Generating Referential Descriptions in Multimedia Environments

Helmut Horacek
Universität des Saarlandes
FB 14 Informatik
D-66041 Saarbrücken, Deutschland
horacek@cs.uni-sb.de

Abstract

All known algorithms dedicated to the generation of referential descriptions use natural language alone to accomplish this communicative goal. Motivated by some limitations underlying these algorithms and the resulting restrictions in their scope, we attempt to extend the basic schema of these procedures to multimedia environments, that is, to descriptions consisting of images and text. We discuss several issues in this enterprise, including the transfer of basic ingredients to images and the hereby reinterpretation of language-specific concepts, matters of choice in the generation process, and the extended application potential in some typical scenarios. Moreover, we sketch our intended area of application, the identification of a particular object in the large visualization of mathematical proofs, which has some characteristic properties of each of these scenarios. Our achievement lies in extending the scope of techniques for generating referential descriptions through the incorporation of multimedia components and in enhancing the application areas for these techniques.

1 Introduction

All known algorithms dedicated to the generation of referential descriptions[1] use natural language alone to accomplish this communicative goal. This task by itself is difficult enough, as a variety of achievements obtained through intensive research demonstrate:

- finding an adequate interpretation of minimality concerning the components of the referring expression to be produced; this interpretation should satisfy computational as well as psychological requirements,
- achieving a reasonable coverage through integrating relations to other referents, controlled recursion, and psychologically motivated concepts, such as inferability and basic level categories into the description, and
- enabling flexible processing through measurements that allow for a widely free descriptor choice and that ensure expressibility of the chosen set of descriptors in natural language terms in a reasonable way.

Despite these achievements, all existing algorithms still have some serious limitations which originate from:

1. *An implicit, simplifying assumption*
   The addressee is not only assumed to understand familiar terms that appear in a description, but he/she is also assumed to be able to recognize the associated object properties under all environmental conditions.

2. *A crucial concept missing*
   In addition to identificational properties, also navigational information would be urgently needed for obtaining comprehensible descriptions (see (Reiter, Dale, 1992)). In larger environments, when referential descriptions could easily become too complex, the algorithms may easily fail to behave adequately.

We believe that extending these algoithms to environments where not only language expressions, but also annotated images contribute to a referential description could not only make many descriptions simpler, but also more reliable (see the first item above) and wider applicable (see the second item above). In our enterprise to adapt the basic schema underlying these algorithms to multimedia environments, we discuss several issues, including

- the transfer of basic ingredients to images,
- the reinterpretation of language-specific concepts,
- matters of choice in the generation process, and
- the extended application potential of multimedia.

This paper is organized as follows. We first review the main concepts shared by the existing algorithms. Then we describe how these concepts can be transferred to images, and we discuss their incorporation into a process schema underlying the existing algorithms. We also outline the potential of extensions obtained through combining identificational and navigational information. Finally, we demonstrate a typical example from the area of proof presentation, which is our intended domain of application.

---

[1] The term 'referential description' is due to Donellan (Donellan, 1966). This notion signifies a referring expression can serve the purpose of letting the hearer identify a particular object out of a set of objects assumed to be in the current focus of attention.

2 Basics of Existing Algorithms

Basically, the issue of producing a distinguishing description requires selecting a set of descriptors according to criteria which reflect humans preferences and verbalizing these descriptors while meeting natural language constraints. Over the last decade, (Dale, 1989, Dale, Haddock, 1991, Reiter, 1990b, Dale, Reiter, 1995), and others[2] have contributed to this issue (see the systems NAOS (Novak, 1988), EPICURE (Dale, 1988), FN (Reiter, 1990a), and IDAS (Reiter, Dale, 1992)). Recently, we have introduced several improvements to these methods (Horacek, 1996, 1997).

