# A Computational Model of Reference

**MSc Thesis** (*Afstudeerscriptie*)

written by

**Ori Garin**
(born June 30th, 1976 in Tel Aviv, Israel)

under the supervision of **Dr. Henk Zeevat**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam*.

| **Date of the public defense:** | **Members of the Thesis Committee:** |
|---|---|
| *June 28, 2007* | Dr. Henk Zeevat |
| | Dr. Paul Dekker |
| | Dr. Maria Aloni |
| | Dr. Paul Egré |
| | Prof. dr. Dick de Jongh |

Institute for Logic, Language and Computation

# Contents

**Abstract**

This paper is an interdisciplinary study toward a theory of information status and representation of objects. The motivation is a computational problem: in Natural Language Generation, the task of Generating Referring Expressions (GRE) tacitly assumes that the referent is part of the user's focus of attention and must be distinguished from other entities in that context. The generation of Hearer-New entities is hopelessly flawed due to a problem which Kronfeld [Kronfeld, 1990] identifies as the standard-name assumption: every domain entity is mutually known to all discourse participants. Under this assumption, referring means finding a description which identifies the internal presentation of the intended referent. Hearer-new referents do not fit this model.

Kronfeld addresses the external perspective of reference, which is crucial for overcoming the standard-name assumption, and provides a theory of referring in which the basis of a successful referring act is mutual individuation: the mutual belief of interlocutors that they are both thinking about the same object. The current paper is focused on the epistemic aspect of Kronfeld's model, aiming at a realistic internal representation of domain entities. To this aim, several epistemic distinctions are made. Finally, applications to GRE and linguistics are discussed.

# Chapter 1

# Introduction

Natural Language Generation (NLG) is a research field between Artificial Intelligence and linguistics which aims at designing and implementing systems that can generate coherent natural language text from non-linguistic data. One of its key tasks is the Generation of Referring Expressions (GRE), responsible for constructing a suitable description for a given database entity. A database entity is an internal representation of some (external) entity the system has knowledge about; a suitable description is a list of properties which can be linguistically realized as a definite description, and which together identify the intended referent to the potential user. In short, GRE takes a uniquely-identifiable database input and needs to generate a uniquely-identifiable linguistic output.

But what does it mean to identify the target referent to the user? Every GRE algorithm I'm aware of assumes the following criterion: "a referring expression satisfies the referential communicative goal if it is a distinguishing description" [Dale and Reiter, 1995]. A distinguishing description is an accurate description of the target referent and of no other entities in the *context set*, the set of entities the user is assumed to be attending to. Formally, the standard input to GRE consists of the target referent $r$ and the context set $C$, and the desired output is a set of properties which together distinguish $r$ from the other entities in $C$ [Dale and Reiter, 1995, Reiter and Dale, 2000].

Several assumptions lurk behind this simple definition: Requiring a distinguishing description assumes that the system always knows precisely which object it wishes to refer to, and that referring requires epistemic identification of that object by the user. Moreover, since the user is assumed to be attending to the entities in the context set (including the target referent), these entities must be familiar to the user, if not more than that —at the focus of his attention. Such assumptions are incompatible with several elementary uses of noun phrases, for instance:

(1.1) Indefinite descriptions, whether referring to a specific object or not:

a. I met *a friend* yesterday.

b. I want to buy *a light bulb*.

(1.2) Attributive uses of definite descriptions, and definite descriptions which lack sufficient content to uniquely identify the referent:

a. *The winner in the race* will get 1000$.

b. I went to that restaurant once. *The waitress* was nice. [I didn't say which one I'm referring to]

These limitations of GRE do not just derive from how the task is defined but also from the way objects are represented —more specifically, due to *the standard name assumption* [Appelt and Kronfeld, 1987, Kronfeld, 1990]. This is the assumption that all objects in the domain have standard names that are known to all participants in discourse. In other words, each individual in the system is assumed to represent a single entity in the actual world, and the user is assumed to have a matching representation. So, the act of referring comes down to associating a description with an internal representation. By the standard name assumption, a description which is suitable for an internal representation of an object is automatically suitable both for the actual object and for the user's internal representation of it.

[Kronfeld, 1990] is a computational model of referring which is free from this assumption. Kronfeld makes an important distinction between the internal perspective to reference, which has to do with co-reference, and the external perspective, which has to do with the connection between a referring expression and the actual object being referred to. The standard-name assumption ignores the external aspect completely, assuming that the internal representations match reality perfectly. So, for instance, it offers no answer to how new symbols could be created for new, previously unfamiliar referents. Considering the external perspective —which is crucial for such cases— brings up the age old philosophical problem of reference: how can thoughts and sentences that articulate them be about objects? A computational implementation which does not deal with this question opts automatically for the standard-name assumption.

An accurate model of referring requires accurate internal representations. The current paper extends Kronfeld's theory by providing an account of how objects should be represented in order to avoid the standard name assumption. The paper is structured as follows. Section 2.1 provides a short overview of NLG and an analysis of GRE, discusses some relevant extensions and points out its limitations in detail. Section 2.2 introduces Kronfeld's theory. Section 3.1 presents my main contribution, a model of internal representation of domain entities. Section 3.2 discusses applications of this model in linguistics, by recasting The Givenness Hierarchy from [Gundel et al., 1993] in terms of this epistemic model, and in NLG, by providing an account of how to appropriately refer to referents of a given status.

# Chapter 2

# Background

## 2.1 The Generation of Referring Expressions

Natural Language Generation (NLG) is a subfield of Artificial Intelligence which aims at developing computer systems capable of producing natural language. The input to NLG systems consists of:

**Communicative Goal** the purpose of the text

**Knowledge of User** records what the system knows about the target audience. Such information may include lexical preference, dialects, and even world knowledge and common knowledge within the community of the user, personal knowledge which the system has about a user, history of previous discourses.

**Discourse Status** information about the status and history of the current discourse.

The architecture of NLG normally consists of three stages [Reiter and Dale, 2000]:

| **Document Planning** | **Microplanning** | **Surface Realization** |
|---|---|---|
| - Content determination | - Lexicalization | - Linguistic realization |
| - Document structuring | - Aggregation | |
| | - Generation of referring expressions | |

In Document Planning, the content of the text is selected and organized in a high level, almost non-linguistic manner. Content determination selects the information that will be used, and Document structuring organizes it, usually in a tree-like structure. This is equivalent of picking a topic for an essay and organizing ideas into paragraphs. The selection and organization of content depends on the communicative goal, knowledge of the user and previous discourse.

The Microplanning phase is also known as Sentence Planning. In this phase the so-called 'paragraphs' are organized into 'sentences'. This is the actual linguistic translation of a specific discourse purpose into a sequence of semantic sentences. Lexicalization maps the content to specific linguistic constructions. Aggregation groups information together to sentences, and GRE creates expressions which identify a domain entity to the hearer.

### 2.1.1 The Basics

The Generation of Referring Expressions (GRE) is an important task in the Microplanning phase. Its goal is to identify a domain entity to the user. The generator gets as input a target referent and a set of potential distractors —also known as the contrast set. These distractors are other domain entities which may be "confused" with the target referent. In other words, the task of GRE is to single out the target referent in context, where the context for GRE is a set of entities which the user is assumed to be attending to, and this context set contains exactly the target referent and the distractors. For instance, in a program that talks about icons on a computer screen, the user would be assumed to be observing the screen and the icons would be part of the context, so when the system chooses one icon to refer to, it becomes the target referent and the other icons become the distractors. The task of GRE becomes bringing the target referent into focus by finding a description that rules out the distractors, leaving the target referent as the only referent.

Domain entities are characterized in terms of a collection of attributes and their values. These attribute-value pairs are the properties by which a the target referent can be distinguished from the distractors. Formally, GRE's output is a list of properties (attribute-value pairs) such that the target referent has all those properties and the distractors don't. For instance, consider the following context, where the user is assumed to be attending to three entities:

(2.1)

| $d_1$ | $d_2$ | $d_3$ |
|---|---|---|
| ⟨type, dog⟩ | ⟨type, dog⟩ | ⟨type, cat⟩ |
| ⟨color, white⟩ | ⟨color, black⟩ | ⟨color, white⟩ |
| ⟨size, big⟩ | ⟨size, small⟩ | ⟨size, small⟩ |

There are several possible distinguishing expressions for entity $d_1$ in the context of (2.1) (where $\{d_2, d_3\}$ is the set of distractors), for instance:

(2.2)
a. {⟨type,dog⟩, ⟨color,white⟩}
b. {⟨type,dog⟩, ⟨size,big⟩}
c. {⟨type,dog⟩, ⟨color,white⟩, ⟨size,big⟩}
d. {⟨color,white⟩, ⟨size,big⟩}

It's easy to see for each list contains properties which are all true for $d_1$ but not all true for $d_2$ and $d_3$. So (2.2a-d) all fit the requirements presented so far. Another crucial requirement is the ability to realize the chosen expression

linguistically. So, for instance, example (2.2d) would be rejected as it lacks a 'type' attribute. Only 'type' attributes can be realized as head nouns, hence (2.2d) cannot be realized as a noun phrase as it doesn't offer a head noun.

[Dale and Reiter, 1995] summarizes some of the requirements for optimal outputs of GRE algorithms, based on Grice's cooperation principle [Grice, 1975]

- The referring expression should not include unnecessary information (the Maxim of Quantity).
- The referring expression should only specify properties that have some discriminatory power (the Maxim of Relevance).
- The referring expression should be short (the Maxim of Brevity).
- The referring expression should use basic-level and other lexically preferred classes whenever possible (Lexical Preference).

Another consideration is computational efficiency. Dale and Reiter show that finding the absolute smallest set of properties which distinguish the target referent is an NP-Hard problem, hence impractical. Moreover, according to psycholinguistic experiments, referring expressions in human discourse are not always the shortest ones possible (ibid. p.14)

The Incremental Algorithm, which was presented in [Dale and Reiter, 1995] and used for several extensions of GRE algorithms since then, has the following requirements:

- Every entity is characterized in terms of a collection of attributes and their values.
- Every entity has 'type' as one of its attributes. This is the kinds of properties that are typically realized by head nouns
- The knowledge base may organize some attribute values in a subsumption taxonomy. Such a taxonomy might record, for example, that animal subsumes dog, and that red subsumes scarlet.

A significant amount of work has been aiming at finding an optimal subset of properties/relations, such that the resulting description not only individuates the intended referent, but also is short, does not lead to false implicatures, gives preference to basic-level words, is computationally efficient, uses a salience ordering of the properties etc.

### 2.1.2 Extensions

So far, GRE seems to have little interaction with discourse. The set of distractors is part of the input; the process of building a distinguishing description simply traverses a list of properties, and while this traversal may be context-sensitive, the requirements from the algorithm are always the same. As a result, GRE doesn't take into consideration the information status of the referent or

discourse history. For instance, a recently mentioned (generated) referent can normally be generated by a pronoun or a reduced (non uniquely distinguishing) description. GRE, however, performs the same process for both discourse-old as well as discourse-new referents, always aiming at a complete distinguishing description.

[Krahmer and Theune, 2002] claims that the notion of the contrast set —the set of distractors— is defined rather vaguely ("the set of entities that the user is assumed to be attending to"), and that it needs to be addressed and formalized. For instance, the authors argue, the contrast set may vary depending on the status of the referent: if a referent is discourse-new, it must be distinguished from all other entities in the domain (assuming that all objects in the domain are equally prominent). However, if a referent has just been mentioned, it seems to have a special status, as the speaker's attention is already directed to it. In such a case, a reduced description can often be used —i.e. a description which doesn't necessarily distinguish the referent from the contrast set, but that distinguishes the referent from other equally salient (and more salient) objects. Krahmer & Theune introduce a model based on salience weights to account for the effects of the relative prominence of discourse referents. So, instead of seeking a definite description which matches exactly the target referent (uniquely identifies it), the algorithm seeks a definite description such that, in the set of entities it matches, the target referent is the most salient entity.

A different approach, by Striegnitz ([Striegnitz, 2003]), deals with bridging descriptions, that is, building descriptions for referents which are inferrable ([Prince, 1981]) in the context. In the following example, "the window" is inferrable from world knowledge that rooms have windows. Hence, since only one room is mentioned, the complete description, "the window of the room", can be reduced:

(2.3) I entered the room. *The window* was open.

The main contribution of Striegnitz's work is the ability to generate descriptions for referents which are not part of the immediate context.

Striegnitz takes bridging descriptions as referring to discourse-new entities whose existence can be inferred on the basis of some discourse old entity —*the anchor*— and general knowledge that entities belonging to the ontological category the anchor belongs to are normally related to entities that fit the description. So, in (2.3), the anchor is "the room" and the general knowledge involved is that entities that are rooms are normally related to entities that are windows.

Striegnitz formalizes licensing conditions for bridging descriptions.

- Familiarity of anchor: the anchor is part of the context set.
- Uniqueness of anchor: the description correctly determines the anchor as intended by the speaker.
- Uniqueness of inferrable: it is plausible to assume that a given anchor is related to only one entity that fits the description.

