Integrating Representation Systems : Language Usage in Graphics Communication

Umata, Ichiro

Integrating Representation systems
Language Usage in Graphics Communication

Ichiro Umata

March 27, 2001

Contents

8 Conclusion **90**

Abstract

Graphical representations like maps and diagrams play an important role in everyday communication settings. In such communication, people effectively integrate two independent systems of representation: spoken language and graphics. This thesis presents an empirical investigation of language usage in graphical communication. Drawing on our dialogue data, we will show that the presence of a graphical representation significantly changes the way the spoken language is used, extending its expressive capacity or affecting the perspectives on events. We will first report the phenomena of *mediated reference* and *dual description* as two remarkable uses of language affected by graphics, illustrating them with actual examples from our data . A quantitative analysis of our data will show that these special uses of language are indeed as common as conventional uses of language in the presence of graphical representations. We will also propose that the mediated use of language contributes to efficient communication through integrating sequentially information of a representation and information conveyed by a linguistic expression. Finally, we show that the configuration in graphics affects linguistic expressions of motion, and this fact demonstrates that the configuration in graphics has an influence on the perspectives of event conceptualizations.

Chapter 1

Introduction

Daily communication is by its nature multi modal. Conversational exchanges often involves not only information from language but also information from other representation systems like maps, diagrams and gestures. Most linguistic studies, however, have not focused on the facet of language as a communicative tool interacting with other representation systems. The purpose of this study is to show the interactive nature of language as a communicative tool, not just as a self-contained symbol system, through empirical investigations of conversation involving graphical representations like map and diagrams.

Conversational exchanges accompanied by external graphical representations are fairly common in our daily lives. People give and ask directions by referring to maps, and they draw diagrams and pictures in discussing where to place living room furniture. As a representation system which interacts with other systems, it is natural that language usage is affected by information from graphical representations. Actual dialogue data shows that language often describes the target situations mediatedly via some other representation, and language usage is influenced significantly by the presence and the configuration of graphical representations.

1

Language system and other representation system like graphics are sequentially integrated in such cases, with language system dominating other systems. We will examine the impact of graphics on the use of language when language system and a graphical representation are sequentially integrated.

First, the presence of a graphical representation significantly changes the way linguistic expressions describe their targets. Linguistic expressions denote objects and relations in the world. This denotation relation is governed by language conventions. An utterance of a linguistic expression carries information about the world through these conventions. This is what the standard view of the semantics of language tells us. However, when we look at speech in conversational exchanges involving external representations such as maps, diagrams or pictures, regardless of the language used, we will immediately notice utterances that do not conform to this standard picture. We will focus on two remarkable uses of language, called "mediated reference" and "dual description," that we found through an examination of actual two-party dialogue data. Both phenomena are clearly specific to dialogues involving some graphical representation, or at least, some external representation other than speech (see Umata, Shimojima & Katagiri (2000a)).

Briefly, *mediated reference* is a case where a linguistic expression reaches its "final" referent due to the fact that its "immediate" referent has a referential connection to this final one in the system of graphics. For example, our subjects often use the indexical "kore" (this) to refer to a building or some other landmark, although its immediate referent is clearly an icon on the map; the icon refers to the landmark in the system of map, and this fact somehow enables the indexical

expression to do so too. We will discuss more examples of mediated reference later, and introduce three more varieties of the phenomenon.

Dual description is a case where a declarative sentence is used to describe a fact that holds in the graphic as well as the corresponding fact in the situation represented by the graphic. Suppose, when asked about the number of stations between two particular stations, one counts the number of icons on a railroad map and says, "There are three stations in-between them." Is this report concerned with the map itself, or with the mapped railroad? Is it reporting that the railroad map has three station icons between two particular station icons, or that the railroad system has three stations between the two stations? Whichever the answer may be, it seems clear that the speaker has managed to describe both facts with this sentence. Note that, on the semantics associated with the railroad map, the first fact means the second fact, and this semantic relation somehow underwrites the duplicative use of the sentence.

Mediated reference and dual description not just extend the descriptive capacity of language, but also provides an effective way of information flow. *Information integration* is a phenomenon often found in those uses of language. When people refer to some real-world building by words like "kore" (this) via a map icon, the information given to the listener is not just the reference relation. The information presented on the map (i.e. the spatial relations between the icon directly referred to and other icons, and so on) is integrated with the linguistic information and provides rich information such as spatial relations. The mediated use of language induces integration of the linguistic and graphical information and provides amplified information to the listeners (Umata, Shimojima & Kata-

3

giri (2000b), (2000c)).

Both uses of language are so natural and common in a dialogue involving a graphical representation that people may not even be aware of the phenomena. In fact, their frequent occurrence in such settings suggests that they are not a deviant but rather a perfectly legitimate use of language. Several researchers in the field of linguistics have been analyzed such mediated uses of language. Jackendoff (1975) noted a similarity in the characteristics and behaviors between picture sentences and belief sentences, and he proposed an analysis of picture sentences based on the close connection between the two. Lakoff & Johnson (1980) and Lakoff & Turner (1989) gave a fairly detailed analysis of the use of metaphors and metonymies in literary works as well as in everyday language use. They also developed a cognitive model based on mapping between conceptual entities, either across or within conceptual domains. Another scholar, Faucounnier (1985), proposed an account of various types of indirect descriptions in terms of his theory of Mental Spaces. Most studies, however, have not focused on the systematic nature of information integration when the mediated uses of language occur. There are also several empirical studies in integration of linguistic and graphical representations which focus on the issue of how speech is used to disambiguate a graphic (Neilson and Lee 1994) or how a graphic is used to disambiguate speech (Lee and Zeevat 1990). The linguistic-graphic integration has been also studied from a logical point of view, but the focus has been on how a graphic expresses what cannot be easily expressed by a linguistic representation (Barwise and Etchemendy 1996, Shimojima 1999). For both views, the fundamental form of linguistic-graphic integration is a *parallel* one, where each mode of representation expresses infor-

4

mation in its own way, but since one mode of representation expresses what the other form does not, they may work complementarily to each other.

This study attempts to explain the mechanism of mediated uses of language based on *sequential* integration of two independent representation systems, namely language and a graphical representation. This results in both reducing and extending the power of linguistic theories. The descriptive power of an utterance is brought not solely from language system but also from the interaction of the language system and a graphical representation. However, language system can dominate other representation systems like graphical representations and convey much richer information when it is sequentially integrated with other systems.

When language system and a graphical representation are sequentialy integrated, people describe and capture the target world events or situations through that graphical representation. The sequential integration provides us a convenient and confortable way of information exchange because of the handiness of graphics: people are seeing and talking through graphics at hand. This provides us yet another kind of perspectives on events or situations of target world; perspectives via graphics. The existence of graphics may affect the way of conceptualizing those events and situations because of the availability of these perspectives, and this may results in influencing human activities such as inference, problemsolving and understanding (see Umata, Katagiri & Shimojima (2001)).

It has been proposed that the existence of graphics not only serves as a memory aide, but also affects strategies of problem-solving and understanding. For example, Schwartz (1995) observed the effect that diagrammatic fidelity has on inference. Suwa and Tversky (1997) examined focus shifts and successful explo-

5

ration of related thought by conducting protocol analyses of designers' reflection on their own sketching behavior. However, little is known about the effect graphical representations have on language usage when people are collaboratively working on a task.

The latter part of this thesis shows how the availability and the configuration of graphics affects language usage in communication and problem-solving. We will focus on the influence of graphical representations on the perspectives from which people conceptualize motion events. Suppose that John and Mary are at Goodge Street tube station, discussing where to have dinner together. Mary might suggest a place by saying (1) below, but she would not do it by saying (2):

(1) Let's go down to Waterloo Station via Northern Line, and visit Crescent.

(2) Let's come down to Waterloo Station via Northern Line, and visit Crescent.

The current position where the two people are located becomes the reference point of the movement in this case, and the movement can only be conceptualized as a movement away from the reference point, and hence the use of "go." Suppose, on the other hand, that John and Mary are discussing their night plan over a map of London Underground shown in Figure 1.1. Mary could use, in this case, either (1) or (2). The availability of the map and the configuration of icons on the map affect the conceptualization of the movement here: the nearness of the Waterloo Station icon from them makes it possible for her to conceptualize the movement, in addition to the previous distal movement conceptualization, as a movement *in the map-world* toward the reference point, their current position. Graphical representation can have influence on language usage.
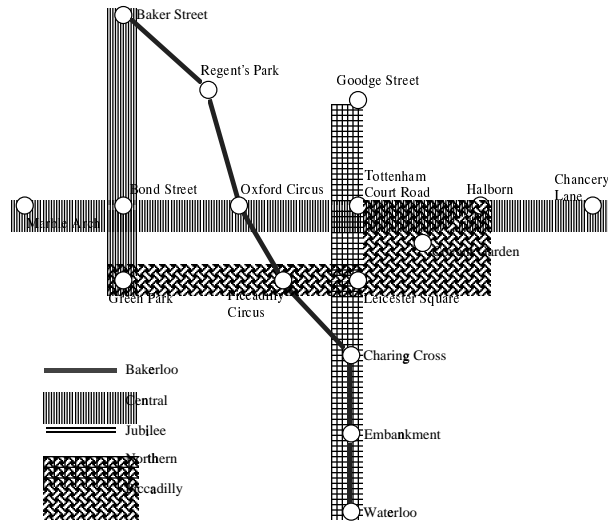
Figure 1.1: Route Map of London's Underground System

The use of "come" in (2) is possible because the map and the graphical objects contained in it are readily available to the speaker as a resource to formulate messages to be communicated and problems to be reasoned about. The locations and arrangements of objects can be expressed in terms of the relationships between graphical objects and the speaker, as well as those between objects themselves and the speaker. This availability, or the ease of accessibility, of graphical representations should work to amplify our communicative and reasoning capabilities by providing us with a novel set of possibilities to construct perspectival event conceptualizations. We will examine the effect of graphical representations on perspectival event conceptualizations through the empirical analysis of the use of motion verbs in actual two-party task-oriented dialogues that involve maps and diagrams.

The next chapter is a qualitative description of the phenomenon of mediated

references, where we illustrate it with examples of language use drawn from our dialogue data. The phenomenon of mediated reference is classified into four different types. The following chapter shows possible types of disambiguation strategies for deciding the final referent of a mediated reference. Four types of such strategies are presented with examples from our data. The phenomenon of dual description is illustrated in chapter four. We will analyze the mechanism underlying dual description with utterances from actual graphical communication dialogues. Chapter five is a quantitative description of the two phenomena, where we use "content phrasal unit" to quantify the frequency of mediated reference and dual description in the dialogue process. As it turns out, these graphic-oriented uses of language occur as frequently as standard uses in our dialogue data, indicating that the speakers are quite ready to exploit the graphical representations at hand to extend the expressive capacity of their language. Chapter six illustrates the phenomenon of information amplification induced by integrated use of language and graphics, and then inquires the mechanism of the flow of information. Chapter seven is devoted to an investigation of the effect of graphical representations on perspectival event conceptualizations. We first examine motion verbs in the Edinburgh HCRC Map Task Corpus, and then look into the Japanese dialogue data from our corpus involving a "Missionaries and Cannibals" type puzzle. The effect of graphics on language usage demonstrates that the configuration of graphics has an influence on the perspectives of event conceptualizations.

Chapter 2

Mediated References

Previous linguistic studies on metaphor and metonymy have not focused on the systematic nature of mediated references involving the world and representation systems. Studies on the use of multimodal information in reasoning and communication have mostly focused on the complementary or parallel form of integration and have not closely examined the sequential form of integration of multiple modalities. Particularly, little is known about the directionality of such integration. In this chapter, we will inquire into the phenomena of mediated references assuming that language, graphical representations and the real world are working as an integrated systems, and that we can describe situations of each local system using other ones.

In our data, two reference directions were observed: mediated references via a representation system (a sketch map, a diagram, etc.) to the world and mediated references via the world to a representation system. We will call the former a *forward* mediated reference, and the latter a *backward* mediated reference. Further, mediated references can be observed between not only individuals but also between relations. We will look into a simple example first, and then examine

the phenomena of mediated references observed in our graphical communication experiments.

2.1 The Phenomenon

Suppose two men, *A* and *B*, are talking about how to get around in London, with a small route map of London's underground system such as the one in Figure 1.

Pointing to an icon of Baker station, speaker *A* utters the following sentence:

(3) The train leaves every 10 minutes from <u>here</u>.

In example (3), the word "here" literally denotes the Baker station icon on the map. However, the train leaves from real Baker station in the real world, not from the icon of it on the map. Therefore, the utterance is describing the real world situation as in (4).

(4) In the real world, the train leaves every 10 minutes from the station which corresponds to the icon on the map that the word "here" literally denotes.

Although we have no means to know what "here" in (3) really refers to without the map, the word is actually referring not to the icon on the map but to the Baker station in the real world. The map serves to referring to the real world object and people are seeing the real world situation through the map, contingent on the systematic correspondency between the map and the world. Here, the word "here" *reffers mediatedly* to an object in the world through an object in the map. In the following, we will first look through the methods through which we collected our data, and then examine various kinds of mediated references observed from our corpus.

2.2   Method

The conversational data analyzed in this paper were gathered from a series of graphical communication experiments, which investigated the interactions between cognitive/communicative factors and graphical representations. In those experiments, each pair of subjects was asked to work on a task cooperatively with the partner. The subjects were video-taped during the task, and the conversational data recorded were partially transcribed.

The tasks were to draw or to read some graphical representation. We had three different kinds of tasks: a map sketching task[1], a Graduate Record Exam (GRE) task[2] and a graph task. In the first task, the goal for the subjects was just to draw a map. In the second and third tasks, the task had two phases; the goal was not merely to draw a graphical representation, but also to solve questions using the representation the subjects drew by themselves. Thus, there were many examples of reasoning with the diagram found in the second and third tasks.

A graphical representation expresses certain real world relations by corresponding spatial relations on the representation. Each kind of graphical representations used here had different nature in what relation in the real world were described. A map expressed the spatial relations among places in the real world by the spatial relations on the map. A GRE diagram expressed the topological relations among the cities in some fictive world by the spatial relation on the diagram. A graph expressed some abstract relations among several countries in the

---

[1]These experiments were designed by Patrick G.T. Healey, Nik Swoboda, Ichiro Umata and Yasuhiro Katagiri.

[2]These experiments were designed by Patrick G.T. Healey, Nik Swoboda, Ichiro Umata and Yasuhiro Katagiri.
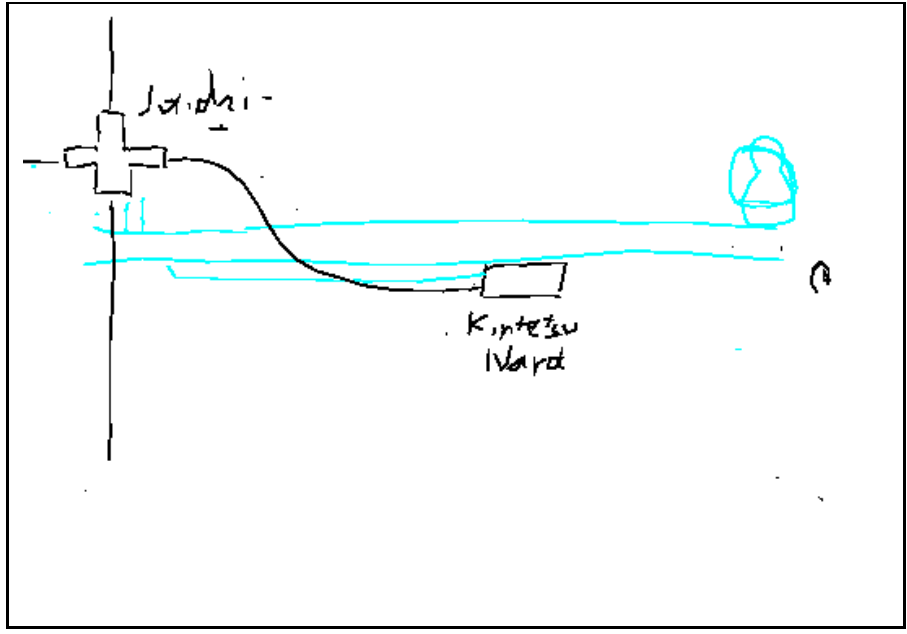
Figure 2.1: An example of the maps subjects drew

real world by the spatial relation on the graph. In all of thses tasks, the correspondence between the graphical representation and the world descrived by it was generally well-established.