In some more detail, the goal is to produce a referring expression that constitutes a *distinguishing description*, that is a description of the entity being referred to, but not to any other object in the current *context set*. A *context set* is defined as the set of entities the addressee is currently assumed to be attending to – the *contrast set* is the same except to the *intended referent*; an equivalent term is the set of *potential distractors* (McDonald, 1981). This is similar to the set of entities in the focus spaces of the discourse focus stack in Grosz and Sidner's theory of discourse structure (Grosz, Sidner, 1986). The existing algorithms attempt to identify the intended referent by determining a set of descriptors attributed to that referent, that is, a set of attributes. Some algorithms also include descriptors in the description that are attributed to other entities related to the original referent, that is, relations from the point of view of the intended referent. Attributes and relations by themselves are mere predicates which still need to be mapped onto proper lexical items, not necessarily in a simple one-to-one fashion. Some of the associated problems and a proposal to systematically incorporate this mapping are described in (Horacek, 1997).

Viewed in procedural terms, the algorithms have to consider three issues:

1. A cognitively motivated pre-selection of descriptors, which is based on psychologically motivated criteria that should reflect human preferences.
2. The ultimate selection of descriptors, which can overrule the cognitively motivated pre-selection of the next descriptor due to linguistic phenomena such as implicature and due to other interference problems with previously chosen descriptors.
3. Adequately expressing the chosen set of descriptors in lexical terms.

---

[2] The approach undertaken by Appelt and Kronfeld (Appelt, 1985a, Appelt, 1985b, Kronfeld, 1986, Appelt, Kronfeld, 1987) is very elaborate but it suffers from limited coverage, missing assessments of the relative benefit of alternatives, and notorious inefficiency.

The first two issues are rather well understood for attributes only, but not so much for relations. The third issue is widely neglected – it is simply assumed that the chosen set of descriptors can be expressed adequately.

For some time, there was a debate about various optimization criteria for comparing the suitability of alternative sets of descriptors, but we feel this issue is settled now in favor of the *incremental algorithm interpretation* (Reiter, Dale, 1992): preferred descriptors are sequentially included in the referring expression to be produced provided each descriptor leads to the exclusion of at least one potential distractor. In comparison to other interpretations, it is the weakest one; it has still polynomial complexity but it is independent of the number of attributes available for building a description.

3 Concepts in Existing Algorithms

Abstracting from details, the algorithms producing a distinguishing description rely on three basic *concepts*:
- the notion of a *focus space*, which delimits the scope in which referents and related entities are to be found,
- the notion of a *descriptor*, by which referents can be described and ultimately identified,
- the notion of a *context set*, which helps distinguishing referents from one another on the basis of sets of descriptors.

In addition, a number of issues are taken into account by these algorithms in one or another way:
- incorporating phenomena, such as *basic-level categories* for objects and *inferability* of properties, such as non-*prototypicality* of mentioned properties,
- *search strategies* and *complexity* considerations, such as interaction between pre-selection and ultimate selection and choices among local referent candidates (selecting among alternative relations as descriptors),
- adequate *expressibility* of the chosen set of descriptors, in terms of naturally composed surface expressions that convey the intended meaning, thereby avoiding, for instance, scope ambiguities or misinterpretations.

In the following, we attempt to transfer the basic concepts to multimedia environments or, in case where this is not possible in a meaningful way, we propose a reasonable reinterpretation better suited to images.

4 Transferring Basic Concepts

As far as the notion of a *focus space* is concerned, the transfer seems to work in a widely straightforward manner. Given some image of a scenery in which some object is to be identified, the focus space is simply the entire picture. There is, however, a principled difference in the way how a focus space is established for concrete images and for abstract language contexts: in a pure

language environment, the conversational setting determines which referents are considered to be within the focus space, which may occasionally be unclear for a few referents. In a multimedia environment, this depends on some application properties. If a specific picture constitutes the situational context, the area and the shape of that picture are precisely determined, as is the associated focus space. Otherwise, the precise boundaries of the image and the associated focus space are subject to minor uncertainties, as in the abstract linguistic context.