According to the familiarity rule, to make sense of "the window" (in a context lacking windows) the user must be familiar with something which is related to a window —or more precisely, some entity whose existence licenses the existence of a window. The uniqueness rules can be understood by the two following examples, where these rules are broken, resulting in infelicity. In (2.4) the hearer cannot infer from the description which restaurant is the anchor —the anchor is not unique, and in (2.5) the hearer cannot infer a unique page, as a book normally has more than one:

(2.4) There is an Italian restaurant on Market Street and a Chinese restaurant on Church Street. ??*The cook* is excellent.

(2.5) I picked up a book. ??*The page* fell out.

The basic approach just described applies to necessary bridging relations — relations such as necessary part (e.g. room:ceiling) and necessary role (e.g. murder:murderer)—, and it cannot account for cases such as (1.2b) where uniqueness is not required (here, "the waitress" is still felicitous even if the speaker knows there was more than one).

### 2.1.3 Limitations

GRE is a very well-defined task: "take a database entity and build a description which identifies it to the hearer in the given context". However, the underlying assumptions are numerous, and they point to serious limitation of GRE's capabilities. These assumptions can be summarized as follows:

i. Attention: the basis for individuation is the set of entities which the user is assumed to be attending to, hence the target referent must be not only familiar to the user but also activated[1].

ii. Identification: the goal of GRE is identification, and GRE must fulfil it regardless of other goals it might fulfil. If GRE cannot find a distinguishing list of properties then it returns "failure".

iii. Extensionality: each domain entity represent a specific object in the world, and therefore domain entities are exclusive - two different domain entities cannot represent the same object.

The so-called distinguishing description doesn't have to be unique. That is, it doesn't have to identify the target entity against the actual world in the sense that the description "the shortest spy" (supposedly) denotes a unique individual. Rather, the referent must be individuated against the context of the hearer's *attention*; the description must distinguish it from all the entities the hearer is attending to, or to be a bit more precise, the referent must be the most salient entity (Cf. [Krahmer and Theune, 2002]) which fits the description among the entities the hearer is attending to. This presupposes, of course, that the target

[1]This assumption has been relaxed in Striegnitz's work on bridging descriptions.

referent itself is one of those entities. Therefore, GRE generates expressions for entities that are in the focus of attention [Horacek, 1997]. This is the Attention assumption.

The work on bridging descriptions [Striegnitz, 2003] seems to relax this last assumption a bit. Anchors must be in the focus of attention, but the target referent doesn't have to be. In fact, if the anchor is familiar and unique and is only connected to a single entity via the bridging relation, then the target referent doesn't even have to be familiar to the user[2]. However, the type of bridging relations which Striegnitz supports is, to put it in Clark's terms [Clark, 1977], only *forward inferences*. That is, when the anchor (e.g. room) is uttered, the bridge (e.g. window) can already be inferred, as its existence is entailed by the existence of the anchor. Compare this kind of (forward) inference (2.6) with *backward inference* (2.7):

(2.6) We love our new *house*, but *the front door* is broken.

(2.7) We love our new *house*, but *the mortgage* is killing us.

In backward inferences, existence of the bridge is not entailed by the anchor, but is presupposed when the bridging description is uttered: normally, both the bridging relation as well as the anchor must be recognized in order to fulfil the uniqueness presupposition. In the case of forward inferences, there seems to be at least some degree of activation of the bridge, given that it is often expected, as demonstrated by example (2.8) in which the bridge (door) is sufficiently activated to license its omission. In conclusion, the Attention assumption dictates that the target referent must be activated, or at least familiar to the user.

(2.8) I reached *the house* and knocked [on *the door*].

An interesting point that comes out of the work on generating bridging descriptions concerns the representation of domain entities. GRE, as its name suggest, aims at descriptions which are, in Donnellan's sense [Donnellan, 1966], referential rather than attributive (see [Reiter and Dale, 1992], [Horacek, 1997]). Hence, assuming that each entity in the Knowledge Base is a potential target referent for GRE, each entity in the database represents a physical object. This is the Extensionality assumption; this is also part of Kronfeld's standard name assumption, which is presented in the next section.

What's problematic about the Extensionality assumption? Not so much if the context set is representative of the user's attention and the target referent is part of the user's attention. However, when bridging descriptions are considered, an extensional representation becomes questionable. If GRE was to generate an example like (2.6), "the front door" would be a target referent and therefore represented in the knowledge base. But does this mean that all objects that can be inferred from objects represented in the knowledge base, should also be represented in the knowledge base? down to the subatomic level? If, however,

[2]Arguably, the referents of bridging descriptions *are* familiar to the hearer (Cf. [Prince, 1981]).

only bridges that are used in discourse should be represented, then the representation is incomplete, as the attention of the user would have been shifted to the bridge and its immediate context, and that context might not be correctly represented.

If GRE is successful, the resulting description is realized (in English) with a definite article —i.e. as a definite description. If the target referent is genuinely Hearer-New, no description can be found. It could be argued that in such cases an indefinite should be generated. Consider a simple example:

(2.9) I met an interesting photographer yesterday.

If the photographer is Hearer-New, the system may decide to generate an indefinite. However, there are no rules to guide the generation of indefinite descriptions. Having visual memory and perhaps other knowledge of the photographer gives many possibilities to introduce her into conversation, depending on what the speaker wants to say. However, neither the noun "photographer" nor the adjective "interesting" have anything to do with identification; they have everything to do with *introduction* —introducing a new referent.

The problem is that GRE is only geared toward choosing properties that distinguish the target referent from other entities; if identification is not required there are no rules to dictate which properties should be chosen. Intuitively, if we want to introduce a new referent into discourse it is usually not relevant how this new referent is different from existing referents. In conclusion, GRE is limited almost exclusively to specific objects, in contexts where all the relevant objects are mutually known by the system and the user.

## 2.2 Kronfeld's work

Amichai Kronfeld's book, Reference and Computation [Kronfeld, 1990], is a book about applied philosophy of language. It aims to provide a computational answer to the age-old philosophical problem of linguistic reference.

Kronfeld makes a distinction between two perspectives of the speech act of referring: the internal perspective, which is concerned with coreference between symbols, and the external perspective, which is concerned with the connection between symbols and the actual objects they refer to. Consider two simple examples:

(2.10) A man walks in the park. *He* whistles.

(2.11) *Bill* is walking in the park.

From the internal perspective, referring is about matching two symbols, that is, between a noun phrase and a mental representation. In (2.10), the success of the referring act depends on the correct resolution of the pronoun to its antecedent —to the mental representation of it. Similarly, in (2.11) it depends on the identification of "Bill" with the mental representation of Bill —that is,

Bill must to be familiar to the hearer. It is possible to understand a piece of discourse perfectly while not having a clue who the discourse referents are.

The external perspective is concerned with the philosophical problem of reference: "How can thoughts (and sentences that articulate them) be about objects?" (p. 13). In the normal use of (2.10) the link from the mental representation of the man to an actual object is nonexistent (or at least unknown, underspecified), whereas in (2.11) the link is important, because having a representation of Bill presupposes that such a link exists. In other words, the speaker of (2.11) intends to refer to a specific entity, so the success of the referring act depends in some way on the hearer's focusing on that entity, which in turn requires that the hearer's representation of "Bill" actually determines Bill. The link from mental representations to objects is necessary because language is used to represent reality; only the external perspective can explain how language use triggers physical action.

Reference in NL systems is based entirely on the internal perspective, and it works under what is called *the standard-name assumption*: "*all* objects in the domain have standard names that are known to all participants in the discourse" (ibid., p. 4). This assumption trivializes the external perspective as it presupposes that discourse participants always, automatically, know every object in the domain; hence, the act of referring in NL systems boils down to associating one symbol (internal representation) with another symbol (noun phrase). In other words, the system has a label (an internal representation) for each object and the user is assumed to have an equivalent label, so:

- a correct match between a noun phrase and a label (system's internal representation of the object) is equivalent to
- a correct match between the noun phrase and the actual object, which, in turn, is equivalent to
- a correct match between the noun phrase and the user's label (his mental representation of the object).

Hence, according to the standard-name assumption, (1) a referring act is successful if and only if the system matches a noun phrase with an internal representation; and (2) the external perspective is subsidiary (or even trivial).

Any model of referring under the standard-name assumption is a poor model of how referring really works. First, the user might not have the same representations as the system, for instance in the case of Oedipus who has one representation for his mother and a separate one for his wife. Most objects people talk about don't have standard names. When a new object is introduced, it normally has no name —no label—; under the standard-name assumption, a new label might be generated in such a case, but this doesn't explain how the object is recognized nor how the discourse participants could be thinking about the same thing. With the standard name assumption, there's no difference between a variable (e.g. "the farmer" in "if a farmer owns a donkey it beats it") and an actual, physical object (e.g. "Barney" in "if Barney owns a donkey it beats

it") —they are all discourse referents. Finally, if a system combines linguistic and nonlinguistic action, a representation of an object will be of little use if the actual object is not recognized.

### 2.2.1 The problem of linguistic reference

Taking the external perspective as crucial not only for the philosophy of language but also for AI, Kronfeld begins by distinguishing four distinct entities, four aspects of the question of how thoughts and language can be about objects (p. 18):

**Referring Expression** How are referring expressions related to objects?

> This is a semantic question of what the meaning of referring expressions is —what is the (w.r.t. the objects they designate). This aspect is epitomized by the philosophical debate about proper names: according to Frege, names have a sense; according to Searle names are backed by a set of identifying descriptions; while according to Kripke names lack any descriptive content.

**Propositions** What propositions are expressed by sentences containing referring expressions?

> This is a question of logical form. A prominent issue here is whether a proposition is general or singular. Singular propositions contain references to particular things, while general propositions are devoid of such references. For instance, according to the standard analysis, proper names lack any descriptive content ([Kripke, 1980]), therefore sentences containing proper names are prototypically singular while sentences containing definite descriptions are prototypically general, as definite descriptions have a quantificational interpretation, in a Russellian analysis.

**Beliefs** What is the role of mental representation in beliefs about objects?

> This is a question of cognitive psychology. A crucial distinction here is between a de dicto belief, a belief that a certain general proposition is true, and a de re belief, a belief about a particular object that it has a certain property.

**Speech acts** What is the correct analysis of the speech act of referring?

A well-known puzzle deals with the content of de re beliefs. In Kronfeld's version, Ralph points to Wiley on a certain occasion, say at the beach, and says: 'I believe this man is a spy', and on another occasion, say at the supermarket, he points to him again and says: 'I believe this man is not a spy'. It cannot be that those beliefs are the singular propositions **Spy**(wiley) and ¬**Spy**(wiley), as it breaks up the logical premise which Kronfeld names *the trivial principle*: it is impossible to both hold and not to hold the same belief. Kronfeld introduces *modes of presentation* as follows: "If the singular proposition is not the complete

content of Ralph's belief, some element of content is missing. Let Ralph's mode of presentation (of Wiley) be *by definition* that missing element" (p. 39).

Kronfeld then presents the following principles (p. 39-40):

**Basic constraint** For every mode of presentation $M_1$ and $M_2$, if Ralph believes Wiley to be a spy under $M_1$ and doesn't believe Wiley to be a spy under $M_2$, then $M_1 \neq M_2$

**Individuation principle** If Ralph believes Wiley is a spy under $M$, then in any possible world compatible with Ralph's beliefs, $M$ presents exactly one object to Ralph.

The basic constraint is a restatement of the motivation behind modes of presentations (if the basic constraint doesn't hold than the trivial principle doesn't necessarily hold), and from it the individuation principle is derived.

The individuation principle means that "modes of presentation –whatever they are– must carry out individuation within one's network of beliefs" (p. 40). They must present a single object in each world compatible with the agent's beliefs. The individuation principle is a weaker version of what Kronfeld calls *the Fregean principle*, under which modes of presentation determine what a belief is about.

Under both principles, Ralph's mode of representation of Wiley determines the referent in all of Ralph's belief worlds. However, the Fregean principle (and traditional descriptivism) argues that modes of presentation also determine the referent in the actual world. This difference provides a point of departure for the causal and the descriptive theories. Causal theories that acknowledge the individuation principle would most likely be forced to reject the Fregean principle, while Kronfeld's descriptive theory seeks to provide an account of why the Fregean principle does hold.

Kronfeld's thesis is not a refutation of the causal theory or a proof that the Fregean principle must hold, but a philosophically plausible theory of referring which is practical for AI —it is a descriptive *research program*, rather than a theory. However, dealing with both the *external* problem of reference and the *internal* problem of referring leads to a convincing answer for both.

### 2.2.2 The descriptive research program

Taking both the internal and external perspective, Kronfeld observes that Donnellan's attributive/referential distinction has different aspects as well. From the internal perspective, Kronfeld identifies the mental-state criterion —whether the speaker has a particular object in mind. From the external perspective, Kronfeld identifies the denotation criterion —whether the denotation of the description must fit the object uniquely:

**Mental-state criterion** If the speaker has a particular object in mind when he refers, his use of the referring expression is taken to be referential. Otherwise it is attributive.

