A Model for Multimodal Reference Resolution

Luis Pineda[*] and Gabriela Garza

An important aspect of the interpretation of multimodal messages is the ability to identify when the same object of the world is the referent of symbols in different modalities. To understand the caption of a picture, for instance, one needs to identify the graphical symbols that are referred to by names and pronouns in the natural language text. One way to think of this problem is in terms of the notion of anaphora; however, unlike linguistic anaphoric inference in which antecedents for pronouns are selected from a linguistic context, in the interpretation of the textual part of multimodal messages the antecedents are selected from a graphical context. According to this, resolving multimodal references is like resolving anaphora accross modalities. Another view of the same problem consists in looking at pronouns in texts about drawings as deictic. In this second view, the context of interpretation of a natural language term is defined as a set of expressions of a graphical language with well defined syntax and semantics. Natural language and graphical terms are thought of as standing in a relation of translation similar to the translation relation that holds between natural languages. In this paper a theory based on this second view is presented. In this theory the relations between multimodal representation and spatial deixis, on the one hand, and multimodal reasoning and deictic inference, on the other, are discussed. An integrated model of anaphoric and deictic resolution in the context of the interpretation of multimodal discourse is also advanced. In Section 1 of this paper a multimodal representational system is presented. The system is motivated with the help of two simple examples. In Section 2 a formalization of the theory for the interpretation of simple maps is presented. The theory is developed along the lines of the theoretical framework provided by Montague's semiotic programme and its general theory of translation. In Section 3 an algorithm for finding out the translations of individual constants of both natural and graphical languages (i.e., for the resolution of deictic inferences) is introduced. In Section 4 how the theory can be extended to handle multimodal discourse is illustrated. In the concluding section some remarks about the adequacy of this theory are discussed.

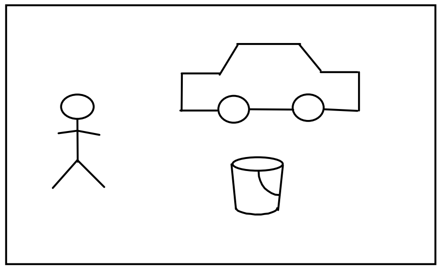
1. Reference, Spatial Deixis and Modality

In this paper a model for the resolution of multimodal references is presented. This is the problem of finding out the referent of a symbol in one modality in terms of information present either in the same or in other modalities. A model of this kind can be useful both for implementing intelligent multimodal tools (e.g., authoring tools to input natural language and graphics interactively for the automatic constructions of tutorials or manuals) and from the point of view of human-computer interaction (HCI) where it can help in the design of computer interfaces in which interpretation constraints of multimodal messages should be taken into account.

Consider Figure 1 (adapted from an example presented by Thomas Rist in the workshop on IMMPS at ECAI 96) in which a message is expressed through two different modalities, namely text and graphics. The figure illustrates a kind of reasoning required to understand multimodal presentations: in order to make sense of the message, the interpreter must realize what individuals are referred to by the pronouns *he* and *it* in the text. For the sake of argument, it is assumed that the graphical symbols in the figure are understood directly in terms of a graphical lexicon, in the same way that the words *he*, *it* and *washed* are understood in terms of the textual lexicon. It can easily be seen that given the graphical context *he* should resolve to the man, and *it* should resolve to the car. However, this inference is not valid since the information inferred to is not contained in the overt graphical context and the meaning of the words involved.

One way to look at this problem is as a case of anaphoric inference. Consider that the information provided by graphical means can be expressed also through the following piece of discourse: *There is a man, a car and a bucket. He washed it*. With Kamp's discourse representation theory (DRT) (Kamp 1981, Kamp et al., 1993) a discourse representation structure (DRS) in which the reference to the pronoun *he* is constrained to be the man can be built. However, the pronoun *it* has two possible antecedents, and for selecting the appropriate one, conceptual knowledge is required. In particular, the knowledge that a man can wash objects with water, and that water is carried on in buckets must be employed. If these concepts are included in the interpretation context like DRT conditions (which

[*] Dept. Computer Science, IIMAS, UNAM, México. E-mail: luis@leibniz.iimas.unam.mx.



“He washed it”

Figure 1
Instance of linguistic anaphor with pictorial antecedent.

should be retrieved from memory rather than from the normal flow of discourse), the anaphora can be solved. In terms of this analogy, situations like the one illustrated in Figure 1 have been labeled as problems of anaphors with pictorial antecedents in which the interpretation context is built not from a preceding text but from a graphical representation which is introduced with the text (André and Rist, 1994).

Consider now the converse situation shown in Figure 2 (adapted from Rist as above) in which a drawing is interpreted as a map thanks to the preceding text. The dots and lines of the drawing, and their properties, do not have an interpretation and the picture in itself is meaningless. However, given the context introduced by the text, and also considering the common sense knowledge that Paris is a city of France, and Frankfurt a city of Germany, and that Germany lies to the east of France (to the right), it is possible to infer that the denotations of the dots to the left, middle and right of the picture are Paris, Saarbrücken and Frankfurt, respectively, and that the dashed lines denote borders between countries, and in particular, the lower segment denotes the border between France and Germany. In this example, graphical symbols can be thought of as “variables” of the graphical representation or “graphical pronouns” that can be resolved in terms of the textual antecedent. Here again, the inference is not valid as the graphical symbols could be given other interpretations or none at all.

“Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt.”

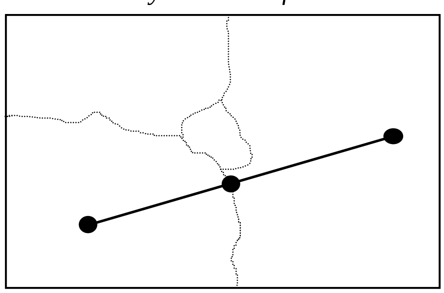


Figure 2
Instance of a pictorial anaphor with linguistic antecedent.

The situation in Figure 2 has been characterized as an instance of a pictorial anaphor with linguistic antecedent, and further related examples can be found in (André and Rist, 1994). This situation, however, cannot be modeled that easily in terms of Kamp’s DRT because the “pronouns” are not linguistic objects, and lacking a proper formalization of the graphical information, there is not a straightforward way to express in a discourse representation structure that a dot representing “a variable” in the graphical domain has the same denotation as a natural language name or description introduced from text in a DRS. Furthermore, consider that the situation in Figure 1 can be thought of as anaphoric only if we ignore the modality of the graphics, as was done above, but if the notion of modality is to be considered at all in the analysis, then the situation in Figure 1 poses the same kind of problems as the one in Figure 2. In general, graphical objects, functioning as constant terms or as variables, introduced as antecedents or as pronouns, cannot be expressed in a DRS, as the rules constructing these structures (the so-called DRS-construction rules) are triggered by specific syntactic configurations of the natural language in which the information is expressed. However, this limitation can be overcome if graphical information can be expressed in a language with well-defined syntax and semantics.

An alternative view on this kind of problems consists in looking at them in terms of the traditional linguistic notion of deixis (Lyons, 1968). This notion has to do with the orientational features of language which are relative to the spatio-temporal situation of a utterance. In this regard, and in connection with the notion of the so-called graphical anaphora discussed above, it is possible to mention the deictic category of demonstrative pronouns: words like *this* and *that* permit us to make reference to extralinguistic objects. In Figure 1, for instance, the pronouns *he* and *it* can be supported by overt pointing acts at the time the expression *he washed it* is uttered. Note that the purpose of the pointing act is to provide the referents for the pronouns directly, simplifying greatly the resolution process. However, the deictic use of a pronoun does not have to be supported by a physical ostension necessarily, as deictic use is characterized, more generally, by the identification of the referent in a metalinguistic context. Ambiguity of this kind of words is not unusual, as they also function as anaphoric if they are preceded by a linguistic context, and even as determiners with a deictic component (e.g., *this car*). Additionally, not only demonstratives and pronouns but also proper names, definite descriptions and even indefinites can be used deictically. As contextual factors involved in the interpretation of a deictic expression are conceivable of a great veriety of kinds, ostensive acts, although prominent, should be thought of as one particular kind of contextual factor. In summary, the denotation of a deictic term is the individual that is picked up by the human interpreter considering the interpretation context[2]. Consider that in the same way that an anaphoric inference is required for identifying the antecedent of an anaphoric term, an inference process is required for interpreting a term used deictically. We refer to such a process as a "deictic inference". The inference by which one gets to know that *he* and *it* are the man and the car is, accordingly, a deictic inference.

For our purpose it is important to investigate what is the relation between the notions of deixis and modality, on the one hand, and multimodal reasoning and inference, either deictic or anaphoric, on the other. According to Kamp (Kamp, 1981), the difference between deictic and anaphoric pronouns is that,

> ...deictic and anaphoric pronouns select their referents from certain sets of antecedently available entities. The two pronouns uses differ with regard to the nature of these sets. In the case of a deictic pronoun the set contains entities that belong to the real world, whereas the selection set for an anaphoric pronoun is made up of constituents of the representation that has been constructed in response to antecedent discourse.

Our concern here is how "the set of entities that belong to the real world" is accessible to the interpreter. In normal deictic spatial situations the referent of a deictic term is perceived directly through the visual modality, and as a result of such a visual interpretation process, the object is represented by the subject. The question is how the information can be expressed in this intermediate "visual" representation. A plausible answer is that there is a coding system and a medium associated to each particular modality. Our suggestion is that the notion of modality is a representational notion, and not a sensory one, as it is normally considered in psychological discussion. In our sense, a modality is a formal language, with a lexicon and well-defined syntactic and semantic structures, with an associated medium in which the expressions of the modality are written on. Multimodal reasoning consists in a process involving information expressed in the languages associated to different modalities, and this is achieved with the help of a translation relation that is similar to the relation of translation between natural languages. Performing a multimodal rasoning process is possible if the translation relation between expressions of different modalities is available. However, for particular multimodal reasoning tasks, the translation relation between individual constants of different modalities cannot be stated beforehand and has to be worked out dynamically through a deictic inferential process, as will be argued in the rest of this paper.

1.1 A Model for Multimodal Representation

This view of multimodal representation and reasoning can be formalized in terms of Montague's general semiotic programme (Dowty, 1985). Each modality in the system can be captured through a particular language, and relations between expressions of different modalities can be modeled in terms of translation functions from basic and composite expressions of the source modality into expressions of the target modality. In a system of this kind, interpreting examples in Figures 1 and 2 in relation to the linguistic modality consists in interpreting the information expressed through natural language directly when enough information is available, and completing the interpretation process by means of translating expressions of the graphical modality into the linguistic one, and vice versa.

[2] An operator called DTHAT for mapping deictic terms into their referents in an interpretation context is introduced in Kaplan's logic of demonstratives (Kaplan, 1978).

Consider Figure 3 —developing on (Pineda, 1989, 1998), (Klein et al., 1990) and (Santana et al., 1997)— in which a multimodal representational system for linguistic and graphical modalities is illustrated.

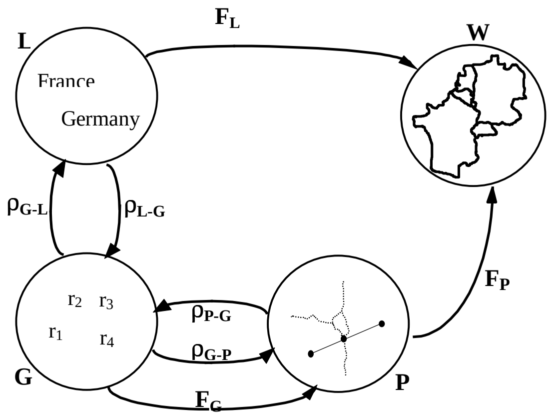


Figure 3
Multimodal representational system for linguistic and graphical modalities.

The circles labeled **L** and **G** in Figure 3 stand for sets of expressions of the natural and graphical languages respectively, and the circle labeled **P** stands for the set of graphical symbols constituting the graphical modality proper (i.e., the actual symbols on a piece of paper or on the screen). Note that two sets of expressions are considered for the graphical modality: the expressions in **G** belong to a formal language in which the geometry of pictures is represented and reasoned about, and **P** contains the overt graphical symbols which can be seen and drawn but cannot be manipulated directly. The functions $\rho_{\mathbf{L-G}}$ and $\rho_{\mathbf{G-L}}$ stand for the translation mappings between the languages **L** and **G**, and the functions $\rho_{\mathbf{P-G}}$ and $\rho_{\mathbf{G-P}}$ stand for the corresponding translation between **G** and **P**. The translation function $\rho_{\mathbf{P-G}}$ maps well-defined objects of the graphical modality into expressions of **G** where the interpretation process is performed. The translation $\rho_{\mathbf{G-P}}$, on the other hand, maps geometrical expressions of **G** into pictures; for every well-defined term of **G** of a graphical type (e.g., *dot*, *line*, etc.) there is a graphical object or a graphical composition that can be drawn or highlighted with the application of geometrical algorithms associated to operators of **G** in a systematic fashion. The circle labeled **W** stands for the world and together with the functions $\mathbf{F_L}$ and $\mathbf{F_P}$ constitutes a multimodal system of interpretation. The ordered pair <**W**, $\mathbf{F_L}$> defines the model $\mathbf{M_L}$ for the natural language, and the ordered pair <**W**, $\mathbf{F_P}$> defines the model $\mathbf{M_P}$ for the interpretation of drawings. The interpretation of expressions in **G** in relation to the world is defined either by the composition $\mathbf{F_L} \circ \rho_{\mathbf{G-L}}$ or, alternatively, by $\mathbf{F_P} \circ \rho_{\mathbf{G-P}}$. The denotation of the word *France* in **L**, for instance, is the same as the denotation of the corresponding region of the map of Europe which denotes France, the country, as both refer to the same individual. The denotation of the symbol r_1 in **G** that is related to the word *France* in **L** through $\rho_{\mathbf{G-L}}$, and to a particular region in **P** through $\rho_{\mathbf{G-P}}$, is also France, as translation is a meaning preserving relation between expressions. The interpretation functions $\mathbf{F_L}$ and $\mathbf{F_P}$ relate basic expressions, either graphical or linguistic, with the objects or relations of the world that these expressions happen to represent, and the definition of a semantic algebra for computing the denotation of composite graphical and linguistic expressions is required.

An important consideration for the scheme in Figure 3 is that the symbols of **P** have two roles: on the one hand they are representational objects (e.g., a region of the drawing represents a country), but on the other, they are also geometrical objects that can be talked about as geometrical entities. The geometrical region of the map representing France, for instance, is itself represented by the constant r_1 in **G**. In this second view geometrical entities are individual objects of the world of geometry, and as such they have a number of geometrical properties that are independent of whether we think of graphical symbols as objects in themselves or as symbols representing something else. The same duality can be stated from the point of view of the expressions of **G** as the set of individual geometrical objects (i.e., **P)** constitutes a domain of interpretation for the language **G**. This is to say that expressions of **G** have two interpretations: they represent geometrical objects, properties and relations directly, but they also represent the objects of the world (e.g., France, Germany, etc.) indirectly through the translation relation and interpretation of symbols in **P** taken as a language (i.e., the composition $\mathbf{F_P} \circ \rho_{\mathbf{G-P}}$). The ordered pair <**P**, $\mathbf{F_G}$> defines the model $\mathbf{M_G}$ for the geometrical interpretation of **G** as geometrical objects; the geometrical interpretation function $\mathbf{F_G}$ assigns a denotation for every constant of **G**; the denotation of individual constants of **G** are the graphical symbols themselves, and the denotation of operators and function symbols of **G** denoting graphical properties and relations will be given by predefined geometrical algorithms commonly used in computational geometry and computer graphics —see, for instance, (Shamos, 1978). The semantic interpretation of composite expressions of **G,** on the other hand, is defined through a semantic algebra, as will be shown below in Section 2.3.2. The definition of this geometrical interpreter will allow us to perform inferences about the geometry of the drawing in a very effective fashion. Consider that to state explicitly all true and false geometrical statements about a drawing would be a very cumbersome task, as the number of statements that would have to be made even for small drawings would be very large. Note as well that although a map can be an incomplete representation of the world (e.g., some cities might have been omitted), the geometrical algorithms associated to operators of **G** will always provide complete information of the map as a geometrical object.

1.2 Multimodal interpretation

For the kind of problems exemplified in Figures 1 and 2 the objects in **L**, **P** and **G** are given, and the function $\mathbf{F_L}$ establishes the relation between linguistic constants and the objects of the world that such constants happen to refer, and for interpreting these multimodal messages what is required is to make explicit $\mathbf{F_P}$. If one asks *who is he?* looking at Figure 1, for instance, the answer is found by computing $\rho_{\mathbf{G\text{-}P}}(\rho_{\mathbf{L\text{-}G}}(he))$ whose value is the picture of the man on the drawing. Once this computation is performed the picture can be highlighted or signaled by other graphical means. However, in other kinds of situations the knowledge of $\mathbf{F_P}$ might be available and the purpose of the interpretation process could be to identify $\mathbf{F_L}$. If one points out the middle dot in Figure 2 at the time the question *what is this?* is asked, the answer can be found by applying the function $\rho_{\mathbf{G\text{-}L}} \circ \rho_{\mathbf{P\text{-}G}}$ to the pointed dot (i.e., $\rho_{\mathbf{G\text{-}L}}(\rho_{\mathbf{P\text{-}G}}(\bullet))$), whose value would be the word *Saarbrücken*. A similar situation arises in the interpretation of multimodal referring expressions. Consider the following example —also from (André and Rist, 1994)— in which a multimodal message is constituted by a picture of an expresso machine that has two switches and by the textual expression *the temperature control.* In this scenario, the denotation of the natural language expression can be found by the human interpreter if the corresponding switch is identified in the picture through visual inspection (e.g., if the switch is highlighted). In general, multimodal coreference can be established if $\rho_{\mathbf{L\text{-}G}}$ and $\rho_{\mathbf{G\text{-}L}}$ are defined, as $\mathbf{F_P}$ can be made explicit in terms of $\mathbf{F_L}$ and vice versa.

In situations in which all theoretical elements illustrated in Figure 3 are given, questions about multimodal scenarios can be answered through the evaluation of expressions of a given modality in terms of the interpreters of the languages involved and the translation functions. However, when one is instructed to interpret a multimodal message, like Figures 1 and 2, not all information in the scheme of Figure 3 is available. In particular, the translation functions $\rho_{\mathbf{L\text{-}G}}$ and $\rho_{\mathbf{G\text{-}L}}$ of the graphical and linguistic individual constants mentioned in the texts and the pictures of the multimodal messages are not known, and the crucial inference of the interpretation process has as its goal to find out the definition of these functions, (i.e., to establish the relations between names of **L** and **G**). It is important to highlight that in order to find out $\rho_{\mathbf{L\text{-}G}}$ and $\rho_{\mathbf{G\text{-}L}}$ the information overtly provided in the multimodal message is usually not enough, and in order to carry out such an interpretation process it will be required to consider the grammatical structure of the languages involved, the definition of translations rules between languages, and also conceptual knowledge stored in memory about the interpretation domain.

An additional consideration regarding the scheme in Figure 3 is related to the problem of ambiguity in the interpretation of multimodal messages. In the same way that an ambiguous natural language expression of a source language can be translated into a number of expressions of a target language, an ambiguous natural language expression can be translated into several graphical expressions and vice versa. In the literature of intelligent multimodal systems ambiguity is commonly seen from the perspective of human-users. A multimodal referring expression constituted by the text *the temperature control* and a drawing with two switches is said to be ambiguous, for instance, if the human-user is not able to tell which one is the temperature control. A well-designed presentation should avoid this kind of ambiguities by suplementing additional information either in a textual form (e.g., the temperature control is the switch on the left) or by a graphical focusing technique, and an important motivation in the design of intelligent presentation systems like WIP (Wahlster, 1993) and COMET (Feiner and McKeown, 1993), is to generate graphical and linguistic explanations in which these kinds of ambiguities are avoided[3]. Note, however, that rather than a problem of ambiguity, such a kind of situation can be characterized as a problem of

[3] It is also worth noticing that systems like WIP and COMET do not interpret multimodal messages input by human users through the interaction and, therefore, there is no ambiguity to be resolved.

underspecification, as the expression *the temperature control* has only one syntactic structure and one meaning, and the referent can be identified in a given context if enough information is available.

The formalization of graphical representations through the definition of graphical languages with well-defined syntax and semantics allows us, on the other hand, to face the problem of ambiguity directly in terms of the relation of translation between the natural and graphical language. Consider that an ambiguous natural language expression in **L** can have several analyses, and the corresponding syntactic structures can be translated into **G** and could be rendered as different graphical configurations in **P**; similarly, a graphical configuration in **P** can have several syntactic analyses which can be translated into different expressions of **G**. Furthermore, a map can be represented by a number of expressions of **G** that refer to each individual graphical object in **P**, or by an expression of **G** that translates into a graphical composition, depending on whether the focus of the interpretation process is the whole of the drawing or one of its parts. The problem of whether a pointing act refers to a part or to the whole of a drawing has also been addressed in a number of intelligent multimodal systems like XTRA (Wahlster, 1991) and AlFresco (Stock, 1993), but the lack of a formalized notion of graphical language (and also a better understanding of indexical expressions), has prevented a deeper analysis of this kind of ambiguity. Although problems of ambiguity of graphics are not the focus of this paper, and despite that very little can be said about genuine graphical ambiguities in which a picture taken as a whole can receive different interpretations, like Wittgenstein's duck-rabbit picture (Wittgenstein, 1953), we believe that the scheme in Figure 3 offers a point of departure to address this kind of investigation.

