Towards a Framework for Bidirectional Optimality Theory in Dynamic Contexts

Anton Benz

toni.benz@german.hu-berlin.de

Abstract

In this paper we study context–sensitive versions of bidirectional Optimality Theory (OT). We discuss a problem presented by Jason Mattausch which shows that context–sensitivity may lead into circularity. In order to be able to represent it we introduce a fundamental structure for bidirectional OT which we call *Blutner structure*. The discussion of Mattausch's Example especially leads us to combine bidirectional OT with Dynamic Semantics.

1 Introduction

Bidirectional Optimality Theory (OT)[1] has been suggested as a framework which explains how speaker and interpreter co-ordinate their choice of preferred forms and preferred interpretations. Recently[2] this theory has been applied to anaphora resolution. In an example like **(1)** the theory explains why it is the best way to refer to *Marion* and *Jo* with *she* and *he*:

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

If we assume that it is more economic for the speaker to produce a pronouns 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. The aim of our paper is to outline a framework which describes the reasoning of bidirectional OT in contexts known from Dynamic Semantics.

In bidirectional OT it is usual 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 co-ordinate on form–meaning pairs which are most preferred from both perspectives. In (Jäger, 2000) the mechanism which leads to optimal

[1](Blutner, 1998, 2000; Blutner, Jäger, 2000; Zeevat, 2000; Beaver, 2000)

[2](Beaver, 2000). Beaver's version of a two–sided OT is in some respects different from the version cited above.

form–meaning pairs is discussed in greater detail[3]. 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 always 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 following example presented by *Jason Mattausch* (2000, pp. 33–36) shows that there is no guarantee that optimal form–meaning pairs exist:

(2) Assume that Marion is a male person, and Jo a female one. 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.

Intuitively, c) is the right way to put it. Mattausch assumes that pronouns have to agree with the natural gender of the person referred to, and that pronouns are preferred over names. On the other side, the hearer prefers an interpretation where *Marion* is female, and *Jo* male. These constraints lead into the following circle:

The speaker starts with the meaning *pulling–hair–out*(*Jo*, *Marion*), hence, he has to choose *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 bidirectional 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. Mattausch then argues that we have to give up bidirectional OT in the style of (Blutner, 2000). We want to show that such a conclusion is premature. But then we have to outline how to apply bidirectional reasoning in contexts known from Dynamic Semantics. The problems behind Mattausch's Example will lead us on our way.

The example shows that there may exist circular structures. *Gerhard Jäger* (2000, Lem. 2) shows that no circularity can arise as long as we assume that preferences are induced by a system of ranked OT–constraints. This result leads him to use OT–systems as they have been introduced for unilateral OT as underlying structures for bidirectional OT. They are pairs $\langle \text{Gen}, C\rangle$ where Gen represents a set of grammatical form–meaning pairs, and C a set of functions representing a system of ranked constraints. If we use them as fundamental structures, then circularity is ruled out a priori. Although we will show that

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

the circularity in Mattausch's Example is only apparent, we think that OT–structures should be flexible enough to represent it. We will see that Mattausch squares can't be generated by OT–systems because *all* constraints are ranked in one line. If we divide them into constraints for the speaker and constraints for the interpreter, and rank them separately, then we can avoid this limitation. We introduce structures — we call them *Blutner structures* — which allow to represent Mattausch's Problem for anaphora resolution, and to solve it within bidirectional OT. Our examples show that we need constraints which are not only sensitive to forms and meanings but also to some facts about the real world. This leads us to consider Blutner structures with arbitrary context–sensitive constraints. These facts may be introduced by previous discourse. Our aim is to integrate bidirectional OT into (a fragment of) Dynamic Semantics. We will first introduce the static part, then apply this framework to Mattausch's Example, motivate thereby the modifications for dynamic contexts, and define them precisely in the last section.

We discuss OT–systems in more detail in Section 2. In Section 3 we introduce *Blutner structures* which are generated by two OT–systems over the same set of grammatical form–meaning pairs. There, we have to show that central notions of bidirectional OT like *optimality* and *weak optimality* can be generalized for these structures, how these structures are related to OT–systems, and how we can handle arbitrary context–sensitive constraints.

In Section 4 we will use Blutner structures to study Mattausch's Example. It turns out that in switching from the speaker's to the interpreter's role we have to be cautious about the contexts and the information the hearer has about contexts. It may well be that the interpreter prefers for a given form a meaning which is only grammatical in a context different from the actual one. Exactly this happens in Mattausch's Example. If we then switch to the speaker's role again, it follows that he can't find for this meaning an appropriate form. We call such a situation a *dead end*. We argue that we have to add to context–dependent bidirectional OT a principle which postulates that the speaker has to avoid forms which necessarily lead into dead ends.

This shows that it is essential for Mattausch's Example that the interpreter has only a limited knowledge about the actual context. In a Dynamic Semantic setting we can assume that he has always *less* information than the speaker. Hence, he has more form–meaning pairs to consider — all the form–meaning pairs which are grammatical in any of his epistemically possible contexts. We assume that these epistemic possibilities are given by the information state defined by the previous discourse. We will introduce Blutner structures for dynamic contexts in Section 5. This leads us to a system which allows to integrate bidirectional OT and Dynamic Semantics.

2 OT–Systems

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 get 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. This defines preferences on forms and meanings.

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 *an* F' *are ranked equal.* We first fix some terminology concerning transitive relations:

Definition 2.1 *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 $m \in M$ 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.

A constraint can be represented most naturally by a formula $\varphi(v_1, \ldots, v_n)$ with free variables. This is equivalent with a representation as a binary valued function c with n argument positions, where $c(a_1, \ldots, a_n) = 0$ iff $\varphi(a_1, \ldots, a_n)$ holds, and $c(a_1, \ldots, a_n) = 1$ iff $\varphi(a_1, \ldots, a_n)$ does not hold. If we assume in addition, that a constraint does also *rank* the tuples $\langle a_1, \ldots, a_n\rangle$, then we can identify it with a function c which maps $\langle a_1, \ldots, a_n\rangle$ into the set of natural numbers $\mathbf{N}$. If we concentrate on form–meaning pairs, then c should be defined on a set Gen of pairs $\langle F, M\rangle$ which are *generated* by the rules of an underlying grammar. These constraints induce a ranking on Gen.

We can bring these parts together. The resulting structure is called *OT–system*[4]:

Definition 2.2 (OT–System) *An OT–system is a pair $\mathcal{O} = \langle \mathrm{Gen}, C\rangle$, where* Gen *is a relation, and $C = (c_\alpha)_{1 \le \alpha < \beta}$ is a sequence of functions from* Gen *to* $\mathbf{N}$*, β an ordinal number.*

1. *Let $a \approx_{\mathcal{O}} b$ iff for all $\alpha < \beta$ $c_\alpha(a) = c_\alpha(b)$.*
2. *Let $a, b \in \mathrm{Gen}$. $a <_{\mathcal{O}} b$ iff there is an γ with $1 \le \gamma < \beta$ such that $c_\gamma(a) < c_\gamma(b)$ and for all $\alpha < \gamma$: $c_\alpha(a) = c_\alpha(b)$.*

We write $a \le_{\mathcal{O}} b$ for $a \approx_{\mathcal{O}} b \vee a <_{\mathcal{O}} b$.

[4]We follow (Jäger, 2000, Def. 4)

For technical reasons the preference relation is read here in inverse order, i.e. $a <_{\mathcal{O}} b$ means that a is preferred over b. This implies that c_1 is the strongest constraint, then comes c_2, etc.

Jäger (2000) suggests that these structures underlie also bidirectional OT. The *two–sidedness* of bidirectional OT is captured by a difference in the constraints. There are two important classes, the class of *input markedness constraints*, and the class of *output markedness* constraints[5].

Definition 2.3 *Let $\mathcal{O} = \langle \text{Gen}, C \rangle$ be an OT–system.*

- *A constraint c is an* output *markedness constraint, iff*

$$\langle F, M \rangle, \langle F', M \rangle \in \text{Gen} \Rightarrow c(\langle F, M \rangle) = c(\langle F', M \rangle).$$

- *A constraint c is an* input *markedness constraint, iff*

$$\langle F, M \rangle, \langle F, M' \rangle \in \text{Gen} \Rightarrow c(\langle F, M \rangle) = c(\langle F, M' \rangle).$$

An advantage of this move is that it allows to generalise results known for one–sided OT to bidirectional OT, which is the concern of Jäger's paper.

Jäger shows that the ranking relation $<_{\mathcal{O}}$ on Gen is well–founded for all OT–systems.