But "having in mind" itself has two aspects, the epistemic aspect which is out of the speaker's control, and the modal aspect which is affected by the speaker's referring intentions:

**epistemic aspect** *If the speaker knows who or what the referent is, the use is referential, in contrast to his lacking such knowledge.*

**modal aspect** *If the speaker intends the referring expression to be interpreted as a rigid designator, the use is referential, in contrast to considering the referent qua having a particular property.*

**Denotation criterion** *If the description must denote one and only one object for the utterance to be about anything, its use is attributive. Otherwise, it is referential.*

Kronfeld argues that Donnellan's distinction has always tacitly assumed that the two criteria are equivalent: if a use is referential according to the denotation criterion then it is assumed to also be referential according to the mental-state criterion and vice versa. However this is not always the case. Kronfeld presents a variation on Donnellan's famous example (where Smith, the most loveable person in the world is found brutally murdered). John, a well-trained police officer, finds Smith's body next to the murder weapon, a knife. He finds fingerprints on the knife, apparently belonging to the culprit, and exclaims: *"The man whose fingerprints these are, whoever he is, whoever he may be, is insane!"*. As it happens, the real murderer wore gloves and the fingerprints actually belong to a certain Mad Max. In such a case, the use is not prototypically referential, as the speaker has no one in mind, and it is not prototypically attributive, as the description does not denote the intended referent.

The epistemic aspect of the mental-state criterion is related to the notion of de-re belief: to have something in mind is to have a de-re attitude towards it. For instance, the epistemic aspect determines whether (2.13) entails (2.12):

(2.12) Ralph believes that Smith's murderer is Smith's murderer

(2.13) There is someone of whom Ralph believes that he is Smith's murderer

To model the epistemic aspect, Kronfeld regards modes of presentation (MoP's) as mental representation which an agent uses to individuate objects: according to the individuation principle, each MoP determines a unique object in all of the agent's belief worlds. Kronfeld defines an *individuating set* (IS) as an exhaustive set of presentation modes, all taken by the agent to denote the same object. That is:

1. as far as the agent's beliefs and knowledge are concerned, each presentation mode uniquely refers

2. each presentation mode the agent has belongs to a single IS of the agent.

3. as far as the agent's beliefs and knowledge are concerned, all presentation modes of a single IS refer to one and the same object

4. the agent doesn't know of any other presentation mode for that object.

The modal aspect concerns ways of referring as well as logical necessity and possibility. Mainly, it concerns the speaker's choice of fixing the referent, referring to it as a rigid designator vs. choosing to refer to a certain (contingent) aspect of it. Kaplan's artificial operator, **DTHAT**, is an explicit interpretation of the referential use under the modal aspect: it acts as a demonstrative, fixing the referent to the denotation in the actual world of the description to which the operator is applied. So, for instance, (2.14) is (or at least could be) about "The inventor of the light bulb, whoever he is", whereas (2.15) is about Edison.

(2.14) The inventor of the light bulb was a genius

(2.15) **DTHAT**("The inventor of the light bulb") was a genius

The choice a speaker has in referring to Edison as a rigid designator (referential use) vs. qua the inventor of the light bulb (attributive use) is part of the modal intuition.

The denotation criterion concerns choice of referring expression —not about what to say but about how to say it. Once the speaker decides on an IS and, possibly, a relevant aspect (MoP) of it to refer to, she could directly express the descriptive content of the MoP. However, in some cases the speaker can use another MoP, for instance for variety (instead of repeating the same description) or for accomplishing other communicative goals. Whether the use is attributive (the definite description must denote a unique object) depends on whether the description expresses the presentation mode by which the speaker wishes to refer to the referent. If (as part of the modal aspect) the speaker chooses to refer to the referent qua having a certain property —by virtue of a certain MoP—, and the description she uses expresses that MoP, then her use is attributive, otherwise it is referential. So, for instance, if the speaker intends to refer to the IS as a rigid designator (referential use), then the denotation aspect is also referential (as there is no description which expresses the whole IS). This explains Kronfeld's fingerprint example: since "the man whose fingerprints these are" is believed to be the same person as "Smith's murderer", it can be chosen to express the MoP Smith's murderer; the use is referential in the denotation aspect.

### 2.2.3 Belief reports and de-re thoughts

Kronfeld's theory of referring needs to explain how thoughts are related to objects; that is, the problem of de-re propositional attitudes. The descriptive answer to this question is: a thought about an object contains an individuating representation which denotes the object. Kronfeld begins with the most straightforward descriptive approach to de-re thought, which he calls the Fregean view:

(2.16) Ralph has a de re belief about $o$ that it is $F$ if and only if

$$\exists\phi \qquad (\exists x\ \phi(x) \wedge (\forall y\ \phi(y) \rightarrow y = x) \wedge x = o)\ \wedge$$
$$\mathrm{Bel}_{Ralph}(\exists z\ \phi(z) \wedge (\forall y\ \phi(y) \rightarrow y = z) \wedge F(z))$$

The Fregean view takes this as both a necessary and sufficient condition on a thought being de-re. Appropriately, it has been attacked on both grounds.

Kronfeld discusses several objections to the descriptive theory:

**Two-squares** Suppose Ralph is seated before a screen of uniform color, large enough to fill his field of vision entirely. Two squares are displayed on the screen; they are identical in every aspect except their positioning on the screen. Asked to name the squares, Ralph decides to call the top one Alpha and the bottom one Beta. However, unbeknownst to Ralph, he has been fitted with spectacles which invert his field of vision, such that the top square actually denotes Beta. In such a case, Ralph has no individuating representation of the squares

**Making-a-mess** Imagine yourself in a supermarket following a trail of sugar, pushing a cart down one aisle and back up the other, in search of the shopper with the torn sack, until it dawns on you that you are that shopper. "I'm making a mess" is the realization, the belief, which triggers a change in your behavior. As Perry shows ([Perry, 1979]), the change in behavior must be explained by a change in belief. However, the new (de-se) belief cannot contain any individuating representation, and there is no equivalent belief that does contain an individuating representation which (necessarily) explains the change in behavior.

Kronfeld's answer to these objections, and in defense of the descriptive approach, incorporates Russell's notion of knowledge by acquaintance (vs. knowledge by description like in (2.16) expresses). Within Russell's theory, a description by which we know a given object may contain reference to another object which we know by acquaintance, and so, an individuating representation may individuate relative to an object known by acquaintance. This counters the two-squares objection, as Ralph has an individuating representation relative to himself. Perry's example can be explained under the same umbrella of knowledge by acquaintance. There is no individuating representation denoting "I" in "I'm making a mess", but Kronfeld claims that none is needed. Singular propositions (e.g. making-a-mess($perry$)) can be the content of beliefs of objects known by acquaintance. Limiting himself to physical objects, Kronfeld claims that his thesis that individuating representations are necessary for de-re beliefs does not apply to objects known by acquaintance (the self, own thoughts, pain etc.)

Therefore, Kronfeld modifies the condition of de-re belief:

(2.17) Ralph has a de re belief about $o$ that it is $F$ if and only if

1. Ralph has knowledge by acquaintance of $o$ and believes that $o$ is $F$, OR
2. Ralph has knowledge by description of $o$ and believes that $o$ is $F$; that is:

$$\exists\phi \qquad (\exists x\ \phi(x) \wedge (\forall y\ \phi(y) \rightarrow y = x) \wedge x = o)\ \wedge$$
$$\mathrm{Bel}_{Ralph}(\exists z\ \phi(z) \wedge (\forall y\ \phi(y) \rightarrow y = z) \wedge F(z))$$

Note that in the Russellian view, $\phi$ is not restricted to general propositions. It may contain reference to objects known by acquaintance, relative to which $o$ is individuated.

In Kronfeld's view, it is the entire individuating set and not any single individuating (modes of) representation in it that determine the referent. This concept lets Kronfeld explain a problematic argument (from [Schiffer, 1978]) for any descriptive theory [Kronfeld, 1990, p. 134]:

1. Sister Angelica of the Holy Names Convent holds the world record for eating the most spaghetti at one sitting.
2. Ralph mistakenly believes himself to be the holder of the record.
3. Hence, since Ralph believes that he had sex with his wife, he also believes that the holder of the record had sex with her.

The Fregean and Russellian views warrant the absurd conclusion that Ralph believes the Sister Angelica had sex with his wife.

Kronfeld answer is that, according to the modal intuition, Ralph's belief is not about a specific property; that is, Ralph doesn't believe that the holder of the record, whoever s/he is, must have had sex with his wife; rather, Ralph's belief is about the IS containing the MoP "holder of the world record..." as a whole. Mistaken beliefs can of course occur, and in this case, Ralph's IS of himself is contaminated with a MoP which denotes sister Angelica. Kronfeld doesn't offer a conclusive answer as to which IS's can be said to be contaminated but still determine an object (by virtue of the other, non-contaminated, MoP's) and which IS's are confused and determine nothing. Intuitively, in this case the aforementioned IS does determine Ralph.

### 2.2.4 A computational model of referring

The analysis Donnellan's distinction as a three-tiered model provides a general architecture for computational implementation:

| Theoretical aspect | Computational component |
|---|---|
| Epistemic aspect | Database |
| Modal aspect | Planner |
| Denotational aspect | Utterance generator |

The epistemic aspect is covered by the database, which contains individuating sets. There is a distinction between *quasi-permanent* vs. *local* individuating sets. Quasi-permanent individuating sets are part of the system's knowledge representation, and can be created or modified independently of any discourse. Local individuating sets, on the other hand, represent discourse referents and are subsequently short-lived: they get constructed, merged with other individuating sets, modified or discarded during discourse.

The modal aspect is part of the intentions the system (speaker) has in referring to the individuating set. Kronfeld sees communication as goal-oriented behavior. He distinguishes between goals whose recognition by the hearer satisfies them (such as the goal of congratulating the hearer), which he call literal goals, and goals underlying those goals, which are part of the discourse purpose. In the case of referring, the literal goal is that as a result of the hearer's recognition of the noun phrase as a referring expression, the hearer will generate a local individuating set for the referent. The discourse purpose is that the hearer will apply various operations to the newly created individuating set so that it will meet appropriate identification constraints.

One of the important notions in the formalization of the literal goal of referring (see [Kronfeld, 1990, p. 162]) is the notion of mutual individuation. Kronfeld argues against the notion of *mutual belief* as a condition for the success of referring acts. Instead, he claims that what's common in all referring acts is that both speaker and hearer need to mutually believe that they are focusing on the same object. For instance, it could be that there is no presentation mode which the participants share (as in the case of Hearer-New referents).

The discourse purpose requires the recognition of identification constraints. Kronfeld suggests a list of probable candidates (p. 80):

- the requirement that the individuating set contain a perceptual presentation mode.
- the requirement that the new local individuating set be merged with a preexisting quasi-permanent one (Note that this implies that the referent is Hearer-Old).
- the requirement that the new local individuating set be merged with a preexisting local one (Note that this implies that the referent is Discourse-Old).
- the requirement that the individuating set contain one or more presentation modes that are privileged with respect to the goals of the speaker (This corresponds to the intention to refer to the individuating set as a rigid designator)
- the null identification constraint, under which the success of the literal goal fulfils the discourse purpose (e.g. in referring to a Hearer-New referent)

# Chapter 3

# The Epistemic Aspect

## 3.1 Knowledge Representation

Kronfeld defines the epistemic aspect as whether the speaker knows who or what the referent is (or at least has knowledge of the referent). However, while human speakers usually know whether they know who they are talking about, this definition is too vague for a computer implementation. Obviously, we always know something of the referent by virtue of there being a referent to begin with. The problem is that, while Kronfeld can avoid the problem of 'Knowing who' when it comes to hearer identification, there's no escape from at least partially addressing the issue of a system's (speaker's) own epistemic identification when planning a referring expression. So while it is a truism *"one cannot simply decide to possess knowledge"* [Kronfeld, 1990, p. 56], there is some distinction which human speakers make that NL systems will also have to.

Consider Kronfeld's example, poor Oedipus who doesn't know that his wife is really his mother. Oedipus may have rich IS's for both his wife and his mother, but these IS's are distinct and thus represent two different objects in Oedipus's model of reality. Now, we could all say *"Oedipus doesn't know who his wife really is"*. A speaker of this sentence may have the following IS: {mother-of-Oedipus, wife-of-Oedipus}. Referring to this IS attributively (Oedipus's wife), the answer to the question of knowing who Oedipus's wife is, is "Oedipus's mother", which the speaker obviously knows, and Oedipus doesn't. We could imagine that Oedipus has countless MoP's for his wife, but none of them is the answer to the "tricky" question of knowing who in the context. Such trickery can almost always be conceived. For instance, there always something I know about some object that someone else doesn't (e.g. any private belief I have or can make up about the referent).