1.3 Multimodal Generation

An important motivation for the study of the interpretation of multimodal messages is the definition of multimodal presentation or explanation systems in which users are able to identify the referent of graphical and linguistic expressions easily. In WIP, for instance, a central concern is whether the human-user is able to "activate" the relevant "representations" (presumably in his or her mind) and resolve the graphical and linguistic ambiguities and anaphoras (using WIP's terminology) present in multimodal messages. This is possible, in general, if the message conveys to the human-user explicit interpretation paths from the information that is available overtly to the information that the user is expected to infer. The production of multimodal referring expressions in this kind of systems depends on the use of presentation strategies defined in terms of rhetorical structures and intentional goals —e.g., along the lines of Rhetorical Structure Theory (RST) (Mann and Thompson, 1988), and its computational implementation (Moore, 1992)— and the use of a particular presentation strategy in a multimodal explanation (e.g., in WIP) depends crucially on whether the expressions generated on the basis of such a strategy satisfies the conditions defined to activate the expected representations in the user's mind (an intentional goal). Furthermore, some rhetorical structures are designed explicitly to provide additional information to activate the expected representations if the conditions for the identification of the referent of an expression are not met. Consider again the resolution of the so-called ambiguity in the interpretation of *the temperature control* example in WIP in which the presentation strategy provides the information required by the human user to indentify the referent, either through the text *the temperature control is the switch on the left* or highlighting or pointing the corresponding switch in the drawing. WIP is able to tell whether the presentation would be ambiguous for the human-user if additional information were not provided because it has a representation of the actual situation and a simple model of the user's beliefs.

Although the main representation structure of multimodal presentation and explanation systems is defined at a rhetorical level, the use of presentation strategies relies on algorithms for the generation of graphical and linguistic referring expressions. For instance, the "activate" presentation strategy of WIP (André and Rist, 1994), which has the purpose to establish a mutual belief between the human-user and the system about the identity of an object, employs an algorithm for the generation of referring expressions based on an incremental interpretation algorithm proposed by Reiter and Dale (Reiter and Dale, 1992). It is interesting to note that presentations generated by WIP and other multimodal explanation systems like COMET (Feiner and McKeown, 1993), or TEXPLAN (Maybury, 1993), are limited to the production of definite descriptions only, despite that the use of indefinite descriptions can be natural in multimodal communication. However, this restriction can be overcome with a more solid representational framework as the one illustrated in Figure 3. Consider that basic or composite expressions of the languages **G** and **L** can be translated to basic or composite expressions of the other language, depending on the definition of the translation function. So, if one needs to refer linguistically to a graphical configuration, for instance, it would be only required to find an expression of **G** which expresses the relevant graphical properties of the desired object in the most simple fashion, and then translate it to its corresponding expression in **L**. The resulting natural language expression could be used directly or embedded in a larger natural language expression containing words that refer to abstract objects or properties. The descriptions obtained through this strategy employ explicitly the concrete and graphical properties of the representation, as expressions of **G** are constituted by constants and operators that directly describe the geometry of objects and configurations.

Consider the natural language text *Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt*. This sentence contains the definite description *the intersection between the border between France and Germany and a line from Paris to Frankfurt*, which in turns contains *the border between France and Germany* and *a line from Paris to Frankfurt*. Finding the graphical referents of these expressions requires the identification of a dot, a curve and a line in the map (i.e., the corresponding graphical objects). These graphical objects can be referred to directly through language; however, there are additional graphical entities in the map in Figure 2 that have an interpretation but are not mentioned explicitly in the text of the multimodal message. In Figure 4, for instance, Belgium is represented by region r_4, and the curve c_6 represents the border between France and Belgium. Once a picture has been interpreted one would be entitled to ask not only for graphical objects that have been mentioned in the textual part of the message, but also for any meaningful graphical object. So, if one points to the curve c_6 in Figure 2 at the time the question *What is this?* is asked, the answer could be *the border between France and Belgium*, or alternatively, the indefinite *a border*. As some graphical objects named by constants of the graphical language do not have a proper natural language name, the translation function $\rho_{\mathbf{G\text{-}L}}$ must associate a basic constant of **G** with a composite expression of **L**. The process of inducing such a translation function is closely related to the process of generating the corresponding natural language descriptions, and this relation will be explored further in Section 3.

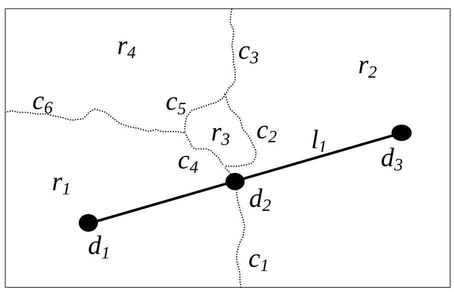


Figure 4
Labeling the graphical objects in Figure 2.

In the rest of this paper, some results on how the scheme for multimodal representation and interpretation in Figure 3 can be carried out are presented. In Section 2, a formalization of the languages **L, P** and **G** with their corresponding translation functions, along the lines of Montague's general semiotic programme, is presented. In this section the process of multimodal interpretation is explained, and the translation of expressions of one modality in terms of the other is illustrated. However, such a process can be carried out only if the translation functions are known and, as was mentioned above, in the interpretation of multimodal messages that is not normally the case. In Section 3, how such functions can be induced in terms of the message, constraints on the interpretation conventions of the modalities, and constraints on the general knowledge about the domain, is presented. In this section the process of generating graphical and linguistic descriptions, which is associated with the induction of the translation functions, is also shown. In Section 4 a discussion on how to extend Kamp's DRS with multimodal structures is illustrated. Finally, in Section 5 some concluding remarks and some lines for further work are presented.

2. A Multimodal System of Representation

In this section the definition of the syntax and semantics of the languages **L**, **P** and **G** to express the multimodal message of Figure 2 is presented. The language **L** is designed to produce expressions useful to refer to objects, properties and relations commonly found in discourse about maps. In particular, the natural language expression of Figure 2 can be constructed in a compositional fashion. The syntactic structure of **P**, on the other hand, imposes a restriction on the possible geometries of the family of drawings in the interpretation domain. The language **G** is a logical language in which interpretation and reasoning about geometrical configurations can be carried out. It is an interlingua representation for information expressed in both of the modalities.

The definitions of **L**, **P** and **G** follow closely the general guidelines of Montague's semiotic programme. As a first step of the syntactic definition of a language the set of categories or types is stated. A number of constants —or basic expressions— for each type is defined, and the combination rules for producing composite expressions are stated. For each type of a source language a corresponding type in the target language is assigned. Basic expressions of the source language can be mapped either to basic or to composite expressions of the corresponding type in the target language and vice versa. Associated to each syntactic rule of a source language, a translation rule for mapping the expression formed by the rule into its translation in the target language is defined. In Section 2.1 the definition of the language **L** is presented. In Section 2.2 the syntactic definition and the semantic interpretation of **P** are stated. The syntactic definition of **G** and the specification of its geometrical interpreter are given in Section 2.3. In this section some examples to illustrate the process of geometrical interpretation are also shown. The translations between **L** and **G** formally specified in Section 2.4 and, finally, the translations between **G** and **P** are described in Section 2.5.

2.1 Definition of the Language L

The language **L** is designed to produce expressions like *Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt* in a compositional fashion.This means that all constants like *France* and *Germany*, and also all subexpressions of the former sentence, like *the border between France and Germany* or *a line from Paris to Frankfurt* can also be produced. In addition, language **L** can produce expressions like *France is a country*, *Frankfurt is a city of Germany* or *Germany is to the east of France* which express common sense knowledge required in the interpretation of maps. Next, the syntax and semantics of **L** are presented in Sections 2.1.1 and 2.1.2, respectively.

2.1.1 Syntactic definition of L

The set of syntactic categories of **L** is as follows:

1. The basic syntactic categories of **L** are *t*, *IV*, *ADJ*, *CN* and *CN'* where *t* is the category of sentences, *IV* is the category of intransitive verbs, *ADJ* is the category of adjectives, and *CN* and *CN'* are two categories of common nouns.
2. If *A* and *B* are syntactic categories then *A/B* is a category.[4]

Traditional syntactic categories of natural language like transitive verbs (*TV*), terms (*T*), prepositional phrases (*PP*) and determiners (*T/CN*) can be derived from the basic categories.

The table in Figure 5 illustrates the constants of **L** with their category names and category definition. Common nouns are divided into *CN* and *CN'*. The reason for this distinction, as will be seen in Section 2.4.1, is that expressions of category *CN* translate into graphical predicates (sets of graphical objects) while expressions of category *CN'* translate into abstract concepts, as expressions of this category do not have a concrete graphical representation. For instance, *city* translates into a set of dots representing cities, but *east* translates into a geometrical function from regions to zones (e.g., if the region representing France is the argument of this function, the zone to the right of that region is the function value). Prepositional phrases are divided into *PP* and *PP'* due to the clasification of common nouns into *CN* and *CN'*. There are no basic constant of categories *PP*, *PP'* and *IV*, as prepositional words are introduced syncategorematically and intransitive verb phrases are always composite expressions in this grammar. Transitive verbs are defined in a standard fashion, and the constant *be* of category *IV/ADJ* is used to form attributive sentences.

[4] An expression of category *A/B* combines with an expression of category *B* to give an expression of category *A*.

Constant	Category name	Category definition
Paris, Frankfurt, Saarbrücken, France, Germany	*T*	*t/IV*
city, country, border, line, intersection	*CN*	*CN*
east	*CN'*	*CN'*
big	*ADJ*	*ADJ*
be, lie at, be to	*TV*	*IV/(t/IV)*
be	*IV/ADJ*	*IV/ADJ*
a, the	*T/CN*	*(t/IV)/CN*
	PP	*CN/CN*
	PP'	*CN/CN'*
	IV	*IV*

Figure 5
Constants of language **L**.

Next, the syntactic rules of **L** are presented. Each rule is shown in a box containing the purpose of the rule, the syntactic rule itself and some examples of expressions that can be formed with the rule. Following Montague, syntactic rules and syntactic operations for combining symbols (for instance, F_{L1}) associated to each rule are separated. In the following, P_C is the set of expressions of category C.

SENTENCES

$S1_L$. If $\alpha \in P_T$ and $\beta \in P_{IV}$, then $F_{L1}(\alpha,\beta) \in P_t$, where $F_{L1}(\alpha,\beta) = \alpha\ \beta^*$, and β^* is the result of replacing the first *verb* in β by its third person singular present form.

Examples:
-*Paris is a city of France*
-*Germany is to the east of France*
-*a country is big*
-*Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt*

TRANSITIVE VERB PHRASES

$S2_L$. If $\alpha \in P_{TV}$ and $\beta \in P_T$, then $F_{L2}(\alpha,\beta) \in P_{IV}$, where $F_{L2}(\alpha,\beta) = \alpha\ \beta$.

Examples:
-*be a city*
-*be to the east of France*

ATTRIBUTIVE VERB PHRASES

$S3_L$. If $\alpha \in P_{IV/ADJ}$ and $\beta \in P_{ADJ}$, then $F_{L2}(\alpha,\beta) \in P_{IV}$.

Examples:
-*be big*

TERMS

$S4_L$. If $\alpha \in P_{T/CN}$ and $\beta \in P_{CN}$ or $P_{CN'}$, then $F_{L3}(\alpha,\beta) \in P_T$, where $F_{L3}(\alpha,\beta) = \alpha^*\ \beta$, and α^* is α except in the case where α is *a* and the first word in β begins with a vowel; here, α^* is *an*.

Examples:
-*a city*
-*a city of France*
-*the border between France and Germany*
-*a line from Paris to Frankfurt*
-*the east of France*

COMMON NOUNS

S5$_L$. If $\alpha\in P_{CN}$ and $\beta\in P_{PP}$, or $\alpha\in P_{CN'}$ and $\beta\in P_{PP'}$, then $F_{L2}(\alpha,\beta) \in P_{CN}$.

Examples: -*city of France*
-*east of France*
-*border between France and Germany*
-*intersection between the border between France and Germany and a line from Paris to Frankfurt*

of PREPOSITIONAL PHRASES[5]

S6$_L$. If $\alpha\in P_T$, then $F_{L4}(\alpha) \in P_{PP}$ or $P_{PP'}$, where $F_{L4}(\alpha)$ = *of* α.

Examples: -*of France*
-*of Germany*
-*of a country*

between PREPOSITIONAL PHRASES

S7$_L$. If $\alpha,\beta\in P_T$, then $F_{L5}(\alpha,\beta) \in P_{PP}$, where $F_{L5}(\alpha,\beta)$ = *between* α *and* β.

Examples: -*between France and Germany*
-*between France and a country*
-*between the border between France and Germany and a line from Paris to Frankfurt*

from-to PREPOSITIONAL PHRASES

S8$_L$. If $\alpha,\beta\in P_T$, then $F_{L6}(\alpha,\beta) \in P_{PP}$, where $F_{L6}(\alpha,\beta)$ = *from* α *to* β.

Example: -*from Paris to Frankfurt*

2.1.2 Semantic definition of L

The semantics of **L** is given in a model-theoretic fashion as follows. The interpretation domain is the world **W** = {Paris, Saarbrücken, Frankfurt, France, Germany, the border between France and Germany, ...}. Let D_x be the set of possible denotations for expressions of type *x*, and for any types *A* and *B*, $D_{A/B} = D_A^{D_B}$ (i.e., the set of all functions from D_B to D_A). Let $\mathbf{F_L}$ be an interpretation function that assigns to each constant of type *A* a member of D_A. For the example in Figure 2 $\mathbf{F_L}$ is defined as shown in Figure 6.

Constant α	$\mathbf{F_L}(\alpha)$
Paris, Frankfurt, Saarbrücken, France, Germany	Paris, Frankfurt, Saarbrücken, France, Germany
city	{Paris, Frankfurt, Saarbrücken, ...}
country	{France, Germany, ...}
border	{border between France and Germany, ...}
line	{line from Paris to Frankfurt, ...}
intersection	{intersection between the border between France and Germany and a line from Paris to Frankfurt, ...}
east	
be, lie at, be to	
a, the	

Figure 6
Interpretation of constants of language **L**.

As can be seen, not every constant of **L** has an interpretation assigned by $\mathbf{F_L}$; in particular, words like *east*, *be*, *lie at* and *be to* have no interpretation defined directly in **L**. In principle the definition of these constants could be stated

[5] Although *of*, *between* and *from* have been introduced syncategorematically in **L** for simplicity, they could have been defined as constants of some category of **L**, and their translations into **G** would have been a composite expression of some graphical type.

as an object of the appropriate semantic type but this is not a straightforward enterprise. Consider, for instance, that the constant *east* of category *CN'* is a basic object (a kind of predicate), but the individual objects in its extension are not overtly defined in the interpretation domain. Furthermore, it is more natural to talk about the interpretation of composite predicates, like *east of France*, in which *east* is a part of. However, even the interpretation of this latter kind of objects is problematic as they have a vague spatial meaning. For these reasons, the interpretation of these constants is not defined explicitly as a part of the function $\mathbf{F_L}$, but in terms of its translation into **G** where a spatial meaning can be formally defined, as will be shown below. A similar strategy is used for the interpretation for spatial prepositions; although *of*, *between*, and *from-to,* were introduced syncategorematically in the syntax of **L,** they could have been defined as objects of an appropriate category and their semantics could have been given explicitly through $\mathbf{F_L}$ or, alternatively, in terms of the translation into Intensional Logic along the lines of PTQ. However, the semantic type of this kind of objects has an extraordinary degree of complexity, and the actual definition of these constants is seldom seen in the literature[6]. In our system the interpretation of spatial prepositions will also be given in terms of the translation into **G** and the interpretation of **P**. Note as well that no interpretation has been defined for the determiners *a* and *the*. One strategy for assigning a denotation would be to translate these constants into Intensional Logic, but this would be required only if a larger fragment of English in which reference to space were not the focus of study. In our approach the determiners will be interpreted in terms of their translations into **G** in which high-order functions can be expressed.

In summary, the semantics of some constants and all composite expressions of **L** will be given in terms of their translations into **G** and **P**. Note that according to the scheme in Figure 3 if the translations between **L** and **G,** and **G** and **P** are defined, and the semantic interpretation of **P** is overtly defined, the interpretation of the natural language expressions can be found. Although the semantics of **L** is not further discussed in this paper we consider that the interpretation of linguistic expressions referring to spatial situations could be embedded in a larger fragment of English, and a full semantic interpretation by translating English into Intensional Logic would have to be given. In such a model the semantic value of spatial prepositions would be left undefined, expressions referring to spatial configurations would be translated into **G**, and the interpretation of expressions of **G** would be embedded within the interpretation of Intensional Logic.

2.2 Definition of the Language P

In this section, the syntax and semantics of the language **P** are formally defined. The purpose of these definitions is to characterize the family of drawings that can be interpreted as maps, and to discriminate these drawings from other kinds of graphical configurations constituted by dots, curves and regions. This notion of multimodal system of representation in which objects in the graphical modality are formalized through a well-defined language is similar to the notion of graphical language introduced by Mackinlay for the automatic design of graphical presentations (Mackinlay, 1987), where a number of graphical languages (e.g., the languages of bar charts, area and positions graphs, scatter plots, etc.) are formally specified. In Mackinlay's work expression of graphical languages are related to the objects of the world that they represent through an *encodes* relation with three arguments: the graphical constant or expression performing the representation, the object of the world that is represented through the graphical expression, and the graphical language to which the graphical expression belongs to[7]. The formalization of **P** permits to define a precise statement of expressivenes of a graphical language, as follows: "... a set of facts is expressible in a language (graphical) if the language containts a sentence that encodes every fact in the set and does

[6] In PTQ prepositions —of category (*IV*/*IV*)/*T*)— are treated semantically as functions which apply to set of properties to give functions from properties to properties, but no explicit example of the actual semantic value of any of these constants is provided. In our system it will be possible to compute the semantic value of spatial prepositional phrases in an effective manner, yet the approach is fully compatible with Intensional Logic.

[7] Incidentally, a similar encoding relation *Encodes* is used in the WIP system to relate the representational object with the object that it represents, but the third argument of this relation in WIP is a so-called context space that allows to use the same presentation in different perspectives (e.g., an expresso machine may refer to an individual machine in a context space, or alternatively it can be seen as the prototype of expresso machines in a different context space). The encodes relation in WIP and in Mackinlay's is similar to the translation relation between objects of **P** (or **G**) and **L** in our system, and we can thought of a graphical language as a language encoding the information that is intended to communicate. However, it is interesting to note that the status of the "linguistic" argument of the encodes relation is different in WIP and in Mackinley's system. In the former, it is an "internal representation" —a psychological notion— while in the latter it stands for an object or a relation in the world itself —a semantic notion. In our approach, on the other hand, there are no "internal representations" and the translation relates graphical and linguistic expressions which are both "external" and both refer to the world through a well-defined semantics.

not encodes any additional facts” (Mackinlay, 1987, pp.54). The formalization allows, additionally, to make empirical studies to determine how effectively a human-user can interpret expressions of a particular graphical language in relation to another in which the same set of facts is encoded. Although all graphical languages studied by Mackinlay are conventional and have a precise geometrical characterization, the notions of expressiveness and effectiveness of graphical languages can be applied to more unruly graphical domains (e.g., maps are analogical representations with a diagrammatic conventional component) as long as a formalization for the family of drawings can be approximated. Here, the question of wherther arbitrary families of graphical objects can be formalized through a well-defined syntax is left open, and although it is possible to think of many families of drawings with very arbitrary geometries, some important efforts have been made in the characterization of design and other kinds of objects —see, for instance, shape grammars (Stiny, 1975). Another related issue that is relevant for the construction of multimodal interactive systems is whether it is possible and useful to input expressions of **P** directly, and to obtain their syntactic structure through graphical parsing techniques (Wittenburg, 1997). In summary, the purpose of formalizing **P** is to be able to talk about maps as a modality as, in our sense, a modality is a code system for the symbols expressed in a medium, and a multimodal system of representation relates information expressed through different code systems in a systematic fashion.

2.2.1 Syntactic definition of P

The types of **P** are *dot*, *line*, *curve*, *region*, *zone*, *composite_region*, *dot_set*, *line_set*, *map*. Let C_s be the set of constants of type *s*, and E_s the set of well-formed expressions of graphical type *s*. Although the constants of **P** are the actual graphical marks on the screen or a piece of paper, a number of labels for facilitating the presentation are illustrated Figure 7.

Constant	Type
$d_1, d_2, d_3, ...$	*dot*
$l_1, l_2, l_3, ...$	*line*
$c_1, c_2, c_3, ...$	*curve*
$r_1, r_2, r_3, ...$	*region*
$z_1, z_2, z_3, ...$	*zone*
$cr_1, cr_2, cr_3, ...$	*composite_region*
$\emptyset, ds_1, ds_2, ...$	*dot_set*
$\emptyset, ls_1, ls_2, ...$	*line_set*
$m_1, m_2, m_3, ...$	*map*

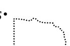
Figure 7
Constants of language **P**.

For the syntactic definition of **P** we capitalize the distinction introduced by Montague between syntactic rules and syntactic operations. This distinction is based on the observation that “... syntactic rules can be thought of as comprising two parts: one which specifies under what conditions the rule is to be applied, and the other which specifies what operation to perform under those conditions” (Dowty et al., 1995, pp. 254). While a syntactic rule comprises both parts and defines the syntactic structure of an expression, the syntactic operation is a rule that depends on —or at least takes into account— the shape of the symbols and the medium in which the symbols are substantially realized. For instance, the syntactic operation F_{L5} in the rule $S7_L$ (i.e., $F_{L5}(\alpha,\beta)$ = *between* α *and* β) combines the symbols *between* and *and* with the arguments to conform the linear string indicated by the operation. For the definition of syntactic operations of **P** we generalize the operations that manipulate strings of symbols into general geometrical operations on the shapes of the graphical symbols on the paper or the screen, and these manipulations are defined according to certain geometrical conditions.

The definition of well-formed expressions of **P** is as follows:

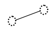
CONSTANT

$S1_P$. If $\alpha \in C_s$ then $\alpha \in E_s$.

Examples: •, /, 

LINE

S2$_P$. If $\alpha,\beta \in E_{dot}$ then $F_{P1}(\alpha,\beta) \in E_{line}$ where $F_{P1}(\alpha,\beta)$ is a line from α to β.

Example:  (the resulting graphical expression is only the line)

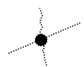
CURVE

S3$_P$. If $\alpha,\beta \in E_{region}$ such that α and β are adjacent then $F_{P2}(\alpha,\beta) \in E_{curve}$ where $F_{P2}(\alpha,\beta)$ is the curve between α and β.

