatory indefinite article. However, as we will see in the next section, the distribution of yi 'one' in actual discourse is similar to that of the indefinite article in Spanish. Most singular referential indefinites and some nonreferential ones are introduced by yi.¹⁹

Bolinger 1980 (Abstract) notes that 'in Old English the indefinite article, now required in the John is a lawyer type of sentence, was optional'. Bolinger also points out that the indefinite article is still optional in Modern English 'in certain peripheral constructions' such as He'll never make (a) captain and He's running for (an) office. Chinese, English, and Spanish thus appear to be at different stages in the development of the indefinite article from the numeral 'one'. Chinese represents a very early stage in which the indefinite article is optional and is generally restricted to referential contexts. In Spanish the indefinite article is optional in Some nonreferential uses; but it has become grammaticalized not only for referential indefinites, but for most non-referential ones as well.

5. THE GIVENNESS HIERARCHY AND NATURAL LANGUAGE DISCOURSE. The Givenness Hierarchy and the associated forms in Table 1 predict that a particular form will be inappropriate if the required cognitive status is not met. Since the statuses are implicationally related, this analysis also predicts that a form can appropriately encode the necessary and sufficient status (the status immediately above the form in the table) as well as all higher statuses (statuses to the left). For example, the referent of an NP with the definite article the in English may be just uniquely identifiable, or it may also be familiar, activated, or in focus; and the referent of a demonstrative pronoun may be simply activated or it may also be in focus. We would thus expect forms to be distributed across more than one status in actual discourse. We tested these predictions by investigating the distribution of different forms of reference in naturally occurring discourse for each of the five languages in our study.²⁰ The results of this investigation are given in Tables 2-6, which show the distribution of the different forms according to the highest cognitive status met by that form in the particular context in which it was used.²¹

¹⁹ Li & Thompson (1981:132) also note that 'the numeral yi 'one'', if it is not stressed, is beginning to function as a.'

²⁰ Our data comes from a variety of spoken and written sources which differ in formality and degree of planning. These include novels, short stories, magazine articles, news broadcasts, interviews, and casual conversations. In addition, for all languages except Russian, we also analyzed narrative film descriptions which were collected for another study. The methodology here was similar to that of the Pear Stories (Chafe 1980). Speakers viewed a silent film called the *The Golden Fish* and, immediately after viewing the film, described it to another native speaker of their language.

²¹ Cognitive status involves assumptions that a coöperative speaker can reasonably make regarding the addressee's knowledge and attention state in the context in which an expression is used. As discussed in §2 (particularly note the discussion of examples 10 and 11), the cognitive

		IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE	TOTAL
	ø	25	1					26
	tā	40						40
	zhè	-	2					2
	nèi							
	zhè N	12	26	1				39
	nèi N	1	7	2				10
	vi N					17	2	19
	N	12	17	14	49	2	10	104
•	TOTALS	90	53	17	49	19	12	240

TABLE 2. Distribution of Chinese forms according to highest status.

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE	TOTAL
it	214	1					215
HE		I					1
this		15					15
that	1	17					18
this N	ľ	11					12
that N		10	7				17
the N	30	95	47	108			280
lefinite							
this N					1		1
a N					41	55	96
TOTALS	246	150	54	108	42	55	655
	TABLE 3. I	Distribution of	English forn	ns accordin	g to highest stat	us	

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE	TOTALS
ø	87	1					88
kare	4						4
kore	1	1					2
sore		I					1
are							
kono N	1	7	1				9
sono N	18	15	1				34
ano N		1	1				2
N	14	32	17	71	45	44	223
TOTALS	125	58	20	71	45	44	363
TUTALS	TABLE 4	. Distribution of	of Japanese fo	orms accord	ing to highest sta	tus.	505

status of a referent is not uniquely determined by syntactic structure. This is reflected in the methods we used for analysis. Two trained coders analyzed each transcript. While there were coding guidelines based on syntax and recency of mention, decisions on cognitive status were not completely mechanical, but also involved judgments based on relevance and the shared knowledge and beliefs of the speaker and hearer. The two coders agreed on approximately 90% of the tokens examined. Most disagreements were between familiar vs. activated or activated vs. in focus. We believe this is because the boundaries between statuses involving attention state are not discrete, even though they map onto discrete forms. Disagreements were resolved by discussion among the coders.

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	Түре	TOTALS
ø	18						18
ono .	51	2					53
ONO		1					1
èto	2	9					11
tot							
èto N	1	7	2				10
to N							
N	25	29	22	66	21	28	191
TOTALS	97	48	24	66	21	. 28	284
	TABLE	5. Distribution	of Russian fo	rms accordi	ng to highest sta	tus.	

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	Туре	TOTALS
ø	63						63
él (subj)	30	3					33
él (nonsubj)	57	2					59
éste		1					1
ése		4					4
aquél		1					1
este N							
ese N	1	2	3				6
aquel N							
el N	23	45	22	42			132
un N					22	10	32
TOTALS	174	58	25	42	. 22	10	331

TABLE 6. Distribution of Spanish forms according to highest status.

As predicted, all but a few occurrences of referring expressions examined met the necessary conditions posited for that expression for the language in question (see Table 1). For example, in Spanish all referents of zero pronominals were in focus; all referents of overt pronouns (both personal pronouns and demonstratives) were at least activated;²² all phrases with the demonstrative determiner *ese* were at least familiar; all phrases with the definite article *el* were at least uniquely identifiable; and all phrases with the indefinite article *un* were at least type identifiable.

When the referent of a particular form does not have the required cognitive status, the result may be infelicitous, as in 39, where the addressee was not able to identify the referent of the pronoun *these* because it wasn't activated.

- (39) M: These. Do these go in here or there?
 - K: These?
 - M: The ones I just got done writing.23 [Fredrickson tapes]

²² Because we had tapes for only a small portion of our data, we were not able to distinguish stressed from unstressed pronouns in most cases. Note, however, that the referents of all but a very few personal pronouns in any of the languages were in focus. Either a stressed or an unstressed personal pronoun would therefore have been appropriate.

²³ This example and some others in the paper are drawn from data outside the study reported on in Tables 2–6. The 'Frederickson tapes' are transcribed recordings of conversations during family gatherings, collected by Karen Frederickson of the University of Minnesota Linguistics Department in 1975–1987. A more typical result of using a form whose referent does not meet the required status is that the addressee accommodates and is able to associate the correct referent with the form in spite of the fact that it was used inappropriately. This is illustrated in 40, the one instance of an unstressed personal pronoun in our English data where the referent is not in focus.

- (40) K.1: Barb got it.
 - N.2: Catmopolitan?
 - K.3: Yeah.
 - N.4: Catmopolitan.
 - K.5: She got it.
 - N.6: Yup. I suspicion she was a cat in her other life.
 - K.7: Oh did I tell you that THEY have a cat, they have two cats; one is Maynard and one's Dudley. [Frederickson tapes]

The referent of *they* in K7 is Barb and her husband. Although the addressee knows that Barb is married (i.e., the couple is familiar), it is probably not the case that mentioning Barb automatically brought her husband into focus (or that the couple had even been activated in this context). But the reference succeeded nevertheless.

The results in Tables 2–6 thus clearly support the necessary conditions hypothesized for the forms in Table 1. Moreover, we found that forms were not only used in coding noun phrases whose referents met the minimal required status; they were also used in coding higher statuses. For example, in English there are some tokens of *the N* for all statuses to the left of, and including, uniquely identifiable, and there are some tokens of the demonstrative pronoun *that* for both activated and in focus. Similarly, in Russian, bare nouns with no preceding determiner are found for all statuses; and there are tokens of the demonstrative determiner *èto* for referents that are activated, as well as for those that are familiar but not activated. This is as expected, since each of the cognitive statuses in the Givenness Hierarchy entails all lower statuses, and a particular status thus implies the possibility of reference with forms associated with lower (entailed) statuses.²⁴

However, the distribution across statuses varies considerably for different forms. For example, we found a relatively high occurrence of phrases with the definite article in English and Spanish and of a bare noun in the other languages for all statuses that imply uniquely identifiable. By contrast, demonstrative pronouns are rarely used for referents that are in focus, and demonstrative determiners are rarely used for referents that are familiar. Similarly, there were no occurrences of indefinite articles for statuses higher than referential, even though all statuses meet the necessary conditions for the use of this form. What

²⁴ The fact that all occurrences of the Chinese distal demonstrative determiner $n\dot{e}i$ were at least familiar may be just an accident of the data we examined; or it may be that the use of $n\dot{e}i$ for referents that are not familiar is relatively rare in Chinese, even though necessary conditions for this form are the same as for the definite article in English and Spanish (see §3). In any case, these findings are not inconsistent with the claim that the determiner $n\dot{e}i$ requires the referent to be uniquely identifiable, but not necessarily familiar.

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

then explains the choice among forms when necessary conditions for the use of more than one form are met? This question will be addressed in the next and final main section of the paper.

6. THE GIVENNESS HIERARCHY AND GRICE'S MAXIM OF QUANTITY. Since each of the cognitive statuses in the Givenness Hierarchy entails all lower statuses, a particular form can often be replaced by forms which require a lower status. For example, the proximal demonstrative determiner *these*, which requires that its referent be activated, could be replaced by *those*, which requires only familiarity; by *the*, which requires only unique identifiability; and sometimes even by an indefinite article (or zero if it is plural), which requires only type identifiability. This is illustrated in 41, where *these systems* was the form used in the cited magazine article, but any of the other alternatives would also have been acceptable in this context.

(41) These incredibly small magnetic bubbles are the vanguard of a new generation of ultradense memory-storage systems.