The next ingredient to consider are the *descriptors*, which reveal a fundamental difference between texts as an abstract medium and images as a concrete medium. Transferring the notion of a descriptor to images in a direct way would lead to a very unnatural way of communicating identificational information by a picture, especially when several descriptors are to be presented in sequence to achieve the ultimately intended identification goal. Acting this way would mean that all objects to which the first descriptor applies must be highlighted in some way, then all to which also the second descriptor applies, and so forth. Obviously, this procedure would be more confusing than helpful to an observer. Moreover, simply highlighting the intended referent might do the job, but this action alone may not always work satisfactorily, if the intended referent is badly recognizable or even invisible.

Because of the inadequacy of adapting the notion of a descriptor to images in a direct fashion, we consider an alternative way of describing the intended referent: a *region* of the picture where the intended referent can be found or, at least, whose identification helps in finding the intended referent. More precisely, a region can either be the area minimally surrounding a specific object, or it can merely be some connected area, specified by its surroundings or by a pointer to a central position in that area. In the first case, the area is precisely defined, but it may be considerably vague in the second case.

In some sense, regions and descriptors cover the focus space in an orthogonal way: while the former cover a connected area on a picture, the latter typically appear there as a set of islands. As opposed to that, a descriptor covers a connected area in the abstract descriptor-referent space, while a region typically appears there as a set of islands. In some occasions, locality descriptors may do a similar job as regions, but this would probably be less effective in many cases, when multiple locality descriptors are required. As a consequence, the selection of an adequate region differs in some crucial sense from the selection of an adequate descriptor: a candidate descriptor is chosen from a set of distinct alternatives, while determining a candidate region is more a matter of accuracy and precision in terms of appropriately fixing the borderlines of the region which lies around the intended referent

or some other entity related to it. Altogether, a region typically comprises the equivalent of several descriptors as far as the contribution to the identification task is concerned: either a category of the object enclosed by the region, accompanied by a set of further descriptors, if necessary, or a suitable combination of locality descriptors.

Once we have "reinterpreted" the notion of a descriptor in terms of regions as building elements of distinguishing descriptions for images, we have to deal with regions in computing the *context set*. For this concept, extending the algorithm does not prove to be difficult. Since both, descriptors and regions restrict the context set in view of the entire focus space or some previously restricted part of it, although in a complementary way, the computation of the context set modified by a newly introduced region works analogously to the pure language environment.

5 Changes in the Algorithms

When extending the existing algorithms to multimedia environments, we discuss choices between regions and descriptors as well as their coordination in the existing processing schema. We first restrict our considerations to single images – allowing the incorporation of multiple images might easily complicate matters so that temporal presentation aspects additionally come into play, requiring the design of animations. Nevertheless, accomplishing the communicative goal in an environment consisting of a single image only is not always trivial in the sense that the intended referent just needs to be annotated or highlighted in some way. That entity may be invisible or badly recognizable so that pointing at it is simply impossible or unlikely to convey the message properly.

As far as the issues involved in composing a description are concerned, some crucial differences between the media considered exist. Basic-level categories are exclusively relevant for language, and inferability is, apart from language, relevant for abstract images only. The expressibility issue, when being reinterpreted for regions of an image, yields problems, too, but they are entirely different from the expressibility problems in language generation: for images, visibility and various aspects of recognizability, such as sufficient degrees of salience in terms of shape, granularity, and brightness come into play. Judging the adequacy of these aspects is a typical issue in presenting information by a picture and, hence, can be considered the visual counterpart of expressibility on the language side.

When choosing between a descriptor and a region as two candidates to focus on some portion of the environment, some principled preferences seem to be plausible when brevity of the resulting expression is envisioned:

- An 'exact' region, taken by a specific object, is probably better conveyed by the picture component, especially if several similar objects are in the focus of attention.
- However, if the object is either very small (almost invisible) or extremely large (almost covering the entire focus space), choosing language as the medium seems to be more appropriate.
- A 'generic' region, that is, a region which nearly perfectly fits a locality descriptor (see (Gapp, 1995) for an operationalization of degrees of applicability), is better described by language, especially when some other region can be used more beneficially as a component of the referential description.
- For ordinary regions, however, images are generally the preferred medium.