Kronfeld's theory, however, doesn't require hearer identification to be epistemic, but *pragmatic*. That is, successful referring requires only a mutual belief between the speaker and the hearer that they are thinking about the same object; it doesn't require the hearer's "knowing who". Let us look back at Don-

nellan's example: Smith, the most lovable person in the world, was brutally murdered. Upon seeing Smith's body, one detective exclaims to another:

(3.1) Smith's murderer is insane.

Now, the hearer doesn't need to *identify* any individuating representation or prior knowledge of the referent. For the referring act to be successful the hearer needs to assume a mutual belief that he and the speaker are thinking about the same object. In (3.1), given that the use of the referring expression was attributive, the mutual belief is secured on basis of world knowledge and linguistic knowledge, not on prior knowledge of the referent.

Nevertheless, it does matter whether the *speaker* knows the referent or not, as it determines whether the referring expression is referential or attributive from the epistemic aspect. In (3.1), by using "Smith's murderer" attributively, the speaker *intends* that pragmatic identification will be reached on basis of semantic meaning and not prior knowledge of the referent. This intention is part of planning (Kronfeld's modal aspect), but obviously, one cannot plan something one does not know, and one cannot simply decide to possess knowledge. The speaker of (3.1) somehow *knows* that she doesn't know who the referent is, despite having several descriptions of that referent (e.g. "Smith's murderer", but also "the man whose fingerprints these are", "the man who broke in here today with the intent to steal" etc.) . A computer implementation should also be able to know whether it knows (referential use) or doesn't know (attributive use) the referent it is referring to.

Sometimes hearers need to infer whether the speaker knows the referent or not, but of course this is not always possible. Consider the following well-known example:

(3.2) a student in this class cheated on the final exam.

According to standard analysis, (3.2) has two readings. In the specific reading, the speaker refers to a particular student and asserts that he cheated, and the nonspecific reading is equivalent to saying that it's not the case that no student cheated (i.e. someone must have cheated). In the latter case that the speaker doesn't know who the referent is —technically, there is no referent[1]. However, it is not necessarily the case that in the former reading (specific student) the speaker knows who the referent is. In other words, the specific reading has two possible interpretations: one in which the speaker personally knows who the student is and another in which she has evidence that points to a specific student but doesn't know who that student is (e.g. the tests were examined anonymously).

Jørgensen, who picks up on Kronfeld's work, provides a definition for the term "referring expression" [Jørgensen, 2000]. His account predicts correctly

[1] If (3.2) expresses the belief that

$$\neg\forall x\ (\text{Student}(x) \rightarrow \neg\ \text{cheated-in-final-exam}(x))$$

then there is no object (student) which the speaker is talking about; hence, it is wrong to even discuss "the referent" as such discussion presupposes that there is one

that the nonspecific reading of "a student" in (3.2) is not a referring expression. Appropriately, a substantial part of Jørgensen's work deals with specificity. The two possible interpretations just presented (the two sub-readings of the specific reading) exemplify two of the various definitions of specificity, which Jørgensen describes as (p. 77-78):

(3.3) A. The speaker believes the referent to be a unique, possibly complex, entity, and has the means to identify it

B. The speaker believes the referent to be a unique, possibly complex, entity, but does not have the means to identify it.

Making the distinction between (3.3A) and (3.3B) is part of answering the question of "Knowing Who". However, this part is about the hearer's *epistemic* identification, and therefore not (always) required for the success of the referring act, which according to Kronfeld is based on the hearer's *pragmatic* identification of the referring expression. Sometimes these two go hand in hand. For instance, if someone says "I want to buy a book" then either both (3.3A) and (3.3B) hold (specific reading) or neither does (nonspecific).

Some epistemic distinction does seem to play a role in the ability of a speaker to distinguish between referential use and attributive use, and this distinction is necessary for an account of referring. When a human speaker decides to refer to "The Queen of England" referentially or not, or to "a student" specifically or not, her decision obviously depends on what she knows. A computational model of referring depends on predicting whether the system can be said to have knowledge *of* the referent. Having knowledge *of* something is intuitively similar to having a de-re belief about it. However, according to the descriptive theory, I have a de-re belief about the shortest spy, given the description I have, the MoP, which denotes him. As a generalization of this argument, I have a de-re belief of the denotation of any unique description, past, present, future, unbounded. Hence this account doesn't offer an answer to the question. The question is: Why is it that to the current Queen of England I am able to refer referentially, but to the actual shortest spy I must refer attributively?

### 3.1.1 Two kinds of presentation modes

I see the backbone of Donnellan's *pragmatic* distinction between referential and attributive uses as an *epistemic* distinction between what I call *extensional* MoP's and *intensional* MoP's. Part of the reason I use these terms is in order to separate *use* of referring expressions (which may be attributive or referential) with *representation* (and belief) of entities. For the sake of introduction let's assume that MoP's only contain descriptive content; that is, they are of the form $\lambda x \phi(x)$, where $\phi(x)$ is a general proposition.

From the epistemic aspect, an intensional MoP exists by virtue of a belief about the property its descriptive content expresses; namely, that it uniquely denotes. In other words, an intensional MoP corresponds to a belief that the MoP singles out a unique object, but the belief is not based on knowledge

of the extension of the MoP but on knowledge of its intension. Conversely, an extensional MoP exists by virtue of a belief about some particular object; namely, that that object has the property expressed by descriptive content of the MoP. So an extensional MoP also corresponds to a belief that the MoP singles out a unique object, but the belief is not based on knowledge of the intension of the MoP but on knowledge of its extension.

It seems tempting to associate extensional MoP's with de-re beliefs and intensional MoP's with de-dicto beliefs, but that would be wrong. The distinction between de-re and de-dicto beliefs is not an epistemic but an *ontological* one. First, de-re beliefs entail their de-dicto counterparts: $\exists x\ B\phi(x) \rightarrow B\exists x\ \phi(x)$. The real distinction is between beliefs about objects that are actual (real) and beliefs about objects that are nonfactual (imaginary). Both extensional MoP's and intensional MoP's are believed to have a denotation —that is, the object represented is believed to exist. However, the object which is represented by the MoP, whatever/whoever it may be, may not actually be. Both extensional MoP's as well as intensional MoP's can represent actual entities, expressing a de-re belief that they exist, and both can represent nothing, expressing a de-re belief about non actual entities —hence not a de-re belief in the actual world. The truth is not in the mind the of the agent, but "out there".

For example, consider two of my own MoP's: "the shortest spy" and "Kripke", associated with the two following beliefs.

(3.4) I believe that there actually exists a unique individual who is the shortest spy.

(3.5) I believe that Kripke actually exists

Are these two beliefs de-re or (only) de-dicto? Do they differ in that respect? Certainly they are expressed differently. For instance, asserting "I believe that the shortest spy actually exists" sounds somewhat *vacuous* and therefore inaccurate as a report of the belief reported in (3.4). The assertion is vacuous because the description "the shortest spy" presupposes the existence of one, yet the utterance explicitly asserts the same thing —the utterance asserts what it presupposes, hence it asserts nothing.

Concerning the shortest spy, "the shortest spy" most probably denotes a unique entity, but it might not (think cloning, parallel universes). However, regardless of whether "the shortest spy" denotes, (3.6) is false:

$$\exists x\ \text{Bel}(\text{shortest-spy}(x)) \tag{3.6}$$

The standard analysis is: since I don't know who the shortest spy is, in one of my belief worlds it could be John and in the other Bill; so, there is no one in the actual world who is the shortest spy in all my belief worlds. Nevertheless, according to Kronfeld's descriptive theory, knowledge by description is sufficient for de-re belief. The de-re belief (provided there is a unique shortest spy) is expressed as follows:

$$\exists x\ \text{shortest-spy}(x) \quad \wedge \quad \text{Bel}\left(\exists x\ \text{shortest-spy}(x)\right) \tag{3.7}$$

This looks like an accurate expression of my knowledge (belief) of the shortest spy, in case there is such a unique object in the actual world. Namely, there is someone who is the object of this existence belief. On the other hand, my belief is not *of* that object that it exists but of the description that it uniquely denotes.

On the other end, if "Kripke" denotes an entity in the actual world, then I have a de-re belief and (3.8) would be true.

$$\exists x \;\; \text{Bel}(x = \text{Kripke}) \tag{3.8}$$

If, however, it turns out that Kripke was a pseudonym of a group of philosophers working together, then *in the actual world*, I have no such de-re belief, as in the actual world there is no such object and (3.8) would be false.

This analysis of (3.8) is, however, problematic —it only works if we assume that proper names are rigid designators in belief contexts. But, it is not a name that makes or breaks a de-re belief but the connection with the actual object. In the case of Ralph seeing Wiley on two occasions believing he's a spy on one occasion (say, on the beach) and not a spy on the other (say, at the supermarket), the existence beliefs would be (assuming Ralph is not dreaming):

$$\exists x \;\; \text{Bel}_{Ralph}(x = \text{man-on-beach}) \tag{3.9}$$

$$\exists x \;\; \text{Bel}_{Ralph}(x = \text{man-in-supermarket}) \tag{3.10}$$

However, there is something wrong with expressing de-re beliefs in this way. Suppose Ralph actually knows Wiley and recognizes him in the supermarket but not on the beach. If names are rigid designators, then Wiley denotes the same object in all of Ralph's belief worlds, but then this leaves out the question of how Ralph believes the man on the beach to be a spy. That is, if (3.9) is true, then given that Wiley is the man in the actual world, and that proper names are rigid designators, Wiley is the man on the beach in Ralph's belief worlds.

[Aloni, 2001] solves this problem by positing pragmatic conceptual covers which mediate between the actual object and the objects in the belief worlds. Her work provides a formalization of the idea that different methods of identification are relevant or salient in different contexts, and they determine the semantic analysis of sentences. In this case, we can view conceptual covers as a special kind of counterpart relations —any given *extensional* MoP determines a conceptual cover. For instance, a MoP which is a proper name determines the conceptual cover of names —in which names are rigid designators. Back to our example, at the beach, the conceptual cover relevant for Ralph's belief about Wiley —the connection between the actual Wiley and the spy in each of Ralph's belief worlds— is perceptual. Under this conceptual cover, names are not rigid designators, and Wiley in the actual world is not Wiley in all possible worlds compatible with Ralph's belief.

In conclusion, the distinction between extensional and intensional MoP's dictates *how* the de-re/de-dicto distinction is resolved, and therefore also how the relation between a belief about an object and that object is determined.

Extensional MoP's are de-re or de-dicto based on the causal theory: the relation between belief and object is determined by a causal link. If the chain ends in an actual object then the belief that the object exists is de-re. But the chain might end up with someone's imagination or misconception, in which case the belief is not de-re as the object of the belief actually does not exist.

Intensional MoP's are de-re or de-dicto based on the descriptive theory: the relation between belief and object is the relation of denotation (of the MoP). If the description denotes a unique individual, then the belief is de-re. Beliefs corresponding to intensional MoP's are not *of* objects, but of the meaning of the description. Therefore (3.4) is not a belief of someone that he is the shortest spy. There is no entity which is the object of my belief.

### 3.1.2 Mutual individuation

NLG systems normally contain a user model containing knowledge about the user's knowledge, beliefs and preferences [Reiter and Dale, 2000]. [Jørgensen, 2000] extends Kronfeld's theory by positing such a user model, standing parallel to the system's internal model. In his model, the system maintains:

- a local internal model (containing the system's local IS's)
- a local user model (representing the user's local IS's)
- a quasi-permanent internal model (containing the system's quasi-permanent IS's)
- a quasi-permanent user model (representing the user's quasi-permanent IS's)

With a user model, the system can model the effects the use of a referring expression bears on the user. If the system decide to generate a referring expression for a given IS, then it needs to know whether this IS already exists in the user database (either the local or quasi-permanent user model). This is an implementation of the hearer-new vs. hearer-old distinction. Intuitively, if the system knows that the user knows the referent, the IS is hearer-old and therefore can be generated with a definite.

However, there are several reasons why positing a user-model is problematic. First, the user model is part of the system's internal model —the user model only contains what the system knows that the user knows. If the system's internal knowledge is disjoint from its knowledge/assumption of the user's knowledge, then this internal knowledge must be very basic —limited to facts. Otherwise, if the internal model already contains the user's assumed knowledge, positing an extra user model is superfluous and unwarranted. It also raises a difficult issue of identity —it's not clear how the system would know whether an internal IS matches a user IS.

Moreover, if a separate user model is required then it could not be all that is required. Since hearers draw inferences (and implicatures) from speech acts, a generation system also needs to reason about the user's knowledge of the

system's knowledge in order to produce text which will be understood correctly —e.g. which doesn't lead to false implicatures. The problem with a separate user model is therefore: either the internal model already contains everything the user model contains, or a user model is not enough: if a separate model is needed for reasoning about the system's knowledge of the user's knowledge, then another separate model is needed for reasoning about the system's knowledge of the user's knowledge of the system's knowledge. And so forth ad infinitum

The point I'm getting at is not new or surprising. It is not merely the speaker's knowledge of what the hearer knows that plays a role in referring, but the speaker's assumptions of the *common ground*: "Two people's common ground is, in effect, the sum of their mutual, common or joint knowledge, beliefs and suppositions" [Clark, 1996, p.93].