2.2.1 Map Sketching Task

In this task, each pair of the subjects was asked to draw a sketching map of four landmarks in Nara (a local town). Pairs worked together to do this. We had sixteen pairs, and all of them drew sketch maps which roughly preserved information of distance and direction (Figure 2.1).

Each pair was seated in a separate, soundproof room and worked together using a shared virtual whiteboard and a full duplex audio connection. All input

to the screen was by stylus, and any writing or erasing by one participant would appear simultaneously on their partner's screen. The subjects were video-taped during the task.

The subjects were asked to imagine that they were drawing a sketch of a map for a tourist who did not read Japanese and who wished to find the landmarks on the list. The time limit for this task was five to eight minutes.

2.2.2   Graduate Record Exam (GRE) Task

The design was similar to that used in the map sketching task. This time, each pair was asked to solve a logical reasoning problem from the GRE. The problem was on the possibilities of route selections in a hypothetical truck delivery area with five fictive towns (ex. "What is the maximum number of towns, including the towns where the trip begins and ends, that can be served by a driver without overnight stop during the course?" etc.). The subjects were given the topological conditions on the area (ex. "Kitamati town and Hasimoto town are connected by a route and the distance is one hundred kilometer," etc.), and two conditions for drivers; (i) The maximam distance a driver can travel without making an overnight stop is 300 kilometers, and (ii) Overnight stops must be made in one of the given five towns.

Each pair was required to work on the problem together and it was suggested that drawing a diagram on the screen might help them to answer the question. All pairs drew a diagram and eight pairs among the nine drew a map-like one (Figure 2.2).

As for these eight pairs, working on this task consisted of two phases; to draw

13
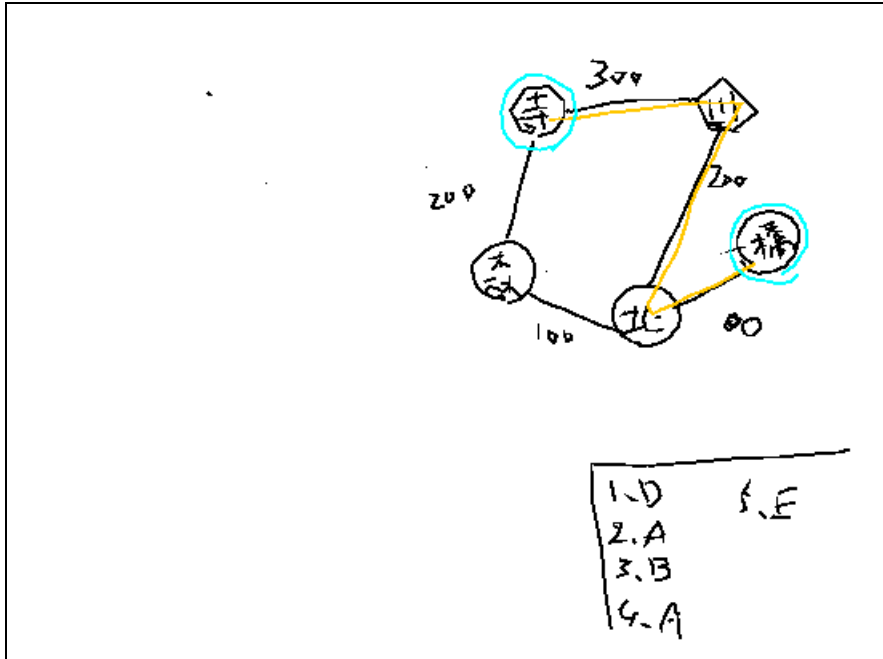
Figure 2.2: An example of the diagrams subjects drew

a map-like diagram and to answer the questions reading the diagram. The pair communicated with each other through a full duplex audio connection and worked on a shared whiteboard. The time limit for this task was eight minutes.

2.2.3 Graph Task

This task also had two phases. First, each pair of the subjects was given a scatterplot expressing the birth and death rate of certain countries, a data list of the birth and death rates of some of these countries, and conditions which held among some countries (ex. "the birth rate of Egypt is higher than China and lower than Nepal," etc.). The scatterplot was incomplete, with some of its dots left unlabeled and a few dots lacking. Each pair was asked to complete the graph properly (see Figure 2.3).

After completing the graph, all the data other than the graph itself were collected. Then, each pair was asked to answer the question relating to the birth and death rates of those countries, reading the graph they drew (ex. "How many countries are there which have higher birth rates than Thailand," etc.). The subjects were allowed to draw anything which could be a help for answering the questions. All the pairs newly drew something on the graph that they had completed in the previous part of the task (see Figure 2.4). There was no time limit in this graph task and people finished it in twenty minutes to forty minutes.

2.3 Individual Mediated Reference

2.3.1 Forward Individual Mediated Reference

Consider the following utterances:

15
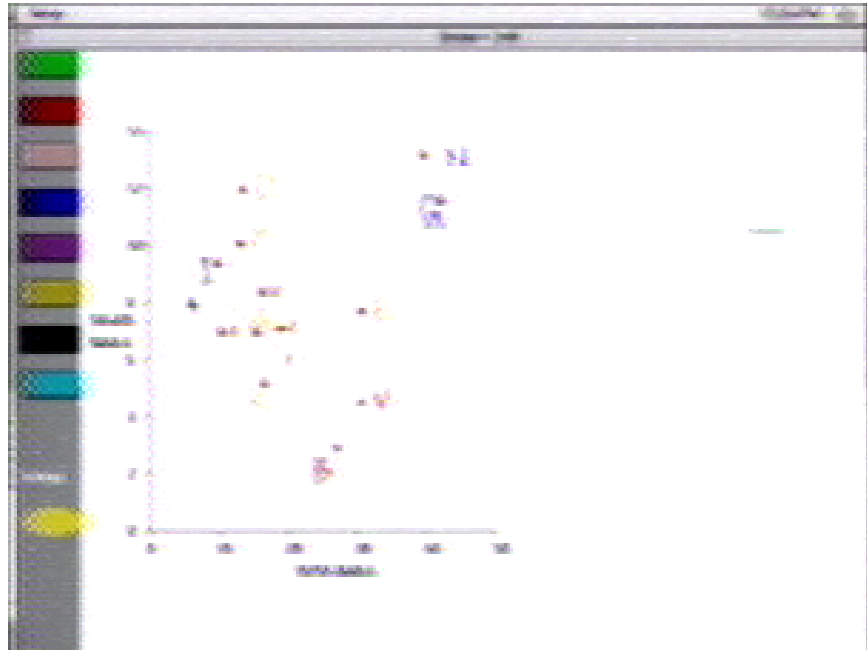
Figure 2.3: An example of the graphs subjects drew
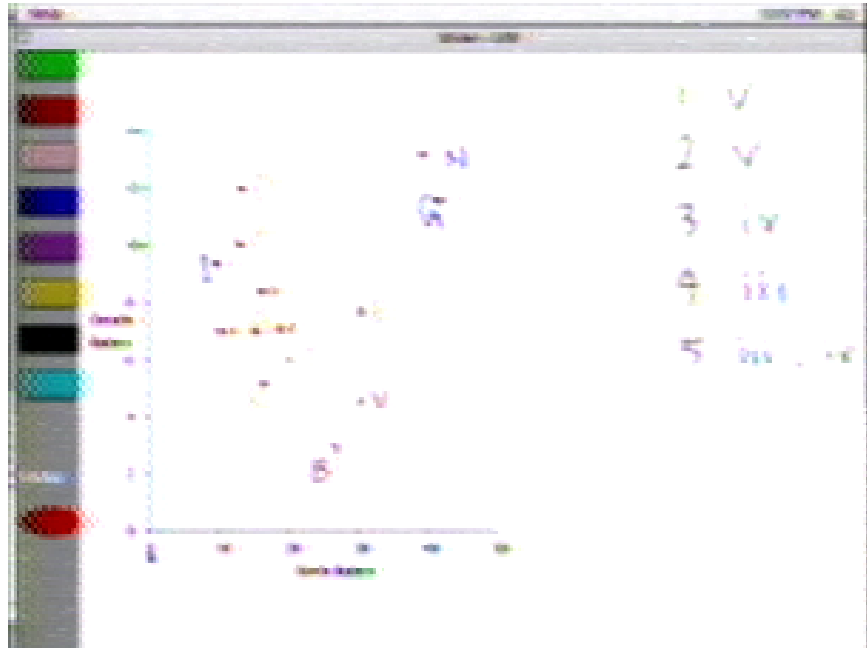
Figure 2.4: An example of the graphs subjects drew

(5) (From the map data: pointing to a part of the map with the stylus.)

de,   <u>*koko*</u>*-ni-ne,*   *tasika*   *Deiri-Sutoa-ga-ne,*
and   here-to   probably   Daily Store-NOM
<u>*kono*</u>   <u>*kado*</u>*-ni*   *atta.*
this   corner-to   was

"And I think there was Daily Store on this corner."

(6) (From the map data: pointing to a part of the map with the stylus.)

<u>*kotti*</u>*-ni*   *ittara*   *sika inai?*
this-way-to   go   deer   is-NOT

"You can find deer around here, can't you?"

(7) (From the GRE data, pointing to a path on the map with the stylus.

<u>*kore*</u>*-ga*   *100*   *desuka?*
this-NOM   100   is

"This is 100km, isn't it?"

The subjects were trying to draw a map of Nara combining their knowledge about the area in the map sketching task. In (5), the speaker was pointing to a part of the map, and the linguistic expression "*koko* (here)" and "*kono kado* (this corner)" literally denoted a part of the map. However, there was just a blank space on this part of the map and there were no symbols which could be regarded as an icon of a store. If we assume that the speaker was talking about the map situation as in (8), this utterance would be regarded as simply meaningless or at most false:

(8) There is an icon for Daily Store on the pointed part of the Map.

In this task setting, it is unlikely that the speaker was talking nonsense or lying. Therefore, this utterance was a statement discribing not the map but the real world situation as in (9).

18

(9) In the real world, there is Daily Store on the corner corresponding to the pointed part of the map.

Thus, the linguistic expressions "*koko*" and "*kono kado*" in (5) referred to some place in the world via the place on the map. Similarly, no deer-like pictures or signs could be seen on the map in the case of utterance (6), and no referent of the expression "*sika* (deer)" could be found on the map. Consequently, this utterance was also describing some real world situation in a specific place, not a place on the map. GRE data also have such references to the world object via an object on a diagram. There were no signs showing the distance on the diagram in the case of utterance (7), and no suitable properties for the referent of the expression "*100*" could be found on the diagram. This utterance was also a description of some situation regarding the delivery route, not a part on the diagram. In these cases, the reliable correspondence between the spatial configuration of a graphical representation and a place in the target world enabled *forward* mediated references: references to places in the target world through places on the graphical representation.

2.3.2   Backward Individual Mediated Reference

We can also find examples of *backward* mediated references in the data. Some of them are as follows:

(10) (From the map data: pointing to the icon of Nara Park on the map.)

ja,	*kore,*	*moo-tyotto*	<u>*kooen*</u>	*okkiku*	*suru?*
So	this	a-little-more	park	big	make

"So, shall we make this park a little bigger?"

19

(11) (From the map data: after realizing that they made a mistake.)

Sarusawaike-wo   *kesite, ...*
Sarusawa Pond-ACC   erase,

"Let's erase Sarusawa Pond, and..."

(12) (From the graph data: instructing his partner to mark an icon.)

Tairando   *kuroku*   *maru*   *sitoke*
Thailand   black   circle   do

"Mark Thailand with a black circle."

The linguistic expression "*kooen* (park)" in utterance (10) literally denoted a real park. However, one can not usually make a real park bigger, so this utterance can not be regarded as a statement about the world. Rather, utterance (10) was a statement about some operation on the map, and "*kooen*" referred not to the real Nara Park but rather to the icon on the map. Similarly, utterance (11) suggested erasing the map icon, not erasing the real Sarusawa Pond of the world. In utterance (12), the speaker instructed his partner to mark an icon with a black circle, not the real Thailand of the world. In these statements, the objects in the world played an intermediate role, and the linguistic expressions referred mediatedly to the icons on the map.

2.4   Relation Mediated Reference

2.4.1   Forward Relation Mediated Reference

So far, we have concentrated on mediated references between individuals. However, there are also examples of mediated references between relations in our data. Consider the following examples:

(13) (From the map data: pointing to a part of the map with the stylus.)

> *kokorahen-ni   Toodaiji-ga   aru*
> around-here-to   Todaiji-temple-NOM   is
> *kara,   kono   sita-no   hoo-kana?*
> because   this   below-GEN   direction-I wonder
> "Because Todaiji is around here, it (Kasuga-shrine) is probably below

this, isn't it?"

(14) (From the map data: looking at his partner's drawing on the map.)

> *aa,   soko   zutto   ue   nobotteiku-to,*
> Yes   there   straight   up   go
> *nyugakusiki-no   toko   dayo.*
> entrance ceremony-GEN   place   is
> "Yeah, if you go straight up there, you can find the place where we

had the entrance ceremony."

In dialogue (13), the speaker intended to show the listener where Kasuga-shrine was. However, because there were no Kasuga-shrine icons on the map, the subjects were talking about a real-world situation via the map. However, the linguistic expression "*sita*," which roughly means under or below, can not be regarded as referring directly to the real world relations; in the real world, Kasuga shrine is not under Todaiji temple but south of it. Here "*sita*" referred to the real world relation (i.e. to-the-south-of) mediatedly via the relation on the map (under). This was based on the semantic correspondence established between the map and the world. Similarly, in the case of utterance (14), there were no icons for "*nyugakusiki-no toko*" (the place where we had the entrance ceremony) on the map. Therefore, this utterance was describing a state in the real world, and the expression "*ue*" referred to the spatial relation in the real world (i.e. to the north) via the relation on the map (i.e. up) in (14).

21

2.4.2 Backward Relation Mediated Reference

Excerpt (15) includes an example of a *backward* mediated reference to a relation, as well as examples of individual backward references.

(15) (From the map data: revising the position of the Nara Station icon.)

> *Hanna Way-no*   *ue-ni*   *agattya*
> Hanna Way-GEN   above-to   go up
> *akan-ttekoto?*
> no good-Q
> *Hanna Way-no*   *yori*   *kita-ni*   *ittya*
> Hanna Way-GEN   than   north-to   go
> *akan?*
> no good-Q
> "So, it can't be above Hanna Way - we can't draw it north of Hanna
>
> Way?"

Since one can not change the place of the real train station, utterances in (15) were about the situation on the map. Thus, the linguistic expression "*Hanna Way*" in (15) referred backwardly to the icons on the map, not to a real world object. Similarly, the linguistic expression "*kita-ni*" (to the north of) makes a backward mediated reference to the relation of the places on the map (i.e. above) via the real world relation to the north of in this utterance[3]. Due to the preservation of the configurational constraints between the map system and the world, such mediated references to spatial relations are quite natural and are commonly found in conversations in which maps are used.