In addition to the choice between a descriptor and a region as the next ingredient for narrowing the focus space, adequate coordination of the participating media is a crucial concern. In our environment, this task is largely simplified because of the restriction to a single image. However, at least some sort of annotation should be given to support the coherence of the overall description. In more complex cases, several regions of an image need to be coordinated as well, which might even require their temporal synchronization.

In addition to dealing with these local preference and choice criteria, we need to incorporate the selection among descriptors and regions into a process where several selections are made in a coordinated way until the intended referent is identified. This process should widely follow the schema based on the incremental algorithm interpretation of minimality of the number of descriptors. By adopting this schema, we maintain the psychologically motivated search strategy and the reasonable computational complexity associated with that schema.

Since descriptors and regions are fundamentally different, a multimedia version of the algorithm requires two choice components to be designed, one for choosing the best descriptor, and the other for choosing the best region. In addition, a referee component could be designed to make the final decision. Such choices could be repeated until a region has outscored its competing descriptor or until the communicative goal is accomplished. This way, a region can describe the intended referent directly or indirectly, that is, in terms of other entities. Because regions may have an entirely different contribution to the restriction of the focus space, a region is usually a proper alternative to a descriptor, rather than a mere substitute.

In view of the environment that is given by a common focus space, that is, by a single image, a simpler strategy may even turn out to be better: a region considered most suitable is selected by the responsible component, and, if necessary, further descriptors are selected until the communicative goal is achieved. Apart from these descriptors, the language part of the description should also entail a reference to the pictorial part, such as an object category or a deictic reference to that region. Even if the region alone already accomplishes the communicative goal, such a reference phrase should be built, in order to clarify the purpose of highlighting the region. The rationale behind this strategy is that in a single image one region is usually sufficient to restrict the context set as much as possible by the pictorial component.

6 Extending the Algorithm's Coverage

So far, we have only considered environments consisting of a single image and language descriptions. If we move on to more complex environments in which several images may contribute to a description, we are definitively leaving the scope of the existing algorithms, since we are not just facing a single focus space, but a set or a chain of focus spaces (when considering only one image at a time). The connection among these focus spaces may vary significantly according to the way how the corresponding images interact. The following constellations seem to be of interest:

1. *An image and some sort of a focused part*
   There could be an image providing a global view of a scenery, combined with images presenting views on portions of that scenery that are invisible on the overview. The subsidiary images may present referents behind an obstacle, or inside some other object, or objects only partially visible in the overview. Moreover, we could be confronted with an image that shows some portion of a larger image (such as a portion of a large map), and the intended referent is located in another part of the whole image. In order to navigate between disjoint portions of a picture, two strategies seem to be promising: either presenting a sequence of pictures that gives some impression of scrolling, or presenting an overview first before moving on to the part that entails the intended referent. In both cases, these images contribute to bridging differences in locality.
2. *An abstracted view and some concrete images*
   There could be an abstract image providing an overview of some sort (such as the map of a city) and several concrete images that refer to one or another part of the abstract image (such as a group of buildings or a square in the city). The abstract image may then be used to direct the addressee's attention to a particular area of the focus space, while the concrete images support the proper identification task.
3. *Images in largely varying degrees of granularity*

There could be an image providing an overview of a large scenery in which individual objects appear in a too small size to be recognizable. In addition to the strategy applied to the abstract overview and the concrete images, a smoother transition seems to be a promising alternative. Depending on the degree of condensation between the overview image and images that present objects in an easily recognizable format, using a few images of intermediate size might be a suitable means to support orientation.

In order to make these concepts more concrete, a lot of testing in connection with concrete applications is required. Moreover, it seems to be much harder to formulate a reasonably concrete schematic procedure and suitable criteria for a multimedia version of the algorithms discussed, because images are associated with a higher degree of freedom than language. However, if we compare the discussion in this section with the original environment underlying the generation of referential descriptions, it becomes apparent that we have left the scope of what is commonly considered as the task of generating referential descriptions in a number of places – but such an effect may easily happen in extending a method to multimedia environments.

