A Model for Multimodal Representation and Inference[1]

Luis Pineda and Gabriela Garza
Department of Computer Science, IIMAS, UNAM,
Mexico City, Mexico

luis@leibniz.iimas.unam.mx

Abstract

In this paper some applications of a theory for representation and inference in multimodal scenarios is presented. The theory is focused on the relation between natural language and graphical expressions. A basic assumption is that graphical expressions belong to a language with well-defined syntax and semantics: a graphical language. A second assumption is that the relation between expressions of different modalities is similar to the relation of translation that holds between expressions of different natural languages. In this paper a multimodal system of representation and inference based on this view of modality is described. First, a brief introduction to the representational structures of the multimodal system is presented. Then, a number of multimodal inferences supported by the system are illustrated. These examples show how the multimodal system of representation can support the definition and use of graphical languages, perceptual inferences for problem-solving and interpretation of multimodal messages. Finally, the intuitive notion of modality underlying this research is discussed.

1. Multimodal Representation

The system of multimodal representation that is summarized in this paper is illustrated in Figure 1. The notion of modality in which the system is based is a representational notion: information conveyed in one particular modality is expressed in a representational language associated with the modality. Each modality in the system is captured through a particular language, and relations between expressions of different modalities are captured in terms of translation functions from basic and composite expressions of the source modality into expressions of the object modality. This view of multimodal representation and reasoning has been developed in [13], [17], [9], [18] and [19], and it follows closely the spirit of Montague's general semiotic programme [5].

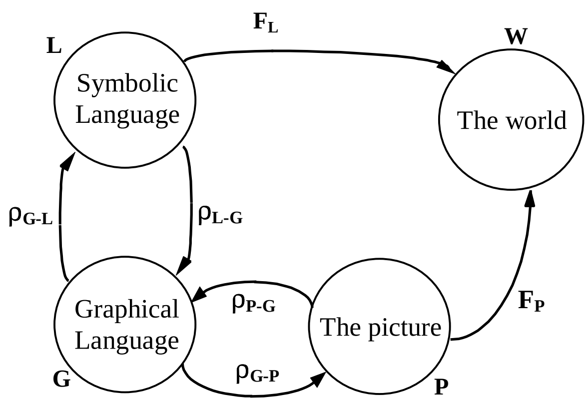


FIGURE 1. Multimodal system of representation.

The theory is targeted to define natural language and graphical interactive computer systems and, as a consequence, the model is focused in these two modalities. However, the system is also used to express conceptual information in a logical fashion and, depending on the application, the circle labeled **L** might stand for first-order logic or any other symbolic language as long as the syntax is well-defined and the language is given a model-theoretical semantic interpretation.

The circles labeled **L** and **G** in Figure 1 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

[1] To be publish also in "Visual Representations and Interpretations", Springer-Verlag, 1998.

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 but which is expressive enough to express the translation of natural language expressions. It is an interlingua that permits to relate the natural language syntactic structures with the structure of graphics which is captured with a graphical grammar. **P** contains the overt graphical symbols which can be seen and drawn but cannot be manipulated directly and captures the underlying structure of graphical expressions. The functions $\rho_{\mathbf{L\text{-}G}}$ and $\rho_{\mathbf{G\text{-}L}}$ stand for the translation mappings between the languages **L** and **G**, and the functions $\rho_{\mathbf{P\text{-}G}}$ and $\rho_{\mathbf{G\text{-}P}}$ stand for the corresponding translation between **G** and **P**. The translation function $\rho_{\mathbf{P\text{-}G}}$ maps well-defined objects of the graphical modality into expressions of **G** where the interpretation process is performed. The translation $\rho_{\mathbf{G\text{-}P}}$, on the other hand, maps geometrical expressions of **G** into pictures. 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 $<\mathbf{W}, \mathbf{F_L}>$ defines the model $\mathbf{M_L}$ for the natural language, and the ordered pair $<\mathbf{W}, \mathbf{F_P}>$ 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 $\mathbf{F_L} \circ \rho_{\mathbf{G\text{-}L}}$ or, alternatively, by $\mathbf{F_P} \circ \rho_{\mathbf{G\text{-}P}}$. The denotation of a name in **L**, for instance, is the same as the denotation of the corresponding graphical object in **G**, as both refer to the same individual. 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. The functions $\rho_{\mathbf{G\text{-}P}}$ and $\rho_{\mathbf{P\text{-}G}}$ define homomorphisms between **G** and **P** as basic and composite terms of these two languages can be mapped into each other.

