On Coordinating Interpretations - Optimality and Rational Interaction

Anton Benz

ab@anton-benz.de
http://www.anton-benz.de

12. February 2003

Abstract

In this paper we investigate some questions about coordination and interpretation which have been addressed by *bidirectional* Optimality Theory (Bi–OT). We consider examples from anaphora resolution, and there especially the role of epistemic contexts and expectations. We look at communicative acts as joint projects following H.H. Clark (1996) and model them by using multi–agent systems with joint goals. We especially try to clarify the relation between a model which approaches anaphora resolution within a theory of rational interaction and bidirectional Optimality Theory. We hope that this will lead to a clearer understanding of the reasoning behind Bi–OT if applied to interpretation tasks in dialogue.

1 Introduction

Bidirectional Optimality Theory (Bi–OT)[1] has been suggested as a framework which explains how speaker and addressee coordinate their use and interpretation of language. The speaker has to choose a form and the hearer has to interpret it. The expressions which the speaker can choose are ordered according to his preferences, and the set of interpretations may be ordered by preferences of the addressee. One of the main applications of this theory has been anaphora resolution[2].

(1) Marion was frustrated with Jo. She/Marion/the girl was pulling his/Jo's/the boy's hair out.

In this example the speaker has to choose between the forms *she/Marion/the girl* and *his/Jo's/the boy's*. What is the best expression for the speaker to choose? If we assume that it is more economic for the speaker to produce a pronoun than a name, and better to repeat the same name than to produce a definite description, and if we assume that the hearer prefers an interpretation where *Marion* denotes a female and *Jo* a male person, then *she* and *his* are the optimal choices for referring back to Marion and Jo. According to Bi–OT participants coordinate on optimal form–meaning pairs. This was originally not meant as a principle which applies directly in dialogue interpretation. Bi-OT was used to explain why e.g. *kill* tends to denote a *direct* killing whereas *caused to die* a *indirect* killing[3]:

(2) a) Black Bart killed the sheriff.
     b) Black Bart caused the sheriff to die.

Kill is the less marked form, and if we assume that speakers prefer less marked forms over marked forms, then *kill* is the optimal way to denote a killing event. If we further assume that direct killing is the normal and expected way of killing, then *kill* and *direct killing* build an optimal form–meaning
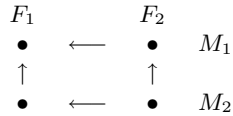
---

[1] (Blutner, 1998, 2000; Blutner, Jäger, 2000; Zeevat, 2000; Beaver, 2000)
[2] E.g. (Beaver, 2000); the following example is due to (Mattausch, 2000).
[3] See e.g. (Blutner, 2000).

pair. The theory predicts that the marked form tends to denote the less expected meaning, i.e. *cause to die* tends to denote an *indirect killing*. In general, if F_1 and F_2 are forms and M_1 and M_2 are meanings where F_1 is preferred over F_2 and M_1 over M_2, then F_1 tends to denote M_1 and F_2 to denote M_2:

$$
\begin{array}{ccccc}
F_1 & & F_2 & & \\
\bullet & \longleftarrow & \bullet & & M_1 \\
\uparrow & & \uparrow & & \\
\bullet & \longleftarrow & \bullet & & M_2
\end{array}
$$

If we apply this to our example **(1)**, then we should expect that the use of a non–optimal form has some impact on meaning:

(3) **a)** Marion was frustrated with Jo. She was pulling his hair out.

 b) Marion was frustrated with Jo. The girl was pulling the boy's hair out.

If we assume that these sentences are uttered in a situation where it is common ground that the hearer does not know who are Marion and Jo, then he should expect that Marion is female and Jo male. Hence, in both cases he should prefer an interpretation where Marion is female. But then we should expect that the less preferred form *the girl* has the tendency to go together with the less expected situation where Marion is male, and therefore should refer to Jo. But the marked form Marion has no tendency to mean that Marion is male. Instead, we find a different phenomenon: We can observe that after the use of *she* and *the girl* it is part of the common ground that Marion is female, whereas after the use of *Marion* it is still only expected that Marion is female:

(4) **a)** Marion was frustrated with Jo. Jo is the sister of the boy Marion. She was pulling his hair out.

 b) Marion was frustrated with Jo. She was pulling his hair out. Jo is the sister of the boy Marion.

 c) Marion was frustrated with Jo. Jo was pulling Marion's hair out. Jo is the sister of the boy Marion.

In version a) *she* must be interpreted as referring back to Jo. In b) *she* will be interpreted as referring to Marion with the implication that Marion is female. This contradicts the third sentence which makes b) confusing.

The role of coordination and interaction has been very much emphasised in H.H. Clark's theory of dialogue (Clark, 1996). He analysed dialogue predominantly in terms of *joint projects*, i.e. every contribution of the speaker is seen as part of an activity where he and the addressee must work together towards a joint goal. Non–linguistic examples for joint projects are *playing a piano duet*, or *paddling a canoe*. In both cases the participants have to coordinate their actions with those of the other participant, or else they may fail to reach their goal. One of the central tasks for dialogue participants is to make sure that they both update their representation of the common ground in the same way. Especially, they have to interpret dialogue contributions in the same way, or else misunderstandings will arise. An example where this becomes especially clear is the use of anaphoric expressions. The speaker has to choose an expression, and the addressee's part is to interpret it. The joint goal is to select the same antecedent. Bi-OT seems to fit nicely into this picture.

In this paper we consider the coordination tasks for anaphora resolution where the interlocutors have to take into account their common expectations about the situation being talked about. Our first aim is to account for the phenomena related to Examples **(3)** and **(4)**, and to show how they are related to the coordination task of interlocutors. Our second aim is to clarify the reasoning behind Bi–OT if applied to interpretation tasks in dialogue. In Section 2 we consider the Bi–OT solution to this problem in more detail. We reconsider Example **(3)**, and we show that we need a framework which is more context–sensitive than ordinary versions of Bi–OT.

H.H. Clark proposes to analyse dialogue acts in terms of joint projects. We will follow this idea. It has been noted that optimal form–meaning pairs can be seen as *Nash equilibria* in the sense of game theory[4]. I.e. one can look at the situation as a problem of rational choice where the

---

[4](Dekker & v. Rooy, 2000)

speaker has to choose the best form and the hearer the most preferred meaning. Then, optimal form-meaning pairs are the possible candidates which rational agents can agree to choose. This move allows us to make use of theories about coordination and knowledge in multi–agent systems[5]. In Section 4 we introduce a simple framework for our examples, which we work out in more detail in Section 5. There we show that the coordination problem is always solved if the interlocutors adhere to the rules of semantics and a number of pragmatic constraints. In Section 6 we compare the structures introduced in Section 5 with OT–structures.

2   Bi–OT on Coordination

In bidirectional OT it is common to assume that there is a set $\mathcal{F}$ of *forms* and a set $\mathcal{M}$ of *meanings* (Blutner, 2000). The speaker has to choose for his next utterance a form which then must be interpreted by the hearer. It is further assumed that the speaker has some ranking on his set of forms, and the hearer on the set of meanings. Blutner (2000) introduced the idea that the speaker and interpreter coordinate on form–meaning pairs which are most preferred from both perspectives. In (Jäger, 2000) the mechanism which leads to optimal form–meaning pairs is discussed in greater detail[6]. The speaker has to choose for a given meaning M_0 a form F_0 which is optimal according to his ranking of forms. Then the interpreter has to choose for F_0 a meaning M_1 which is optimal according to his ranking of meanings. Then again the speaker looks for the most preferred form F_1 for M_1. A form–meaning pair is optimal if ultimately speaker and hearer choose the same forms and meanings. If $\langle F, M \rangle$ is optimal in this technical sense, then the choice of F is the optimal way to express M such that both speaker's and interpreter's preferences are matched.

The OT–mechanism which allows to calculate the optimal form–meaning pairs does not make reference to context and perspectives of participants. In a normal dialogue situation the participants have only partial knowledge about the situation described and about each other. The following example shows that this poses some problems. It was first discussed by J. Mattausch (2000, pp. 33–36).

(5) Assume that Marion is a male person, and Jo a female. The speaker wants to express with the second sentence that Jo was pulling Marion's hair out:

   a) Marion was frustrated with Jo. She was pulling his hair out.
   b) Marion was frustrated with Jo. He was pulling her hair out.
   c) Marion was frustrated with Jo. Jo was pulling Marion's hair out.

   We assume that the addressee knows nothing about Marion and Jo except for what the speaker told him.

Intuitively, c) is the right way to put it. We assume that pronouns have to agree with the natural gender of the person referred to, and that the hearer prefers an interpretation where *Marion* is female and *Jo* male. These constraints lead into a circle: The speaker starts with the meaning *pulling–hair–out*(*Jo, Marion*), hence, he has to choose the form *She was pulling his hair out*. The hearer will interpret this form according to his preferences as *pulling–hair–out*(*Marion, Jo*). But this content should be expressed by the speaker as *He was pulling her hair out*. For this form the hearer should prefer the interpretation *pulling–hair–out*(*Jo, Marion*). And here the circle closes. We never reach a situation where speaker and hearer will always choose the same form and meaning. This means that Bi–OT can't provide for an optimal form–meaning pair, and if the speaker wants to communicate that *Jo was pulling Marion's hair out*, then it fails to predict that exactly this sentence is the optimal one.

