Abducing Coreference by Model Construction

Peter Baumgartner, *Institut für Informatik, Universität Koblenz–Landau, Rheinau 1, D–56075 Koblenz, Germany. E-mail: peter@uni-koblenz.de*

Michael Kühn, *Institut für Informatik / Computerlinguistik, Universität Koblenz–Landau, Rheinau 1, D–56075 Koblenz, Germany. E-mail: kuehn@uni-koblenz.de*

Abstract

In this paper, we argue that the resolution of anaphoric expressions in an utterance is essentially an *abductive* task following [12] who use a weighted abduction scheme on horn clauses to deal with reference. We give a semantic representation for utterances containing anaphora that enables us to compute possible antecedents by abductive inference. We extend the disjunctive model construction procedure of *hyper tableaux* [3, 14] with a clause transformation turning the abductive task into a model generation problem and show the completeness of this transformation with respect to the computation of abductive explanations. This abductive inference is applied to the resolution of anaphoric expressions in our general model constructing framework for incremental discourse representation which we argue to be useful for computing information updates from natural language utterances.

Keywords: Model construction, abduction, coreference, anaphora resolution

1 Introduction

There is a lot of work in computational linguistics dealing with anaphora, frameworks describing anaphora resolution into the larger context of discourse coherence [11] but also statistical approaches, e.g. [8]. In logical approaches to anaphora, much attention has been paid to the question which expressions *cannot* serve as antecedent referent due to structural constrains e.g. in conditional or negated clauses [10, 13]. The challenging question how to establish the coreference between anaphora and possible antecedents in a logical framework is attacked e.g. in [17, 16]. We follow the line of Hobbs et al. who use a weighted abduction scheme on horn clauses to deal with reference [12], but instead of working on horn clauses, we incorporate an abductive inference into the disjunctive model construction procedure of *hyper tableaux* [3, 14] which we argue to be useful for computing information updates [21] from natural language utterances [15].

The plan of this paper is as follows: After giving some basic notions, we present in Section 3 the hyper tableau calculus [3] that is used in our general model–generating framework for discourse interpretation of Section 4. Section 5 states anaphora resolution as an inference task and examines a solution which finally leads to an abductive restatement of the problem. In Section 6, we define a clause transformation that enables the hyper tableau calculus to compute abductive explanations. It is employed in the model construction algorithm with anaphora resolution in Section 7. Finally, we compare our results to related work and think about future directions.

2 Preliminaries

We are mainly using clause logic. A clause $B_1 \vee \cdots B_n \vee \neg A_1 \vee \cdots \vee \neg A_m$ (for some $m,n \geq 0$), i.e. an implicitly universally quantified disjunction of literals, is also written in implication style as $B_1 \vee \cdots \vee B_n \leftarrow A_1 \wedge \cdots \wedge A_m$; the subclause $B_1 \vee \cdots \vee B_n$ is called the (possibly empty) *head* and $A_1 \wedge \cdots \wedge A_m$ the (possibly empty) *body* of the clause. Occasionally, it facilitates presentation to confuse a clause with an expression $\mathcal{B} \leftarrow \mathcal{A}$, where $\mathcal{B} = \{B_1, \ldots, B_n\}$ and $\mathcal{A} = \{A_1, \ldots, A_m\}$ are multisets of atoms. By $\mathcal{L}$ we denote a given predicate logic language underlying the considered clause sets. We assume that $\mathcal{L}$ is finite, and that $\mathcal{L}$ contains at least one constant symbol. Whenever appropiate, we omit the mentioning of the logic language and assume that $\mathcal{L}$ is implicitely given.

Concerning semantics, we use throughout this paper Herbrand interpretations [6]. Since the domain of any Herbrand interpretation is fixed by the set of $\mathcal{L}$-ground terms, we can uniquely represent a Herbrand interpretation $\mathcal{I}$ as a (possibly infinite) subset of the underlying Herbrand base, i.e. the set of all ground $\mathcal{L}$-atoms. As usual, the members of $\mathcal{I}$ are exactly those atoms that are *true* in $\mathcal{I}$. Evaluation of literals, clauses and clause sets wrt. interpretations is defined as usual, and we write $\mathcal{I} \models \varphi$ iff φ is *true* in $\mathcal{I}$. In particular $\mathcal{I} \models \mathcal{B} \leftarrow \mathcal{A}$ iff every $\mathcal{L}$-ground instance of $\mathcal{B} \leftarrow \mathcal{A}$ is *true* in $\mathcal{I}$.

Reasoning in Herbrand interpretations is quite different from reasoning in "general" interpertations. For instance, suppose that $\mathcal{L}$ contains just one constant a. Then $P(a) \wedge \exists x\, \neg P(x)$ is unsatisfiable wrt. Herbrand interpretations, but it is satisfiable wrt. general interpretations. This example also demonstrates that *skolemization* is *not* a valid principle wrt. Herbrand interpretation, since the skolemized form $P(a) \wedge \neg P(b)$ is both satisfiable and Herbrand-satisfiable.

Of particular interest of this paper is the special case that $\mathcal{L}$ contains only 0-ary function symbols (i.e. constants). Then, since the Herbrand base is finite, any formula can be transformed in principle to propositional logic: any $\exists$-quantified ($\forall$-quantified) formula can be expanded into a disjunction (conjunction) of ground instances by replacing the $\exists$-quantified variables ($\forall$-quantified variables) by constants in all possible ways. More explicitly, any $\exists$-quantified formula $\exists x.\varphi(x)$ can be expanded into the finite disjunction $\varphi(c_1) \vee \ldots \vee \varphi(c_n)$, where $c_1, \ldots, c_n$ are all the constants in $\mathcal{L}$. However, to keep the search space smaller, we propose *not* to expand $\forall$-quantified formulas (e.g. the given clauses of the background theory) in this way, but to treat $\forall$-quantifiers directly in the calculus.