---

[3]In this utterance, "*agaru (aggattya)*" and "*iku (ittya)*" expressed the movement of the icon on the map. Actually, this was not a real movement. The subjects had erased the old icon and began to draw a new one. These phrases can be regarded as examples of metaphorical expressions, a subject that is beyond the scope of this paper.
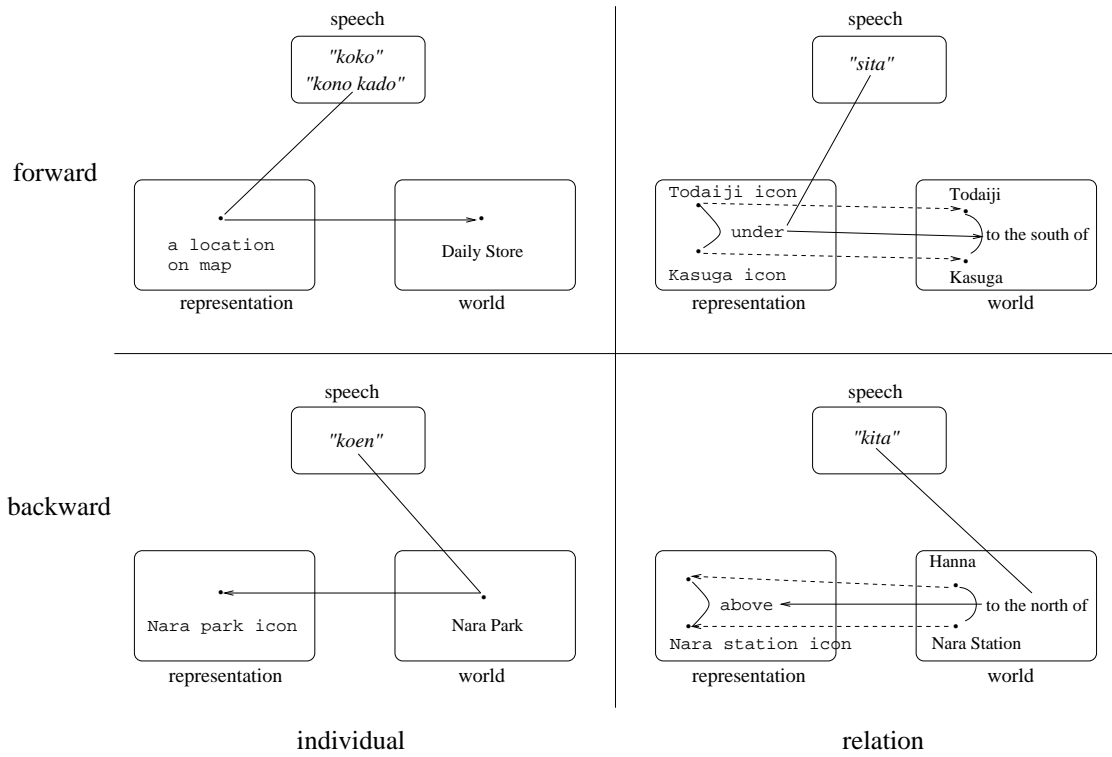
Figure 2.5: Four categories of mediated references

Four varieties of mediated references we found in our data are summarized in Figure 2.5.

Chapter 3

Disambiguation Strategies

Thus, when a graphical representation accompanies a dialogue, speakers can make a wide variety of mediated references, either individual or relational, forward or backward, through the systematic semantic relations established between the map and the mapped area. From the speaker's point of view, this means increased freedom of reference with a limited vocabulary, but from the listener's point of view, this might mean an increased likelihood that an utterance will become ambiguous in regards to the map itself or the mapped region. Although speakers have only one referent for each linguistic expression in their mind, hearers often have more than two candidates for the referent and have to resolve such ambiguity, due to the possibility of mediated references.

Fortunately, listeners can often rely on pragmatic cues to resolve such ambiguity, as we have seen in the cases of (5)–(15). Generally, listeners can reject an interpretation of a statement if, on that interpretation, an utterance is to perform a speech act that is not felicitous in that context. For example, it is unlikely that speakers are talking about a graphical representation when it includes no candidates for referents of some linguistic expressions they are using. In the utterances

(5), (6), (13) and (14), some linguistic expressions have no candidates for their referents on the map and thus they were considered not to be about the map but to be about the mapped region. Another pragmatic cues found in above examples are mismatches between the world and relations (or predicates) referred to. In the case of (10), (11), (12) and (15), speakers are requesting or otherwise discussing an operation on the mapped region that is impossible to address in the real world. These utterances are considered not to be about the mapped region but to be about the map. Utterance (7) mentions some property that is available not on the diagram but in its target world, and this should be considered to be about the the target world.

Thus, the pragmatic cues observed above are closely related to mismatchs between the domain (i.e. the world, a graphical representation) and possible referents (i.e. individuals, relations/ predicates). Mismatches between individuals and the domain are observed when candidates for individual referents of some linguistic phrase are not found in some domain. We call these *missing individuals* cases. Mismatches between relations/predicates and the domain are found when some rerations/predicates are inappropriate to be interpreted in some domain. We call these case *unsuitable relations/predicates*. The possible varieties of the disambiguation strategies are as follows:

Mismatches between the Domain and Possible referents		
	Graphical Representations	The World
Individuals	Missing Individuals in a Graphical Representation	Missing Individuals in the World
Relations/ Predicates	Unsuitable Relations/Predicates in a Graphical Representation	Unsuitable Relations/ Predicates in the World

All four varieties listed above are found in our data. In the following part of

25

this section, we will look into further examples from Map, GRE and the scatter plot task corpus, and examine possible disambiguation strategies of hearers.

3.1 Missing Individuals in a Graphical Representation

First, we will look into the cases in which we can find no candidates for the referent of some noun phrase in a relevant graphical representation. Consider these examples:

(16) (From the map data: pointing to a part of the map with the stylus.)

koko-ni-mo	*nanka*	*tera*	*rasiki*	*mono-ga*
here-too-to	something	temple	like	thing-NOM
atta	*kara.*			
was	because			

"There was some temple-like thing here, too."

(17) (From the map data: pointing to a part of the map with the stylus.)

Kasugataisya-tte	*zuutto*	*kotti-no*	*hoo*	*dayo.*
Kasuga Shrine-TOP	far	this-way-GEN	direction	is

"Kasuga Shrine is far to this direction."

(18) (From the map data: pointing to a part of the map with the stylus.)

e,	*ano,*	*Nara Kooen-no*	*tikaku-ni*	
uh	well	Nara Park-GEN	near-to	
atta-no-ga,		*Kintetu Nara*		*yannne.*
was-thing-NOM		Kintetsu Nara Station		is

"Uh, well, the station near from Nara Park is Kintetsu Nara Station, isn't it?"

26

In example (16), the speaker was pointing to a part of the map when he said "*koko* (here)," and these two expressions literally denoted that part of the map. However, as in example (5), there were no symbols drawn at that part and the linguistic expression "*tera rashiki mono* (*someting like a temple*)" could not find any candidates for its referent. It is infelicitous to talk nonsense during they are cooperating to do some task, and thus the hearer understood that this utterance was describing not the map but some real world situation. Here, the linguistic expressions "*koko*" should be interpreted referring mediatedly to some place in the world.

Similarly, the speaker was pointing to just a blank space on the map when he said "*kotti-no hoo* (this side)" in exapmle (17). Here again, no candidates for the referents of "*Kasugataisya* (Kasuga shrine)" were found on the map and the only possible interpretation for the hearer was that this utterance was describing some real world situation. The expression "*kotti-no hoo*" allowed the hearer only one interpretation in which the phrase referred to some part of the world via the pointed part of the map.

In the case of (18), there was the icon of Nara Park drawn on the map, and the noun phrase "*Nara Kooen* (Nara Park)" may have been referring either to real Nara Park directly or to the icon of it on the map mediatedly. However, the noun phrases "*Nara Kooen-no tikaku-ni atta-no* (the thing which was near from Nara Park)" and "*Kintetu Nara* (Kintetu Nara Station)" had no referents on the map, and the only possible interpretation was that the utterance was describing the real world situation. Thus, the noun phrases "*Nara Kooen-no tikaku-ni atta-no*" and "*Nara Kooen*" were interpreted as referring to the place of the real world

27

directly.

Thus, hearers can reject the possibility of reference to an object in a graphical representation when there are mismatches found between the graphical representation and individuals. When some of the linguistic expressions have no candidates for their referents in the graphical representation, the utterance is understood as describing the real world. In such cases, expressions literally refer to an object in a graphical representation are understood referring mediatedly to the world, and expressions literally refer to an object are understood as it is. Thus, hearers can resolve the ambiguity of direct/mediated references depending on lack of referents in some domain.

3.2 Missing Individuals in the World

Next, we will examine the cases in which there are no candidates for the referent of some noun phrase in the real world. Examples are as follows:

(19) (From the graph reading data: The subjects were tring to work out if there were any countries that had almost the same birth and death rate other than Italy. One of them thought that the birth and death rate of Denmark might be also close, and that they should check it with the inclination on the graph. She suggested putting the problem sheet on the screen like a ruler to make a line between the origin and the point marked as "Italy," representing real world Italy.)

atetemi,	*genten-to*	*Itarii-de.*
put	origin-and	Italy-LOC

"Put the problem sheet along the origin and Italy."

(20) (From the graph reading data: In a similar situation as in (12). One of the pair drew a line between the origin and the point marked D, representing real world Denmark.)

A: *kono-saa,*   zero-kara-no-saa,   ensen-joo-de-saa,   <u>D</u>-ga.
    this   zero-from-GEN   product-on-LOC   D-NOM
B: <u>D</u>-ga   *haittere-ba,*   <u>D</u>   *haitteru-naa.*
    D-NOM   is-on-IF   D   is-on

A: "On the product from zero, if D ..."

B: "If D is on the product, yeah, D is on it."

In the utterance (19), the noun phrase "*Itarii* (Italy)" was ambiguous: it may either refer directly to the real world Italy or refer mediatedly to the point on the scatterplot representing Italy via real Italy in the world. However, the noun phrase "*genten* (the origin)" had no candidates for its referent in the discourse relevant domain of the real world. Therefore, the only possibility left for the hearer was that the utterance was mentioning some operation on the graph, with the noun phrase "*Itarii*" mediatedly referring to the point on the graph.

The linguistic expression "*D*" in excerpt (20) was also ambiguous: it may either refer directly to the point on the scatterplot representing Denmark or refer mediatedly to real Denmark via the point on the graph. In this case, the noun phrase "*zero* (zero)" had no candidates for its referent in the discourse relevant domain of the real world. This lack of the referent candidate in the real world enabled the hearer to understand that the utterance was about the graph. Thus, he could resolve the ambiguity and interpret correctly that the expression "*D*" referred directly to the point of the graph.

29

These examples show that hearers can also resolve the ambiguity when there are mismatches found between the world and Individuals. When some of the linguistic expressions have no candidates for their referents in the discourse relevant domain of the real world, the utterance is understood as describing a graphical representation.

3.3 Unsuitable Relations/Predicates in a Graphical Representation

Mismatches between a graphical representation and relations also serve as cues to resolve reference ambiguities. Consider the following examples.

(21) (From the GRE data: The subjects were discussing the possibility of overnight stops when a truck driver traveled along a certain route. One of the pair pointed to the Kawabata icon saying "*koko-de*," and pointed to the Teramati icon saying "*Teramati-de*.")

saitee	*ippaku.*	*syukuhaku*	*nasi-de*	*unten*
at least	one overnight stop	overnight stop	without	drive

dekiru-no-ha,	*300*	*kiro*	*nande,*	*koko-de*
what one can do	300	kilometers	because	here-LOC

ippaku	*site,*	*Teramati-de*	*moo*
one overnight stop	do-and	Teramachi town-LOC	already

tuiteiru-to.
have arrived

"At least one overnight stop. Because you can't drive more than 300km without overnight stop, you have to stay overnight here. After that, you'll arrive at Teamachi town soon."

30

(22) (From the GRE data: Another pair was working on the same problem as in (21). One of the pair pointed to the Kawa icon saying "*koko-de*".)

ee,	*soo*	*desu-yo-nee,*	*Hasimoto,*	*Kawabata-de*
yes	so	is	Hashimoto town	Kawabata town-LOC

300-de,	*koko-de*	*yasumantoikemasen-yonee.*
300-is	here-LOC	have to take rest

"Yes, that's true. It's already 300km to Kawabata-town via Hashimoto town. You have to take rest here."

The linguistic expression "*koko* (here)" in (21) was ambiguous; it may have been referring either directly to the Kawabata icon on the map or mediatedly to the Kawabata town in the world. However, there were no predicates on the map which could be a candidate for the referent of the linguistic expression "*...-de ippaku site* (stay overnight at ...)," for we can not stay overnight in a icon of the town on the map. Therefore, the utterance was about some real world situation and the linguistic expression "*koko*" mediatedly referred to the Kawabata town in the real world. The word *Teramati* was also ambiguous, but the hearer could resolve the ambiguity and understand that it referred not mediatedly to the Teramati icon but directly to Teramati town in the world.

In (22), the linguistic expression "*koko* (here)" could refer either directly to an icon on the map or mediatedly to Kawabata icon in the world. However, the linguistic expression "*...-de yasumu* (take a rest or sleep at ...)" had no candidates for its referent. In this case also, the utterance described some real world situation, not that of the map. Thus, the word "*koko*" in (22) could be interpreted only as referring mediatedly to the real world.

31

As is shown above, mismatches between a graphical representation and relations/predicates play some important role when hearers try to resolve reference ambiguities.

3.4 Unsuitable Relations/Predicates in the World

Finally, we will look into the cases where mismatches between the world and relations/predicates. Consider the following examples:

(23) (From the graph drawing data: without any pointing.)

Tairando, kuroku maru sitoke.
Thailand black circle draw

"Circle Thailand in black."

(24) (From the graph reading data: without any pointing.)

Benezuera-wa kuro, ten tuketa.
Venezuela-TOP black point marked

Oosutoraria-no migiue, tyotto.
Australia-GEN to the upper right of a little

"I marked Venezuela with a black point. It's just to the upper right of

Australia."

(25) (From the graph drawing data: The speaker noticed that the hearer erased

the point expressing Brunei, and said.)

Burunei, kieteru-de.
Brunei disappeared

"Brunei has been erased."

In example (23), the linguistic expression "*Tairando* (Thailand)" had two reference possibilities: direct reference to the world or mediated reference to a point of the graph via Thailand in the real world. However, the verb phrase "*kuroku maru sitoke* (mark . . . with a black circle)" could not find any appropreate referents in the real world. Because one can not mark real Thailand with a black circle, it was nonsense to think the utterance was describing some real world situation. This incompatibility between the world and the predicate enabled the hearer to resolve the reference ambiguity of the word "*Tairando*" and understand the word was referring mediatedly to the point on the graph via real Thailand in the world.

Similarly, the linguistic expressions "*Benezuera* (Venezuela)" and "*Oosutoraria* (Australia)" in (24) could refer either directly to real Venezuela in the world or mediatedly to a point on the graph. Here again, the linguistic expressions "*kuro* (is black)" and "*ten tuketa* (put a dot)" could not find any appropriate referents in the real world. The linguistic expression ". . .*-no migiue* (is to the upper right of . . .)" was also incompatible with the real world for two reasons: one was that Venezuela is not to the upper right of Austrtalia in the real world, and the other was that the spatial relation of the two countries in the real world was totally irrelevant to the context of the utterance. Thus, the hearer could understand that the utterance was about some situation on the graph and that the linguistic expressions "*Benezuera*" and "*Oosutoraria*" could refer only to the points on the graph, not to the countries in the real world.

In (25), the word "*Burunei* (Brunei)" was also ambiguous. However, the verb phrase "*kieteru* (was erased)" could not find any appropriate referents in the real world. The word "*Burunei*" could be interpreted as referring only to the point on

33

the map which represents real Brunei.

These examples show that mismatches between the domain (i.e. the world, a graphical representation) and possible referents (i.e. indivisuals, relations/predicates) serve to resolve reference ambiguities in conversation. Thus, people can communicate effectively making various kind of references, without misunderstanding what the speaker refers to.

Chapter 4

Dual Descriptions

Thus, listeners can often rely on pragmatic cues to resolve reference ambiguities in graphical communication, as we have seen in the last chapter. However, our data also contain a number of utterances *not* subject to even such disambiguations. In the following, we will start with describing the phenomenon intuitively with a simple example, and then examine real-life examples of dual description. We will discuss communicative functions played by dual description and give a model of the informational mechanism responsible for such speech based on the data.