Lemma 2.4 *Let $\mathcal{O}$ be an OT–system. Then $\leq_{\mathcal{O}}$ is a linear and well–founded pre–order.*

Proof: It is clear that $\leq_{\mathcal{O}}$ is linear, and that it is a pre–order. Assume that there exists $(a_n)_{n \in \omega}$ such that $\forall n\, a_{n+1} <_{\mathcal{O}} a_n$. Let $\gamma'_n := \min\{\gamma < \beta | c_\gamma(a_{n+1}) < c_\gamma(a_n)\}$ and $\gamma_n := \omega\gamma'_n + c_{\gamma'_n}(a_n)$. It is clear that all γ_n exist and that $\gamma_{n+1} < \gamma_n$. But this contradicts the well–foundedness of the ordering of ordinals. □

This implies that we can never represent a Mattausch square in $\langle \leq_{\mathcal{O}}, \text{Gen} \rangle$:

$$\begin{array}{ccc} \langle F_1, M_1 \rangle & \longleftarrow & \langle F_2, M_1 \rangle \\ \downarrow & & \uparrow \\ \langle F_1, M_2 \rangle & \longrightarrow & \langle F_2, M_2 \rangle \end{array}$$

Where the form–meaning pairs are as follows:

$$\langle F_1, M_1 \rangle := \langle \text{She was pulling his hair out}, \textit{pulling–hair–out}(j, m) \rangle$$

$$\langle F_1, M_2 \rangle := \langle \text{She was pulling his hair out}, \textit{pulling–hair–out}(m, j) \rangle$$

$$\langle F_2, M_1 \rangle := \langle \text{He was pulling her hair out}, \textit{pulling–hair–out}(j, m) \rangle$$

$$\langle F_2, M_2 \rangle := \langle \text{He was pulling her hair out}, \textit{pulling–hair–out}(m, j) \rangle$$

Mattausch postulates the following constraints in order to generate this square: 1) Pronouns have to agree with the natural gender of the person referred to. 2) Pronouns are preferred over names. 3) The interpreter prefers a default interpretation concerning the gender of the bearer of names.

We first see that neither 1) nor 3) are input or output markedness constraints in the sense defined above. Nevertheless, they are natural constraints and should

[5](Jäger, 2000, p. 54–55)

not be excluded from bidirectional OT. The definitions above imply that the function $c \in C$ depend either on $\mathcal{F}$ or on $\mathcal{M}$. Hence, they have in fact only one true argument. The first thing which we can learn is that the two–sidedness of bidirectional OT should not be captured as a restriction on the type or number of arguments. But even if we do not restrict the class of constraints, we get only well–founded orders $\leq_{\mathcal{O}}$ on Gen. We think that OT should allow for the representation of Mattausch squares. If so, then OT–systems can not be the right structures.

Reconsidering the constraints used by Mattausch we see that they naturally divide into group 1), 2) of constraints responsible for the choice of a from F by a speaker, and constraint 3) which governs the choice of a meaning M by the interpreter. The preferences for forms depend on a meaning argument, and the choice of meanings on a form argument. Hence the constraints for the speaker induce a preference relation $\preceq_M$ for each meaning between forms, and those for the interpreter a preference relation $\preceq_F$ for each form between meanings. We collect these parts in a structure which we will call *Blutner structure*. We study them in the next section.

It is of course not clear from the beginning that these structures improve over OT–systems. But as they allow for the representation of Mattausch squares it is clear that they can't be reduced to OT–systems. We will show that Blutner structures are essentially generated by *two* OT–systems, one for the producer, and one for the interpreter. This gives us also an intuitive reason why OT–structures don't allow for circularity. We assume for sake of the argument that all four form–meaning pairs are grammatical in the given situation. Of course, all four form–meaning pairs pass the test with the second constraint. Hence, we need only to consider 1) and 3). First we assume that the constraints are ordered as 1) < 3), i.e. a pair $\langle pronoun, person\rangle$ is better for a fixed utterance context w if the gender of the pronoun agrees with the real sex of the person, than if the pronoun agrees only with the expected sex of this person. This implies that the hearer has to check at a crucial point a *speaker's* constraint: The speaker wants to express that M_1. As $\langle F_2, M_1\rangle$ fails the test with constraint 1), it follows that it is outranked by $\langle F_1, M_1\rangle$, which violates only the third constraint. Then we take the perspective of the interpreter. He again has to check all three constraints. And this means that he has to check whether the pronoun agrees with the natural gender of the person referred to. But this means that he first has to know that the pronoun has to agree with the gender of *Jo*, and he has to know that Jo is a woman. But this presupposes that he has to have access to the speaker's knowledge. He can't check constraint 1)! If we invert the order of constraints and assume that 3) < 1), then we face a similar problem for the speaker. At a crucial point he has to check an *interpreter's* constraint: If he chooses between F_1 and F_2 in order to express M_1, he will first test with 3) and get F_2 as optimal form. But intuitively this form should be outranked by *his* preference for 1).

These considerations show that we have to consider the effect of the interpreter's partial knowledge. In a Dynamic Semantics framework we can identify the hearer's knowledge with the information introduced by the previous discourse. Hence, this will play a role if it comes to integrate bidirectional OT and Dynamic Semantics. But first, we introduce Blutner structures for static contexts. The detailed discussion of Mattausch's Example in Section 4 will then motivate the necessary modifications for dynamic contexts.

3 Blutner Structures

In the last section we have seen that it is reasonable to look for structures which make the idea explicit that the constraints divide into a ranked group which provides for the speaker's preferences on forms, and a second ranked group which provides for the hearer's preferences on meanings. We call these structures *Blutner structures*. If these structures shall be acceptable, we have to show that we can model the usual reasoning in bidirectional OT with these structures. Therefore, we have to generalise the definitions of *optimality* and *weak optimality*. We essentially copy the explications of Jäger (2000). We then show how these structures are related to OT–systems. Especially, we show that they are generated by two OT–systems over the same set of grammatical form–meaning pairs, one system for the speaker and one system for the interpreter. Finally, we show how to represent arbitrary context–sensitive constraints within this framework.

Blutner Structures

OT–systems are pairs $\langle \text{Gen}, C \rangle$, where Gen is a set of form–meaning pairs which is *generated* by a given grammar. We use the same underlying structure but make the set of *forms* $\mathcal{F}$ and *meanings* $\mathcal{M}$ explicit, i.e. the underlying structure has the form $\langle \mathcal{F}, \mathcal{M}, \text{Gen} \rangle$. The following definition of *Blutner structure* can be motivated by the way how speaker and interpreter find an optimal form–meaning pair. As explained in the introduction, the speaker starts with a meaning M and searches for this meaning a grammatical form F which fits best to his preferences. This means that he has to search for a fixed M the set

$$R(M) := \{F \mid \langle F, M \rangle \in \text{Gen}\}.$$

Then, his preferences must rank the elements in this set, i.e. they must define a binary relation $\preceq_M$ on $R(M)$. We assume that $\preceq_M$ is at least a linear pre–order. A symmetric assumption has to be made for the interpreter. If we collect all the preference relations $\preceq_M$ and $\preceq_F$, then we end up with structures of the following type:

Definition 3.1 (Blutner Structure)
A Blutner structure *is a tuple* $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq \rangle$ *where*

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

We use the following terminology: We call $\mathcal{B}$ *a* two–sided *Blutner structure, iff* $P = \mathcal{F} \cup \mathcal{M}$. *If* $P = \mathcal{F}$, *then we call* $\mathcal{B}$ $\mathcal{M}$*–sided, and if* $P = \mathcal{M}$, *then* $\mathcal{F}$*–sided. We call* $\mathcal{F}$ *a set of* forms, *and* $\mathcal{M}$ *a set of meanings.*

Of course, $\mathcal{F}$ and $\mathcal{M}$ may denote any sets of objects. In case of Mattausch's Example we will define $\mathcal{F}$ as the set of sentences of natural language, and the set $\mathcal{M}$ as the set of their translations into a language of formal logic. The speaker has to choose an object F in $\mathcal{F}$ for an $M \in \mathcal{M}$, and the interpreter has to choose some object M in $\mathcal{M}$ for an $F \in \mathcal{F}$. We assume that they can only choose an F, or an M, if the resulting form–meaning pair $\langle F, M \rangle$ is an element of Gen, i.e. if it is grammatical. Now, we claimed that Blutner structures make the idea explicit that constraints divide into a group which provides for the speaker's preferences on forms, and one group which provides for the hearer's preferences on meanings. It may seem that this is not captured by the definition above because there it is assumed that for the speaker there exists for *every* form F a preference relation on meanings, and that for the interpreter there exists for *every* meaning a preference relation on forms. At first sight this is much more fine–grained. But we will see in Lemma 3.7 that Blutner structures are really generated by *two* ranked sequences of constraints provided that the inverse relations of $\preceq_F$ and $\preceq_M$ are all well–founded.