But this circle is not a true circle if we take the epistemic context into account, i.e. the knowledge of speaker and hearer. If the speaker chooses between *Jo, the girl*, and *she*, then he knows that Jo is a girl. Hence he should choose, according to our constraints, the pronoun *she*. Then we have

---

[5]Our sources are (Fagin e.al., 1995; Meyer & v. d. Hoek, 1995).

[6]We describe the procedure which provides for a *strong z–optimal* form–meaning pair. (Blutner, 1998, 2000) introduced in addition *weak* optimality, also called *superoptimality*, see (Jäger, 2000, p.45) and below.

to switch to the hearer, and he can decide only on the basis of what he has been told so far. The assumption is that he will choose a referent according to his expectations, and because he has no knowledge about Marion and Jo he will expect Marion to be the girl. But this means that he has to assume now that the actual context is such that Marion is a girl and Jo is a boy. If we now switch back to the speaker, then we have to consider the contexts which the hearer believes to be possible, and that means that we have to judge the choice of the pronoun *she* in contexts where Marion is actually the girl, and where Marion pulls Jo's hair out. In this situation *She was pulling his hair out* is the optimal form for *pulling–hair–out*(*Marion, Jo*) and vice versa. Hence, it is an optimal form–meaning pair, and we have left the circle. This shows that we need to represent world knowledge and knowledge about each other in our framework.

The following picture provides a graphical solution for Mattausch's Problem. The first row lists the different forms the speaker can choose. In the first column we list the different contexts. In the second column we list the different formulas which represent the possible translations for the forms. We use the following abbreviations:

$\varphi(v_1, v_2)$ *iff pull–hair–out*(v_1, v_2),

$\mu(v_1, v_2)$ *iff* $\varphi(v_1, v_2)$ & *female*(v_1) & *male*(v_2),

$\lambda(v_1, v_2)$ *iff* $\varphi(v_1, v_2)$ & *male*(v_1) & *female*(v_2).

The worlds w_i are those worlds where *Marion*$(x) \wedge Jo(y) \wedge$ *frustrated–with*(x, y), and where the formulas listed in the second column hold. The speaker knows the actual context, and the hearer knows that it belongs to the set $\sigma = \{w_0, w_1, w_2, w_3\}$. We indicate the fact that the addressee cannot distinguish between $w_1, \ldots, w_4$ by the dashed box around the worlds. The horizontal arrows show the preferences of the speaker, the vertical ones the expectations of the interpreter. The big dots represent the form–meaning pairs which can be generated in the context listed in the first column. The two circles around the big dots in the second and third column indicate the optimal form–meaning pairs.
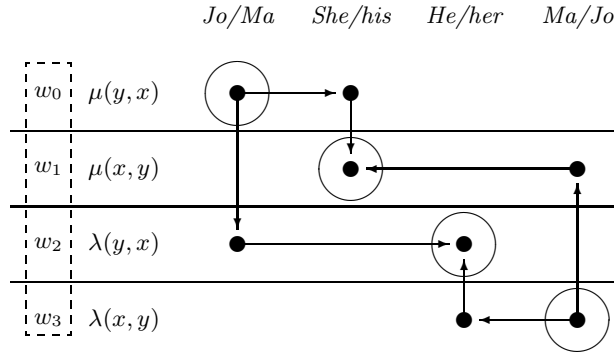
(Blutner, 1998) and (Blutner, 2000) introduced in addition to the notion of (strongly) optimal form–meaning pairs the notion of *weakly* optimal form–meaning pairs. Lets reconsider the situation of Example **(2)**. F_1 and F_2 are forms and M_1 and M_2 are meanings. F_1 is preferred over F_2 and M_1 over M_2:

We can easily check that $\langle F_1, M_1 \rangle$ is an optimal form–meaning pair. The speaker has no reason to use any other form to express that M_1 is the case, and the hearer has no reason to interpret F_1 differently from M_1. Hence we may eliminate for the speaker all form–meaning pairs $\langle F, M_1 \rangle$ where F is different from F_1, and for the hearer all pairs $\langle F_1, M \rangle$ where M is different from M_1. This means that we eliminate all nodes in the column below $\langle F_1, M_1 \rangle$ and in the row right of $\langle F_1, M_1 \rangle$.

Clearly, in the remaining graph $\langle F_2, M_2 \rangle$ is optimal. These optimal nodes of the reduced graph are then called *weakly optimal*.

In the picture representing Mattausch's Example we have indicated the weakly optimal form–meaning pairs by the circles around the big dots in the first and fourth column. This means that the theory suggests that *Jo was pulling Marion's hair out* has in this context the meaning that Jo is a girl and Marion a boy. But the addressee would still expect Jo to be a boy:

(6) Marion was frustrated with Jo. Jo was pulling Marion's hair out. She is the girl–friend of my daughter.

The addressee should interpret *she* as referring to Marion. Example **(4)**, repeated as **(7)**, did show that there is a difference in information between both forms:

(7) **a)** Marion was frustrated with Jo. She was pulling his hair out. Jo is the sister of the boy Marion.

  b) Marion was frustrated with Jo. Marion was pulling Jo's hair out. Jo is the sister of the boy Marion.

(7) a) is bad. We will argue that it is bad because the pronouns trigger an update with the expected facts, i.e. with *female*(*Marion*) & *male*(*Jo*). If we add the neutral translations to the graphical representation of Mattausch's Example, then we get the following picture:
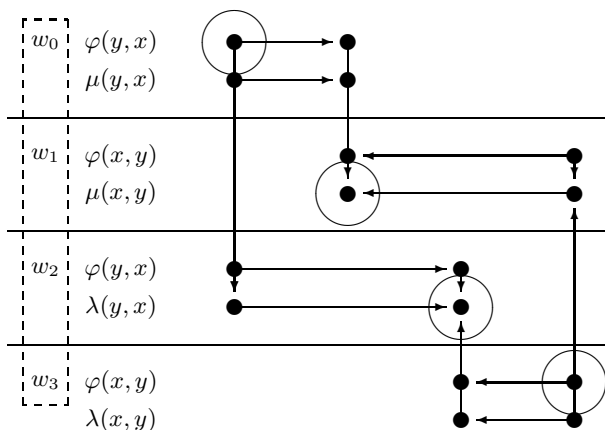
The abbreviations are the same as for the last picture. The horizontal arrows show the preferences of the speaker, the vertical ones the expectations of the interpreter. The circles around the big dots indicate the intuitively optimal form–meaning pairs. We have to explain why they are optimal and suggest a general principle which explains why in some cases we have to update with expected facts. We look at the situation of Example **(5)** as a problem of rational choice where the speaker has to choose the best form and the hearer the most preferred meaning. The coordination problem is solved if speaker and hearer can make sure that it is *common information* that they both get the same interpretation for an asserted natural sentence. In the next section we provide an informal explanation of the interpretations in Example **(5)**.

3   Interpretation and Accommodation

All our examples are assertions, and we assume that it is the goal of an assertion to inform the addressee that ψ is the case for some formula ψ chosen by the speaker. Let $\mathcal{L}$ be a first order language which contains representations for all predicates the interlocutors can use to talk about a described situation, and *NL* the set of sentences of a natural language. Let $\mathcal{C}$ be a set of *contexts*. We assume that there are two structures which define the semantics of $\mathcal{L}$ and *NL*:

$\langle \mathcal{C}, \mathcal{L}, \models \rangle$ defines the *static* semantics for $\mathcal{L}$ in the usual way.

$\langle \mathcal{C}, NL, \mathcal{L}, \text{Gen} \rangle$ with $\text{Gen} \subseteq \mathcal{C} \times NL \times \mathcal{L}$ defines for each context the possible logical forms which represent a natural sentence; i.e. if $\langle c, F, \varphi \rangle \in \text{Gen}$, then φ is a translation for sentence F in context c.

The contexts should contain enough information to make the translation unique. E.g. it should always be clear which variable the interpreter must choose for a pronoun if he has full knowledge about the situation. Therefore we can define a partial function which provides us with the translations of natural sentences: If $\langle c, F, \varphi \rangle \in \text{Gen}$, then we write $\varphi_{F,c}$ for φ.

A context c divides into three components: Two for the interlocutors, and one for the *environment* including the situation talked about. This means that a context is of the form $\langle e, c_S, c_H \rangle$, where e denotes the state of the environment, c_S the state of the speaker, and c_H the state of the hearer. We assume that the interlocutors represent the common ground by a DRS[7]. The coordination task is to translate every sentence into the same formula, and merge it with the old DRS such that the common ground is always the same for both participants. We assume that sentences with anaphoric NPs translate into formulas where the argument position for this NP is filled with a variable which is already interpreted. Normally, the set of epistemically possible contexts will contain more than one dialogue situation. But this implies that the set of possible translations for a natural sentence F may contain different formulas φ for different contexts, i.e. the translation is underspecified.

If there is only one way to interpret expressions, then the optimal solution to the coordination task is trivial: The speaker chooses the most preferred form from the expressions which have the desired interpretation. If there is more than one interpretation then in addition common expectations about the state of the world must be considered.