These systems Those systems The systems New generation ultradense memory-storage systems

are extremely rugged: they are resistant to radiation and are nonvolatile. [Gordon Graff, Better bubbles. Popular Science 232(2):68 (1988)]

As noted in §5, however, the distribution of forms across statuses which meet necessary conditions for their appropriate use is not random. And some forms rarely occur, even when necessary conditions for their use have been met. Thus, as seen in Tables 2, 3, and 6, all noun phrases in our data coded with the indefinite article in Chinese, English, and Spanish are at most referential or type identifiable, even though the status 'type identifiable' is entailed by all other statuses on the hierarchy and therefore all statuses meet necessary conditions for use of the indefinite article. Similarly, entities in focus are generally coded by unstressed pronouns or by zero, even though a demonstrative pronoun in all five languages requires only activation, and any entity that is in focus is necessarily also activated. Moreover, a given form is often inappropriate, or conveys some special effect, even when necessary conditions for its use have been met. Thus, it in 42b is naturally interpreted as referring to the topic, Simplified English, which is in focus at this point in the discourse. But if it is replaced by this, as in 42b', the most natural interpretation is one where this refers to the whole statement about Simplified English.

- (42) a. Simplified English disallows the use of passive, progressive, and perfective auxiliary verbs, among other things.
 - b. It requires engineers to break up long compound nouns and technical expressions into chunks of three or less elements. [message from electronic news group]
 - b'. This requires engineers to break up long compound nouns and technical expressions . . .

In the remainder of this section, we will argue that such facts involve conversational implicatures which result from interaction of the Givenness Hierarchy with Grice's maxim of quantity, stated in 43.²⁵

(43) MAXIM OF QUANTITY (Grice 1975):

Q1 Make your contribution as informative as required (for the current purposes of the exchange).

Q2 Do not make your contribution more informative than is required.

The classic examples that give rise to quantity implicatures are ones which form an implicational scale (see Horn 1972, 1984, Hirschberg 1985). In Q1based implicatures, use of a weaker (entailed) form conversationally implicates that a stronger (entailing) form does not obtain. For example, while 44 entails (and is therefore consistent with) 45, a speaker who utters 45 normally implicates the negation of 44, i.e., she implicates that she does not agree with every-thing the addressee said.

(44) I agree with all of what you said.

(45) I agree with some of what you said.

In Q2-based implicature, by contrast, use of a weaker (entailed) form implicates a stronger (entailing) form. For example, 46 entails 47, but not vice versa. However, a speaker who utters 47 normally intends to convey 46, i.e., she implicates that she will go to the meeting only if the addressee goes.

(46) I'll go to the meeting if and only if you go.(47) I'll go to the meeting if you go.

Atlas & Levinson (1981) propose that forms which give rise to such implicatures are usually ones where the meaning associated with the stronger form is stereotypical. Thus, since conditional relations discussed in everyday discourse are stereotypically biconditional, the biconditional form (*if and only if*) would be 'more informative than is required'.

Since the statuses in the Givenness Hierarchy form an implicational scale, we find, as expected, that use of forms which overtly signal different cognitive statuses gives rise to quantity implicatures as well.

6.1. Q1. As noted above, we found no examples of the indefinite article in Chinese, English, or Spanish for statuses above referential, even though all statuses meet necessary conditions for this form. Moreover, use of an indefinite article typically implicates by Q1 that the referent is not uniquely identifiable

²⁵ Following Grice 1975, we use the term 'conversational implicature' to refer to nondeductive inferences that arise when the maxims of conversation are being observed. These are, of course, not restricted to conversation, but are assumed in all coöperative uses of language, both spoken and written. There has been important work in recent years which aims to reduce one or both parts of the quantity maxim to more general principles (cf., for example, Atlas & Levinson 1981, Levinson 1983, Horn 1984, and Sperber & Wilson 1986). We have returned to Grice's original formulation because we find it adequate for explicating the facts at issue here (see also Levinson 1987). However, we highly recommend the recent work for essential insights into ways in which the two parts of the quantity maxim interact (e.g. Horn's Division of Pragmatic Labor).

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

and hence also not familiar, activated, or in focus.²⁶ Thus, while we claim that the conventional meaning of the indefinite article (what it explicitly signals) is simply that the referent is at least type identifiable, use of this form conversationally implicates by Q1 that the addressee cannot uniquely identify the referent.

Since conversational implicature (unlike entailment) is not a necessary inference, our analysis correctly predicts that the referent of an indefinite noun phrase can be uniquely identifiable, or even familiar, in some contexts. For example, in 48 the definite article (the form actually used in the original citation) is appropriate in referring to the 137.7-million-dollar Host Senate office building because the latter is uniquely identifiable. However, since the descriptive information necessary to identify the referent is fully encoded in the nominal itself, an explicit marker of identifiability is unnecessary and an indefinite article would have been equally appropriate here.

(48) The senate on August 10 voted 50 to 48 to spend \$736,400 for a third Senate Gymnasium due to be built in THE/A 137.7 MILLION-DOLLAR HOST SENATE OFFICE BUILDING OPENING IN JANUARY. (U.S. News and World Report, 3 August 1982, 9)

Similarly, as discussed in Dahl 1984, the phrase *a doctor* in 49 does not introduce a new entity into the discourse. Since it is the property of being a doctor, and not the identity of this particular doctor, which is relevant here, an indefinite article can be used without implicating that the referent is not familiar. If nonidentifiability (and therefore nonfamiliarity) were part of the conventional meaning associated with the indefinite article, 49 would necessarily have the interpretation that the speaker believes exercise helps because she heard it from someone other than Dr. Smith.

(49) Dr. Smith told me that exercise helps. Since I heard it from A DOCTOR, I'm inclined to believe it.

Our analysis also predicts correctly that the interpretation of nonidentifiability (or nonfamiliarity) associated with the indefinite article can be cancelled without contradiction, as in 50, and reinforced without redundancy, as in 51.

- (50) I met a STUDENT before class. A STUDENT came to see me after class as well—in fact it was the same student I had seen before.
- (51) But forged provenance papers still did not mean that the kouros was fake.... The Getty decided that the fake documents were not reason enough to ask Mr. Becchina, the Basel dealer who had sold

²⁶ Grice himself observed that 'Anyone who uses a sentence of the form ''X is meeting a woman this evening'' would normally implicate that the person to be met was someone other than X's wife, mother, sister, or perhaps even close platonic friend' (1975:168). In other words, although Grice does not explicitly state it in these terms, use of *an N* implicates that the referent is not someone uniquely identifiable to the addressee. Note, however, that the indefinite article does not implicate nonreferentiality. This is so because, with the exception of indefinite *this* in colloquial English, the languages we examined have no separate form that signals referentiality. The indefinite article would thus be the strongest possible form for coding something which is referential but not uniquely identifiable. the kouros, to take back the sculpture. (Attempts by The Times to reach Mr. Becchina were unsuccessful.) Then last April, an independent scholar in London, Jeffrey Spier, was shown a photograph of a fake torso of a kouros, belonging to A BASEL DEALER (NOT MR. BECCHINA), that looked similar to the Getty's sculpture. [The New York Times, 4 August 1991, 2:24]²⁷

We also find Q1 operating in the coding of noun phrases whose referents are in focus. As Tables 2–6 show, the number of in-focus referents coded by the strongest, most restrictive possible form (zero or unstressed pronoun) ranges from 71% (Russian) to 87% (English); and a demonstrative pronoun was rarely used in such cases.²⁸ In English, for example, 214 of the 246 noun phrases that were in focus were unstressed personal pronouns, and only one was a demonstrative pronoun. Similarly, in Japanese 87 of the 125 noun phrases whose referents were in focus were coded by a zero pronominal, and only 5 by a demonstrative pronoun. This is so even though all forms meet the necessary conditions for coding an in-focus referent, since the status 'in focus' entails all other statuses. Moreover, use of a demonstrative pronoun, which requires only activation, often implicates by Q1 that the referent is not currently in focus; that is, it signals a focus shift (cf. Isard 1975, Linde 1979, Bosch 1983, Sidner 1983).²⁹

²⁷ Example 50 is adapted from Hawkins (1991:419), who argues, as we do here, that use of the indefinite article conversationally implicates the negation of what is conventionally signalled by the definite article. Hawkins explicitly defines the conventional implicature of *the* as follows: '*The* conventionally implicates that there is some subset of entities {P}, in the universe of discourse which is mutually manifest to S & H on-line and within which definite referents exist and are unique' (1991:414). We believe that the extension of this definition is essentially equivalent to what we mean by 'uniquely identifiable'.

Example 51 has another interpretation where the phrase *a Basel dealer* is used to express the inability of the London scholar to identify the dealer as Mr. Becchina. As pointed out to us by an anonymous reviewer, the expression *not Mr. Becchina* does not reinforce an implicature for the reader on this interpretation.

According to the native speakers we consulted, the facts illustrated in 48–51 are similar in Chinese and Spanish. It is tempting to propose that nonidentifiability or nonfamiliarity is conversationally (not conventionally) implicated by the indefinite article in all languages which have such an article. This would be the case if articles, personal pronouns, and demonstratives conventionally signal a cognitive status on the givenness hierarchy in all languages, as they do for the languages we have investigated here. In any case, this is an interesting empirical claim which is worthy of further investigation.

²⁸ Although all the languages but English are 'pro-drop', we find some interesting differences among them in the distribution of pronoun vs. zero. While 70% of referents in focus were coded with zero in Japanese, Chinese had only 28% zeroes, and Russian had only 19%. The preferred form for in-focus referents in Chinese and Russian appears to be unstressed personal pronouns (44% in Chinese and 53% in Russian). In Spanish, which allows zero only in subject position, 68% of in-focus pronominal subjects were zeros and 32% were overt subject pronouns. The relatively high number of zero pronominals in Japanese is predicted by our analysis, since only a zero argument requires the referent to be in focus in this language (i.e., Japanese has no unstressed personal pronouns). The facts in the other languages are consistent with our analysis, but we have no explanation for the differences in frequency of zero pronominals.

²⁹ Since the topic of the immediately preceding utterance is always in focus at the beginning of the next utterance (see §2), a shift in topic is generally also a focus shift.

Some illustrations of the focus-shift function of pronominal demonstratives are given in 52–56.

(52) Chinese:

- a. Xião háizi hěn gāoxìng. Suöyǐ tā bă tā dài gěi hóng small child very happy so he ом he bring give red jīnyú de yīzhī huā fàng zài yúgāng lǐmiàn. goldfish NOM one flower put in bowl LOC '(The) small child was very happy. So he put a flower he had brought for (the) goldfish into (the) bowl.'
- b. Zhèi jiù shì zhèi-ge gùshi.