Herbrand interpretations are particularly well suited for our discourse representation tasks for the following reasons: first, in Herbrand interpretations different terms, say "Mia" and "Sally", denote different objects. This property is also known as *unique name assumption*. And second, in Herbrand interpretations *only* those objects exist in the domain that are talked about, i.e. that are contained in $\mathcal{L}$. This property is known as *domain closure assumption*. This is an adequate property when resolving anaphoric expressions, since these refer only to objects introduced so far. However, for indefinite descriptions like "a man" in a non–generic sense as well as for deictic references, it is reasonable to introduce new elements into the discourse universe.

3 Hyper tableau calculus

The purpose of this section is to recall the calculus of *hyper tableaux*, which is the basic inference engine in this paper. We begin with some usual model-theoretical and syntactical definitions. However, we assume the reader to be familiar with classical first–order logic and refer to standard text books such as [6].

The calculus of *hyper tableaux* has been introduced in [3] as a forward chaining, model–generating proof procedure for clause logic. Therefore, we suppose that the theory Σ is given as a set of clauses. We assume in the rest of this paper, that the given Herbrand universe is finite, i.e. it contains only 0-ary function symbols. Then, according to the comments of the last section on a finite Herbrand base, instead of skolemization we sometimes also expand a $\exists$-quantifed formula into a finite disjunction.

To present the calculus, we need the notion of a *pure clause*: A clause $C = B_1 \vee \cdots \vee B_n \leftarrow A_1 \wedge \cdots \wedge A_m$ is called *pure* iff no atoms in the head of a clause share variables, i.e. $Var(B_i) \cap Var(B_j) = \emptyset$ for each i, j with $i \neq j$. A substitution π is a *purifying substitution* for a clause C iff $C\pi$ is pure. Obviously, every non-pure clause can be turned into a pure instance thereof by application of some purifying substitution.

In this paper, we identify the commonly used tree structure of tableaux with sets of branches, where a branch is a finite set of literals. More specifically, hyper tableaux Θ are defined as just the objects which can be produced by the proof procedure below. It employs the following definitions: given a branch β, the *extension* of β with a disjunction $B_1 \vee \cdots \vee B_n$ of atoms is defined by $\eta = \bigcup \{\{B_i\} \cup \beta\}$. The case $n = 0$ is permissible and results in $\emptyset$. In this case we say that β is *closed*.

We say that a clause $B_1 \vee \cdots \vee B_n \leftarrow A_1 \wedge \cdots \wedge A_m$ is *applicable* to β with substitution σ iff there are atoms $A'_1, \ldots, A'_m$, each being a new variant of some atom in β, such that $A_i\sigma = A'_i\sigma$[1]. We say that a disjunction $B_1 \vee \cdots \vee B_n$ of atoms is *redundant* in β iff there is a variant B' of an atom $B \in \beta$ and a substitution δ such that $B'\delta = B_i$, for some i. In other words, some head literal has to be subsumed by a variant of an atom from B.

These definitions are purely syntactical, as is feasible for a proof procedure, but the intuition is justified semantically: for a branch β we define $[\![\beta]\!]$ to be the set of ground atoms of all elements of β. Thus, $[\![\beta]\!]$ is nothing but a Herbrand interpretation represented compactly by the "patterns" in β. But then, the applicability condition just means, if it applies, that there is at least one instance $(\mathcal{B} \leftarrow \mathcal{A})\sigma$ of the considered clause $\mathcal{B} \leftarrow \mathcal{A}$ the body of which is satisfied. In fact, the body of any further instance $(\mathcal{B} \leftarrow \mathcal{A})\sigma\pi$ is satisfied in $[\![b]\!]$. Thus, the clause $(\mathcal{B} \leftarrow \mathcal{A})\sigma\pi$ possibly is *false* in $[\![b]\!]$, but it certainly is *true* if $\mathcal{B}\sigma\pi$ is redundant in $[\![b]\!]$.

Now, these considerations can be used to intuitively understand the following *hyper tableau procedure* as an attempt to generate a model for the given clause set Σ:

$\Theta := \{\{\}\}$
while $\Theta \neq \emptyset$ **do**
    select $\beta \in \Theta$
    if for some clause $(\mathcal{B} \leftarrow \mathcal{A}) \in \Sigma$ and for some substitutions σ, π:
      $\mathcal{B} \leftarrow \mathcal{A}$ is applicable to β with σ, and
      $\mathcal{B}\sigma\pi$ is pure, and
      $\mathcal{B}\sigma\pi$ is not redundant in β
      then
          $\eta := \bigcup_{B \in \mathcal{B}} \{\{B\sigma\pi\} \cup \beta\}$;; Extension of β by $B\sigma\pi$.
          $\Theta := (\Theta \setminus \{\beta\}) \cup \eta$
       else return "β is a model for Σ"
    fi

[1] Any *most general* σ, which is sufficient, can be computed by incrementally unifying $A_i\sigma_1\cdots\sigma_{i-1}$ with A'_i by σ_i and finally setting $\sigma = \sigma_1\cdots\sigma_m$.