The first sentence of Example **(5)**, *Marion was frustrated with Jo*, restricts the possibilities to the set of all world–assignment pairs where a formula of the form

$$frustrated\text{--}with(x, y) \,\&\, Marion(x) \,\&\, Jo(y)$$

is true. This means that no information with respect to the sex of Marion and Jo is added. Hence, in some possible contexts the pronouns *she* and *he* translate into the variables x and y for Marion and Jo, in others into y and x. It is common knowledge that the models where Marion is female and Jo male are highly preferred. In such a situation we assume that the use of the pronouns *she* and *he* by the speaker triggers an *accommodation* of $female(x) \,\&\, male(y)$. What was after the first sentence only a defeasible expectation becomes thereby part of the common ground. If this is correct, then the versions **(8)**b) and **(9)**b) should be better because in the a) versions the third sentence contradicts the information which must be accommodated in order to interpret the second one.

(8) **a)** The doctor kissed the nurse. She is beautiful. The doctor there is a woman.

b) The doctor kissed the nurse. The doctor there is a woman. She is beautiful.

The same holds for cross–speaker anaphora.

(9) *A* was told that a doctor kissed a nurse. He has no evidence whether the doctor is male or not. *B* knows that.

a) *A*: C told me that the doctor kissed a nurse. *B*: Did C tell you her name? All doctors there are women.

b) *A*: C told me that the doctor kissed a nurse. *B*: All doctors there are women. Did C tell you her name?

The situation is similar to the case where there is only one antecedent. In this case it is clear which fact to accommodate:

(10) Smith entered the room. She greeted everybody.

---

[7]It is not important that it is really a DRS. But we need a representation for discourse referents.

We postulate the following principle:

If it is common knowledge

1. that the interpreter can find possible contexts where a natural sentence F translates into a formula φ_1 and contexts where it translates into a different formula φ_2,

2. that the (defeasible) expectations based on common knowledge imply a fact χ which we have to accommodate in order to translate F into φ_1,

then the assertion of F triggers the accommodation of χ and the update with φ_1.

Condition 1) and the role of expectations have been motivated before. The condition that expectations must be based on *common knowledge* is motivated by examples like the following **(11)**:

(11) A: Do you know when the guests Andrea and Maria will arrive? B: I've phoned with *her/*the woman/Andrea. They arrive tomorrow.

In **(11)** it is crucial that *Andrea* is interpreted as a name of a female person. But this is only a defeasible inference. It will normally hold if the conversation takes place in a German community but not if Andrea and Maria are known to be Italians. In the latter case it will be common knowledge that the hearer will assume that Andrea is male, hence, the speaker can refer with *the woman* to Maria. If **(11)** takes place in a hotel which is frequented only by German guests, then B can't use *she* or *the woman* even if B (privately) knows that Andrea and Maria are an Italian couple.

With this principle at hand we can explain the interpretations in Example **(4)**. The only interesting case is **(4)** b). After the first sentence, *Marion is frustrated with Jo*, it is expected but not necessary that Marion is the girl and Jo the boy. The second sentence *She was pulling his hair out* can be interpreted only if the hearer knows who is the girl and who is the boy. If the translation of the first sentence is $Marion(x) \wedge Jo(y) \wedge frustrated\text{--}with(x, y)$, then the possible translations of the second sentence are

$$\varphi_1 : \ pull\text{--}hair\text{--}out(x, y) \qquad \varphi_2 : \ pull\text{--}hair\text{--}out(y, x).$$

In order to resolve the anaphors we need to find a male and a female person in the common ground. Hence, we need to know which of the following formulas is true:

$$\chi_1 : \ female(x) \,\&\, male(y) \qquad \chi_2 : \ female(y) \,\&\, male(x)$$

Now, the contexts where χ_1 is true are highly preferred. This is part of the common ground, hence, our principle implies that the addressee accommodates χ_1 and updates then with φ_1.

We want to show how this principle follows from a general principle of rational decision and how it fits into the perspective which looks at interpretation as a *joint project*. Here, we follow an idea of *H.H. Clark* about communicative acts. According to Clark (1996, pp. 140-153) a communicative act comes in a hierarchy of joint actions, a so-called *action ladder*. He distinguishes four levels, but we are only interested in the two highest levels. At the lower of the two levels (level 3) the speaker presents a signal, and the hearer has to recognise it. For our examples this means that the speaker presents a sentence of *natural language* which is a signal for some formula φ, and the hearer has to recognise this formula. We call this level the *interpretation level*. At the higher level (level 4) the speaker proposes a mutual update of the common ground, and the hearer has to take up this project. We will assume that both participants maintain a DRS–like representation of the common ground. The mutual update then consists of the update of the representing DRSes. We call this level the *update level*. Success at the higher level implies success at the lower level. We will describe a joint project by a multi–agent system together with a joint goal. Hence, we will introduce two multi–agent systems, one for each level.

Before we introduce multi–agent systems we want to be more precise about the context–dependent grammatical form–meaning pairs $\langle c, F, \varphi \rangle \in$ Gen. We made the assumption that there is at most one translation $\varphi_{F,c}$ of F in context c. We assume that this translation is provided by some mechanism like DRT[8]. This translation is unique relative to a resolution function which

---

[8](Kamp & Reyle, 1993; v. Eijck & Kamp, 1997)

7

tells us which anaphoric expression refers to which discourse referent in the previously established common DRS. We assume that the resolution function in each context is given by the speaker's meaning. This means, he always knows how to resolve anaphors correctly. We furthermore assume that we have a background theory which tells us which facts must be accommodated by the addressee in order to find the correct resolution. This means for our example that Gen must tell us: If the first sentence of (**5**) is true and translates into $Marion(x) \land Jo(y) \land frustrated\text{--}with(x,y)$, then (1) in all contexts where Marion is a girl and Jo a boy it follows that $\chi_1 :\equiv female(x) \,\&\, male(y)$ must be accommodated, and (2) in all contexts where Marion is a boy and Jo a girl it follows that $\chi_2 :\equiv female(y) \,\&\, male(x)$ must be accommodated. We include this in the translation: In contexts c which belong to class (1) the sentence *she was pulling his hair out* translates into $\chi_1 \land \varphi_1$, and in those which belong to class (2) it translates into $\chi_2 \land \varphi_2$. We use again the following abbreviations:

$\varphi(v_1, v_2)$ *iff pull--hair--out*(v_1, v_2),

$\mu(v_1, v_2)$ *iff* $\varphi(v_1, v_2) \,\&\, female(v_1) \,\&\, male(v_2)$,

$\lambda(v_1, v_2)$ *iff* $\varphi(v_1, v_2) \,\&\, male(v_1) \,\&\, female(v_2)$.

For Example (**5**) these assumptions lead to the following situation:
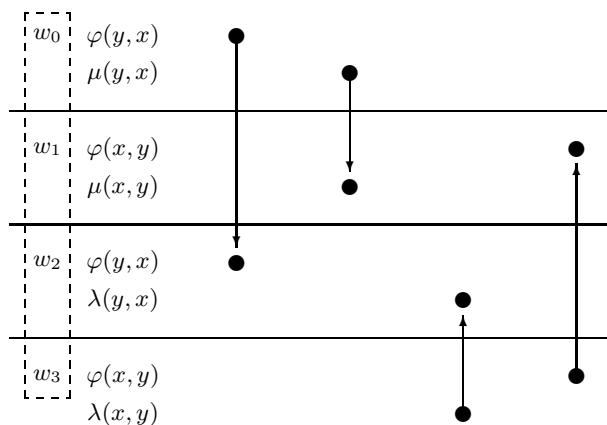
Again, the worlds w_i represent the worlds where $Marion(x) \land Jo(y) \land frustrated\text{--}with(x,y)$, and where the formulas listed in the second column hold. The speaker knows the actual context, and the hearer that it belongs to the set $\sigma = \{w_0, w_1, w_2, w_3\}$. We again indicate the fact that the addressee cannot distinguish between $w_1, \ldots, w_4$ by the dashed box around the worlds. The vertical arrows represent the expectations of the interpreter concerning the actual state of affairs. The big dots again represent the form--meaning pairs which can be generated in the context listed in the first column. If we add the arrows representing the preferences of the speaker on forms, then we see that they are no longer horizontal. In Bi--OT it is common to assume that going from meaning to form means to choose for a fixed meaning M the most preferred form F from the set of forms which translate into M. Now we see that the sentences do not need to translate all into the same formula φ.

It is not necessarily due to an ambiguity in the first sentence that the speaker has to compare forms with different logical interpretations:

(**12**)   **a)** My daughter was frustrated with my son. She was pulling his hair out.

   b) My daughter was frustrated with my son. Marion was pulling Jo's hair out.

   c) My daughter was frustrated with my son. Marion was pulling the boy's hair out.

All three choices seem to be acceptable. Hence, we have to characterise exactly the set of forms from which the speaker can make his choice. This set depends on the goals which the speaker pursues on the update level.

4 Dialogues as Multi–Agent Systems and Joint Projects

We follow Herbert H. Clark in analysing dialogue exchanges in terms of joint projects. We describe a joint project by a multi–agent system together with a joint goal. A multi–agent system[9] consists of the following components

1. A set $\mathcal{C}$ of *global states*.

2. A set ACT of possible dialogue acts.

3. A function P which tells us which dialogue acts can be performed in which dialogue situations. Hence, $P : \mathcal{C} \longrightarrow \mathcal{P}(\mathrm{ACT})$.

