Exploiting the Addressee's Inferential Capabilities in Presenting Mathematical Proofs

Detlef Fehrer Helmut Horacek
Fachbereich Informatik
Universitat des Saarlandes
D-66041 Saarbriicken, Germany
{fehrer,horacek}@cs.uni-sb.de

Abstract

Proof presentation systems and, in some more general context, many natural language generation systems suffer from a crucial problem: they present too much information explicitly which the intended audience could more naturally infer from a less detailed text. Moreover, proofs in mathematical textbooks make extensive use of building chains of inferences in specialized notations, which is not sufficiently taken into account by proof presentation systems. Encouraged by these observations, we present a model for presenting mathematical proofs that (1) features the implicit conveyance of information through concise texts, (2) organizes major lines in the proof presentation around focused chains of inferences in a specialized notation, (3) can adapt its output to some of the capabilities of its audience. The methods described in this paper allow us to present proofs of moderately complex size in a quality approaching that of proofs found in mathematical textbooks.

1 Introduction

Proof presentation systems and, more generally, many natural language generation systems suffer from a crucial problem: they present too much information explicitly which the intended audience could more naturally infer from a less detailed text. In contrast, mathematical proofs as typically found in textbooks express lines of reasoning in a rather condensed form by leaving out several elementary, but logically necessary inference steps. Moreover, the proofs emphasize conciseness by making extensive use of building chains of inferences in specialized notations, such as series of inequations. However, when presenting a mathematical proof to less trained people, those parts which require increased experience to be understood should be expressed in closer detail.

Motivated by these observations, we have developed a model in which we try to mimic the properties of mathematical textbook proofs to a significant extent. Our model also supports more verbose presentations to meet the needs of formally less trained addressees. It

1. features the implicit conveyance of information through concise texts,
2. organizes major lines in the proof presentation around focused chains of inferences in a specialized notation,
3. can adapt its output to some of the capabilities of its audience.

The paper is organized as follows: After discussing the role of inferences in the larger context of natural language generation, we briefly describe how the results obtained by a theorem prover are prepared for presentation. Then we introduce our inference model which is particularly dedicated to understanding mathematical proofs, and further motivate and describe our user model. Finally, we illustrate our results by an example.

2 Dealing with Inferences in Generation

In the larger context of natural language generation the role of inferences in texts differs significantly across genres. It has not been without reason that the description of database schemata was chosen as the application domain for the first natural language generation system to produce paragraph-length text rather than just sentences [McKeown, 1985]. Texts in this genre are characterized by skillfully ordered facts, each of them providing some new piece of information, with inferences across individual facts being nearly absent. Explanatory texts, on the contrary, are likely to include boring redundancies, when the addressees' inferential capabilities are neglected, such as in [Moore and Paris, 1993].

Only in some more recent approaches, inferences are modelled explicitly [Lascarides and Oberlander, 1992; Green and Carberry, 1994; Zukerman and McConachy, 1993; Horacek, 1997]. These methods, however, are only of limited use for (automatically generated) mathematical proofs. [Lascarides and Oberlander, 1992], like severa! others, address the derivation of implicitly entailed discourse relations, rather than the contextual infer ability of entire propositions, as we do. [Green and Carberry, 1994] focus on inferences that deal with advanced concepts like enablement in everyday situations, which are beyond the kinds of proofs we have investigated. [Zukerman and McConachy, 1993] pertain to handling descriptions rather than argumentation, and the inferential structure used operates on a less uniform level. Only [Horacek, 1997] deals with similar sorts of inferences as we consider here. However, the hypothetical reasoning about sets and preferences aimed at in that work would be an overshot for our enterprise. Therefore, we build a simplified inference model, based on these concepts.

Moreover, the presentation of proofs is oriented on crucial properties of the domain of mathematics:

1. The conceptual information to be presented, i.e. the mathematical proofs, are based on a formal calculus that is well established and extremely detailed.
2. There exist established domain specific presentation techniques that are applied whenever reasonably possible (for instance, chains of equations). The resulting text structure deviates in some aspects from the familiar notion of coherence.
3. The domain concepts used in mathematical proofs are also established and precisely defined as opposed to many other scientific and everyday domains.