13 **<u>od</u>**
14 return "Σ is unsatisfiable"

Some comments are due: in line 1 the tableau under construction is initialised with one single model candidate, namely $\{\}$. In line 3, an arbitrary model candidate β is selected. If the condition of the if-statement (line 4–7) does *not* apply, then, according to the semantics just explained, any candidate $(\mathcal{B} \leftarrow \mathcal{A})\sigma\pi$ to be falsified by $[\![b]\!]$ certainly is *true* in $[\![b]\!]$. In other words, the procedure may stop and report a model in line 11. In this case the branch β is called *finished*.

Otherwise, if the condition applies, then the clause instance $(\mathcal{B} \leftarrow \mathcal{A})\sigma\pi$ identified in the condition possibly is *false* in $[\![\beta]\!]$. Therefore, we have to modify β such that $(\mathcal{B} \leftarrow \mathcal{A})\sigma\pi$ becomes satisfied. To do so, in line 9 and 10, β is replaced by the posibilities to assign **true** to one of the head literals (i.e. extension). Then, the procedure enters the loop again to find possibly more *false* clauses.

The purifying substitution π (cf. line 6) has to be applied for soundness reasons, assuring that reporting "unsatisfiable" in line 14 is indeed correct. For example, the extension of the empty branch $\{\}$ by the clause $\forall x,y\ P(x,y) \vee Q(x)$ has to yield the branch set, say, $\{\{P(a,y)\},\ \{Q(a)\}\}$ by means of the purifying substitution $\pi = \{a/x\}$, because, essentially, $\forall x,y\ (P(x,y) \vee Q(x))$ does not entail $(\forall x,y\ P(x,y)) \vee \forall x\ Q(x)$ (but $(\forall y\ P(a,y)) \vee Q(a)$ is entailed).

This procedure is shown to be sound and refutationally complete in [3].

4 Model construction for discourse representation

In general, the term *discourse model* is used to denote an abstract representation of a discourse. In this work, we use this term in the stricter sense of a logical model for a set of formulas representing the discourse to be interpreted, more specific a Herbrand interpretation, i.e. a set of relations between involved individuals or discourse referents, that satisfies the formulas representing the given utterances. Besides the advantage that logical models offer a natural way to represent discourses described by the logical semantic representation of utterances, in this section we argue for a model–oriented discourse interpretation because of certain computational and inferential features model generation has to offer concerning the *incrementality* and the need for dealing with *ambiguities* of natural language discourses [13].

Natural language utterances are often ambigue at various levels, beginning with homonymies at the lexical level up to quantifier raising problems at the semantic structural level, and thus any approach to natural language processing has to provide techniques to deal with this indefiniteness that may lead to exponentially many discourse interpretations. Model–based deduction offers an approach to deal with several alternative interpretations without the need to represent them all at once by restricting the search space to only one model candidate at a time. Thus, model generation already offers a solution to deal with many alternative readings of a natural language utterance without forcing a combinatorial explosion of readings to be considered.

Another inherent feature of natural language discourses is the need of incremental processing as texts or discourses are given in sequence and often only can be disambiguated by checking certain constraints on the background of former utterances. Two of these constraints are *consistency* and *informativity* [20, 4]: A reading represented by φ is *inconsistent* if and only if $\neg\varphi$ logically follows from the former discourse represented by Φ with a background theory Ψ encoding suited world knowledge. A reading φ is *non–informative* if and only if φ already logically follows from $\Psi \cup \Phi$. Readings corresponding to one kind of these definitions should be rejected as discourse continuations as they are violating conversational maximes [9].

We illustrate the construction of discourse models satisfying consistency and informativity by two examples where coreference is indicated by same indices.

$$\text{“Mia}_1\text{’s husband}_3 \text{ loves Sally}_2\text{. She}_i \text{ is unmarried.”} \tag{4.1}$$

Assume that we have already identified Mia's husband and found out that “She” is either “Mia” or “Sally”, i.e. i equals 1 or 2. Let the background theory contain the knowledge that someone who has a husband is married and that no one is at the same time married and unmarried: $\Psi = \{Husband(x,y) \rightarrow Married(y), Married(z) \wedge Unmarried(z) \rightarrow \bot\}$. The representation of the first sentence is $\Phi = \{Husband(h,m), Loves(h,s)\}$. Now, the second sentence has two different readings: one where “She$_i$” is mapped to Mia and another reading with the mapping to Sally. We represent the two readings of the second sentence by the formulas $\varphi = \{Reading_{1.1} \vee Reading_{1.2}, Reading_{1.1} \rightarrow Unmarried(m), Reading_{1.2} \rightarrow Unmarried(s)\}$. The figure below depicts a hyper tableau constructed by the model construction procedure applied to this clause set. It consists of two model branches, and the shared atoms are written in the middle.

The model branch representing the first reading is discarded i.e. removed from the set of interpretations in this hyper tableau, as the contained information that Mia is unmarried contradicts the former information that Mia is married by virtue of the background theory. This is precisely what the consistency condition for new utterances demands. Now, let's have a look at a similar example concerning informativity:

$$\begin{array}{c|c} \multicolumn{2}{c}{Husband(h,m)} \\ \multicolumn{2}{c}{Married(m)} \\ \multicolumn{2}{c}{Loves(h,s)} \\ Reading_{1.1} & Reading_{1.2} \\ Unmarried(m) & Unmarried(s) \\ \star & \end{array}$$

FIG. 1. Hyper tableau for (4.1)

$$\text{“Mia}_1\text{’s husband}_3 \text{ loves Sally}_2\text{. She}_i \text{ is married.”} \tag{4.2}$$