This then be this-CLF story

'This then is the (lit. 'this') story.'

(53) Japanese:

a. Toori e dete shibaraku hashitteku. street to go.out for.some.time run

'He goes out onto the street and runs for some time.'

- b. To nanka yatai mitaina omise ga atte. and something stall seem shop NOM be 'There is a shop like a stall.'
- c. KARE wa sono omise no toko e itte. he TOP that shop GEN place to goes
 - 'He goes to that shop.'
- (54) K 1: And .. So what he did was ... came in, set up the tree ... 2: And then he made wassail, with rum in it?
 - 3: And .. made it in coffee cans and heated it on the stove in the graduate lounge.
 - A 4: Oh, gee.
 - K 5: And THIS was the solstice tree. [Frederickson tapes]
- (55) When Snepp makes a speech he has to submit a text to CIA censors first. When he wrote a book review for the Los Angeles Times, he had to show it to the agency before he sent it to the newspaper, and when the editor asked for a change, he had to show THAT to the censors too. [Anthony Lewis, Secrecy policy has no sense. Minneapolis Star and Tribune 4/14/87]
- (56) Anyway going on back from the kitchen then is a little hallway leading to a window, and across from the kitchen is a big walk-through closet. On the other side of THAT is another little hallway leading to a window [personal letter]

In the Chinese example in 52, the proximal demonstrative pronoun *zhèi* in (b) implicates that focus has shifted from the boy—who is the center of attention at the end of (a)—to the story. In the Japanese example in 53, focus shifts from the boy in (a) to the shop in (b). Thus, when the boy is reintroduced in (c), the pronoun *kare* (which, like demonstrative pronouns, requires only activation) shifts focus back to the boy again. In 54, the tree has been activated by its mention in K1, thus licensing the use of a pronominal in K5. However, since the tree is not in focus in K2, K3, and A4, reference to it in K5 constitutes a

focus shift and thus requires a stressed demonstrative form. The fact that demonstrative pronouns, unlike unstressed personal pronouns and zero, can be used to refer to something which is activated but not currently in focus follows from the necessary conditions posited for appropriate use of these forms. What is of interest here is that demonstratives not only don't require the referent to be in focus, but often implicate that the referent is not currently in focus. This is particularly striking in cases where there is potential ambiguity of reference. For example, *this* in 54 would not normally be interpreted as the currently infocus wassail; the referent of *that* in 55 is interpreted as something other than the first version of the book review; and the referent of *that* in 56 is not understood as the kitchen. If *that* in 56 were replaced by *it*, however, the kitchen would be the only possible referent since it is in focus here.

We believe that use of pronominal *this* and *that* in referring to previous statements (cf. Webber 1988) is just a special case of focus shift, since the focus of attention at the point after a statement is made is typically not the event or state of affairs described by that statement but rather the entity which is the topic of the statement. Thus, in 42b above, use of *it* continues the topic and refers to Simplified English; but use of *this* in 42b' implicates that the referent is not in focus, and is interpreted as referring to the whole statement about what Simplified English disallows.³⁰

6.2. Q2. We have seen in §6.1 that the choice among forms when the necessary cognitive status for more than one form is met can be partly explained in terms of the first part of the quantity maxim—be as informative as required. Interaction of this maxim with the Giveness Hierarchy correctly predicts that (a) an indefinite article will normally not be used for referents that are uniquely identifiable, since this form explicitly signals only type identifiability, and (b) for referents that are in focus, an unstressed personal pronoun or zero, which explicitly delimits the set of possible referents to those that are in focus, will normally be chosen over a demonstrative pronoun, which gives less information about cognitive status because it only requires that the referent be activated.

However, as is evident from Tables 2–6, use of a particular form doesn't always conversationally implicate that necessary conditions for a form requiring a higher status don't obtain. Over 85% of full noun phrases whose referents were at least familiar were introduced by a definite article in English and Span-

³⁰ While we don't necessarily want to claim that interaction of cognitive status and implicature accounts for all functions of particular forms of referring expression, we believe that at least some frequently discussed functions of particular forms may insightfully be explained in terms of focus shift. The tendency for overt (stressed) pronouns in pro-drop languages to signal emphasis, contrast, or the opening of a new thematic unit; the tendency for use of a distal demonstrative to signal closure of a thematic unit; and the tendency of demonstratives to signal contrast may all be subcases of focus shift. In (i), for example, the speaker could have referred to the currently in-focus travel journal with pronominal *ii*. Use of *this* implicates by Q1 that the referent should be viewed in a new way, thus implicitly bringing into focus other travel journals with which it might be contrasted.

(i) I've been working some more on my book, on my travel journal from '85. That sounds funny but I have all my notes. However, I want to write a little better than I am, put a little more effort, make this a little more quality effort than the last one was. [Frederickson tapes]

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

ish and by a bare noun in Russian, even though necessary conditions for a demonstrative determiner (the stronger, more restrictive, form) were met. While demonstrative determiners occurred somewhat more frequently in Chinese and Japanese, a bare noun (the weaker form) was also common for full familiar NPs in these languages (47% in Chinese and 59% in Japanese).³¹ Thus, use of a definite article or a bare noun clearly doesn't implicate by Q1 that the referent is not familiar; rather, it is the second part of the Quantity Maxim—don't be more informative than required—that is relevant here.

The question that naturally arises is: why should choice among definite determiners be primarily dictated by Q2 when, as just argued, choice among definite pronominals is primarily dictated by Q1? We believe the answer to this guestion is related to the fact that pronominal forms have little if any descriptive content, so that information about cognitive status is crucial in delimiting the set of possible referents. For full noun phrases, however, signalling identifiability is often sufficient for identifying the referent, given the descriptive content of the noun and its modifiers, and an explicit signal of a more restrictive cognitive status is therefore unnecessary. Moreover, since familiarity is the most common basis for identifiability, the Q2 implicature here would follow from the observation that the second part of the quantity maxim induces stereotypical interpretations (cf. the observation in Atlas & Levinson 1981 noted above.) Since most references which are uniquely identifiable in a discourse are also at least familiar, explicitly signalling a status higher than uniquely identifiable would be more informative than required. Thus, while some researchers consider familiarity to be part of the conventional meaning of the definite article and treat the nonfamiliar cases as exceptional (cf. Heim 1982), we propose that the definite article conventionally signals only that the referent is uniquely identifiable and that familiarity is conversationally implicated by Q2. Since the definite article explicitly signals that the speaker expects the addressee to uniquely identify the referent, the more restrictive cognitive statuses associated with demonstrative determiners often have little information value and do not need to be signalled explicitly. This is especially true when the referent has just been introduced by a phrase which is at least partially identical in form, as in the following examples from English and Spanish.32

³¹ The higher frequency of demonstrative determiners in Chinese and Japanese may be due partly to the fact that these languages lack a separate form for the definite article. It is of interest, however, that Russian, which also lacks a definite article, patterns more like English and Spanish here, with demonstratives accounting for less than 15% of the full definite NPs. The fact that the highest number of demonstrative determiners was found in the Chinese data would appear to support observations that the demonstrative determiner is beginning to function like a definite article in Chinese (see §4.1). But notice that it is the proximal rather than the distal demonstrative which occurs most frequently. Our data thus suggests that both demonstratives are beginning to function like a definite article in Chinese, with the proximal form predominating for referents that are at least activated.

³² Notice that Q2 can be invoked here only to explain why the definite article is used instead of a demonstrative determiner. It cannot explain why a full NP rather than a pronominal is used in the activated cases. Some possible reasons for this, including ambiguity resolution and global focus shift, are discussed in Marslen-Wilson et al. 1982, Guindon 1985, and Fox 1987.

(57) 'How in the world,' demanded Harriet, 'did you get here?'
'Car,' said Lord Peter, briefly. 'Have they produced the body?'
'Who told you about THE BODY?' [Dorothy Sayers, Have his carcase, 1986:38]

(58) Spanish:

Y él atrajo mucho al pez, y él miraba AL PEZ, and he attracted a.lot to.the fish and he looked at.the fish

- y EL PEZ miraba a él.
- and the fish looked at him

'And he was attracted to the fish, and he looked at the fish, and the fish looked at him.'

Examples 59-61 illustrate the use of a bare noun for an activated referent in Chinese, Japanese, and Russian.

(59) Chinese:

Zhèi shíhou qiàqiǎo jīnyú yǐjīng hé niǎor zài yóuxì. this time so.happen goldfish already with bird at play

Zài yóuxi-de shíhou JINYÚ yťjing tiào-dào le at play-NOM time goldfish already jump ASP zhuōmiàn-shàng.

table-top

'At this time, (it) so happened that (the) goldfish was already playing with (the) bird. At (the) time (they) were playing, (the) goldfish had already jumped onto the table top.'

(60) Japanese:

De sono kake de otokonoko ga katte.

and that gambling in boy NOM win

Akai kingyo ga sono otokonoko no toki wa kingyo ga red goldfish NOM that boy GEN time TOP goldfish NOM kantan.ni patto toreta to ju kanji de:

easily rapidly caught Q say seem be

de отоколоко wa vorokonde uchi e

and boy TOP happily home to

modotte iku-tte iu bamen ga atte.

go.back go-comp say scene Nom is

'And the boy wins the bet. It seems that the red goldfish is easily caught in that boy's turn; and there's a scene where the boy goes home happily.'

(61) Russian:

Etot kot poterjal vsjakuju sovest'. Koτ— brodjaga i bandit. this cat lost every conscience cat vagabond and bandit 'This cat has lost all conscience. The cat is a vagabond and a bandit.'

Since Q2 predicts use of the weakest possible form for full definite NPs, when demonstrative determiners do occur there is often a good reason for conveying the stronger cognitive-status information. For example, a demonstrative determiner that requires familiarity but not activation (the proximal form in Russian and the distal form in the other languages) often facilitates

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

comprehension by serving as a signal to the addressee to search long-term memory for a familiar referent. In such cases, which we refer to as 'reminder *that*', the determiner *that* explicitly signals that the referent is familiar and conversationally implicates by Q1 that the referent is not activated. Examples of 'reminder *that*' in English and Russian are given in 62 and 63.³³

(62) Exxon Oil claims it will take several million dollars to clean up THAT OIL SPILL OFF THE COAST OF ALASKA. [beginning of radio newscast]

(63) Russian:

A sošli my na ostanovke kotoraja nazyvalas' and descended we at stop which was.called sorok-vtoroj kilometr, a spustilis' v ETOT KAN'ON. forty-second kilometer and went.down in this canyon
'And we got off at a stop which was called forty-second kilometer, and (we) went down into that (lit. 'this') canyon.'