An important consideration for the scheme in Figure 1 is that the symbols of **P** have two roles: on the one hand they are representational objects, but on the other, they are also geometrical objects that can be talked about as geometrical entities. 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 indirectly through the translation relation and interpretation of symbols in **P** taken as a language (i.e., the composition $\mathbf{F_P} \circ \rho_{\mathbf{G\text{-}P}}$). The ordered pair $<\mathbf{P}, \mathbf{F_G}>$ 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. The semantic interpretation of composite expressions of **G,** on the other hand, is defined through a semantic algebra. 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, for instance, 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.

The purpose of this paper is to provide an overview of the functionality of the system and for that reason in the next section a number of examples involving multimodal inferences in different application domains are illustrated. Our purpose is to show that inferences related to reasoning with graphical languages, solving problems involving interpretation of pictures, interpreting multimodal messages like pictures with their associated captions, can all be explained with the help of a common underlying representational framework and involve a small set of basic but powerful inferential strategies. The formalization of the multimodal representational system is presented elsewhere (e.g., [19]).

2. Multimodal Inference

In this section a number of problems involving multimodal representation and inference in different domains are illustrated. Once these examples are shown a summary of the kinds of multimodal inferences involved is presented.

2.1. Graphical Languages

In this section the definition and interpretation of a graphical language in relation to the multimodal system of representation is illustrated. Consider the picture in Figure 2 a) in which there are two triangles and two rectangles that have been assigned an interpretation through a graphical and natural language dialogue supported by pointing acts. The setting is such those triangles are interpreted as students and rectangles as subjects; additionally it is stated that if a student is in a subject he or she studies that subject, and if a student studies both subjects he or she is clever. According to this interpretation the picture in Figure 2 a) is a graphical expression that expresses that both students are clever, but if the picture is manipulated as shown in Figure 2 b), a graphical expression is formed which expresses the fact that only John is clever.