4.1   The Phenomenon

Suppose two men, A and B, talk about how to get around in London, with a small route map of London's underground system such as the one in Figure 4.1 sited below again.

Pointing to a particular area of the route map, the speaker A utters the following sentence:

(26) See, Baker is connected to Waterloo by Piccadilly.
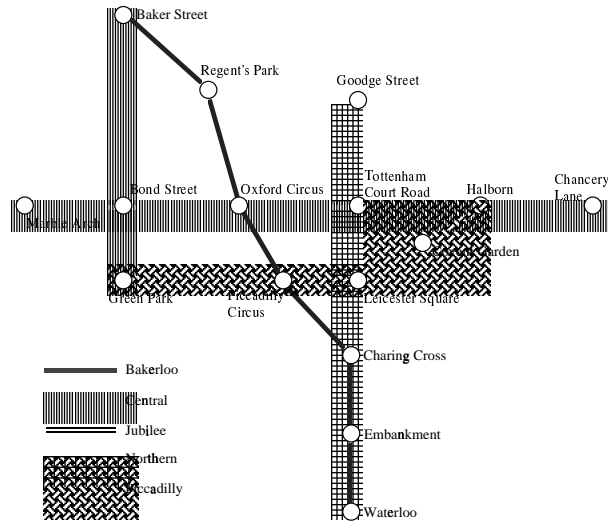
Figure 4.1: A Route Map of London's Underground System

What does this utterance exactly mean? What is B expected to "see" at this point? One possibility is that B is expected to see the map *itself*, and find the following information:

(27) The icon for the Baker station is connected to the icon for the Waterloo station by the band representing the Piccadilly line.

The other possibility is that he is expected to see *through* the map, to find:

(28) The Baker station is connected to the Waterloo station by the Piccadilly line.

In the first case, the part of A's utterance after "See" is an assertion about the structure of the map itself, while in the second case, it is an assertion about the London underground system depicted by the map. Which of these assertions does the utterance make? Or is the utterance simply ambiguous as to which?

36

The most intuitive answer, it appears to us, is that B is expected to "see" *both* information (27) and information (28), rather than a single one of these; the utterance therefore should be taken as an assertion about both the map and the mapped railway system, or more precisely, as informing B of the mapped system by informing us of the map itself. Thus, the information put forward by the utterance is not ambiguous, but determinate—it is just that the information is dual. In fact, our central claim in this chapter is that an utterance, when used with a different representation system, can convey two pieces of information in a systematic fashion.

Before proceeding any further to establish this main claim, however, we want to verify the correctness of our basic intuition, namely, that two pieces of information are put forward and made relevant by the above utterance of A. For this purpose we propose the following test.

Imagine that the partner B replies to A's utterance in question by denying it, using the sentential pronoun "that." Thus the dialogue goes as follows:

A: See, Baker is connected to Waterloo by Piccadilly.

B: No, no, that's not right.

Now what is B possibly denying in his reply? One possibility is that B is denying that the Baker icon is connected to the Waterloo icon by the Piccadilly band. In this case, B is simply correcting A for misreporting the map's structure.

This, however, is not the only possibility. B might be a native Londoner, and knows that the Baker station is not connected to the Waterloo station by the Piccadilly line. In this case, the target of B's denial may well be the information

about the connection between the Baker station and the Waterloo station in the real train system, as opposed to the connection between the Baker icon and the Waterloo icon on the map. Thus, B is correcting A for misreporting London's underground system depicted by the map.

Note that in the first case, the pronoun "that" refers to the information (27), while in the second case, the pronoun refers to the information (28). Thus, after A's utterance, B has the freedom to use the pronoun "that" to refer to either of the two pieces of information. Note that B would not have this freedom if more than one piece of information were not put forward by A's utterance. Thus, this test shows that in the utterance in question A puts forward two pieces of information for examination. A's speech is an instance of dual description, as we call it.

Generally, we can verify the occurrence of dual description in an utterance u accompanied by an external representation s in the following way. Suppose, after u, one replies to u with an utterance containing a sentential pronoun such as "that." If the use of the pronoun would be appropriate, and it could be used to refer to information about s itself or information about the target represented by s, then u is a case of xeroxing speech. Otherwise u is a case of singular informing.

4.2   Examples from the Corpus

Upon reflection, this type of utterance seems frequent in everyday dialogue involving some graphical representation. We do not always makes clear to listeners, or even to ourselves, whether our assertion or report is about the picture at hand or about the situation the picture depicts. In some of these cases, we end up making assertions on both the picture and the pictured situation, and communication

naturally continues as though such dual descriptions were a trivial semantic event.

This point is attested by numerous instances of utterances in our data, which pass the pronoun test discussed above and hence can be considered cases of dual description. This section looks over some of those instances to obtain a sense of how ubiquitous the phenomenon is and how it is being exploited in the actual contexts of communication.

The following dialogue from our GRE data was conducted just after the partners drew a graph-like map showing the routes connecting various towns, including Kawabata, Kitamati, and Hasimoto. The map is reproduced in Figure 4.2. The speakers are concerned with how many towns a truck driver can pass through in one day.

(29)

A: *kazoemasukanee?*

(Shall we count?)

B: *soosuruto.*

(If we do so, then....)

A: *kazoeruto 3-tu kanaa. 4-tu-wa tyotto muridesuyonee.*

(On my counting, it is three, I suppose. Four is not feasible, is it?)

B: *uun.*

(Hmm)

A: *Kawa toka dattara, kore moo sudeni 300 toka dakara, moo, Kita, kotti, Kita-ni*

39
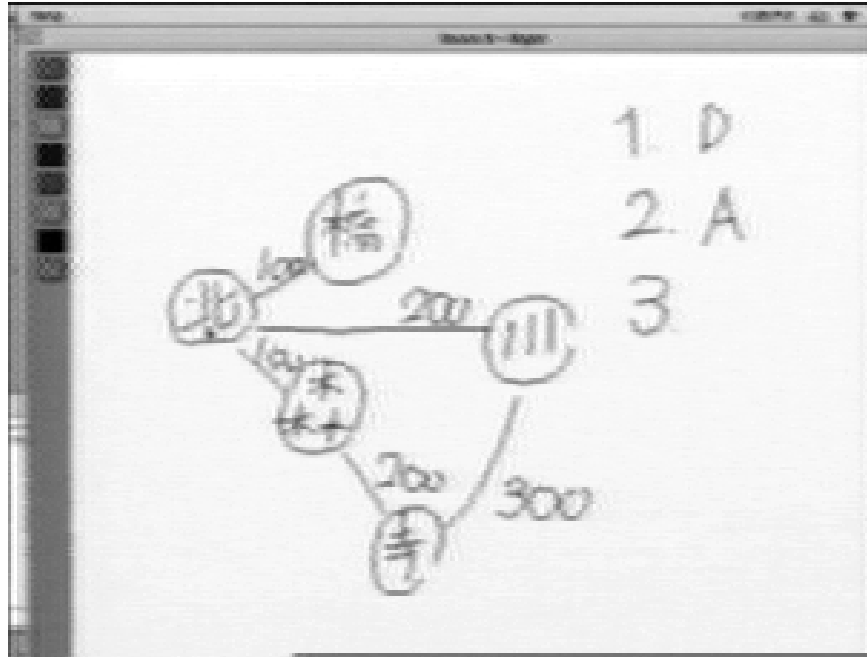
Figure 4.2: The Diagram Involved in the Dialogue (29)

(If this is Kawa or something, and if this is already 300 or so, well, Kita, here, to Kita
[Mumbling indistinctively.])

Kawa-kara Kita-ni itte Hasi-de, kore 3-tu desuyone.

(Going from Kawa to Kita and then to Hasi, that's three, isn't it?)

   The case in point is the last utterance of speaker A, which is underlined. Imag-
ine that you replies to this utterance by saying, "No, that's not true"a (or "iie
soreha chigaimasu" if you prefer speaking Japanese). What could you be denying
with this utterance? One possibility is that you be denying A's observation on the
map, whose content may be specified in the following way:

(30) There are three town icons on the path: the Kawa icon, the Kita icon and

the Hasi icon.

In this case, you are disputing the number of town icons on the particular path on the map. Alternatively, however, you could use the same sentence to dispute the number of towns on the real route denoted by the path, denying the accuracy of the following information:

(31) There are three towns on the route: Kawabata town, Kitamati town and Hasimoto town.

Thus, the sentential pronoun "that" used in your sentence could refer to either (30) or (31). This indicates that A's utterance in the last line of (29) passes our test, and A's utterance can be taken as xeroxing speech, carrying both information (30) and information (31).

In fact, the excerpt (29) shows that, prior to this utterance, the subjects have explicitly agreed to count the number of the town icons on the particular path on the map, and hence reporting information such as (30) is a speech act that A may well perform at this point. On the other hand, recall that the present problem for the subjects is the maximum number of towns that a truck driver can pass through within one day. Reporting information such as (31) is directly relevant to the solution of this problem, and hence is a speech act whose performance is desirable at this point. According to our analysis, A is reporting both in the single utterance, satisfying two related, but distinct informational needs at the same time. The dialogue (29) thus clearly illustrates that dual description significantly can contribute to the efficiency of communication.

41

There are also cases where the speaker exploits the informational duality not just for efficient communication, but also for efficient argumentation. The following is an example from our GRE task data, where the speaker works on the same truck-driver problem as in the example (29):

(32) (While speaker A is pointing to the map just drawn:)

A: *Mittuga gendodesyoo*

(Three is the maximum, I suppose.)

Yottudato sitara

(Since if it were four,)

Hyaku-hyaku-hyakuga naito ikenaikedo

(There must be [a sequence of] 100, 100, and 100, but)

Soreha sonzai sinaikara

(such [a sequence] does not exist.)

The utterance, "*Soreha sonzai sinaikara*," on the fourth line is the case in point. Imagine that in response to this, you utter the sentence, "No, that's not true." You can use this sentence to dispute the non-existence of a particular type of path on the map, but alternatively, you can also use the same sentence to dispute the non-existence of a particular type of route on the region represented by the map. You are denying the accuracy of the information (33) in the first case, while you are denying the accuracy of (34) in the second case.

(33) A path consisting of the segments respectively labeled with "100," "100," and "100" does not exist.
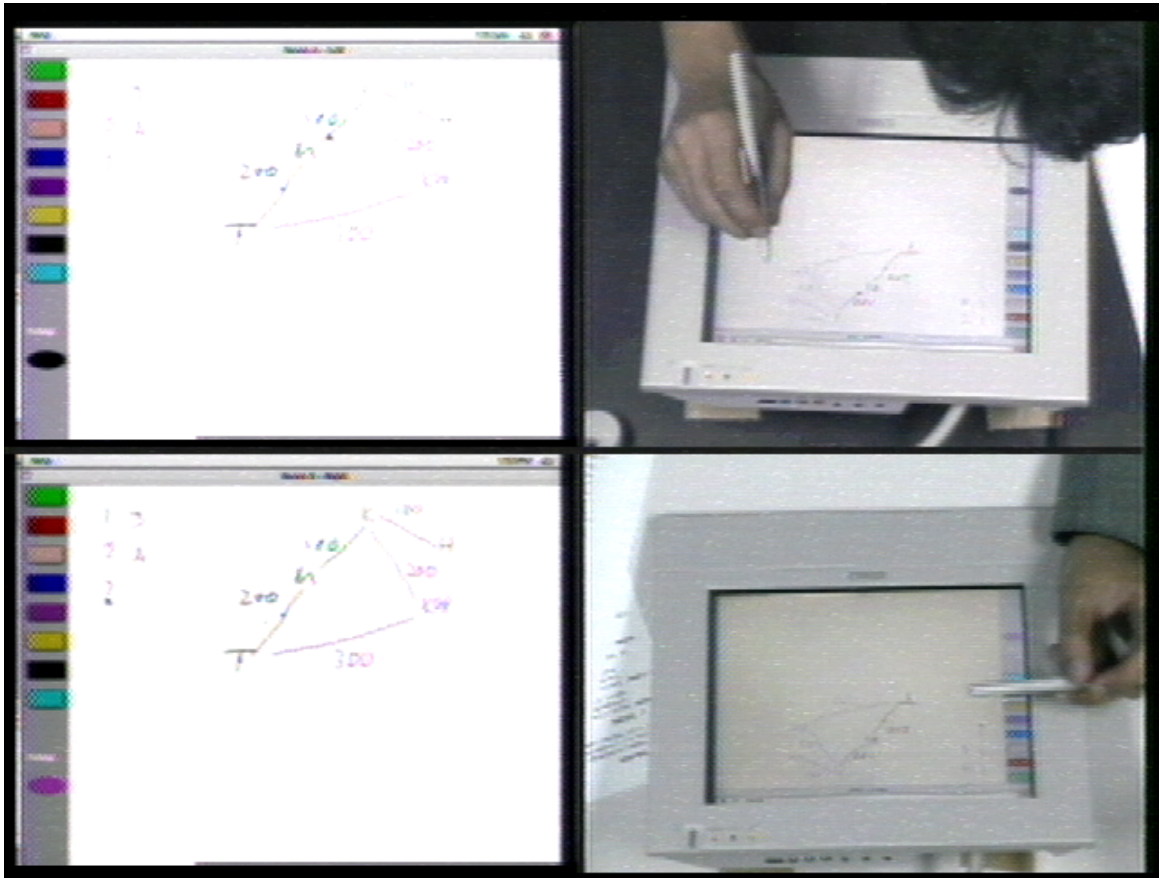
42

Figure 4.3: The Diagram Involved in the Dialogue (32)

(34) A route consisting of the sub-routes of 100 km, 100 km, and 100 km does not exist.

The utterance in the last line of (32) thus passes our test, and it can be taken to be xeroxing speech, carrying both information (33) and information (34). Note that here the speaker does not simply convey the two pieces of information (33) and (34) in a parallel manner. Rather, he puts them in an argumentative relation, using the information (33) about the map as evidence for the information (34) about the mapped region. The speaker then uses (34) to argue for the final conclusion that "three is the maximum," which is concerned with the traffic routes in the mapped region. If the utterance did not carry information (33) about the map, the speaker's point (34) would have no support, and if the utterance did not carry information (34) about the mapped region, the final conclusion would have no support. Thus, the logical link between (33) and (34), both presented in this utterance, is an integral part of the speaker's overall argument in this context.

Our data on the map-drawing task contains an example of still another way of exploiting the informational duality of an utterance.

(35) (The speaker has just finished drawing a map, reproduced in Figure 4.4.)

A: *Ja mazu kinnaraga kokoni aruyone*

   (Now, first of all, Kinnara is here, you see?)

B: *Ha*

   (Huh?)

A: *De Je Jeiaarunaraga kokoni atte*

   (And J, JR Nara is here, and)

44

B: *Ha*

(Huh)

A: *De Koohukujiga kokoni atte*

(And the Kohukuji Temple is here, and)

B: *Ha*

(Huh)

A: *Sarusawaikega kokoni atte*

(The Sarusawa Lake is here, and)

B: *Ha*

(Huh)

A: *Sinoomiyaga kokoni arutte iu huuni kaitanndawa*

(Sinoomiya is here—this is how I drew it.)

This sequence began immediately after speaker A has finished drawing a partial map. Now suppose the speaker B had replied to A's first utterance by saying, "No, that's not true," instead of the simple "Ha (huh?)." What could B have been possibly disputing with this reply? First of all, B could have been objecting to A's assertion on the location of an icon on the map, namely, to (36) below (where x stands for the position on the map the speaker is pointing to).

 (36) The icon for the Kintetu Nara station is at x.