Example:  (the resulting graphical expression is only the curve)

INTERSECTION

S4$_P$. If $\alpha \in E_{curve}$ and $\beta \in E_{line}$ then $F_{P3}(\alpha,\beta) \in E_{dot}$ where $F_{P3}(\alpha,\beta)$ is the dot in the intersection between α and β.

Example:  (the resulting graphical expression is only the dot)

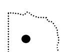
RIGHT

S5$_P$. If $\alpha \in E_{region}$ then $F_{P4}(\alpha) \in E_{zone}$ where $F_{P4}(\alpha)$ is the zone to the right of the region α (the interpretation of "right" will be given below in the semantics of language **G**).

Example:  (the resulting graphical expression is only the gray zone)

DOT INSIDE A REGION

S6$_P$. If $\alpha \in E_{region}$ then $F_{P5}(\alpha) \in E_{dot}$ where $F_{P5}(\alpha)$ is the drawing of a dot inside α.

Example:  (the resulting graphical expression is only the dot)

COMPOSITE REGION (1)[8]

S7$_P$. If $\alpha,\beta \in C_{region}$ such that α and β are adjacent then $F_{P6}(\alpha,\beta) \in E_{composite\_region}$ where $F_{P6}(\alpha,\beta)$ is the drawing of α and β.

COMPOSITE REGION (2)

S8$_P$. If $\alpha \in C_{region}$ and $\beta \in E_{composite\_region}$ such that α and β are adjacent then $F_{P6}(\alpha,\beta) \in E_{composite\_region}$.

SET OF DOTS

S9$_P$. If $\alpha \in E_{dot\_set}$ and $\beta \in C_{dot}$ then $F_{P6}(\alpha,\beta) \in E_{dot\_set}$.

SET OF LINES

S10$_P$. If $\alpha \in E_{line\_set}$ and $\beta \in C_{line}$ then $F_{P6}(\alpha,\beta) \in E_{line\_set}$.

MAP

S11$_P$. If $\alpha \in E_{composite\_region}$, $\beta \in E_{dot\_set}$ and $\delta \in E_{line\_set}$ then $F_{P7}(\alpha,\beta,\delta) \in E_{map}$ where $F_{P7}(\alpha,\beta,\delta)$ is the drawing of α, β and δ.

With the help of this grammar it is possible to draw maps as the one illustrated in Figure 2. Note that the basic object in this particular graphical construction is the region. The idea is to construct successfully a map out its

[8] Examples for the rules S7$_P$ to S11$_P$ are included in the construction of the map in Figure 8, as explained below.

constituting regions (i.e., like in a jigsaw puzzle) until the full map is produced. Once the map is constructed other kind of objects with conventional meanings, like dots and lines, can be drawn upon the assembly of regions. As an illustration, consider Figure 8 in which the syntactic structure of the map in Figure 4 is shown. Note that the decision to use regions as basic objects in the graphical composition is not mandatory, and alternative constructions are possible; for instance, we could have designated curves as basic objects and obtain regions as compositions made out of curves. As this decision is conventional, the set of graphical symbols included in a graphical syntactic tree of a map will be designated as "the base". For instance, the base of the map in Figure 8 is the set $\{d_1, d_2, d_3, l_1, r_1, r_2, r_3, r_4\}$. The base is just the set of graphical objects that are taken as "the atoms" of the graphical composition in each particular interpretation task, and different graphical grammars would select different types of graphical objects for the base.

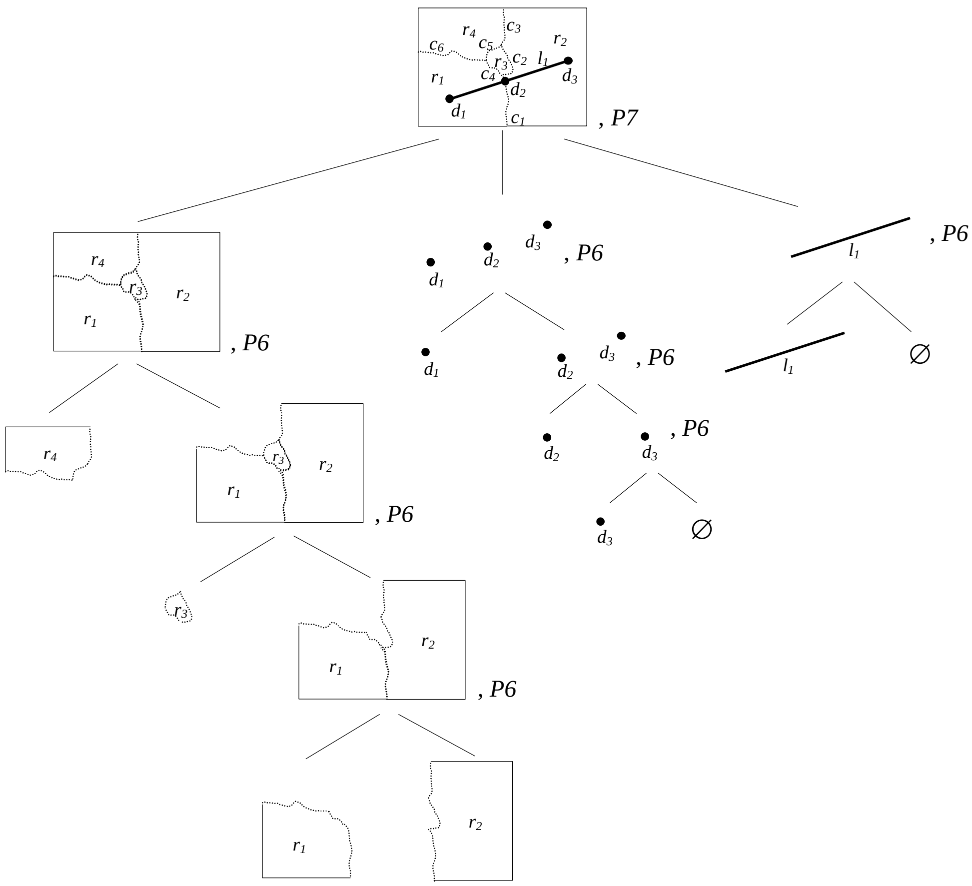


Figure 8
Construction of a map.

The purpose of this grammar is illustrative and as we have discussed no intuitive criterion to be able to say what counts as a map, neither we have made an empirical investigation to characterize families of maps, **P** imposes very few constraints on graphical expressions, and many configurations that can be produced with these rules might not count as maps; in addition, **P** is not expressive enough to characterize a large number of objects that would be normally interpreted as maps. Another consideration is that graphical objects can be used as basic building blocks of the construction, or as objects produced by graphical compositions (which we call "emergent objects"); for instance, in the grammar of **P** regions are basic objects but curves are produced by graphical compositions. Additionally, in some contexts the interpretation of the graphical expression as a whole may be required but in others only the interpretation of some of the parts may be relevant; for instance, although curves are not a part of the syntactic tree in Figure 8 they can be generated and translated into **G** when required through rules $S3_P$ and $T3_{P\text{-}G}$ as long as the composition is made out of regions included in the base of the map. Had the grammar allowed the generation of composite regions out of regions of the base, these emergent objects could also be used for the generation of curves. Another consideration is that expressions of type *map* are in general ambiguous as they have several syntactic analyses, but as this feature is harmless for the current discussion we do not pursue further this issue. A final remark is that alternative grammars could be defined for characterizing the same kind of drawings with different consequences in the syntax and the semantics. One possibilitity, for instance, is to define a syntactic operation that takes two adjacent regions and produces the union of the regions as one single emerging region, instead of the set of the two regions as it is currently defined. Such a rule would be similiar to the rule that combines two regions to produce a curve, and it would be useful in applications like XTRA (Wahlster, 1991) in which the ambiguity of pointing to a part or the whole is intended to be resolved.

2.2.2 Semantic definition of P

The semantics of **P** is given in a model-theoretic fashion as follows. Let $\mathbf{W} = A_{city} \cup A_{line} \cup A_{border} \cup A_{country} \cup A_{zone}$ be the world. Let D_x be the set of possible denotations for expressions of type *x*, such that $D_{dot}=A_{city}$, $D_{line}=A_{line}$, $D_{curve}=A_{border}$, $D_{region}=A_{country}$, $D_{zone}=A_{zone}$, and, for any types *a* and *b*, $D_{<a,b>} = D_b^{D_a}$ (i.e., the set of all functions from D_a to D_b). Let $\mathbf{F_P}$ be an interpretation function that assigns to each constant of type *a* a member of D_a. The interpretations of the constants are presented in Figure 9.

Constant α	$\mathbf{F_P(\alpha)}$
$d_1, d_2, d_3, \ldots$	Paris, Saarbrücken, Frankfurt, ...
$l_1, l_2, l_3, \ldots$	line from Paris to Frankfiurt, ...
$c_1, c_2, c_3, \ldots$	border between France and Germany, ...
$r_1, r_2, r_3, \ldots$	France, Germany, ...
$z_1, z_2, z_3, \ldots$	east of France, east of Germany, ...
$cr_1, cr_2, cr_3, \ldots$	region formed by France and Germany, ...
$\emptyset, ds_1, ds_2, \ldots$	sets of cities
$\emptyset, ls_1, ls_2, \ldots$	set of lines
$m_1, m_2, m_3, \ldots$	maps

Figure 9
Semantics of constants of **P**.

Following Montague, we adopt the notational convention by which the semantic value or denotation of an expression α with respect to a model *M* is expressed as $[[\alpha]]^M$. The semantic rules for interpreting the language **L** are the following:

CONSTANT
M1$_P$. If $\alpha \in C_s$ then $[[\alpha]]^M = \mathbf{F_P}(\alpha)$.

LINE
M2$_P$. If $\alpha,\beta \in E_{dot}$ then $[[F_{P1}(\alpha,\beta)]]^M$ = is a line from $[[\alpha]]^M$ to $[[\beta]]^M$.

CURVE
M3$_P$. If $\alpha,\beta \in E_{region}$ such that α and β are adjacent then $[[F_{P2}(\alpha,\beta)]]^M$ is the border between $[[\alpha]]^M$ and $[[\beta]]^M$.

INTERSECTION
M4$_P$. If $\alpha \in E_{curve}$ and $\beta \in E_{line}$ then $[[F_{P3}(\alpha,\beta)]]^M$ is the intersection between $[[\alpha]]^M$ and $[[\beta]]^M$.

RIGHT
M5$_P$. If $\alpha \in E_{region}$ then $[[F_{P4}(\alpha)]]^M$ is the east of $[[\alpha]]^M$.

DOT INSIDE A REGION
M6$_P$. If $\alpha \in E_{region}$ then $[[F_{P5}(\alpha)]]^M$ is a city of $[[\alpha]]^M$.

COMPOSITE REGION (1)
M7P. If $\alpha,\beta \in C_{region}$ such that α and β are adjacent then $[[F_{P5}(\alpha,\beta)]]^M$ is the union of $\{[[\alpha]]^M\}$ and $\{[[\beta]]^M\}$.

COMPOSITE REGION (2)
M8P. If $\alpha \in C_{region}$ and $\beta \in E_{composite\_region}$ such that α and β are adjacent then $[[F_{P5}(\alpha,\beta)]]^M$ is the union of the sets $\{[[\alpha]]^M\}$ and $[[\beta]]^M$.

SET OF DOTS
M9P. If $\alpha \in E_{dot\_set}$ and $\beta \in C_{dot}$ then $[[F_{P5}(\alpha,\beta)]]^M$ is the union of the sets $[[\alpha]]^M$ and $\{[[\beta]]^M\}$.

SET OF LINES
M10P. If $\alpha \in E_{line\_set}$ and $\beta \in C_{line}$ then $[[F_{P5}(\alpha,\beta)]]^M$ is the union of the sets $[[\alpha]]^M$ and $\{[[\beta]]^M\}$.

MAP
M11P. If $\alpha \in E_{composite\_region}$, $\beta \in E_{dot\_set}$ and $\delta \in E_{line\_set}$ then $[[F_{P6}(\alpha,\beta,\delta)]]^M$ is the union of the sets $[[\alpha]]^M$, $[[\beta]]^M$ and $[[\delta]]^M$.

2.3 Definition of the Language G

In this section the syntax and semantics of the graphical language **G** are formally stated. **G** is defined along the lines of Intensional Logic, and it is expressive enough to refer to graphical symbols and configurations, on the one hand, and to express the translation of quantified expressions of **L**, on the other. As was mentioned, the interpretation of expressions of **G** is defined in relation not to the world **W** but to a domain constituted by the graphical objects in **P**, and for this reason, we refer to the interpreter of **G** as a geometrical interpreter, and to the process of interpreting expressions of **G** as a geometrical interpretation process.

2.3.1 Syntactic definition of G

The types of the language **G** are as follows[9]:

(1) *e* is a type (graphical objects).
(2) *t* is a type (truth values).
(3) If *a* and *b* are any types, then $<a, b>$ is a type[10].
(4) Nothing else is a type.

Let V_s be the set of variables of type *s*, C_s the set of constants of type *s*, and E_s the set of well-formed expressions of graphical type *s*. The constants of **G** are presented in Figure 10. Note that constants like *right*, *curve_between*, etc. have an associated $right_*$, $curve\_between_*$, etc. The unsubscripted version of these constants denotes a relation between sets of properties of graphical individuals and the subscripted version denotes the corresponding geometrical relation between individuals; the type-raised version is used for preserving quantification properties in the translation process from **L** into **G**, while the subscripted version is used for computing the geometry associated to the corresponding relation, as will be shown below in Section 2.3.2.

G is a formal language with constants and variables for all types, functional abstraction and application, and existencial and universal quantification. The syntactic rules of **G** are as follows:

1. If $\alpha \in C_s$, then $\alpha \in E_s$.
2. If $\mu \in V_s$, then $\mu \in E_s$.
3. If $\alpha \in E_{<a,\, b>}$ and $\beta \in E_a$, then $\alpha(\beta) \in E_b$.
4. If $\alpha \in E_a$ and $u \in V_b$, then $\lambda u[\alpha] \in E_{<b,\, a>}$.
5. If $\mu \in V_s$ and $\beta \in E_t$ then $\exists\mu(\beta) \in E_t$.

---

[9] A simplifying assumption rests on the consideration that the interpretations of all expressions included in these languages depend only on the current graphical state and no intensional types are included in the system. However, this analysis can be extended on the lines of Intensional Logic to be able to deal with a more comprehensive fragment of English.
[10] An expression of type <*a,b*> combines with an expression of type *a* to give an expression of type *b*.

Constant	Type
$d_1, d_2, d_3, r_1, r_2, r_3, r_4, l_1$	e
dot, region, curve, line, intersection	$<e, t>$
right	$<<<e, t>, t>, <e, t>>$
lie_at, be_in_zone, inside	$<<<e, t>, t>, <e, t>>$
$=$	$<e, <e, t>>$
$\wedge, \vee, \leftrightarrow$	$<t, <t, t>>$
curve_between, intersection_between, line_from_to	$<<<e, t>, t>, <<<e, t>, t>, <e, t>>>$
$right_*$	$<e, e>$
lie_at_*, $be\_in\_zone_*$, $inside_*$	$<e, <e, t>>$
$curve\_between_*$, $intersection\_between_*$, $line\_from\_to_*$	$<e, <e, e>>$

Figure 10
Constants of language **G**.

6. If $\mu \in V_s$ and $\beta \in E_t$ then $\forall\mu(\beta) \in E_t$.

G is a very expressive language and not every well-formed expression has a translation into **L** as will be further discussed in Section 2.5. Useful translation are, for instance, names and descriptions of geometrical objects and configurations. Next, the definition of expressions of **G** that have a translation into **L** is presented. For clarity, the abbreviations in Figure 11 are used.

Abbreviation	Formal expression	Type
A	$\lambda P\lambda Q\exists x[P(x)\wedge Q(x)]$	$<<e, t,>, <<e, t>, t>>$
THE	$\lambda P\lambda Q\exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]$	$<<e, t,>, <<e, t>, t>>$
BE$_\mathrm{a}$	$\lambda\mathcal{P}\lambda x\mathcal{P}(\lambda y[x=y])$	$<<<e, t>, t>, <e, t>>$
BE$_\mathrm{b}$	$\lambda P\lambda xP(x)$	$<<e, t>, <e, t>>$
D$_i$	$\lambda P[P(d_i)]$	$<<e, t>, t>$
R$_i$	$\lambda P[P(r_i)]$	$<<e, t>, t>$
OF$_\mathrm{a}$	$\lambda x_{<<e, t>, t>}\ \lambda y_{<e, t>}\ \lambda z_e[y(z) \wedge inside(x)(z)]$	$<<<e, t>, t>, <<e, t>, <e, t>>>$
OF$_\mathrm{b}$	$\lambda x_{<<e, t>, t>}\ \lambda y_{<<<e, t>, t>, e>}\ \lambda z_e[y(x)(z)]$	$<<<e, t>, t>, <<<<e, t>, t>, e>,$ $<e, t>>>$
BETWEEN$_\mathrm{a}$	$\lambda x_{<<e, t>, t>}\ \lambda y_{<<e, t>, t>}\ \lambda z_{<e, t>}\ \lambda u_e\ [z(u) \wedge curve\_between(x)(y)(u)]$	$<<<e, t>, t>, <<<e, t>, t>,$ $<<e, t>, <e, t>>>>$
BETWEEN$_\mathrm{b}$	$\lambda x_{<<e, t>, t>}\ \lambda y_{<<e, t>, t>}\ \lambda z_{<e, t>}\ \lambda u_e\ [z(u) \wedge intersection\_between(x)(y)(u)]$	$<<<e, t>, t>, <<<e, t>, t>,$ $<<e, t>, <e, t>>>>$
FROM_TO	$\lambda x_{<<e, t>, t>}\ \lambda y_{<<e, t>, t>}\ \lambda z_{<e, t>}\ \lambda u_e\ [z(u) \wedge line\_from\_to(x)(y)(u)]$	$<<<e, t>, t>, <<<e, t>, t>,$ $<<e, t>, <e, t>>>>$

Figure 11
Short-hand definitions.

Two geometrical interpretations are given for the spatial prepositions *of* and *between.* Although the characterization of the meaning of these words is a very complex problem that is beyond current discussion, we allow that spatial prepositions can be interpreted in more than one way, as long each interpretation is stated in terms of a geometrical algorithm explicitly defined in **G**. Consider, for instance, the spatial meaning of *of* is different in *city of France* and *east of France*. In the former term *of* denotes an spatial inclusion relation (OF$_\mathrm{a}$) but in the latter it denotes a relation of adjacency (OF$_\mathrm{b}$). Similarly, the spatial meaning of *between* in *border between France and Germany* and *intersection between the border between France and Germany and a line from Paris to Frankfurt* is different, as it denotes a curve in the first case (BETWEEN$_\mathrm{a}$) and a dot in the second (BETWEEN$_\mathrm{b}$).

The restrictions for the expressions of **G** that can be translated into **L** are as follows. In rules S6$_\mathrm{G}$ to S8$_\mathrm{G}$ Q stands for either the quantifier A or THE.

SENTENCES

S1$_G$.

If $\alpha \in E_{<<e,\, t>,\, t>}$ y $\beta \in E_{<e,\, t>}$, then $F_{G1}(\alpha,\beta) \in E_t$, where $F_{G1}(\alpha,\beta) = \alpha(\beta)$.

Examples:
- D$_1$ (BE$_a$ (A (OF$_a$(R$_1$) (*dot*))))
- R$_3$ (*be_in _zone* (THE (OF$_b$(R$_1$) (*right*))))
- A (*region*) (BE$_b$(*big*))
- D$_3$ (*lie_at* (THE (BETWEEN$_b$(THE (BETWEEN$_a$(R$_1$) (R$_3$) (*curve*)))
(A (FROM_TO(D$_1$) (D$_3$) (*line*)))
(*intersection*))))

TRANSITIVE VERB PHRASES

S2$_G$. If $\alpha \in E_{<<<e,\, t>,\, t>,\, <e,\, t>>}$ and $\beta \in E_{<<e,\, t>,\, t>}$ then $F_{G1}(\alpha,\beta) \in E_{<e,\, t>}$.

Examples:
- BE$_a$ (A (*dot*))
- *be_in_zone* (THE (OF$_b$(R$_1$)(*right*)))

ATTRIBUTIVE VERB PHRASES

S3$_G$. If $\alpha \in E_{<<e,\, t>,\, <e,\, t>>}$ and $\beta \in E_{<e,\, t>}$ then $F_{G1}(\alpha,\beta) \in E_{<e,\, t>}$.

Example: - BE$_b$ (*big*)

TERMS

S4$_G$. If $\alpha \in E_{<<e,\, t>,\, <<e,\, t>,\, t>>>}$ and $\beta \in E_{<e,\, t>}$, then $F_{G1}(\alpha,\beta) \in E_{<<e,\, t>,\, t>}$.

Examples:
- A (*dot*)
- A (OF$_a$(R$_1$)(*dot*))
- THE (BETWEEN$_a$(R$_1$) (R$_2$) (*curve*))
- A (FROM_TO(D$_1$) (D$_3$) (*line*))
- THE (OF$_b$(R$_1$)(*right*))

COMMON NOUNS

S5$_G$. If $\alpha \in E_{<<e,\, t>,\, <e,\, t>>}$ and $\beta \in E_{<e,\, t>}$, or $\alpha \in E_{<<<<e,\, t>,\, t>,\, e>,\, <e,\, t>>}$ and $\beta \in E_{<<<e,\, t>,\, t>,\, e>}$, then $F_{G1}(\alpha,\beta) \in E_{<e,\, t>}$.