The first two properties widely determine the way how the proof presentation is organized, and the third one influences the incorporated inference model. Taken together, these properties significantly facilitate the control over the inference process in our domain, because of the precise and commonly accepted terminology. In order to generalize our inference model to real world domains, concepts for dealing with vague terminology and default expectations need to be added.

3 Preparing a Proof for Presentation

The tasks of organizing the presentation of a proof are realized within the mathematical assistant OMEGA [Benzmiiller *et* a/., 1997], an interactive environment for proof development. Within ΩMEGA, 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 verbalization, which requires two tasks to be accomplished:

- An appropriate level of granularity must be selected on which elements of the proof are communicated.
- A backbone structure must be imposed on the line of proof which is oriented at domain specific techniques, such as building series of inequations, with auxiliary information added appropriately.

The first task addresses a central issue in generating natural language texts, which is to put domain concepts intended to be communicated to humans in a form that suits the structures and concepts underlying natural language, to narrow the "generation gap" [Meteer, 1990]. Domain property 1, as introduced in the previous section, allows one to accomplish this task for mathematical proofs in a principled and uniform way.

The second task imposes a particular sort of structure on the elements of a proof to be presented. Mathematical proofs frequently emphasize chains of inferences, as in "lemma *X*. proposition_1 -> --> proposition_n", where lemma *X* is an additional justification for one of the propositions (cf. domain property 2, as introduced in the previous section). In ordinary texts, breaking the implication chain for explicitly stating the lemma at the precise location where it is referred to is often preferred. In some sense, techniques for presenting mathematical proofs *blur the scope* of lemmata.

Note, that this procedure of organizing a proof's presentation is in some contrast to established ways of text planning. Usually, a coherent text structure tree is composed from some sort of weakly structured data, whereas in our case the original tree structure of the proof is reorganized into some condensed but less explicitly interconnected chunks.

An appropriate level of granularity is selected by condensing groups of inference steps to yield proofs built from "macro-steps"[1]. This is called the *assertion level* and dealt with in detail in [Huang, 1996]. The realization is part of the PROVERB System [Huang and Fiedler, 1997; 1996]. A typical example of an assertion level step is e.g. the application of a lemma.

Once an appropriate level of granularity is chosen, the modified tree needs to be reorganized by building a backbone structure. The fact that proofs containing equations can be more naturally expressed by chains of equations rather than trees, has already been noted by e.g. [Lingenfelder and Pracklein, 1990; Denzinger and Schulz, 1994]. However, these approaches are constrained to purely equational proofs. We have shown that this can be generalized to inequations, for the crucial property of the respective (binary) relation is *transitivity,* which holds as well for inequality [Fehrer and Horacek, 1997]. In the case of equations, one can pursue the strategy of directly appending the two subproofs because of the transitivity of equality. Given proofs of $r = s$ and $s = t$ one thus immediately gets a proof of $r = t$. This is called "flattening" in [Denzinger and Schulz, 1994]. The same

[1]This is motivated by rules of the natural deduction calculus [Gentzen, 1935].

holds for other transitive binary relations, such as inequalities. The necessary reordering can be performed purely automatically, as shown in [Fehrer and Horacek, 1997].

After this structural reorganization is completed, decisions have to be made how the resulting intermediate representation should best be presented to the addressee (cf. next section). Finally, PROVERB [Huang and Fiedler, 1997] produces appropriate surface forms.

4 The Inference Model

Presenting a proof adequately requires an inference model that takes care of at least the following issues:

- Determining trivial, that is, easily inferable, parts of the proof. These should be left implicit in the presentation, because the addressee can be assumed to be able to recover the associated content from the material presented explicitly. This issue aims at reducing the length of a proof presentation.
- Determining key parts which encapsulate the main proof concept. These parts should be highlighted in the presentation in order to guide the attention of the addressee.

Both issues generally aim at discharging the addressee in the comprehension task by taking his inferential capabilities and the associated limitations into account. In accordance with the model presented in [Horacek, 1997], we concentrate our efforts on determining and leaving out easily inferable proof parts in the presentation, which by itself should lead to a significant improvement in comparison to prior approaches.