Redescriptions of an activated referent provide a compelling example of a stronger than strictly necessary form being required. Determiner *this*, which requires activation, becomes crucial in identifying the referent in such cases by serving as an explicit signal that the referent has already been activated. Examples of such uses of *this* N in English, which appear to be restricted to more formal, written genres, are given in 64–66.

- (64) Nearly lost in the polemic was Judge Kennedy himself. That was ironic, because in many ways THIS FORMER SMALL-CITY LAWYER with the stable marriage and three attractive children and the fine reputation appears to personify just those values that made the image of Ronald Reagan so attractive after the convulsions of the 1960's and 1970's. [New York Times 11/15/87, 4:1]
- (65) <u>Poll Return</u> The attachment feature sends THIS INBOUND TAG to the series/1 channel controls to indicate a poll capture for interrupt servicing or nonburst cycle steal servicing. It is not used to signal a burst transfer. [technical document]
- (66) One valuable outcome of these organizational studies was the refinement of our notions of three different approaches that could be incorporated in an automated message filtering system. We refer to THESE TECHNIQUES as the cognitive, social, and economic approaches to information filtering. [Thomas Malone et al., Intelligent information sharing systems. Communications of the ACM 30(5):391 (1987)]

As seen in Tables 2–6, demonstrative forms (both pronouns and determiners) are relatively infrequent in the five languages we investigated. Similar observations have been made by other investigators (cf. Ariel 1988, for example). The analysis we have proposed here provides an explanation for why this should be the case. Since demonstrative pronouns require only activation, they signal a weaker, less restrictive cognitive status than unstressed personal pronouns or zero, which require the referent to be in focus. Demonstrative pro-

³³ Other special effects associated with demonstratives, such as emotional uses discussed, for example, in Lakoff 1974, may also be attributed to quantity implicatures.

nouns are thus less informative than are unstressed personal pronouns, because anything which is in focus is also activated, but not vice versa. And since demonstrative determiners (other than the distal demonstrative in Chinese) require the referent to be at least familiar, they signal a stronger, more restrictive cognitive status than the definite article or zero determiner, which require only identifiability. Demonstrative determiners are thus more informative than the definite article or zero determiner, because anything which is familiar is also (uniquely) identifiable, but not vice versa. The application of Q1 (give as much information as necessary) for definite pronouns and Q2 (don't give more information than necessary) for full definite NPs thus conspires to result in relatively low frequency of demonstratives, both pronoun and determiner, in natural language discourse.

7. CONCLUSION. We have proposed that six implicationally related cognitive statuses are relevant for describing speakers' ability to appropriately use and interpret different forms of reference in natural language discourse. We have shown that each of the statuses is a necessary and sufficient condition for the use of one or more different forms, and that interaction of these form-status correlations with the Gricean maxim of quantity allows us to account for facts regarding the actual distribution of different forms of reference (both within and across languages) which remain unexplained in previous analyses.

The form of referring expressions is only one of a number of linguistic phenomena which have been shown to depend on factors relating to speakers' assumptions about the addressee's knowledge and attention state. Others include intonation, topic/focus marking particles, and a wide range of syntactic structures. (See, for example, Prince 1985, Lambrecht 1986, Sgall et al. 1986, Gundel 1988, Kuno 1989, Givón 1990, Hedberg 1990, Rochemont & Culicover 1990, Steedman 1991, Ward et al. 1991, and other references cited in Green 1989:127–40.) We hope that the theory of cognitive statuses outlined in this paper will contribute to more adequate and insightful analyses of these phenomena as well.

REFERENCES

ARIEL, MIRA. 1988. Referring and accessibility. Journal of Linguistics 24.67-87.

- ATLAS, JAY D., and STEPHEN C. LEVINSON. 1981. It-clefts, informativeness, and logical form. Radical pragmatics, ed. by Peter Cole, 1–61. New York: Academic Press. BACH. KENT. 1981. Referential/attributive. Synthese 49.219–44.
- BIRNER, BETTY J. 1991. Discourse entities and the referential/attributive distinction. Paper presented at the 1991 Annual Meeting of the Linguistic Society of America, Chicago.
- BOLINGER, DWIGHT. 1980. Syntactic diffusion and the indefinite article. Bloomington: Indiana University Linguistics Club.
- —. 1986. Intonation and its parts: Melody in spoken English. Stanford, CA: Stanford University Press.
- BOSCH, PETER. 1983. Agreement and anaphora: A study of the role of pronouns in syntax and discourse. New York: Academic Press.
- CHAFE, WALLACE L. 1976. Givenness, contrastiveness, definiteness, subjects, topics

and point of view. Subject and topic, ed. by Charles N. Li, 25-55. New York: Academic Press.

- (ed.) 1980. The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. Norwood, NJ: Ablex.
- —. 1987. Cognitive constraints on information flow. Coherence and grounding in discourse, ed. by Russell Tomlin, 21–51. Amsterdam: John Benjamins.
- CHASTAIN, CHARLES. 1975. Reference and context. Minnesota studies in the philosophy of science, vol. VII: Language, mind, and knowledge, ed. by K. Gunderson, 194– 269. Minneapolis: University of Minnesota Press.
- CHOMSKY, NOAM. 1971. Deep structure, surface structure and semantic interpretation. Semantics: An interdisciplinary reader in linguistics, philosophy and psychology, ed. by Danny Steinberg and Leon Jakobovits, 183-216. Cambridge: Cambridge University Press.
- CLARK, HERBERT H., and SUSAN E. HAVILAND. 1977. Comprehension and the given-new contract. Discourse processes: Advances in research and theory, volume 1: Discourse production and comprehension, ed. by Roy Freedle, 1–40. Norwood, NJ: Ablex.
- CLARK, HERBERT H., and CATHERINE R. MARSHALL. 1981. Definite reference and mutual knowledge. In Joshi et al., 10–63.
- DAHL, DEBORAH A. 1984. Recognizing specific attributes. Paper presented at the Annual Meeting of the Linguistic Society of America, Baltimore.
- —. 1986. Focusing and reference resolution in Pundit. Paper presented at AAAI-86, Philadelphia.
- DONNELLAN, KEITH S. 1966. Reference and definite descriptions. Philosophical Review 75.281-304.
- ENC, MURVET. 1991. The semantics of specificity. Linguistic Inquiry 22.1-25.
- ERKŪ, FERIDE, and JEANETTE K. GUNDEL. 1987. The pragmatics of indirect anaphors. The pragmatic perspective: Selected papers from the 1985 International Pragmatics Conference, ed. by Jef Verschueren and Marcella Bertuccelli-Papi, 533-45. Amsterdam: John Benjamins.
- FILLMORE, CHARLES J. 1975. Santa Cruz lectures on deixis. Bloomington: Indiana University Linguistics Club.
- —. 1982. Toward a descriptive framework for spatial deixis. Speech, place, and action, ed. by R. J. Jarvella and W. Klein, 31–59. Chichester, NY: John Wiley & Sons.
- FODOR, JANET D., and IVAN SAG. 1982. Referential and quantificational indefinites. Linguistics and Philosophy 5.355–98.
- FOX, BARBARA. 1987. Interactional reconstruction in real-time language processing. Cognitive Science 11.365–87.
- GARROD, S. C., and A. J. SANFORD. 1982. The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. Journal of Semantics 1.21–41.
- GERNSBACHER, MORTON ANN, and SUZANNE SHROYER. 1989. The cataphoric use of the indefinite *this* in spoken narratives. Memory and Cognition 17.536-40.
- GIVÓN, T. 1983. Topic continuity in discourse: A quantitative cross-language study. Amsterdam: John Benjamins.
- —. 1990. Syntax: A functional-typological introduction, vol. 2. Amsterdam: John Benjamins.
- GREEN, GEORGIA M. 1989. Pragmatics and natural language understanding. Hillsdale, NJ: Lawrence Erlbaum.
- GRICE, H. P. 1975. Logic and conversation. Speech acts, ed. by Peter Cole and Jerry Morgan, 41-58. New York: Academic Press.
- GROSZ, BARBARA. 1981. Focusing and description in natural language dialogues. In Joshi et al., 84-105.