Examples:
- OF$_a$(R$_1$)(*dot*)
- OF$_b$(R$_1$)(*right*)
- BETWEEN$_a$(R$_1$) (R$_2$) (*curve*)
- BETWEEN$_b$ (THE(BETWEEN$_a$(R$_1$)(R$_2$)(*curve*))) (A(FROM_TO(D$_1$)(D$_3$)(*line*))) (*intersection*)

of PREPOSITIONAL PHRASES

S6$_G$. If $\alpha \in E_{<<e,\, t>,\, t>}$ such that α is either R$_i$ or Q(*region*), then $F_{G2}(\alpha) \in E_{<<e,\, t>,\, <e,\, t>>}$ and $F_{G3}(\alpha) \in E_{<<<<e,\, t>,\, t>,\, <e,\, t>>,\, <e,\, t>>}$, where $F_{G2}(\alpha)$ = OF$_a$(α) and $F_{G3}(\alpha)$ = OF$_b$(α)

Examples:
- OF$_a$(R$_1$)
- OF$_b$(R$_2$)
- OF$_a$(A(*region*))

between PREPOSITIONAL PHRASES

S7$_G$. (a) If α, β∈ $E_{<<e,\ t>,\ t>}$ such that α,β are either R_i or Q(*region*), then $F_{G4}(\alpha,\beta) \in E_{<<e,\ t>,\ <e,\ t>>}$, where $F_{G4}(\alpha,\beta)$ = BETWEEN$_a$(α)(β).

(b) If α, β∈ $E_{<<e,\ t>,\ t>}$ such that α is either C_i or Q(*curve*) and β is either L_i or Q(*line*), then $F_{G5}(\alpha,\beta) \in E_{<<e,\ t>,\ <e,\ t>>}$, where $F_{G5}(\alpha,\beta)$ = BETWEEN$_b$(α)(β).

Examples: - BETWEEN$_a$(R_1) (R_2)
- BETWEEN$_a$(R_1) (A(*region*))
- BETWEEN$_b$ (THE (BETWEEN$_a$(R_1) (R_2) (*curve*))) (A (FROM_TO(D_1) (D_3) (*line*)))

from-to PREPOSITIONAL PHRASES

S8$_G$. If α, β∈ $E_{<<e,\ t>,\ t>}$ such that α,β are either D_i or Q(*dot*), then $F_{G5}(\alpha,\beta) \in E_{<<e,\ t>,\ <e,\ t>>}$, where $F_{G5}(\alpha,\beta)$ = FROM_TO(α)(β)

Example: - FROM_TO(D_1) (D_3)

2.3.2 Semantic definition of G

The semantics of **G** is given in a model-theoretic fashion as follows. Let $\mathbf{P_{base}} = \{d_1, d_2, d_3, r_1, r_2, r_3, r_4, l_1\}$ be the set of basic graphical objects shown in Figure 4. Let **P** be the union of $\mathbf{P_{base}}$ and all graphical objects that can be produced from $\mathbf{P_{base}}$ with the help of geometrical functions: the emergent objects. Emergent objects can also be produced on the basis of other emergent objects previously generated. A particular kind of emergent object that is interesting for the current discussion is the zone of a map that is considered to be the east of a region. For the production of emergent objects in **P** there is a well-defined computational geometry algorithm associated to an operator symbol of **G** as will be seen below.

Let D_x be the set of possible denotations for expressions of type *x*, such that D_e = **P**, D_t = {1, 0}, and, for any types *a* and *b*, $D_{<a,b>} = D_b^{D_a}$. Let $\mathbf{F_G}$ be an interpretation function that assigns to each constant of type *a* a member of D_a. For every graphical object φ in $\mathbf{P_{base}}$ there is a constant α of type *e* such that $\mathbf{F_G}(\alpha)=\varphi$; for our example $\mathbf{F_G}$ assigns the objects $d_1, d_2, d_3, r_1, r_2, r_3, r_4, l_1$ to the constants $d_1, d_2, d_3, r_1, r_2, r_3, r_4, l_1$, respectively. The interpretation (assigned by $\mathbf{F_G}$) of the geometrical-type predicates *dot*, *region*, *curve*, *line*, *intersection* are the sets containing the corresponding graphical objects. The constants *right*$_*$, *lie_at*$_*$, *be_in_zone*$_*$, *inside*$_*$, *curve_between*$_*$, *intersection_between*$_*$ and *line_from_to*$_*$ are interpreted as geometrical functions. If the arguments of these geometrical functions are of an appropriate type, expressions containing these constants can be properly interpreted through geometrical algorithms; otherwise, these expressions have no denotation in **G** and, as a consequence, their translation into **L** lack a denotation too. For further discussion of the interpretation of graphical expressions which have no proper graphical referent in the interpretation state see (Pineda, 1992).

Following Montague, the interpretation of variables is defined in terms of an assignment function *g*. We adopt the notational convention by which the semantic value or denotation of an expression α with respect to a model *M* and a value assignment *g* is expressed as $[[\alpha]]^{M,g}$.

The semantic rules for interpreting expressions of **L** are the following:

1. If $\alpha \in C_s$, then $[[\alpha]]^M = \mathbf{F_G}(\alpha)$.

2. If $\mu \in V_s$, then $[[\mu]]^{M,g} = g(\mu)$.

3. If $\alpha \in E_{<a,b>}$, and $\beta \in E_a$, then $[[\alpha(\beta)]]^{M,g} = [[\alpha]]^{M,g}([[\beta]])^{M,g}$

4. If $\alpha \in E_a$ and $u \in V_b$, then $[[\lambda u[\alpha]\ ]]^{M,g}$ is that function *h* from D_b into D_a such that for all objects *k* in D_b, *h*(*k*) is equal to $[[\alpha]]^{M,g'}$, where *g'* is exactly like *g* except that *g'*(*u*)=*k*.

5. If $\mu \in V_s$ y $\beta \in E_t$ then $[[\exists\mu(\beta)]]^{M,g}$ =1 iff for some value assignment *g*' such that *g*' is exactly like *g* except possibly for the individual assigned to μ by *g*', $[[\beta]]^{M,g'}$ =1.

6. If $\mu \in V_s$ y $\beta \in E_t$ then $[[\forall\mu(\beta)]]^{M,g}$ =1 iff for every value assignment *g*' such that *g*' is exactly like *g* except possibly for the individual assigned to μ by *g*', $[[\beta]]^{M,g'}$ =1.

As was mentioned, graphical predicates like *be_in_zone*, *curve_between* and *inside* have type-raised arguments; for instance, the expression *curve_between*$(\lambda P[P(r_1)])(\lambda P[P(r_2)])(x)$ refers to the curve *x* between regions r_1 and r_2;

however, the first two arguments of this predicate refer not to the regions themselves, but to the set of properties that such regions have. Similarly, the expression $inside(\lambda P \exists y[region(y) \wedge P(y)])(z)$ denotes that the dot z is inside a region y, but the first argument denotes the set of properties P that a region has instead of y directly. In order to complete the geometrical interpretation process, the first of these expressions must be transformed into $curve\_between_{*}(r_1,r_2)=x$ and the second into $\exists y[region(y) \wedge inside_{*}(z,y)]$, where $curve\_between_{*}$ and $inside_{*}$ denote geometrical functions whose arguments are graphical entities. Although graphical terms are type-raised for preserving the quantification properties along the translation process, the graphical entities themselves are the objects of concrete geometrical computations.

In order to assign an appropriate semantical value to expressions, syntactic transformations like those mentioned above are carried out by means of the following *meaning postulates*:

MP1. $\forall x \forall \mathsf{P}\ [\delta(\mathsf{P})(x) \leftrightarrow \mathsf{P}(\lambda y[\delta_{*}(x,y)])$ where $\delta \in \{lie\_at, be\_in\_zone, inside\}$

MP2. $\forall x \forall \mathsf{P}\ [\delta(\mathsf{P})(x) \leftrightarrow \mathsf{P}(\lambda y[\delta_{*}(y)=x])$ where $\delta \in \{right\}$

MP3. $\forall x \forall \mathsf{P}_1 \forall \mathsf{P}_2\ [\delta(\mathsf{P}_1)(\mathsf{P}_2)(x) \leftrightarrow \mathsf{P}_2(\lambda u[\mathsf{P}_1(\lambda v[\delta_{*}(v,u)=x])])$ where $\delta \in \{curve\_between, intersection\_between, line\_from\_to\}$

where P, P_1 and P_2 are variables ranging over sets of properties (i.e., of type $<<e, t>, t>$), and x, y, u, v over individuals. Meaning postulate MP1 establish, for instance, that a geometrical relation that holds between a set of properties of an individual a and an individual b stands in one to one correspondence with the relation that holds between the individuals a and b themselves, as the only property of a that is relevant for the geometrical interpretation process is the property of being in such geometrical relation with the object b (i.e., that the object a lies at, is in a zone or is inside the object b). Similarly for meaning postulates MP2 and MP3.

In order to illustrate how the graphical interpreter works the following five examples are presented:

Example 1

Consider the interpretation of the expression A(*region*) (BE (*big*)), which is the translation of *a country is big*. The expression can be reduced as follows:

(1) $\lambda P \lambda Q \exists x[P(x) \wedge Q(x)]\ (region)\ (\lambda P \lambda z P(z))\ (big)$
(2) $\lambda P \lambda Q \exists x[P(x) \wedge Q(x)]\ (region)\ (\lambda z\ big(z))$
(3) $\lambda Q \exists x[region(x) \wedge Q(x)]\ (\lambda z\ big(z))$
(4) $\exists x[region(x) \wedge \lambda z\ big(z)\ (x)]$
(5) $\exists x[region(x) \wedge big(x)]$

Expression (5) is interpreted through the standard quantification rules of the geometrical interpreter without the help of meaning postulates. The interpretation of *big* is an algorithm that computes the average area of all regions in the map and returns the set of all regions whose area is larger than the average. This is a simple convention for illustrative purposes and alternative conventions could be chosen. Although the purpose of this paper is not to explore issues related to the interpretation of vague terms, it is interesting to note that within the present framework specific algorithms related to specific application domains which take into account the graphical context could be defined for the construction of practical applications.

Example 2

Consider the interpretation of THE (BETWEEN_a (R_1) (R_2) (*curve*)) —which is the translation of *the border between France and Germany* as will be shown in Section 2.4.1. The expression without the abbreviations is:

(1) $\lambda P \lambda Q \exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]$
$(\lambda x_{<<e,t>,t>}\ \lambda y_{<<e,t>,t>}\ \lambda z_{<e,t>}\ \lambda u_e\ [z(u) \wedge curve\_between(x)(y)(u)]\ (\lambda P[P(r_1)])\ (\lambda P[P(r_2)])\ (curve))$

which can be reduced as follows:

(2) $\lambda P \lambda Q \exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]\ \ (\lambda u[curve(u) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P[P(r_2)])(u)])$
(3) $\lambda Q \exists y[\forall x[\lambda u[curve(u) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P[P(r_2)])(u)](x) \leftrightarrow x=y] \wedge Q(y)]$
(4) $\lambda Q \exists y[\forall x[(curve(x) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P[P(r_2)])(x)\ ) \leftrightarrow x=y] \wedge Q(y)]$
(5) $\lambda Q \exists y[\forall x[(curve(x) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P[P(r_2)])(x)\ ) \leftrightarrow x=y] \wedge Q(y)]$
(6) $\lambda Q \exists y[\forall x[(curve(x) \wedge \lambda P[P(r_2)](\lambda u[\lambda P[P(r_1)](\lambda v[curve\_between_{*}(v,u)=x])])\ ) \leftrightarrow x=y] \wedge Q(y)]$
(7) $\lambda Q \exists y[\forall x[(curve(x) \wedge \lambda P[P(r_2)](\lambda u[\lambda v[curve\_between_{*}(v,u)=x](r_1)])\ ) \leftrightarrow x=y] \wedge Q(y)]$
(8) $\lambda Q \exists y[\forall x[(curve(x) \wedge \lambda P[P(r_2)](\lambda u[curve\_between_{*}(r_1,u)=x])\ ) \leftrightarrow x=y] \wedge Q(y)]$
(9) $\lambda Q \exists y[\forall x[(curve(x) \wedge \lambda u[curve\_between_{*}(r_1,u)=x](r_2)\ ) \leftrightarrow x=y] \wedge Q(y)]$
(10) $\lambda Q \exists y[\forall x[(curve(x) \wedge curve\_between_{*}(r_1, r_2)=x) \leftrightarrow x=y] \wedge Q(y)]$

Note that expression (5) cannot be further reduced unless the types of the arguments of the predicate *curve_between* are lowered with the help of meaning postulate MP3. The geometrical functions in expression (10) can be evaluated directly. Expression (10) is a denoting concept which refers to the curve between the regions r_1 and r_2 and cannot be further reduced. Consider that the expression *the border between France and Germany* is a definite description and, in order to obtain a truth value, it has to be combined with a predicate. However, if we were interested in the graphical object referred to by (10), it could be identified regardless the nature of the predicate Q, as this predicate is not used for picking up the object referred to by the definite description[11]. We call the object referred to by the denoting concept its "concrete extension". The concrete extension of (10) can be identified, for instance, by interpreting the denoting concept taking off the predicative abstraction Q (i.e., $\exists y[\forall x[(curve(x) \wedge curve\_between_*(r_1, r_2)=x) \leftrightarrow x=y]]$) in relation to the graphical domain; if the denoting concept is indefinite, we take any object satisfying the expression as its concrete extension.

Example 3

Consider the interpretation of an expression similar to the one in example 2, but in which an indefinite is included. The expression is THE (BETWEEN$_a$ (R$_1$) (A(*region*)) (*curve*)), which is the translation of *the border between France and a country*. The full expression is:

(1) $\lambda P\lambda Q\exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]$ $(\lambda x_{<<e,t>,t>}\ \lambda y_{<<e,t>,t>}\ \lambda z_{<e,t>}\ \lambda u_e\ [z(u) \wedge curve\_between(x)(y)(u)]$ $(\lambda P[P(r_1)])\ (\lambda P\exists z[region(z) \wedge P(z)])\ (curve)))$

the reduction is as follows:

(2) $\lambda P\lambda Q\exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]\ (\lambda u[curve(u) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P\exists z[region(z) \wedge P(z)])(u)])$

(3) $\lambda Q\exists y[\forall x[\lambda u[curve(u) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P\exists z[region(z) \wedge P(z)])(u)](x) \leftrightarrow x=y] \wedge Q(y)]$

(4) $\lambda Q\exists y[\forall x[(curve(x) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P\exists z[region(z) \wedge P(z)])(x)\ ) \leftrightarrow x=y] \wedge Q(y)]$

(5) $\lambda Q\exists y[\forall x[(curve(x) \wedge \lambda P\exists z[region(z) \wedge P(z)](\lambda u[\lambda P[P(r_1)](\lambda v[curve\_between_*(v,u)=x])])\ ) \leftrightarrow x=y] \wedge Q(y)]$

(6) $\lambda Q\exists y[\forall x[(curve(x) \wedge \lambda P\exists z[region(z) \wedge P(z)](\lambda u[\lambda v[curve\_between_*(v,u)=x](r_1)])\ ) \leftrightarrow x=y] \wedge Q(y)]$

(7) $\lambda Q\exists y[\forall x[(curve(x) \wedge \lambda P\exists z[region(z) \wedge P(z)](\lambda u[curve\_between_*(r_1,u)=x])\ ) \leftrightarrow x=y] \wedge Q(y)]$

(8) $\lambda Q\exists y[\forall x[(curve(x) \wedge \exists z[region(z) \wedge \lambda u[curve\_between_*(r_1,u)=x](z)]\ ) \leftrightarrow x=y] \wedge Q(y)]$

(10) $\lambda Q\exists y[\forall x[(curve(x) \wedge \exists z[region(z) \wedge curve\_between_*(r_1,z)=x]\ ) \leftrightarrow x=y] \wedge Q(y)]$

Meaning postulate MP3 is used for reducing from (4) to (5). Expression (10) is a denoting concept similar to the final expression in example 2, but one which has an embedded quantified expression. Meaning postules MP1 to MP3 are defined in such a way that terms preserve quantificational properties along the reduction process.

Example 4

Consider the expression R$_2$(*be_in_zone*(THE(OF$_b$(R$_1$)(*right*)))) —which is the translation of *Germany is to the east of France*. The reduced expression is the following:

(1) $be\_in\_zone\ (\lambda Q\exists y[\forall x[right(\lambda P[P(r_1)])(x) \leftrightarrow x=y] \wedge Q(y)])\ (r_2)$

by meaning postulate MP2:

(2) $be\_in\_zone\ (\lambda Q\exists y[\forall x[\lambda P[P(r_1)](\lambda z[right_*(z)=x]) \leftrightarrow x=y] \wedge Q(y)])\ (r_2)$

(3) $be\_in\_zone\ (\lambda Q\exists y[\forall x[\lambda z[right_*(z)=x](r_1) \leftrightarrow x=y] \wedge Q(y)])\ (r_2)$

(4) $be\_in\_zone\ (\lambda Q\exists y[\forall x[right_*(r_1)=x \leftrightarrow x=y] \wedge Q(y)])\ (r_2)$

by meaning postulate MP1:

(5) $\lambda Q\exists y[\forall x[right_*(r_1)=x \leftrightarrow x=y] \wedge Q(y)]\ (\lambda z[be\_in\_zone_*(r_2, z)])$

(6) $\exists y[\forall x[right_*(r_1)=x \leftrightarrow x=y] \wedge \lambda z[be\_in\_zone_*(r_2, z)](y)]$

(7) $\exists y[\forall x[right_*(r_1)=x \leftrightarrow x=y] \wedge be\_in\_zone_*(r_2, y)]$.

Expression (7) is a first-order formula which can be directly evaluated by the intepreter of **G**. The operator $right_*$ is interpreted as a geometrical algorithm which computes the centroid (x_c,y_c) of a region r and returns the semi-plane to

[11] As argued by Kaplan, contextual factors have to be considered for the identification of the referent of a definite description used referentially instead of attributively (Kaplan, 1978). If the referent is identified deictically, as in the current example, the referent is found through the translation of the definite description into the graphical language, where the shape of the object is available directly. Note as well that as expressions of **G** have an interpretation not only in relation to the graphical domain but also in relation to the world, through the translation into **P** and the semantics of **P**, the referent of a definite description in **L** can be found by computing the geometrical interpretation of its translation into **G**.

the right of the centroid of r (i.e., the set of all ordered-pairs of reals $<x_i, y_i>$ such that $x_i > x_c$). This convention captures whether a graphical object is to the right of a region, or whether the object is in the right part of a region[12]. The graphical predicate *be_in_zone*$_*$ checks whether r_2 is within y —i.e., the zone to the right of r_1.

Example 5

Consider the interpretation of the translation into **G** of the textual part of the multimodal message in Figure 2. The translation of *Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt* is shown in (1), its reduction in (2), and its final reduction applying the meaning postulates in (3):

(1) D_3 (*lie_at* (THE ($\mathrm{BETWEEN_b}$ (THE ($\mathrm{BETWEEN_a}$ (R_1) (R_3) (*curve*)))
(A (FROM_TO(D_1) (D_3) (*line*)))
(*intersection*))))

(2) $lie\_at(\lambda Q \exists y[\forall x[intersection(x) \wedge intersection\_between$
$(\lambda Q \exists u[\forall v[(curve(v) \wedge curve\_between(\lambda P[P(r_1)])(\lambda P[P(r_2)])(v)\ ) \leftrightarrow v{=}u] \wedge Q(y)])$
$(\lambda Q \exists z[line(z) \wedge line\_from\_to(\lambda P[P(d_1)])(\lambda P[P(d_3)])(z) \wedge Q(z)])$
$=x \leftrightarrow x{=}y] \wedge Q(y)])\ (\mathrm{D}_3)$

(3) $\exists y[\forall x[(intersection(x) \wedge \exists z[line(z) \wedge line\_from\_to_*(d_1, d_3){=}z \wedge$
$\exists u[\forall v[(curve(v) \wedge curve\_between_*(r_1, r_2){=}v) \leftrightarrow v{=}u] \wedge$
$intersection\_between_*(u,z){=}x])\ \leftrightarrow x{=}y] \wedge lie\_at_*(d_2, y)].$

Expression (3) is true if the position of dot d_3 is the same as the position of the intersection between the curve between r_1 and r_2 and the line from d_1 to d_3, as is the case in Figure 2.

As a final remark in this section it is worth emphasizing that as the five examples illustrate, the reason for type-raising graphical terms is to be able to translate natural language quantified expression into the graphical domain.

2.4 Translations between L and G.

In this section the translation funtions $\rho_{\mathbf{L\text{-}G}}$ and $\rho_{\mathbf{G\text{-}L}}$ are defined. Although in the interpretation of a multimodal message like the one in Figure 2 the translation of individual constants are not known, and the purpose of the multimodal interpretation process is to find out such translations, as was discussed in Section 1, in this section we assume that the translation is fully defined in order to illustrate all theoretical elements of the scheme in Figure 3. The induction of the translation of individual constants, on the other hand, will be shown in Section 3.

For each syntactic category of **L** there is a corresponding type in **G**. The correspondence between linguistic categories and geometrical types resembles the translation from English to Intensional Logic (Dowty, 1985) and it is defined in terms of the function f as follows:

1. $f(t) = t$.
2. $f(CN) = f(IV) = f(ADJ) = <e, t>$.
3. For any categories A and B, $f(A/B) = <f(B), f(A)>$.

In Section 2.4.1 the translation from **L** into **G** is presented, and in Section 2.4.2. the useful translations from **G** into **L** are shown.