Optimality and Weak Optimality

We first want to show that some central notions of bidirectional OT can be redefined for Blutner structures. $\preceq$ induces a pre–order $\leq$ on Gen. It is the (inverse) counterpart to the order $\leq_{\mathcal{O}}$ induced by an OT–system.

Definition 3.2 *Let* $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq \rangle$ *be a Blutner structure. Then we can define a pre–order* $\leq$ *on* Gen $\times$ Gen *if we build the transitive closure of the relation defined by:* $\langle F, M \rangle \leq \langle F', M' \rangle$ *iff*

$$(F = F' \;\&\; M \preceq_F M') \vee (M = M' \;\&\; F \preceq_M F').$$

Clearly, we can find a Blutner structure where we can represent a Mattausch square in the associated $\langle \leq, \text{Gen} \rangle$. This implies, that in general we can't generate $\langle \leq, \text{Gen} \rangle$ by an OT–system. We show how to generalise the usual Blutner–Jäger definitions for optimality and weak optimality to Blutner structures.

Definition 3.3 (Optimality)
Let $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq \rangle$ *be a Blutner structure, and* $\langle \leq, \text{Gen} \rangle$ *the associated pre–order. A pair* $\langle F, M \rangle \in$ Gen *is* optimal*, iff it is a maximal element in* $\langle \leq, \text{Gen} \rangle$.

Optimality is a central notion of bidirectional OT because bidirectional OT makes the assumption that speakers and interpreters agree to choose only optimal form–meaning pairs, or *weakly optimal* form–meaning pairs. That they choose optimal form–meaning pairs can be expected from general considerations about rationality. Weak optimality is the empirically more interesting notion because its implications are less expected. But it will play no role in our future considerations. Therefore, we just show that the usual definition of Blutner and Jäger can be generalised for Blutner structures. For more motivation we refer to (Blutner, 2000). The crucial condition for weak optimality is that of Def. 3.4.

Definition 3.4 (Weak Optimality Set)
Let $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq \rangle$ *be a Blutner structure.*

A Set $S \subseteq$ Gen *is a* weak optimality set*, iff for all* $\langle F, M \rangle \in$ Gen $\langle F, M \rangle \in S$ *implies that the following two conditions hold*

1. *there is no pair* $\langle F', M\rangle \in S$ *such that* $F \prec_M F'$,

2. *there is no pair* $\langle F, M'\rangle \in S$ *such that* $M \prec_F M'$.

Intuitively, a pair $\langle F, M\rangle$ is weakly optimal, iff there is no weakly optimal $\langle F', M\rangle$ or $\langle F, M'\rangle$ which is preferred by one of the interlocutors. As it stands, this definition is circular, and, in fact, there are many sets S which have the property of Definition 3.4. We show that we can construct a weak optimality set. We can start with the set of optimal form–meaning pairs, then take away all pairs which are immediately *dominated* by an optimal form–meaning pair (the first of the disjunction in the following definition of D_α), and add then the new set of optimal pairs. This process can be repeated until no new optimal form–meaning pairs emerge. It is possible to extend the process even beyond this point (here we need also the second part of the disjunction), but the choice of new elements for S is absolutely arbitrary.

Lemma 3.5 *Let* $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \mathrm{Gen}, \preceq\rangle$ *be a Blutner structure, and* $\langle \leq, \mathrm{Gen}\rangle$ *the associated pre–order.*

1. *The set of optimal elements is a weak optimality set.*

2. *Every weak optimality set can be extended to a maximal weak optimality set by the following construction:*

   Let S_0 *be any weak optimality set,* $\mathrm{Gen}_0 := \mathrm{Gen}$*, and for* $\alpha > 0$ *we set:*

   (a) $S_{<\alpha} := \bigcup_{\beta<\alpha} S_\beta$,

   (b) $D_\alpha := \{\langle F, M\rangle \in \mathrm{Gen} \mid \exists \langle F', M'\rangle \in S_{<\alpha} : (F = F' \vee M = M') \wedge (\langle F, M\rangle \leq \langle F', M'\rangle \vee \langle F', M'\rangle \leq \langle F, M\rangle)\}$,

   (c) $\mathrm{Gen}_\alpha := \mathrm{Gen} \setminus D_\alpha$,

   (d) $S_\alpha := S_{<\alpha}$ *together with some arbitrary element of* Gen_α*, if there is one.*

3. *If we start the construction with the set of all optimal elements, and replace* S_α *by*

   $$S_\alpha := S_{<\alpha} \text{ together with the set of optimal elements for } \langle \leq, \mathrm{Gen}_\alpha\rangle\,,$$

   then the maximal set WOT *which we can construct in this way is a weak optimality set, and it is the intersection of all weak optimality sets which extend* WOT*.*

Proof: That the set of optimal elements is a weak optimality set follows trivially by definition.

For 2: The union $\bigcup_{\beta<\alpha} S_\beta$ and all S_α are weak optimality sets. This follows directly by definition. Furthermore, $\beta < \alpha$ implies $S_\beta \subseteq S_\alpha$. It follows by the Lemma of Zorn that we ultimately find a maximal weak optimality set.

For 3: That WOT is a weak optimality set follows by induction over α. Assume that S is a weak optimality set which extends WOT. Let $\langle F, M\rangle \in S \setminus WOT$. Then, there is no $\langle F', M'\rangle \in WOT$ such that $(F = F' \vee M = M') \wedge (\langle F, M\rangle \leq \langle F', M'\rangle \vee \langle F', M'\rangle \leq \langle F, M\rangle)$, and $\langle F, M\rangle$ can't be an optimal element for $\langle \preceq, \mathrm{Gen}_\infty\rangle$, where we set $D_\infty := \{\langle F, M\rangle \in \mathrm{Gen} \mid \exists \langle F', M'\rangle \in$

$WOT : (F = F' \vee M = M') \wedge (\langle F, M\rangle \leq \langle F', M'\rangle \vee \langle F', M'\rangle \leq \langle F, M\rangle)\}$, and $\text{Gen}_\infty := \text{Gen} \setminus D_\infty$. Hence, there must exist a $\langle F', M'\rangle \in \text{Gen}_\infty (F = F' \vee M = M') \wedge (\langle F, M\rangle \leq \langle F', M'\rangle)$. As $\langle F', M'\rangle \in \text{Gen}_\infty$ we can start with WOT, choose $\langle F', M'\rangle$, and construct a maximal weak optimality set S'. But $\langle F, M\rangle \notin S'$ by definition of $D_{\infty+1}$, hence, $\langle F, M\rangle \notin S \cap S'$. □

Definition 3.6 (Weak Optimality)

We call a form–meaning–pair $\langle F, M\rangle$ weakly optimal *for a Blutner structure $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq\rangle$, iff $\langle F, M\rangle$ is an element of WOT.*

Blutner Structures and OT–Systems

Now we want to show how Blutner structures are related to underlying OT–systems. We will see that every Blutner structure $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq\rangle$ where the inverse of $\preceq_F$ and $\preceq_M$ are well–founded relations can be generated by two OT–systems over the same set of grammatical form–meaning pairs, i.e. they are generated by pairs $\langle \mathcal{O}_\mathcal{F}, \mathcal{O}_\mathcal{M}\rangle$ of OT–systems $\mathcal{O}_X = \langle \text{Gen}, C_X\rangle$. As both systems have the same set Gen they differ only with respect to the sequence C of constraints. The construction below shows that $C_\mathcal{F}$ is responsible for the definition of the preferences $\preceq_M$ on $\mathcal{F}$, and $C_\mathcal{M}$ for the preferences $\preceq_F$ on $\mathcal{M}$. But this means that the interpreter's preferences for meanings are defined by a sequence of constraints, and the speaker's preferences for forms by a separate sequence. This shows that Blutner structures really incorporate the idea that preferences are defined relative to *two* separate sequences of constraints, one for the speaker and one for the interpreter. We will see in the next section that we have to weaken the condition that both OT–systems are defined over the *same* set Gen if we want to capture dynamic contexts.