- —, and CANDACE L. SIDNER. 1986. Attention, intentions, and the structure of discourse. Computational Linguistics 12.175–204.
- —; ARAVIND K. JOSHI; and SCOTT WEINSTEIN. 1983. Providing a unified account of definite noun phrases in discourse. Proceedings of the Annual Meeting of the Association for Computational Linguistics 21.44–50.
- GUINDON, RAYMONDE. 1985. Anaphora resolution: Short-term memory and focusing. Proceedings of the Annual Meeting of the Association for Computational Linguistics 23.218–27.
- GUNDEL, JEANETTE K. 1978. Stress, pronominalization and the given-new distinction. University of Hawaii Working Papers in Linguistics 10/2,1-13.
- -. 1985. 'Shared knowledge' and topicality. Journal of Pragmatics 9.83-107.
- —. 1988. Universals of topic-comment structure. Studies in syntactic typology, ed. by Michael Hammond, Edith Moravcsik, and Jessica R. Wirth, 209–39. Amsterdam: John Benjamins.
- -; NANCY HEDBERG; and RON ZACHARSKI. 1988. On the generation and interpretation of demonstrative expressions. Proceedings of the International Conference on Computational Linguistics 12.216–21.
- -; ---: ---: 1989. Givenness, implicature and demonstrative expressions in English discourse. Chicago Linguistic Society 25/2.89–103.
- ; ----; ----. 1990. Givenness, implicature, and the form of referring expressions in discourse. Berkeley Linguistics Society 16.442-53.
- Налсоvá, Eva. 1987. Focussing—a meeting point of linguistics and artificial intelligence. Artificial Intelligence II: Methodology, systems, applications, ed. by Ph. Jorrand and V. Sgurev, 311–21. Amsterdam: North-Holland.
- HALLIDAY, M. A. K. 1967. Notes on transitivity and theme in English II. Journal of Linguistics 3.199-244.
- , and RUQAIYA HASAN. 1976. Cohesion in English. London: Longman.
- HAWKINS, JOHN. 1978. Definiteness and indefiniteness: A study in reference and grammaticality prediction. Atlantic Highlands, NJ: Humanities Press.
- —. 1984. A note on referent identifiability and copresence. Journal of Pragmatics 8.649-60.
- . 1991. On (in)definite articles. Journal of Linguistics 27.405-42.
- HEDBERG, NANCY ANN. 1990. Discourse pragmatics and cleft sentences in English. Minneapolis: University of Minnesota dissertation.
- HEIM, IRENE R. 1982. The semantics of definite and indefinite noun phrases. Amherst: University of Massachusetts dissertation.
- HINDS, JOHN. 1973. Anaphoric demonstratives in Japanese. Journal of the Association of Teachers of Japanese 2.1–14.
- HIRSCHBERG, JULIA. 1985. A theory of scalar implicature. Philadelphia: University of Pennsylvania dissertation.
- HORN, LAURENCE R. 1972. On the semantic properties of logical operators in English. Los Angeles: UCLA dissertation.
- —. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature in context. Meaning, form, and use in context, ed. by Deborah Schiffrin, 11-42. Washington, DC: Georgetown University Press.
- ISARD, STEPHEN, 1975. Changing the context. Formal semantics of natural language, ed. by Edward L. Keenan, 287–96. Cambridge: Cambridge University Press.
- JACKENDOFF, RAY. 1972. Semantic interpretation in generative grammar. Cambridge: MIT Press.
- JOSHI, ARAVIND K.; BONNIE L. WEBBER: and IVAN A. SAG (eds.) 1981. Elements of discourse understanding. Cambridge: Cambridge University Press.
- Камеулма, Медими. 1986. Zero anaphora: The case of Japanese. Stanford, CA: Stanford University dissertation.
- KARTTUNEN, LAURI. 1976. Discourse referents. Syntax and semantics 7: Notes from the

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

linguistic underground, ed. by James D. McCawley, 363-86. New York: Academic Press.

- KITAGAWA, CHISATO. 1979. A note on sono and ano. Explorations in linguistics: Papers in honor of Kazuko Inoue, ed. by George Bedell, E. Kobayashi, and M. Muraki, 232-43, Tokyo: Kenkyusha.
- KRIPKE, SAUL. 1977. Speaker's reference and semantic reference. Contemporary perspectives in the philosophy of language, ed. by P. French, T. Uehling and H. Wettstein, 6-27. Minneapolis: University of Minnesota Press.
- KRONFELD, AMICHAI. 1990. Reference and computation. Cambridge: Cambridge University Press.
- KUNO, SUSUMU. 1973. The structure of the Japanese language. Cambridge, MA: The MIT Press.
- ——. 1989. Functional syntax: Anaphora, discourse and empathy. Chicago: University of Chicago Press.
- KURODA, S-Y. 1965. Generative grammatical studies in the Japanese language. Cambridge, MA: MIT dissertation.
- LAKOFF, ROBIN. 1974. Remarks on this and that. Chicago Linguistics Society 10.345-56.
- LAMBRECHT, KNUD. 1986. Topic, focus and the grammar of spoken French. Berkeley: University of California dissertation.
- LEVINSON, STEPHEN C. 1983. Pragmatics. Cambridge: Cambridge University Press.

— 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. Journal of Linguistics 23.379–434.

LI, CHARLES N., and SANDRA A. THOMPSON. 1981. Mandarin Chinese: A functional reference grammar. Berkeley: University of California Press.

LINDE, CHARLOTTE. 1979. Focus of attention and the choice of pronouns in discourse. Discourse and syntax, ed. by T. Givón, 337-54. New York: Academic Press.

LUDLOW, PETER, and STEVEN NEALE. 1991. Indefinite descriptions: In defense of Russell. Linguistics and Philosophy 14.171-202.

MACLARAN, Rose. 1982. The semantics and pragmatics of the English demonstratives. Ithaca: Cornell University dissertation.

MARSLEN-WILSON, WILLIAM; ELENA LEVY; and LORRAINE K. TYLER. 1982. Producing interpretable discourse: The establishment and maintenance of reference. Speech, place, and action, ed. by Robert J. Jarvella and Wolfgang Klein, 339-78. Chichester, NY: John Wiley & Sons.

MILLIKAN, RUTH GARRETT. 1984. Language, thought, and other biological categories. Cambridge, MA: MIT Press.

NUNBERG, GEOFFREY. 1978. The pragmatics of reference. Bloomington: Indiana University Linguistics Club.

PARTEE, BARBARA. 1970. Opacity, coreference, and pronouns. Synthese 21.359-85.

PERLMAN, A. 1969. This as a third article in American English. American Speech 44.76-80.

PRINCE, ELLEN. 1981a. On the reference of indefinite-this NPs. In Joshi et al., 231-50.

—. 1981b. Toward a taxonomy of given-new information. Radical pragmatics, ed. by Peter Cole, 223-56, New York: Academic Press.

-----. 1985. Fancy syntax and 'shared knowledge'. Journal of Pragmatics 9.65-81.

ROCHEMONT, MICHAEL, and PETER CULICOVER. 1990. English focus constructions and the theory of grammar. Cambridge: Cambridge University Press.

- RUSSELL, BERTRAND. 1919. Introduction to mathematical philosophy. London: George Allen & Unwin.
- SEARLE, JOHN R. 1979. Referential and attributative. Monist 62.190-208.
- SGALL, PETR; EVA HAJICOVA; and EVA BENESOVA. 1973. Topic, focus, and generative semantics. Kronberg: Scriptor Verlag.

SGALL, PETR; EVA HAJICOVÁ; and JARMILA PANEVOVÁ. 1986. The meaning of the sentence in its semantic and pragmatic aspects. Dordrecht: Reidel.

REFERRING EXPRESSIONS IN DISCOURSE

- SIDNER, CANDACE L. 1983. Focusing in the comprehension of definite anaphora. Computational models of discourse, ed. by Michael Brady and Robert C. Berwick, 267– 330. Cambridge, MA: MIT Press.
- SPERBER, DAN, and DEIRDRE WILSON. 1986. Relevance: Communication and cognition. Cambridge, MA: Harvard University Press.

STEEDMAN, MARK, 1991. Structure and intonation. Lg. 67.260-96.

STRAWSON, P. F. 1950. On referring. Mind 61.320-44.

WARD, GREGORY; RICHARD SPROAT; and GAIL MCKOON. 1991. A pragmatic analysis of so-called anaphoric islands. Lg. 67.439–74.

WEBBER, BONNIE LYNN. 1983. So what can we talk about now. Computational models of discourse, ed. by Michael Brady and Robert C. Berwick, 331–71. Cambridge, MA: MIT Press.

—. 1988. Discourse deixis and discourse processing. Technical Report. Philadelphia: University of Pennsylvania Department of Computer and Information Science.

WRIGHT, S., and T. GIVON. 1987. The pragmatics of indefinite reference: Quantified text-based studies. Studies in Language 11.1-33.

Jeanette K. Gundel Department of Linguistics 142 Klaeber Ct. University of Minnesota Minneapolis, MN 55455 [Received 11 October 1991; revision received 22 June 1992; accepted 24 October 1992.]

306

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

been met. For example, the definite article *the* signals 'you can identify this', the demonstrative determiner *that* signals 'you are familiar with this, and therefore can identify it', and so on. We thus agree with Garrod & Sanford 1982 and Ariel 1988 that the different forms serve as processing signals to the addressee. However, while these authors (as well as others who have proposed degrees or types of 'givenness') view the statuses signalled by different forms as mutually exclusive, in the model we propose here the statuses are implicationally related (by definition), such that each status entails (and is therefore included by) all lower statuses, but not vice versa. The statuses are thus ordered from most restrictive (in focus) to least restrictive (type identifiable), with respect to the set of possible referents they include. For example, an entity which is in focus is necessarily also activated, familiar, uniquely identifiable, referential, and type identifiable. However, not all uniquely identifiable entities are familiar and not all familiar entities are either activated or in focus.

In presenting this framework, we make only minimal assumptions about reference processing and about the representation of referents in long- and short-term memory. None of these are particularly controversial. The individual statuses are characterized below.

TYPE IDENTIFIABLE: The addressee is able to access a representation of the type of object described by the expression. This status is necessary for appropriate use of any nominal expression, and it is sufficient for use of the indefinite article a in English. Thus, $a \ dog$ in 2 is appropriate, only if the addressee can be assumed to know the meaning of the word dog and can therefore understand what type of thing the phrase $a \ dog$ describes.

(2) I couldn't sleep last night. A dog (next door) kept me awake.

REFERENTIAL: The speaker intends to refer to a particular object or objects. To understand such an expression, the addressee not only needs to access an appropriate type-representation, he must either retrieve an existing representation of the speaker's intended referent or construct a new representation by the time the sentence has been processed. The status 'referential' is necessary for appropriate use of all definite expressions, and it is both necessary and sufficient for indefinite *this* in colloquial English.² Thus, while 2 can have an

² There has been considerable debate in the philosophical and linguistic literature concerning the referential status of both definite and indefinite expressions (cf., for example, Russell 1919, Strawson 1950, Donnellan 1966, Partee 1970, Chastain 1975, Fodor & Sag 1982, and Ludlow & Neale 1991). To the extent that we are concerned here primarily with uses of referential expressions, i.e. with speaker reference, rather than with referential interpretation in a purely semantic sense (cf. Kripke 1977), our work is independent of much of this debate. Thus, like Chastain (1975) and Fodor & Sag (1982), we believe that indefinites may be used either referentially or nonreferentially; but we agree with Chastain (1975), Searle (1979), Bach (1981), and Birner (1991) that definite expressions are always used referentially in the sense that speakers intend to refer to a particular entity in using them—either one they are acquainted with and intend to refer to irrespective of whether the description actually fits (Donnellan's 'referential' use), or one which the description actually fits, irrespective of whether the speaker is directly acquainted with it (Donnellan's 'attributive' use).