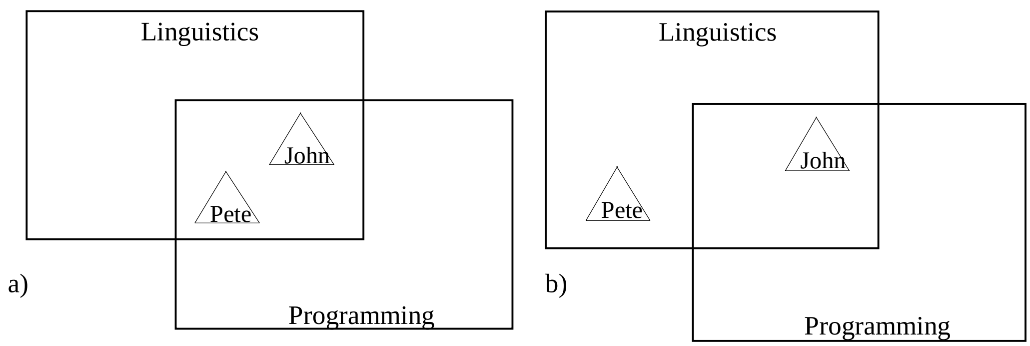


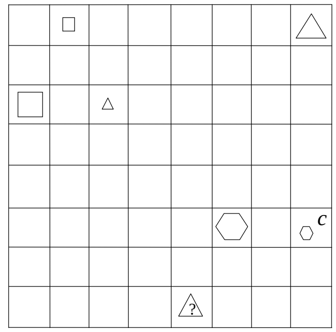
Figure 2. Graphical Expressions

The question is how this knowledge is represented and, in particular, what is the relation between the expression of the abstraction (i.e., that a student is clever), and the geometrical fact that the symbol representing the student is contained within the rectangle representing a subject. For the interpretation of this particular situation the linguistic preposition *in* is interpreted as a geometrical algorithm that computes the relation in the graphical domain. To answer the question whether a student is clever or whether all students are clever, a deductive reasoning process is performed upon the representational structures in the language L; however, when the interpretation of the spatial preposition and its arguments is required to complete the inference, there is no knowledge available in L and the corresponding expression has to be translated into a expression in G in the graphical domain, which in turn can be evaluated by the geometrical interpreter with the help of a geometrical algorithm that tests the geometrical predicates involved. The result of this test is translated back into the language L to allow the reasoning process to succeed. As can be seen, in this kind of inference the picture functions as a recipient of knowledge that can be extracted on demand by the high-level reasoning process performed at the symbolic level. This kind of inference has been characterized as *predicate extraction* by Chandrasekaran in ([4]), and it is commonly used in graphical reasoning systems and the interpretation of expressions of visual languages, where large amounts of information are represented through graphics and geometrical computations improve considerably the efficiency of the reasoning process. For further discussion of this notion of graphical language see [12] and [13].

2.2. Perceptual Inference

One important feature of the multimodal interpretation and reasoning strategies used in the scenario of Section 2.1 is that the translation functions between expressions of **L** and **G** are defined in advance. The multimodal interpretation and reasoning cycle must move across modalities in a systematic fashion and this is achieved through the mappings defined in terms of the translation functions. However, there are situations in which the interpretation of a multimodal message or the solution of a problem involving information in different modalities requires to establish such an association in a dynamic fashion.

Consider, for instance, a problem typical of the Hyperproof system for teaching logic ([2]) in which information is partially expressed through a logical theory and partially expressed through a diagram, as shown in Figure 3.



given:

$large(a) \vee small(a)$
$hex(b) \wedge below(a,b)$
$\forall x(triangle(x) \wedge large(x) \rightarrow left\_of(d,x))$
$\neg\exists x(small(x) \wedge below(x,c))$

prove:

$square(d) \vee small(d)$

FIGURE 3. Multimodal problem.

As can be seen the problem consists in finding out whether the object named *d* is either a square or small. This inference would be trivial if we could tell by direct inspection of the diagram what object is *d*, but that information is not available. Note, on the other hand, that under the constraints expressed through the logical language the identity of *d* could be found by a "valid" deductive inference. Note in addition that the information expressed in the diagram in Figure 3 is incomplete. In the Hyperproof setting, the question mark on the bottom triangle indicates that we know that the object is in fact a triangle but its size is unknown to us. However, the conceptual constraints expressed in the logical language do imply a particular size for the occluded object which can be made explicit through the process of multimodal problem-solving. This situation is analogous to the interpretation of images in which some objects are occluded by some others.

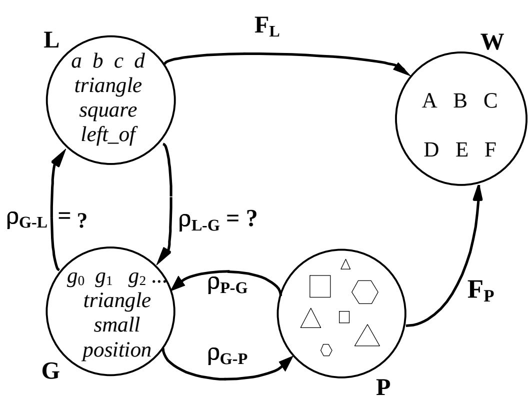


FIGURE 4 Multimodal representation system for the Hyperproof problem**.**

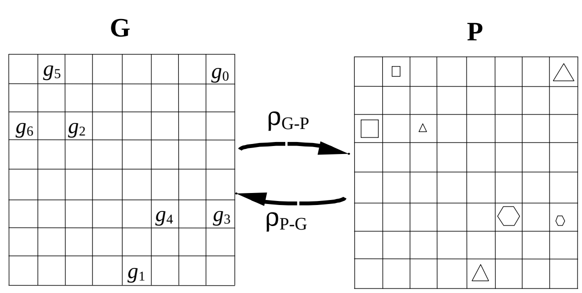


FIGURE 5 Relation between **G** and **P**.