Alternatively, B could have been disputing the location of the real Nara station, and in this case, the target of B's objection is the following information:
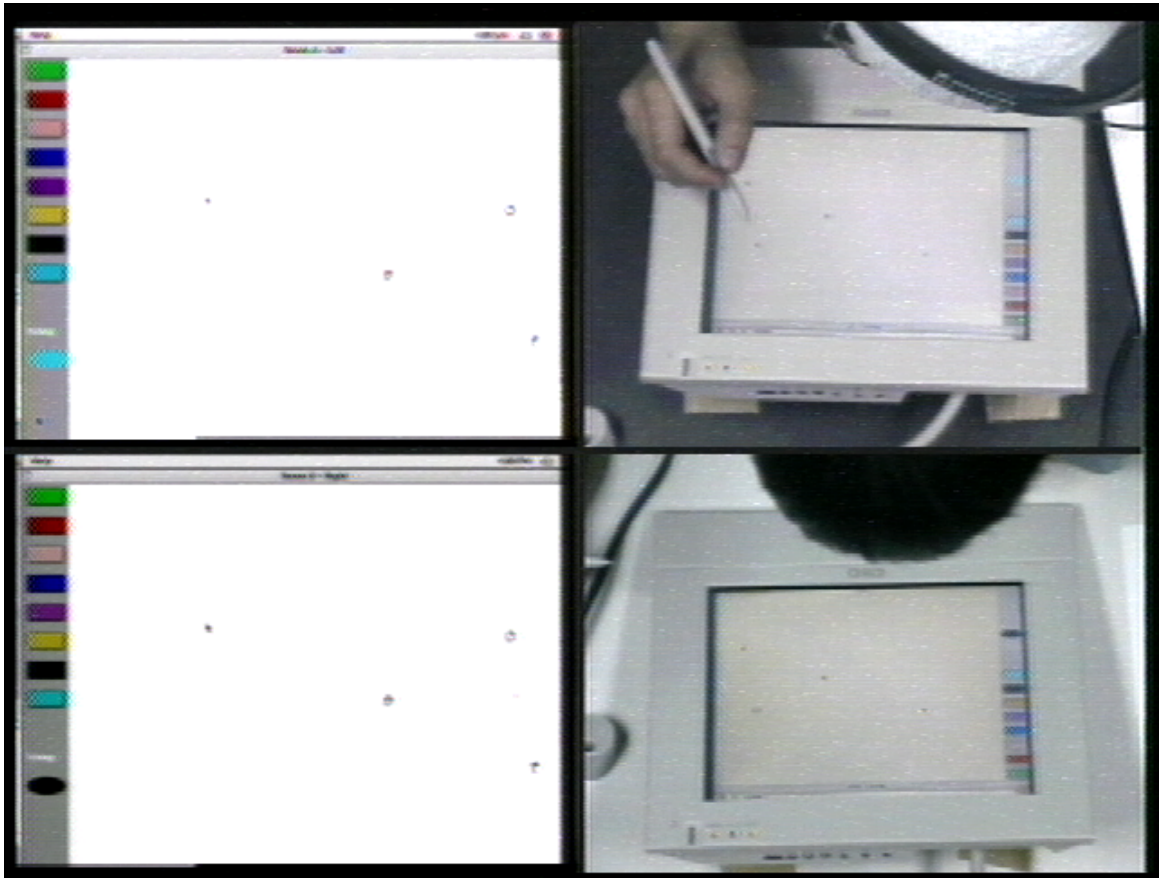
45

Figure 4.4: The Map Involved in the Dialogue (35)

(37) The Kintetu Nara station is in the position represented by x.

With this test, we can see that A's first utterance in the excerpt (35) is a case of dual description, putting forward both information (36) and information (37).

Unlike the previous example (32), the speaker here does not use the information in (36) about the map as evidence for the information in (37) about the mapped region. Still, he uses the logical connection of the two: if (37) is false, then (36) must also be false, if the map is to be accurate. Thanks to this logical connection, speakers can do a partial check of the accuracy of the map collaboratively: identify the particular property (36) of the map, and then evaluate the truthfulness of the information (37) carried by this property. (Speakers can do a full check of the map's accuracy by repeating this procedure for each property on the map carrying information about the mapped region.) In the above dialogue, speaker A proposes to do this collaborative checking: he uses his statements to identify the particular property (36) of the map, and proposes to check its appropriateness by evaluating the other information (37) presented in the utterance. The listener is supposed to evaluate (37) on the basis of his own knowledge, and to suggest a removal of (36) if he judges (37) to be false.

In fact, the rest of the dialogue can be viewed as a continuation of this procedure to a fuller check the map's accuracy. Thus, each underlined utterance serves as a proposal for the subjects to collaboratively check different properties of the map. As it turned out, the listener had little knowledge about the geography of Nara, and could not take part in the proposed act. This is why he keeps saying "*ha* (huh)" in reply.

Our final example is taken from the data on the graph-reading task. It is in-

47

tended to show that xeroxing speech could occur with a wide variety of graphical representations, not just maps. In the following dialogue, the two speakers use the scatter plot reproduced, in Figure 4.5, to find out what countries have lower birth rates and higher deathrates than the US.

(38) (The speaker A has aligned the left edge of her problem sheet vertically to the dot denoting the US, hiding the dots for the countries with higher birth rates than the US's.)

  A: *Kono mittsu desho*

    (<u>These three</u>, I suppose.)

    <u>Poorando Denmaaku Italii</u>

    (<u>Poland, Denmark, and Italy</u>.)

  B: *Souyane Poorando Denmaaku Italii*

    (That sounds right, Poland, Denmark, and Italy.)

Here the speaker B responds affirmatively to A's report, but if he said, "No, that's doesn't sound right," then it could be the denial of either A's report on the class of dots appearing in a particular area of the scatter dot, or else A's assertion on the class of countries that fall in the designated ranges of birthrates and deathrates. Thus, the underlined part of A's utterance can be considered to be xeroxing speech, where A reports the fact (39) about the scatter plot and thereby makes the assertion (40) about the situation depicted in the plot.

(39) The dotes labeled "P," "D," and "I" are to the upper left of the dot labeled "U."

Figure 4.5: The Scatter Plot Involved in the Dialogue (38)

(40) Polland, Denmark, and Itally have lower birthrates and higher deathrates than the US.

In summary, we found that the speaker can exploit the informational duality of dual description in the following ways:

1. Reporting two pieces of information in a single utterance, when both information about the graphic at hand and information about the depicted situation are required.

2. Making a simple argument in a single utterance, where the information about the graphic serves as evidence to the information about the depicted situation.

3. Doing a collaborative check of the accuracy of the graphic, where a speaker reports information about the graph at hand in view of the fact that the accuracy of the corresponding information about the depicted situation serves as a partial confirmation of the accuracy of the graphic.

4.3   Mechanism of Xeroxing

We now have seen several real-life examples of duasl description, where a single utterance appears to carry information about the graphical representation accompanying it along with information about the situation depicted by it. In this section we consider how such a thing is ever possible.

Briefly, dual description is possible because representation is a transitive relation. Due to this principle, whenever an utterance represents a picture as having

a property α, and this property α on the picture in turn represents the depicted object as having property β, the original utterance ends up also representing this object as having property β.

Recall the example of the route map of the London underground system. The utterance (28) can be considered to represent the route map as having the structural property of (29), and due to the semantic convention associated with the map, a map with the property (29) represents the mapped region as having the structural property (30). Thus, by transitivity, the utterance also represents the mapped region having the property of (30). This is how a single utterance carries two pieces of information, one about the map and the other about the mapped region.

This mechanism may be made clearer with an analogy to a copy machine. Suppose you make a copy d' of a document d, and then make a copy d'' of the copy d' that you just made. The copy $d,''$ being a copy of the copy $d,'$ represents d' more or less accurately, carrying information about $d.'$ Notice that this copy d'' also carries information about the original document d—we can look at the second copy d'' and learn what the original document d is like. (In fact, this is usually the main use of the second copy: we look at it in order to get information about the original document, often forgetting that it also carries information about the first copy.) Thus, the second copy d'' carries two pieces of information, one about the first copy, and the other about the original document. The second copy d'' carries the latter *via* the first copy, thanks to the transitivity of representation.[1]

Our claim is that the same thing happens in the case of the utterance (28),

---

[1]The idea that carrying information is a transitive relation is called "the Xerox Principle" by Dretske (1981), and has been a focus of interest in situation theory (Barwise and Perry 1983) and a subsequent development of qualitative information theory (Barwise and Seligman 1997).

where the London underground system is the original document d, the route map of it is the first copy d' of d, and the utterance (28) is the copy d'' of d'. The utterance (28) carries information about the London underground system via the route map, just as d'' carries information about d via d'. The utterance carries duplicated information (29) and (30) about the route map and the mapped region, just as d'' carries duplicated information about d' and d.

On this account, whenever an utterance reports a feature of a graphical representation, which in turn indicates a certain fact about the depicted situation, then the utterance carries information both about the picture and the depicted situation. Presumably, this special feature is best considered a *potential* possessed by such an utterance, something that may or may not be fully exploited in individual contexts of communication.

Nevertheless, the examples discussed in the previous section are clearly cases where the informational potential of such an utterance is exploited to serve various communicative purposes, and we are now in the position to be fairly precise about the informational mechanism underlying those examples. The general pattern is: (i) an utterance u primarily reports a feature α of the graphic at hand, (ii) due to the semantic convention associated with the graphic, α indicates another piece of information β about the situation depicted by the graphic, and (iii) by the transitivity of information carrying, the utterance u carries both information α and information β. Table 4.1 specifies each components of this mechanism as applied to the actual examples from our corpus.

Note that each semantic convention appealed to in the third column of Table 4.1 is quite natural to be associated with the graphic representation involved in

Utterance	Primary content, about the graphic	Relevant semantic convention of the graphic	Secondary content, about the depicted
Last utterance of (29)	(30)	(30) $\Rightarrow$ (31)	(31)
Last utterance of (32)	(33)	(33) $\Rightarrow$ (34)	(34)
First utterance of (35)	(36)	(36) $\Rightarrow$ (37)	(37)
A's utterance in (38)	(39)	(39) $\Rightarrow$ (40)	(40)

Table 4.1: Informational Model of Xeroxing Speech Applied to the Examples from the Corpus

the relevant case. Also, according to the present analysis, each instance of dual description discussed in the previous section primarily describes a feature of the graphic at hand, as opposed to the situation depicted by it. Thus, the last utterance of the dialogue (29) is considered to primarily report the result of the speaker A's counting of town icons; the last utterance of the excerpt (32) to primarily report the non-existence of a particular type of path on the map; the first utterance of the dialogue (35) to report the position of the Kinnara icon on the map; the speaker A's utterance in the dialogue (38) to report the result of her inspection of the labels of the dots falling in a particular area of the scatter plot at hand. We leave the reader to check the context of each dialogue to see that this interpretation of the primary concern of each utterance is plausible.

Thus, the model of dual description presented here appears to give a simple explanation of how a single speech may carry two pieces of information. Whether and how a speaker might be able to exploit this informational potential of an utterance to make it play specific communicative functions is a different issue, however. One might exploit it to serve two different informational needs in a single

utterance, as in the dialogue (29); one might use it to construct a little argument in a single utterance, as in the excerpt (32); one might use it to effectively check the accuracy of the graphic at hand, as in (35). Comprehensive research on what variety of communicative functions are served by dual description is, however, beyond the scope of this thesis, except for the few initial steps taken in the previous section.

Chapter 5

Quantitative Analysis

We have observed that a combination of graphical representation and linguistic representation in a graphical communication setting provides us with a novel sequential method for integrating of the linguistic and graphical modalities in the form of mediated and dual references. The analysis so far has been concerned with classifications and functions of instances of these new types of references.

In order to further establish that the sequential integration actually provides us with a viable and effective mechanism for communication, we conducted a quantitative analysis on the relative frequencies of the "new" forms of references, both mediated and dual references; we performed comparison with "conventional" direct references within our data obtained in our Map and GRE experiment. Furthermore, the different characteristics of each task were expected to result in a different distribution of the final referents of linguistic phrases. The Map corpus was expected to have more instances referring solely to the object in the graphics domain, because the aim of the task was to complete a map. On the other hand, the GRE corpus was expected to include fewer of such instances, because the aim was to solve the problems of the world domain and the graphics simply assist in
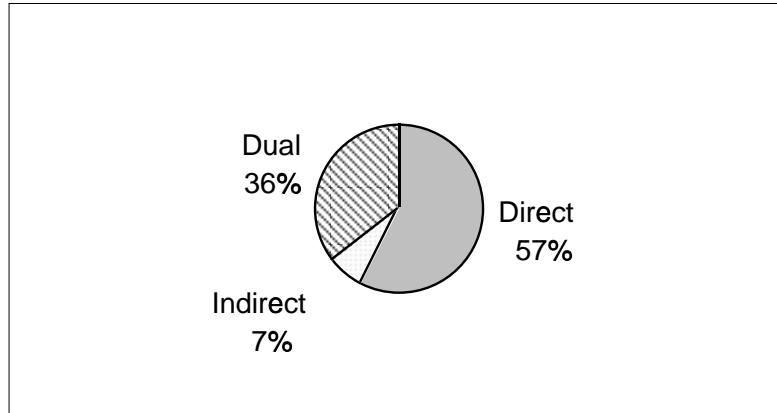
Figure 5.1: Relative usage frequencies of the direct, mediated references and dual descriptions

that purpose.

We analyzed a part of our corpus which consists of Map and GRE dialogues. It contains 14,011 words (9,179 for the Map and 4,832 for the GRE), and the number of content phrasal units[1] was 5,325 (3,394 for the Map and 1,931 for the GRE). Of them, 4,667 units were the ones describing the situations of the graphics and/or the world domain (2,875 for the Map and 1,792 for the GRE). We classified these units into the three categories shown above: direct, mediated and dual references.

Figure 5.1 shows the relative usage frequencies of the three types of references: direct, mediated and dual. Of all the reference occurrences, 57% were instances of direct references and 43% were instances of either mediated references or dual descriptions. Thisclearly shows that mediated references and dual descriptions are not mere theoretical possibilities or exceptional phenomena, but rather are mundane mechanisms routinely employed in actual communication.

---

[1]A content phrasal unit is a minimum phrasal unit that has a content word as its head.
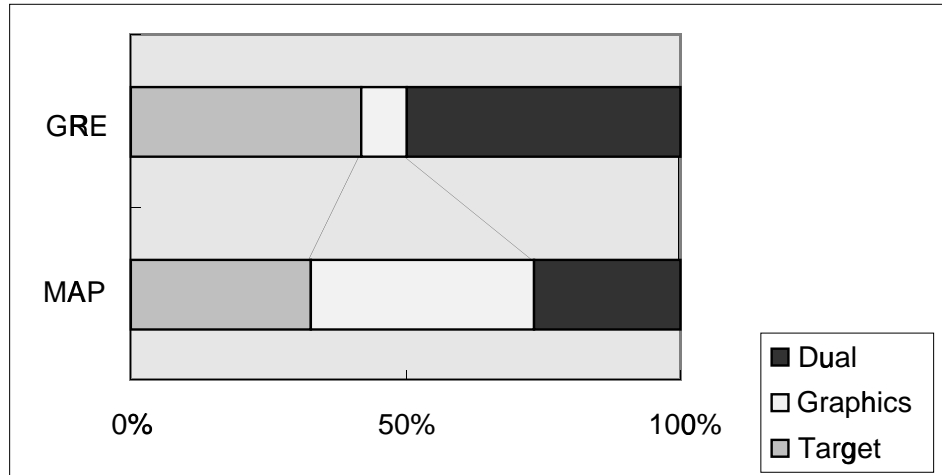
Figure 5.2: Distributions of final target domains of references for the Map and GRE data

Task characteristics of the Map sketching task and the GRE task can also be captured in quantitative terms. Figure 5.2 shows the distribution of final target domains of reference for the two tasks. A direct reference to the world and a forward mediated reference through the graphics to the world share the world as their final target domain of reference. Similarly, a direct reference to the graphics and a backward mediated reference through the world to the graphics eventually refer to the graphics as their final target domain. A dual description is indeterminate as to its final target domain. The final target domains exhibit significantly different distributions between the two tasks ($\chi^2{}_{(4)} = 595.60, p < .001$). More concretely, (1) the Map data had more instances of graphic-only references, (adjusted residual: Map = 23.75, GRE = $-$ 23.75); (2) the GRE data had more instances of world-only references, (adjusted residual: Map = $-$ 6.27, GRE =6.27); (3) the GRE data had more instances of dual descriptions (adjusted residual: Map = $-$

57

16.02, GRE = 16.02). Thus the assumption that the GRE data would have more world referents and fewer graphic referents than the Map data was supported. Furthermore, it is likely that dual references are strongly related to inferences on graphics, given the abundant instances of dual descriptions in the GRE data.