The crucial question is, what is it that qualifies a (sequence of) step(s) as considered trivial? Logically, given a calculus and a genuine theorem, *every* proof step is trivial, in the sense that the theorem is entailed by the premises and therefore is inferable in the (complete) calculus. However, the fact that the reader's inferences are subject to limited resources, actually *changes the calculus* in which his deductions are carried out. The resulting calculus is in general weaker than the one with unrestricted resources[2], thus becoming incomplete. Therefore, a distinction between trivial and non-trivial inference steps can not be based on the logical calculus alone; it must be grounded on the addressee's acquaintance with the following sorts of concepts:

1. Domain knowledge, which consists of signature (terminology), definitions, axioms, and lemmata.
2. The mechanism underlying an inference step, which is given by the (abstract) calculus.
3. The complexity of an inference step, which is measured in terms of the numbers of premises used and of the number of intermediate steps needed to refer the inference step's conclusion back to known premises (basic axioms or communicated facts).

[2] We do not treat here the possibility that it could as well be stronger, because resource limitations may cause *incorrect* inference steps to be applied.

Note that these concepts, particularly the second and third one, become meaningful in the context of a precise calculus with well defined domain concepts. In order to be credited with the ability to perform a certain inference step, the addressee of a proof presented in partial explicitness must be assumed to be

1. acquainted with the pieces of domain knowledge involved in that inference step,
2. capable of mentally applying the mechanism underlying an inference step in principle, without the need to "guess" an appropriate instantiation of a theorem or axiom,
3. able to effectively perform the inference at hand with reasonable effort (unlike the other two criteria, this feature must be assessed in quantitative terms).

The domain knowledge comprises concepts that are particular to more or less specialized mathematical (sub)theories such as group theory, as well as concepts that are valid over a significantly large set of varying theories, such as the laws of associativity, commutativity etc.

As far as the mechanism underlying an inference step is concerned, we consider a variety of special and in some sense simpler forms of substitution:

1. *Generalizations of concepts*

   Generalizations, such as an assertion holding for semi-groups implies that the same assertion also holds for groups; formally, if $C_1(x)$ and $C_1(x) \rightarrow C_2(x)$ hold, neither $C_2(x)$ nor $C_1(x) \rightarrow C_2(x)$ are expressed explicitly if knowledge about the truth of $C_2(x)$ is required[3].

2. *Some sort of weaker constraints*

   Relations involving a weaker constraint differing from an established one only in a single constant, such as $x > 0$ when $x > 1$ holds; formally, if $R_1(x, c_1)$ holds, and $R_1(x, c_1) \wedge (R_2(c_1, c_2) \rightarrow R_1(x, c_2))$ holds, neither $R_1(x, c_2)$ nor $R_2(c_1, c_2) \rightarrow R_1(x, c_2)$ are expressed explicitly if knowledge about the truth of $R_1(x, c_2)$ is required.

3. *Generalizations of relations*

   Relations whose truth logically follows from known relations, such as a number being even if it can be divided by four. Formally, if $R_1(x, y)$ holds, and $R_1(x, y) \rightarrow R_2(x, y)$ is a theorem, neither $R_2(x, y)$

[3] Since unary predicates can be viewed as *sorts* in a sorted logic this corresponds to simply omitting sort inferences.

nor $R_1(x,y) \rightarrow R_2(x,y)$ are expressed explicity if knowledge about the truth of $R_2(x,y)$ is required.

4. ***Abstraction from individuals***
   Concluding the existence of a proposition abstracted from the individuals it applies to; $P(c_1,..,c_n) \rightarrow \exists x_1,..x_n\, P(x_1,..x_n)$, in formal terms.

5. *Modus Ponens*
   The propositional variant without substitutions.

As far as the choice of appropriate instantiations is concerned, an interesting difference to the real world domain situations treated in [Horacek, 1997] can be observed. In real world domains, a sole generic rule, such as "Group leaders must be assigned to single rooms", is adequate as an explanation if the communicative context allows one to conclude which persons are meant to be group leaders, which is frequently the case[4]. In the context of proof presentation, however, an assertion like "apply the law of associativity" would at best be considered as a hint to construct a partial proof, but not as a proof presentation. Because of that, chains of inequations cannot be shortened with confidence in a presentation, with application of the law of associativity as a special exception (this simply amounts to leaving out parentheses).

As far as assessing the effort associated with actually performing an inference step of some complexity is concerned, we believe that this assessment can only be done on an empirical basis; appropriate measurements can be obtained indirectly by examining proofs in text books with varying styles and intended readers. We have not done this systematically yet, but for the mathematical proofs we have considered so far, no limitation was apparent. Typically, additional assumptions appeared between basic premises and the conclusion of a proof, so that chainable inferences were separated into chunks that are small enough (i.e. so that no further structure imposing processes need to be applied prior to presentation). For more complex proofs, we expect this criterion to become of significantly greater importance.