4. A (partial) transition operation τ with domain $\{\langle \mathtt{act}, c \rangle \mid \mathtt{act} \in P(c)\}$ and values in $\mathcal{C}$. It models the effect of the performance of dialogue acts.

5. A set of initial dialogue situations $\mathcal{C}_0$.

We identify interpretation processes with sequences $D = \langle c_0, \mathtt{act}_0, \ldots, \mathtt{act}_{n-1}, c_n \rangle$ where c_0 is an initial dialogue situation, and:

- $\mathtt{act}_i \in P(c_i)$, i.e. $\mathtt{act}_i$ is possible in c_i.

- $c_{i+1} = \tau(\mathtt{act}_i, c_i)$.

We denote the set of all processes by $\mathcal{D}$.

We have already noted that we will model global dialogue states by tripartite structures $c = \langle e, c_S, c_H \rangle$ where e represents the state of the *situation talked about*, c_S the speaker's state, and c_H the addressee's state. We will be more precise about these components below.

We assume that dialogue acts split up into the speaker's act and the addressee's act. Hence, we represent every $\mathtt{act} \in \mathrm{ACT}$ as a pair $\langle \mathtt{act}_S, \mathtt{act}_H \rangle$. We apply this to our interpretation problem in the following way: The possible dialogue acts consist of pairs $\langle F, \varphi \rangle$ where F denotes the natural sentence asserted by the speaker, and φ represents its interpretation as a formula chosen by the addressee.

P tells us which act can be performed in which context. The underlying semantics represented by Gen defines a function P by: $P(c) := \{\langle F, \varphi \rangle \mid \langle c, F, \varphi \rangle \in \mathrm{Gen}\}$.

We will characterise a class of possible dialogues by using these structures. We want to characterise the interpretation process as a joint project. We do this as follows: We formulate the joint goal of speaker and hearer in the interpretation project. Then we introduce some general pragmatic principles which imply that the joint goal is reached in the actual dialogue. We will show moreover that the participants can be sure that they have reached the joint goal, and that they can be sure that they both can be sure that they have reached it, and that they both can be sure that they both can be sure that they have reached it, etc... I.e. we will show that they have *common information* that the joint goals have been reached.

There are well worked out theories about knowledge and common knowledge in multi–agent systems[10]. It is standard to identify the knowledge of an agent in a multi-agent system with the set of all global states which are *indiscernible* from the actual state. It is assumed that two global states are indiscernible for an agent X, iff his local states are identical. This is essentially a possible worlds approach. We don't want to call what we represent in this way *knowledge* but more neutrally *information*. This means that we identify the information of a participant X in a context $c = \langle e, c_S, c_H \rangle$ with the set of all contexts c' where the local state c_X is identical with c'_X. But this would mean that we must represent all necessary information about the history of the actual dialogue in the local states. For example, we would have to represent all former local states and all utterances of the speaker. This is not a principal problem but it leads to cumbersome representations. Instead we put this information into the indiscernability relation. I.e. a participant X should not be able to discern dialogues D and D' where the sequence of his local states and the publicly performed acts are the same. This induces an equivalence relation on dialogues. Hence, we assume that for each dialogue $D \in \mathcal{D}$ there are sets $I(S, D) \subseteq \mathcal{D}$ and $I(H, D) \subseteq \mathcal{D}$ which tell us

---

[9] Our presentation of multi–agent systems follows (Fagin e.al., 1995).

[10] (Fagin e.al., 1995; Meyer & v. d. Hoek, 1995)

which dialogues are *indiscernible* for speaker S and addressee H after D; i.e. if $D' \in I(S, D)$, then the speaker's local state in the final situation of D is identical with its final local state in D', and all publicly performed acts are the same for both dialogues D and D'. If we characterise a property of dialogues by a set M of dialogues, then the speaker S can be sure that this property holds after D if $I(S, D) \subseteq M$.

This leads directly to the following representation of the *common information* $\mathrm{CI}(D)$ after Dialogue D: Let $M^0 := \{D\}$, $M^{n+1} := \bigcup\{I(X, D') \mid X = S, H \ \& \ D' \in M^n\}$, and

$$\mathrm{CI}(D) := \bigcup_{n \in \mathbf{N}} M^n.$$

If M characterises some property of dialogues, i.e. if $M \subseteq \mathcal{D}$, then it will be *common information* that the actual Dialogue D has this property, iff $\mathrm{CI}(D) \subseteq M$.

We now want to be more precise about the contexts which represent a dialogue situation. They are structures $c = \langle e, c_S, c_H \rangle$ where e represents the state of the *situation talked about*, c_S the speaker's state, and c_H the addressee's state. We said in the last section that we assume that speaker and hearer maintain DRS–like representations of the common ground. Therefore, we may assume that their local states split up into a DRS which represents their *private* knowledge, and a DRS which represents the *public* knowledge. But for our problem at hand — anaphora resolution — we can simplify matters and represent only the participant's public DRS. We assume further that all discourse referents refer to specific objects, i.e. they behave like free variables. The referent of each variable should be determined by the previous dialogue, e.g. by the speaker's meaning when they first have been introduced. To simplify matters even more, we assume that all possible states of the environment contain the same objects. We represent the state of the environment by a first order model s plus an assignment function f, i.e. it is a world–assignment pair (s, f). We denote the set of all possible world–assignment pairs by $\mathcal{W}$. f is an assignment for a DRS D, iff it maps the discourse referents of D to objects. A DRS D is a pair $\langle \mathcal{U}_D, \mathrm{Con}_D \rangle$ of a set of discourse referents and a set Con_D of formulas of $\mathcal{L}$. We assume further that the participants introduced the same variables as discourse referents into their public DRSes. Hence, all global states have the form $\langle (s, f), D_S, D_H \rangle$ such that:

- D_S and D_H are DRSes with $\mathcal{U}_{D_S} = \mathcal{U}_{D_H}$,
- f is an assignment function for D_S and D_H,
- s is a first order model for the language $\mathcal{L}$.

We can define truth of a DRS D for a world–assignment pair (s, f), and the meaning relative to an assignment function f:

- $(s, f) \models D$ iff $(s, f) \models \varphi$ for all $\varphi \in \mathrm{Con}_D$.
- $[\![\varphi]\!]_f = \{(s, f) \mid (s, f) \models \varphi\}$ and $[\![D]\!]_f = \{(s, f) \mid (s, f) \models D\}$.

Where the truth of a formula φ is defined as usual.

Finally, we characterise a set of *ideal* dialogue contexts by a set $\mathcal{I}$. These are all contexts where the DRSes D_S and D_H are identical and true in the context:

$$\mathcal{I} := \{\langle (s, f), D_S, D_H \rangle \mid (s, f) \models D_S \ \& \ D_S = D_H\}$$

5 Coordination of Interpretation

In this section we want to show how we can solve the coordination problems in Example **(5)**. We use the general interpretation principle which we introduced in Section 3.

We have mentioned before that a communicative act comes in a hierarchy of joint actions. Clark[11] distinguishes four levels. We are interested in the two highest levels. At the lower of the two levels (level 3) the speaker presents a signal, and the hearer has to recognise it. At this level the speaker presents a sentence of *natural language* which is a signal for some formula φ, and the

---

[11](1996, pp. 140-153)

hearer has to recognise this formula. We call this level the *interpretation level*. At the higher level (level 4) the speaker proposes a mutual update of the common ground, and the hearer has to take up this project. The mutual update consists of the update of the DRSes which represent the public information. We call this level the *update level*.

A multi–agent system describes the possible acts and their effects. A joint project is defined by the task to reach a joint goal. We represent such a goal by a set G of global states, i.e. the joint goal is reached if the agents choose their actions in such a way that their effects lead to a situation in G. It is not necessary that both interlocutors *know* the set G.

At every level speaker and hearer both perform an act. As there are two levels we represent the process by two successive pairs of acts:

$$\langle c_0, \langle \text{presenting}(S, F), \text{recognise}(H, F) \rangle, c_1, \langle \text{update}(S, \varphi_{F,c}), \text{update}(H, \varphi) \rangle, c_2 \rangle.$$

Where $\text{presenting}(S, F)$ means that the speaker presents some natural sentence $F \in NL$, and $\text{recognise}(H, F)$ that the addressee interprets this sentence as some formula φ. Then they both update their DRSes, the speaker with the interpretation $\varphi_{F,c}$ of F in context c as it is given by Gen, and the addressee with the formula φ which was his interpretation of F. We write the whole sequence shorter as

$$\langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \varphi_{F,c}, \oplus_H \varphi \rangle, c_2 \rangle.$$

We describe the two levels separately by two multi–agent systems.