Chapter 6

Amplification of Information Contents

Thus far we have examined examples of both mediated references and dual descriptions and observed that such usage of language provides rich reference possibilities and communicative functions. Now we will observe that such usage provide us yet another way of amplifying information. Because linguistic expressions used in such ways indicate something about a graphic, they are in the position of exploiting the strong informational potentials of the graphic and thus they can convey more information than when an utterance directly describes the target. We will examine how amplification of information happens through sequential integration of representation systems in this chapter.

First, let us back to the case of London map shown in Figure 1.1. Consider the following utterances where A is pointing to the Goodge Street icon.

(41)

  A: Warren Street is above this.

B: So, it's to the north of Leicester Square, isn't it?

Since there is no icon of Warren Street on the map, the word "this" in A's utterance refers to real Goodge Street Station via the map icon. However, what is happening here is not just a mediated reference through the map. The information which A's utterance conveys is integrated with the map information so that the hearer can obtain the information about the spatial relation between Warren Street and Leicester Square. Neither the words nor the map alone can provide this information; the integration of language and graphics is indispensable in this case. In the following, we will examine the phenomena of information integration found in our corpus.

6.1 A simple Model of Information Integration

Now we will examine the phenomena of mediated references and information integration of representation systems with their targets. First, for linguistic expressions to refer target world objects via some representation, there should be some semantic relations between the representation system and the target world. For example, an icon for a station on the map represents some real station in the target world. These semantic relations between representation systems and their targets hold not only for individuals, but also relational predicates and propositions. A map relation **above** represents the target world relation "to the north of", and a map proposition "Goodge Street station icon is above the Leicester station icon" represents a target world proposition "Goodge Street station is to the north of Leicester Square station." Semantic relations like these are basically independent from the context of linguistic utterances; each icon represents its target building

even when there are no linguistic utterances. Notice that the semantic relations are not always truth preserving. A wrong map may contain propositions which do not hold in the target world. In this respect, a map can be regarded as a set of assertions on the target world which can be true or false.

Given these semantic relations between a representation system and its target, linguistic expressions can refer to objects of the target via its representation. Thus, mediated references make use of two different kinds of semantic relations transitively; one is the semantic relations between language systems and representation systems, the other is the semantic relations between representation systems and their targets.

Now let us consider the following exchange drawn from our MAP data.

(42) (From the map data: B is pointing to an icon on the map with the stylus.)

   B: *kore-ga Kintetu Nara?*
      (Is this Kintetsu Nara Station?)

   A: *tabun.*
      (Probably.)

Figure 6.1 shows the state of the shared map at this point, where the small rectangle near the right edge is the icon referred to as "kore (this)" by the speaker B. This exchange took place immediately after the speaker A finished drawing the icon. Given this context, A's obscure response, "Tabun (Probably)," is rather strange, since B's question was simply concerned with A's own intention about the icon that A had just drawn himself. What explains this indefinite answer?

61

Figure 6.1: Map Used in Dialogue (42)

Let us call the icon in question "a." If the speaker A intended a to be something other than the Nara Station icon or had no intention about what a is to be, then the answer should have been definitely negative. So, it seems that A indeed intended a to be the Nara Station icon, and he might well have answered, "Yes, it is." Suppose he had. Then, it would have indicated that the icon a is the Nara Station icon. Now, as Figure 6.1 shows, a was located to the right of the Todaiji Temple icon on the map (the other rectangle), so if a were the Nara Station icon, then the map would end up having the Nara Station icon to the right of the Todaiji Temple icon. According to the semantic conventions associated with the map, this last fact would indicate that Nara Station is to the east of Todaiji Temple. Thus, saying "Yes, it is" in this timing would commit the speaker to this extra information.

Figure 6.2(a) is a schematic view of this informational chain, where $\vdash$ is the

entailment relation holding on the map and $\Rightarrow$ is the indication relation determined by the semantic conventions associated with the map.

In fact, since the icon a has a variety of spatial relations with many other map elements, the sentence "Yes, it is" in this timing would have indicated much more extra information about the mapped region. For example, as the icon a is close to the band denoting Hanna Avenue, this mechanism would have made the sentence indicate the extra information that Nara Station is on Hanna Avenue. Similarly, the sentence would have also indicated Nara Station is to the east of Route 24 (denoted by the vertical line near the left edge). Thus, if one is unsure about one of the pieces of information thus produced, one might well hesitate to utter "Yes, it is" in this timing. The speaker A's hesitation to issue a definite positive answer in the dialogue (42) should be explained in terms of the amplified information content that such an utterance would have.

This same amplification phenomenon is observed in the following example:

(43) (From the map data: A is drawing an icon above the band for Hanna Avenue.)

A: *de:: Todaiji-ha konohen-da, konohen desho?*

(And Todaiji Temple is around here—around here, isn't it?)

B: *a sokka Hanna-no ue-ka.*

(Oh, yes, it's above Hanna Avenue.)

The speaker A was describing the location of Todaiji Temple as "konohen (around here)," while drawing an icon in a particular area b in the map. Since

63

"Yes, it is"

P: The icon a is to   Q: The icon a is   ⊢   R: The Nara Station icon
the right of the   the Nara Station   is to the right of the
Todaiji icon.   icon.   Todaiji icon.

R': Nara Station is to the
east of Todaiji Temple.

(a) "Yes, it is" in Dialogue (42)

α

Σ σ ⊢ σ'

θ'
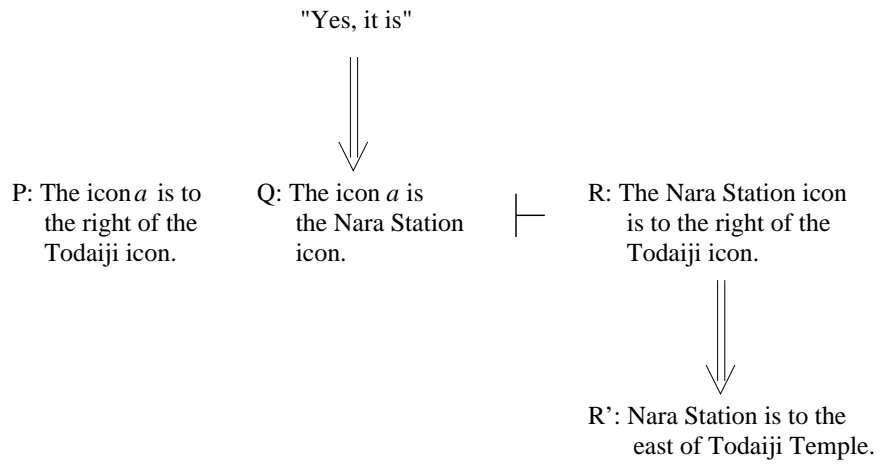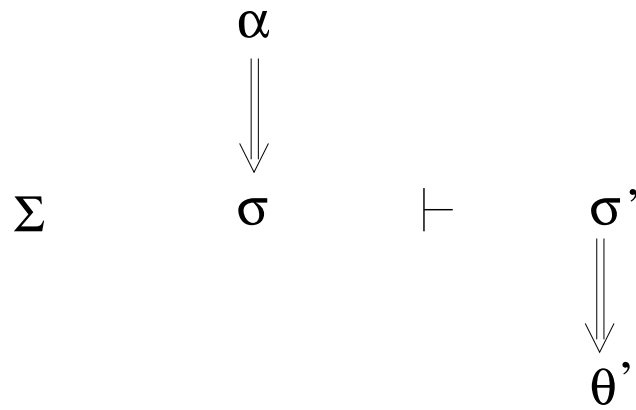
(b) General case

Figure 6.2: Informational chain resulting in amplification of utterance content.

the area b was above the band standing for Hanna Avenue, being in b in this map entails being above that band. By the semantic conventions associated with the map, being above that band indicates being to the north of Hanna Avenue. Thus, A's utterance ends up indicating that Todaiji Temple is to the north of Hanna Avenue. Apparently, the speaker B's response in the above dialogue was directed toward this extra information carried by A's preceding utterance. It should be clear that the informational chain involved in this extra content is isomorphic to the one depicted in Figure 6.2(a)[1].

The generalized schema underlying this mechanism is shown in Figure 6.2(b). Let α be a sentence and Σ be the set of all states of affairs holding in a map. Then, α indicates the information θ' as extra content if there are states of affairs σ, σ' about the map such that:

- α indicates σ_1 by the semantic conventions associated with the language,

- Σ plus σ entail σ' by a structural constraint on the map,

- σ' indicates θ' by the semantic conventions associated with the map.

Note that the states of affairs Σ already holding in the map and the entailment relation $\vdash$ holding on the map are essential in this informational chain. In other words, the sentence α obtains the added content θ' *only when* it is combined with a graphical representation in which Σ and the constraint "$\Sigma, \sigma \vdash \sigma'$" hold. Conversely, the sentence obtains as extra content regardless of *whatever* information

---

[1]The difference from the earlier example is that this involves the entailment relation at the level of properties (such as "being in b" or "being above the Hanna band"), rather than at the level of propositional information (such as "a is the Nara Station icon" or "a is to the left of the Todaiji icon"). An exact model of the content amplification would thus require us to distinguish these two cases in terms of, say, the distinction between closed and open formula or between non-parametric and parametric situation-types, but we will not get into this issue here.

θ' satisfies this scheme. As we have seen in examples (42) and (43), this easily results in much amplified information content of a sentence. Thus, our model shows that language becomes an extremely powerful informational device when combined with graphics.
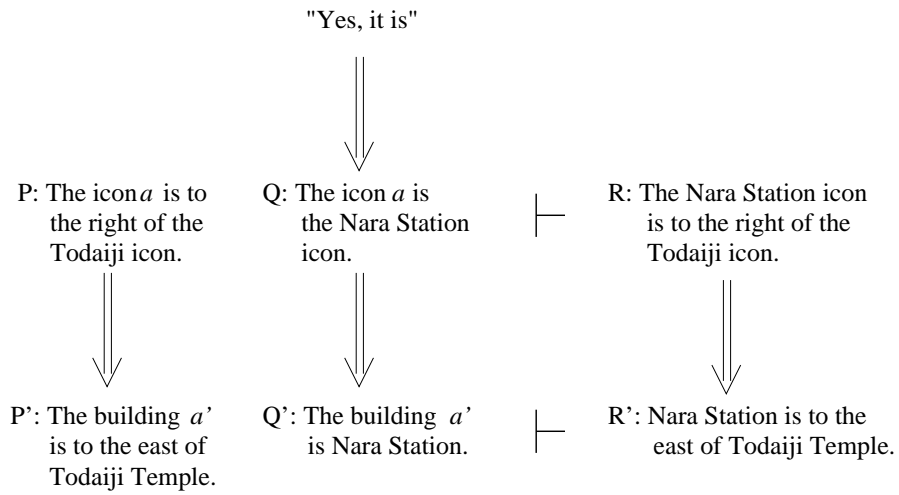
Corresponding to the informational chain depicted in Figure 6.2(b), however, there is generally an alternative route to the extra information θ'. Think of (42), again. According to the semantic conventions associated with the map, the states of affairs P and Q in Figure 6.2(a) indicate the following state of affairs about the mapped region (where a' is the building denoted by the icon a):

P': The building a' is to the west of Todaiji Temple.

Q': The building a' is Nara Station.

Notice that P' and Q' jointly entail R'. This means that, when we interpret the content of the sentence "Yes, it is," there can be two independent inference paths to the extra content R', as is shown in Figure 6.3(a). In one case, we first make an inference in the map domain from the information P and Q about the map to obtain the information R also about the map. We then project R to the target and obtain it's corresponding piece of information R' in the target domain. In another case, we first map each piece of information P and Q about the map to the target and obtain the corresponding pieces of information P' and Q' about the target domain. The conclusion R' is then drawn from them within the target domain. We call the former a *source-oriented* inference, and the latter a *target-oriented* inference. Generalizing this observation, we can add the target-oriented route in Figure 6.2(b) to obtain the revised general picture shown in Figure 6.3(b). Here,

66

"Yes, it is"

$$\Downarrow$$

P: The icon a is to   Q: The icon a is $\vdash$ R: The Nara Station icon
   the right of the   the Nara Station   is to the right of the
   Todaiji icon.   icon.   Todaiji icon.

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad\qquad \Downarrow$$

P': The building a' Q': The building a' $\vdash$ R': Nara Station is to the
    is to the east of   is Nara Station.   east of Todaiji Temple.
    Todaiji Temple.

(a) "Yes, it is" in Dialogue (42)

$$\alpha$$

$$\Downarrow$$

$$\Sigma \qquad\qquad \sigma \qquad \vdash \qquad \sigma'$$

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad\qquad \Downarrow$$

$$\Sigma' \qquad\qquad \theta \qquad \vdash \qquad \theta'$$
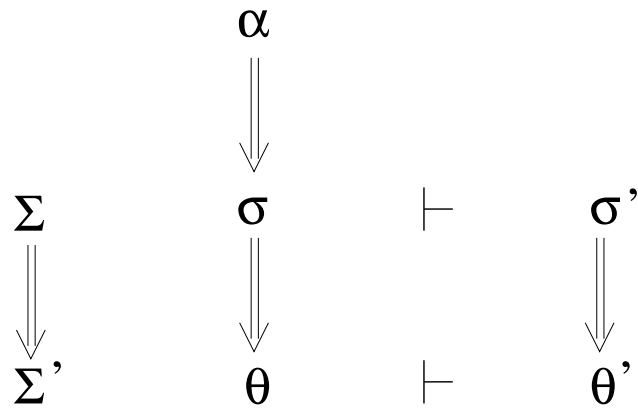
(b) General case

Figure 6.3: Mechanism of information integration of sentence α's content.

Σ' is the set of all pieces of information indicated by the facts Σ according to the semantic conventions associated with the relevant graphical representation.

In the target-oriented case, the information Σ' originating in the graphic and the information σ' originating in the sentence are combined at the final representation layer. In this sense, the target-oriented case is an instance of ordinary information integration of a linguistic expression and a graphical representation. In contrast, the source-oriented case is where the information P originating in the map and the information Q originating in the sentence are combined in the intermediate map layer.

These two inference paths should result in the same conclusion as long as the (upper) constraints $\vdash$ holding on the domain of graphical representation matches with the (lower) constraint $\vdash$ holding on the domain of targets. Therefore, from a strictly logical point of view, the new inference path made possible by the intervening representation layer does not produce any new information. However, if we think of actual situations in which people utilize graphical representations in reasoning, the difference in availability of information between the map and the target does make a difference in the relative ease with which people can perform inference.

In many cases where people utilize some form of graphical representations in reasoning, graphics is readily available, but the problem domain is hard to get at. A map, a picture or a diagram is presented on a sheet of paper, on a whiteboard or on a computer screen, all of which are easily accessible. But the problem domains themselves, the place represented by the map, the object depicted by the picture or the problem described by the diagram, often lie distant from them. Un-

der such situations it is quite natural for people to rely on graphics at hand and to apply a source-oriented inference rather than to go all the way to the distant problem domain and apply a target-oriented inference. The former must be easier, more efficient and less prone to errors. Their ease of access is not the only benefit of having intervening graphical representations. We can also directly manipulate them through various operations. We add and erase elements of graphics. We count and compare graphical objects. The results of operations on graphics become immediately available to us and can be exploited in further inferencing about the problem domain. In (43) above, the drawing made by A facilitates B's source-oriented inference by making various pieces of information easily accessible to him. This information is newly produced through the integration of the information already on the map together with that resulting from the drawing, namely, information on the relative position of the Todaiji Temple icon to the position of other icons in the map. Direct operations on graphical representations provide rich opportunities for source-oriented inferences.