2.4.1 Translation from L into G

In Figure 12 the translation for constants of **L** is presented. As can be seen, simple terms like the names of cities and countries translate into expressions denoting characteristic functions of sets of graphical entities. This graphical type is interpreted as the set of “properties” that an individual named by the term has (for the purpose of this dicussion a property is just the set of individuals, as no intensional types are considered). So, as a city is represented through a dot in the graphical domain, the translation of Paris, for instance, is the set of geometrical properties that the dot representing Paris has in the intepretation state. Common nouns of category *CN* and *CN’* translate into predicates and functions from sets of properties to individuals, respectively. Adjetives occurring in attributives sentences are

[12] This is an arbitrary convention defined for illustrative purposes and alternative conventions could be chosen. Similar conventions could be taken to interpret whether other kind of graphical objects stand in a right-of relation. Furthermore, several conventions for the interpretation of this kind of words can be used and a particular geometric algorithm can be defined for each of such interpretations. These algorithms need not be fully quantitative and more qualitative approaches can be employed as long as the computation returns a semantic value of an appropriate kind.

translated as set of individuals. Note that there are two constants *be*: one combines with a term and the other with an adjective and both combinations produce intransitive verbs. The translation corresponding to these constants are functions from sets of properties to sets of individuals, and from sets of individuals to set of individuals, respectively. Transitive verbs like *lie_at* and *be_to* translate into geometrical operators whose type is a function from sets of properties to sets of individuals. Determiners are translated in a standard fashion.

constant of L: α	**Category name**	**Category definition**	**Translation into G:** $\rho_{\mathbf{L\text{-}G}}(\alpha)$	**Corresponding type in G**
Paris	*T*	*t/IV*	$\lambda P[P(d_1)]$	$<<e, t>, t>$
Frankfurt	*T*	*t/IV*	$\lambda P[P(d_3)]$	$<<e, t>, t>$
Saarbrücken	*T*	*t/IV*	$\lambda P[P(d_2)]$	$<<e, t>, t>$
France	*T*	*t/IV*	$\lambda P[P(r_1)]$	$<<e, t>, t>$
Germany	*T*	*t/IV*	$\lambda P[P(r_2)]$	$<<e, t>, t>$
city	*CN*	*CN*	*dot*	$<e, t>$
country	*CN*	*CN*	*region*	$<e, t>$
border	*CN*	*CN*	*curve*	$<e, t>$
line	*CN*	*CN*	*line*	$<e, t>$
intersection	*CN*	*CN*	*intersection*	$<e, t>$
east	*CN'*	*CN'*	*right*	$<<<e, t>, t>, e>$
big	*ADJ*	*ADJ*	*big*	$<e, t>$
be	*TV*	*IV/(t/IV)*	$\lambda P\lambda xP(\lambda y[x=y])$	$<<<e, t>, t>, <e, t>>$
be	*IV/ADJ*	*IV/ADJ*	$\lambda P\lambda xP(x)$	$<<e, t>, <e, t>>$
lie at	*TV*	*IV/(t/IV)*	*lie_at*	$<<<e, t>, t>, <e, t>>$
be to	*TV*	*IV/(t/IV)*	*be_in_zone*	$<<<e, t>, t>, <e, t>>$
a	*T/CN*	*(t/IV)/CN*	$\lambda P\lambda Q\exists x[P(x) \wedge Q(x)]$	$<<e, t,>, <<e, t>, t>>$
the	*T/CN*	*(t/IV)/CN*	$\lambda P\lambda Q\exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]$	$<<e, t,>, <<e, t>, t>>$

Figure 12
Translation of constants of **L** into **G**.

The translation rules for composite expressions are as follows.

SENTENCES

T1$_{\text{L-G}}$. If $\alpha\in P_T$ and $\beta\in P_{IV}$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then $\rho_{\mathbf{L\text{-}G}}(F_{L1}(\alpha,\beta)) = \alpha'(\beta')$, that is to say, the function α' applied to the argument β'.

Examples: $\rho_{\mathbf{L\text{-}G}}$(*Paris is a city of France*) = D_1 (BE_a (A (OF_a(R_1) (*dot*))))
$\rho_{\mathbf{L\text{-}G}}$(*Germany is to the east of France*) = R_3 (*be_in _zone* (THE (OF_b(R_1) (*right*))))
$\rho_{\mathbf{L\text{-}G}}$(*a country is big*) = A(*region*) (BE_b(*big*))
$\rho_{\mathbf{L\text{-}G}}$(*Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt*) =
D_3 (*lie_at* (THE (BETWEEN_b (THE (BETWEEN_a (R_1) (R_3) (*curve*)))
(A (FROM_TO(D_1) (D_3) (*line*))) (*intersection*))))

TRANSITIVE VERB PHRASES

T2$_{\text{L-G}}$. If $\alpha\in P_{TV}$ and $\beta\in P_T$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then$\rho_{\mathbf{L\text{-}G}}(F_{L2}(\alpha,\beta)) = \alpha'(\beta')$.

Examples: $\rho_{\mathbf{L\text{-}G}}$(*be a city*) = BE_a (A (*dot*))
$\rho_{\mathbf{L\text{-}G}}$(*be to the east of France*) = *be_in_zone* (THE (OF_b(R_1)(*right*)))

ATTRIBUTIVE VERB PHRASES

T3$_{L\text{-}G}$. If $\alpha\in P_{IV/ADJ}$ and $\beta\in P_{ADJ}$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then $\rho_{\mathbf{L\text{-}G}}(F_{L2}(\alpha,\beta)) = \alpha'(\beta')$.

Example: $\rho_{\mathbf{L\text{-}G}}$(*be big*) = BE_b(*big*)

TERMS

T4$_{L\text{-}G}$. If $\alpha\in P_{T/CN}$ and $\beta\in P_{CN}$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then $\rho_{\mathbf{L\text{-}G}}(F_{L3}(\alpha,\beta)) = \alpha'(\beta')$.

Examples: $\rho_{\mathbf{L\text{-}G}}$(*a city*) = A (*dot*)
$\rho_{\mathbf{L\text{-}G}}$(*a city of France*) = A ($\mathrm{OF}_a(\mathrm{R}_1)$(*dot*))
$\rho_{\mathbf{L\text{-}G}}$(*the border between France and Germany*) = THE ($\mathrm{BETWEEN}_a$ (R_1) (R_2) (*curve*))
$\rho_{\mathbf{L\text{-}G}}$(*a line from Paris to Frankfurt*) = A (FROM_TO(D_1) (D_3) (*line*))
$\rho_{\mathbf{L\text{-}G}}$(*the east of France*) = THE ($\mathrm{OF}_b(\mathrm{R}_1)$(*right*))

As can be seen the term *the east* can be formed by the rule S4$_L$, but it cannot be translated into **G** because there is a type restriction in the definition of T4$_{L\text{-}G}$ (i.e., $\beta\in P_{CN}$ but *east* $\in P_{CN'}$). This restriction prevents the translation of terms like *the east* as these expressions have no concrete graphical representation; however, *the east of France* can be generated, translated into **G** and interpreted through the geometry as was shown in Section 2.3.2. In general, natural language expressions denoting abstract concepts do not have a graphical representation (i.e., *the population of France*), and although in this grammar we have focused on expressions that can be translated into **G**, the language can be extended with linguistic terms that would be interpreted only in the linguistic modality.

COMMON NOUNS

T5$_{L\text{-}G}$. If $\alpha\in P_{CN}$ and $\beta\in P_{PP}$, or $\alpha\in P_{CN'}$ and $\beta\in P_{PP'}$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then $\rho_{\mathbf{L\text{-}G}}(F_{L2}(\alpha,\beta)) = \beta'(\alpha')$.

Examples: $\rho_{\mathbf{L\text{-}G}}$(*city of France*) = $\mathrm{OF}_a(\mathrm{R}_1)$(*dot*)
$\rho_{\mathbf{L\text{-}G}}$(*east of France*) = $\mathrm{OF}_b(\mathrm{R}_1)$(*right*)
$\rho_{\mathbf{L\text{-}G}}$(*border between France and Germany*) = $\mathrm{BETWEEN}_a$ (R_1) (R_2) (*curve*)
$\rho_{\mathbf{L\text{-}G}}$(*intersection between the border between France and Germany and a line from Paris to Frankfurt*) = $\mathrm{BETWEEN}_b$ (THE ($\mathrm{BETWEEN}_a$ (R_1) (R_2) (*curve*))) (A (FROM_TO(D_1) (D_3) (*line*))) (*intersection*)

of PREPOSITIONAL PHRASES

T6$_{L\text{-}G}$. If $\alpha\in P_T$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, then $\rho_{\mathbf{L\text{-}G}}(F_{L4}(\alpha))$ is either $\mathrm{OF}_a(\alpha')$ or $\mathrm{OF}_b(\alpha')$.

Examples: $\rho_{\mathbf{L\text{-}G}}$(*of France*) = $\mathrm{OF}_a(\mathrm{R}_1)$
$\rho_{\mathbf{L\text{-}G}}$(*of Germany*) = $\mathrm{OF}_b(\mathrm{R}_2)$

between PREPOSITIONAL PHRASES

T7$_{L\text{-}G}$. If $\alpha,\beta\in P_T$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then $\rho_{\mathbf{L\text{-}G}}(F_{L5}(\alpha,\beta)) = \mathrm{BETWEEN}(\alpha')(\beta')$.

Examples: $\rho_{\mathbf{L\text{-}G}}$(*between France and Germany*) = BETWEEN(R_1) (R_2)
$\rho_{\mathbf{L\text{-}G}}$(*between the border between France and Germany and a line from Paris to Frankfurt*) = $\mathrm{BETWEEN}_b$(THE($\mathrm{BETWEEN}_a(\mathrm{R}_1)(\mathrm{R}_2)$(*curve*))) (A(FROM_TO($\mathrm{D}_1$)($\mathrm{D}_3$) (*line*)))

from-to PREPOSITIONAL PHRASES

T8$_{L\text{-}G}$. If $\alpha,\beta\in P_T$, and $\rho_{\mathbf{L\text{-}G}}(\alpha)=\alpha'$, $\rho_{\mathbf{L\text{-}G}}(\beta)=\beta'$ then $\rho_{\mathbf{L\text{-}G}}(F_{L6}(\alpha,\beta)) = \mathrm{FROM\_TO}(\alpha')(\beta')$.

Example: $\rho_{\mathbf{L\text{-}G}}$(*from Paris to Frankfurt*) = FROM_TO (D_1) (D_3)

2.4.2 Translation from G into L

In this section the translation function $\rho_{\mathbf{G\text{-}L}}$ is defined. The translation of expressions of **G** into **L** are shown in Figures 13 and 14. Note that constants of **G** in Figure 13 translate into constants of **L**; however, the translations shown in Figure 14 are more complex as composite expressions of **G** can translate into basic or composite expressions of **L**.

Constant of G: α	**Translation into L:** $\rho_{\mathbf{G\text{-}L}}(\alpha)$
dot	*city*
region	*country*
curve	*border*
line	*line*
intersection	*intersection*
right	*east*
big	*big*
lie_at	*lie at*
be_in_zone	*be to*

Figure 13
Translation of constants of language **G** into **L**.

Expression of G: δ	**Translation into L:** $\rho_{\mathbf{G\text{-}L}}(\delta)$
$\lambda P[P(d_1)]$, $\lambda P[P(d_2)]$, $\lambda P[P(d_3)]$	*Paris, Frankfurt, Saarbrücken*, respectively
$\lambda P[P(r_1)]$, $\lambda P[P(r_2)]$	*France, Germany*, respectively
$\lambda P[P(c_1)]$	*the border between France and Germany*
$\lambda\mathcal{P}\lambda x\mathcal{P}(\lambda y[x=y])$	*be*
$\lambda P\lambda xP(x)$	*be*
$\lambda P\lambda Q\exists x[P(x) \wedge Q(x)]$	*a*
$\lambda P\lambda Q\exists y[\forall x[P(x) \leftrightarrow x=y] \wedge Q(y)]$	*the*

Figure 14
Translation of some composite expressions of **G** into constants of **L.**

The translation from **G** into **L** is as follows. In rules $\mathrm{T6}_{\text{G-L}}$ to $\mathrm{T8}_{\text{G-L}}$ Q stands for either the quantifier A or THE.

SENTENCES

$\mathrm{T1}_{\text{G-L}}$. If $\alpha \in E_{<<e,\ t>,\ t>}$ and $\beta \in E_{<e,\ t>}$, and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G1}(\alpha,\beta)) = \alpha'\ \beta''$ (the concatenation), where β'' is the result of replacing the first *verb* in β' for its third person singular present form.

Examples: $\rho_{\mathbf{G\text{-}L}}$**(** D_1 (BE_a (A ($\mathrm{OF}_a(\mathrm{R}_1)$ (*dot*)))) **) =** *Paris is a city of France*
$\rho_{\mathbf{G\text{-}L}}$**(** R_3 (*be_in _zone* (THE ($\mathrm{OF}_b(\mathrm{R}_1)$ (*right*)))) **) =** *Germany is to the east of France*
$\rho_{\mathbf{G\text{-}L}}$**(** A(*region*) ($\mathrm{BE}_b$(*big*)) **)** = *a country is big*
$\rho_{\mathbf{G\text{-}L}}$**(** D_3 (*lie_at* (THE ($\mathrm{BETWEEN}_b$ (THE ($\mathrm{BETWEEN}_a$ (R_1) (R_3) (*curve*)))
(A (FROM_TO(D_1) (D_3) (*line*))) (*intersection*)))) **) =**
Saarbrücken lies at the intersection between the border between France and Germany and a line from Paris to Frankfurt

TRANSITIVE VERB PHRASES

$\mathrm{T2}_{\text{G-L}}$. If $\alpha \in E_{<<<e,\, t>,\, t>,\, <e,\, t>>}$ and $\beta \in E_{<<e,\, t>,\, t>}$, and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G1}(\alpha,\beta)) = \alpha'\ \beta'$.

Examples: $\rho_{\mathbf{G\text{-}L}}$**(** BE_a (A (*dot*)) **) =** *be a city*
$\rho_{\mathbf{G\text{-}L}}$**(** *be_in_zone* (THE ($\mathrm{OF}_b(\mathrm{R}_1)$(*right*))) **)** = *be to the east of France*

ATTRIBUTIVE VERB PHRASES

T3$_{G\text{-}L}$. If $\alpha \in E_{<<e,t>,<e,t>>}$ and $\beta \in E_{<e,t>}$, and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G1}(\alpha,\beta)) = \alpha'\ \beta'$.

Example: $\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ BE$_b$(*big*) **)** = *be big*

TERMS

T4$_{G\text{-}L}$. If $\alpha \in E_{<<e,t>,<<e,t>,t>>>}$, $\beta \in E_{<e,t>}$, and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G1}(\alpha,\beta)) = \alpha''\ \beta'$, where α'' is α' except in the case where α' is *a* and the first word in β begins with a vowel; here, α'' is *an*.

Examples: $\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ A (*dot*) **)** = *a city*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ A (OF$_a$(R$_1$)(*dot*)) **)** = *a city of France*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ THE (BETWEEN$_a$ (R$_1$) (R$_2$) (*curve*)) **)** = *the border between France and Germany*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ A (FROM_TO (D$_1$) (D$_3$) (*line*)) **) =** *a line from Paris to Frankfurt*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ THE (OF$_b$(R$_1$)(*right*)) **)** = *the east of France*

COMMON NOUNS

T5$_{G\text{-}L}$. If $\alpha \in E_{<<e,t>,<e,t>>}$ and $\beta \in E_{<e,t>}$, or $\alpha \in E_{<<<<e,t>,t>,e>,<e,t>>}$ and $\beta \in E_{<<<e,t>,t>,e>}$, and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G1}(\alpha,\beta)) = \beta'\ \alpha'$.

Examples: $\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ OF$_a$(R$_1$)(*dot*) **) =** *city of France*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ OF$_b$(R$_1$)(*right*) **)** = *east of France*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ BETWEEN$_a$ (R$_1$) (R$_2$) (*curve*) **) =** *border between France and Germany*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ BETWEEN$_b$ (THE(BETWEEN$_a$(R$_1$)(R$_2$)(*curve*)))
(A(FROM_TO(D$_1$)(D$_3$)(*line*))) (*intersection*) **)** = *intersection between the border between France and Germany and a line from Paris to Frankfurt*

of PREPOSITIONAL PHRASES

T6$_{G\text{-}L}$. If $\alpha \in E_{<<e,t>,t>}$ such that α is either R$_i$ or Q(*region*) and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G2}(\alpha)) = \rho_{\mathbf{G\text{-}L}}(F_{G3}(\alpha)) =$ *of* α'

Examples: $\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ OF$_a$(R$_1$) **)** = *of France*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ OF$_b$(R$_2$) **)** = *of Germany*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ OF$_a$(A(*region*)) **)** = *of a country*

between PREPOSITIONAL PHRASES

T7$_{G\text{-}L}$. If α, $\beta \in E_{<<e,t>,t>}$ such that
(a) α, β are either R$_i$ or Q(*region*) or
(b) α is either C$_i$ or Q(*curve*) and β is either L$_i$ or Q(*line*),
and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G4}(\alpha,\beta))$ = *between* α' *and* β'.

Examples: $\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ BETWEEN$_a$ (R$_1$) (R$_2$) **)** = *between France and Germany*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ BETWEEN$_a$ (R$_1$) (OF$_a$(A(*region*))) **)** = *between France and a country*
$\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ BETWEEN$_b$ (THE (BETWEEN$_a$ (R$_1$) (R$_2$) (*curve*))) (A (FROM_TO(D$_1$) (D$_3$) (*line*))) **) =** *between the border between France and Germany and a line from Paris to Frankfurt*

from-to PREPOSITIONAL PHRASES

T8$_{G\text{-}L}$. If $\alpha,\beta \in E_{<<e,t>,t>}$ such that α,β are either D$_i$ or Q(*dot*) and $\rho_{\mathbf{G\text{-}L}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}L}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}L}}(F_{G5}(\alpha,\beta))$ = *from* α' *to* β'.

Example: $\rho_{\mathbf{G\text{-}L}}\mathbf{(}$ FROM_TO(D$_1$) (D$_3$) **)** = *from Paris to Frankfurt*

As was mentioned above **G** is a very expressive language and not all expressions of **G** can be translated into expressions of **L**, and rules $T1_{G\text{-}L}$ to $T8_{G\text{-}L}$ define the expressions that do have a translation. Instances of expressions that cannot be translated are individual constants (e.g., d_1), equality relations between individuals (e.g., $d_1 = d_2$), and conjunctions or disjunctions (e.g., $dot(d_1) \wedge dot(d_2)$). Other examples are expressions of the form $\lambda P[P(e_1) \vee P(e_2) \vee ... \vee P(e_n)]$, where e_i is an individual constant, which denote the set of properties that one or another individual has. However, this latter kind of expresions could be translated if the expressiveness of **L** were augmented by allowing conjoined term phrases in the grammar.

2.5 Translations between G and P

To conclude with the presentation of the theoretical elements of the system of multimodal representation the translation funtions $\rho_{\mathbf{G\text{-}P}}$ and $\rho_{\mathbf{P\text{-}G}}$ are defined in this section. For each type of **P** there is a corresponding type in **G** and it is defined in terms of the function $f_{P\text{-}G}$ as follows:

1. $f_{P\text{-}G}(dot) = f_{P\text{-}G}(line) = f_{P\text{-}G}(curve) = f_{P\text{-}G}(region) = f_{P\text{-}G}(zone) = f_{P\text{-}G}(composite\_region) = f_{P\text{-}G}(dot\_set) = f_{P\text{-}G}(line\_set) = f_{P\text{-}G}(map) = <e, t>$.
2. For any types a and b, $f_{P\text{-}G}(<a,b>) = <f_{P\text{-}G}(a), f_{P\text{-}G}(b)>$.

2.5.1 Translation from P into G

The translations of the constants of **P** into **G** are presented in Figure 15.

Constant of P: α	**Translation into G:** $\rho_{\mathbf{P\text{-}G}}(\alpha)$
$d_1, d_2, d_3, ...$	$\lambda P[P(d_1)], \lambda P[P(d_2)], \lambda P[P(d_3)], ...$
$l_1, l_2, l_3 ...$	$\lambda P[P(l_1)], \lambda P[P(l_2)], \lambda P[P(l_3)],...$
$c_1, c_2, c_3, ...$	$\lambda P[P(c_1)], \lambda P[P(c_2)], \lambda P[P(c_3)], ...$
$r_1, r_2, r_3, ...$	$\lambda P[P(r_1)], \lambda P[P(r_2)], \lambda P[P(r_3)], ...$
$z_1, z_2, z_3, ...$	$\lambda P[P(z_1)], \lambda P[P(z_2)], \lambda P[P(z_3)], ...$

Figure 15
Translation of constants of language **P** into **G**.

In the following definitions Q stands for either the quantifier A or THE. The translation rules are as follows:

CONSTANT[13]

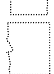
$T1_{P\text{-}G}$. If $\alpha \in C_s$ where $s \in \{dot, line, curve, region, zone\}$ then
(a) $\rho_{\mathbf{P\text{-}G}}(\alpha)$ is as shown in Figure 15.
(b) $\rho_{\mathbf{P\text{-}G}}(\alpha)$=Q($s$).

Examples:

$\rho_{\mathbf{P\text{-}G}}(\bullet) = \lambda P[P(d_1)]$

$\rho_{\mathbf{P\text{-}G}}(/) = \lambda P[P(l_1)]$

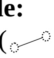
$\rho_{\mathbf{P\text{-}G}}($  $) = \lambda P[P(r_1)]$

$\rho_{\mathbf{P\text{-}G}}($  $) =$ A($region$)

LINE

$T2_{P\text{-}G}$. If $\alpha,\beta \in E_{dot}$, and $\rho_{\mathbf{P\text{-}G}}(\alpha)=\alpha$' and $\rho_{\mathbf{P\text{-}G}}(\beta)=\beta$' then $\rho_{\mathbf{P\text{-}G}}(F_{P1}(\alpha,\beta))$ = Q(FROM_TO(α')(β')($line$)).

Example:

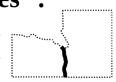
$\rho_{\mathbf{P\text{-}G}}($$)$ = A(FROM_TO ($\lambda P[P(d_1)]$) ($\lambda P[P(d_3)]$) ($line$))

[13] Rule (b) allows that the "concrete extension" of a graphical object in **P** is represented as its corresponding denoting concept in **G**.

CURVE

T3$_{\text{P-G}}$. If $\alpha,\beta \in E_{region}$ such that α and β are adjacent, and $\rho_{\textbf{P-G}}(\alpha)=\alpha'$ and $\rho_{\textbf{P-G}}(\beta)=\beta'$ then $\rho_{\textbf{P-G}}(F_{P2}(\alpha,\beta)) = \text{Q}(\text{BETWEEN}_{\text{a}}(\alpha')(\beta')(\textit{curve}))$.