The Interpretation Level

We describe first the interpretation level as a multi–agent system $\text{MAS}_I = \langle \mathcal{C}_I, \text{ACT}_I, \tau_I, P_I, \mathcal{I} \rangle$. We consider only dialogues with assertions. We simplify matters and assume that the speaker does not introduce new discourse referents with his new assertion. We can identify the set of possible actions the speaker can perform with the set of all natural sentences and the corresponding acts of the interpreter with formulas in $\mathcal{L}$. ACT_I represents all possible joint actions on the interpretation level. Hence,

$$\text{ACT}_I := \{ \langle F, \varphi \rangle \mid F \in NL \,\&\, \exists c \, \langle c, F, \varphi \rangle \in \text{Gen} \}.$$

The interpretation level is intended to represent the system defined by the pure semantics of the language NL. Hence, the speaker is allowed to assert F in a given context $c = \langle (s, f), D_S, D_H \rangle$, iff it has a translation in this context, and the hearer can interpret F as φ if there is a context c' which he cannot distinguish from c such that there F translates into φ:

$$P_I(c) := \{ \langle F, \varphi \rangle \in \text{ACT}_I \mid \exists c' \exists \varphi \, (D_H^c = D_H^{c'} \wedge \langle c', F, \varphi \rangle \in \text{Gen}) \}.$$

If the speaker utters F and the hearer interprets F as φ, then they should store this information somewhere in their local states. Therefore we extend the local states and define the update effect of an action $\langle F, \varphi \rangle$ in context $c = \langle (s, f), D_S, D_H \rangle$ by:

$$\tau_I(\langle F, \varphi \rangle, c) = \langle (s, f), (D_S, F), (D_H, \varphi) \rangle$$

Furthermore, we consider only ideal situations as initial states. We denote the possible courses of events on the interpretation level by $\mathcal{D}_I$.

We assume that the speaker always knows the objects being talked about and how the sentence F must be translated according to our semantics represented by Gen. Moreover, both participants should know in the resulting state that F has been uttered. We represented this information in the indiscernability relation on dialogues, i.e. for a dialogue $D = \langle c_0, \langle F, \varphi \rangle, c_1 \rangle$ we set:

- $I(S, D) = \{ D' \in \mathcal{D}_I \mid (D_S^{c_1'}, F') = (D_S^{c_1}, F) \,\&\, f^{c_1'} = f^{c_1} \,\&\, \varphi_{F', c_1'} = \varphi_{F, c_1} \}$,

- $I(H, D) = \{ D' \in \mathcal{D}_I \mid (D_H^{c_1'}, \varphi') = (D_H^{c_1}, \varphi) \,\&\, F = F' \}$,

where D' has the form $\langle c_0', \langle F', \varphi' \rangle, c_1' \rangle$ with $c_i' = \langle (s^{c_i'}, f^{c_i'}), (D_S^{c_i'}, F'), (D_H^{c_i'}, \varphi') \rangle$.

11

For a joint project a joint goal is needed. It is the task at this level to interpret the asserted natural sentence in the same way. As we assume that the speaker always chooses $\varphi_{F,c}$ as interpretation, the aim is reached if the hearer finds the correct translation $\varphi_{F,c}$ for F in c. The joint goal is given by the set:

$$G_I := \{\langle c_0, \langle F, \varphi \rangle, c_1 \rangle \in \mathcal{D}_I \mid \varphi = \varphi_{F,c_0}\}.$$

The Update Level

At the update level both interlocutors should mutually update their representation of the common ground. We denote the update of a DRS D with a formula φ by $D \oplus \varphi^{12}$. We describe the update level by a multi–agent system $\text{MAS}_u = \langle \mathcal{C}_u, \text{ACT}_u, \tau_u, P_u, \mathcal{C}_I \rangle$. The acts are all possible updates of the local DRSes:

$$\text{ACT}_u := \{\langle \oplus_S \psi, \oplus_H \psi' \rangle \mid \psi, \psi' \in \mathcal{L}\}.$$

Where $\oplus_X \psi$ is meant to denote the update of agent X' local DRS with ψ. We define here a system which is only restricted by (dynamic) semantics. An action $\langle \oplus_S \psi, \oplus_S \psi' \rangle$ is possible in a situation $c = \langle (s, f), D_S, D_H \rangle$, iff the updates $D_S \oplus \psi$ and $D_H \oplus \psi'$ are defined by our underlying DRT framework which is represented in Gen. We set:

$$P_I(c) = \{\langle \oplus_S \psi, \oplus_H \psi' \rangle \mid \exists F, F' \in \mathcal{F} \ \langle c, F, \psi \rangle, \langle c, F', \psi' \rangle \in \text{Gen}\}.$$

The transition operation is defined by the updates for DRSes:

$$\tau_u(\langle \oplus_S \psi, \oplus_H \psi' \rangle, c) = \langle (s, f), D_S \oplus \psi, D_H \oplus \psi' \rangle,$$

with $c = \langle (s, f), D_S, D_H \rangle$. We denote the possible courses of events on the update level by $\mathcal{D}_u$.

The joint goal on the update level is to update the DRSes in the same way and with the formula intended by the speaker. Hence, we define the joint goal for this level by:

$$G_u := \{\langle c_1, \langle \oplus_S \psi, \oplus_H \psi' \rangle, c_2 \rangle \in \mathcal{D}_u \mid \psi = \psi'\}.$$

Now, it is easy to see how the solution for the coordination problem on the update level depends on the solution of the coordination problem on the interpretation level. If both participants should update their representations of the common ground in the same way, then this presupposes that the addressee interprets the sentence F in the correct way.

We can now combine both levels again and describe the whole process of update and interpretation by sequences $\langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \psi, \oplus_H \varphi \rangle, c_2 \rangle$ such that:

$$\langle c_0, \langle F, \varphi \rangle, c_1 \rangle \in \mathcal{D}_I \ \& \ \langle c_1, \langle \oplus_S \psi, \oplus_H \varphi \rangle, c_2 \rangle \in \mathcal{D}_u \ \& \ \psi = \varphi_{F,c_0},$$

where φ_{F,c_0} is the translation of F in context c_0. This provides us with the set of all sequences which describe the interpretation process with anaphora resolution. We denote the set of these sequences with $\mathcal{D}_a$.

The definition of $\mathcal{D}_a$ implies that the update level is determined by the interpretation level. But then: What is our reason for distinguishing interpretation and update level? We have seen in Example **(12)** that the forms which the speaker may choose do not need to translate all into the same formula. The update with the translating formula is only a means to reach his private goals, i.e. at least to inform the addressee about some fact represented by a formula ψ. This goal is reached at the update level. The speaker has to check at this level whether the update of the old DRS with the translation of his sentence implies ψ. There may be different sentences with different translating formulas which have this property.

---

[12]For update operations see (v. Eijck & Kamp, 1997).

Pragmatic Constraints

We have described the communicative act of interpreting an asserted sentence F and thereby resolving anaphoric expressions by processes of the form $\langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \varphi_{F,c_0}, \oplus_H \varphi \rangle, c_2 \rangle$. These sequences are defined by a combination of two multi–agent systems MAS_I and MAS_u. These systems are determined by an underlying DRT–like (dynamic) semantic. It is easily seen that the semantic part for itself can not solve the coordination problem, i.e. it cannot make it common information that both participants represent F by the same formula. We will show that the coordination problem is solved if we assume that the interlocutors adhere to three very general pragmatic constraints:

- **Rational Choice** If in a situation c an agent X has to make a choice between actions $\mathtt{act}_1, \dots, \mathtt{act}_n$, and if he believes that action $\mathtt{act}_i$ has the highest expected chance of success, then he will choose $\mathtt{act}_i$. We will assume that the expectations depend on the common ground.

- **Sincerity** The speaker does not want to misinform the addressee.

- **Certainty** The speaker will only perform an act if he can be sure that it has success.

Rational Choice: We will apply this principle in the following form: Assume that σ is the set of contexts which are compatible with the knowledge of the hearer. In each context a sentence F has at most one interpretation as a formula $\varphi_{F,c}$. Hence, the formulas $\varphi_{F,c}$ define the following *partition* of σ:

$$
\begin{aligned}
P_F(\sigma) \quad &:= \quad \{N \subseteq \sigma \mid \exists \varphi \in \mathcal{L} \; N = \{c' \in \sigma \mid \varphi_{F,c'} = \varphi\}\} = \\
&\quad \{P_I^{-1}[\varphi] \cap \sigma \mid \exists c \in \sigma \; \langle c, F, \varphi \rangle \in \mathrm{Gen}\}.
\end{aligned}
$$

I.e. a sentence F divides σ into the sets of contexts where F has the same interpretation. The addressee's choice between possible translations is successful in a context c if he chooses F's *real* translation $\varphi_{F,c}$. If F has in all contexts in σ the same translation, then his choice is clear. But if there is more than one possibility, then he is most likely to be successful if he chooses the translation of the most expected context. We assume that his *expectations* about contexts determine one set $N \in P_F(\sigma)$ as the most probable; i.e. they define a function $\exp$ with $\exp(F, \sigma) \in P_F(\sigma)$. The principle of *rational choice* says that the addressee will always choose the formula φ as interpretation of F for which $P_I^{-1}[\varphi] \supseteq \exp(F, \sigma)$.

We have made the assumption that the hearer cannot distinguish between situations where his DRS for the public information is the same as in the actual one, and where the same sentence has been uttered. According to our previous definition of the *indiscernability*, the set σ of the hearer H's epistemic possibilities in context c is equal to $\{c_1 \mid \exists \varphi \; \langle c_1, \langle F, \varphi \rangle, c_2 \rangle \in \mathcal{D}_I \;\&\; D_H^{c_1} = D_H^c\}$. But this means that $\exp$ depends only on F and $D_H^c = D_S^c$, the local representations of the public information in the initial situation. Hence, we write $\exp(F, D)$, where D is a DRS[13].