6.2   Other Examples of Information Integrations

We have argued that mediated references make use of semantic relations transitively and that the transitive use of semantic relations invokes information integrations of the representation and its target. Now we will look into some of the examples of mediated references presented above more closely. Consider the following utterances again:

(5) (From the map data: pointing to a part of the map with the stylus.)

de,	*koko-ni-ne,*	*tasika*	*Deiri-Sutoa-ga-ne,*
and	here-to	probably	Daily Store-NOM
kono	*kado-ni*	*atta.*	
this	corner-to	was	

"And I think there was Daily Store on this corner."

(6) (From the map data: pointing to a part of the map with the stylus.)

kotti-ni	*ittara*	*sika inai?*	
this-way-to	go	deer	is-NOT

"You can find deer around here, can't you?"

The linguistic expressions "*koko*" and "it kono kado" in (5) referred to some place in the world via the place on the map, as we observed before. However, the pointed part of the map was not just a relay; the spatial relations between the point and the other landmarks on the map provided enough information for the hearer to know what part of the world was exactly referred to. In this case, the information related to the pointed part on the map was integrated with the information in the real world and provided new information to the hearer (ex. "Daily Store is to the south of Shin-Omiya station"). Thus, the use of mediated references through some representation involves information integration of information of the representation system into its target world.

In (6), the speaker was pointing to a part on the map, and this part had the property of being inside the Nara Park icon. By the semantic relation, this introduces to the scene of the target world the information that the real world place corresponding to the pointed part on the map had the property of being a part of Nara Park. The partner could get the information that Nara Park is the place where she could find deer integrating the map information into the target information, though she did not have enough information of the target world. The speaker

70

intended to describe the real world situation in this utterance and the word "*kotti*" referred to the area in the real world via the map.

Next we will examine the case of backward individual mediated reference. Consider (10) cited below again:

(10) (From the map data: pointing to the icon of Nara Park on the map.)

> *ja, kore, moo-tyotto   <u>kooen</u> okkiku suru?*
> So   this   a-little-more park   big   make
> "So, shall we make this park a little bigger?"

Before this utterance, the partner had pointed out that Kofukuji temple should be inside Nara Park in the real world. However, the map icon of Nara Park was too small to be able to include the Kofukuji icon without any inconsistency with the real world configuration, and the speaker suggested making the icon a litle bigger. The word "*kooen*" referred to the icon via real Nara Park, and the properties which real Nara Park had were introduced by the semantic relation between the real world and the map. In this case, the scene of the real world worked as a kind of representation of the map and the information introduced to the map by the semantic relation should be true for the map to be correct one. The newly introduced information served as the reason why they had to revise the size of the Nara Park icon, and the information integration provided an effective way of communication here.

Now we will look into the cases of the relation mediated references. Let us examine example (13) again:

(13) (From the map data: pointing to a part of the map with the stylus.)

71

>   *kokorahen-ni*   *Toodaiji-ga*   *aru*
>   around-here-to   Todaiji-temple-NOM   is
>   *kara,*   *kono*   *sita-no*   *hoo-kana?*
>   because   this   below-GEN   direction-I wonder
>   "Because Todaiji is around here, it (Kasuga-shrine) is probably
>
>   below this, isn't it?"

In this utterance, the speaker tried to show his partner that Kasuga-shrine was to the south of Todaiji. The word "*kono sita*" denoted the real world relation via the map relation. To stand in this relation implied to be below the Todaiji icon on the map. This introduced the information that the object in question should be to the south of Todaiji in the target world. Thus the speaker could communicate the information that Kasuga-shrine should be to the south of Todaiji in a effective way using the map.

Example (15) is the case of backward relation mediated references, cited again below:

(15) (From the map data: revising the position of the Nara Station icon.)

>   *Hanna Way-no*   *ue-ni*   *agattya*
>   Hanna Way-GEN   above-to   go up
>   *akan-ttekoto?*
>   no good-Q
>   *Hanna Way-no*   *yori*   *kita-ni*   *ittya*
>   Hanna Way-GEN   than   north-to   go
>   *akan?*
>   no good-Q
>   "So, it can't be above Hanna Way - we can't draw it north of
>
>   Hanna Way?"

Just before the utterance, the partner had pointed out that the configuration of the map was not in right order. Then they began to revise the map, and the speaker

asked the partner how the map should be to be correct. The phrase "*Hanna way-no kita*" referred to the map relation above the Hanna Way icon via the real world relation to the north of Hanna Way, and the information of the real world configuration concerning with Hanna Way was transferred to the information of the map configuration by the semantic relation. Though the speaker's knowledge of the real world configuration was not enough to add some new information to the map, the partner had enough information and pointed out where the Nara Station icon should be on the map. Here the question worked in quite effective way; the speaker could get the information where the Nara Station icon should be on the map to reflect the real world configulation properly.

Chapter 7

Graphical Representations and Perspectives of Motion Events

We have observed the phenomena of mediated references, dual descriptions and integration of information induced by such language usage. The sequential integration of language and a graphical representation enables efficient communication, providing rich reference possibilities and amplified information. However, the effects that sequential integration causes are not only those. Because of the handiness of graphical representations, people see, talk and think about the target world situations via graphics. This provides us with yet another kinds of viewpoint from which we capture the target world situations; the perspectives via a graphical representations. These perspectives often affect the way people capture the target world situations. In this chapter, we will look into this phenomenon forcusing on the concepturization of motion events in graphical communication settings. Usage of motion verbs in HCRC Map Corpus and our "Missionaries and Cannibals" Type Puzzle Corpus will be examined in the following, preceded by a section which presents categorizations of possible perspectives in graphical

communication.

7.1 Perspectives in Graphical Communication

When we expand the domain of discourse to include maps and graphical objects so as to encompass a whole set of communicative behaviors in graphical communications, we need to classify four types of perspectival event conceptualizations:

(a) Problem perspective

> The problem setting determines a uniform direction, from the initial state to the final goal state, throughout the entire space, the source of which makes the reference point of all instances of movements. The perspective can belong either to the real-world or to the map space.

(b) Protagonist perspective

> A movement is conceptualized from the viewpoint of an imaginary agent in a narrative world. In the graphical communication situations, a map provides the narrative domain for this perspective. The agent can be identified with either the speaker or the listener. This perspective belongs solely to the narrative world.

(c) Observer-to-World perspective

> A movement is taken as a movement in the real-world and conceptualized from the viewpoint of the observer within the real-world. This perspective concerns solely with the real-world.

(d) Observer-to-Map perspective

> A movement is taken as a movement in the map space and con-
> ceptualized from the viewpoint of the observer relative to the
> map. This perspective concerns both the real-world and the map
> space and makes the bridging between the two.

Among the four perspectives above, the Problem perspective is available in the problem domain, and makes a graphics-oriented sub-type when a graphical domain is taken as the problem domain. The Protagonist perspective typically works in fictitious stories. When used in graphical communication, the map space becomes the narrative space and makes a graphics-oriented sub-type by providing concrete and tangible graphical objects upon which the perspectival conception is laid out. The Observer-to-Map perspective presupposes the use of graphical representations, and is available only in graphical communication. These three perspectives constitute a set of perspectival event conceptualizations specific to graphical communication behaviors.

7.2 Analysis of the HCRC Map Corpus

7.2.1 Motion Verbs: "*come*" and "*go*"

Verbs like *come* and *go* reflect a speaker's reference point, as is shown by (1) and (2). When a speaker's reference point is set to the origin of the motion, the motion is expressed with the verb *go*. On the contrary, *come* expresses events in which a speakers' reference point is the goal of the motion (Figure 7.1).
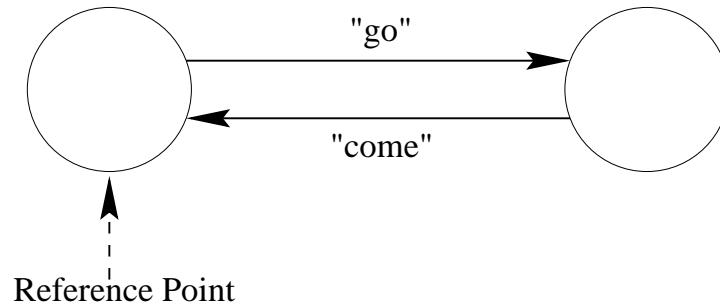
"go"

"come"

Reference Point

Figure 7.1: Reference point and Motion Verbs

A reference point is set from one of the perspectives shown in the previous section. For example, if one says, "Now we're going up north," the reference point is set at the speaker's current position from the Protagonist view. An utterance such as "Then you'll come to a meadow on the bottom of this map," expresses an event conceptualized from the Observer's view, whose reference point is set to the destination.

In conversation involving a graphical representation, people describe the target-world information through the graphical representation (see Umata, Shimojima and Katagiri (2000)). The target world is captured via its representation in such cases, and our prediction is that the configuration of a graphical representation will affect the conceptualization of an event in its target world. We examined the HCRC Map Task corpus by focusing on the usage of the verbs *come* and *go*.

7.2.2 Data

The data analyzed here is from the HCRC Map Corpus. This Map Task is a cooperative one involving two participants. The two speakers sit opposite one another, and one speaker gives instruction of a route to the other one. Each has

a map that the other cannot see, and a route is marked on the Instruction Giver's map while no route is marked on the Instruction Follower's one. The speakers are told that their goal is to reproduce the Giver's route on the Follower's map. Their maps are not identical and the speakers are told this explicitly at the beginning of their first session. It is, however, up to them to discover how the two maps differ. The maps describe fictitious areas. We selected and analyzed 16 conversations (non-eye-contact, unfamiliar pair condition) from the entire corpus.

7.2.3 Analysis

Of all the occurrences of *come* and *go*, only those that describe motion were analyzed here. The occurrences of "fictive motion" expressions such as "the bay goes like that," were not considered. The direction of the motion they describe are analyzed for all the occurrences.

The maps used in this task was fictitious one, and the subjects had no direct access to the target world of the map. The Observer-to-World perspective was not available in this task setting. The general motion direction set in this task was of course from the start to goal. Subjects would use *go* if they grasped motion from the Problem perspective. Almost all of the Instruction Givers adopted the strategy of giving their Followers local instructions step-by-step along their route on the map, which finally lead the Followers to their goals. Each step is motion from the current position of the Follower to some landmark in this case. The subjects were also likely to use *go* more than *come* if they captured mortion from the Protagonist's perspective. Therefore, *go* is expected to be generally prominent.

If we assume that the configuration of graphics affects the conceptualization

of motion events under the Observer-to Map perspective, then it is likely that the distance between the speaker and objects on graphics play the key role. The occurrence of *go* would be prominent anyway because of the reason described above. Therefore, *go* would be used widely to express either motion toward the speaker (*toward motion*) or motion away from the speaker (*away motion*) in the graphics. Because *come* is less likely to occur, it may be considered partly because of the configurational effect of graphics when it occurrs. Consequently, *come* is expected to be used more frequently to describe motion toward the speaker than motion away from the speaker.

7.2.4 Results

There were 238 occurrences of verb *go* and 56 occurrences of verb *come* used to describe motion. The distribution of motion described is shown in Table 7.1.

Table 7.1: Distribution of *come* and *go*

	away motion	toward motion	else
come	11	28	17
go	65	84	89

The frequency of *go* is higher than that of it come as was expected. The frequency of *away* motion is significantly smaller in *come* occurrences, but not so much in *go* occurrences. Thus the assumption that the usage of *come* is affected by the configuration of graphics is supported.

We cannot separate the occurrence of *go* into those with the Problem perspective and those with the Protagonist perspective, but the compound of those two

groups is prominent in the corpus. The distribution of *come* shows the configuration of graphical representation affects the conceptuarization of motion events.

7.3 Dialogues Involving a "Missionaries and Cannibals" Type Puzzle

We have demonstrated that the availability of graphics affects the usage of English motion verbs in the last section. However, it is still not clear how the possible perspectives interact. The goal of the HCRC Map Task was to reproduce the Giver's route on the Follower's map, and their movement on the map was almost always in one way progress up to the finish point. Therefore, the Problem perspective and the Protagonist perspective often shared their direction so much that it was difficult to distinguish these two from linguistic data. The dialogue data analyzed in this section were taken from collaborative problem solving experiment which involved back and forth movement. The Problem perspective and the Protagonist perspective were often expected to conflict in this setting. We analyzed the interaction between these perspectives making use of this conflict.

Another important feature of this task was that it involved two real world places which subjects were familiar with. The maps of HCRC Map Task describes fictitious world to which subjects were not directly accessible. We will also examine how much effect the Observer-to-World perspective has on the usage of motion verbs.

"iku (go)"
"tsurete-iku (take)"

"kuru (come)"
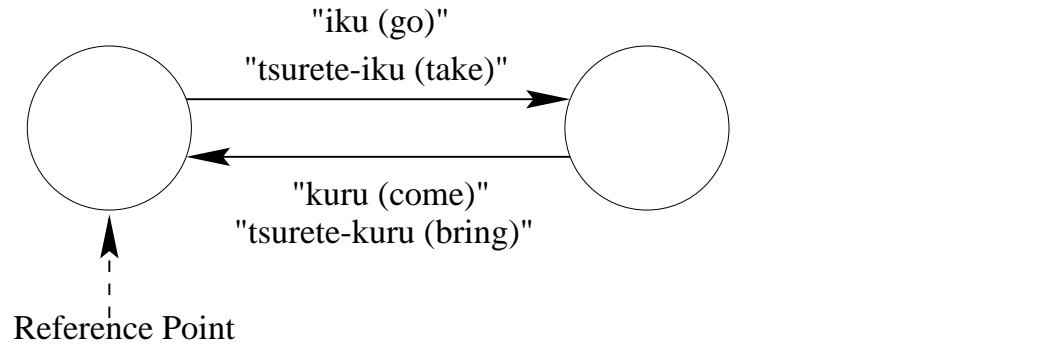"tsurete-kuru (bring)"

Reference Point

Figure 7.2: Reference point and Japanese Movement Verbs

7.3.1 Motion Verbs in Corpus

The Japanese language also has a pair of motion verbs similar to English *come* and *go*; *kuru* and *iku*. The reference point is set to the origin of the motion in the cases of *iku* (go) and *tsurete-iku* (take), and to the goal in the cases of *kuru* (come) and *tsurete-kuru* (bring)[1]. There are several verbs that can be classified into these two classes. We examined the usage of these two classes of verbs in the following two experiments.

7.3.2 Data

The data analyzed here is gathered from experiments involving problem solving. In this task, two subjects collaboratively worked on "Missionaried and Cannibals" type puzzles using a diagram given to them. The structure of the puzzle was basically the same as the original one, except it involved two actual places that the subjects were familiar with. The subjects were seated in separate, soundproof

---

[1] There is one clear difference, though. When a speaker is trying to go to the hearer, s/he will say, "I'll come to you," where *iku* (go) is used rather than *kuru* (come) in Japanese. However, this difference will not be relevant to the point here.

81

rooms and worked together using a shared virtual whiteboard and a full duplex audio connection. The diagram was shown on the whiteboards. All inputs to the screen were by stylus, and any writing or erasing by one participant would appear simultaneously on the partner's screen. The subjects were video-taped during the task.

7.3.3 The Mortorcycle Gang Task

The puzzle was almost the same as the original one, except that we used two actual places and replaced the missionaries and cannibals with two teams of mortorcycle gangs. The subjects were told to work out how all the members of both gangs could be transported safely. This task involves just two kinds of motion: forward and backward motion between two places. The time limit was seven minutes, including the time they used to read the problem sheet.