The actual common ground is an abstract, theoretical construct which represents mutual agreement, hence it is external —not a mental model. As Clark notes (given the background of him and his son observing a conch shell together):

> Common ground is not information I have by myself or that my son has by himself. Only an omniscient being can say "It is common ground for the two of them that there is a conch shell between them". [Clark, 1996, p.96]

Therefore, the actual common ground has explanatory force only in hindsight, only after it is already known what piece of information was or wasn't common ground in a particular case.

According to Clark's theory of contribution to discourse [Clark and Marshall, 1978, Clark and Schaefer, 1989], each participant in conversation constructs and maintains a representation of what he believes to be mutually believed by all the participants. This may be called the *subjective* common ground —a mental model, a representation an agent has of what he believes the objective common ground is. The sum of each participant's subjective common ground determines the actual common ground. Namely, the actual, objective, common ground, taken as a set of propositions, is the intersection of the interlocutors' subjective common grounds. If Clark assumes that it is common ground between him and his son that $p$ and his son assumes so too, then $p$ really is common ground.

To make things formal, let $\mathcal{A}$ be a set of agents participating in discourse. Let $\mathsf{CG}_i^{\mathcal{A}}$ denote the set of propositions which the agent $A_i$ believes to be common ground among $\mathcal{A}$ —$A_i$'s subjective common ground. Assume that the common ground among $\mathcal{A}$ is represented as an infinite set of *belief sentences*:

$$\phi \in \mathsf{CG}^{\mathcal{A}} \iff \mathrm{Bel}_{x_1}\mathrm{Bel}_{x_2}\ldots\mathrm{Bel}_{x_n}\phi \text{ is true for any } n \geq 1, \forall i \leq n\colon x_i \in \mathcal{A}$$

then the subjective common ground of agent $a$ is the subset of that set in which all sentences begin with $\mathrm{Bel}_a$:

$$\phi \in \mathsf{CG}_a^{\mathcal{A}} \iff \mathrm{Bel}_a\mathrm{Bel}_{x_1}\mathrm{Bel}_{x_2}\ldots\mathrm{Bel}_{x_n}\phi \text{ is true for any } n \geq 0, \forall i \leq n\colon x_i \in \mathcal{A}$$

Hence, the objective common ground is the intersection of the interlocutors' subjective common grounds:

$$\mathsf{CG}^{\mathcal{A}} = \bigcap_i \mathsf{CG}_i^{\mathcal{A}} \tag{3.11}$$

That is, if $\phi$ is believed to be common ground by every agent in $\mathcal{A}$, then it really is common ground, and viceversa.

How does the notion of common ground apply to referring acts? According to [Clark and Marshall, 1978], the beliefs involved with successful referring are not restricted to the speaker's beliefs about the hearer's beliefs, but require instead an infinite set of beliefs. Clark and Marshall's thesis is that a speaker successfully refers to referent $R$ by description (term) $t$ if and only if both speaker and hearer believe that the following statements are true:

(3.12)
1. $t$ is $R$

2a. $S$ believes that $t$ is $R$

2b. $H$ believes that $t$ is $R$

3a. $S$ believes that $H$ believes that $t$ is $R$

3b. $H$ believes that $S$ believes that $t$ is $R$

...

Kronfeld finds this requirement too strict, and argues for what he calls *mutual individuation* [Kronfeld, 1990, p. 155]:

> The mutual belief that is crucial for referring in general is not what is mutually believed about the *referent*. Rather, it is what the speaker and hearer mutually believe about *each other*. What is common to *all* referring acts is that *both speaker and hearer need to mutually believe that they are focusing on the same object*.

I think Kronfeld's analysis is correct for the following reason: the propositions in (3.12) are singular propositions: they use a single symbol —$R$— to refer to both the speaker's and the hearer's beliefs; therefore, Clark and Marshall's proposal suffers from the standard-name assumption.

The speaker must have a MoP which the description $t$ expresses: it could be a local MoP, constructed before or during the utterance (in which $t$ is uttered), and it could be a quasi-permanent MoP. In order to use $t$ to refer to $R$ the speaker needs to believe that the interlocutor has an equivalent MoP, that these two MoP's co-refer, and that it is common ground that they co-refer. Mutual individuation can also be explicitly spelled out, as an infinite set of MoP's all belonging to the same IS. Consider a subset of my IS representing Kripke[2]:

$$\{\text{Kripke}, \iota x \text{ author-of}(x, \text{Naming-and-Necessity})\}$$

[2]The iota ($\iota$) operator encodes uniqueness [Partee, 1987].

Now, if I want to discuss Kripke's view of proper names with a colleague, there is an infinite set of MoP's which are part of my IS of Kripke:

$$\iota x \text{Bel}_{you}(\text{author-of}(x, \text{Naming-and-Necessity}))$$
$$\iota x \text{Bel}_{you}\text{Bel}_I(\text{author-of}(x, \text{Naming-and-Necessity}))$$
$$\iota x \text{Bel}_{you}\text{Bel}_I\text{Bel}_{you}(\text{author-of}(x, \text{Naming-and-Necessity}))$$
$$...$$

However, these MoP's are not —and could not all be— explicitly represented in my mind, as they all stem from my belief that the MoP "author of Naming and Necessity" ($\iota x$ author-of$(x, \text{Naming-and-Necessity})$) is common ground —this MoP is part of my subjective common ground. This common ground can be expressed as follows:

$$\iota x \text{ author-of}(x, \text{Naming-and-Necessity}) \in CG_S^{S,H} \quad \text{if and only if}$$
$$\text{Bel}_S \exists! x \bigwedge_{n \in \mathbb{N}} \bigwedge_{A_i \in \{S,H\}} \text{Bel}_{A_1} \ldots \text{Bel}_{A_n} (\text{author-of}(x, \text{Naming-and-Necessity}))$$

Expressing mutual individuation in this way is necessary in order to make sure that the MoP mutually individuates the entity the speaker believes it refers to. The fact that it can only be expressed as an infinite formula is no more or less psychologically implausible than expressing common ground of a proposition as an infinite set of propositions.

Mutual individuation should also work with intensional MoP's, as there is a sense in which the same object, *whoever it is*, is mutually believed. However, the object itself is arbitrary, that is, it is determined by denotation —outside of any agent's mental model. It doesn't make sense, therefore, that the object itself is mutually believed. What plays a role here, I suggest, is mutual belief about the intension of the MoP rather then its extension. And this mutual belief is simply that the MoP uniquely denotes. For instance:

$$\iota x \text{ shortest-spy}(x) \in CG_S^{S,H} \quad \text{if and only if}$$
$$\forall n \in \mathbb{N}; A_i \in \{S, H\}\colon \text{ Bel}_S\text{Bel}_{A_1} \ldots \text{Bel}_{A_n} (\exists! x \text{ shortest-spy}(x))$$

Describing the common ground in terms an infinite set of belief statements (or an infinite belief statement) is useful for getting the message across in a standard logical way. However, it could not be a model of how common ground is actually represented (cf. [Clark, 1996]), nor a plausible model for computer implementation —updating the common ground is not an infinite process. For the common ground to be updated, it seems that some extralinguistic assumptions are required. According to Clark and Marshall, in the case of immediate physical copresence, "simultaneity, attention, and rationality" are required [Clark and Marshall, 1978, p.61]. And immediate physical copresence is actually the most straightforward of the contexts which the authors present.

Common ground is interesting both to investigate as well as to model. This work, however, doesn't aim to contribute to any of these directions beyond

this point. Having laid out what I think is the basis of (referential) subjective common ground, I take it from this point as a primitive, a binary epistemic feature: a MoP can either be common ground or not common ground. If a MoP is common ground then it is Hearer-old, and the appropriate belief statement applies —depending on whether the MoP is extensional or intensional.

### 3.1.3 Presentation modes in discourse

Kronfeld makes a distinction between the quasi-permanent database which corresponds to long-term memory, and the local database which corresponds to short-time memory and serves as discourse model. Respectively, there are two kinds of MoP's (and also IS's): quasi-permanent MoP's and local MoP's. According to the individuation principle, MoP's must uniquely identify an object to the agent; however, nominal expressions in discourse do not always have enough content (and context) to uniquely identify. Instead of choosing a different epistemic representation altogether for discourse, I would like to propose that *local* MoP's should be allowed to violate the individuation principle, and I introduce another epistemic feature called *determinacy*: a MoP is determinate if and only if the agent believes that the content of the MoP determines the referent —in other words, if and only if the individuation principle holds for this MoP.

As an example of nondeterminate intensional MoP's, consider the following specimen of the famous donkey sentences:

(3.13) if Pedro owns *a donkey*, he beats it

It should be obvious that the use of the proper name "Pedro" introduces a local MoP into the common ground, but, I think, so does the use of "a donkey", despite the fact that it is only a variable under quantification. One reason is that it is referred back to by the use of the pronoun in (3.13). Furthermore, the hearer can refer back to it, for instance by asking: "What if it's very young?", read "if Pedro owns a donkey and the donkey is very young, does Pedro still beat it?". It is clear that the use of "a donkey", however it may be represented, must have a limited scope and must be constrained. Unlike the MoP of "Pedro", this MoP cannot be freely referred back to.

The MoP introduced by "a donkey" has no place in the quasi-permanent database. One reason is economy: anything we might know about this MoP is entailed by the following standard interpretation of (3.13):

$$\forall x(\text{donkey}(x) \wedge \text{owns}(\text{Pedro}, x) \rightarrow \text{beats}(\text{Pedro}, x))$$

The use of "a donkey" in (3.13) is nonspecific, hence there is no (specific) entity which can be stored; there is only a variable. The MoP which represents "a donkey" can be called *nondeterminate*. Such MoP's don't determine a unique object to the agent. Uses of intensional nondeterminate MoP's are normally not considered as referring expressions. This depends, of course, on how the term 'referring expression' is defined, but at the minimum, there must be something which is referred to —a specific referent— (Cf. [Jørgensen, 2000]). Following

Kronfeld (and Jørgensen), I view the property of being a referring expression as a feature of noun phrases, and while I do not seek to provide a definition of this term, the current analysis suggests that intensional nondeterminate MoP's are not referring expressions, and that all other types of MoP's are, or at least could be[3].

What sense of determinacy could be applied to extensional MoP's? Let's consider a straightforward example:

(3.14) Let me tell you about *an interesting conversation* I had yesterday.

For ease of discussion, call the speaker Yuval and the hearer Ran. According to Kronfeld's analysis, such cases are referring expressions, which, upon their recognition as such, cause the hearer to generate a new local MoP:

> The one and only $x$ such that $x$ is the object the speaker wants to say something about.

In terms of the current analysis, Ran's new *common ground* MoP can be something such as:

(3.15) The one and only conversation which Yuval wants to say something about".

In the current discourse, (3.15) does make sense. But it makes less sense as a quasi-permanent MoP. According to the individuation principle, MoP's must present a unique object to the agent. For (3.15) to satisfy the individuation principle, some reference to the referring act in which (3.14) was uttered seems to be required. Despite finding it questionable or at least artificial that MoP's should be forced into uniqueness in such a way, I concede that the MoP, which represents what the use of "a conversation" in (3.14) refers to, can be made unique. This is the essence of mutual individuation —assuming that it is common ground that Yuval has a specific object in mind, mutual individuation is secured.

I believe determinacy is nevertheless relevant for extensional MoP's —namely, as a feature of a speaker's local MoP. Determinacy distinguishes between MoP's which are intended to uniquely identify an object to the hearer, and MoP's which are intended to just introduce a new referent into discourse without the purpose of *epistemically* identifying that referent. Definite descriptions are prototypical examples of determinate MoP's and indefinite descriptions are prototypical examples of nondeterminate MoP's, as the following examples from [Abbott, 2005] suggest:

(3.16) i. Mary asked *the oldest student in the class* to explain everything.

ii. * Mary asked *an oldest student in the class* to explain everything.

[3] An alternative definition would be to call intensional nondeterminate MoP's variables, other intensional MoP's denoting description, and reserve the term 'referring expression' to extensional MoP's.

However, the following examples suggest that nondeterminate MoP's can sometimes be realized with definite descriptions:

(3.17) Yesterday I went to the new Italian restaurant around the corner. *The waitress* was nice.

(3.18) *The kids outside* are making too much noise and I can't study.

In these examples, uniqueness or maximality are possible, but not necessary. In the first example, the waitress might be the only waitress in the restaurant but it might not. Moreover, the context does not suggest any unmentioned property which necessarily complements the description and makes it unique: both "the waitress who served you" and "the waitress who took your order" may be wrong; similarly "the waitress you thought was nice" could also be non-unique. The second example can be explained along similar lines. A denoting description must be relative to the referring act.