Again, the second sentence has two readings, mapping “she$_i$” to Mia resp. Sally, where the first reading should be discarded as it is *non–informative*, i.e. it does not add any new information to the considered discourse model branch since the information $Married(m)$ is already contained in the left branch. Technically, the clause head $Married(m)$ is redundant in the left branch. To deal with non–informativity in this context, a method is needed to achieve that the currently considered reading is rejected (i.e. the resulting branch of updating it with this reading is discarded) if the reading is already entailed by the discourse history so far. The solution we propose is to incorporate a consistency checking operator $\Diamond\varphi$ similar to the one used in *update semantics* [21] with the underlying meaning that φ is consistent wrt. the model under consideration. Using the $\Diamond$ operator, the informativity constraint is formalized on an axiomatic level by saying that a formula φ is informative if and only if it holds that $\Diamond\neg\varphi$, i.e. it is possible that the negation holds. For space reasons, we refer the reader to [15] for further details how this technique can be incorporated into hyper tableaux.

Thus, model construction gives a well–suited framework for incrementally interpreting utterances, representing several alternative readings, checking additional constraints and inferring new knowledge. But, returning again to the representation of utterance (4.2). How can the coreference between “She” and “Mia” resp. “Sally” be established in this model generating approach? This is the topic of the following section.

5 Abducing coreference

In the *predicate logic with anaphora (PLA)* [7], pronouns are represented by a new type of terms p_i added to the logical language which are interpreted by a special assignment function mapping the i–th pronoun term to the last but i–th existentially introduced term. In our treatment of anaphoric expressions, we go one step further by formalizing also the antecedent assignments in first–order axioms and thus enable an abductive computation of possible anaphora antecedents. Further, the antecedent assignments are not fixed wrt. the reverse order of occurences in discourse, but are established by the possible abductive explanations.

Consider again the second sentence of utterance (4.2). It is represented in predicate logic with anaphora by $Married(p_0)$ where p_0 is a special pronoun term mapped to the last introduced discourse referent. The formalization we use is the clause $Anaph_0(x) \rightarrow Married(x)$ where the intuitive meaning of $Anaph_0(x)$ is "this anapher is mapped to x" (not necessarily the last term introduced). Here, $Anaph_0$ is a new predicate name that has not yet occured so far. The information that the antecedent for this pronoun actually exists and that "she" corresponds to a female referent then is encoded by the formula $\exists x.(Anaph_0(x) \wedge Female(x))$.

Thus, our representation of utterance (4.2) is as follows, stating that the anaphoric expression points to a female referent which is married:

$$a)\ \exists x.(Anaph_0(x) \wedge Female(x)) \text{ and } b)\ \forall x.(Anaph_0(x) \rightarrow Married(x)) \tag{5.1}$$

There is a crucial difference in purpose between the existentially quantified formula (5.1.a) and the clause in (5.1.b): while the latter expresses the new fact contained in utterance (4.2) that the referent of "she" is married, formula (5.1.a) contains the information needed to find the antecedent of this anaphoric expression. Now, one approach to deal with these predicates might be stated as follows: to resolve an anaphoric expression, take all antecedent candidates and reject those leading to a contradiction.

E.g. for the utterance in (4.2), we take the disjunction $Reading_{1.1} \vee Reading_{1.2} \vee Reading_{1.3}$ plus $Reading_{1.1} \rightarrow Anaph_0(m)$, $Reading_{1.2} \rightarrow Anaph_0(s)$ and $Reading_{1.3} \rightarrow Anaph_0(h_0)$:

$Reading_{1.1}$	$Reading_{1.2}$	$Reading_{1.3}$
$Anaph_0(m)$	$Anaph_0(s)$	$Anaph_0(h_0)$
$Female(m)$	$Female(s)$	$Female(h_0)$

FIG. 2. Alternative readings for (4.2)

Now, if the background theory contains the knowledge that no one can be female and male at the same time and that a husband is male, the last model branch is discarded and the desired pronoun resolvants survive.

Unfortunately, the check for consistency only works if the additional information about the anaphor is mutually exclusive like it is for the genus in "he, she, it". If we have to resolve richer anaphoric expressions like definite references, consistency is not enough. Consider utterance (5.2).

"A politician chased a gangster. The criminal died." (5.2)

The constructed discourse models (ignoring tense information) with the background knowledge that gangsters are criminals, i.e. $Gangster(x) \rightarrow Criminal(x)$, is displayed in the figure below. The first model branch is certainly not intended by utterance (5.2), although $Politician(p)$ and $Criminal(p)$ are not inconsistent *per se*.

The mistake is here that it is not sufficient for antecedent hypotheses to preserve consistency, but that they should also *imply* the definite description that has been used. More formally, in this case we look for one instance of $Anaph_1(x)$ such that $\exists x.(Anaph_1(x) \wedge Criminal(x))$ logically follows. In other words, in order to resolve this anaphoric expression, we have to *abduce* potential antecedent candidates.

We are thus turning now to a form of "abduction" suitable for our needs. In principle, we can use the following common definition [5]:

$$
\begin{array}{c}
Politician(p) \\
Gangster(g) \\
Chase(p,g) \\
Criminal(g) \\
\begin{array}{c|c}
Reading_{2.1} & Reading_{2.2} \\
Anaph_1(p) & Anaph_1(g) \\
Criminal(p) & Die(g) \\
Die(p) &
\end{array}
\end{array}
$$

FIG. 3. Hyper tableau for (5.2)

DEFINITION 5.1 (Abductive Explanation)
Let Ω be a set of sentences called *observations*, Σ a set of sentences called *background theory* and Δ a set of sentences called *allowable hypotheses*. A conjunctively understood set of sentences Γ is an *abductive explanation* of Ω if and only if it satisfies the following three criteria:

1. $\Gamma \subseteq \Delta$,
2. $\Sigma \cup \Gamma \models \Omega$, and
3. $\Sigma \cup \Gamma$ is satisfiable.