To facilitate different presentations of the same proof to a varying readership, each step in a sequence of inequations is annotated with a justification, which, corresponding to the user model at hand, can either be explicitly presented or left out. We could for instance say "*F > G*" without any further explanation, or as well "F > G, because of assumption *X*" or "... because of lemma Y"[5]. Unlike for example specific premises, the reference to theorems as justifications can be made by the referential form (which is preferred for non-standard theorems and non-elementary substitutions) or in the generic form (which is preferred for standard theorems). Either variant should be sufficient to convey the whole story, according to the model in [Horacek, 1997].

[4]This is one of the numerous default expectations typical of everyday domains.

[5]If the assumed addressee is even less trained, the application of the transitivity axiom is mentioned explicitly.

5 The User Model

In order to select among presentation variants in a motivated way, we take into account a number of assumptions about the addressee, which reflect the categories introduced in the previous section. The assumptions in our user model are organized in a small set of stereotypes with increasing or complementary coverage, in accordance to mathematical subtheories. Inheritance from general to more special stereotypes is applied, which works in an orthogonal and conflict free way, motivated by the domain properties of mathematics. The following sketches show some elementary stereotypes:

apprentice

knowledge: integers, ordering relations, ...

inferences: special forms of substitutions

mathematical student

knowledge: associativity, monotony, ...

inferences: general form of substitutions

group theory specialist

knowledge: group definition, unit element, ...

Once we will move on to presenting more complex proofs, we intend to incorporate the complexity attributed to inference steps, too. These simple user models already give us the necessary distinctions to motivate the omission of trivial inference steps for the appropriate group, and to produce different presentations according to varying user expertise. We do, however, not attempt to infer (learn) the user's capabilities from his interaction with ΩMEGA. The intended user model is simply activated by choosing appropriate parameters.

6 An Illustrative Example

The example stems from a German textbook on analysis [Liineburg, 1981]. It is taken from the introductory chapter that deals with the algebraic background, namely the theory of ordered fields.

Theorem 1.11 ([Lüneburg, 1981]) *Let K be an ordered field. If $a \in K$, then $1 < a$ implies $0 < a^{-1} < 1$ and vice-versa.*

We constrain ourselves to prove the direction $(1 < a) \Rightarrow (0 < a^{-1} < 1)$. The proof makes use of the lemmas

Lemma 1.5 ([Lüneburg, 1981]) *If R is an ordered ring, then $1 > 0$.* and

Lemma 1.10 ([Lüneburg, 1981]) *Let K be an ordered field. If $a \neq 0 \in K$, then $0 < a$ implies $0 < a^{-1}$ and vice-versa.*

and simply runs (according to [Lüneburg, 1981]):

Proof: Let $1 < a$. According to lemma 1.10 we then have $a^{-1} > 0$. Therefore $a^{-1} = 1a^{-1} < aa^{-1} = 1$.

For this example the proof transformation component [Huang and Meier, 1997] of PROVERB transforms the solution found by OTTER [McCune, 1994] into a structure on the assertion level as shown below. The items shown in boxes represent premises or instantiations of theorems, the encircled ones derived assertions. Inferences are to be read downward.

Reorganizing the proof by building chains of inferences of the same sort splits the proof into four parts, three lemmas and a chain of inequations:

1. The lemma $0 < a$ is proved by the chain
   $0 < 1$ lemma 1.5
   $< a$ premise
2. Directly from this lemma $0 \neq a$ is obtained by using the trichotomy axiom.
3. Following from this, using the lemma again, as well as lemma 1.10, we get $0 < a^{-1}$.
4. The last part is a single chain of inequations:
   $a^{-1} = 1a^{-1} < aa^{-1} = 1$

Next we proceed by eliminating easily inferable parts, which is extremely successful here. $0 \neq a$ is considered inferable from $0 < a$ because the *relation generalization criterion* (see 4.3) applies. Similarly, $0 < a$ is considered inferable from the premise $1 < a$ because the *weaker constraint criterion* (see 4.2) is applied to ordering relations. Thus, neither part 1 nor part 2 has to be mentioned explicitly; only $0 < a^{-1}$ (part 3) retains the state as a lemma, and it complements the main line of the proof (part 4).