Examples[14]**:**

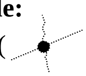
$\rho_{\textbf{P-G}}($  $) = \text{THE}(\text{BETWEEN}_{\text{a}}\ (\lambda P[P(r_1)])\ (\lambda P[P(r_2)])\ (\textit{curve}))$

$\rho_{\textbf{P-G}}($  $) = \text{THE}(\text{BETWEEN}_{\text{a}}\ (\lambda P[P(r_1)])\ (\text{A}\ (\textit{region}))\ (\textit{curve}))$

INTERSECTION

T4$_{\text{P-G}}$. If $\alpha \in E_{curve}$ and $\beta \in E_{line}$, and $\rho_{\textbf{P-G}}(\alpha)=\alpha'$ and $\rho_{\textbf{P-G}}(\beta)=\beta'$ then $\rho_{\textbf{P-G}}(F_{P3}(\alpha,\beta)) = \text{Q}(\text{BETWEEN}_{\text{b}}(\alpha')(\beta')(\textit{intersection}))$.

Example:

$\rho_{\textbf{P-G}}($  $) = \text{THE}(\text{BETWEEN}_{\text{b}}\ (\text{THE}(\text{BETWEEN}_{\text{a}}\ (\lambda P[P(r_1)])\ (\lambda P[P(r_2)])\ (\textit{curve})))$
$(\text{A}\ (\text{FROM\_TO}\ (\lambda P[P(d_1)])\ (\ \lambda P[P(d_3)])\ (\textit{line})))$
$(\textit{intersection}))$

RIGHT

T5$_{\text{P-G}}$. If $\alpha \in E_{region}$ and $\rho_{\textbf{P-G}}(\alpha)=\alpha'$ then $\rho_{\textbf{P-G}}(F_{P4}(\alpha)) = \text{Q}(\text{OF}_{\text{b}}(\textit{right})(\alpha'))$.

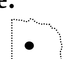
Example:

$\rho_{\textbf{P-G}}($  $) = \text{THE}(\text{OF}_{\text{b}}\ (\textit{right})\ (\lambda P[P(r_3)]))$

DOT INSIDE A REGION

T6$_{\text{P-G}}$. If $\alpha \in E_{region}$ and $\rho_{\textbf{P-G}}(\alpha)=\alpha'$ then $\rho_{\textbf{P-G}}(F_{P5}(\alpha)) = \text{Q}(\text{OF}_{\text{a}}(\alpha')(\textit{dot}))$.

Example:

$\rho_{\textbf{P-G}}($  $) = \text{A}(\text{OF}_{\text{a}}\ (\lambda P[P(r_1)])\ (\textit{dot}))$

COMPOSITE REGION (1)[15]

T7$_{\text{P-G}}$. If $\alpha,\beta \in C_{region}$ such that α and β are adjacent, and $\rho_{\textbf{P-G}}(\alpha)=\lambda P[\alpha']$ and $\rho_{\textbf{P-G}}(\beta)=\lambda P[\beta']$ then $\rho_{\textbf{P-G}}(F_{P6}(\alpha,\beta)) = \lambda P[\alpha' \vee \beta']$.

COMPOSITE REGION (2)

T8$_{\text{P-G}}$. If $\alpha \in C_{region}$ and $\beta \in E_{composite\_region}$ such that α and β are adjacent, and $\rho_{\textbf{P-G}}(\alpha)=\lambda P[\alpha']$ and $\rho_{\textbf{P-G}}(\beta)=\lambda P[\beta']$ then $\rho_{\textbf{P-G}}(F_{P6}(\alpha,\beta)) = \lambda P[\alpha' \vee \beta']$.

SET OF DOTS

T9$_{\text{P-G}}$. (a) If $\alpha \in E_{dot\_set} = \varnothing$ and $\beta \in C_{dot}$, and $\rho_{\textbf{P-G}}(\beta)=\lambda P[\beta']$ then $\rho_{\textbf{P-G}}(F_{P6}(\alpha,\beta)) = \lambda P[\beta']$.
(b) If $\alpha \in E_{dot\_set} \neq \varnothing$ and $\beta \in C_{dot}$, and $\rho_{\textbf{P-G}}(\alpha)=\lambda P[\alpha']$ and $\rho_{\textbf{P-G}}(\beta)=\lambda P[\beta']$ then $\rho_{\textbf{P-G}}(F_{P6}(\alpha,\beta)) = \lambda P[\alpha' \vee \beta']$.

[14] These two example expressions correspond to the abbreviated expressions in examples 2 and 3, respectively, presented in Section 2.3.2.

[15] Examples of the application of the rules T7$_{\text{P-G}}$ to T11$_{\text{P-G}}$ are included in the translation of a map shown in Figure 16, as explained below.

SET OF LINES

T10$_{P\text{-}G}$. (a) If $\alpha \in E_{line\_set} = \varnothing$ and $\beta \in C_{line}$, and $\rho_{\mathbf{P\text{-}G}}(\beta) = \lambda P[\beta']$ then $\rho_{\mathbf{P\text{-}G}}(F_{P6}(\alpha,\beta)) = \lambda P[\beta']$.

(b) If $\alpha \in E_{line\_set} \neq \varnothing$ and $\beta \in C_{line}$, and $\rho_{\mathbf{P\text{-}G}}(\alpha) = \lambda P[\alpha']$ and $\rho_{\mathbf{P\text{-}G}}(\beta) = \lambda P[\beta']$ then $\rho_{\mathbf{P\text{-}G}}(F_{P6}(\alpha,\beta)) = \lambda P[\alpha' \vee \beta']$.

MAP

T11$_{P\text{-}G}$. If $\alpha \in E_{composite\_region}$, $\beta \in E_{dot\_set}$ and $\delta \in E_{line\_set}$, and $\rho_{\mathbf{P\text{-}G}}(\alpha) = \lambda P[\alpha']$, $\rho_{\mathbf{P\text{-}G}}(\beta) = \lambda P[\beta']$, $\rho_{\mathbf{P\text{-}G}}(\delta) = \lambda P[\delta']$, then $\rho_{\mathbf{P\text{-}G}}(F_{P7}(\alpha,\beta,\delta)) = \lambda P[\alpha' \vee \beta' \vee \delta']$.

As an instance of the translation of a map from **P** into **G** by rule T11$_{P\text{-}G}$ consider Figure 16. As can be seen a map is interpreted in **G** as the set of properties that one or another graphical object in the base of the map has.

For the interpretation of the picture in Figure 4 it is not required to compute and translate all possible syntactic structures that can be generated in **P** on the basis of the overt graphical symbols of the drawing. The translation rules permit to map a large number of syntactic structures into **G** and they can be used as required. However, for the interpretation of a map we will only translate a designated expression ζ of type *map* that results of parsing a full drawing in terms of the graphical objects in the base. ζ will be called "the map".This criterion ensures that the drawing belongs to the map modality **P**. In addition, the graphical terms in the disjuntion of the body of expressions of type *map* are used in **G** to define the interpretation domain $\mathbf{P_{base}}$. When this set is defined the semantic rules to interpret expressions of **G** can be evaluated.

2.5.2 Translation from G into P

As was mentioned in Section 1 in relation to the scheme in Figure 3 the purpose of this translation is to draw the graphical symbols that are referred to in **G**. To picture the full map the only symbols that have to be drawn are the symbols of the base ($\mathbf{P_{base}}$) as emerging symbols do not have an independent pictorial realization. According to this the only translations that have to be defined are the translations of the symbols contained in the expression ζ (i.e., the map). We also have to consider that graphical terms occuring in expressions of **G** can have a graphical realization, which may be required for specific purposes. For instance, if one needs to highlight the region to the east of France the term of **G** denoting that region should be translated and depicted in **P**. In the definition of the rules below Q stands for either the quantifier A or THE.

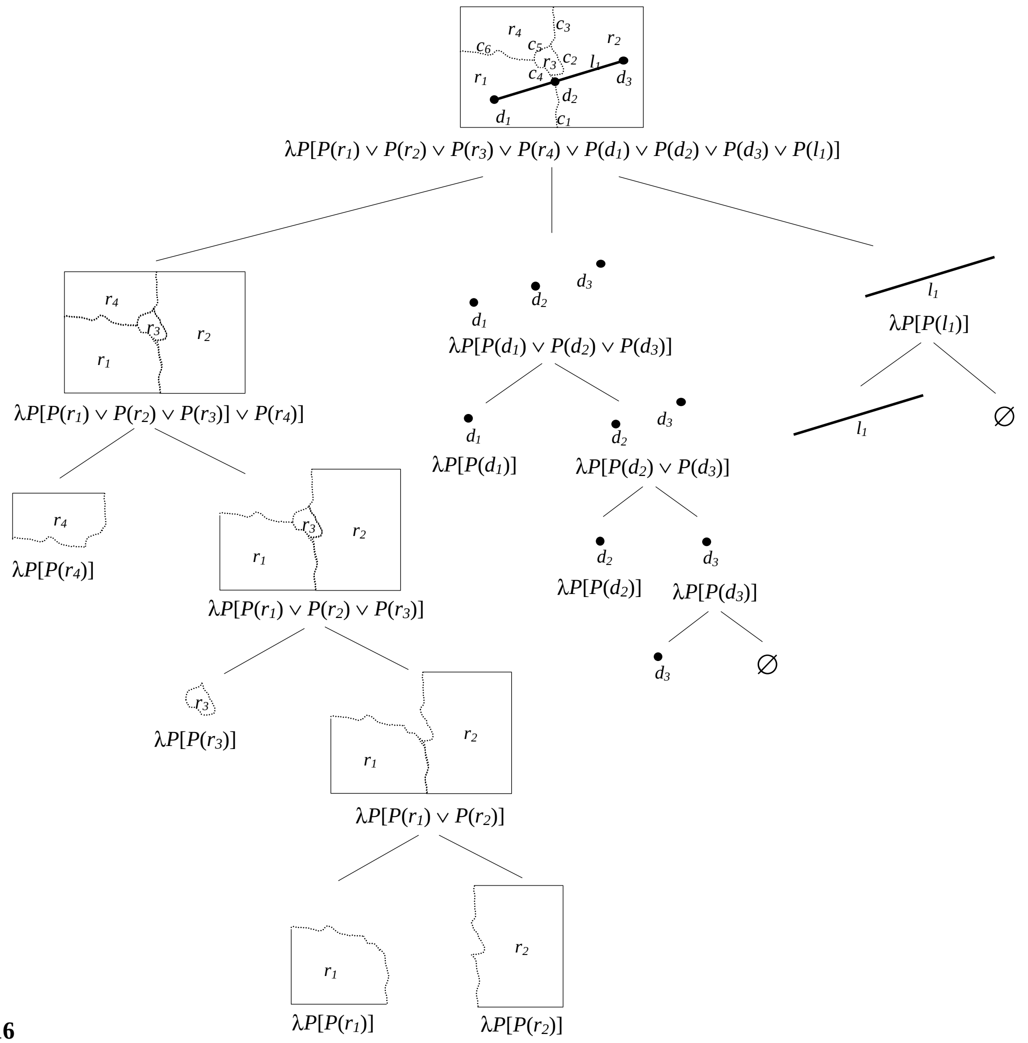


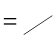
Figure 16
Translation into **G** of a map.

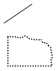
CONSTANT[16]

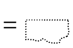
T1$_{\text{G-P}}$. (a) If $\alpha = \lambda P[P(\alpha^*)]$ and $\alpha^* \in C_e$ then $\rho_{\mathbf{G\text{-}P}}(\alpha)$ is the drawing of α^*.
(b) If $\alpha = Q(s)$ where $s \in \{dot, line, curve, region, zone\}$ then $\rho_{\mathbf{G\text{-}P}}(\alpha)$ is the drawing of whatever graphical object in C_s.

Examples:

$\rho_{\mathbf{G\text{-}P}}(\lambda P[P(d_1)]) = \bullet$

$\rho_{\mathbf{G\text{-}P}}(\lambda P[P(l_1)]) =$ 

$\rho_{\mathbf{G\text{-}P}}(\lambda P[P(r_1)]) =$ 

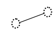
$\rho_{\mathbf{G\text{-}P}}(\mathrm{A}(region)) =$ 

[16] Rule (b) allows that a graphical denoting concept in **G** is represented in **P** as its “concrete extension”.

LINE

T2$_{\text{G-P}}$. If $\alpha,\beta \in E_{<<e,\ t>,\ t>}$ such that α,β are either D_i or $Q(\mathit{dot})$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$ and $\rho_{\mathbf{G\text{-}P}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}P}}(Q(\mathrm{FROM\_TO}(\alpha)(\beta)(\mathit{line}))) = F_{P1}(\alpha',\beta')$.

Example:

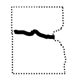
$\rho_{\mathbf{G\text{-}P}}(\ \mathrm{A}(\mathrm{FROM\_TO}\ (\lambda P[P(d_1)])\ (\lambda P[P(d_3)])\ (\mathit{line}))\ ) =$ 

CURVE

T3$_{\text{G-P}}$. If $\alpha,\beta \in E_{<<e,\ t>,\ t>}$ such that α,β are either R_i or $Q(\mathit{region})$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$ and $\rho_{\mathbf{G\text{-}P}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}P}}(Q(\mathrm{BETWEEN}_a(\alpha)(\beta)(\mathit{curve}))) = F_{P2}(\alpha',\beta')$.

Examples[17]**:**

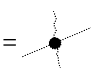
$\rho_{\mathbf{G\text{-}P}}(\ \mathrm{THE}(\mathrm{BETWEEN}_a\ (\lambda P[P(r_1)])\ (\lambda P[P(r_2)])\ (\mathit{curve}))\ ) =$ 

$\rho_{\mathbf{G\text{-}P}}(\ \mathrm{THE}(\mathrm{BETWEEN}_a\ (\lambda P[P(r_1)])\ (\mathrm{A}\ (\mathit{region}))\ (\mathit{curve}))\ ) =$ 

INTERSECTION

T4$_{\text{G-P}}$. If $\alpha,\beta \in E_{<<e,\ t>,\ t>}$ such that α is either C_i or $Q(\mathit{curve})$ and β is either L_i or $Q(\mathrm{line})$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$ and $\rho_{\mathbf{G\text{-}P}}(\beta)=\beta'$ then $\rho_{\mathbf{G\text{-}P}}(Q(\mathrm{BETWEEN}_b(\alpha)(\beta)(\mathit{intersection}))) = F_{P3}(\alpha',\beta')$.

Example:

$\rho_{\mathbf{G\text{-}P}}(\ \mathrm{THE}(\mathrm{BETWEEN}_b\ (\mathrm{THE}(\mathrm{BETWEEN}_a\ (\lambda P[P(r_1)])\ (\lambda P[P(r_2)])\ (\mathit{curve})))$
$(\mathrm{A}\ (\mathrm{FROM\_TO}\ (\lambda P[P(d_1)])\ (\ \lambda P[P(d_3)])\ (\mathit{line})))\ \ (\mathit{intersection}))\ ) =$ 

RIGHT

T5$_{\text{G-P}}$. If $\alpha \in E_{<<e,\ t>,\ t>}$ such that α is either R_i or $Q(\mathit{region})$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$ then $\rho_{\mathbf{G\text{-}P}}(Q(\mathrm{OF}_b(\mathit{right})(\alpha))) = F_{P4}(\alpha')$.

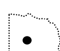
Example:

$\rho_{\mathbf{G\text{-}P}}(\ \mathrm{THE}(\mathrm{OF}_b\ (\mathit{right})\ (\lambda P[P(r_3)]))\ ) =$ 

DOT INSIDE A REGION

T6$_{\text{G-P}}$. If $\alpha \in E_{<<e,\ t>,\ t>}$ such that α is either R_i or $Q(\mathit{region})$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$ then $\rho_{\mathbf{G\text{-}P}}(Q(\mathrm{OF}_a(\alpha)(\mathit{dot}))) = F_{P5}(\alpha')$.

Example:

$\rho_{\mathbf{G\text{-}P}}(\ \mathrm{A}(\mathrm{OF}_a\ (\lambda P[P(r_1)])\ (\mathit{dot}))\ ) =$ 

COMPOSITE REGION (1)[18]

T7$_{\text{G-P}}$. If $\alpha=\lambda P[P(\alpha^*)]$ and $\beta=\lambda P[P(\beta^*)]$ such that $\mathit{region}(\alpha^*)$ and $\mathit{region}(\beta^*)$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}P}}(\beta)=\beta'$, and α' and β' are adjacent then $\rho_{\mathbf{G\text{-}P}}(\lambda P[P(\alpha^*) \vee P(\beta^*)]) = F_{P6}(\alpha',\beta')$.

COMPOSITE REGION (2)

T8$_{\text{G-P}}$. If $\alpha=\lambda P[P(\alpha^*)]$ and $\beta=\lambda P[\beta'']=\lambda P[P(\beta_1)\vee P(\beta_2)\vee ... \vee P(\beta_n)]$ such that $\mathit{region}(\alpha^*)$ and $\mathit{region}(\beta_i)$, and $\rho_{\mathbf{G\text{-}P}}(\alpha)=\alpha'$, $\rho_{\mathbf{G\text{-}P}}(\beta)=\beta'$, and α' and β' are adjacent then $\rho_{\mathbf{G\text{-}P}}(\lambda P[P(\alpha^*) \vee \beta''] = F_{P6}(\alpha',\beta')$.

[17] These two example expressions correspond to the abbreviated expressions in examples 2 and 3, respectively, presented in Section 2.3.2.

[18] Examples for rules T6$_{\text{G-P}}$ to T11$_{\text{G-P}}$ are included in Figure 16 above.

SET OF DOTS

T9$_{\text{G-P}}$. (a) If $\beta=\lambda P[P(\beta^*)]$ such that $dot(\beta^*)$ then $\rho_{\mathbf{G-P}}(\beta) = F_{P6}(\emptyset,\beta)$.
(b) If $\alpha = \lambda P[\alpha''] = \lambda P[P(\alpha_1)\vee P(\alpha_2)\vee...\vee P(\alpha_n)]$ and $\beta=\lambda P[P(\beta^*)]$ such that $dot(\alpha_i)$ and $dot(\beta^*)$, and $\rho_{\mathbf{G-P}}(\alpha)=\alpha'$ and $\rho_{\mathbf{G-P}}(\beta)=\beta'$ then $\rho_{\mathbf{G-P}}(\lambda P[\alpha'' \vee P(\beta^*)]) = F_{P6}(\alpha',\beta')$

SET OF LINES

T10$_{\text{G-P}}$. (a) If $\beta=\lambda P[P(\beta^*)]$ such that $line(\beta^*)$ then $\rho_{\mathbf{G-P}}(\beta) = F_{P6}(\emptyset,\beta)$.
(b) If $\alpha = \lambda P[\alpha''] = \lambda P[P(\alpha_1)\vee P(\alpha_2)\vee...\vee P(\alpha_n)]$ and $\beta=\lambda P[P(\beta^*)]$ such that $line(\alpha_i)$ and $line(\beta^*)$, and $\rho_{\mathbf{G-P}}(\alpha)=\alpha'$ and $\rho_{\mathbf{G-P}}(\beta)=\beta'$ then $\rho_{\mathbf{G-P}}(\lambda P[\alpha'' \vee P(\beta^*)]) = F_{P6}(\alpha',\beta')$

MAP

T11$_{\text{G-P}}$. If $\alpha\in E_{<<e,t>,t>} = \lambda P[\alpha''] = \lambda P[P(\alpha_1)\vee P(\alpha_2)\vee...\vee P(\alpha_m)]$, $\beta\in E_{<<e,t>,t>} = \lambda P[\beta''] = \lambda P[P(\beta_1)\vee P(\beta_2)\vee...\vee P(\beta_n)]$ and $\delta\in E_{<<e,t>,t>} = \lambda P[\delta''] = \lambda P[P(\delta_1)\vee P(\delta_2)\vee...\vee P(\delta_r)]$ such that $region(\alpha_i)$, $dot(\beta_i)$ and $line(\delta_i)$, and $\rho_{\mathbf{G-P}}(\alpha)=\alpha'$, $\rho_{\mathbf{G-P}}(\beta)=\beta'$ and $\rho_{\mathbf{G-P}}(\delta)=\delta'$ then $\rho_{\mathbf{G-P}}(\lambda P[\alpha'' \vee \beta'' \vee \delta'']) = F_{P7}(\alpha', \beta', \delta')$.

With this, the specification of the system of multimodal representation in Figure 3 is concluded. In this system it is possible to express natural language and graphical information about maps and translate expressions between both of the modalities. Natural language can be seen as stating or imposing an interpretation upon graphical representations, making the graphics meaningful. Alternatively, graphics can be seen as representing knowledge in an effective fashion. Expressions of the languages **L** and **P** can be translated through the interface language **G** in which both the semantics of **L** and the geometrical structure of **G** can be represented and reasoned about in an integrated fashion.

The system provides a solid semantic ground to state and resolve problems of reference in multimodal scenarios. The syntactic and semantic structures of the three languages permit to express and interpret information in each of the modalities and to find correlated expressions in a different modality with the same semantic values in a systematic fashion. As a consequence it is possible to state formally what does it mean to resolve a multimodal reference: according to this theory to resolve a multimodal reference is to find the semantic value of an expression in terms either of the information expressed in the modality or in terms of information expressed in other modalities with the help of the translation functions. In a fully interpreted multimodal system as the one illustrated in this section interpreting a multimodal message is a matter of evaluating the multimodal expression. However, as was argued in Section 1, the relation between individual constants input through different modalities have to be established before multimodal expressions can be evaluated. To establish this relationship is the crucial part of the interpretation process and how can this be done is illustrated in Section 3.

3 Resolution of Deictic Inference by Constraint Satisfaction

In the theory developed in Section 2 it was assumed that the translation of constants of all categories from **L** into **G** and vice versa were available, and then multimodal interpretation could be carried out; however, in the interpretation of multimodal messages, natural language and graphics are input from different sources, and working out the meaning of a multimodal message is by no means trivial. As was discussed in Section 1, resolving the references and inducing the translation between graphical and linguistic terms can be thought of as the same problem. Consider, for instance, the situation of reading a book with words and pictures: when the associations between textual and graphical symbols is realized by the reader, the message as a whole has been properly understood. However, it cannot be expected that such an association can be known beforehand.