In terms of our system of multimodal representation the task is not, like in the previous example, to make explicit information that was expressed only implicitly by predicate extraction or graphical simulation, but to find out what are the translations between basic constants of the logical language, the names, and the graphical objects of which they are names of. The problem is to induce the translations between basic constants of both representational languages. This situation is illustrated in Figure 4 in which the translation functions have been labeled with a question mark.

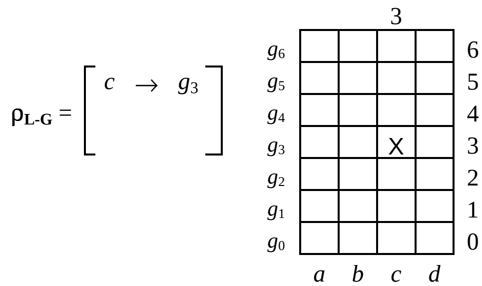


FIGURE 6. Initial interpretation function.

Another way to look at this is thinking of the graphical objects as the domain of interpretation for the logical theory. The multimodal inference consists in finding out all consistent models for the theory, and these can be found through a process of incremental constraint satisfaction.

Consider Figure 5 in which a constant of **G** has been assigned to every graphical object (i.e., the objects of **P** properly). At the starting point of the interpretation process only the identity of the block *c* is known as can be seen in Figure 3. Accordingly, the interpretation of the linguistic theory is partially defined only. To see this consider Figure 6 in which a table relating the names of the theory in the horizontal axis with the names of the graphical objects in the vertical one is shown. This table can be interpreted as a partial function from individual constants of **L** to individual constants of **G** if no more than one square in each column is filled up. The interpretation task consists in completing this function by assigning a graphical object to each name in a manner that is consistent with the first-order logical theory expressed in **L** in the conceptual domain.

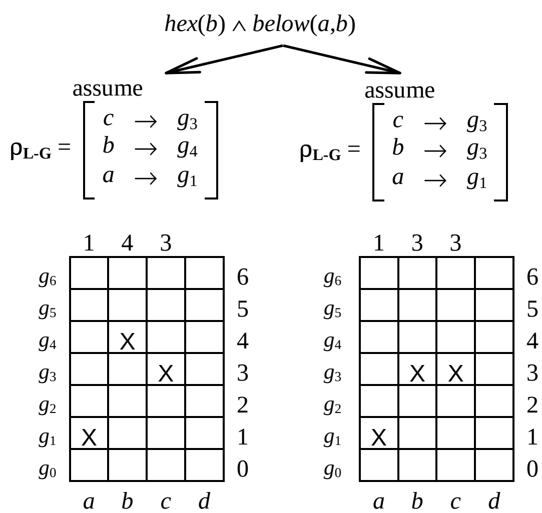


Figure 7. Two possible ways for extending the interpretation function $\rho_{\mathbf{L\text{-}G}}$.

The strategy will be to find the set of consistent models incrementally in a cycle in which a formula of the theory is assumed to be true and all consistent models for such an assumption are found out through geometrical verification. Each cycle of assumption and verification is concluded with an abstraction phase in which all consistent models computed in the cycle are subsumed into a single complex object.

To exemplify this cycle of model construction consider that the formula $hex(b) \wedge below(a,b)$ –in Figure 3– of the theory can be assumed to be true. With this assumption it is possible to extend the function in Figure 6 in two possible ways, which represent consistent models with the assumption and the given facts, as shown in Figure 7.

To end the incremental constraint satisfaction cycle it can be noticed that the two partial models in Figure 7 are similar in the denotations assigned to the objects *a* and *c*, and only differ in the denotation assigned to object *b*. Then, these two models can be subsumed into a structure by simple superposition as shown in Figure 8 in which the column for *b* that is filled with two marks is taken to represent either of both functions. This incremental constraint satisfaction cycle can be continued until the set of models for the theory is found and expressed as an abstraction, as was discussed above.