For a group theory specialist, the entire proof is verbalized as "$0 < a^{-1}$ follows from $1 < a$ and Lemma 1.10 and $a^{-1} = 1a^{-1} < aa^{-1} = 1$ holds", as the reader is assumed to know the axioms underlying the justifications in each step of this chain of inequations.

It turns out that this presentation comes very near to the book proof. For a less experienced addressee, however, the proof can be is presented as

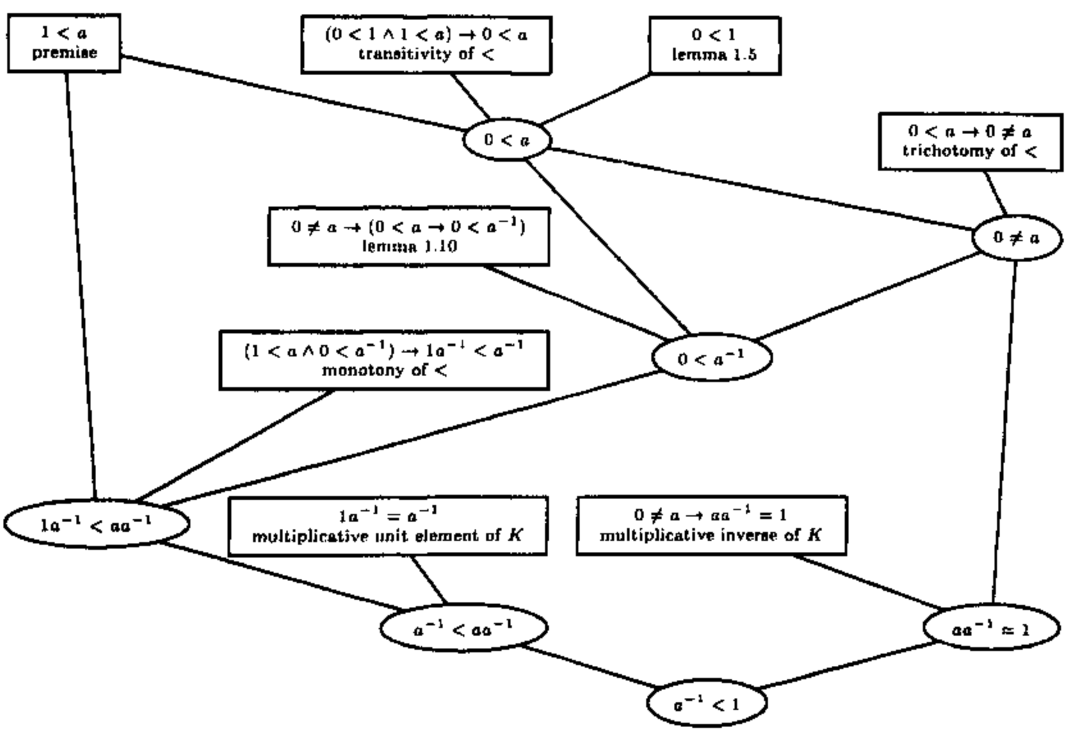
$0 < a^{-1}$ (from $1 < a$ and lemma 1.10).

Therefore

a^{-1}	$=$	$1a^{-1}$	(unit element of K)
	$<$	aa^{-1}	($1 < a$, $0 < a^{-1}$, and monotony)
	$=$	1	(inverse element of K for $a \neq 0$)

holds, q.e.d.

As this example demonstrates, our techniques are already sufficient to obtain a natural proof presentation.



7 Conclusion and Further Work

We have described a model for the presentation of mathematical proofs which constitutes significant progress over previous approaches. Condensed machine-generated proofs are reorganized around focused chains of inferences, whose presentations can be adapted to the addressee's inferential capabilities. An exceptional feature is the inference model, which is not only based on the addressee's knowledge, but also on his inferential skill. The set of examples considered so far (including the introductory chapter from [Llineburg, 1981], plus selected examples from [Deussen, 1971]) demonstrates that the quality of the resulting proof presentations is approaching that of moderately complex proofs found in mathematical textbooks.