The sense of 'referential' that we define here is not to be confused with the sense of 'specific',

interpretation where the speaker intends to say something about a particular dog or one where she is simply asserting that there is at least one dog (next door) that kept her awake, 3 is appropriate only if the speaker intends to say something about a particular dog.³

(3) I couldn't sleep last night. This dog (next door) kept me awake.

UNIQUELY IDENTIFIABLE: The addressee can identify the speaker's intended referent on the basis of the nominal alone. This status is a necessary condition for all definite reference, and it is both necessary and sufficient for appropriate use of the definite article *the*.⁴ Identifiability may be based on an already existing representation in the addressee's memory, as would probably be the case in 4 without the material in parentheses, but, as Hawkins (1978) and others have pointed out, identifiability does not have to be based on previous familiarity if enough descriptive content is encoded in the nominal itself. For example, the phrase *the dog next door* in 4 would be perfectly felicitous even if the addressee had no previous knowledge that the speaker's neighbor has a dog.

(4) I couldn't sleep last night. The dog (next door) kept me awake.

Thus, expressions which are referential but not uniquely identifiable require the addressee to construct a new representation as determined by the content of the referential expression along with the rest of the sentence. For expressions which are both referential and uniquely identifiable, on the other hand, the addressee is expected to construct or retrieve a representation on the basis of

whereby the phrase *a student in the syntax class* in (i) is necessarily specific since it can only have a wide-scope existential reading (i.e. *There is a student in the syntax class who* . . .).

(i) A student in the syntax class cheated on the final exam.

This phrase could be used either referentially or nonreferentially here, since a person who utters (i) 'might be intending to assert merely that the set of students in the syntax class who cheated on the final exam is not empty [the nonreferential reading]; or he might be intending to assert of some particular student, whom he does not identify, that this student cheated [the referential reading]' (Fodor & Sag 1982;356). See Enç 1991 for a discussion of some other senses in which the term 'specific' has been used.

³ A number of researchers have found that referents of indefinite noun phrases introduced with *this* are more likely to be continued in subsequent sentences than referents of phrases introduced with *a* (Prince 1981a, Wright & Givón 1987, and Gernsbacher & Shroyer 1989). Such findings would be expected if speakers always intend to refer to a particular entity when using phrases introduced by *this*, whereas phrases introduced by indefinite *a*, which requires only type identifiability, are ambiguous between a referential and a nonreferential interpretation. Indefinite *this* is very likely an extension of the cataphoric use of the proximal demonstrative, i.e. its use in referring to an object which will not be activated for the addressee until the next sentence is processed, as in *What I wanted to tell you is this. Last night*... (see also Perlman 1969 and Maclaran 1982 for discussion).

⁴ The Givenness Hierarchy thus allows us to define explicitly the notion of definiteness. A noun phrase is definite if its referent is necessarily at least uniquely identifiable. Since 'type identifiable' and 'referential' are the only statuses that don't entail 'uniquely identifiable', it follows that all forms listed under statuses to the left of 'referential' (i.e., all but a and indefinite *this*) are associated with noun phrases that are definite.

the referring expression alone (See Webber 1983 and Millikan 1984 for further discussion).⁵

FAMILIAR: The addressee is able to uniquely identify the intended referent because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has). This status is necessary for all personal pronouns and definite demonstratives, and it is sufficient for appropriate use of the demonstrative determiner *that.*⁶ Thus 5, unlike 4, is appropriate only if the addressee already knows that the speaker's neighbor has a dog.

(5) I couldn't sleep last night. That dog (next door) kept me awake.⁷

ACTIVATED: The referent is represented in current short-term memory. Activated representations may have been retrieved from long-term memory, or they may arise from the immediate linguistic or extralinguistic context.⁸ They therefore always include the speech participants themselves. Activation is necessary for appropriate use of all pronominal forms, and it is sufficient for the demonstrative pronoun *that* as well as for stressed personal pronouns. The pronoun *that* in 6 can thus be used appropriately to refer to the barking of a dog only if a dog has actually been barking during the speech event or if barking had been introduced in the immediate linguistic context.

(6) I couldn't sleep last night. That kept me awake.

Activation is also necessary for appropriate use of the definite demonstrative determiner *this*.

⁵ This does not mean, of course, that the content of the remainder of the sentence cannot assist the addressee in correctly identifying the speaker's intended referent.

⁶ There is a relatively infrequent use of demonstrative *that* which does not require familiarity. For example, the referent of the demonstrative phase in (i) is uniquely identifiable but not necessarily familiar.

 (i) It has great potential value for those who must read technical documents. [message from electronic newsgroup]

We believe this to be a special 'precision' use of the demonstrative *that*, which emphasizes the exhaustiveness of the referent. For example, the presence of *those* in (i) forces and emphasizes an interpretation where the referent is ALL the people that meet the description. There were no instances of 'precision' *that* in the present study, and only 2% of the tokens of demonstrative *that* in our 'previous studies were of this type.

⁷ Our characterization of the distinction between demonstrative determiners and the definite article appears to disagree with that of Hawkins (1978), who proposes that 'identifiability' is a necessary property of demonstratives, but not of the definite article. However, this disagreement may be more terminological than substantive, since Hawkins' definition of identifiability is closer to our 'familiar' than to our 'uniquely identifiable'. Thus, like Hawkins, we maintain that familiarity is not part of the conventional meaning of the definite article. This point is discussed further in §5.

⁸ Cf. Sgall et al. 1973, Chafe 1976, 1987, and Gundel 1978 for earlier uses of the term 'activated'. A number of important questions arise concerning the nature of representations in memory and their relation to (representations of) linguistic forms. For example, does the process of retrieving a representation from long-term memory involve first constructing a representation from the discourse and then somehow matching this one to one in memory, or are the previous representations accessed directly? While such questions need to be addressed in a complete theory of reference processing, they are beyond the scope of the present study.

Both determiner and pronominal *this* require the referent to be not only activated, but speaker-activated, by virtue of having been introduced by the speaker or otherwise included in the speaker's context space (cf. Lakoff 1974, Fillmore 1975, 1982, Halliday & Hasan 1976). The phrase *this dog* in 7 is therefore inappropriate in the context of A's question.

(7) A: Have you seen the neighbor's dog?

B: Yes, and $\left\{ \begin{array}{c} ??this \ dog \\ that \ dog \end{array} \right\}$ kept me awake last night.

But in 8, where the dog has been introduced by the speaker, either *this* or *that* is appropriate.

(8) My neighbor has a dog. $\begin{cases} This \\ That \end{cases}$ dog kept me awake last night.⁹

IN FOCUS: The referent is not only in short-term memory, but is also at the current center of attention. This status is necessary for appropriate use of zero and unstressed pronominals. The entities in focus at a given point in the discourse will be that partially-ordered subset of activated entities which are likely to be continued as topics of subsequent utterances. Thus, entities in focus generally include at least the topic of the preceding utterance, as well as any still-relevant higher-order topics.¹⁰

To the extent that syntactic structure and prosodic form encode topic-comment structure and serve to highlight constituents whose referents the speaker wants to bring into focus, membership in the in-focus set is partially determined by linguistic form (cf. the centering and focusing algorithms of Grosz et al. 1983, Sidner 1983, and Dahl 1986). For example, subjects and direct objects of matrix sentences are highly likely to bring a referent into focus, whereas this is not the case for elements in subordinate clauses and prepositional phrases. Thus, in 9a the bull mastiff is not currently in focus because it has

⁹ When *this* is used to refer to an entity not activated by the speaker, the speaker-activation condition is being exploited to convey special effects, such as solidarity. One such use is in polite interruptions, generally clarification questions, like (i) (see Hedberg 1990).

(i) This is Chris vou're talking about, right? [Frederickson tapes]

Another example is in expressions like *this is true* where, as Georgia Green has suggested to us, the speaker, by using *this* rather than *that*, appropriates an idea introduced by the addressee.

The speaker-activation condition may also be extendable to uses of *this* for extralinguistic objects relatively close to the speaker and for intervals including speech time, and uses of *that* for objects relatively far away from the speaker and for times prior to speech time.

¹⁰ By 'topic' we mean what the speaker intends a sentence to be primarily about. While the topic is often in subject position, it does not have to be. In fact it need not be overtly represented in the sentence at all (see Gundel 1985, 1988). The term 'focus' has been used in two distinct ways in the literature (see Hajičová 1987). We use 'in focus' here to refer to the psychological notion of focus of attention (Hajičová's focus_{A1}—cf. Linde 1979, Grosz & Sidner 1986). This is to be distinguished from the notion of focus as the position of linguistic prominence in the part of the sentence that expresses the comment (Hajičová's focus_L—cf. Halliday 1967, Chomsky 1971, Jackendoff 1972). These two senses of focus are related, however, in that elements tend to be linguistically focussed because the speaker wants to bring them into the focus of attention. In addition, like the topic of a sentence, the referent of a linguistically focussed element is likely to be in focus in subsequent utterances in the discourse.

not been mentioned in the immediately preceding discourse. But since it is introduced in matrix subject position (and is most likely also the topic) in 9a, it is brought into focus, and can therefore be appropriately referred to with either *that* or *it* in 9b. But in 10, where the bull mastiff has been introduced in a prepositional phrase that functions primarily to restrict the referent of the indirect object, reference with *it* is inappropriate.

(9) a. My neighbor's bull mastiff bit a girl on a bike.

- b. $\begin{cases} II's \\ That's \end{cases}$ the same dog that bit Mary Ben last summer.
- (10) a. Sears delivered new siding to my neighbors with the bull mastiff. $(\#\mu)_{s=0}$
 - b. $\begin{cases} \#It's \\ That's \end{cases}$ the same dog that bit Mary Ben last summer.¹¹
 - c. Anyway, this siding is real hideous and ...