7   Our Future Application Area

In the near future, we intend to apply our approach to interface a graphical component by which we can visualize machine-generated mathematical proofs and related data structures. The task of our present interest, the identification of a particular object in the trace of a proof, is one of the issues in presenting mathematical proofs in multimedia environments. In some occasions, even groups of objects and their relations to one another may be subject to identification, which constitutes another kind of extension to the algorithms for generating referential descriptions.

Proof presentation is realized within the mathematical assistant $\Omega mega$ (Benzmüller et al., 1997), an interactive environment for proof development. Within $\Omega mega$, automated prover components such as *Otter* (McCune, 1994) can be called on problems considered as manageable by a machine. The result is a proof tree which needs to be fundamentally reorganized prior to presentation, because the refutation graph underlying the original proof is much too detailed to be comprehensible to humans, even to experts in mathematics. Therefore, an appropriate level of granularity is selected by condensing groups of inference steps to yield proofs built from "macro-steps", which is motivated by rules of the natural deduction calculus (Gentzen, 1935). This is called the *assertional level* and dealt with in detail in (Huang, 1996). A typical example of an assertion level step is e.g., the application of a

lemma. Once a proof is transformed to the assertional level, it can be verbalized suitably by the *Proverb* System (Huang, Fiedler, 1996). Another possibility to present a proof is to visualize the proof tree, which is the kind of presentation we address in this paper.

Even at the assertional level, traces of machine-found proofs may grow very large even for problem of medium complexity. Therefore, a number of measurements to support identification are required, for instance, moving from an overview of the proof tree to a focused part of it. Moreover, moving from abstract to concrete environments may apply here to cases where the object to be identified lies in some detailed information about axioms or theorems, to which some node in the trace gives access.

The following Figures show the trace of a moderately complex proof. The proof demonstrates the truth of the following axiom: the transitive closure of the union of two sets is identical to the transitive closure of the union of the transitive closures of the two sets, in terms of formulas: $(\rho \cup \sigma)^* = (\rho^* \cup \sigma^*)^*$. Figure 1 shows an overview of the whole proof, and Figures 2 and 3 selected portions of it, at a larger size. While individual nodes are still identifiable in the proof tree overview in Figure 1, the recognizability of nodes may easily be lost in larger proof trees, which motivates focusing on tree portions.
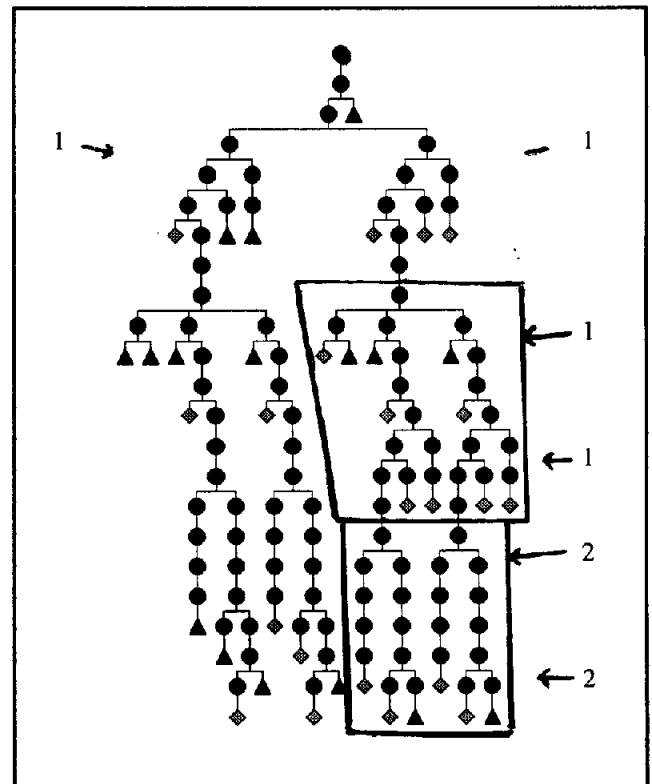