Major activities in the future concern the application and adaptation of our model to more complex proofs, in particular with regard to limitations in the complexity of inferences that the addressee can carry out. For the proofs considered so far, reorganizing a proof and eliminating inference steps considered trivial proved to be quite adequate to obtain a satisfactory result. For more complex proofs, we also expect a need to complement this "destructive" strategy by "constructive" ones, such as partitioning the presentation of a proof into meaningful portions, and putting emphasis on its key parts. For usage in real world domains, concepts for imprecise terminology and default expectations have to be added.

References

[Benzmiiller *et al,* 1997] Christoph Benzmuller, Lassaad Cheikhrouhou, Detlef Fehrer, Armin Fiedler, Xiaorong Huang, Manfred Kerber, Michael Kohlhase, Karsten Konrad, Andreas Meier, Erica Melis, Wolf Schaarschmidt, Jorg Siekmann, and Volker Sorge. OMEGA: Towards a Mathematical Assistant. In *Proceedings of CADE 97,* 1997.

[Denzinger and Schulz, 1994] Jorg Denzinger and Stephan Schulz. Analysis and representation of equational proofs generated by a distributed completion based proof system. SEKI Report SR-94-05, Fachbereich Informatik, Universitat Kaiserslautern, 1994.

[Deussen, 1971] Peter Deussen. *Halbgruppen und Automaten.* Heidelberger Taschenbiicher. Springer Verlag, 1971.

[Fehrer and Horacek, 1997] Detlef Fehrer and Helmut Horacek. Presenting Inequations in Mathematical Proofs. In *Proceedings of JCIS 97,* March 1997.

[Gentzen, 1935] Gerhard Gentzen. Untersuchungen uber das logische SchlieBen I. *Mathematische Zeitschrift,* 39, 1935.

[Green and Carberry, 1994] N. Green and S. Carberry. A hybrid reasoning model for indirect answers. In *Proceedings of ACL-94,* Las Cruces, New Mexico, 1994.

[Horacek, 1997] Helmut Horacek. A Model for Adapting Explanations to the User's Likely Inferences. To appear in: *User Modeling and User-Adapted Interaction,* Kluwer Academic, Dordrecht, 1997.

[Huang and Fiedler, 1996] Xiaorong Huang and Armin Fiedler. Presenting machine-found proofs. In *Proceedings of the 13th CADE,* 1996.

[Huang and Fiedler, 1997] Xiaorong Huang and Armin Fiedler. Proof presentation as an application of NLG. In *Proceedings of IJCAI 97,* 1997.

[Huang and Meier, 1997] Xiaorong Huang and Andreas Meier. Natural deduction proofs can be as short as resolution proofs. To appear as SEKI report, 1997.

[Huang, 1996] Xiaorong Huang. Translating machine-generated resolution proofs into ND-proofs at the assertion level. In Norman Foo and Randy Goebel, editors, *Proceedings of PR1CAI-96,* LNAI 1114, pages 399-410. Springer, 1996.

[Lascarides and Oberlander, 1992] A. Lascarides and J. Oberlander. Abducting temporal discourse. In R. Dale, E. Hovy, D. Rosner, and O. Stock, editors, *Aspects of Automated Language Generation,* pages 167 182. Springer, Berlin, 1992.

[Lingenfelder and Pracklein, 1990] Christoph Lingenfelder and Axel Pracklein. Presentation of proofs in an equational calculus. SEKI Report SR-90-15, Fachbereich Informatik, Universitat Kaiserslautern, 1990.

[Liineburg, 1981] Heinz Limeburg. *Vorlesungen iiber Analysis.* BI Wissenschaftsverlag, 1981.

[McCune, 1994] William McCune. Otter 3.0 reference manual and guide. Technical Report ANL-94/6, Argonne National Laboratory, 1994.

[McKeown, 1985] K. McKeown. Discourse strategies for generating natural language text. *Artificial Intelligence,* 27:1-41, 1985.

[Meteer, 1990] M. Meteer. *Expressibility and the problem of efficient text planning.* Pinter, London, 1990.

[Moore and Paris, 1993] J. Moore and C. Paris. Planning text for advisory dialogs: Capturing intentional and rhetorical information. *Computational Linguistics,* 19(4):651~694, 1993.

[Zukerman and McConachy, 1993] Ingrid Zukerman and Richard McConachy. Generating concise discourse that adresses a user's inferences. In *Proceedings of IJCAI-93,* pages 1202-1207, Chamber y, France, 1993.