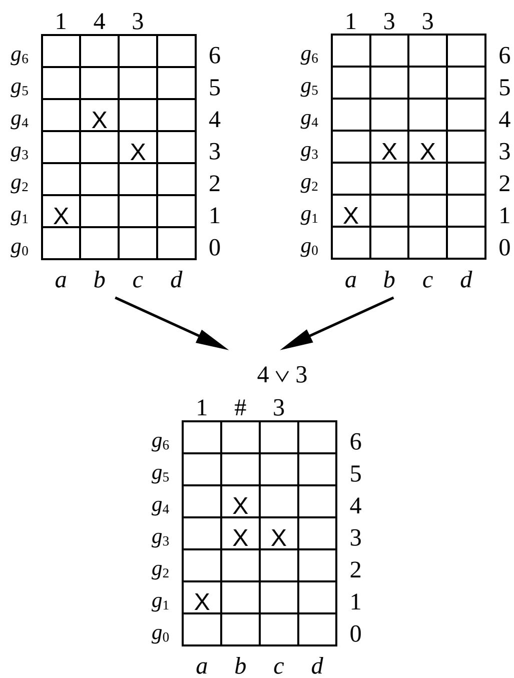


FIGURE 8. Subsumption of two models.

There is an additional way in which we can profit from the process. Consider that in the original stipulation of the problem the graphical information is incomplete, as the size of the bottom triangle is unknown. However, with the partial model obtained after the first inference cycle, in which such a block has been identified as *a*, the theory constraints the size of the block which can be found by an inferential cycle involving logical deduction in **L** and graphical verification in **G**. For this particular example, and in relation to the partial model in Figure 8, the proof that the size of such a block must in fact be large is given in Figure 9. This inference requires a cycle of assumption, deduction in **L** and verification in **G** which we refer as heterogeneous inference.

Prove (problem statement):	(0) $large(a) \vee small(a)$
Assume from theory:	(1) $\neg\exists x(small(x) \wedge below(x,c))$
Axiom:	(2) $\neg\exists x(\mathrm{P}(x)) \leftrightarrow \forall x(\neg\mathrm{P}(x))$
From (1) and (2):	(3) $\forall x(\neg(small(x) \wedge below(x,c)))$
Universal instantiation from (3):	(4) $\neg(small(a) \wedge below(a,c)))$
Morgan's law from (4):	(5) $\neg small(a) \vee \neg below(a,c)$
Direct inspection of the diagram:	(6) $below(a,c)$
From (5) and (6):	(7) $\neg small(a)$
From (0) and (7):	(8) $large(a)$.

FIGURE 9. Heterogeneous inference.

Another way to refer to this in the terminology of Chandrasekaran [4] is as *predicate projection* as the predicative information flows not from the picture to the logical theory, as the situation that was referred above as *predicate extraction*, but from the conceptual knowledge expressed through **L** into the graphical theory in **G**.

In summary, the incremental constraint satisfaction cycle involves the following steps:

1) Visual verification (geometrical interpretation)
2) Assumption and verification of theory (identification of consistent models)
3) Heterogeneous inference
4) Abstraction

With the application of this cycle it is possible to find the set of consistent models for the problem stated in Figure 3, which is represented by the abstraction in Figure 10, and corresponds to the six graphical configurations shown in Figure 11.

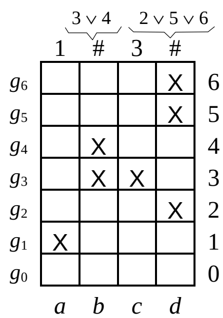


FIGURE 10. Abstraction.

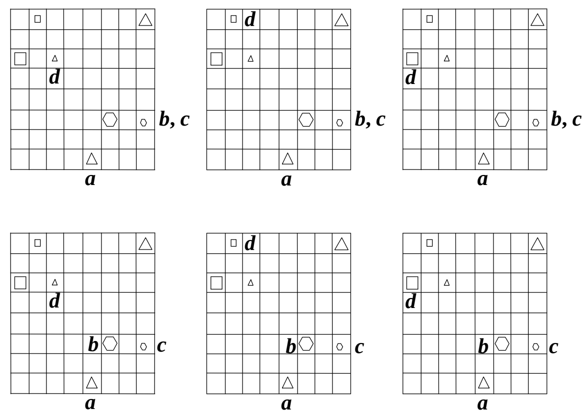


FIGURE 11. Set of possible interpretations.