Figure 1: An overview of a proof tree

In these proof trees, a root node represents the lemma to be proved (a root node of a subtree represents some supporting lemma), and the leaf nodes represent assumptions, axioms, or coreferences to specific subtrees in the proof. Moreover, proof derivations join nodes and their successors in upward direction. The geometric figures in the proof tree represent types of nodes: circles stand for ordinary nodes, triangles for assumptions or axioms, and squares for coreferences. The annotations in the Figures are made here by hand, to illustrate focused steps in the proof. In the implementation, a formula associated with an individual node can be viewed by clicking on that node so that the formula appears in a separate window (though in a less convenient predicate-like format rather than in the more common mathematical notation). In addition, the formulas are marked by numbers that also appear in the corresponding node of the proof tree.

As an adds-on to this graphical presentation, we intend to incorporate a variety of interactive explanation facilities. One part of these facilities comprises various sorts of identification issues:

- one specific object in the proof tree,
- some formula or subformula associated with a specific node in the proof tree; this constellation is an instance of a concrete entity associated with some part of an abstract overview – see the second item in the extension categories introduced in Section 6,
- a formula associated with a specific node in the proof tree or some part of it, that is not shown in the visible portion of the tree; this constellation is an instance of a referent that lies outside the scope of the focus space – see the first item in the extension categories,
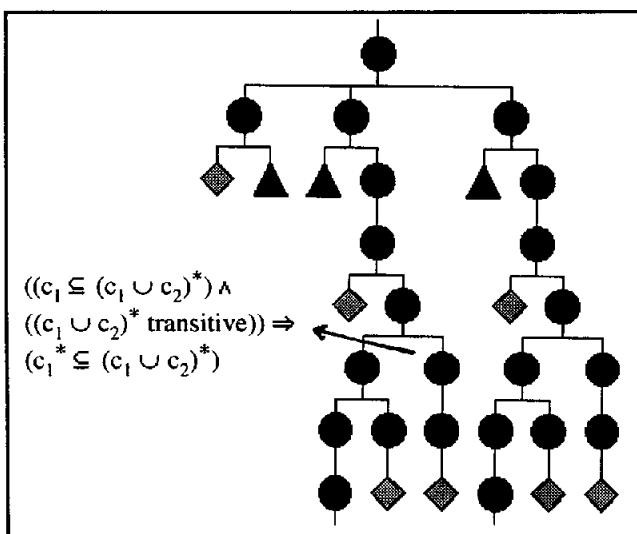
- some part of a formula associated with a specific node in the proof tree, which appears in a too small size to be recognizable; this is an instance of a referent which needs to be zoomed at to be recognizable – see the third item in the extension categories.

Moreover, multiple objects may be subject to any of the above identification issues. In the following, we illustrate these identification categories by a few examples including suitable graphical displays and associated verbal descriptions.

Let us assume that the whole proof tree (as an overview) is in the current focus of attention, and the user asks: "Where is the lemma '$((x \subseteq y) \wedge (y \text{ transitive})) \Rightarrow (x^* \subseteq y)$' used in the proof?" As an answer, the regions where the three instantiations of this lemma appear in the proof are marked (see the arrows labeled by 1 in Figure 1), and their instantiations are given as formulas in the associated verbal description. Moreover, the regions of one or several of these instantiations could be illustrated by a focused picture, such as in Figure 2. A suitable accompanying verbal description would be: "That lemma is applied three times (see the annotations in the overview labeled by 1), one of these instantiations appears in the part proving $(\rho \cup \sigma)^* \supseteq (\rho^* \cup \sigma^*)^*$, where x is instantiated to c_1 and y is instantiated to $(c_1 \cup c_2)^*$, (see the annotation in the enhanced tree portion, corresponding to the tree portions marked by 1 in the overview)."