If Ω has a non–zero finite number of explanations $\Gamma_0, \ldots, \Gamma_n$, then the *cautious explanation* $\hat{\Gamma}$ is their disjunction $\hat{\Gamma} = \bigvee_i \Gamma_i$.

Let P be a predicate symbol. The abductive explanation Γ is called *P-minimal* if additionally there is no explanation Γ' that contains strictly less P-atoms than Γ.

It should be emphasized that items (2) and (3) above are to be understood wrt. Herbrand-interpretations, but not as usual wrt. general interpretations (cf. Section 2 for a motivation on Herbrand interpretations).

Now, returning to our example, how are Ω, Σ and Δ instantiated for utterance (5.2)? As said before, the observation that has to be explained is the information that there is a discourse referent available who is a criminal:

$$\Omega = \exists x.(Anaph_1(x) \wedge Criminal(x))] \tag{5.3}$$

The allowable hypotheses or *abducibles* are all ground $Anaph_1$–atoms since this is the discourse referent we want to determine, so

$$\Delta = \{Anaph_1(p), Anaph_1(g)\}.$$

Let the background theory Σ in this case contain the knowledge given by the first sentence of utterance (5.2), namely $Politician(p)$, $Gangster(g)$ and $Chase(p,g)$ plus the rule that gangsters are criminals, $Gangster(x) \rightarrow Criminal(x)$

As explained, any explanation for anapher resolution should be *true* of *exactly one* $Anaph_1$-atom. In order to accomplish this, we use the minimality criterion of Definition 5.1, and furthermore accept only $Anaph_1$-minimal explanations consisting of *exactly one* $Anaph_1$-atom. Notice that here we make use of Herbrand-interpretations: since different constants denote different objects of the universe, we can be sure that if the anapher can be resolved at all with one discourse referent, the explanation must be a singleton.

Now, how do the abductive explanation for Ω look like? According to property 1 in Definition 5.1 there are four candidates for Γ, namely

$$\Gamma_1 = \{\} \qquad \Gamma_3 = \{Anaph_1(g)\}$$
$$\Gamma_2 = \{Anaph_1(p)\} \qquad \Gamma_4 = \{Anaph_1(p), Anaph_1(g)\}$$

The candidate Γ_1 is excluded immediately, because then $\exists x.Anaph_1(x)$ in Ω would not be entailed by $\Sigma \cup \Gamma_1$ as required by property 2 in Definition 5.1. The candidate Γ_4 is not an explanation either, because it does not satisfy the mentioned minimality criterion. The candidate Γ_2 is excluded, since it does not entail with Σ the last part of the observation Ω, namely $\exists x.Criminal(x)$. Thus, only $\Gamma_3 = \{Anaph_1(g)\}$ is a minimal (and also cautious) explanation of Ω. It is also the intuitively expected one, satisfying as well the existential observation $\exists x.Anaph_1(x)$ as well as the information the information that g is a criminal.

If we had also been given the fact that the politician is corrupt and therefore a criminal, too, i.e. $Criminal(p)$, then Γ_2 would have also been a minimal explanation and $\Gamma_2 \vee \Gamma_3$ the cautious explanation for Ω. But in contrast, given the utterance "A murderer shot a gangster. The criminal slept.", the reading where "the criminal" is mapped to the murderer would be discarded by inconsistency as explanation, if it is derivable from the background theory Σ that no one can shoot and sleep at the same time. Once the referents are identified for the desired cautious explanation, the discourse models constructed so far are extended by this possibly disjunctive formula and deductively closed, yielding a set of updated model branches. All of them would be kept, until there are good reasons to give up one of them later. In the following section we define a calculus that allows to compute explanations. It is the base of our main algorithm in Section 7 which realizes the incremental consistency principle.

6 Computing abductive explanations by hyper tableaux

We are now turning to the question how to compute abductive explanations in our hyper tableau framework. In order to avoid blind guessing of abductive explanations, we transform Σ and Ω in a first step replacing each abducible literal in a clause by it's complementary literal. The underlying motivation for this is that abductive reasoning can be understood as *modus ponens* in the wrong direction. So, instead of assuming that the abducible facts are already stated by the theory Σ, the generated hyper tableaux for the transformed clause set Σ^Δ will contain hypotheses, which of the abducibles may serve as an explanation for the observations at hand.

For this transformation, we assume that a set of *abducible predicate symbols* $\mathcal{P}^\Delta$ is given, taken from the language $\mathcal{L}$ under consideration, and that the set of abducibles Δ is given as the set of all ground atoms with a predicate symbol of $\mathcal{P}^\Delta$. One trivial solution would be to guess for given Ω, Σ and Δ explanation candidates $\Gamma \subseteq \Delta$ and check for the properties stated in Definition 5.1. A drawback of this approach is its undirectness in generating the explanation candidates $\Gamma \subseteq \Delta$. We circumvent this problem by transforming Σ in a first step and then generating candidates with hyper tableaux in a more directed fashion.

Now, the announced transformation renames the literals with abducible predicate symbols in a clause such that every (possibly non–ground) atom A with a predicate symbol $P \in \mathcal{P}^\Delta$ in the body (resp. head) of C is moved to the head (resp. body) as $A^\neg$ where $A^\neg$ contains the same argument terms as A but a different predicate symbol $P^\neg$ that is new in $\mathcal{L}$.