2.3. Multimodal Interpretation

The next kind of multimodal inference is related to one of the central problems of multimodal communication which we refer as the problem of multimodal reference resolution. This is the problem of finding out the reference of a symbol in one modality in terms of information present in other modalities. In this section we discuss how our model of multimodal representation and interpretation illustrated in Figure 1 can also be applied to the problem of multimodal reference resolution.

Consider situation in Figure 12 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 of 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 non at all.

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

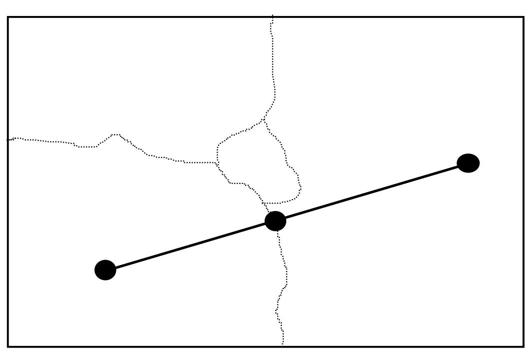


FIGURE 12. Instance of pictorial anaphor with linguistic antecedent.

The situation in Figure 12 has been characterized as an instance of a pictorial anaphor with linguistic antecedent and further related examples can be found in [1]. An alternative view on this kind of problems consists in looking at them in terms of the traditional linguistic notion of deixis [11]. This notion has to do with the orientational features of language which are relative to the spatio-temporal situation of an utterance. To appreciate the deictic nature of the example consider that the inference required to identify the graphical symbols would be simplified greatly if at the time the words *Paris*, *Frankfurt*, *Saarbrücken*, *France* and *Germany* are mentioned overt pointing acts are performed by the speaker. In such a situation the overt ostension would be one factor of the interpretation context among many others. In this respect we can say that pointing is like describing. However, the opposite is also true: the names in the natural language text are like pointers to the graphical symbols and in order to identify the referents of the linguistic terms an inference process is required. For carrying on with such an identification process the context, including graphics and common sense knowledge about the geography of Europe, needs to be considered. For that reason, if we think of the names or other linguistic terms, like pronouns or descriptions, as pointers whose referent can be found out in terms of the context the situation is deictic. We call the inference process that has as a purpose to identify the referent of a graphical or a linguistic term in a multimodal context a *deictic inference*. This notion contrasts with the notion of anaphoric inference in which the referent of a term is found in terms of a context constructed out of expressions of the same modality of the term.

In our model, interpreting the text in Figure 12 consists in interpreting the information expressed through the linguistic modality 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.

In order to see how the multimodal system of representation works for the interpretation of messages with texts and graphics consider, for instance, that the denotations of the word *Saarbrücken* and the dot on the intersection between the straight line and the lower segment of curve representing the border between France and Germany in Figure 12 are the same, which is the city of Saarbrücken itself. If one points out the middle dot at the time the question *what is this?* is asked, the answer is found by applying the function $\rho_{G\text{-}L}$ to the pointed dot, whose value would be the word *Saarbrücken*.

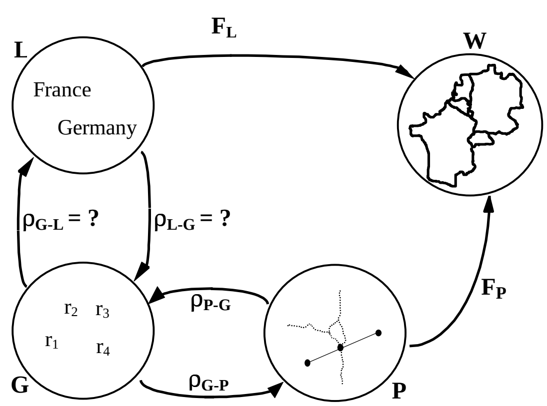


FIGURE 13. Multimodal representational system for linguistic and graphical modalities.