We first show how OT–systems over a fixed set Gen of form–meaning pairs generate Blutner structures. Let $\mathcal{O} = \langle \text{Gen}, C\rangle$ be an OT–systems over a set of form–meaning pairs $\text{Gen} \subseteq \mathcal{F} \times \mathcal{M}$ with $C = \langle c_\alpha\rangle_{\alpha<\beta}$. We can then use the result of Jäger to define for each $M \in \mathcal{M}$ and $F \in \mathcal{F}$ an associated linear well–founded pre–order $<_{\mathcal{O}_M}$ on $R(M) := \{F \in \mathcal{F} \mid \langle F, M\rangle \in \text{Gen}\}$, and $<_{\mathcal{O}_F}$ on $R(F) := \{M \in \mathcal{M} \mid \langle F, M\rangle \in \text{Gen}\}$. For $\langle F, M\rangle, \langle F', M'\rangle \in \text{Gen}$ we set:

1. $F <_{\mathcal{O}_M} F'$ iff there is an γ with $1 \leq \gamma < \beta$ such that $c_\gamma(F, M) < c_\gamma(F', M)$ and for all $\alpha < \gamma : c_\alpha(F, M) = c_\alpha(F', M)$.
2. $M <_{\mathcal{O}_F} M'$ iff there is an γ with $1 \leq \gamma < \beta$ such that $c_\gamma(F, M) < c_\gamma(F, M')$ and for all $\alpha < \gamma : c_\alpha(F, M) = c_\alpha(F, M')$.

Then, we can define $\preceq_F$ as the inverse of $<_{\mathcal{O}_F}$, and $\preceq_M$ in the same way. If we assume furthermore that $\mathcal{F}$ and $\mathcal{M}$ are disjoint, and set $\preceq = (\preceq_p)_{p\in P}$ for $P = \mathcal{F}$ or $P = \mathcal{M}$, then $\langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq\rangle$ is a one–sided Blutner structure.

It is easy to see that two one–sided Blutner structures $\mathcal{B}_\mathcal{F}$ and $\mathcal{B}_\mathcal{M}$, where $\mathcal{B}_\mathcal{F}$ is $\mathcal{F}$–sided, and $\mathcal{B}_\mathcal{M}$ is $\mathcal{M}$–sided, generate a two–sided Blutner structure $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, (\preceq_p)_{p\in P}\rangle$ if we set $P = P_\mathcal{F} \cup P_\mathcal{M}$. Hence, we see that any pair $\langle \mathcal{O}_1, \mathcal{O}_2\rangle$ with $\mathcal{O}_i = \langle \text{Gen}, C_i\rangle$ generate a two–sided Blutner structure.

Now we show also the inverse. Again it is easy to see that we can separate a two–sided Blutner structure in a $\mathcal{F}$–sided and a $\mathcal{M}$–sided structure.

Hence, we show that we can find for any $\mathcal{F}$–sided Blutner structure $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq\rangle$ where the inverse of $\preceq_M$ is a well–founded pre–order for all $M \in \mathcal{M}$ an underlying OT–system $\langle \text{Gen}, C_\mathcal{F}\rangle$ which generates $\mathcal{B}$.

Lemma 3.7 *For every $\mathcal{F}$–sided Blutner structure $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq \rangle$ where the inverse $\leq_M$ of $\preceq_M$ is a well–founded pre–order for all $M \in \mathcal{M}$ there exists an OT–system $\langle \text{Gen}, C_{\mathcal{F}} \rangle$, such that for all $\langle F, M \rangle, \langle F', M' \rangle \in \text{Gen}$:*

$$F \leq_M F' \iff F \leq_{\mathcal{O}_{\mathcal{F}}} F'.$$

Proof: Let $\leq_M$ be the inverse relation of $\preceq_M$ on $R(M)$. By assumption, $\leq_M$ is well–founded. For each $F \in R(M)$ we can define equivalence classes $[F]_M := \{F' \in R(M) \mid F \leq_M F' \;\&\; F' \leq_M F\}$, where $[F]_M \leq_M [F']_M \Leftrightarrow F \leq_M F'$. Hence $\langle \{[F]_M \mid F \in R(F)\}, \leq_M \rangle$ is well–ordered for each $M \in \mathcal{M}$. Let $f(M, F)$ be the associated order type of $[F]_M$. Then we set:

$$\begin{aligned} c_\alpha(M, F) := 0 \quad &\iff \quad f(M, F) = \alpha \\ c_\alpha(M, F) := 1 \quad &\iff \quad f(M, F) \neq \alpha. \end{aligned}$$

We set $\mathcal{O}_{\mathcal{F}} := \langle \text{Gen}, C_{\mathcal{F}} \rangle$ with $C_{\mathcal{F}} := \langle c_\alpha \rangle_{\alpha < \beta_{\mathcal{F}}}$, where $\beta_{\mathcal{F}} := \sup_{M \in \mathcal{M}} (\beta_M + 1)$, and $\beta_M := \sup_{F \in R(M)} (f(M, F) + 1)$. Let $F, F' \in R(M)$, $\gamma := f(M, F)$ and $\gamma' := f(M, F')$. Then we find:

$F \leq_M F'$ *iff* $\gamma \leq \gamma'$ *iff* $(\forall \alpha < \beta_M \, c_\alpha(M, F) = c_\alpha(M, F')) \vee \forall \alpha < \gamma \, (c_\alpha(M, F) = c_\alpha(M, F') \wedge c_\gamma(M, F) < c_\gamma(M, F'))$ *iff* $F \leq_{\mathcal{O}_{\mathcal{F}}} F'$. □

The same result holds for $\mathcal{M}$. Hence, we can sum up our result as:

Lemma 3.8 *Let $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq \rangle$ be a Blutner structure where the inverse of $\preceq_F$ and $\preceq_M$ are well–founded relations. Then, there is a pair $\langle \mathcal{O}_{\mathcal{F}}, \mathcal{O}_{\mathcal{M}} \rangle$ of OT–systems $\mathcal{O}_{\mathcal{F}} = \langle \text{Gen}, C_{\mathcal{F}} \rangle$ and $\mathcal{O}_{\mathcal{M}} = \langle \text{Gen}, C_{\mathcal{M}} \rangle$ which generate $\mathcal{B}$.*

Blutner Structures with Context–Sensitive Constraints

Our structures represent cases where the constraints depend on arguments of the form $\langle F, M \rangle$. We want to extend Blutner structures in order to be able to handle arbitrary contexts given by a set $\mathcal{C}$. In case of Mattausch's Example we considered a constraint which says that the gender of a pronoun must agree with the sex of the person referred to. The real person is neither part of a natural sentence nor of its translation into a sentence of formal logic. Later, in Sections 4 and 5 we define structures where the real world and the information state of the interpreter are possible arguments for constraints. But we will show that Blutner structure with contexts $\mathcal{C}$ are not really new structures, i.e. that they can be considered to be a Blutner structure in the sense of Def. 3.1.

Definition 3.9 (Blutner Structure with Contexts)
A Blutner structure with contexts *is a tuple $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \mathcal{C}, \text{Gen}, \preceq \rangle$ where*

- *$\mathcal{F}$ and $\mathcal{M}$ are disjoint sets.*
- *$\mathcal{C}$ is a set.*
- 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 $\{M \mid \langle c, F, M \rangle \in \text{Gen}\}$.*
  - *$\preceq_{c,M}$ is a linear pre–order on $\{F \mid \langle c, F, M \rangle \in \text{Gen}\}$.*

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

We show that every Blutner structure $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \mathcal{C}, \mathrm{Gen}, \preceq \rangle$ with contexts can be represented as a Blutner structure in the sense of Definition 3.1.

Let $\langle \mathcal{F}, \mathcal{M}, \mathcal{C}, \mathrm{Gen}, \preceq \rangle$ be given. Then we set:

$$\begin{aligned} \mathcal{M}' &:= \{\langle c, M\rangle \mid \exists F \in \mathcal{F}\ \langle c, F, M\rangle \in \mathrm{Gen}\}, \\ \mathcal{F}' &:= \{\langle c, F\rangle \mid \exists M \in \mathcal{M}\ \langle c, F, M\rangle \in \mathrm{Gen}\}, \\ \mathrm{Gen}' &:= \{\langle \langle c, F\rangle, \langle c, M\rangle\rangle \mid \langle c, F, M\rangle \in \mathrm{Gen}\}. \end{aligned}$$

Then we can define $\preceq$:

$$\begin{aligned} \langle c, M\rangle \preceq'_{\langle c,F\rangle} \langle c, M'\rangle \quad &\textit{iff} \quad M \preceq_{c,F} M' \\ \langle c, F\rangle \preceq'_{\langle c,M\rangle} \langle c, F'\rangle \quad &\textit{iff} \quad M \preceq_{c,M} M'. \end{aligned}$$

$\preceq'_{\langle c,F\rangle}$ and $\preceq'_{\langle c,M\rangle}$ are linear pre–orders on the sets $\{\langle c, M\rangle \mid \langle c, F, M\rangle \in \mathrm{Gen}\}$ and $\{\langle c, F\rangle \mid \langle c, F, M\rangle \in \mathrm{Gen}\}$ respectively. Hence, $\mathcal{B}' = \langle \mathcal{F}', \mathcal{M}', \mathrm{Gen}', \preceq' \rangle$ is a Blutner structure in the sense of Def. 3.1.