DEFINITION 6.1 (Abducibles renaming)
Let $C = \mathcal{B} \leftarrow \mathcal{A}$ be a clause from Σ and let $\mathcal{P}^\Delta$ be a set of abducible predicate names. For an

atom $A = P(t_1,\ldots,t_n)$, let $\mathrm{Pred}(A) = P$ and $A^{\neg} = P^{\neg}(t_1,\ldots,t_n)$ with a new predicate symbol $P^{\neg}$. The *abducibles renaming* of C wrt. $\mathcal{P}^{\Delta}$ is $C^{\Delta} = \mathcal{B}^{\Delta} \leftarrow \mathcal{A}^{\Delta}$ where

$$\mathcal{B}^{\Delta} = (\mathcal{B} \setminus \{B \mid \mathrm{Pred}(B) \in \mathcal{P}^{\Delta}\}) \cup \{A^{\neg} \mid A \in \mathcal{A} \text{ with } \mathrm{Pred}(A) \in \mathcal{P}^{\Delta}\} \text{ and}$$
$$\mathcal{A}^{\Delta} = (\mathcal{A} \setminus \{A \mid \mathrm{Pred}(A) \in \mathcal{P}^{\Delta}\}) \cup \{B^{\neg} \mid B \in \mathcal{B} \text{ with } \mathrm{Pred}(B) \in \mathcal{P}^{\Delta}\}.$$

For a clause set Σ, the abducibles renaming Σ^{Δ} is defined as the renaming of all its members.

Note that by a suitable change of the language $\mathcal{L}$ to $\mathcal{L}^{\Delta}$, the transformed objects are again clauses. Henceforth, when the hyper tableau calculus is applied to a renamed clause set, always the modified language $\mathcal{L}^{\Delta}$ is meant. Further, we will also make use of the fact that for any interpretation $\mathcal{I}$ fitting to $\mathcal{L}$ there is an interpretation $\mathcal{I}^{\Delta}$ fitting to $\mathcal{L}^{\Delta}$ with $\mathcal{I}^{\Delta}(A^{\neg}) = \mathrm{true}$ if and only if $\mathcal{I}(A) = \mathrm{false}$ such that holds: $\mathcal{I} \models C$ iff $\mathcal{I}^{\Delta} \models C^{\Delta}$.

Now we can turn to the computation of abductive explanations. As said at the beginning, we assume in the following that the language $\mathcal{L}$ contains only constants and thus the Herbrand base is finite. For a set of observations Ω, let $\overline{\Omega}$ denote the negated set of observations transformed into clausal normal form where existentially quantified formulas are translated into their corresponding finite disjunction as described in Section 2. Then, the abductive explanations for a set of observations Ω, a theory Σ and a set of abducible predicate symbols $\mathcal{P}^{\Delta}$ are computed by constructing a finished hyper tableau Θ for $(\Sigma \cup \overline{\Omega})^{\Delta}$ and to collect from every branch of Θ one renamed abducible $A_i^{\neg}$. An abductive explanation then is obtained by taking a ground–instantiated conjunction of the corresponding non–renamed abducibles A_i.

DEFINITION 6.2 (Abductive explanation cut)
Let Σ be a theory in clausal normal form, $\overline{\Omega}$ the negation of the set of observations in clausal normal form and $\mathcal{P}^{\Delta}$ a set of abducible predicate symbols.

If Θ is a finished hyper tableau with n open branches $\beta_1,\ldots,\beta_n$, then a set $\mathcal{C} = \{A_1,\ldots,A_m\}$ with each $\mathrm{Pred}(A_j) \in \mathcal{P}^{\Delta}$ and $m \le n$ is said to be an *abductive explanation cut* (or *AE–cut*) of Θ iff for each $\beta_i \in \Theta$ there is a corresponding $A_j \in \mathcal{C}$ such that $A_j^{\neg} \in \beta_i$.

That is, an AE-cut is obtained from Θ by picking one renamed $\mathcal{P}^{\Delta}$-atom from each open branch in Θ and stripping off the negation symbol. The following theorem states that AE–cuts deserve their name as each abductive explanation corresponds to an instantiation of a computed AE–cut.

THEOREM 6.3 (Abductive explanations in a hyper tableau)
Let $\Gamma \subseteq \Delta$ be a minimal abductive explanation of Ω wrt. Σ and $\mathcal{P}^{\Delta}$. If Θ is a finished hyper tableau for $(\Sigma \cup \overline{\Omega})^{\Delta}$, then there is an AE-cut $\mathcal{C}$ of Θ such that $\mathcal{C}\gamma = \Gamma$ for some ground substitution γ of $\mathcal{C}$.

Notice that AE-cuts need not necessarily exist, as there might be an open finished branch that contains no single negated $\mathcal{P}^{\Delta}$ atom. It can be shown that in this case there is *no* abductive explanation for Ω then, as the existence of an open branch indicates that the negation of Ω is consistent with Σ, whatever Γ is chosen, and hence Ω cannot be entailed as required.

7 Coreference and model construction

In this section we show how the general solution to abductive reasoning with hyper tableaux presented in Section 6 fits to the problem of anaphora resolution in our framework of discourse interpretation by model construction. The anaphora resolution problem is stated as a abductive problem instance with specific properties. One interesting property is that the abducible renaming only occurs in the observation formulas. Further, at last we answer the soundness criterion left open at the end of the last section in our algorithm for anaphora resolution.