Sincerity: We have described the communicative act of asserting as a joint project with joint goals. The speaker also has his private goals which he pursues in communication. We assume as a minimal[14] goal that he wants to inform the hearer about some fact. We again identify such a goal G_S with the set of all dialogues $D \in \mathcal{D}_a$ where the goal is reached. As a minimal condition, we demand that the speaker should not want to misinform the hearer. This is equivalent with the constraint that for all $D \in G_S$ the final global state is an *ideal* situation, i.e. if $\langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \varphi_{F,c_0}, \oplus_H \varphi \rangle, c_2 \rangle \in G_S$, then $c_2 \in \mathcal{I}$, where $\mathcal{I} = \{\langle (s,f), D_S, D_H \rangle \mid (s,f) \models D_S \;\&\; D_S = D_H\}$ denotes the set of ideal situations.

---

[13] We should note at this point that we can make this definition only if we assume that we can neglect *private* knowledge. This is only justified because we can assume that the expectations of the addressee are part of the common ground.

[14] It may be that this is only a means which serves further ends.

Certainty   Certainty claims that the speaker is sure that he has success. This means that he will perform an act only if he is sure that all his goals will be reached. If G defines the set of dialogues where all his goals are reached, then his choice of a sentence F in a situation c is restricted to the set of forms where after every possible choice of the hearer the resulting dialogue will belong to G.

We can look at the whole process as a *game*. The speaker S chooses first a form F_i, and then the hearer H a translation φ_{ij} for F. The hearer *wins* if he chooses the right formula φ_{ij}. As we assume that the speaker always chooses the correct translation $\varphi_{F,c}$ in context c, it means that the hearer wins if he finds $\varphi_{F,c}$. The speaker *wins* if the hearer finds $\varphi_{F,c}$ and the mutual update with $\varphi_{F,c}$ leads to an ideal situation where the hearer is informed about the fact represented by G_S:
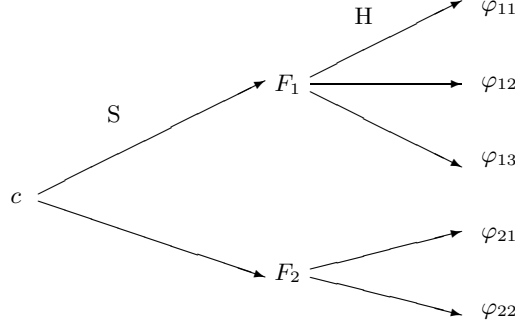
This *game* is described by the multi–agent systems plus the joint and private goals. The pragmatic principles lead to a restriction on the possible game moves. This can be formally represented as a restriction on the P function of the multi–agent systems.

Rational Choice: This principle defines a restriction on the interpretation level. Hence, we redefine P_I for $c = \langle (s,f), D_S^c, D_H^c \rangle$ as follows:

$$P_I^1(c) := \{\langle F, \varphi \rangle \in P_I(c) \mid \exists c' \in \exp(F, D_H^c) : \varphi = \varphi_{F,c'}\}.$$

This defines a restricted set $\mathcal{D}_a^1$ of sequences which describe interpretation processes for assertions with anaphora resolution. For $D = \langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \psi, \oplus_H \varphi \rangle, c_2 \rangle \in \mathcal{D}_a$ it holds that $D \in \mathcal{D}_a^1$ iff $\langle F, \varphi \rangle \in P_I^1(c_0)$.

Sincerity & Certainty: This defines a restriction on the possible choices of F in the interpretation game. Let $\mathcal{D}(F,c)$ be the set of sequences in $\mathcal{D}_a^1$ which start with the speaker's choice of the form F in situation c. Let G denote the intersections of the sets which represent the joint goal at the interpretation and update level, and his private goal G_S. Then, the new condition is:

$$P_I^2(c) = \{\langle F, \varphi \rangle \in P_I^1(c) \mid \mathcal{D}(F,c) \subseteq G\}.$$

This defines again a restricted set $\mathcal{D}_a^2$ of sequences which describe interpretation processes for assertions with anaphora resolution. For $D = \langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \varphi_{F,c_0}, \oplus_H \varphi \rangle, c_2 \rangle \in \mathcal{D}_a^1$ it holds that $D \in \mathcal{D}_a^2$ iff $\langle F, \varphi \rangle \in P_I^2(c_0)$.

The central claim is that the coordination problem is always solved for the system $\mathcal{D}_a^2$. It is for all sequences $D \in \mathcal{D}_a^2$ common information that the joint projects are successful:

Lemma 5.1 *Let D_a^2 be as above. Let G be the set of all $\langle c_0, \langle F, \varphi \rangle, c_1, \langle \oplus_S \varphi_{F,c_0}, \oplus_H \varphi \rangle, c_2 \rangle \in \mathcal{D}_a$ such that $\langle c_0, \langle F, \varphi \rangle, c_1 \rangle \in G_I$ and $\langle c_1, \langle \oplus_S \varphi_{F,c_0}, \oplus_H \varphi \rangle, c_2 \rangle \in G_u$, where G_I and G_u represent the joint goals on the interpretation level and update level:*

- *$G_I = \{\langle c_0, \langle F, \varphi \rangle, c_1 \rangle \in \mathcal{D}_I \mid \varphi = \varphi_{F,c_0}\}$,*

- *$G_u = \{\langle c_1, \langle \oplus_S \psi, \oplus_H \psi' \rangle, c_2 \rangle \in \mathcal{D}_u \mid \psi = \psi'\}$.*

Then, it holds for all $D \in \mathcal{D}_a^2$ that $CI(D) \subseteq G$.

14

The proof is by induction. In the induction step we have to show that $D \in G \Rightarrow I(S, D) \cup I(H, D) \subseteq G$, which follows with the certainty condition.

So it may seem that the certainty condition is very strong. Is it not too strong? It claims that the speaker can only assert a sentence if he knows that the hearer will interpret it in the same way as he does. How can the speaker be certain? The next lemma gives an answer. It is important that the initial utterance situation is an ideal situation (elements of $\mathcal{I}$), especially that they both have the same representations D_S and D_H of the common ground. Furthermore, it is important that the expectations are determined by the common ground and the uttered sentence. In addition to the joint goals, the speaker pursues his private goals. When can he be certain to reach them? We assume here that we can identify his private goal with the goal to inform the hearer about some fact ψ. We denote by $G(\psi)$ the set of all sequences in $\mathcal{D}$ where this goal is reached. The sincerity condition implies that the final states of sequences in $G(\psi)$ have to be ideal situations. Let $\mathcal{D}^1(F, c)$ be the set of sequences in $\mathcal{D}_a^1$ which start with the speaker's choice of the form F in situation c. We find:

Lemma 5.2 *Let $c = \langle (s, f), D_S, D_H \rangle \in \mathcal{I}$ and let $\exp(F, D)$ be as defined above. Let F be such that $\exists \varphi \ \langle c, F, \varphi \rangle \in \mathrm{Gen}$, hence $\mathcal{D}^1(F, c) \neq \emptyset$, then:*

$$\mathcal{D}^1(F, c) \subseteq \mathcal{D}_a^2 \cap G(\psi) \iff c \in \exp(F, D_S) \ \& \ [\![Con_{D_S}]\!]_f \cap [\![\varphi_{F,c}]\!]_f \subseteq [\![\psi]\!]_f.$$

If D is in $\mathcal{D}_a^1$ then it is known that the addressee adheres to the principle of rational choice based on common expectations. As the initial situation is an ideal situation, the DRSes for public information are identical, hence, $\exp(F, D_S) = \exp(F, D_H)$. If the real situation c is an element of $\exp(F, D_H)$, it follows that the addressee will take $\varphi_{F,c}$ as interpretation of F. This shows that the coordination problem on the interpretation level is solved, hence, it is also on the update level. The addressee will update with $\varphi_{F,c}$, hence his information relative to the assignment function f is given by $[\![Con_{D_S}]\!]_f \cap [\![\varphi_{F,c}]\!]_f$. It follows that the addressee is informed about ψ. As $c \in [\![D_S]\!]_f \cap [\![\varphi_{F,c}]\!]_f$, it follows that the final state is again an ideal state. If $c \notin \exp(F, D_H)$, then speaker and addressee will choose different formulas as translations for F. Hence, the speaker violates the condition of certainty or sincerity if he chooses F. If $[\![Con_{D_S}]\!]_f \cap [\![\varphi_{F,c}]\!]_f \not\subseteq [\![\psi]\!]_f$, then the addressee is not informed that ψ.

Mattausch's Example Reconsidered

We reconsider Mattausch's Example **(5)** repeated here as **(13)**:

(13) **a)** Marion was frustrated with Jo. She was pulling his hair out.

b) Marion was frustrated with Jo. Jo was pulling Marion's hair out.

We consider a context where *Marion was pulling Jo's hair out*. We assume again that the first sentence translates into $Marion(x) \wedge Jo(y) \wedge frustrated\text{--}with(x, y)$. Let f be the assignment function with $f(x) = Marion$ and $f(y) = Jo$. Hence, we assume that both interlocutors maintain a DRS D for public information with discourse referents $\mathcal{U}_D = \{x, y\}$, and a set of conditions $Con_D = \{Marion(x), Jo(y), frustrated\text{--}with(x, y)\}$. We assume further that the speaker knows who Marion and Jo are, that Marion is a girl and Jo a boy, and that he wants to inform the addressee about the fact that *Marion was pulling Jo's hair out*. In this context, the sentences *Marion was pulling Jo's hair out* and *She was pulling his hair out* should be successful.