We now show that they both induce the *same* preference relation on Gen and Gen$'$. The map

$$e:\ \mathrm{Gen} \longrightarrow \mathrm{Gen}', \langle c, F, M\rangle \mapsto \langle \langle c, F\rangle, \langle c, M\rangle\rangle$$

is a bijection. Let $\leq$ be the pre–order on Gen induced by $\mathcal{B}$, and $\leq'$ the pre–order on Gen$'$ induced by $\mathcal{B}'$. Then, for all $a, b \in \mathrm{Gen}$ it holds that:

$$a \leq b \iff e(a) \leq' e(b).$$

This follows by:

$$\begin{aligned} &\langle c, F, M\rangle \leq \langle c', F', M'\rangle \ \Leftrightarrow \\ &\quad \Leftrightarrow c = c' \ \&\ (M = M' \wedge F \preceq_{c,M} F' \vee F = F' \wedge M \preceq_{c,F} M') \\ &\quad \Leftrightarrow c = c' \ \&\ (M = M' \wedge \langle c, F\rangle \preceq_{\langle c,M\rangle} \langle c, F'\rangle \vee F = F' \wedge \langle c, M\rangle \preceq_{\langle c,F\rangle} \langle c, M'\rangle) \\ &\quad \Leftrightarrow \langle \langle c, F\rangle, \langle c, M\rangle\rangle \leq' \langle \langle c', F'\rangle, \langle c', M'\rangle\rangle \end{aligned}$$

This allows us to generalise all results to Blutner structures with contexts.

4 Mattausch's Example

In this section we show how to solve Mattausch's Problem (2000) within a Bi-OT approach using Blutner structures. Beaver (2000) applied a version of a two–sided OT to predict choice and explain resolution of anaphoric expressions. Mattausch (2000) improves on this approach, and his discussion of bidirectional OT in the version of (Blutner, 2000), (Blutner, Jäger, 2000) or (Jäger, 2000) leads him to consider the following example (Mattausch, 2000, pp.33–36):

(3) Assume that Marion is a male person, and Jo a female one. 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.

The considered form–meaning pairs are

$$\langle F_1, M_1\rangle := \langle \text{She was pulling his hair out}, \textit{pulling–hair–out}(j, m)\rangle$$

$$\langle F_1, M_2\rangle := \langle \text{She was pulling his hair out}, \textit{pulling–hair–out}(m, j)\rangle$$

$$\langle F_2, M_1\rangle := \langle \text{He was pulling her hair out}, \textit{pulling–hair–out}(j, m)\rangle$$

$$\langle F_2, M_2\rangle := \langle \text{He was pulling her hair out}, \textit{pulling–hair–out}(m, j)\rangle$$

$$\langle F_3, M_1\rangle := \langle \text{Jo was pulling Marion's hair out}, \textit{pulling–hair–out}(j, m)\rangle$$

We need the following constraints: 1) Pronouns have to agree with the natural gender of the person referred to. 2) Pronouns are preferred over names. 3) *Marion* is interpreted by default as female, and *Jo* as male. 1) and 2) define preferences on forms, 3) on meanings. These constraints lead into the following circle:

> In case of *pulling–hair–out*(j, m) 'She was pulling his hair out' is the preferred form over 'He was pulling her hair out'.
>
> In case of 'She was pulling his hair out' *pulling–hair–out*(m, j) is the preferred meaning over *pulling–hair–out*(j, m).
>
> In case of *pulling–hair–out*(m, j) 'He was pulling her hair out' is the preferred form over 'She was pulling his hair out'.
>
> In case of 'He was pulling her hair out' *pulling–hair–out*(j, m) is the preferred meaning over *pulling–hair–out*(m, j).

This results into a circular structure:

$$\begin{array}{ccccc} \langle F_3, M_1\rangle & \longrightarrow & \langle F_1, M_1\rangle & \longleftarrow & \langle F_2, M_1\rangle \\ & & \downarrow & & \uparrow \\ & & \langle F_1, M_2\rangle & \longrightarrow & \langle F_2, M_2\rangle \end{array}$$

The situation where the speaker and hearer have to co–ordinate their choice of an optimal form–meaning pair is given by the context generated after the utterance of *Marion was frustrated with Jo*. The second sentence contains pronouns, hence, it is natural to seek for a dynamic framework to handle Mattausch's Example. In a first step we will combine Dynamic Semantics with bidirectional OT using Blutner structures for (static) contexts. The underlying idea will be that Dynamic Semantics accounts for the meaning of formulas in a given context, and OT provides for the pragmatic constraints which govern anaphora resolution.

That OT should be seen in context with Dynamic Semantics has been suggested by Blutner (2000). The idea that the two theories should work together in anaphora resolution goes back to (Beaver, 2000), who was followed by Mattausch (2000).

Let $\mathcal{L}$ be a set of well–formed formulas of a first order language. We assume that for all sentences of natural language there are translations in $\mathcal{L}$. This translation should be unique up to substitutions of free variables. We denote the set of all translations for a natural sentence F by F^*. Let W be a set of models for $\mathcal{L}$ representing the possible states of affairs. Let $\mathcal{W}$ be a set of pairs (w, f) where f is a partial assignment function for variables and $w \in W$. We denote the set of variables by Var . In Dynamic Semantics it is usual to identify the *meaning* of a sentence with its *update potential*, i.e. with a relation between world–assignment pairs. Hence, we assume that for every formula $\varphi \in \mathcal{L}$ $[\![\varphi]\!]$ is a set of pairs $\langle\sigma, \tau\rangle$ such that $\sigma, \tau \subseteq \mathcal{W}$. Furthermore, we assume that $[\![\varphi]\!]$ is a function. This means that the update effect of a sentence of natural language is given once we have resolved anaphora. This allows us to define the *meaning* $[\![\varphi]\!]_\sigma$ *of a formula* φ *in context* σ: It is defined if $\sigma \in \mathrm{dom}[\![\varphi]\!]$, and $[\![\varphi]\!]_\sigma = \tau \Leftrightarrow \sigma[\![\varphi]\!]\tau$. The subsets σ, τ of $\mathcal{W}$ are called *information states*. They represent — and this is important for the following representation of contexts — the *hearer's* knowledge.

The task for speaker and hearer in Mattausch's Example is to find the correct resolution for and choice of the pronouns *he/she* and *his/her* in the environment *... was pulling ... hair out*.

In general, the set of forms $\mathcal{F}$ for our Blutner structure is given by the set of all syntactically correct sentences of natural language. But for the present purpose we can restrict our considerations to the set:

$$\mathcal{F} := \{\underbrace{\text{She was pulling his hair out}}_{F_1}, \underbrace{\text{He was pulling her hair out}}_{F_2}\}$$

The set $\mathcal{M}$ is the set of all possible translations:

$$\mathcal{M} := \{\varphi \mid \varphi \in \mathcal{L}\} = \mathcal{L}$$

As the information of the hearer is represented by an information state $\sigma \subseteq \mathcal{W}$, his choice of an optimal translation should be determined by this set. Whereas the speaker should be able to rely for his choice of the best form also on the true state of affairs. Hence, we consider contexts of the following form: $c = \langle w, \sigma\rangle$ where w represents the actual world, and $\sigma \subseteq \mathcal{W}$ the information state of the hearer, i.e.

$$\mathcal{C} := \{\langle w, \sigma\rangle \mid w \in W \,\&\, w \in \sigma \subseteq \mathcal{W}\}.$$

Now we have to say which form–meaning pairs can be generated in which contexts. The elements in Gen have the form $\langle c, F, \varphi\rangle$, where c is a pair $\langle w, \sigma\rangle$. We have to consider the constraints imposed by our dynamic semantic and the fact that φ must be a possible translation for F. This leads to the following definition:

$$\mathrm{Gen} := \{\langle\langle w, \sigma\rangle, F, \varphi\rangle \mid \varphi \in F^* \,\&\, \sigma \in \mathrm{dom}[\![\varphi]\!] \,\&\, \exists f\, (w, f) \in [\![\varphi]\!]_\sigma\}.$$

Gen is the set of all $\langle c, F, \varphi\rangle$ such that 1) φ is a possible translation for the natural sentence F, 2) the information state of the hearer can be updated with φ, and 3) the existential closure of φ must be a true sentence in w. $\mathrm{dom}[\![\varphi]\!]$ denotes the domain of φ. Hence, if $\sigma \in \mathrm{dom}[\![\varphi]\!]$, then $[\![\varphi]\!]_\sigma$ is defined. Gen is the set of all proper *grammatical* form–meaning pairs.