While linguistic form plays an important role in determining what will be brought into focus, actual inclusion in the 'in-focus' set depends ultimately on pragmatic factors, and is not uniquely determinable from the syntax. For example, a large wind energy project in 11a is in a syntactic position similar to that of the bull mastiff in 10a, but its referent, unlike that of the bull mastiff, is brought into focus because of its importance in this context. Subsequent reference with the unstressed pronoun *it*, as in 11b, is therefore appropriate here.

- (11) a. However, the government of Barbados is looking for a project manager for a large wind energy project.
 - b. I'm going to see the man in charge of it next week. [personal letter]

3. INFERRABLES. In her seminal work on givenness, Prince (1981b) proposes different types of givenness/newness and suggests that these are related in the following preference hierarchy, though she does not explicitly link the statuses with particular forms.

(12) FAMILIARITY SCALE (Prince 1981b):



An important distinction between the statuses in 12 and those in the Givenness Hierarchy is that the Familiarity Scale does not distinguish between 'activated' and 'in focus'; the status 'evoked' covers both. Furthermore, while statuses in both scales are ranked according to degree of givenness (from most familiar to least familiar), the relation between statuses in the Givenness Hierarchy is one of entailment, while statuses in the Familiarity Scale are mutually exclusive. Some of Prince's statuses correspond to set differences between ours. For example, 'unused' corresponds roughly to 'familiar' but not 'activated'; 'containing inferrable' corresponds to 'identifiable' but not 'familiar';

¹¹ We use # here to indicate unacceptability in the given context.

and 'brand new' corresponds to 'type identifiable' but not 'uniquely identifiable'. As we show below in §6, entailment relations among the statuses allow for a straightforward explanation of the interaction of the Givenness Hierarchy with Grice's Maxim of Quantity in predicting the actual distribution of forms in discourse.

Of particular concern to us here is the category Prince calls 'inferrable' (see also the 'bridging inferences' of Clark & Haviland 1977, the 'associated anaphors' of Hawkins 1978, and the 'indirect anaphors' of Erkü & Gundel 1987). In such cases, the speaker assumes that the hearer can infer an entity 'via logical—or more commonly, plausible—reasoning from discourse entities already Evoked or from other Inferrables' (Prince 1981b:236). For example, the referents of *a whole paragraph* in 13 and *the pulse* in 14 are inferrables.

- (13) [Boss to secretary who just typed an affidavit he is reading] 'Miss Murchison,' said Mr Urquhart, with an expression of considerable annoyance, 'do you know that you have left out A wHOLE PARAGRAPH?' [Dorothy Sayers. Strong Poison (1977)]
- (14) Members of the jury—there is no need, I think, for me to recall the course of Philip Boyes' illness in great detail. The nurse was called in on June 21st, and during that day the doctors visited the patient three times. His condition grew steadily worse... On the day after, the 22nd, he was worse still—in great pain, THE PULSE growing weaker, and the skin about the mouth getting dry and peeling off. [Dorothy Sayers, Strong Poison (1977;21)]

To account for such examples, Garrod & Sanford (1982) distinguish between explicit focus, which contains representations of entities directly mentioned in a discourse, and implicit focus, which contains information from situational scenarios that is not specifically mentioned but is directly relevant to something which is mentioned. Similarly, Chafe (1987) suggests that an entity can be 'semiactive' (in an individual's consciousness but not in focus) by being a member of a 'set of expectations associated with a schema' evoked by the discourse. Thus, for these authors, as for Prince, inferrable entities have a separate cognitive status on a par with different types or degrees of givenness.

- An interesting property of inferrables is that they typically do not allow reference with a pronominal (see Garrod & Sanford 1982:28–29) or with a demonstrative determiner (see Webber 1988:26). *The pulse* in 14 cannot be replaced with *that pulse* or with *it*, for example. Such facts can be accounted for naturally if 'inferrable' is viewed not as a separate cognitive status but rather as a way that something can achieve a particular status by association with an entity that has been activated. We would thus expect inferrables to have different statuses, and to be coded by different forms, depending on the nature and strength of the link between the inferrable and its associated discourse entity. If, as suggested in Garrod & Sanford 1982, hearers/readers do not automatically construct token representations of entities that can be inferred but are not directly mentioned in a discourse, then most inferrables would have a status lower than familiar, as is the case in 13 and 14 above. Thus, in 13 the

the referring expression alone (See Webber 1983 and Millikan 1984 for further discussion).⁵

FAMILIAR: The addressee is able to uniquely identify the intended referent because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has). This status is necessary for all personal pronouns and definite demonstratives, and it is sufficient for appropriate use of the demonstrative determiner *that.*⁶ Thus 5, unlike 4, is appropriate only if the addressee already knows that the speaker's neighbor has a dog.

(5) I couldn't sleep last night. That dog (next door) kept me awake.⁷

ACTIVATED: The referent is represented in current short-term memory. Activated representations may have been retrieved from long-term memory, or they may arise from the immediate linguistic or extralinguistic context.⁸ They therefore always include the speech participants themselves. Activation is necessary for appropriate use of all pronominal forms, and it is sufficient for the demonstrative pronoun *that* as well as for stressed personal pronouns. The pronoun *that* in 6 can thus be used appropriately to refer to the barking of a dog only if a dog has actually been barking during the speech event or if barking had been introduced in the immediate linguistic context.

(6) I couldn't sleep last night. That kept me awake.

Activation is also necessary for appropriate use of the definite demonstrative determiner *this*.

⁵ This does not mean, of course, that the content of the remainder of the sentence cannot assist the addressee in correctly identifying the speaker's intended referent.

⁶ There is a relatively infrequent use of demonstrative *that* which does not require familiarity. For example, the referent of the demonstrative phase in (i) is uniquely identifiable but not necessarily familiar.

 (i) It has great potential value for those who must read technical documents. [message from electronic newsgroup]

We believe this to be a special 'precision' use of the demonstrative *that*, which emphasizes the exhaustiveness of the referent. For example, the presence of *those* in (i) forces and emphasizes an interpretation where the referent is ALL the people that meet the description. There were no instances of 'precision' *that* in the present study, and only 2% of the tokens of demonstrative *that* in our 'previous studies were of this type.

⁷ Our characterization of the distinction between demonstrative determiners and the definite article appears to disagree with that of Hawkins (1978), who proposes that 'identifiability' is a necessary property of demonstratives, but not of the definite article. However, this disagreement may be more terminological than substantive, since Hawkins' definition of identifiability is closer to our 'familiar' than to our 'uniquely identifiable'. Thus, like Hawkins, we maintain that familiarity is not part of the conventional meaning of the definite article. This point is discussed further in §5.

⁸ Cf. Sgall et al. 1973, Chafe 1976, 1987, and Gundel 1978 for earlier uses of the term 'activated'. A number of important questions arise concerning the nature of representations in memory and their relation to (representations of) linguistic forms. For example, does the process of retrieving a representation from long-term memory involve first constructing a representation from the discourse and then somehow matching this one to one in memory, or are the previous representations accessed directly? While such questions need to be addressed in a complete theory of reference processing, they are beyond the scope of the present study.

Both determiner and pronominal *this* require the referent to be not only activated, but speaker-activated, by virtue of having been introduced by the speaker or otherwise included in the speaker's context space (cf. Lakoff 1974, Fillmore 1975, 1982, Halliday & Hasan 1976). The phrase *this dog* in 7 is therefore inappropriate in the context of A's question.

(7) A: Have you seen the neighbor's dog?

B: Yes, and $\begin{cases} ??this \ dog \\ that \ dog \end{cases}$ kept me awake last night.

But in 8, where the dog has been introduced by the speaker, either *this* or *that* is appropriate.

(8) My neighbor has a dog. $\begin{cases} This \\ That \end{cases}$ dog kept me awake last night.⁹

IN FOCUS: The referent is not only in short-term memory, but is also at the current center of attention. This status is necessary for appropriate use of zero and unstressed pronominals. The entities in focus at a given point in the discourse will be that partially-ordered subset of activated entities which are likely to be continued as topics of subsequent utterances. Thus, entities in focus generally include at least the topic of the preceding utterance, as well as any still-relevant higher-order topics.¹⁰

To the extent that syntactic structure and prosodic form encode topic-comment structure and serve to highlight constituents whose referents the speaker wants to bring into focus, membership in the in-focus set is partially determined by linguistic form (cf. the centering and focusing algorithms of Grosz et al. 1983, Sidner 1983, and Dahl 1986). For example, subjects and direct objects of matrix sentences are highly likely to bring a referent into focus, whereas this is not the case for elements in subordinate clauses and prepositional phrases. Thus, in 9a the bull mastiff is not currently in focus because it has

⁹ When *this* is used to refer to an entity not activated by the speaker, the speaker-activation condition is being exploited to convey special effects, such as solidarity. One such use is in polite interruptions, generally clarification questions, like (i) (see Hedberg 1990).

(i) This is Chris you're talking about, right? [Frederickson tapes]

Another example is in expressions like *this is true* where, as Georgia Green has suggested to us, the speaker, by using *this* rather than *that*, appropriates an idea introduced by the addressee.

The speaker-activation condition may also be extendable to uses of *this* for extralinguistic objects relatively close to the speaker and for intervals including speech time, and uses of *that* for objects relatively far away from the speaker and for times prior to speech time.

¹⁰ By 'topic' we mean what the speaker intends a sentence to be primarily about. While the topic is often in subject position, it does not have to be. In fact it need not be overtly represented in the sentence at all (see Gundel 1985, 1988). The term 'focus' has been used in two distinct ways in the literature (see Hajičová 1987). We use 'in focus' here to refer to the psychological notion of focus of attention (Hajičová's focus_{AI}—cf. Linde 1979, Grosz & Sidner 1986). This is to be distinguished from the notion of focus as the position of linguistic prominence in the part of the sentence that expresses the comment (Hajičová's focus_{AI}—cf. Halliday 1967, Chomsky 1971, Jack-endoff 1972). These two senses of focus are related, however, in that elements tend to be linguistically focussed because the speaker wants to bring them into the focus of attention. In addition, like the topic of a sentence, the referent of a linguistically focussed element is likely to be in focus in subsequent utterances in the discourse.