It should be clear that if all theoretical elements illustrated in Figure 13 are given, questions about multimodal scenarios can be answered through the interpretation process, as was shown for the interpretations of graphical expressions in Section 2.1. However, when one is instructed to interpret a multimodal message, like the one in Figure 12, not all information in the scheme of Figure 13 is available. In particular, the translation functions $\rho_{L\text{-}G}$ and $\rho_{G\text{-}L}$ for basic constants are not known, and the crucial inference of the interpretation process has as its goal to induce these functions. This is exactly the problem of finding the set of consistent models in the perceptual inferences carried out in the context of the Hyperproof system as illustrated in the previous section. According to our theory, the kind of so-called perceptual inferences performed by users of the hyperproof system can be characterized as deictic inferences.

It is important to highlight that in order to induce $\rho_{L\text{-}G}$ and $\rho_{G\text{-}L}$ the information overtly provided in the multimodal message is usually not enough. Unlike the inference illustrated in relation to the

Hyperproof system, additional conceptual information would have to be brought into consideration, like common sense knowledge required for the interpretation of maps. However, when contextual knowledge is included in the theory through expressions of **L**, the resolution of multimodal reference can be produced through an incremental constraint satisfaction process that is similar to the one illustrated above in relation to perceptual inference, as the basic inferential strategies required for both kinds of problems are the same.

3. Summary of Multimodal Inferences

From the examples in Sections 2.1 to 2.3 a number of inference strategies have been employed. Similar strategies can be found on examples about design (see [6], [15] and [16]). An analogous view of interpretation of pictures is developed in Reiter's Logic of Depiction (see [20]). Reasoning directly on expressions of a particular representational language, like **L** or **G**, corresponds to traditional symbolic reasoning. However, reasoning in **G** involves, in addition to symbolic manipulation, a process of geometrical interpretation as predicates in **G** have an associated geometrical algorithm. Another way to think about the geometrical representation is that it has a number of expressions representing explicit knowledge; however, it has a large body of implicit knowledge that can be accessed not from a valid symbolic inference, but from the geometry.

The multimodal system of representation supports an additional inference strategy that involves the induction of the translation of basic constants between the languages **L** and **G**, and this process is qualitatively different from a simple symbolic manipulation process operating on expressions of a single language. Examples of this kind of inference strategy are perceptual inferences and resolution of multimodal references which, as we have argued, can be characterized as deictic inferences.

In terms of the system, a multimodal inference can be deductive if it involves symbolic processing in both languages in such a way that information is extracted from one modality and used in the other by means of the translation functions. Multimodal inferences involving the induction of translation relations, or the computation of models, on the other hand, are related to deictic inferences. The use of these two main kinds of multimodal inference strategies is the characteristic of a multimodal inference process which has a deictic character.

4. A Notion of Modality

The multimodal system of representation and inference that has been illustrated in this paper has been developed on the basis of an intuitive notion of modality that can be characterized as *representational*. Representational in the sense that a modality is related in our system to a particular representational language, and information conveyed through a particular modality is represented as expressions of the language associated with the modality. The reason for taking this position is that one aim of this research is to be able to distinguish what information is expressed in what modality, and to clarify the notion of multimodal inference. If an inference is multimodal, it should be clear how modalities interact in the inference process.

This view contrasts with a more psychologically oriented notion in which modalities are associated with sensory devices. In this latter view one talks about visual or auditive modality; however, as information of the same modality can be expressed through different senses (like spoken and written natural language), and the same sense can be used to perceive information of different modalities (written text and pictures are interpreted through the visual channel) this psychological view offers little theoretical tools to clarify how modalities interact in an inference process, and the very notion of modality is unclear.

One consequence of our system is that modalities have to be thought of as related in a systematic fashion, and this relation is established in terms of a relation of translation between modality specific representational languages. One of the reasons to adopt Montague's semiotic programme is precisely to model the relation between modalities as translation between languages.