For the current purposes, I suggest the following distinction: there are two different kinds of local extensional MoP's, deriving from two different kinds of intentions a speaker may have in referring to extensional MoP's. On the one hand, there are MoP's that are intended to introduce a new referent into the common ground, without identifying it —these MoP's are nondeterminate. On the other hand, there are MoP's which are intended to identify an object to the hearer —these MoP's are determinate. Note that identification can be by virtue of prior knowledge (familiarity) or by virtue of a description which uniquely determines the referent (uniqueness).

## 3.2 Applications

It is widely recognized that (the choice of) the form of a referring expression depends on the assumed cognitive status of the referent. For instance, pronouns are normally used to refer to salient —recently mentioned— referents, while indefinites are normally used to refer to Hearer-new referents. The cognitive status of a referent is based on the assumptions a speaker can make about the knowledge and attentional state of the hearer, and encompasses issues of familiarity/givenness (Cf. [Prince, 1981], [Gundel et al., 1993]), activation/saliency (Cf. [Ariel, 1990]), and discourse structure (Cf. [Grosz and Sidner, 1986]).

With Kronfeld's 3-tiered model in mind, we can consider the information status of a MoP as part of the epistemic aspect, and define it in terms of the distinctions argued for above. The information status of a MoP not only affects the denotation aspect —that is, the form of referring expression used— but first and foremost the modal aspect —the intentions a speaker may have in realizing —conveying— this MoP.

In this section, the Givenness Hierarchy presented in [Gundel et al., 1993] is recast in terms of the model of reference and the hierarchy is refined. Subsequently, the statuses in the hierarchy are discussed from the modal aspect and it is shown how they correlate to the discourse purpose of referring. And finally, the implication and applications to GRE are discussed.

### 3.2.1 The Givenness Hierarchy

Gundel, Hedberg & Zacharsky (GHZ henceforth) present a *Givenness Hierarchy* with six cognitive statuses which stand in an entailment relation:

| in focus | > | activated | > | familiar | > | uniquely identifiable | > | referential | > | type identifiable |
|---|---|---|---|---|---|---|---|---|---|---|
| *it* | | *that*<br>*this*<br>*this N* | | *that N* | | *the N* | | indef.<br>*this N* | | *a N* |

Each status is, according to GHZ, necessary and sufficient for its corresponding form, and each status entails all its lower statuses, which is to say that, looking at cognitive statuses as properties of the referent, a referent with a given status property has also all the status properties which are lower in the hierarchy, and it is therefore sufficient for their corresponding forms.

As a short introduction: type-identifiable means the addressee is expected to be able to access a representation of the type of the referent. Referential means that the speaker refers to a particular object —the addressee is expected to either retrieve a representation of that referent, or create a new one. Uniquely identifiable means the addressee is expected to be able to identify the referent on the basis of the description and previous knowledge. Familiar means the addressee already has a representation of the referent. Activated means that the referent is represented in short-term memory, and finally, In-focus means that the referent is (not only activated but) also the topic of discussion.

The entailment relation which holds between the statuses provides an implicational scale for the Quantity maxim of Grice's cooperation principle [Grice, 1975]:

Maxim of Quantity

**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.

According the analysis of GHZ, when a form is used which signals a specific status, it is conversationally implicated by Q1 that the referent is not of a higher status, and by Q2 that the referent is not of a lower status. The *defeasible* nature of implicatures allows an "incorrect" form to be used, if there is a reason to cancel the implicature. In the following example, the referent "a doctor" is ACTIVATED. The implicature (that the referent's status is not higher than TYPE-IDENTIFIABLE) can be explained away by the fact that what's relevant here is "the property of being a doctor, and not the identity of this particular doctor" [Gundel et al., 1993, p. 296].

(3.19) Dr. Smith told me that exercise helps. Since I heard it from *a doctor*, I'm inclined to believe it.

In conclusion, the Givenness Hierarchy provides a standard mapping between statuses and forms, and the entailment relation between the statuses provides flexibility and choice —making that mapping defeasible.

We can now cast each status in terms of the epistemic model developed in the previous section. At this point I'm treating the status IN-FOCUS as part of ACTIVATED[4]. Moreover, where GHZ refer to the speaker's assumptions of the hearer, I refer to subjective common ground, which, as argued above, is a more accurate model of the beliefs involved referring. The Givenness hierarchy under modes of presentation is presented in figure (3.1). I use the following notation:

**Ext / Int** means that the intended MoP is extensional / intensional

**+Det / -Det** means that the intended MoP is determinate / nondeterminate

**+Local-CG / +QP-CG / −CG** means that the intended MoP is believed to be in the local common ground / the quasi-permanent common ground / not common ground (private).

The ACTIVATED status consists of referents assumed to be in short-term memory, hence it is equivalent to Prince's EVOKED status ([Prince, 1981]) and includes both textually-evoked (discourse-old) referents as well as situationally-evoked referents. The translation to epistemic features is straightforward: ACTIVATED consists of the MoP's whose IS's are Local-C.G. In this status the other

[4]The difference between IN-FOCUS and ACTIVATED has to do with saliency and, arguably, with topicality. A straightforward way to define IN-FOCUS is the status containing the most salient ACTIVATED referent(s). Topicality has to do with more than givenness status, e.g. discourse purpose, relevance etc. This work however doesn't provide a separate analysis for topics (other than their being ACTIVATED).

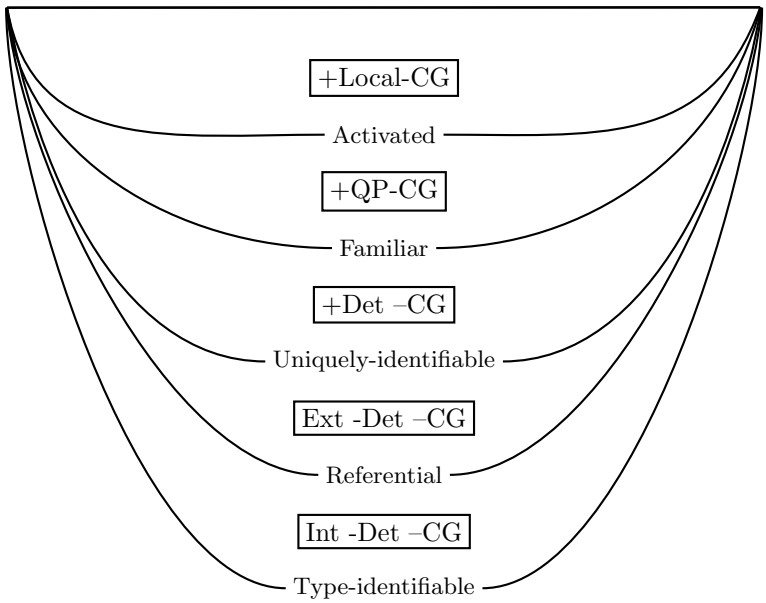


Figure 3.1: Givenness hierarchy under modes of presentation

epistemic features play less of a role. The following examples are of ACTIVATED referents with different values of these features:

(3.20) a. *Int, -Det*: If a farmer owns a donkey, *he* beats *it*.

b. *Int, +Det*: If the shortest farmer owns a donkey, *he* beats it.

c. *Ext, +Det*: If Pedro owns a donkey, *he* beats it.

Hence: ACTIVATED: Local C.G. MoP's

The FAMILIAR status includes referents which the addressee is expected to already have a representation of. Hence, it encompasses all MoP's that are C.G, whether it is the local common ground, where the MoP's are also ACTIVATED), or the quasi-permanent C.G. As I argue above, QP-C.G MoP's must be extensional. Excluding intensional MoP's makes a sharp distinction between the FAMILIAR status and the UNIQUELY-IDENTIFIABLE status.

For instance, if it is common ground that there are spies (and that no two spies have the same height), then 'the shortest spy' might be considered common ground just as 'the Queen of England' is. Defining as 'given' or 'familiar' anything whose existence can be entailed by common ground is a popular way for familiarity-based theories, (e.g. [Prince, 1981, Heim, 1982, Roberts, 2003]), to explain uniqueness properties of definite descriptions in terms of familiarity. However, giving 'the shortest spy' the same familiarity status as the current Queen of England is intuitively incorrect —jumping from plausible (inferred) existence to familiarity is unwarranted. Hence: FAMILIAR: QP C.G. extensional MoP's (and higher status MoP's)

The UNIQUELY-IDENTIFIABLE status includes all referents that the addressee can identify either based on an existing representation or on the descriptive content of the nominal. GHZ present the following example, where the descriptive content is supposedly enough for identification:

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

What sense of 'uniquely-identifiable' is fulfilled by "the dog next door" in case the hearer has no prior knowledge of that dog? Obviously, the hearer couldn't identify it at a dog show. Moreover, it is not clear whether the description uniquely denotes the referent: it could be that the speaker doesn't have or doesn't supply enough information to ensure determinacy. In such a case, the status is REFERENTIAL and not UNIQUELY-IDENTIFIABLE.

This case can be construed as another instance of non-unique and non-familiar uses of definite descriptions presented in [Birner and Ward, 1994]. It is plausible that the speaker of (3.21) only wishes to complain about feeling tired, in which case the definite article can be explained by the fact that it is irrelevant for the purposes of the exchange which particular dog kept the speaker awake.

The UNIQUELY-IDENTIFIABLE status must therefore be more precisely defined. First, this status is entailed by higher statuses, hence it includes all C.G. MoP's. Beyond familiar referents, UNIQUELY-IDENTIFIABLE contains referents which the hearer can identify based on the descriptive content of the referring expression. However, this is the point where the notion of identification and

identifiability becomes vague. What sense of identification or identifiability is fulfilled, for instance, by a Russellian unique description such as "the shortest spy"? Following Kronfeld, the issue here is not necessarily identification in an epistemic sense but (some sense of) mutual individuation.

There are two types of mutual individuation: in the case of intensional MoP's an object is mutually individuated if and only if the description uniquely denotes it —if and only if the MoP is determinate. If an intensional MoP does not determine —denote— the intended referent, then there cannot be an intended referent —the status of the MoP is TYPE-IDENTIFIABLE. In the case of extensional MoP's, the determinate feature marks the difference between UNIQUELY-IDENTIFIABLE and REFERENTIAL MoP's. That is, if a MoP is not already common ground, then the speaker does not believe the object is mutually known, which entails that she doesn't believe that the hearer believes the object is mutually known[5], so mutual individuation is achieved based on the speaker's intention to refer to a particular object. Mutual individuation is sufficient and necessary for the REFERENTIAL status. For UNIQUELY-IDENTIFIABLE, the description must also uniquely determine the referent to the hearer outside the context of its use —i.e. regardless of the speaker's intention of referring uniquely. Hence, UNIQUELY-IDENTIFIABLE: –CG/+Det MoP's (and higher status MoP's).

The status REFERENTIAL applies when the speaker intends to refer to a particular object. This, according to GHZ, is a different notion than Donnellan's referential use. For instance, this status includes all MoP's in the higher statuses, so it also includes, for instance "the shortest spy" and "you". Since the UNIQUELY-IDENTIFIABLE status already contains all determinate MoP's and all C.G. MoP's (given that it is entailed by the higher statuses), this status only adds the extensional MoP's which are not determinate, hence REFERENTIAL: –CG/Ext/-Det (and higher status MoP's).

TYPE-IDENTIFIABLE is the least restrictive status, and necessary for the appropriate use of any nominal expression. The addressee is expected to access a representation of the type of object described by the expressions. If the status of the intended object is strictly TYPE-IDENTIFIABLE then the addressee is only expected to access a representation of its type, which means that the type of object is part of the meaning of the nominal. Any MoP strictly contained in this class must be attributive (otherwise it would be REFERENTIAL), non-determinate (otherwise it would be UNIQUELY IDENTIFIABLE) and not common ground (otherwise it would be FAMILIAR). Hence, TYPE-IDENTIFIABLE: –CG/Int/-Det (and higher status MoP's).

Some conclusions can now be drawn concerning the entailment relation in the Givenness Hierarchy:

- ACTIVATED should not entail FAMILIAR. The reason is that ACTIVATED MoP's can be intensional and nondeterminate (e.g. anaphora in the don-

[5]If the speaker does not believe the object is mutually known but thinks that the hearer believes that it is, then it must be that the speaker believes the hearer is wrong, hence the intended referent is something other than what the hearer believes the description determines.

key sentences), but such MoP's are neither REFERENTIAL, nor UNIQUELY-IDENTIFIABLE, and do not belong in the quasi-permanent database.

- UNIQUELY-IDENTIFIABLE should not entail REFERENTIAL. The reason is that UNIQUELY-IDENTIFIABLE MoP's may be intensional; e.g.: "the shortest spy".

To overcome these objections, I propose a more refined entailment relation. The new hierarchy is presented in figure 3.2. Entailment is explained by the following four principles:

**Specificity** This is the requirement that the speaker intends to refer to a particular object, and not just anything that fits the description. This is a condition which extensional MoP's satisfy and intensional MoP's do not. Hence, a status that contains intensional MoP's also contains extensional MoP's which fulfil the requirements.