For pragmatic reasons this set would be too restrictive. This has to do with the limited *perspectives* of speaker and interpreter. In our example the hearer can rely for his interpretation in a context $\langle w, \sigma\rangle$ only on the information state

σ. Hence, he has to consider all form–meaning pairs $\langle F, \varphi\rangle$ such that there is a context $\langle w', \sigma\rangle$ where $\langle F, \varphi\rangle$ is proper grammatical. This means, that the hearer's set of alternatives is defined by:

$$\text{Gen}_H := \{\langle\langle w, \sigma\rangle, F, \varphi\rangle \mid \exists w' \in W\ \langle\langle w', \sigma\rangle, F, \varphi\rangle \in \text{Gen}\}.$$

Whereas the speaker has full knowledge about the utterance situation. I.e. the set of forms where his choice takes place is determined by the set Gen.

Of course, this is not covered by our framework set up in the last section. It motivates the last modifications which we will provide subsequent to the discussion of Mattausch's Example.

Now that we have set up the general framework we can consider Mattausch's problem in more detail. We denote the object *Marion* by m, and *Jo* by j. As short forms we use:

$$\begin{array}{lll} \varphi(v_1, v_2) & :\Leftrightarrow & \textit{pull–hair–out}(v_1, v_2), \\ \psi(v_1, v_2) & :\Leftrightarrow & \textit{Marion}(v_1) \wedge \textit{Jo}(v_2) \wedge \textit{frustrated–with}(v_1, v_2). \end{array}$$

We assume that the actual world is such that

$$w_0 \models \psi(m, j) \wedge \varphi(j, m) \wedge \textit{male}(m) \wedge \textit{female}(j)$$

We assume that the hearer chooses a variable x for *Marion* and a variable y for *Jo* when processing the sentence *Marion was frustrated with Jo.* His information state is then represented by the set

$$\sigma_0 = \{(w, f) \mid x, y \in \text{dom}f \;\&\; (w, f) \models \textit{Marion}(x) \wedge \textit{Jo}(y) \wedge \textit{frustrated–with}(x, y)\}.$$

This implies that the hearer has to choose the variable x as a translation for a pronoun if it should refer to Marion, and y if it should refer to Jo. With (Mattausch, 2000) we do not assume that $\textit{Marion}(x) \Rightarrow \textit{female}(x)$ or $\textit{Jo}(y) \Rightarrow \textit{male}(y)$ forms part of the meaning of *Marion* and *Jo.* But we have to assume — and here we again follow Mattausch (2000) — that it is part of the default interpretation.

The context where the speaker has to choose the best form, and where the hearer has to find the correct translation is given by the pair $c = \langle w_0, \sigma_0\rangle$.

The speaker wants the hearer to update his information state with:

$$\tau_1 := \{(w, f) \mid x, y \in \text{dom}f \;\&\; f(x) = m \;\&\; f(y) = j \;\&\; (w, f) \models \varphi(y, x)\}.$$

This is the meaning of the formula $\varphi(y, x)$ in σ_0, i.e. $[\![\varphi(y, x)]\!]_{\sigma_0} = \sigma_0 \cap \tau_1$. Now it is the assumption for Mattausch's Example that the best choice for the speaker is a pronoun which agrees with the natural gender of the persons it is intended to refer to. Hence, he should choose *He* for Marion and *She* for Jo. This means that he prefers in context $c = \langle w_0, \sigma_0\rangle$ F_1 over F_2, i.e. $F_2 \prec_{c,\varphi(y,x)} F_1$.

Now we have to take the perspective of the interpreter and choose the best translation for F_1. We assume here that it is part of the semantics of a pronoun whether or not its referent is male or female. Of course, it may be more appropriate to call it a presupposition. But we don't want to enter here into the subtleties of presupposition theory. It follows that the set of translations for F_1

is given by[6]:

$$F_1^* = \{\underbrace{\varphi(v_1, v_2) \wedge \mathit{female}(v_1) \wedge \mathit{male}(v_2)}_{\mu(v_1,v_2)} \mid v_1, v_2 \in \mathrm{Var}\}$$

The set of alternatives where the choice takes place in context $c = \langle w_0, \sigma_0 \rangle$ is defined as:

$$R_c(F_1) = \{\lambda \in \mathcal{L} \mid \exists w \in W \ \langle \langle w, \sigma_0 \rangle, F_1, \lambda \rangle \in \mathrm{Gen}_H\}$$

Hence, $R_c(F_1) = \{\mu(v_1, v_2) \mid v_1, v_2 \in \mathrm{Var} \,\&\, \sigma_0 \in \mathrm{dom}[\![\mu(v_1, v_2)]\!] \,\&\, \exists w \exists f\, (w, f) \in [\![\mu(v_1, v_2)]\!]_{\sigma_0}\} = \{\mu(v_1, v_2) \mid (v_1 = x \wedge v_2 = y) \vee (v_1 = y \wedge v_2 = x)\} = \{\mu(x, y), \mu(y, x)\}$. Here we have to assume that x and y are the only variables interpreted so far by the hearer.

We impose the same constraints as Mattausch for the choice of optimal meanings. Hence, we have to assume that the hearer prefers default interpretations over non–default interpretations. But this means that he has to choose y for *He* and x for *She* because $\sigma_0 \models \mathit{Jo}(y) \,\&\, \mathit{Marion}(x)$. Hence, he will translate F_1 into $\mu(x, y)$, and update his information state with the set:

$$\tau_2 := \{(w, f) \mid x, y \in \mathrm{dom} f \,\&\, (w, f) \models \varphi(x, y) \wedge \mathit{female}(x) \wedge \mathit{male}(y)\}.$$

Now we have to switch to the perspective of the speaker again. He has to choose the optimal sentence in the context $c = \langle w_0, \sigma_0 \rangle$ for $\mu(x, y)$. Hence, in accordance with Definitions 3.9 he has to choose the preferred form in the set

$$R_c(\mu(x, y)) = \{F \in \{F_1, F_2\} \mid \langle c, F, \mu(x, y) \rangle \in \mathrm{Gen}\}.$$

This is the set of all $F \in \{F_1, F_2\}$ such that

$$\mu(x, y) \in F^* \,\&\, \sigma_0 \in \mathrm{dom}[\![\mu(x, y)]\!] \,\&\, \exists f\, (w_0, f) \in [\![\mu(x, y)]\!]_{\sigma_0}.$$

$[\![\mu(x, y)]\!]_{\sigma_0}$ is the set of all $(w, f) \in \sigma_0$ such that:

$$x, y \in \mathrm{dom} f \,\&\, (w, f) \models \varphi(x, y) \wedge \mathit{female}(x) \wedge \mathit{male}(y).$$

Hence, it is the set of all (w, f) where $x, y \in \mathrm{dom} f$, and where in (w, f) holds:

$$\mathit{Marion}(x) \wedge \mathit{Jo}(y) \wedge \psi(x, y) \wedge \varphi(x, y) \wedge \mathit{female}(x) \wedge \mathit{male}(y).$$

But for all (w_0, f) with $x, y \in \mathrm{dom} f$ and $f(x) = \mathit{Marion}$ and $f(y) = \mathit{Jo}$ it holds that $(w_0, f) \models \neg\varphi(x, y) \wedge \neg\mathit{female}(x) \wedge \neg\mathit{male}(y)$. Hence

$$R_c(\mu(x, y)) = \emptyset.$$

[6]Of course, our decision to make the gender of a pronoun part of the meaning, and therefore of its translation, has some repercussions on our previous definitions because now F_1 is no more a possible choice for the producer. This follows because $\varphi(y, x)$ is not a translation for F_1, hence $\langle \langle w_0, \sigma_0 \rangle, F_1, \varphi(y, x) \rangle \notin \mathrm{Gen}$. This problem is due to the strong condition for grammaticality which implies that the formula has to be an exact translation. But we can weaken this condition: The exact translation has to imply the formula. This is without influence on our subsequent discussion. The definition of Gen then is as follows: $\langle \langle w, \sigma \rangle, F, \varphi \rangle \in \mathrm{Gen}$ iff

$$\sigma \in \mathrm{dom}[\![\varphi]\!] \wedge w \in [\![\varphi]\!]_\sigma \wedge \exists \psi \in F^*\ (\sigma \in \mathrm{dom}[\![\psi]\!] \wedge w \in [\![\psi]\!]_\sigma \wedge [\![\psi]\!]_\sigma \subseteq [\![\varphi]\!]_\sigma).$$

I.e. in this context the speaker can find no grammatical sentence to choose. This implies that $\langle\langle w_0,\sigma_0\rangle, F_1, \mu(x,y)\rangle$ is optimal in Gen_H. But the last result also shows that it is not an element of Gen. Intuitively, $\langle F_1, \mu(x,y)\rangle$ should not be counted as an optimal form–meaning pair in context $\langle w_0,\sigma_0\rangle$. Of course, our considerations suggest that this is due to the fact that this pair is an element of $\text{Gen}_H \setminus \text{Gen}$, i.e. because it can't be generated by our grammar for this context. If we restrict the search for optimal pairs to Gen, then we face the problem that $\langle F_1, \mu(y,x)\rangle$ would be optimal. But this is neither an intuitive nor reasonable solution as clearly the hearer should translate F_1 as $\mu(x,y)$. Hence, we postulate the following principle:

> We call the elements of $\text{Gen}_H \setminus \text{Gen}$ *dead ends*, and require for *optimality* that the considered form–meaning pair is maximal in Gen and that it has no dead ends as successors in Gen_H.