This view implies also that perceptual mechanisms are related to representational languages in specific ways: a message can only be interpreted in one modality if the information of the message can be mapped by the perceptual devices into a well-formed expression of the representational language associated with the modality. The algorithms mapping information in **P** to expressions of **G**, for instance, are designed relative to the syntactic structure of **G.** These algorithms might be different for different modalities, but once a multimodal system is set up these algorithms are wired, and are fired automatically if suitable input information is present to the input device. This let us to postulate two kinds of perceptual devices: physical, like the visual or auditive apparatus, and logical or conceptual, which relate information input by physical sensory devices with modality specific representational languages. Whether these views can be held is matter for further research.

References

1. Elisabeth André and Thomas Rist. 1994. Referring to World Objects with Text and Pictures, technical report, German Research Center for Artificial Intelligence (DFKI).
2. Jon Barwise and John Etchemendy. 1994. *Hyperproof*. CSLI.
3. A. Borning. 1981. The Programming Language Aspects of Thinglab, A Constraint-Oriented Simulation Laboratory. *ACM Transactions in Programming Languages and Systems*. 3, No. 4. pp. 353-387.
4. B. Chandrasekaran. 1997. Diagrammatic Representation and Reasoning: some Distinctions. *Working notes on the AAAI-97 Fall Symposium Reasoning with Diagrammatic Representations II*. MIT, November 1997. (Also in this volume).
5. David R. Dowty, Robert E. Wall and Stanley Peters. 1985. *Introduction to Montague Semantics*. D. Reidel Publishing Company, Dordrecht, Holland.
6. E. G. Garza and L. A. Pineda, 1998. "Synthesis of Solid Models of Polyhedra Views using Logical Representations, Expert Systems with Applications, Vol. 14, No. 1. Pergamon 1998.
7. Hans Kamp. 1981. A Theory of Truth and Semantic Representation. *Formal Methods in the Study of Language*, 136 pp. 277-322, Mathematical Centre Tracts.
8. Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publisher, Dordrecht, Holland.
9. Ewan Klein and Luis Pineda. 1990. Semantics and Graphical Information. *Human-Computer Interaction, Interact'90*. pp. 485-491. Diaper, Gilmore, Cockton, Shackel (eds). IFIP, North-Holland.
10. Wm Leler. 1987. *Constraint Programming Languages*. Addison-Wesley Publishing Company.
11. John Lyons. 1968. *Introduction to Theoretical Linguistics*, Cambridge University Press, Cambridge.
12. Luis Pineda, Ewan Klein and John Lee. 1988. Graflog: Understanding Graphics through Natural Language. *Computer Graphics Forum*, Vol. 7(2).
13. 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.
14. Luis Pineda. 1992. Reference, Synthesis and Constraint Satisfaction. *Computer Graphics Forum*. Vol. 2, No. 3, pp. C-333 - C-334.
15. L. A. Pineda, "On Computational Models of Drafting and Design", Design Studies, Vol. 14 (2), pp. 124-156. April, 1993.
16. L. A. Pineda, Santana, J. S., Massé, A. "Satisfacción de Restricciones Geométricas: ¿Problema Numérico o Simbólico?", Memorias de XI Reunión Nacional de Inteligencia Artificial, Universidad de Guadalajara, SMIA, pp. 105 - 123, 1994.
17. Luis Pineda. 1996. Graphical and Linguistic Dialogue for Intelligent Multimodal Systems. In *G. P. Facinti and T. Rist editors, WP32 Proceedings, 12th European Conference on Artificial Intelligence ECAI-96,* Hungary, August. Budapest University of Economic Sciences.
18. J. Sergio Santana, Sunil Vadera, Luis Pineda. 1997. The Coordination of Linguistic and Graphical Explanation in th Context of Geometric Problem-solving Tasks, *technical report on the IIE/University of Salford in-house PhD Programme*.
19. Luis Pineda and Gabriela Garza. 1997. A Model for Multimodal Reference Resolution (submitted to *Computational Linguistics*).
20. Raymond Reiter and Alan K. Mackworth. 1987. The Logic of Depiction, *Research in Biological and Computational Vision*, University of Toronto.