Consider again utterance (5.2), “A politician chased a gangster. The criminal died.”. Assume that the logical representation of the first sentence has been given by

$$\Sigma = \{Politician(p), Gangster(g), Chase(p,g)\}$$

and we have the world knowledge $\Phi = \{Gangster(x) \rightarrow Criminal(x)\}$. As argued in Section 5, the second sentence should be represented by an abductive observation,

$$\Omega = \{\exists x.(Anaph_1(x) \wedge Criminal(x))\}$$

and another formula representing the new content of the sentence $\varphi = \{Anaph_1(x) \rightarrow Die(x)\}$.

The hyper tableau constructed for the first sentence consists of only one model branch, namely $\beta = \{Politician(p), Gangster(g), Chase(p,g), Criminal(g)\}$. $\overline{\Omega}$ is the negation of Ω in clausal normal form, i.e. $\overline{\Omega} = \{Anaph_1(x) \wedge Criminal(x) \rightarrow \bot\}$. Since $Anaph_1$ is the only abducible, we have $\overline{\Omega}^\Delta = \{Criminal(x) \rightarrow Anaph_1^{\neg}(x)\}$. Σ^Δ equals Σ, as $Anaph_1$ is a new predicate symbol introduced in the semantic representation of the second sentence and does not occur in Σ. Now, instead of constructing again a hyper tableau for Σ, the discourse model β constructed so far is reused for computation of the abductive explanation cuts. For this example, $\overline{\Omega}^\Delta$ contains only one new clause that can be applied to β, yielding $\beta' = \beta \cup \{Anaph_1^{\neg}(g)\}$, so the only AE–cut is $\mathcal{C} = \{Anaph_1(g)\}$.

In the following algorithm the abductive approach to anaphora resolution is incorporated into the discourse model construction outlined in Section 4.

ALGORITHM 7.1 (Abducing coreference in discourse models)
Let β be the selected discourse model branch in a tableau Θ finished for the discourse representation Σ so far and Ω be the observation to be explained in a new utterance with an anaphoric expression represented by the predicate $Anaph_i(x)$ and the new content φ.

1 Let $\mathcal{P}^\Delta := \{Anaph_i\}$ and Ω^Δ be the abducible renaming of $\overline{\Omega}$ wrt. $\mathcal{P}^\Delta$
2 Derive a tableau Θ^Δ starting from $\Theta_0 = \{\beta\}$ s.t. Θ^Δ is finished for $\Sigma \cup \Omega^\Delta$
3 Let $\hat{\mathcal{C}} := \{\mathcal{C} \mid \mathcal{C}$ is an abductive explanation cut of Θ' wrt. $\mathcal{P}^\Delta\}$
4 Let $\hat{\Gamma} := \{\mathcal{C}\gamma \mid \gamma$ is a ground substitution for $\mathcal{C} \in \hat{\mathcal{C}}\}$
5 Let $\eta := \bigcup_{\Gamma \in \hat{\Gamma}} \{\Gamma \cup \beta\}$ and $\Sigma' := \Sigma \cup \{\varphi\}$
6 Starting from $(\Theta \setminus \{\beta\}) \cup \eta$, derive a hyper tableau Θ' that is finished for Σ'

Let us give some comments on the algorithm just described:

- The abductive process is performed wrt. the currently considered discourse model branch β. Therefore, only discourse referents occuring in this model branch are taken into account for possible antecedents. If this branch is closed at some point and another discourse model branch is selected, this abductive procedure has to be repeated for the newly selected branch.
- Since the predicate $Anaph_i$ is new and does not occur in Σ, it suffices to rename only $\overline{\Omega}$. This has the advantage that the discourse representation Σ has not to be renamed any time an anaphoric expression is encountered in a new utterance and that the discourse model β computed so far can be reused for abducing the antecedent referent.

The current discourse model branch β is extended by the set of all alternative abductive explanations and φ is added to clause set representing the discourse. By adapting techniques from [1, 2], open branches containing non $Anaph_i$-minimal models can be identifed. In our special case, we can simply close all open branches that contain more than one (ground) $Anaph_i$-atom. The resulting tableau Θ' is the new tableau that contains the discourse models representing the discourse including the last processed utterance.

By Theorem (6.3) we know that $\hat{\Gamma}$ contains all ground instantiated abductive explanations for Ω. However, it might well be that some element in $\hat{\Gamma}$ is *not* an abductive explanation as it is inconsistent with Σ'. Therefore, in the last step we have to extend the original branch β with the set of alternative abductive explanations and to deductively close with the new clause set Σ'. In case that some abductive explanation is inconsistent with Σ', the corresponding model branches are closed. This addresses the soundness question mentioned at the end of Section 6. Because, by way of contradiction, assume that Θ' contains an open finished branch that contains some element Γ from $\hat{\Gamma}$ and $\Sigma \cup \Gamma$ is unsatisfiable. The soundness result of hyper tableaux [3] gives us in the mentioned branch an model of $\Sigma \cup \Gamma$. This is a plain contradiction to the just assumed. Therefore, in sum, Algorithm 7.1 contains in its open finished branches *only* abductive explanations. Thus, soundness of the abductive explanations is established afterwards in this algorithm.

8 Conclusions

In this paper, we developed an abductive solution to anaphora resolution in a model generating framework that preserves the main advantages of the hyper tableaux calculus [3] for discourse interpretation, namely the ability to deal with several alternative discourse histories, the incrementality of the proof procedure that allows a tight integration of semantic interpretation, logical deduction and a selection of most appropiate readings as stated by [15]. We have presented an clausal transformation step for abductive reasoning and established a formal completeness result for hyper tableaux with this abductive transformation. Finally, the algorithm is given that realizes our abductive inference in the incremental setting of hyper tableaux as promised at the beginning.