If this description is followed by a subsequent question "How is the subset definition applied here?", the pictorial presentation needs to move to an adjacent portions of the proof tree, because the referent to be identified lies outside the subtree shown in Figure 2. The overview is then

$((c_1 \subseteq (c_1 \cup c_2)^*) \wedge$
$((c_1 \cup c_2)^* \text{ transitive})) \Rightarrow$
$(c_1^* \subseteq (c_1 \cup c_2)^*)$

Figure 2: A portion of a proof tree showing an axiom
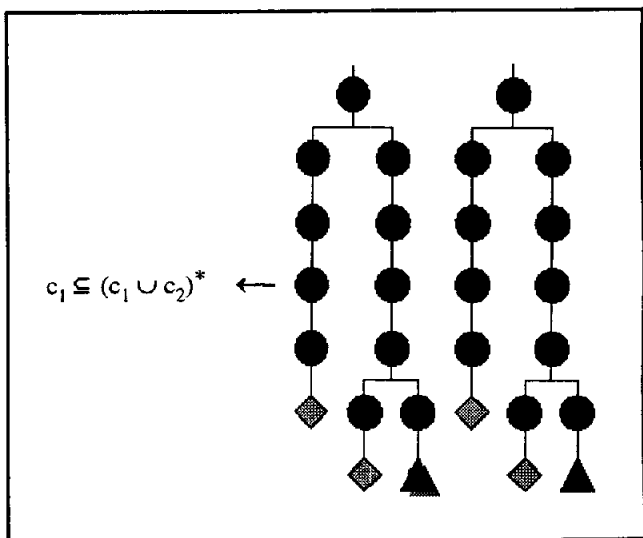

$c_1 \subseteq (c_1 \cup c_2)^*$

Figure 3: A portion of a proof tree showing a definition

shown again, and the annotation in Figure 3 provides additional information, in terms of the instantiations of this definition. A suitable verbalization would be "That lemma is proved in an adjacent part of the tree, where $c_1 \subseteq (c_1 \cup c_2)^*$ is proved, as indicated in the overview (see the annotation labeled by 2 and the tree portion marked by 2 in the overview). The subset definition is instantiated to c_1 and $(c_1 \cup c_2)^*$, respectively."

We believe that these moderate sketches already demonstrate the usefulness of multimedia presentations in the task envisioned. Finally, these examples illustrate the following observations:

- Choices between media become even richer through the possibility to incorporate annotations, which offers itself in the domain of mathematics.
- The identification task is tightly interwoven with providing additional, descriptive information, which we feel to be typical in realistic domains.
- While many of the details in proof presentation are highly domain-specific, the general lines in identifying objects in multimedia environments are valid across a number of domains. However, a characteristic feature that limits the generality and at the same time greatly helps in referring to portions of the proof tree is its strictly hierarchical organization, which may be present in some, but not in many other domains.

In any case, future experience will tell us more about identification techniques in multimedia environments, especially concerning the contribution of each presentation mode and their coordination, as well as about degrees of domain-dependence and independence of the techniques involved.

8 Conclusion

In this paper, we have discussed multimedia extensions to algorithms for generating referential descriptions. In doing this, we have reinterpreted major concepts used in the language-specific algorithms for multimedia environments, which has led to the introduction of regions to identify portions of an image as a counterpart to the language-specific descriptors. In addition to incorporating regions into a descriptions building process, we have categorized some sorts of extensions to the basic form of this process, including the coordination of abstract and concrete images, as well as images of varying size and granularity. We also have exemplified these extensions by applying our techniques to aspects of the presentation of mathematical proofs. Even these preliminary examples demonstrate the enhanced application potential and the extended scope of our method.

References

Doug Appelt. 1985a. Planning English Referring Expressions. *Artificial Intelligence*, 26:1-33.

Doug Appelt. 1985b. Some Pragmatic Issues in the Planning of Definite and Indefinite Referring Expressions. *In 23rd Annual Meeting of the Association for Computational Linguistics*, pages 198-203. Association for Computational Linguistics, Morristown, New Jersey.

Doug Appelt, and Amichai Kronfeld. 1987. A Computational Model of Referring. In Proceedings of the *10th International Joint Conference on Artificial Intelligence*, pages 640-647, Milano, Italy.