Don't force the interpreter to choose an ungrammatical meaning! The principle can also be incorporated into the notion of *weak optimality*.

We first show how the Mattausch problem can be solved with this principle, and then we provide for the precise description of the necessary modifications of Blutner structures.

According to our definition, the pair $\langle$*She was pulling his hair out*$, \mu(x,y)\rangle$ is a dead end in context $c = \langle w_0,\sigma_0\rangle$. Hence, the triple $\langle c, F_1, \mu(y,x)\rangle$ is a maximal element in Gen but it should not be counted as optimal because $\mu(y,x) \prec_{c,F_1} \mu(x,y)$ and $\langle c, F_1, \mu(x,y)\rangle \notin \text{Gen}$. The principle stated above implies that we have to exclude all elements in Gen which necessarily lead into a dead end, i.e. we have to remove all triples of the form $\langle c, F_1, \,.\,\rangle$. In this way we reduce the set of possible choices for the speaker, and we get new optimal form–meaning pairs. Especially, F_1 is no more a possible choice. We now consider Mattausch's problem in its full version with originally three choices for the speaker, i.e. with $\mathcal{F} := \{F_1, F_2, F_3\}$ where

$F_1 :=$ She was pulling his hair out,

$F_2 :=$ He was pulling her hair out,

$F_3 :=$ Jo was pulling Marion's hair out.

Our considerations show that the choice of F_1 for $\varphi(y,x)$ in situation $\langle w_0,\sigma_0\rangle$ leads necessarily into a dead end, hence our mechanism removes it from $\mathcal{F}$. F_2 is neither a possible choice because $\langle\langle w_0,\sigma_0\rangle, F_2, \varphi(y,x)\rangle$ is not an element of Gen. Hence, it remains only F_3 as a possible choice. Of course, $\langle\langle w_0,\sigma_0\rangle, F_3, \varphi(y,x)\rangle$ is grammatical, and it is easy to see that $\varphi(y,x)$ is also the preferred translation for F_3 for the interpreter. Hence, it turns out that

$$\langle\langle w_0,\sigma_0\rangle, \text{Jo was pulling Marion's hair out}, \textit{pull–hair–out}(y,x)\rangle$$

is optimal.

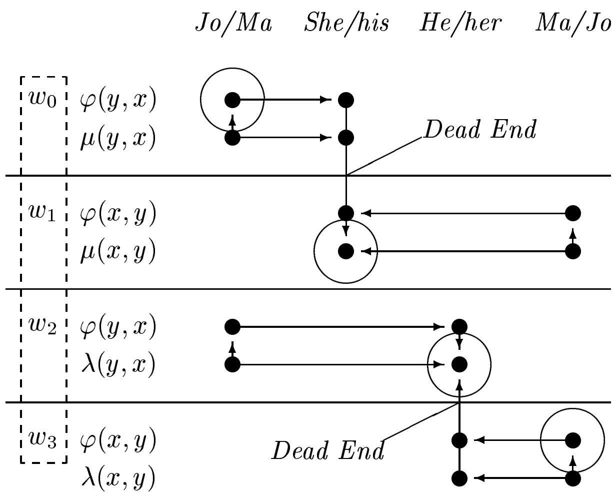
The following picture provides a graphical solution for Mattausch's Problem. In the first row there are listed the different forms the speaker can choose. In the first column we listed the different contexts. In fact, we listed only the world w of the pair $\langle w,\sigma\rangle$ because σ is in all four cases the same set, namely $\sigma_0 = \{w_0, w_1, w_2, w_3\}$. All four situations are indiscernible for the interpreter, indicated by the dashed box around the worlds. In the second column we listed the different formulas which represent the possible translations for the forms. We used 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 $Marion(x) \wedge Jo(y) \wedge frustrated\text{–}with(x, y)$, and where the formulas listed in the second column hold.



The big dots represent the form–meaning pairs which can be generated in the context listed in the first column. The horizontal arrows show the preferences of the speaker, the vertical ones the preferences of the interpreter. The vertical arrows which cross the lines separating the different contexts indicate that they lead to a dead end. Of course, the dead ends itself are not listed in the picture. The circles around the big dots indicate the optimal form–meaning pairs.

5 Bidirectional OT for Dynamic Contexts

It remains to describe the structures which we used to solve Mattausch's Problem in more detail. First we notice that it was essential that the set of meanings where the interpreter can choose the best meaning may depend on a different set Gen_H of possible form–meaning pairs than the speaker's choice. Of course, this set is generated by an equivalence relation on contexts. This has also some consequences for the preference relation on meanings and forms: The preference relation on forms which determines the speaker's choice is defined for the original contexts, whereas the interpreters preference relation on meanings is defined for equivalence classes of contexts.

We have seen in Section 3 that we can identify a Blutner structure with a pair of OT–systems $\langle \mathcal{O}_\mathcal{F}, \mathcal{O}_\mathcal{M} \rangle$, or equivalently with a pair $\langle \mathcal{B}_\mathcal{F}, \mathcal{B}_\mathcal{M} \rangle$ of two one–sided Blutner structures. Both systems are built over the same set of grammatical form–meaning pairs Gen. We use this representation and generalise it in order to cover dynamic contexts.

Definition 5.1 (Blutner Structures for Dynamic Contexts) *Let $\langle \mathcal{B}_{\mathcal{F}}, \mathcal{B}_{\mathcal{M}}, [\ ] \rangle$ be a triple where*

1. $B_{\mathcal{F}} = \langle \mathcal{F}, \mathcal{M}, \mathcal{C}, \mathrm{Gen}, \preceq \rangle$ *is a $\mathcal{F}$–sided Blutner structure.*
2. $\mathcal{B}_{\mathcal{M}} = \langle \mathcal{F}, \mathcal{M}, \mathcal{C}_H, \mathrm{Gen}_H, \preceq \rangle$ *is a $\mathcal{M}$–sided Blutner structure.*
3. $[\ ] : \mathcal{C} \longrightarrow \mathcal{C}_H$, $c \mapsto [c]$, *maps $\mathcal{C}$ onto $\mathcal{C}_H$.*
4. $\mathrm{Gen}_H = \{\langle [c], F, M \rangle \mid \langle c, F, M \rangle \in \mathrm{Gen}\}$

We write $c' \in [c]$ iff $[c'] = [c]$.

As $[\ ]$ is a function, it follows that $c \sim c' \Leftrightarrow [c] = [c']$ defines an equivalence relation. This definition is slightly more general than it is needed for dynamic contexts. There, contexts had the form $\langle w, \sigma \rangle$, where w represents the real utterance situation, and σ the common ground. If we set $[\langle w, \sigma \rangle] := \{\langle v, \sigma \rangle \mid v \in \sigma\}$, or simply $[\langle w, \sigma \rangle] := \sigma$, then Gen_H is of the form:

$$\mathrm{Gen}_H = \{\langle [\langle w, \sigma \rangle], F, \varphi \rangle \mid \langle \langle w, \sigma \rangle, F, \varphi \rangle \in \mathrm{Gen}\}.$$

Hence, for a fixed $F \in NL$ we find that the set of meanings $R_{[\langle w,\sigma \rangle]}(F)$ where the interpreter can make his choice is given by:

$$\begin{aligned} R_{[\langle w,\sigma \rangle]}(F) &= \{\varphi \in \mathcal{L} \mid \langle [\langle w, \sigma \rangle], F, \varphi \rangle \in \mathrm{Gen}_H\} \\ &= \{\varphi \in \mathcal{L} \mid \exists v \in \sigma\ \langle \langle v, \sigma \rangle, F, \varphi \rangle \in \mathrm{Gen}\}. \end{aligned}$$

This means that he has to consider a formula φ as a translation for a sentence F, iff there is any epistemically possible context for him where $\langle F, \varphi \rangle$ is grammatical. This is exactly as intended.