[16] also presents a tableaux calculus to deal whith anaphora resolution. Since they stay in a purely deductive framework, they do not generate referent hypotheses but only check the entailment of a given hypothesis. Their approach is restricted to pronouns with the gender information encoded as a sort attached to variables and considers structural constraints e.g. in [10] concerning negated or conditional sentences. In our work, we do not obey these structural constraints, although some of them hold naturally as e.g. referents in negative scope will never occur in our model branches. Rather, we believe that these problems cannot be solved on sentence level, but have to be attacked in a larger framework of discourse coherence.

Certainly, the most inspiring paper for this work is [12] who introduce a weighted abduction scheme for resolving an abduction schema on horn clauses for resolving anaphoric references, but also for compound nominals, syntactic ambiguities and metonomies. The main difference to our approach is due to the features of our underlying disjunctive model generation. While their approach computes once the best solution wrt. weighted abduction using horn clauses and stores the corresponding facts in a database, in our approach all alternatives remain available unless they eventually become inconsistent. Therefore, we can re-examine the selection of the most appropiate discourse history so far and select another discourse branch if preferences change e.g. by a new utterance .

The question how this branch selection might look like is left open for future work — [12] state themselves that the weightings they use "is highly *ad hoc* at the present time." An interesting task for the future would be how to incorporate statistically determined factors [8] into this selection function or an explanatory coherence function [18]. Two further research lines also suggested in [12] would be on the one hand to investigate the abductive treatment of more reference problems, e.g. events, metonomies and compound nominals, on the other hand the integration of our model–based semantic component into a larger framework with interfaces to syntactic parsing, but also to broader theories of discourse coherence like centering [11].

References

[1] Chandrabose Aravindan and Peter Baumgartner. Theorem proving techniques for view deletion in databases. *Journal of Symbolic Computation*, 1999. To appear.

[2] Peter Baumgartner, Peter Fröhlich, Ulrich Furbach, and Wolfgang Nejdl. Semantically Guided Theorem Proving for Diagnosis Applications. In *15th International Joint Conference on Artificial Intelligence (IJCAI 97)*, pages 460–465, Nagoya, 1997. International Joint Conference on Artificial Intelligence.

[3] Peter Baumgartner, Ulrich Furbach, and Ilkka Niemelä. Hyper Tableaux. In *Proc. JELIA 96*, number 1126 in Lecture Notes in Artificial Intelligence. European Workshop on Logic in AI, Springer, 1996.

[4] Patrick Blackburn, Johan Bos, Michael Kohlhase, and Hans de Nivelle. Inference and computational semantics. CLAUS Report 99, University of the Saarland, Saarbrücken, 1998. To appear in Proceedings of IWCS III (Third International Workshop on Computational Semantics), Tilburg, 1999.

[5] Gerhard Brewka, Jürgen Dix, and Kurt Konolige. *Nonmonotonic Reasoning*, volume 73 of *Lecture Notes*. CSLI Publications, 1997.

[6] C. Chang and R. Lee. *Symbolic Logic and Mechanical Theorem Proving*. Academic Press, 1973.

[7] Paul Dekker. Predicate Logic with Anaphora. In R. Cooper and J. Groenendijk, editors, *Integrating Semantic Theories II*. Dyana–2, 1994.

[8] Niyu Ge, John Hale, and Eugene Charniak. A Statistical Approach to Anaphora Resolution. In *Proc. of the Sixth Workshop on Very Large Corpora*, 1998.

[9] H. Paul Grice. *Studies in the Way of Words*. Harvard University Press, Cambridge, Massachusetts, 1989.

[10] Jeroen Groenendijk and Martin Stokhof. Dynamic Predicate Logic. *Linguistics and Philosophy*, 14:39–100, 1991.

[11] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21(2):203–225, 1995.

[12] J. R. Hobbs, M. E. Stickel, D. E. Appelt, and P. Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142, October 1993.

[13] H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993.

[14] Michael Kühn. Rigid Hypertableaux. In *Proc. of KI '97*, Lecture Notes in Artificial Intelligence. Springer, 1997.

[15] Michael Kühn. A Dynamic View on Model Construction for Discourse Interpretation. In *Proc. of ESSLLI '99, Student Session*, Utrecht, 1999. To appear.

[16] C. Monz and M. de Rijke. A tableau calculus for pronoun resolution. In N.V. Murray, editor, *Automated Reasoning with Analytic Tableaux and Related Methods,* TABLEAUX'99, LNAI. Springer, 1999.

[17] Christof Monz and Maarten de Rijke. A Resolution Calculus for Dynamic Semantics. In Jürgen Dix, Lúıs Fari nas del Cerro, and Ulrich Furbach, editors, *Logics in Artificial Intelligence*, number 1489 in LNAI, pages 184–198. JELIA '98, 1998.

[18] Hwee Tou Ng and R. J. Mooney. On the role of coherence in abductive explanation. In *Proc. of the 8th National Conference on Artificial Intelligence*, pages 337–342, 1990.

[19] Marek A. Suchenek. First-order syntactic characterizations of minimal entailment, domain-minimal entailment, and herbrand entailment. *Journal of Automated Reasoning*, 10:237–263, 1993.

[20] R. A. van der Sandt. Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9, 1992.

[21] Frank Veltman. Defaults in Update Semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

Received September 15, 1999