Robert Dale. 1988. Generating Referring Expressions in a Domain of Objects and Processes. PhD Thesis, Centre for Cognitive Science, University of Edinburgh.

Robert Dale. 1989. Cooking Up Referring Expressions. In *27th Annual Meeting of the Association for Computational Linguistics*, pages 68-75, Vancouver, Canada. Association for Computational Linguistics, Morristown, New Jersey.

Robert Dale, and Nick Haddock. 1991. Generating Referring Expressions Involving Relations. In Proceedings of the *European Chapter of the Association for Computational Linguistics*, pages 161-166, Berlin, Germany.

Christoph Benzmüller, Lassaad Cheikhrouhou, Detlef Fehrer, Armin Fiedler, Xiaorong Huang, Manfred Kerber, Michael Kohlhase, Karsten Konrad, Andreas Meier, Erica Melis, Wolf Schaarschmidt, Jörg Siekmann, and Volker Sorge. 1997. Omega: Towards a Mathematical Assistant. To appear in Proceedings of *Conference on Automated Deduction*, Perth, Australia.

Robert Dale, and Ehud Reiter. 1995. Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19:233-263.

K. Donellan. 1966. Reference and Definite Description. *Philosophical Review*, 75:281-304.

G. Gentzen. 1935. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift* 39.

Klaus-Peter Gapp. 1995. Efficient Processing of Spatial Relations in General Object Localization Tasks. In Proceedings of the *Eighth Australian Joint Conference on Artificial Intelligence*, Canberra, Australia.

Barbara Grosz, and Candace Sidner. 1986. Attention, Intention, and the Structure of Discourse. *Computational Linguistics*, 12:175-206.

Helmut Horacek. 1996. A New Algorithm for Generating Referring Expressions. In Proceedings of the *8th European Conference on Artificial Intelligence*, pages 577-581, Budapest, Hungary.

Helmut Horacek. 1997. An Algorithm for Generating Referential Descriptions With Flexible Interfaces. In *35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain. Association for Computational Linguistics, Morristown, New Jersey.

Xiaorong Huang. 1996. Translating machine-generated resolution proofs into ND-proofs at the assertion level. In Proceedings of *Pacific Rim Conference on Artificial Intelligence*, pages 399-410, LNAI 1114, Springer.

Xiaorong Huang, and Armin Fiedler. 1996. Presenting Machine-Found proofs. In Proceedings of the *13th Conference on Automated Deduction*, pages 577-581, Budapest, Hungary.

Amichai Kronfeld. 1986. Donellan's Distinction and a Computational Model of Reference. In *24th Annual Meeting of the Association for Computational Linguistics*, pages 186-191. Association for Computational Linguistics, Morristown, New Jersey.

W. McCune .1994. Otter 3.0 Reference Manual and Guide. Technical Report ANL-94/6, Argonne National Laboratory.

David McDonald. 1981. Natural Language Generation as a Process of Decision Making Under Constartints. PhD Thesis, *MIT*.

Hans-Joachim Novak. 1988. Generating Referring Phrases in a Dynamic Environment. In M. Zock, G. Sabah, editors, *Advances in Natural Language Generation*, Vol. 2, pages 76-85, Pinter publishers, London.

Ehud Reiter. 1990a. *The Computational Complexity of Avoiding Conversational Implicatures.* In *28th Annual Meeting of the Association for Computational Linguistics*, pages 97-104, Pittsburgh, Pennsylvania. Association for Computational Linguistics, Morristown, New Jersey.

Ehud Reiter. 1990b. Generating Descriptions that Exploit a User's Domain Knowledge. In R. Dale, C. Mellish, M. Zock, editors, *Current Issues in Natural Language Generation*, pages 257-285, Academic Press, New York.

Ehud Reiter, and Robert Dale. 1992. Generating Definite NP Referring Expressions. In Proceedings of the *International Conference on Computational Linguistics*, Nantes, France.