7.3.4 Experiment 1

The first experiment was conducted to examine the interaction between the availability of graphical representations and the Problem perspective. The motion in this task was much more simplified, though there was back and forth motion which was not in the HCRC Map Task. The problem involved motion between two actual places on a motorbike so that subject could also access directly to the real world situation. This task has a general direction of motion: all six boys have to move from *Saidaiji* to *Nara*. Those actual places were at almost the same distance from where the experiment was conducted. The bike was supposed to be able to carry only two people at one time, and someone had to ride back on it. Two kinds
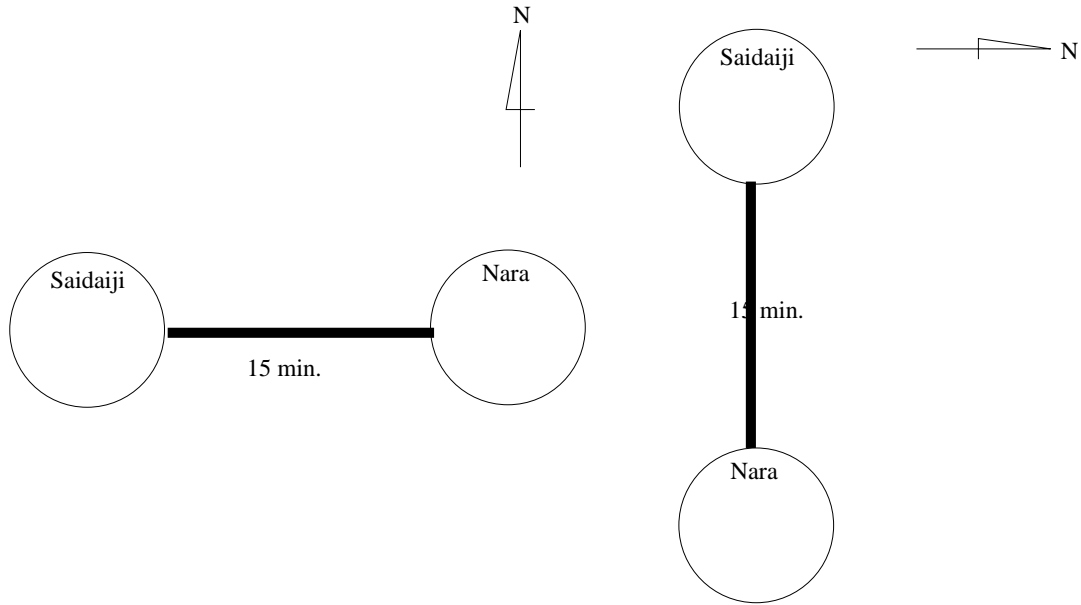
Figure 7.3: Horizontal and Vertical Diagrams

of graphical representations were provided as shown in Figure 7.3. One had horizontal configuration, in which the two icons of the places are at about the same distance from the subject[2]. The other one was with vertical configuration, which had variation in the distance from the subject and each place. Each condition had four pairs of subjects.

The assumption was that *kuru(come)*-type verbs would be used for the motion to Nara more frequently in the vertical condition than in the horizontal one because of the nearness of the Nara icon in the vertical condition.

Results of Experiment 1

The distribution of the motion verbs was as follows:

---

[2]Note that maps whose tops are not north are commonly found and not unnaturall.

Table 7.2: Distribution of *iku*-type and *kuru*-type verbs.

		iku-type	*kuru*-type
Horizontal	Saidaiji → Nara	21	0
	Nara → Saidaiji	2	10
Vertical	Saidaiji → Nara	8	7
	Nara → Saidaiji	1	9

Iku-type was the most frequent for Saidaiji to Nara in the Horizontal condition, and *kuru*-type was the most frequent for Nara to Saidaiji in both conditions. There was no occurrence of *kuru*-type for Saidaiji to Nara in the Horizontal condition, while the frequency of *kuru*-type was almost as high as that of *iku*-type in Saidaiji to Nara in the Vertical condition. There was few occurrence of *iku*-type for Nara to Saidaiji in both conditions.

The result shows that the Problem perspective was prominent: the subjects generally set their reference point to the general origin. However, the effect of the configuration of graphics was also observed. *kuru*-type showed as much frequency as *iku*-type for Saidaiji to Nara in the Vertical condition. This shows that the spatial relation between the speaker and the graphical objects affects the reference point setting. The handiness of graphical representation can be one of the cause of reference point shift. The low frequency of *iku*-type for Nara to Saidaiji suggests that the protagonist's perspective was weakest, considering that the starting point of each boy's movement should be the reference point in that perspective.

Thus, it was shown that the effect of the Problem perspective of task was the most prominent, but the configuration of a graphical representation often affects

the reference point setting of observer's viewpoint.

7.3.5   Experiment 2

The previous experiment showed that the Problem perspective was the most influential, while the configuration of graphics also affects the usage of motion verbs. Now we will look into the effect of the real world configuration in conversation involving graphics. The setup of Experiment 2 is almost the same as the previous one, except that the two places had variation in distance in the real world and that we used different diagrams than the previous ones for each condition. One of the places was the current position of the subject, and the other was a place away from there.

The diagrams used both had vertical configurations. The difference was that one diagram had a configuration consistent with the real-world relationship, while the other did not; that is, the nearer icon in the graphics represented a farther place in the real world. The general starting point was placed at the top of both diagrams. These diagrams are shown in Figure 7.4

If the real-world configuration has some effect on setting the reference point, the motion verbs will show different distributions between the consistent condition and the inconsistent condition. The frequency of *kuru*-type in start-to-goal motion is expected to be lower in the inconsistent condition than in the consistent one. If the real world configuration did not have much effect, then the distribution would be almost the same between these two conditions. Each condition had four pairs of subjects.
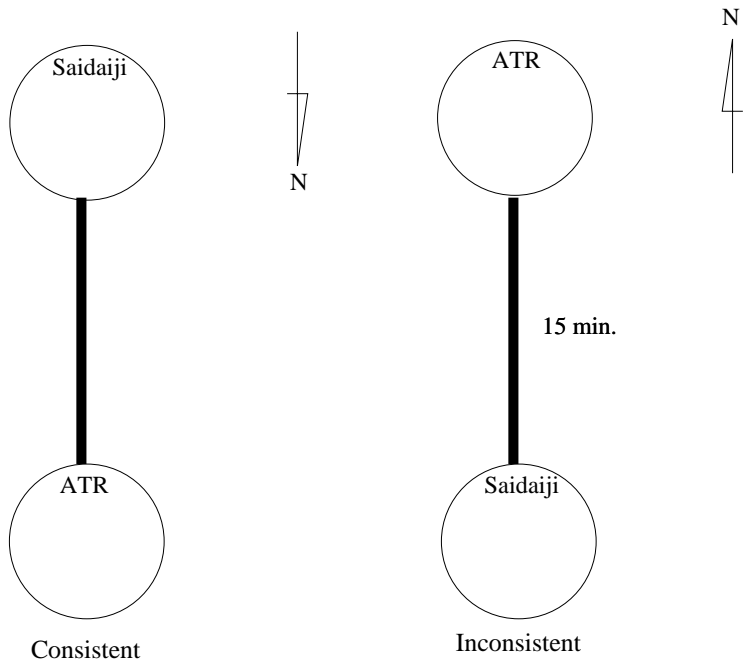
Figure 7.4: Consistent and inconsistent diagrams

Results of Experiment 2

The distribution of the motion verbs was as in Table 7.3.

Table 7.3: Distribution of *iku*-type and *kuru*-type verbs 2.

		iku-type	*kuru*-type
Consistent	Start → Goal	22	9
	Goal → Start	0	16
Inconsistent	Start → Goal	16	19
	Goal → Start	1	18

The distribution was almost the same as in the Vertical condition of Experiment 1. *Kuru*-type was observed in the Start to Goal motion in both conditions. However, *kuru*-type in the Start-to-Goal motion showed higher frequency in the inconsistent condition than the consistent condition. The frequency of *iku*-type for the Nara to Saidaiji motion was quite low again.

The results show that the spatial consistency does not contribute to shifting the reference point. The Observer-to-World perspective does not have strong influence in conversation with diagram. The Protagonist perspective is weak also in this setting. The reason why the frequency is lower in the consistent condition is not clear at the moment. It may be because the current position of the subjects was the general starting point of the task. This might enhance the general perspective by fixing the reference point to the current place. Further research will be needed to clarify the cause of this phenomenon.

At any rate, it is obvious that spatial consistency does not enhance the referent point shift. The spatial property of graphics has stronger effect on event conceptualization than that of its target world. This suggests that the Observer-to-Map

87

perspective is stronger than the Protagonist perspective in graphical communication settings.

7.4 Discussions

The analysis of the dialogue corpora discussed in the previous two sections showed that, among the three graphics-specific perspectives, the Problem perspective was the dominant event conceptualization, and the Observer-to-Map perspective worked as a somewhat weaker alternative conceptualization, particularly for mesial movements.

It might be argued that this apparent dominance relation could be the result of the communication setting, where two participants were seated in separate rooms and could not see each other. They had to guess the relative orientation of their partners toward their maps through their interchange, which might have made it difficult for them to employ, or otherwise discouraged them from employing, relationships between their maps and themselves. In situations where the information about the relative orientation of each participant with respect to her map is all shared among the participants, they might more willingly resort to the Observer-to-Map perspective. Although we need to perform another experiment for a definite answer, the Problem perspective dominance seems to be highly plausible, since it is the most reliable and free from error among the three because of its independence from relative orientation of maps and participants with respect to each other.

The Protagonist perspective seems also to be weaker than the Problem perspective as was indicated by the contrast between the experiment 1 and 2. The

Protagonist perspective often overlaps with the Problem and/or the Observer-to-Map perspective, and our analysis was not successful in systematically separating the Protagonist perspective *per se*. We have identified, however, a few of its specific instances. The following example was found in the HCRC corpus:

(44) Then, go to your right for almost the same, maybe five centimetres, til you come to the lagoon.

Here, the use of "your" in the expression "your right" suggests that the speaker is imagining a protagonist for the listener on the map, and is speaking from its perspective. The direction indicated by "your right" was actually the left from the speaker's Observer-to-Map perspective in this example.

In the following example, the speaker is imagining two protagonists for the speaker and the listener, describes the movement of the listener protagonist from the speaker protagonist's perspective.

(45) . . . then come down to the rife valley . . . which is where I am now . . .

The examination of these instances suggests that the Protagonist perspective has a much finer structure than the other two perspectives, and the speakers seem to employ it in specialized circumstances.

Chapter 8

Conclusion

So far we have observed the sequential form of integration between spoken language and graphics taking place in graphical communication settings. Based on the data of spontaneous spoken dialogues involving graphic representations, we analyzed the semantic behaviors of phrasal units appearing in speech, as well as the pragmatic roles utterances play at the sentential level. We found:

(i) a pre-established semantic relation between a graphic and the situation depicted by it provides the speaker with rich possibilities of mediated references, including forward individual, backward individual, forward relational, and backward relational references.

(ii) the same semantic relation also lets the speaker use a declarative sentence to express dual pieces of information.

(iii) mediated reference and dual description are not exceptional but rather mundane mechanisms routinely employed in actual communication.

(iv) the characteristics of communicative contexts affect the distributions of the final referents of linguistic phrases.

We further suggested that the use of dual descriptions is strongly related to inferences on graphics. These findings indicate that in spontaneous human communications, spoken language and a graphic representation may be used in the *sequential* composition, where the latter affects the usage of the former to extend its expressive capacity. This is in stark contrast to the common view of the interaction between linguistic and graphic modalities, where the integration is made only at the level of multiple pieces of information expressed by the two modalities in individual manners.

We then developed an analysis of information integration provided by sequential integration of language and graphics. We found that a fact already holding in the graphic representation is combined with the sentence to convey an extra piece of information.

The existence of graphics not only enables efficient communication but also provides graphics related perspectives. We have shown that the configuration of a graphical representation affects the reference point setting when people conceptualize motion events, based on the empirical analysis of the uses of motion verbs in actual conversational data.

We proposed a four-way classification of possible perspectival event conceptualization for graphical communication:

- Problem perspective

- Protagonist perspective

91

- Observer-to-World perspective

- Observer-to-Map perspective,

of which all but the Observer-to-World perspective involve graphical representations.

We have found that:

(v) All three perspectival conceptualizations involving graphical representations are employed in actual conversation,

(vi) The Problem perspective is the strongest among the four types of perspectival conceptualizations,

(vii) The Observer-to-Map perspective is the next strongest and

(viii) The real-world perspective does not contribute, in comparison with the graphics perspective, to the reference point shift.

These results suggest that we are mainly grasping an event of the world via its representation rather than the event itself in graphical communication situations. Thus, the point of the graphical representations is the convenience and the ease of access they provide us, which help us to grasp an event through the mediation of graphics, and this mediation makes it possible to talk about distal objects by manipulating their proximal counterparts, thereby facilitating both communication and reasoning processes.

Thus, the sequential composition of language and graphics is a common phenomenon that enables an efficient way of communication and affect event conceptuarization introducing perspectives via graphical representations. Other types of

sequential composition between linguistic and graphical modalities are quite probable. It would also be interesting to see what types of sequential composition take place when other types of graphical representations are involved. Development of a detailed informational model would greatly contribute to further characterize the nature and the workings of the sequential composition. It would also be interesting to look at different modalities to elucidate the possibilities than language and graphics, as well as to try out different ways of combining a set of modalities to elucidate the possibilities and characteristics of the sequential composition phenomena. A *parallel* composition is not the only form of the graphic-linguistic integration, and quite probably, not even a dominant form.

References

Barwise, J., and J. Perry (1983). *Situations and Attitudes*. Cambridge, Mass.: MIT Press.

Barwise, J., and J. Seligman (1997). *Information Flow: The Logic of Distributed Systems*. Cambridge, U.K.: Cambridge University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press

Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, Mass.: MIT Press.

Jackendoff, R. S. (1975). On Belief-Contexts. *Linguistic Inquiry 6*, 53–93.

Lakoff, G (1987). *Women, Fire and Dangerous Things*. Chicago: The University of Chicago Press.

Lakoff, G., and M. Johnson (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Lakoff, G., and M. Turner (1989)] *More than Cool Reason*. Chicago: The University of Chicago Press.

Lee, J., and H. Zeevat (1990). Integrating Natural Language and Graphics in Dialogue. In Diaper, D., Gilmore, D., Cockton, G. and B. Schackel. (Eds.) *Human Covmputer Interaction—INTERACT'90*, 211–234 Amsterdam:

Neilson, I., and J. Lee (1994). Conversations with Graphics: Implications for the Design of Natural Language/Graphics Interfaces *International Journal of Human-Computer Studies 40*, 509–541.

Nunberg, G. (1993). Indexicality and Deixis *Linguistics and Philosophy 16*, 1–43.

Quine, W. V. (1968). Ontological Relativity *Journal of Philosophy 65*, 185–212.

Schwartz, D. L. (1995). Reasoning about the Referent of a Picture versus Reasoning about the Picture as the referent: An effect of Visual Realism *Memory & Cognition 23*, 709–722.

Shimojima, A. (1999). The Graphic-Linguistic Distinction: Exploring Alternatives *Artificial Intelligence Review 13*, 313–335.

Suwa, M., and B. Tversky (1997). How Do Designers Shift Their Forcus of Attention in their Own Sketches? *Papers from the AAAI-97 Fall Symposium*, 102–108. AAAI Press.

Umata, I., Shimojima, A. and Y. Katagiri (2000a) Talking through Graphics: An Empirical Study of the Sequential Integration of Modalities *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 529–534.

Umata, I., Shimojima, A. and Y. Katagiri (2000b) An Informational Analysis of the Mediated Use of Language in Graphical Communication *Proceedings of*

the Workshop on Integrating Information from Different Channels in Multi-Media-Contexts at ESSLLI 2000, 48–55.

Umata, I., Shimojima, A. and Y. Katagiri (2000c) The Twofold Structure of Depiction: Modal Integration of Language and Graphics *BI-Metonymy 6th to 8th of October, 2000 Proceedings*

Umata, I., Katagiri, Y. and A. Shimojima (2001) Graphically Speaking: Do Graphics Affect the Perspectives of Event Conceptualization? unpublished manuscript.