**Determinacy** This is the requirement that enough content is encoded in the MoP such that it determines the referent. This means that a status that contains nondeterminate MoP's also contains the determinate MoP's which fulfil the same requirements.

**Common Ground** This is the requirement that the MoP is common ground, which means that it is also extensional. A status which contains non common ground extensional MoP's also contains common ground extensional MoP's.

**Saliency** Saliency is more an issue of processing. The more salient a discourse referent is the less descriptive content is required to refer to it. This principle means that the conditions required to express a discourse-new referent are met by discourse-old referents.

### 3.2.2 The Givenness Hierarchy and NLG

The ability to generate descriptions for Hearer-New referents depends heavily on the modal aspect; that is, on an implementation of a sentence planner that can realize referential goals other than identification. The Givenness Hierarchy represents the range of referential intentions a speaker can realize when producing a linguistic expression of a mode of presentation. It provides a basic classification of discourse purposes, which corresponds directly to identification constraints [Kronfeld, 1990, p.80].

**Type-identifiable** the speaker has no *referring* intention; the discourse purpose here is describing-as, attributing, or generalizing. Such a discourse referent is a variable or a predicate, rather than a referring expression. Intention: attributing

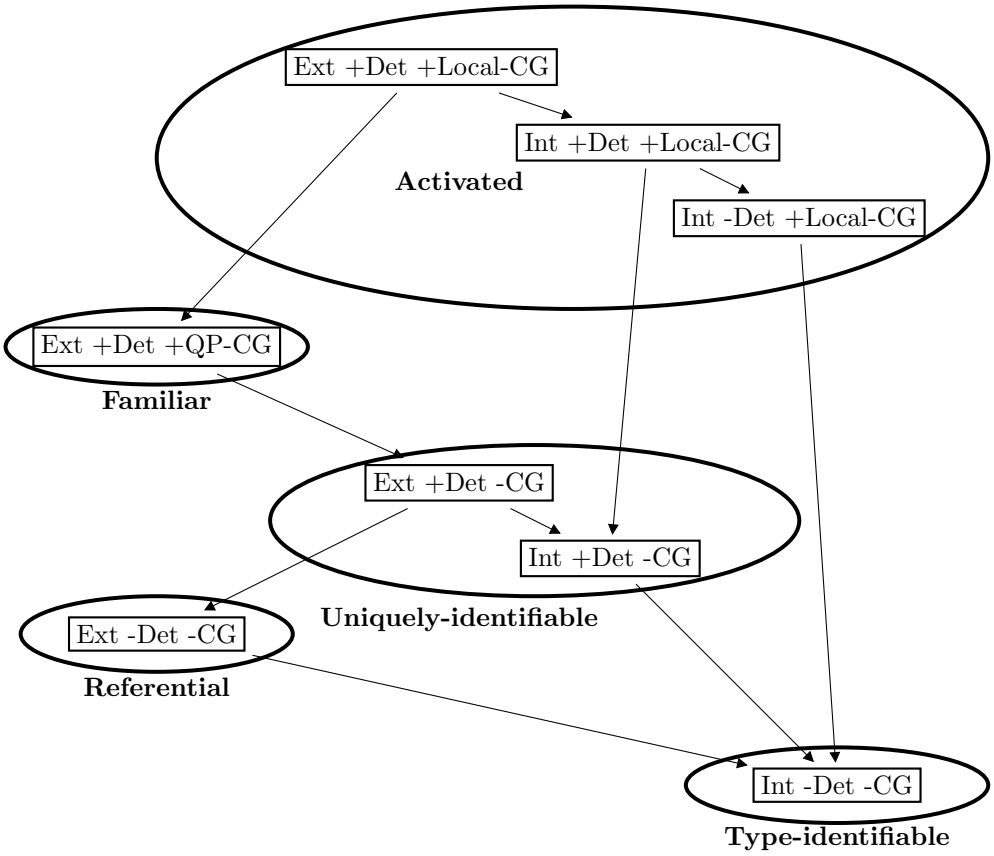


Figure 3.2: Fully-specified Givenness Hierarchy under modes of presentation, with preference relations

**Referential** this class corresponds to the null identification constraint. The hearer's recognition of the literal goal also satisfies the discourse purpose (of referring), which is to let the hearer generate a local IS that determines the object as the one the speaker intends to refer to. Intention: referring

**Uniquely-identifiable** in this case there is a MoP which the speaker intends to be determinate, and mutual individuation is on basis of the content that MoP[6]. This corresponds to the identification constraint requiring that the IS contain a privileged MoP (the basis of mutual individuation). Intention: uniquely referring

**Familiar** the speaker intends to refer to a common ground MoP, corresponding to the constraint requiring that the new IS be merged with a previous quasi-permanent IS. Intention: Hearer-Old referring, or more precisely: referring to Unused referents ([Prince, 1981])

**Activated** the speaker intends to refer to a discourse MoP, corresponding to the constraint requiring that the new IS be merged with a previous local IS. Intention: Discourse-Old referring, or more precisely: referring to Evoked referents ([Prince, 1981])

Another intention which isn't explicit in the hierarchy but is very prominent in NLG, is the intention to refer to an object in the visual field of the hearer, corresponding to the perceptual identification constraint. This intention is part of the intention to uniquely refer (the UNIQUELY-IDENTIFIABLE class) —the privileged MoP in this case is the perceptual MoP which the speaker intends as context. Note that, this context doesn't have to be familiar, as it can be in the future (based on the speaker's private knowledge), as in the following example from [Birner and Ward, 1994]:

(3.22) Can you go get *the bag of potato chips* on the bed in my room?

The intention associated with TYPE-ATTRIBUTIVE is to introduce an arbitrary *variable* into the discourse history, or to describe-as —attributing a property to a discourse referent. That is, the type of MoP which the referring act[7] is intended to produce for the hearer is intensional and nondeterminate. As Kronfeld observes, a speaker can intend to refer attributively despite having knowledge of the referent; but more than that, a speaker can intend to *attribute* something to a referent, by referring attributively without uniquely-identifying, thereby referring to a property of the referent. By this, a speaker indirectly refers to some referent, as the following example shows:

(3.23) Dr. Smith told me that exercise helps. Since I heard it from *an idiot*, I just ignored it.

[6] If the MoP is extensional the speaker is referring to a specific object which is not necessarily the object which satisfies the content (the speaker might be wrong). Still, the content of the MoP is a basis of mutual individuation.

[7] This intention is not exactly an intention to refer, as there's normally no intended referent. For lack of a better word, I stick with the word referring: referring acts, generating referring expression, referring intention etc.

This example, as well as the original one, (3.19), seem to suggest that the intention to attribute is not as simple as just introducing a nondeterminate intensional MoP, as there is something which connects such a non-referring MoP with an actual referent. However, I don't have anything to say about how this indirect referring happens.

As for the more standard case in this class —the intention to introduce an arbitrary variable—, this intention is applicable in negated contexts, quantified contexts, de-dicto belief attributions and other opaque contexts, as the following variations of donkey sentences show:

(3.24) a. The farmer down the road doesn't have *a donkey*.

b. If *a farmer* owns *a donkey*, he beats it.

c. Adam had a bad dream and believes there's *a donkey* under his bed.

d. Rufus is looking for *a donkey* to rent.

e. Will you buy me *a donkey* for my birthday?

In realizing this intention, the role of the generator is minimal, as no distinguishing has to take place. The property which has to be realized is determined by the planner: it is part of what to say, not of how to say it.

The intention associated with REFERENTIAL, is a proper referring intention —the speaker intends to refer to a particular object, and not just anything that fits the description. Mutual individuation is based on the hearer's recognition that the speaker is indeed referring to a particular object. As a result of this recognition, the literal goal of referring is satisfied: the hearer generates a nondeterminate extensional MoP. Subsequently, the hearer's new local MoP is causally linked to the speaker's new local MoP, and these two MoP's co-refer. In this case as well no distinguishing has to take place: The planner needs to determine which MoP (or MoP's) is relevant in introducing this referent. The role of the generator is, again, minimal.

The intention associated with UNIQUELY-IDENTIFIABLE is to give the hearer the means to identify the referent, and thereby generate a determinate MoP. However, in contrast with the standard name assumption, the issue here is not to distinguish the target referent from others. Consider two simple example of bridging:

(3.25) I like that restaurant. *The cook* really knows what he's doing

(3.26) When I go to a new restaurant I like to meet *the cook*.

As regards the modal aspect (the aspect of planning), in (3.25) the use of "the cook" might be referential, but it could also be attributive; in (3.26) it is undoubtedly attributive. However, in neither case is there an issue of distinguishing the cook from other objects[8]. The new MoP is intended to be determinate, and the basis of mutual individuation is the MoP which is a functional concept.

[8]There is in general an issue of distinguishing the anchor, but that's another point.

The difference between an implicit bridge like "the cook" in (3.25) and an explicit bridge, like "the cook of that restaurant" has to do with uniqueness and saliency of the anchor, a discourse referent —this is discussed below (discourse referents have a higher status than UNIQUELY-IDENTIFIABLE). The difference between a bridging description and a unique description such as "the shortest spy", is that bridging descriptions express MoP's that are functional concepts.

MoP's that are UNIQUELY-IDENTIFIABLE are functional: they denote a unique object, but, as we have seen in (3.26), this can be relative to some other object (Cf. [Kadmon, 1990]). However, there are two ways a MoP can be functional: by extension and by intension. Uniqueness by extension is determined by referential knowledge: In the case of (3.25), the speaker can intend to refer uniquely to the cook if she has an extensional MoP which is determinate —that is, she believes of someone (the cook she knows of) that he is the sole cook of that restaurant —this belief is not constructed for this referring act but part of her referential knowledge (expressed by the determinate feature) . The exception to this is MoP's that refer to sense data, which require a translation from a mental image to language (it seems implausible that humans automatically generate unique descriptions for anything they perceive). Perceptual MoP's are typical examples for GRE. On the other end are determinate intensional MoP's. These kinds of MoP's are not (necessarily) directly represented in referential knowledge, as they can be inferred —they are entailed by propositional knowledge. With these kinds of MoP's uniqueness must be derived from general knowledge (as [Striegnitz, 2003] does, at least in part), but there is no matter of distinguishing the target referent. Hence, with the exception of perceptual MoP's, generation is determined by the planner.

With FAMILIAR IS's the situation changes, as now there is a context on the basis of which a referent can be distinguished. With the exception of perceptual MoP's, only from this status onward is there any sense in trying to *distinguish* the target referent from so-called distractors, as distinguishing is based on properties which the hearer is assumed to know that the referent has, hence such means of identification presupposes a target referent and distractors which are all common ground. Hence, a Hearer-Old IS can be referred to by distinguishing it from the other common ground IS's in the context.

Referring to ACTIVATED IS's is similar to referring to FAMILIAR w.r.t. distinguishing the target referent, but the context set here is restricted to Discourse-Old referents. Another difference is the issue of referring to discourse referents which are nondeterminate intensional MoP's, that is, referents that do not refer or denote, but serve as variables, which can only be referred to in a restricted context.

# Chapter 4

# Conclusions

## 4.1 Summary

Noun phrases in natural discourse fulfil different types of discourse purposes: from serving as (non-referring) variables to uniquely identifying an actual object, from introducing a Hearer-New referent to referring to the topic of conversation. As a result, the representation of objects in our minds must be very heterogenous: we know ourselves by acquaintance, recognize objects in our environment by our visual memory; of some objects we have uniquely denoting descriptions and proper names, while of others we might have very little information. On top of this, drawing on world knowledge, a virtually infinite number of objects can be conceived of based on the belief that they exist: if we know of a house, we can assume that it has *a door*, that the door is made of some *material*, and so on; if we know there are spies, we may assume there is such a person who is *the shortest spy*.

This paper originally began as an attempt to generate suitable descriptions for Hearer-Old (unfamiliar) referents in the context of Natural Language Generation (NLG). However, it turns out that the Generation of Referring Expressions (GRE) task within NLG makes several tacit assumptions, the most pervasive of which is a problem known as *the standard name assumption*. Under this assumption, any object that can be talked about is conveniently represented as some sort of label which is shared with the user. Connection with the actual object can be "safely" ignored as the referring algorithm carefully seeks a suitable description for the internal label —such a description, if found, is automatically guaranteed to be suitable for the user too. This model of referring Hearer-New referring is hopelessly flawed.

Perhaps the main contribution of Kronfeld's theory is his 3-tiered model of Donnellan's attributive/referential distinction. Considering this distinction from both the internal perspective as well as the external perspective, Kronfeld identifies three different aspects of it: the epistemic aspect, the modal aspect, and the denotation aspect. This theoretical model has an immediate correlation

to computational systems: the epistemic aspect concerns representing domain entities, a role fulfilled by the database; the modal aspect is roughly about planning, and is fulfilled by the sentence planner; the denotation aspect concerns use of referring expression, and is fulfilled by the utterance generator —By GRE, in our case.