We first consider the sentence $F_0 =:$ *Marion was pulling Jo's hair out*. It is natural to assume that F_0 translates in all contexts to

$$\varphi_0 : \ Marion(x) \wedge Jo(y) \wedge pull\text{--}hair\text{--}out(x, y).$$

According to our previous definition of the *indiscernability*, the set σ of the hearer H's epistemic possibilities in context c relative to $\mathcal{D}_I$ is equal to $\{c_1 \mid \exists \varphi \ \langle c_1, \langle F_0, \varphi \rangle, c_2 \rangle \in \mathcal{D}_I \ \& \ D_H^{c_1} = D\}$. This is equal to $\{c \mid c \models D \ \& \ \exists \varphi \ \langle c, F, \varphi \rangle \in \mathrm{Gen}\}$.

Then $P_{F_0}(\sigma) = \{N \subseteq \sigma \mid \exists \varphi \in \mathcal{L} \ N = \{c' \in \sigma \mid \varphi_{F_0, c'} = \varphi\}\}$. Hence, $P_{F_0}(\sigma) = \{\sigma\}$, and there is only one set which can be chosen by $\exp(F_0, D)$. If c_1 denotes the actual situation where Marion

was pulling Jo's hair out, then $c_1 \in \sigma \cap [\![D]\!]_f \cap [\![\varphi_1]\!]_f$. Lemma 5.1 implies that the joint goals of finding the same interpretation for F_0 and updating with the same formula are reached.

Let us now consider the case $F_1 = $ *She was pulling his hair out*. The possible translations of F_1 are

$$\varphi_1 : \; pull\text{-}hair\text{-}out(x,y) \qquad \varphi_2 : \; pull\text{-}hair\text{-}out(y,x).$$

In order to resolve the anaphors the hearer needs to find a male and a female person in the common ground. Hence, he needs to know which of the following formulas is true:

$$\chi_1 : \; female(x) \,\&\, male(y) \qquad \chi_2 : \; female(y) \,\&\, male(x).$$

We have made the assumption that accommodated facts like χ_1 and χ_2 are part of the translating formula[15]. Hence F_2 must either translate to $\chi_1 \wedge \varphi_1$ or $\chi_2 \wedge \varphi_2$. The set of the hearer's epistemic possibilities is then $\{c \,|\, c \models D \,\&\, \exists \varphi \; \langle c, F_1, \varphi \rangle \in \mathrm{Gen}\} = \{c \,|\, c \models D \wedge (\chi_1 \wedge \varphi_1 \vee \chi_2 \wedge \varphi_2)\}$. $P_{F_2}(\sigma)$ divides into the sets:

$$N_1 = \{c \in \sigma \,|\, c \models \chi_1\} \qquad N_2 = \{c \in \sigma \,|\, c \models \chi_2\}.$$

If it is only known that *Marion was frustrated with Jo*, then it is more expected that the first set contains the actual situation. Hence, $\exp(F_1, D) = N_1$. If c_1 denotes the actual situation where Marion was pulling Jo's hair out, then $c_1 \in \exp(F_1, D) \cap [\![D]\!]_f \cap [\![\chi_1 \wedge \varphi_1]\!]_f$. Again, Lemma 5.1 implies that speaker and hearer find the same interpretation for F_1. We see also that the use of *She was pulling his hair out* leads to an update with a stronger formula which implies that Marion is female and Jo male.

Hence, we see that both forms, *Marion was pulling Jo's hair out* and *She was pulling his hair out*, are suitable, and the speaker can now choose the more preferred one, i.e. according to our assumptions *She was pulling his hair out*.

Let us now consider the situation c_2 where Jo was pulling Marion's hair out, and where Jo is female and Marion male. Of course, *Marion was pulling Jo's hair out* is ruled out because there is no true translation. For *Jo was pulling Marion's hair out* the situation is parallel to that for F_0 in the previously considered context. Lets consider $F_1 = $ *She was pulling his hair out* again. In this case the hearer's epistemic possibilities and the expectations based on the common ground remain the same as above. But we find $c_2 \notin \exp(F_1, D)$. If the speaker adheres to the rules and interprets F_1 by its correct translation $\chi_2 \wedge \varphi_2$, the speaker and hearer will end up with different interpretations and a misunderstanding is predicted.

6 Bi–OT Reconsidered

In the last section we introduced a framework for studying the coordination problems related to anaphora resolution. The central idea was to describe the interpretation process by using multi–agent systems and to define pragmatic constraints which account for the role of joint goals and expectations in this process. We now want to show how this defines OT–structures for dynamic contexts. First, we introduce *Blutner structures* as structures for BI–OT. We then compare them with the structures introduced in Section 5.

In this section we provide a precise description of the structures which underlie the version of BI–OT which we have discussed in Section 2. They are more general than the structures proposed by Blutner (2000) and Jäger (2000)[16].

According to OT, producer and interpreter of language use a number of constraints which govern their choice of forms and meanings. These constraints may get into conflict. OT proposes a mechanism for how these conflicts are resolved. It assumes that the constraints are ranked in a linear order. If they get into conflict, then the higher-ranked constraints win over the lower ranked ones. This defines preferences on forms and meanings.

---

[15] See p. 8

[16] We have discussed the reasons for introducing these more general structures in (Benz, 2001). We named them *Blutner Structures*. There we also compared them in detail with the OT–systems as proposed by Jäger (2000).

Preferences can be identified with transitive relations $\preceq$, where we read $F \prec F'$ as F' *is preferred over* F, and $F \approx F'$ as F *and* F' *are ranked equal*. We first fix some terminology concerning transitive relations:

Definition 6.1 *Let M be a set and $\preceq\, \subseteq M \times M$ a relation. We say that $\preceq$ is a* pre–order, *iff*

- $m \preceq m$,
- $m \preceq m' \wedge m' \preceq m'' \Rightarrow m \preceq m''$.

$\preceq$ *is an* order, *iff in addition*

- $m \preceq m' \wedge m' \preceq m \Rightarrow m = m'$.

If $\preceq$ is a pre–order, then the sets $[m] := \{m' \in M \mid m \preceq m' \,\&\, m' \preceq m\}$ are equivalence classes. If we set $[m] \preceq [m'] \Leftrightarrow m \preceq m'$, then $\preceq$ is an order relation on the set of equivalence classes $[m]$.
   An order $\preceq$ is linear, *iff*

$$m \preceq m' \vee m' \preceq m.$$

It is well founded, iff there is for every set $X \subseteq M$ an $m \in X$ such that

$$\forall m' \in X \; m \preceq m'.$$

We call a pre–order well–founded or linear, iff the associated order on the set of equivalence classes $\{[m] \mid m \in M\}$ is well–founded or linear.

The following structures underlie the version of Bi–OT which we have discussed before.

Definition 6.2
A Blutner structure with contexts is a tuple $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \mathcal{C}, \mathrm{Gen}, \preceq \rangle$ where

- $\mathcal{C}$, $\mathcal{F}$ *and* $\mathcal{M}$ *are sets.*
- Gen *is a subset of* $\mathcal{C} \times \mathcal{F} \times \mathcal{M}$.
- $\preceq$ *is a family* $(\preceq_p)_{p \in P}$ *with* $P \subseteq \mathcal{C} \times (\mathcal{F} \cup \mathcal{M})$ *where*
   - $\preceq_{c,F}$ *is a linear pre–order on* $\{M \mid \langle c, F, M \rangle \in \mathrm{Gen}\}$.
   - $\preceq_{c,M}$ *is a linear pre–order on* $\{F \mid \langle c, F, M \rangle \in \mathrm{Gen}\}$.

We call $\mathcal{F}$ a set of forms, $\mathcal{M}$ *a set of meanings, and $\mathcal{C}$ a set of contexts.*

$\preceq_{c,F}$ stands for the preferences of the interpreter in context c given a form F, and $\preceq_{c,M}$ stands for the preferences of the speaker in context c given a meaning M.

An *optimal* form–meaning pair $\langle F, M \rangle$ for a context c is a pair $\langle F, M \rangle$ such that (1) there is no $F' \in \mathcal{F}$ with $\langle c, F', M \rangle \in \mathrm{Gen}$ and $F \prec_{c,M} F'$, and (2) there is no $M' \in \mathcal{M}$ with $\langle c, F, M' \rangle \in \mathrm{Gen}$ and $M \prec_{c,F} M'$.

We compare these structures by defining Blutner structures for the games considered in the last section. Let $\mathrm{MAS} = \langle \mathcal{C}, \mathrm{ACT}, \tau, P, \mathcal{I} \rangle$ be one of the multi–agent systems for the interpretation level defined in Section 5. Let $\langle NL, \mathcal{L}, \mathrm{Gen} \rangle$ represent the underlying semantics and $\mathcal{D}$ the defined set of processes. If we try to characterise the Blutner structures for these components, what are the major changes in comparison to Definition 6.2? We are especially interested in (1) the sets for which the preference relations of speaker and hearer are defined, (2) the additional parameters which we need to determine them, and (3) in the different properties of optimal form–meaning pairs.