The process of inducing the translation functions for constants of **G** and **L** is similar to the computer vision problem of interpreting drawings. A related antecedent is the work on the logic of depiction (Reiter and Mackworth, 1987) in which a logic for the interpretation of maps, to be applied to computer vision and intelligent graphics, is developed. It is argued that any adequate representation scheme for visual (and computer graphics) knowledge must mantain the distinction between knowledge of the image (the geometry) and knowledge of the scene (its linguistic interpretation), and about the depiction relation. In Reiter's system two sets of first order logical sentences representing the scene and the image are employed, and express, respectively, the conceptual and geometrical knowledge about hand drawn sketch maps of geographical regions. In the view adopted here, the depiction relation corresponds to the translation function between constants of **L** and **G** as discussed above. An interpretation in Reiter's system is defined as a model, in the logical sense, of both sets of sentences and the depiction relation, and interpreting a drawing consists in finding out all possible models of such sets of sentences. The domain for these models is determined by the set of individuals in the image and the scene of the picture that is being interpreted. Although computing the set of models of a set of first order logical formulae is a very hard computational problem, the entities constituting a drawing conform, normally, a finite set which is often small. So, the possibility of computing the set of models of a drawing is a matter for empirical research. In particular, Reiter's system employs a constraint satisfaction algorithm to find out all possible interpretation of maps, and the output of his system is a set of labels for curves or chains as rivers, roads or shores, and for areas as land regions or water regions. As was mentioned above, to find the translation functions between **G** and **L** is a similar problem with the same kind of complexity. In Section 3.1 a constraint satisfaction algorithm for the induction of the translation into **G** of individual constants of **L** mentioned explicitly in the text of a multimodal message is presented. In this section it is also shown how composite terms of **L** can be translated into their corresponding graphical expressions of **G** (and subsequently of **P**).

A second consideration in this section is that working out the translation between graphical and linguistic individual constants suggests a method for generating natural language expressions that refer to graphical objects and configurations. Note that inducing the linguistic translation of a graphical term which has not been mentioned overtly in the textual part of a multimodal message is the same as generating a linguistic description for the object denoted by the corresponding graphical term: once one knows the translation between individual constants of both of the modalities the generation of multimodal descriptions can be achieved through the translation rules; for instance, in the map of Figure 4, if one points to the curve c_1 once the translation of individual constants has been found, the expression *the border between France and Germany* can be generated. This strategy to produce natural language descriptions is further discussed below in Section 3.2.

3.1 Resolution of Spatial Deixis by Constraint Satisfaction

As a preliminary consideration for the interpretation of multimodal messages like Figures 1 and 2 it is worth mentioning that from the point of view of our system what is given are expressions of **L** and expressions of **P** and what has to be worked out is the composition $\rho_{\mathbf{L\text{-}G}}{}^\circ\rho_{\mathbf{G\text{-}P}}$ and the reciprocal function $\rho_{\mathbf{P\text{-}G}}{}^\circ\rho_{\mathbf{G\text{-}L}}$. However, note that the expressions of **P** are the graphical symbols on the drawings and parsing a drawing (expression of type *map*) produces a syntactic structure of **P** whose translation into **G** is the expression ζ (which we called the map). Emergent objects can also be represented in **G** as long as they can be produced out of the base through syntactic rules of **P** and their translations into **G**. Consequently, expressions of **G** which refer to graphical objects stand in a one to one relation with the corresponding objects in **P**. Although theoretically expressions in **G** and **P** are different representational objects in actual interpretation processes they always come packed together. The relation between expressions of **G** and **L**, on the other hand, has to be worked out. For this purpose we present an algorithm to establish a relationship between the individual constants of **L** and the graphical constants included in the expression ζ (the map), which correspond to the interpretation domain $\mathbf{P_{base}}$. The algorithm for computing the translation function assigns a graphical constant to all proper names overtly mentioned in the linguistic part of a multimodal message (e.g., the graphical symbols d_1, d_2, d_3, r_1, r_2 to the linguistic constants *Paris*, *Saarbrücken*, *Frankfurt*, *France*, *Germany*, respectively). The set of proper names appearing in a particular multimodal message will be referred to as *Names.* As the translation for linguistic constants of other types are given beforhand, once the translations for proper names are available, it is possible to find the graphical symbols and configurations that correfer with composite natural language descriptions through the translation rules between **L**, **G** and **P**; for instance, once the regions representing France and Germany have been identified the term *the border between France and Germany* can be translated into an expression of **G**, which denotes the corresponding curve, and also into the drawing of the curve in **P** which denotes the border between France and Germany itself. Here, it is important to highlight that the translation for individual constants cannot normally be found with the overt information expressed through the multimodal message only. For working out the interpretation of Figure 2, for instance, it is required, in addition to the text and graphics, knowledge about the geography of Europe and also knowledge about interpretation conventions of maps.

For the definition of the algorithm a table for representing the set of possible functions from linguistics predicates (e.g., *city*, *country*, etc.) to their corresponding graphical types (e.g, *dot*, *region*, etc.) is defined. This table will be referred to as *function table*. For each particular interpretation task a set of appropriate function tables is defined according to the following rule:

For each $\delta \in C_{CN}$ of **L** and $\delta' \in C_{<e,t>}$ of **G** such that $\rho_{\mathbf{L\text{-}G}}(\delta)=\delta'$ create a function table $(X_\delta, Y_{\delta'})$ such that:

$X_\delta = \{x \in C_T \mid [[x \text{ is a } \delta]]^M \textit{ is true} \text{ and } x \in \textit{Names}\}$

$Y_{\delta'} = \{y \in C_e \mid [[\delta'(y)]]^M \textit{ is true}\}$

where X_δ and Y_δ are not empty. In case either of these two sets is empty no function table for the corresponding pair is defined.

The function tables for our example are illustrated in Figure 17.

d_3			
d_2			
d_1			
	Paris	*Saarbrücken*	*Frankfurt*

(X_{city}, Y_{dot})

r_4		
r_3		
r_2		
r_1		
	France	*Germany*

$(X_{country}, Y_{region})$

Figure 17
Funtion tables for the message in Figure 2.

Note that if only one cell of each column of a function table is filled up a function from proper names to graphical constants is defined. Furthermore, if the result of this process is a table in which only one cell of each row is also marked, the function is one to one. According to this, if there are n names and m graphical objects, the first column of a function table can be filled up in m differents ways, the second in m–1 different ways, etc., until n graphical objects have been assigned. As a consequence, each function table with n names and m graphical objects defines $m!/(m-n)!$ possible translation functions[19]. For the example, (X_{city}, Y_{dot}) and $(X_{country}, Y_{region})$ define six and twelve possible functions, respectively. Let T_δ be the set of possible translation functions for the function table $(X_\delta, Y_{\delta'})$, and let Γ be the cross product of all T_δ in an interpretation context (i.e., the set of possible translation models). For our example, $\Gamma = T_{city} \times T_{country}$, where $|\Gamma| = 72$. This set contains 72 ordered pairs of functions, and each one represents a possible translation model for the multimodal message. Translation models can be enumerated by assigning a natural number to every cell in the array Γ. We give the following enumeration for bidimensional translation models: let $\gamma_n=(f_i, g_j)$ be the nth translation model in $\Gamma = T_x \times T_y$ where $0<n<|\Gamma|$ and $f_i \in T_x$, $g_j \in T_y$. For every n, if mod $|T_x| \neq 0$ then $i = (n\text{-}1 \bmod |T_x|)+1$ and $j = (n\text{-}1 \operatorname{div} |T_x|)+1$. Similar expressions can be defined for higher dimensions.

To enumerate the set of possible functions from n names to m graphical objects we use the following procedure: Let N be a list of names and O a list of graphical objects, and let *F-INDEX* be a n digits string containing the n digits of a numeral in base m. Every string in *F-INDEX* codifies a total function in which the j graphical object m_j in O (where $0 \leq j < m$) is assigned to the n_i name in N by the rule *F-INDEX*(i) = j. The set of possible entries in *F-INDEX* codes the m^n possible functions from n names to m graphical objects. One to one functions are those in which no m_j occurs more than once in a given value of *F-INDEX*. The functions sought are the $m!/(m-n)!$ one to one functions that result from enumerating in base m all possible values for *F-INDEX* from 0 to m^n–1, and filtering out all numbers in which the same digit occurs more than once, in the enumeration order. Consider the graphical illustration of the *F-INDEX* scheme for identifying the functions corresponding to a function table with three names and four graphical objects in Figure 18. This function table has 4^3=64 possible functions out of which 4!/(4-3)! = 24 are one to one. The graphical object m_j in O is associated to the name n_i in N by marking the corresponding cell in the table, where j is placed in the corresponding cell of *F-INDEX*. The string in *F-INDEX* is the numeral 000 in base 4 and represents the function in which the graphical object o_0 is assigned to all three names.

[19] In general, if graphical objects can receive more that one name —e.g., as it is the case in the multimodal interpretation scenarios related to the Hyperproof system (Barwise et al., 1994)— the number of possible translation functions will be m^n, where m is the number of graphical objects and n is the number of names.

O			
o_3			
o_2			
o_1			
o_0	X	X	X
	n_1	n_2	n_3 *N*

F-Index	0	0	0

Figure 18
Set of functions associated to a function table.

In Figure 19 some examples of the enumeration of functions are shown. The first table shows function 12 (base 4) which is the smaller index for a total function; the second table shows function 123 which associates names n_1, n_2 and n_3 to the graphical objects o_1, o_2 and o_3, respectively; the third table illustrates the function 321 which is the largest index for a total function in the set; and finally, the fourth table illustrates the function 333 which is a constant function assigning the object o_3 to all three names.

O			
o_3			
o_2			X
o_1		X	
o_0	X		
	n_1	n_2	n_3 *N*
F-Index	0	1	2

O			
o_3			X
o_2		X	
o_1	X		
o_0			
	n_1	n_2	n_3 *N*
F-Index	1	2	3

O			
o_3	X		
o_2		X	
o_1			X
o_0			
	n_1	n_2	n_3 *N*
F-Index	3	2	1

O			
o_3	X	X	X
o_2			
o_1			
o_0			
	n_1	n_2	n_3 *N*
F-Index	3	3	3

Figure 19
Examples of function index.

Armed with these concepts we can define an algorithm for working out the interpretation of a multimodal message, as follows: Let $message_{\mathbf{L}}$ be a sentence of **L** (the textual part of the multimodal message), $\vartheta_{\mathbf{G}}$ an empty set of expressions of **G,** and Γ the set of possible translation models for $message_{\mathbf{L}}$. Then, for each $\gamma_i \in \Gamma$ assume that γ_i is a translation model for $message_{\mathbf{L}}$ and include its translation $message_{\mathbf{G}}$ under γ_i in $\vartheta_{\mathbf{G}}$ —i.e., $\rho_{\mathbf{L\text{-}G}}(message_{\mathbf{L}}) = message_{\mathbf{G}}$. If the semantic value of all expressions $\vartheta_{\mathbf{G}}$ in relation to the geometrical domain $\mathbf{P_{base}}$ is true, then γ_i is a translation model for $message_{\mathbf{L}}$; otherwise, exclude γ_i from Γ. Once all translation models have been tested check whether there is only one γ_j in Γ. If that is the case, such γ_j is the translation function; otherwise, select a new appropriate expression of **L** (a common sense knowledge constraint) and include its translation into **G** in $\vartheta_{\mathbf{G}}$, and repeat the process until there is only one γ_j in Γ.

For our example four translations out of the 72 γ's in Γ will come out true for the first cycle of the algorithm in which the multimodal message is used as the only constraint (example 5 in Section 2.3), as shown in Figure 20.

To continue with the algorithm it is required to consider some knowledge about the goegraphy of Europe. For our problem the relevant constraints for interpreting the message are illustrated in Figure 21.

As can be seen, the idea of the algorithm is simply to take constraints one at a time and produce the interpretation of the message incrementally. Considering constraint 1 in Figure 21, the translation functions (2) and (4) in Figure 20 can be removed; the translation function (3), in turn, can be ruled out either through constraints 2 or 3 (the interpretation of the translation of constraint 1 into **G** is shown in example 4 in Section 2.3). For the example, only three cycles of the algorithm are required to rule out all but the correct translation model in Γ, which is the translation function (1) in Figure 20.

(1)

	Paris	*Saarbrücken*	*Frankfurt*
d_3			x
d_2		x	
d_1	x		

	France	*Germany*
r_4		
r_3		
r_2		x
r_1	x	

(2)

	Paris	*Saarbrücken*	*Frankfurt*
d_3			x
d_2		x	
d_1	x		

	France	*Germany*
r_4		
r_3		
r_2	x	
r_1		x

(3)

	Paris	*Saarbrücken*	*Frankfurt*
d_3	x		
d_2		x	
d_1			x

	France	*Germany*
r_4		
r_3		
r_2		x
r_1	x	

(4)

	Paris	*Saarbrücken*	*Frankfurt*
d_3			x
d_2		x	
d_1	x		

	France	*Germany*
r_4		
r_3		
r_2	x	
r_1		x

Figure 20
Possible translation models without additional constraints.

1.	*Germany is to the east of France*
2.	*Paris is a city of France*
3.	*Frankfurt is a city of Germany*

Figure 21
General knowledge of geography.

This concludes the presentation of the procedure for interpreting proper names deictically in relation to a graphical context. Although only the interpretation of this kind of constants was required for our example, the interpretation of other kinds of terms, like pronouns, can be carried out in the present framework. Consider that to be able to cope with multimodal messages in which pronouns were included in the textual part, like in Figure 1, a more general definition of the language **L** would be required, but in such an extension both proper names and pronouns would be constants of category *T* in the grammar. In the present framework pronouns would be interpreted along the lines of proper names. For the definition of functions tables, each pronoun present in a multimodal text would be included in the set *Names*, and as a first approximation, it would be a member of the domain of all function tables; different instances of the same pronoun would be considered two different objects in the interpretation process (e.g., he_o, he_1, ..., etc.), and the interpretation would be worked out as was shown above in this section. It is also possible to think of a situation in which there are two or more graphical objects with the same name; in this kinds of context, proper names would be considered as kinds of pronuons, and from the point of view of **L**, a different subscripted constant of category *T* ($name_0$, $name_1$,...) would be assigned to each such graphical objects. To differentiate these objects alternative definite description could be obtained through the translation from constants of **P** into expressions of **G**, as will be argued in Section 3.2, and such descriptions could be used in the context of the particular rhetorical structures and communicative purposes of multimodal messages. A further consideration is that not only the constants of category *T* in a grammar can be used deictically; definite and indefinite descriptions can also be interpreted in this way. Consider that the textual part of the multimodal message in Figure 1 could have been either *John washed it*, *the man washed it* or even *a man washed it* and all three terms *John*, *the man* and *a man* would have to be interpreted deictically in relation to the graphical context. To be able to deal with this latter situation, descriptions can be interpreted deictically in our approach if terms of this kind are also included in the set *Names* for the construction of function tables. More generally, our interpretation procedure defines a function from terms into individuals of the world through the graphical context. This is so because although function tables define translation models between linguistic and graphical terms, graphical objects in **P** denote the corresponding individuals in the world. We can think of our deictic interpretation procedure as a specific implementation for our graphical domain of Kaplan's operator DTHAT —in our simplified extensional language— which takes a term and maps it into an individual of the world in the interpretation context whenever the term is used deictically (Kaplan, 1978).

The interpretation of proper names and definite descriptions has been the source of interesting semantic problems. Consider that linguistic terms serve the purpose to identify individuals, and whenever they are used, the individual they denote should exist. However, as pointed out by Donnellan and commented by Kaplan, "using a definite description referentially a speaker may say something true even though the description correctly applies to nothing" (Kaplan, 1978). The example is as follows: if someone says *His wife is kind to him* when a bachelor enters into a room accompanied by a woman that has been misintroduced as the bachelor's wife, having the speaker noticed the woman solicitous attention to the man, the speaker used the description *his wife* to refer to a bachelor's wife!, and nevertheless what he said was true. Here, one would be inclined to say that *his wife* applies to nothing, but if the woman is in the visual field of the speaker, it would be more proper to say that *his wife* applies deictically to the woman. Consider that if the expression *she is kind to him* had been used instead, or simply, the speaker had pointed to the woman at the time the expression *is kind to him* were uttered, the deictic nature of the reference would be easily revealed. According to Kaplan, whenever a description is used referentially (as opposed to attributively), describing can be taken as a form of pointing, and as he suggested, instead of taking the sense of a description as the subject of a proposition, the sense is used only to fix the denotation which is then taken directly as the subject of the proposition. Similarly, although a proper name is usually thought of as related to an individual (the bearer) in an intimate fashion through an interpretation function in model theory, and it is often stated that proper names are related to the same individual through all world and time indices (i.e., as rigid designators (Kripke, 1972)), proper names can be also interpreted deictically to fix a referent which can then be taken directly as the subject of a proposition.

3.2 Generation of Natural Language Descriptions

The multimodal representation scheme and the resolution of deictic inferences presented above permit the generation of multimodal descriptions in a simple and systematic fashion. Once a multimodal representational system is fully defined the generation of graphical and linguistic expressions can be achieved through the translation rules directly. As the crucial piece of knowledge to be able to use the translation rules is the translation model, the deictic inference required to identify an individual and the inference required to generate a description for such an individual, are not but two sides of the same coin.

If a graphical object is pointed out on the screen, a number of natural language descriptions to refer to such an object can be produced. To find an appropriate description several strategies are available depending on whether the object pointed at is on $\mathbf{P_{base}}$ or whether it is an emergent object. Another consideration is whether the object has a proper name or can be referred to either by a definite or an indefinite description. Suppose that a graphical cursor for pointing to graphical objects is available. The cursor itself is modeled as a graphical object of type *dot* within the graphical domain. With this interactive device we can identify dots if the position of a dot in a drawing is the same as the position of the cursor, lines and curves if the cursor lies on the line or the curve, and regions if the cursor is inside or in the border of a region. With this device we can select all basic or emergent graphical objects that are identified by an individual pointing act. Basic objects will be identified directly, and the emergent objects selected by a pointing act will be those that can be produced by the grammar of **P** and satisfy the geometrical conditions associated to the cursor. Objects identified by a cursor will be terms of **P** which can be translated into **G** as proper names, definite descriptions or indefinite descriptions. To refer to a graphical object we can use the following simple strategy: if a graphical object can be translated into **L** as a proper name, use the proper name; if the graphical object can be translated as a description and it is the only one satisfying such a description, use a definite description; otherwise, use an indefinite description.

Consider a pointing action in which the dot d_1 or the region r_1 is selected. The translations from **P** into **G** of these objects, according to rule $T1_{P\text{-}G}$ (a), are $\lambda P[P(d_1)]$ and $\lambda P[P(r_1)]$ respectively. As these expressions can be translated into **L** as shown in Figure 14, these objects can be referred to in **L** as *Paris* and *France*, respectively. However, consider that as these objects are the concrete extension of a number of denoting concepts of **G**, these objects can also be translated into **G** through rule $T1_{P\text{-}G}$ (b) as A/THE(*dot*), and A/THE(*region*), respectively, and they can be translated into **L** as *a/the city* and *a/the country*. Another possible translation for the dot d_1 whenever it is produced by rule $S4_P$ and translated into **G** through $T4_{P\text{-}G}$ is A/THE(OF_a(R_1)(*dot*)), which in turn can be translated into **L** as *a/the city of France*.

There may be constants of **P** that cannot be translated into **L** as proper names. Consider, for instance, the line l_1 in Figure 4; the translation of this line into **G** is $\lambda P[P(l_1)]$, but as can be seen in Figure 14, there is no proper name that corresponds to this expression in **L**. However, this constant could be translated by rule $T1_{P\text{-}G}$ (b) as the denoting concepts A/THE(*line*). As the line is also an emergent object that can be produced through the syntactic rule $S2_P$, its translations according to rule $T2_{P\text{-}G}$ are A/THE(FROM_TO(D_1)(D_3)(*line*)); subsequently, such an object can be referred to in **L** as *a/the line* and *a/the line from Paris to Frankfurt*. A similar example is the description of the curve c_1 which is produced by rule $S3_P$ and can be translated into **L** through **G** as *a/the border between France and Germany*.

As was mentioned in Section 1, the generation of descriptions is required within the context of specific rhetorical and intentional structures, like the *activate* structure of the WIP system which employs Reiter and Dale's algorithm for the production of definite descriptions on demand. Our system can be used to support the generation of descriptions either definite or indefinite, and even pronouns used deictically, in multimodal generation systems with a solid semantic base. These descriptions would be used according to particular rhetorical and intentional structures related to specific application domains. The advantage of such an approach would be that the choice of the expressions to be used in multimodal presentations could be made not only on the basis of predifined heuristics, but also on the basis of the semantic value of these expressions in the context of use. In addition, the decision about what kind of knowledge is expressed through either modality for the production of coordinated natural language and graphical explanations could be made taking into account not only the kind of heuristics that are currently employed in systems like WIP and COMET, but also on the basis of the expressiveness and effectiveness criteria of natural and graphical languages.

4 Multimodal Discourse Representation Theory

The ability to interpret individual multimodal messages is a prerequisite for interpreting sequences of multimodal messages occuring in the normal flow of interactive conversations. In the same sense that discourse theories, like DRT, are designed to interpret sequences of sentences, it is desirable to have a theory in which sequences of multimodal messages can be interpreted. Such a kind of theory would have to support anaphoric and deictic resolution models in an integrated fashion, and would have to be placed in a larger pragmatic setting in which intentions and presupositions are considered, and in which mechanisms to retrieve knowledge from memory are also taken into account. To work out such a theory is quite an ambitious goal, however, in the same way that DRT focuses in internal structural processes that govern anaphoric resolution, it is plausible to consider a multimodal discourse representation theory (MDRT) to cope with the resolution of spatial deictic inferences. In the same way that DRT postulates discourse representation structures in which referents and conditions are introduced incrementally through the interpretation of the incoming natural language discourse by means of the application of construction rules, it is plausible to conceive similar multimodal discourse representation structures (MDRS) whose referents and conditions would be introduced by modality depending construction rules acting upon the expressions of the corresponding modality. In these structures, DRS-conditions extracted from different modalities would be kept in separate partitions, but discourse referents would be abstract objects common to the whole MDRS. In particular, MDRS's could help to specify accesibility relations between anaphoric and deictic terms and their antecedents and interpretation context, imposing severe constraints on the possible interpretations, as it is normally done in DRT. The resolution process itself would be accomplished by incremental constraint satisfaction as was shown for deictic inferences. In the rest of this section we present a schematic picture of how a MDRS can be developed, and we examplify the use of this structure in relation to the interpretation of the multimodal message in Figure 2. To start this presentation consider an empty MDRS in Figure 22.