In this paper I set out to provide a realistic representation of objects —a computational model of Kronfeld's epistemic aspect. I propose an epistemic distinction between two types of modes of presentation (MoP's): intensional MoP's and extensional MoP's. Informally, this distinction can be expressed by the question "where does this presentation mode come from?", and the informal answer is: MoP's can come from two sources: from the external world or from the internal world (e.g. via inference). Of course, both types of MoP's are in our minds, both might determine an object in the actual world and they might not; however, the beliefs involved in these two types of MoP's are different: intensional MoP's represent objects by virtue of a belief about the intension of the MoP —the belief that the property expressed by it uniquely denotes; extensional MoP's represent objects which become known to the agent by experience or through discourse, and the belief involved is that an object has the property expressed by the MoP. By doing this, my proposal becomes a blend of the causal theory of reference and the descriptive theory. It is my contention that object representations must be heterogenous in such a way.

With the distinction between intensional MoP's and extensional MoP's I examine reference to Hearer-New objects and present a further distinction between determinate MoP's and nondeterminate MoP's. In the simpler cases, this distinction corresponds to the question "is the description believed to uniquely determine the object?". By this I abuse the notion of MoP as defined by Kronfeld and others, by allowing it to contain descriptions that are not necessarily uniquely identifying. By accepting that the causal theory is (also) right, possession of a unique description is no longer a requirement for referring to a particular object: an agent can refer to an object to which he has a causal chain but not enough descriptive content to identify. By allowing intensional MoP's to be nondeterminate too, we can account for the use of a noun phrase as a non-referring expression —as a variable either under quantification, negation, or in an opaque context.

By representing the system's subjective (assumed) common ground with the user, I show how Appels and Kronfeld's notion of mutual individuation is realized for the two types of MoP's. Mutual individuation, a necessary condition for the goal of referring, is the mutual belief between speaker and hearer that both of them are focusing on the same object. With intensional MoP's, mutual individuation is explained by the descriptive theory: what guarantees mutual individuation is the fact that the descriptive content of the MoP actually *denotes* the object. Mutual individuation in case of extensional MoP's is explained by the causal theory. In the case of Hearer-New extensional MoP's, mutual individuation is based solely on the hearer's assumption that the speaker refers to an actual object. In a computer implementation, a successful reference to a Hearer-New extensional MoP would somehow mark that MoP as common

ground —allowing future references to succeed on the basis of that.

As an application of the new model of the epistemic aspect of referring, the Givenness Hierarchy of Gundel, Hedberg and Zacharski, was recast in terms of the new epistemic variables: intensional vs. extensional, determinacy, common ground, and discourse model vs. referential knowledge. In the process the information statuses in the hierarchy get sharpened. For instance, the status FAMILIAR is defined to exclusively contain extensional MoP's that are assumed to be common ground, excluding MoP's that are inferrable. This, however crude as it may turn out to be, provides a theory that puts uniqueness and familiarity each in its own place.

Finally, the statuses in the givenness hierarchy were formulated into a preliminary model of referring. Since GRE is based on finding a distinguishing description, its role in generating Hearer-New referents —MoP's below the class FAMILIAR— is minimal, as the introduction of new discourse referents is dependent on communicative intentions and conversational relevance. The epistemic foundations laid out in this work for the internal representation of objects would hopefully be incorporated into a more realistic model of referring in the future.

## 4.2 Open issues and further research

There are several points which lack some depth and require further explanation, as is often the case with interdisciplinary studies.

In this work, I have been focused mainly on MoP's and haven't talked so much about IS's. Contrary to Kronfeld I am not sure whether there is a way to refer to an entire IS as a whole. In case of the status FAMILIAR, discourse participants may share a big part of the IS of a given object, and might be said to be referring to the IS as a whole. But referring to an object which is common ground is different than referring to a Hearer-new object. I believe that in introducing a new referent a speaker cannot refer to the IS as a whole. Since I accept the causal theory, I can still keep the original spirit of Kronfeld's modal aspect: an extensional MoP can be referred to as a rigid designator — determined by a causal link.

### Determinacy and bridging

There's some issue which I don't address in my analysis of determinacy. There are, I think, different possible conditions for determinacy: it can be entailed by the agent's private knowledge, and it can be entailed by common ground. Moreover, in the case of intensional MoP's, this entailment can be due to default knowledge.

For instance, as part of the interpretation of (4.1), a hearer should accept that the (extensional) MoP which expresses "owner of El Azteco" is determinate —there is one and only one owner.

(4.1) I met *the owner of El Azteco*.

On the other hand, in the interpretation of (4.2), "the owner" can be successfully inferred as "the owner of the new restaurant around the corner", given the fact that restaurants, and businesses in general, normally have owners. This doesn't necessarily imply that there is one and only one owner.

(4.2) The new restaurant around the corner became popular fast. *The owner* must be very happy.

Further research can discover whether the distinction between intensional and extensional MoP's provides some insights about the interpretation of bridging descriptions.

## Status entailment and surface form

I haven't really explained the implications of the new entailment relation I present for the Givenness Hierarchy (figure 3.2, p. 36). One reason is I am not sure that the conventional mapping proposed by GHZ is correct. For instance, it seems unnatural that demonstratives are part of conventional use of a status, and while GHZ's proposal is not wrong, common sense suggests that conventional use should be the most frequent one. The following is a different, tentative, mapping from statuses to surface forms, in which demonstratives are syntactically marked versions of the statuses, and not part of their conventional meaning:

| activated | familiar | uniquely identifiable | referential | type identifiable |
|---|---|---|---|---|
| **Conventional use:** | | | | |
| *it,* reduced *the N* | *the N* | *the N* | *a N* | *a N* |
| **Marked use:** | | | | |
| *that,* *this,* *this N* | *that N* | *the N* | indef. *this N* | |

This mapping should not come as a surprise. First, the conventional use is straightforward: according to my analysis familiarity (C.G. extensional MoP) entails uniqueness (determinate); these two factors are normally considered as conditions of definiteness (Cf. [Abbott, 2005]). Moreover, the conventional use (in English) of a given status is, as common sense predicts, its most frequently used form (See [Gundel et al., 1993, p. 291]).

The marked versions can be used to explicitly signal the actual status of a MoP when it is not inferrable from the context. Moreover, they can be used for raising the status of a MoP. That is, in some cases a speaker can explicitly signal a higher status than a MoP actually has, by referring to that MoP using the marked version of the higher status. Note that demonstratives can be used to signal various kinds of effects, such as focus shift, contrast, and emotional

effects (Cf. [Gundel et al., 1988]). The entailment relation restricts the possible status raising moves —the intended status of the MoP must entail the epistemic status it actually has[1]. Moreover, status raising is not allowed without warrant. Informally, it depends on the assumptions the speaker can make about the expectations of the hearer.

- A MoP which is nondeterminate —TYPE-ATTRIBUTIVE or REFERENTIAL— can be marked as determinate and raised to UNIQUELY-IDENTIFIABLE if the hearer can be expected to interpret it as the most relevant or most salient object fitting the description in the context, *without knowing which one it is*:

  (4.3) I reached the house and knocked on *the door*.

  (4.4) Yesterday I went to the new Italian restaurant around the corner. *The waitress* was nice.

  In a blank context, (4.3) could be interpreted as the front door of the house, but this interpretation is just a default salient interpretation, and it may not be true. I propose that in such a case, given that epistemic identification (by the hearer) is not relevant, the speaker can assume a salient door (e.g. the one which was nearest to the speaker depending on the direction from which she approached the house), but I think attaching a description to it to make it unique cannot always work. Similarly, in (4.4), a relevant nonidentifiable waitress can be presupposed. This analysis should be taken with a grain of salt, and even so, it doesn't account for some of the difficult examples presented in [Birner and Ward, 1994], especially the ones involving locations.

- A MoP which is not C.G., but which corefers with a C.G. MoP (hence, in the same IS) can be raised to FAMILIAR, signalling to the hearer that the referent is —should be— mutually known:

  (4.5) *That bastard friend of yours* is back in town.

  (4.6) K: ...he wears... *those kind of tennis shoes that are expensive.*
  N: Adidas. [Gundel et al., 1988, p. 219]

Consider the following cases in which forms are not allowed by the new entailment relation:

- A determinate intensional MoP cannot be signalled to have the status FAMILIAR, which is reserved for extensional MoP's. This correctly predicts the infelicity of the following cases:

[1] Note that I'm not dealing here with cases such as (3.19), which are, questionably, some kind of status lowering:

(3.19) Dr. Smith told me that exercise helps. Since I heard it from *a doctor*, I'm inclined to believe it.

(4.7) ?? *That winner in the race* will get 1000$.

(4.8) ?? Whenever I travel to a new country, I try to learn *that spoken language*.

- Similarly, only (nondeterminate) extensional MoP's, and not intensional MoP's, can have the status REFERENTIAL, as explicitly marked by the use of indefinite this.

A more rigorous analysis would need to take into account the other uses of demonstratives.

# Bibliography


[Abbott, 2005] Abbott, B. (2005). DEFINITE AND INDEFINITE. *Encyclopedia of language and linguistics, 2nd ed. New York: Elsevier*.

[Aloni, 2001] Aloni, M. (2001). *Quantification Under Conceptual Covers*. Universiteit van Amsterdam.

[Appelt and Kronfeld, 1987] Appelt, D. and Kronfeld, A. (1987). A computational model of referring. *In Morgen Kaufman, editor, Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, pages 640–647.

[Ariel, 1990] Ariel, M. (1990). Accessing noun-phrase antecedents.

[Birner and Ward, 1994] Birner, B. and Ward, G. (1994). Uniqueness, familiarity, and the definite article in english. In *Proceedings of the Annual Meeting of the Berkeley Linguistic Society*, pages 93–102.

[Clark, 1977] Clark, H. (1977). Bridging. *Thinking: Readings in Cognitive Science*, pages 411–420.

[Clark, 1996] Clark, H. (1996). *Using Language*. Cambridge University Press.

[Clark and Marshall, 1978] Clark, H. and Marshall, C. (1978). Reference diaries. *Proceedings of the theoretical issues in natural language processing*, pages 57–63.

[Clark and Schaefer, 1989] Clark, H. and Schaefer, E. (1989). Contributing to discourse. *Cognitive science*, 13(2):259–294.

[Dale and Reiter, 1995] Dale, R. and Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

[Donnellan, 1966] Donnellan, K. (1966). Reference and Definite Descriptions. *The Philosophical Review*, 75(3):281–304.

[Grice, 1975] Grice, H. P. (1975). Logic and conversation. *P. Cole and J. Morgan, editors, Syntax and Semantics*, 2/3(3, Speech Acts):43–58.

[Grosz and Sidner, 1986] Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

[Gundel et al., 1988] Gundel, J., Hedberg, N., and Zacharski, R. (1988). On the generation and interpretation of demonstrative expressions. *Proceedings of the 12th conference on Computational linguistics-Volume 1*, pages 216–221.

[Gundel et al., 1993] Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

[Heim, 1982] Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts.

[Horacek, 1997] Horacek, H. (1997). An algorithm for generating referential descriptions with flexible interfaces. *Proceedings of the 35th conference on Association for Computational Linguistics*, pages 206–213.

[Jørgensen, 2000] Jørgensen, S. (2000). *Computational Reference*. PhD thesis, PhD thesis, Copenhagen Business School.

[Kadmon, 1990] Kadmon, N. (1990). Uniqueness. *Linguistics and Philosophy*, 13(3):273–324.

[Krahmer and Theune, 2002] Krahmer, E. and Theune, M. (2002). Efficient context-sensitive generation of referring expressions. *Information Sharing: Givenness and Newness in Language Processing*, pages 223–264.

[Kripke, 1980] Kripke, S. (1980). *Naming and Necessity*. Blackwell Publishing.

[Kronfeld, 1990] Kronfeld, A. (1990). *Reference and Computation*. Cambridge: Cambridge University Press.

[Partee, 1987] Partee, B. (1987). Noun phrase interpretation and type-shifting principles. *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, 8:115–143.

[Perry, 1979] Perry, J. (1979). The Problem of the Essential Indexical. *Noûs*, 13(1):3–21.

[Prince, 1981] Prince, E. F. (1981). Toward a taxonomy of given-new information. In *In P. Cole (Ed.), Radical Pragmatics*, pages 223–255. N.Y.:Academic Press, New York.

[Reiter and Dale, 2000] Reiter and Dale (2000). *Building Natural Language Generation Systems*. Cambridge University Press.

[Reiter and Dale, 1992] Reiter, E. and Dale, R. (1992). A fast algorithm for the generation of referring expressions. *Proceedings of the 14th conference on Computational linguistics-Volume 1*, pages 232–238.

[Roberts, 2003] Roberts, C. (2003). Uniqueness in Definite Noun Phrases. *Linguistics and Philosophy*, 26(3):287–350.

[Schiffer, 1978] Schiffer, S. (1978). The basis of reference. *Erkenntnis*, 13(1):171–206.

[Striegnitz, 2003] Striegnitz, K. (2003). *Generating Anaphoric Expressions; Contextual Reasoning in Sentence Planning*. PhD thesis, Universität des Saarlandes.