Now that we know that the pragmatic constraints guarantee that an assertion leads from ideal situations again to ideal situations, we can simplify our contexts. According to our previous definition, an ideal dialogue situation $c = \langle (s, f), D_S, D_H \rangle$ is such that $D_S = D_H$ and $(s, f) \models D_S$. Hence, we can identify it with a pair $\langle (s, f), D \rangle$, i.e. a world–assignment pair and a DRS D which represents the common ground. We denote the set of all ideal situations again by $\mathcal{I}$. For these contexts we want to define a Blutner structure.

Bi–OT explains the use and interpretation of expressions by recursion to their preferences on forms and meanings. The interpretation problem at hand is basically a *translation* problem, where the speaker has preferences on sentences $F \in NL$ of natural language, and the addressee on translating formulas $\varphi \in \mathcal{L}$.

The epistemic alternatives for the interpreter in context $c = \langle (s, f), D \rangle$ are given by $\{c' = \langle (s', f'), D \rangle \mid \exists \varphi \in \mathcal{L} \; \langle c', F, \varphi \rangle \in \text{Gen}\}$. We denote this set by $[c]_F$. Let exp be the function which represents the commonly known expectations with respect to the information represented by a DRS D and the fact that a sentence F has been asserted. exp defines a function prf which provides us with the most preferred translation for a sentence F in a context $c = \langle (s, f), D \rangle$:

$$\text{prf}(F, c) = \varphi \text{ iff } \exp(F, D) = P^{-1}[\varphi] \cap [c]_F.$$

If the expectations define a linear pre–order on $\{P^{-1}[\varphi] \cap [c]_F \mid \exists c' \in [c]_F \; \langle c', F, \varphi \rangle \in \text{Gen}\}$, then this order defines a linear pre–order on

$$R_{F,c}\{\varphi \mid \exists c' \in [c]_F \; \langle c', F, \varphi \rangle \in \text{Gen}\}.$$

This shows that the epistemic perspective of the addressee must be considered. We write for the order $\preceq_{F,c}$, where $\varphi \preceq_{F,c} \psi$ means that ψ is preferred over φ. This accounts for the preferences of the addressee on translating formulas.

Example **(12)** did show that a problem shows up when characterising the set of forms where the speaker can make his choice. It seemed to be a too strong a requirement that all these forms should translate into the same formula. Assume that the speaker wants to inform the addressee that some fact ψ holds. With Lemma 5.2 at hand we can describe the set where he can make his choice in context $c = \langle (s, f), D \rangle$ as follows:

$$R_{\psi,c} := \{F \in \mathit{NL} \mid c \in \exp(F, D) \; \& \; \llbracket D \rrbracket_f \cap \llbracket \varphi_{F,c} \rrbracket_f \subseteq \llbracket \psi \rrbracket_f\}.$$

As we have simplified our contexts, we can write $\exp(F, c)$ instead of $\exp(F, D)$. If we now assume that the speaker's goal is given by a set G of contexts and if we write $\oplus_\varphi(c)$ for the context where the common ground was updated with a formula φ, then we can simplify:

$$R_{\psi,c} = \{F \in \mathit{NL} \mid c \in \exp(F, c) \; \& \; \oplus_{\varphi_{F,c}}(c) \in G\}.$$

The possible choices of forms for the speaker are determined by his private goals and expectations in addition to the pure semantics represented by Gen.

We collect the additional parameters. For the comparison we present the structures in a slightly more abstract form: Let $\langle \mathcal{F}, \mathcal{M}, \mathcal{C}, \text{Gen} \rangle$ be given as in Definition 6.2:

- $[c]_F$: The set of epistemic alternatives for the addressee in situation c given F.

- $\exp(F, c)$: A subset of $[c]_F$ which tells us which facts are expected if F is used in context c.

- $\oplus_M$: An update operation with $\oplus_M(c) \in \mathcal{C}$.

- G: A subset of $\mathcal{C}$ representing the speaker's goal.

What does this mean for optimal form–meaning pairs? Of course, the definition must be basically the same. In 6.2 the set of alternatives where the speaker can make his choice is only restricted by Gen, a context, and a meaning. If we now neglect the influence of goals and only look at the expectations, then the additional restriction which is imposed by them is: Avoid misleading forms! If we reconsider now the picture on page 5, we see that the form–meaning pairs in the first and the fourth column, which looked there like weakly–optimal form–meaning pairs, are really optimal. The pairs without circles around them are ruled out by the condition that $c \in \exp(F, c)$.

7 Summary

We investigated some questions about coordination and interpretation which have been addressed by *bidirectional* Optimality Theory (Bi–OT). Bi–OT assumes that preferences of speaker and hearer play an essential role. Following (Beaver, 2000) and (Mattausch, 2000) we looked at examples for anaphora resolution, and there at the role of epistemic contexts and expectations. Our aim was to clarify the reasoning behind Bi–OT if applied to interpretation tasks.

The role of coordination and interaction has been very much emphasised in H.H. Clark's theory of dialogue (Clark, 1996). He analysed dialogue predominantly in terms of *joint projects*, i.e. every contribution of the speaker is seen as part of an activity where he and the addressee must work together towards a joint goal. The interpretation problem for anaphora resolution is basically a

translation problem, where the speaker has preferences on sentences $F \in NL$ of natural language, and the addressee on translating formulas $\varphi \in \mathcal{L}$.

We looked at the situation as a problem of rational choice where the speaker has to choose the best form and the hearer the most preferred meaning. This move allowed us to make use of theories about coordination and knowledge in multi–agent systems. Our basic move was to consider an assertion as a joint project. Following (Clark, 1996) we divided this project into two dependent subprojects. We could show that pure semantics plus some pragmatic conditions always guarantee that it is mutual knowledge that these projects are successful. At one level, the interlocutors have to agree on the translations of uttered sentences. On another level, they have to reach the conversational goal by a mutual update. We described each level as a multi–agent system. We considered examples where world knowledge, and expected (defeasible) facts about the world define the preferences of the hearer for translations. These enter at the interpretation level. Here, expected facts were accommodated if this was needed to make an interpretation task unambiguous.

In Section 4 we introduced a framework for our examples, which we worked out in more detail in Section 5. We showed that the coordination problem is always solved if the interlocutors adhere to the rules of semantics and a number of pragmatic constraints:

- **Rational Choice** If an agent has to make a choice between actions $\mathtt{act}_1, \ldots, \mathtt{act}_n$, and if he believes that action $\mathtt{act}_i$ has the highest expected chance of success, then he will choose $\mathtt{act}_i$. In applications we assume that expectations must be part of the common ground.

- **Sincerity** The speaker does not want to misinform the addressee.

- **Certainty** The speaker will only perform an act if he can be sure that it has success.

In Section 6 we compared the structures introduced in Section 5 with Bi–OT, and showed how they define OT–structures for dynamic contexts. We were especially interested in (1) a characterisation of the sets of forms and meanings where speaker and addressee can make their choices, (2) the additional parameters which we need to determine them, and (3) the different properties of optimal form–meaning pairs.

References

D. Beaver (2000): *The Optimization of Discourse;* ms. Stanford; to appear in *Linguistics and Philosophy.*

A. Benz (2001): *Towards a Framework for Bidirectional Optimality Theory in Dynamic Contexts;* ROA 465–0901, Rutger's Optimality Archive.

R. Blutner (1998): *Lexical Pragmatics;* Journal of Semantics 15, pp. 115–162.

R. Blutner (2000): *Some Aspects of Optimality in Natural Language Interpretation;* In: Helen de Hoop & Henriette de Swart (eds.) *Papers on Optimality Theoretic Semantics.* Utrecht Institute of Linguistics OTS, December 1999, pp 1-21. Also: *Journal of Semantics* 17, pp. 189-216.

R. Blutner, G. Jäger (2000): *Against Lexical Decomposition in Syntax;* In A.Z. Wyner (ed.): *Proceedings of the Fifteenth Annual Conference*, IATL 7, University of Haifa, pp. 113-137 *Proceedings of IATL 15*, University of Haifa.

H.H. Clark (1996): *Using Language;* Cambridge.

P. Dekker, R. v. Rooy (2000): *Bi–Directional Optimality Theory: An Application of Game Theory;* Journal of Semantics 17, pp. 217–242.

J. v. Eijck, H. Kamp (1997): *Representing Discourse in Context;* in: J. v. Bentham, A. t. Meulen *Handbook of Logic & Language*, Elsevier, Amsterdam.

R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi (1995): *Reasoning About Knowledge;* MIT Press, Cambridge, Massachusetts.

G. Jäger (2000): *Some Notes on the Formal Properties of Bidirectional Optimality Theory;* ms, ZAS Berlin.

H. Kamp, U. Reyle (1993): *From Discourse to Logic;* Dordrecht.

J.–J. CH. Meyer, W. van der Hoek (1995): *Epistemic Logic for AI and Computer Science;* Cambridge University Press, Cambridge.

J. Mattausch (November 2000): *On Optimization in Discourse Generation;* master thesis, Universiteit van Amsterdam.

H. Zeevat (2000): *Semantics in Optimality Theory;* H. de Hoop, H. de Swart (eds.): *Optimality Theoretic Semantics*, OTS preprint, University of Utrecht.