The MDRS is a structure with four partitions and extends traditional DRS with one partition for graphical conditions and another to store the translation models that hold in a particular interpretation state. The partition for linguistic conditions is used as in normal DRS, and the top partition for referents includes a variable for every individual that is referred to in the multimodal message in either of the modalities. Figure 23 illustrates the initial state for the interpretation process of the multimodal message in Figure 2. As can be seen, graphical expressions of **G** (the map) are included in the graphical section of the MDRS and textual conditions, with the associated type information, are included in the linguistic section as in normal DRS. A referent is included in the corresponding partition for every individual that has been introduced through either modality, as referents are considered media

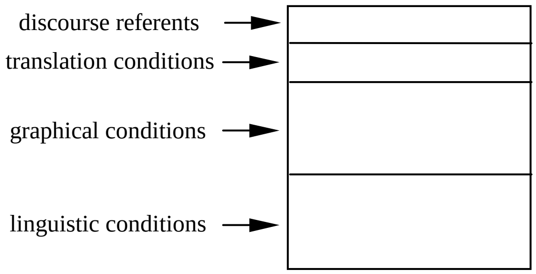


Figure 22 Components of the multimodal discourse representation structure.

independent abstractions. In the same way that the order of processing of linguistic information is not crucial for the definition of the linguistic conditions, we abstract over scanning considerations and assume that graphical expressions are introduced as a single "sentence"[20]. Finally, the partition for the possible translation models is empty in this stage as the coreferring relation between text and graphics has not been established yet.

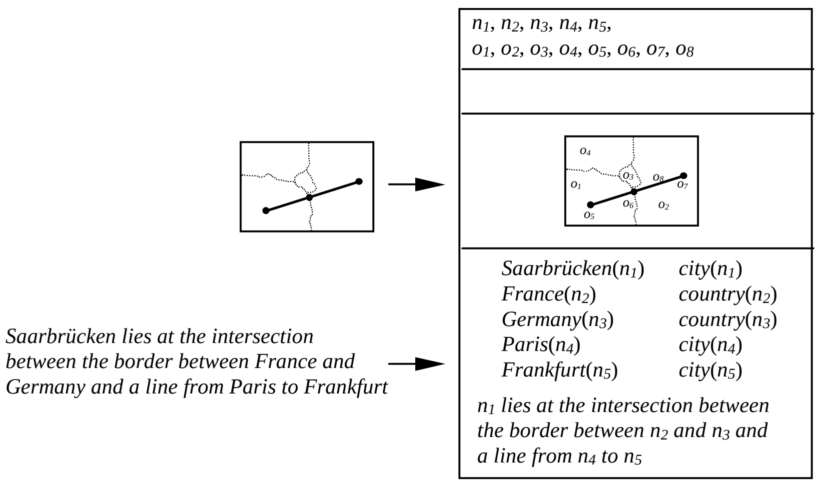


Figure 23 Initial MDRS for the interpretation of the multimodal message.

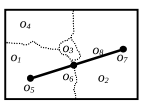
[20] We left for further research whether the analysis of scanning protocols by means of eye-tracking techniques can provide information to impose additional constraints on accesibility relations and possible translation models for the construction of MDRS's. An interesting antecedent for the definition of this kind of constraints can be found in (Faraday and Sutcliffe, 1998).

The interpretation process by constraint satisfaction is illustrated in Figure 24. Figure 24 (a) illustrates the interpretation state after the first cycle of the constraint satisfaction algorithm presented in Section 3.1 has been carried out. In this state, the partition for the translation conditions contains the disjunction of the four possible translation models that are consistent with the message, taking the message itself as the only interpretation constraint. Figure 24 (b) illustrates the interpretation state once the additional constraint that Germany is to the east of France has been considered[21]. The interpretation of the corresponding expression introduces two additional discourse referents (n_6 and n_7) as the terms *Germany* and *France* in the textual part of the message should be resolved anaphorically in relation to the context previously built. However, this anaphoric resolution process is kept within the linguistic section of the MDRS and should take into account the accesibility constraints between anaphor and antecedent as commonly done in DRT. The result of this anaphoric inference is reflected as the equality conditions n_6=n_3 and n_7=n_2. The inclusion of the constraint *Germany is to the east of France* permits to rule out two possible translation models and the result of the second cycle of the constraint satisfaction algorithm is reflected in the new state of the partition for translation conditions of the MDRS. Figure 24 (c) illustrates the final interpretation state in which the constraint that Paris is a city of France has been considered and involves anaphoric and deictic resolution inferences as in the interpretation of the previous constraint. As a result of this last constraint satisfaction cycle only one translation model is left in the partition for translation conditions and reflects the correct interpretation of the multimodal message.

(a)

$n_1, n_2, n_3, n_4, n_5,$
$o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8$

$\{n_1=o_6, n_2=o_1, n_3=o_2, n_4=o_5, n_5=o_7\} \vee$
$\{n_1=o_6, n_2=o_2, n_3=o_1, n_4=o_5, n_5=o_7\} \vee$
$\{n_1=o_6, n_2=o_1, n_3=o_2, n_4=o_7, n_5=o_5\} \vee$
$\{n_1=o_6, n_2=o_2, n_3=o_1, n_4=o_7, n_5=o_5\}$



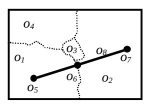
Saarbrücken(n_1) *city*(n_1)
France(n_2) *country*(n_2)
Germany(n_3) *country*(n_3)
Paris(n_4) *city*(n_4)
Frankfurt(n_5) *city*(n_5)

n_1 lies at the intersection between the border between n_2 and n_3 and a line from n_4 to n_5

(b)

$n_1, n_2, n_3, n_4, n_5,$
$o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8$
n_6, n_7

$\{n_1=o_6, n_2=o_1, n_3=o_2, n_4=o_5, n_5=o_7\}$
$\vee$
$\{n_1=o_6, n_2=o_1, n_3=o_2, n_4=o_7, n_5=o_5\}$



Saarbrücken(n_1) *city*(n_1)
France(n_2) *country*(n_2)
Germany(n_3) *country*(n_3)
Paris(n_4) *city*(n_4)
Frankfurt(n_5) *city*(n_5)

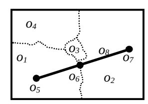
n_1 lies at the intersection between the border between n_2 and n_3 and a line from n_4 to n_5

n_6 is to the east of n_7
$n_6=n_3$, $n_7=n_2$

(c)

$n_1, n_2, n_3, n_4, n_5,$
$o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8$
n_6, n_7
n_8, n_9

$\{n_1=o_6, n_2=o_1, n_3=o_2, n_4=o_5, n_5=o_7\}$



Saarbrücken(n_1) *city*(n_1)
France(n_2) *country*(n_2)
Germany(n_3) *country*(n_3)
Paris(n_4) *city*(n_4)
Frankfurt(n_5) *city*(n_5)

n_1 lies at the intersection between the border between n_2 and n_3 and a line from n_4 to n_5

n_6 is to the east of n_7
$n_6=n_3$, $n_7=n_2$

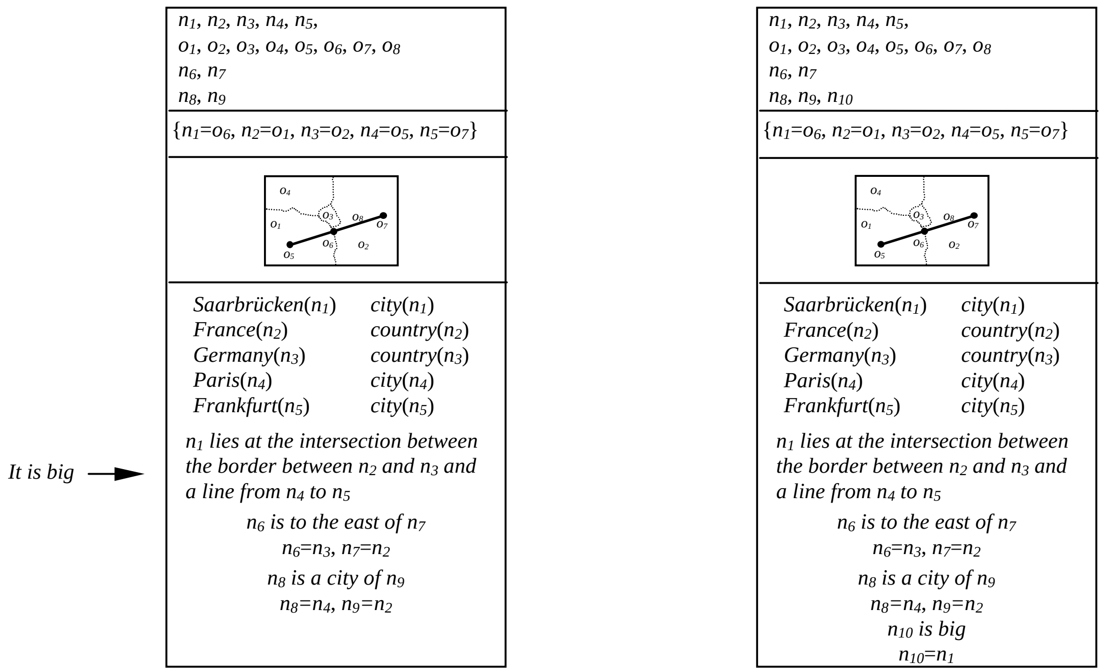
n_8 is a city of n_9
$n_8=n_4$, $n_9=n_2$

Figure 24 Interpretation of multimodal message by constraint satisfaction.

As a last example of the integrated anaphorical and deictical interpretation consider a situation in which the natural language expression *It is big* is mentioned after the multimodal message in Figure 2 has been interpreted, as illustrated in Figure 25. In this latter situation the natural language information would enter into the partition for linguistic conditions and the pronoun *it* should be interpreted anaphorically in relation to the context currently

[21] How this constraint is selected is beyond the scope of this paper and we only make the assumption that the symbols in the graphical and linguistic partitions of the MDRS form a part of the indexing scheme required to retrieve the information from memory. For a prototype implementation, this kind of constraints could be provided by the human-user directly.

provided by the MDRS and could resolve to *Saarbrücken* (although there are several possibilities). However, if the expression is supported by an overt ostention to the city of Paris, for instance, it would have been deictic and its interpretation would have to be worked out with the same machinery; although in this latter situation the translation relation between graphical and linguistic referents could be asserted directly in the translation model as the ostention would spare the constraint satisfaction part of the deictic inference.



(a) (b)

Figure 25 Integrated interpretation of anaphora and spatial deixis.

With this we conclude the presentation of our model for integrated deictic and anaphoric inferences. As can be seen, the distinction between anaphora and deixis is clearly demarcated as the antecedent for a pronoun, a proper name or a description used anaphorically is provided by the discourse interpretation context, while the referent for a deictic pronoun, proper name, definite or indefinite description, or a demonstrative word like *this* and *that* is taken from an intermediate representation of a non linguistic modality like the graphical context, and denotes an individual of the world directly, a view that is consistent with Kamp's distinction quoted in Section 1.

5 Conclusions

In this paper a theory of representation and interpretation for multimodal messages, and a model for multimodal reference resolution has been presented. The model is based on the view that a modality is a code system on a medium which can be characterized through well-defined syntax and semantics. Multimodal interpretation consists on working out corefering relations between terms of different modalities. A central concern for the articulation of this theory is a clear characterization of the resolution process of spatial deitic inference and its relation with the resolution of anaphors in normal flow of discourse. A key theoretical assumption is that graphics are interpreted deictically in opposition to alternative views according to which the interpretation of graphical representations can be characterized as anaphoric.

The theoretical machinery for carrying on with the definition of the syntax and semantics is formally developed along the lines of Montague's semiotic programme and its associated general theory of translation. We have also illustrated an algorithm for finding out the translation between texts and graphics, as messages in these modalities are introduced through independent input channels, and the translation between linguistic terms and their corresponding graphical expressions must be induced dynamically. It was also suggested to extend Kamp's DRT with multimodal discourse structures (MDRS). This model defines, on the one hand, an integrated interpretation model for multimodal messages while, on the other, keeps a clear demarcation between indexical and anaphoric inferential processes. Natural language terms, like proper names, pronouns and descriptions can be interpreted in relation to a model; however, these linguistic terms admit anaphoric and deictic interpretations too.

A final remark before concluding this paper is that although we used a simplified extensional definition for the semantics of natural language and graphical expressions, the system was carefully designed to move smoothly into the intensional domain. Consider that the extensional formulation used in the semantic definition of the graphical language can be easily extended into an intensional one by changing the types of constants, predicates and sentences from individuals, set of individuals and truth values into individual concepts, properties of individuals and propositions, respectively. This is achieved simply by indexing the interpretation of expressions in terms of a possible world and time, and all definitions presented above could be considered relative to the current world and time. The move to the intensional domain would allow the definition of the interpretation of more comprehensive natural language segments.

Intensionality is also relevant for the interpretation of graphical languages, in general, and for the definition of graphical and linguistic interactive systems, in particular. In computer graphics interactive sessions, the interaction states can be considered as possible worlds and the interpretation of graphical constants would depend on particular graphical states. If a graphical object like a dot, for instance, is moved from one position to another in an interactive transaction, we have the intuition that the object before and after the change is the same and denotes the same object of the world and yet not even its position, which one could think of as a essential property of a dot, is the same. Accordingly, in the intensional setting the semantic value of a graphical constant is not an individual but an individual concept. Consider as well that the same graphical description can have different semantic values in different interactive states; for instance, the value of the expression *position of* d_1 will be a different ordered pair before and after dot d_1 is moved. According to this, the interpretation of graphical operators at every index will be a function from sequences of graphical objects of the proper kind into graphical objects; however, unlike normal linguistic situations in which a different function in different index are assigned to operators and predicates, the same function at every index has to be assigned to geometrical operators and predicates, as the geometry is always the same. Moving into the intensional setting is also relevant for our treatment of indexicals. In our current approach the interpretation of a term used deictically is an individual of the world; in the intensional context the interpretation of the same term in one particular interaction state will be the same in every state despite that the description for referring to such an object in the state in which it was selected might pick up a different individual in a different state.

An intriguing path of exploration for further development consists in dealing with a more general fragment of natural language in which temporal expressions are included. In the same way that the language **G** provided a finite and small domain for the interpretation of linguistic spatial prepositions, a similar language **T** for the interpretation of temporal prepositions could be defined. Temporal predicates and operators of this language would be interpreted in terms of arithmetic functions like those presented, for instance, in Allen's temporal logic (Allen, 1983). In the same way that the constraint satisfaction algorithm for the definition of the translation between graphical and linguistic terms helped to solve deictic inferences, a constraint satisfaction algorithm for resolving temporal deictic references in relation to a finite and small domain of actions and events is conceivable. The definition for such a spatial and temporal indexical model could be quite helpful for the implementation of natural language and graphics systems in which actions and events are mentioned in the curse of interactive conversations.

6. Implementation

Although a prototype system for the theory presented in this paper has not been implemented, several aspects of the theory have been implemented in relation to simpler systems. A simpler version of the strategy for multimodal interpretation of the scheme in Figure 3 was implemented in the first version of the Graflog system (Pineda, 1989). Several versions of the graphical language and its geometrical interpreter have been implemented in relation to different application domains (Morales, 1994), (Masse, 1994), (Santana et al., 1997) and (Garza, 1995) with BinProlog and the TCL/TK programming environment.The geometrical interpreter and the strategy of evaluating a set of geometrical constraints incrementally in relation to a graphical domain was used in a later version of Graflog to solve and generate graphical explanations of geometrical constraint satisfaction problems (Pineda, 1992, 1998), and also for the definition of a model (not fully implemented yet) for the production of solids out of the orthogonal views of polyhedra (Garza and Pineda, 1998). The scheme for enumerating functions used in the definition of translation models was implemented for a semantic theorem-proving system written in Prolog to find the possible models satisfying logical theories about graphical scenarions of the Hyperproof system (Barwise et al., 1994) on the light of the present theory.

References

James Allen. 1983. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*. 26(11), pp. 832-843.

Elisabeth André, Thomas Rist. 1994. Referring to World Objects with Text and Pictures, Coling 94.

Jon Barwise and John Etchemendy. 1994. *Hyperproof*. CSLI.

David R. Dowty, Robert E. Wall, Stanley Peters. 1985. *Introduction to Montague Semantics*. D. Reidel Publishing Company, Dordrecht, Holland.

Pete Faraday and Alistair Sutcliffe. 1998. Providing Advice for Multimedia Designers. *Proceedings of the Workshop on Interpretation and Generation in Intelligent Multimodal Systems and Graphical Reasonings in Expert Systems* (L. Pineda, T. Rist and J. Lee Eds.) 4th World Congress on Expert Systems. ITESM Mexico City Campus, Mexico city.

Steven Feiner and Kathleen McKeown. 1993. Automating the Generation of Coordinated Multimedia Explanations. In: *Intelligent Multimedia Interfaces*, pp. 117-239. M. Maybury (ed.), AAAI Press/MIT Press, Menlo Park, California.

Gabriela Garza. 1995. Síntesis de Poliedros a partir de sus Vistas Ortogonales: Un Caso de Estudio acerca del Razonamiento Gráfico. Msc. Thesis, ITESM, Campus Morelos, México.

Gabriela Garza, Luis Pineda. 1998. Synthesis of Solid Models of Polyhedra from their Orthogonal Views using Logical Representations. *Expert Systems with Applications*. Vol. 14. pp. 91-108. Elsevier Science Ltd.

Hans Kamp. 1981. A Theory of Truth and Semantic Representation. *Formal Methods in the Study of Language*, 136 pp. 277-322, Mathematical Centre Tracts.

Hans Kamp, Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publisher, Dordrecht, Holland.

David Kaplan. 1978. DTHAT. *Syntax and Semantics*, Vol. 9.

Ewan Klein, Luis Pineda. 1990. Semantics and Graphical Information. *Human-Computer Interaction, Interact'90*. pp. 485-491. Diaper, Gilmore, Cockton, Shackel (eds). IFIP, North-Holland.

Kripke, Saul. 1972. *Naming and Necessity*. Basil Blackwell, Oxford.

John Lyons. 1968. *Introduction to Theoretical Linguistics*, Cambridge University Press, Cambridge.

Jock Douglas Mackinlay. 1987. Automatic Design of Graphical Presentations. PhD Thesis. Stanford University. University Microfilms International.

Mark Maybury. 1993. Planning Multimedia Explanations Using Communicative Acts. In *Intelligent Multimedia Interfaces*, pp. 59-74. M. Maybury (ed.), AAAI Press/MIT Press, Menlo Park, California.

Johanna Moore. 1995. Participating in Explanatory Dialogues: interpreting and responding to questions. A Bradford Book, The MIT Press, Cambridge.

J. Antonio Masse. 1994. Satisfacción de Restricciones por Referencia Simbólica en Dibujos Geométricos. Bsc. Thesis, ENEP Aragón, UNAM, México.

Rafael Morales. 1995. Pizarrones Interactivos Multimodales para la Enseñanza de Conceptos Matemáticso, Msc. Thesis. ITESM, Campus Morelos, México.

Luis Pineda. 1989. Graflog: a Theory of Semantics for Graphics with Applications to Human-Computer Interaction and CAD Systems. PhD thesis, University of Edinburgh, U.K.

Luis Pineda. 1992. Reference, Synthesis and Constraint Satisfaction. *Computer Graphics Forum*. 11(3), pp. 333-344.

Luis Pineda. 1998. Graphical and Linguistic Dialogue for Intelligent Multimodal Systems. *Expert Systems with Applications*. Vol. 14. pp. 149-157. Elsevier Science Ltd.

Raymond Reiter, Alan K. Mackworth. 1987. The Logic of Depiction, *Technical Reports on Research in Biological and Computational Vision* at the University of Toronto. RCBV-TR-87-18.

Ehud Reiter, Robert Dale. 1992. A Fast Algorithm for the Generation of Referring Expressions. In: Proc. of the 14th. COLING, Volume 1, pp. 232-238, Nantes, France.

Sergio Santana, Sunil Vadera, Luis Pineda. 1997. The coordination of linguistic and graphical explanations in the context of geometric problem-solving tasks. *Technical Report on the IIE/University of Salford in-house PhD programme.*

M. I. Shamos. 1978. Computational Geometry. Ph. Thesis, Yale University. University Microfilms International.

G. Stiny. 1975. *Pictorical and Formal Aspects of Shape Grammars*. Basel, Birkhauser Verlag.

Oliviero Stock et. al. 1993. AlFresco: Enjoying the Combination of Natural Language Processing and Hypermedia for Information Exploration. In: *Intelligent Multimedia Interfaces*, pp. 197-224. M. Maybury (ed.), AAAI Press/MIT Press, Menlo Park, California.

Kent Wittenburg (in press) Visual Language Parsing: If I had a Hammer. In Harry Bunt and Robbert-Jan Beun (eds.), *Cooperative Multimodal Communication*, Kluwer.

L. Wittgenstein, 1953. *Philosophical Investigations*. Oxford: Basil Blackwell , Part II, section 10.

Wolfgang Wahlster. 1991. User and Discourse Models for Multimodal Communication. In *Intelligent User Interfaces*, pp. 45-67. J. W. Sullivan and S. W. Tyler (eds.), Frontier Series. New York: ACM Press.

Wolfgang Wahlster, Elisabeth André, Wolfgang Finkler. Hans-Jürgen and Thomas Rist. 1993. Plan-based integration of natural language and graphics generation, *Artificial Intelligence* 63, pp. 387-427, Elsevier.