A two–sided Blutner structure for dynamic contexts again induces a pre–order $\leq$ on Gen:

Definition 5.2 *Let $\mathcal{B} = \langle \mathcal{F}, \mathcal{M}, \mathrm{Gen}, \preceq \rangle$ be a $\mathcal{F}$–sided Blutner structure, and let $\langle \mathcal{B}_{\mathcal{F}}, \mathcal{B}_{\mathcal{M}}, [\ ] \rangle$ be a Blutner structure for dynamic contexts. Then we define a pre–order $\leq$ on $\mathrm{Gen} \times \mathrm{Gen}$ by: $\langle c, F, M \rangle \leq \langle c', F', M' \rangle$ iff*

$$(c = c' \;\&\; M = M' \;\&\; F \preceq_{c,M} F') \vee (c' \in [c] \;\&\; F = F' \;\&\; M \preceq_{[c],F} M').$$

Let us assume for a while that the sets of forms and meanings are finite, and that the preference relations $\preceq$ are proper linear orders. If we assume furthermore that the induced order $\leq$ on Gen is well–founded, then it is known that there exists an algorithm which leads necessarily to an optimal form–meaning pair. The idea goes back to (Blutner, 1998) and (Blutner, 2000), and is discussed in more detail in (Jäger, 2000). The algorithm starts with some meaning M_1, then the speaker chooses the most preferred form F_1. Then the interpreter chooses his most preferred meaning M_2 for F_1, and then the speaker again chooses a form F_2 for M_2. This defines a sequence $(a_n)_{n \in \mathbf{N}}$ of forms and meanings where $a_{2n} \in \mathcal{M}$ and $a_{2n+1} \in \mathcal{F}$. In case $\langle \leq, \mathrm{Gen} \rangle$ is generated by an

OT–system, Jäger (2000) shows that this sequence ultimately chooses always the same forms and meanings. Of course, this holds as long as $\langle\leq, \text{Gen}\rangle$ is well–founded. If $\langle\leq, \text{Gen}\rangle$ is generated by a Blutner structure $\mathcal{B}$ then this sequence becomes a circle, i.e. there exist $M, N \in \mathbf{N}$ such that for all $m \geq M$ and for all $k \in \mathbf{N}$ $a_{m+k \cdot N} = a_m$. But it may happen that it does not always choose the same forms and meanings, i.e. that $a_{m+2} \neq a_m$. Now, if we consider Blutner structures for dynamic contexts, then a third case can occur: the sequence can run into a dead end, i.e. it may happen that the interpreter chooses a meaning M such that there is no F for the speaker to choose. Hence, it may happen that there is an $m \in \mathbf{N}$ such that the sequence can not be extended beyond a_m. This happens exactly, if the interpreter prefers in context c a meaning M for which there is no grammatical triple $\langle c, F, M\rangle \in \text{Gen}$.

Definition 5.3 (Dead Ends)
Let $\mathcal{B}_{\mathcal{F}} = \langle \mathcal{F}, \mathcal{M}, \text{Gen}, \preceq\rangle$ be a $\mathcal{F}$–Blutner structure, and let $\langle \mathcal{B}_{\mathcal{F}}, \mathcal{B}_{\mathcal{M}}, [\ ]\rangle$ be a Blutner structure for dynamic contexts. Dead ends *are a set of ungrammatical form–meaning pairs for a certain context:*

$$DE := \{\langle c, F, M\rangle \mid \langle [c], F, M\rangle \in \text{Gen}_H \;\&\; \langle c, F, M\rangle \notin \text{Gen}\}$$

A form–meaning pair leads into *a dead end, if it is an element of DE^+, where DE^+ is the set of all $\langle c, F, M\rangle \in \text{Gen}$:*

$$\exists M' : M \prec_{[c],F} M' \;\&\; \forall M''(M' \preceq_{[c],F} M'' \Rightarrow \langle c, F, M''\rangle \in DE).$$

Definition 5.4 (Optimality)
Let $\langle \mathcal{B}_{\mathcal{F}}, \mathcal{B}_{\mathcal{M}}, [\ ]\rangle$ be a Blutner structure for dynamic contexts. Let $\langle\leq, \text{Gen}\rangle$ be the induced order. We call an element $\langle c, F, M\rangle \in \text{Gen}$ optimal *for $\langle \mathcal{B}_{\mathcal{F}}, \mathcal{B}_{\mathcal{M}}, [\ ]\rangle$, iff it is maximal in $\langle\leq, \text{Gen} \setminus DE^+\rangle$.*

Now we can make precise how the preferences of speaker and interpreter enter into the resolution problem of pronouns, and therefore how it determines the information update.

The resolution problem was given by an (underspecified) translation operation * which maps a sentence of natural language to a subset of the sentences of a formal language $\mathcal{L}$, and by a dynamic semantic for $\mathcal{L}$. Hence, we can single out three structures which determine the optimal choice of pronouns and their translation:

> $\langle NL, \mathcal{L}, ^*\rangle$ where $^* : NL \longrightarrow \mathcal{P}(\mathcal{L})$ correlates natural sentences and formal sentences. The translation part.
>
> $\langle \mathcal{L}, \mathcal{W}, [\![\ ]\!]\rangle$ where $\mathcal{W}$ is a set of world–assignment pairs, and where $[\![\ ]\!]$ is a function from $\mathcal{L}$ into sets of pairs $\langle \sigma, \tau\rangle$ of subsets of $\mathcal{W}$. The dynamic semantic part.
>
> $\langle \mathcal{B}_{\mathcal{F}}, \mathcal{B}_{\mathcal{M}}, [\ ]\rangle$ the Blutner structure with contexts $\mathcal{C} = \{\langle w, \sigma\rangle \mid w \in \sigma \subseteq \mathcal{W}\}$. The Blutner structure is determined by $\langle NL, \mathcal{L}, ^*\rangle$ and $\langle \mathcal{L}, \mathcal{W}, [\![\ ]\!]\rangle$ up to preferences, i.e. $\mathcal{F} = NL$, $\mathcal{M} = \mathcal{L}$, and
>
> $$\text{Gen} := \{\langle \langle w, \sigma\rangle, F, \varphi\rangle \mid \varphi \in F^* \;\&\; \sigma \in \text{dom}[\![\varphi]\!] \;\&\; \exists f\, (w, f) \in [\![\varphi]\!]_\sigma\}.$$

They together define a transition system $\langle \mathcal{C}, NL, \longrightarrow \rangle$ such that for all $c = \langle w, \sigma \rangle, c' = \langle w', \sigma' \rangle \in \mathcal{C}$:

$$c \xrightarrow{F} c' \iff w = w' \;\&\; \exists \varphi \in \mathcal{L}\, (\langle c, F, \varphi \rangle \text{ is optimal} \wedge \sigma' = [\![\varphi]\!]_\sigma).$$

Now, if we assume that the preferences induced by the OT–constraints are linear orders (not only pre–orders), then there can be only one φ such that $\langle c, F, \varphi \rangle$ is optimal. Hence, we can derive for a fixed world w a dynamic semantic for *NL*: $\langle \mathcal{W}, NL, [\![\ ]\!] \rangle$ with

$$\sigma [\![F]\!] \tau \iff \langle w, \sigma \rangle \xrightarrow{F} \langle w, \tau \rangle\,.$$

6 Conclusion

The aim of our paper was to outline a framework which allows to combine bidirectional OT and Dynamic Semantics. We introduced Blutner structures which improve over OT–systems. Blutner structures are generated by two OT–systems, one for the speaker and one for the interpreter. They allow

- to generate circular preferences on form–meaning pairs,
- to consider arbitrary context–sensitive constraints,
- to explain the role of the interlocutors' different epistemic perspectives in dynamic contexts.

Our model for Mattausch's Example shows that the circularity is only apparent. If we take utterance context and epistemic context into account, then the circularity vanishes. But this is an empirical finding, no more a restriction imposed a priori by the underlying OT–structure. The consideration of the role of epistemic perspectives lead us to postulate an additional restriction on the set of form–meaning pairs which can count as *optimal* choices. As the interpreter has only limited information about the utterance context he may prefer a meaning M for a form F such that $\langle F, M \rangle$ is not grammatical in the given context. We called such a pair a *dead end*. We claimed that only those form–meaning pairs can count as optimal which don't lead into a dead end, i.e. the speaker is not allowed to choose a form which misleads the interpreter.

References

D. Beaver (2000): *The Optimization of Discourse*; ms, Stanford.

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

R. Blutner (2000): *Some Aspects of Optimality in Natural Language Interpretation*; to appear in *Journal of Semantics*.

R. Blutner, G. Jäger (2000): *Against Lexical Decomposition in Syntax*; to appear in A.Z. Wyner (ed.): *Proceedings of IATL 15*, University of Haifa.

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

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

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

Contents