LANGUAGE, VOLUME 69, NUMBER 2 (1993)

addressee is expected to identify the subclass of paragraphs that the referent of *a whole paragraph* belongs to, namely paragraphs in the affidavit that the speaker is holding. But she is not expected to uniquely identify the particular paragraph in question. That is, *a whole paragraph* is referential (and therefore type identifiable), but not uniquely identifiable (and therefore also not familiar). In 14 the referent of *the pulse* is uniquely identifiable, but not familiar. What must be familiar in order for 14 to be felicitous is the knowledge that patients have pulses, but there is no reason to expect that the mention of a patient will automatically evoke a representation of that particular patient's pulse.

Since most instances of inferrables are of the type in 13 and 14, at most uniquely identifiable, but not familiar, our analysis predicts correctly that they generally cannot be referenced by pronouns (which require their referents to be at least activated) or by demonstrative determiners (which require their referents to be at least familiar). But when the link between an inferrable and its associated discourse entity is strong enough to create (or activate) an actual representation of the inferrable, a demonstrative determiner as in 15, or even a pronoun as in 16, is possible.

- (15) We went to hear the Minnesota Orchestra last night. THAT CONDUCTOR was very good.
- (16) There was not a man, woman or child within sight; only a small fishing-boat, standing out to sea some distance away. Harriet waved wildly in its direction, but THEY either didn't see her or supposed that she was merely doing some kind of reducing exercises. [Adapted from Dorothy Sayers, Have his carcase (1932:15)]

In 15 the link between orchestras and their conductors is strong enough to make a conductor familiar in this case. The conductor can therefore be referenced with a demonstrative determiner. Similarly, in 16 the mention of someone waving towards a boat is enough to create and bring into focus a representation of people in the boat. Reference with a pronoun is therefore possible here.

A similar explanation can be given for the contrast between well-known examples like those in 17a-b.¹²

- (17) a. I dropped ten marbles and found all of them, except for one. It's probably under the sofa.
 - b. #I dropped ten marbles and found only nine of them. It's probably under the sofa.

In 17a, reference with *one* creates and brings into focus a representation of the missing marble, thus licensing reference with the unstressed pronoun *it* in the second sentence. In 17b, by contrast, the fact that there is a missing marble is inferrable from the first sentence, but the possibility of inference is not sufficient to create a representation of that marble. Reference with *it* in the second sentence is thus inappropriate. Since demonstrative pronouns and determiners also require familiarity (i.e., they require an already existing mental representation)

12 Quoted in Heim (1982:21) and originally due to Barbara Partee.

of the referent), we would predict that a demonstrative determiner or pronoun, like a personal pronoun, could appropriately refer to the missing marble in 17a but not in 17b. This prediction is correct, as illustrated in 18.

(18) a. I dropped ten marbles and found all of them, except for one.

$$\begin{cases} That \\ That missing marble \end{cases} 's probably under the sofa. \\ b. I dropped ten marbles and found only nine of them. \\ # \begin{cases} That \\ That missing marble \end{cases} 's probably under the sofa. \end{cases}$$

However, a form with a definite article can appropriately refer to the missing marble in both the (a) and (b) examples, given a description which is sufficient to uniquely identify the marble.

- (19) a. I dropped ten marbles and found all them, except for one. The missing marble's probably under the sofa.
 - b. I dropped ten marbles and found only nine of them. The missing marble's probably under the sofa.

These facts are also predicted by our analysis, since a definite article, unlike a personal pronoun or demonstrative, requires only that the referent be uniquely identifiable, but not necessarily that it be familiar.¹³

4. UNIVERSALITY OF THE GIVENNESS HIERARCHY. We have proposed six implicationally related cognitive statuses, and have shown that each of these is necessary and sufficient for the appropriate use of a different form or set of forms in English. In this section, we discuss correlations between cognitive status and different forms of reference in four additional languages—Mandarin Chinese, Japanese, Russian, and Spanish. These correlations are presented in Table 1.¹⁴ Forms are listed below the highest cognitive status necessary for their appropriate use. For example, the proximal demonstrative determiner or pronoun *zhè* in Chinese requires that the referent be at least activated; a zero (\emptyset) pronoun in Spanish requires that the referent be in focus; and so on.

¹³ As far as we can see, the contrast between 18 and 19 remains unexplained under Heim's 1982 analysis. Note that it would be too strong to require that referents of pronouns and demonstrative phrases must be identical to referents introduced in the linguistic or spatiotemporal context. While correctly ruling out examples like 17b and 18b, such a restriction would also rule out acceptable reference with pronouns and demonstratives, as in 15 and 16.

¹⁴ For simplicity, we have excluded proper names, generics, and indefinite plurals from our study. We have also excluded zero NPs in conjoined and nonfinite clauses, in relativized position, and in special uses of language such as English casual speech (e.g. *smells good*) and recipes (e.g. *bake for five minutes*). Thus, we did not include English among the languages which allow zero (Ø) NPs.

We have included only one form on the chart to represent members of a whole class. In English, for example, *it* in the 'in focus' column represents all unstressed personal pronouns, and HE in the 'activated' column represents all stressed personal pronouns. (As Bolinger 1986 and others have noted, *that* is the typical stressed form of *it*.).

Abbreviations used in the glosses are ACC 'accusative', ADV 'adverb', ASP 'aspect', CLF 'classifier', COMP 'complementizer', GEN 'genitive', NOM 'nominative', LOC 'locative', OM 'object marker', TOP 'topic, and Q 'question'.

ni słóż	IN FOCUS	ACTIVATED	Familiar	UNIQUELY IDENTIFIABLE	REFERENTIAL	Type, Identifiable
CHINESE	ø <i>tā</i> `s/he. it`	TA zhè 'this' nèi 'that' zhè N	a correct. Mexand for	nèi N	ar Physics Th Second Arrest	,viN 'aN' ∦N
ENGLISH	it	HE, this, that, this N	that N	the N	indefinite this N	αN
Japanese	Ø	kare 'he' kore 'this' sore 'that' medial are 'that' distal kono N 'this N' sono N 'that N' medial	ano N 'that N' distal	ЯN		
RUSSIAN	Ø on 'he'	ON èto 'this' to 'that'	èto N to N .	v the garb	ЯN	nine et -
Spanish	Ø él `he`	ÉL éste 'this' ése 'that' medial 'aquél 'that' distal este N	ese N 'that N' medial aquel N 'that N' distal	el N 'the N'	ø N un N	l'a N'

TABLE 1. Correlation between linguistic form and highest required status.

4.1. THE UNIVERSALITY OF COGNITIVE STATUSES. The six statuses in the Givenness Hierarchy appear to be adequate for describing appropriate use of demonstratives, articles, and pronouns in the five languages we examined. However, not all six statuses are required for all the languages. Thus, as seen in Table 1, only English has a form (indefinite *this* N) for which the status 'referential' is both necessary and sufficient.¹⁵ The remaining five statuses are not all required for Chinese, Japanese, and Russian, which lack distinct forms for articles. As illustrated in 20–22, a noun with no preceding determiner in these languages can be interpreted as either uniquely identifiable (definite) or merely referential or type identifiable (indefinite).

(20) Chinese:

Tā zài bĭsài zhōng huò jiǎng. he in game during win prize 'He won a prize in a/the game.

(21) Japanese:

Kare wa akai kingyo o hoshii. he TOP red goldfish ACC want 'He wants a/the red goldfish.'

(22) Russian:

V ruke deržali bilety. in hand held tickets

'In (their) hand(s), (they) held tickets/the tickets.'

¹⁵ But see Enç 1991 for a discussion of how her notion of specificity (close to the notion of referentiality described here) is needed to account for the omission of morphological case markers in Turkish, as well as universal constraints on the use of existential constructions.

These languages differ, however, as to which status is unnecessary. As Table I shows, Japanese and Russian have no forms for which the status 'uniquely identifiable' is both necessary and sufficient. In Chinese, by contrast, this status appears to be sufficient for appropriate use of the distal demonstrative determiner $n \partial i$; but Chinese apparently has no determiner which requires the referent to be familiar, but not necessarily activated. Thus, with respect to cognitive status requirements, $n \partial i$ behaves more like the definite article in English and Spanish than like the distal demonstratives in these languages. According to the Chinese speakers we consulted, 23, unlike its counterpart in the other languages, is appropriate even if the addressee has no previous knowledge that the speaker's neighbor has a dog.

(23) Zuótiān wănshàng wõ shuì-bù-zháo. Gébì-de nèi tiáo yesterday evening 1 sleep-not-achieve next.door that CLF gõu jiáo de lìhai. dog bark ADV extremely
'I couldn't sleep last night. The (lit. 'that') dog next door was bark-ing.'

This supports the observation, made for example by Li & Thompson (1981: 131–32), that the unstressed distal demonstrative in Chinese is beginning to function like a definite article.

4.2. CORRELATIONS BETWEEN COGNITIVE STATUS AND LINGUISTIC FORM. In the languages we investigated, the statuses necessary for appropriate use of corresponding forms are the same for all forms except demonstrative determiners and the indefinite article.¹⁶ It is also of interest that when the required status for a corresponding form is the same across languages, the correlation appears not to be arbitrary. Thus, forms which signal the most restrictive cognitive status (in focus) are always those with the least phonetic content, namely unstressed pronouns, clitics, and zero pronominals (cf. Givón 1983, Kameyama 1986, Levinson 1987, and Ariel 1988 for similar observations). In addition, all pronouns (including demonstrative pronouns) require the referent to be at least activated, which is no doubt related to the fact that the minimal descriptive content of a pronoun provides little if any basis for identifying the referent. Finally, in English and Spanish, the languages which have a definite article, the referent of a phrase introduced by this article must be at least uniquely identifiable, but not necessarily familiar. Thus, the phrases headed by the conclusion in 24a and la sombra in 25a do not require previous familiarity. However, the definite article would be inappropriate without the modifiers in these contexts, as in 24b and 25b, because the referent would no longer be uniquely identifiable from the description.

¹⁶ We are only talking about cognitive status here: there are, of course, other conditions which differ across languages, such as restrictions on the use of definite determiners with generics and proper names. Languages also exploit morphological devices such as noun incorporation and syntactic devices such as preverbal vs. postverbal position to signal cognitive status.