# Some Notes on the Formal Properties of Bidirectional Optimality Theory

GERHARD JÄGER
*Zentrum für Allgemeine Sprachwissenschaft, Jägerstr. 10/11, D-10117 Berlin, Germany*
*E-mail: jaeger@zas.gwz-berlin.de*
*phone: +49-30-20192-408*

**Abstract.** In this paper, we discuss some formal properties of the model of bidirectional Optimality Theory that was developed in Blutner (2000). We investigate the conditions under which bidirectional optimization is a well-defined notion, and we give a conceptually simpler reformulation of Blutner's definition. In the second part of the paper, we show that bidirectional optimization can be modeled by means of finite state techniques. There we rely heavily on the related work of Frank and Satta (1998) about unidirectional optimization.

**Keywords:** Optimality Theory, Formal Languages, Regular Languages, Finite State Transducers

**Keywords:** Optimality Theory, Formal Languages, Regular Languages, Finite State Transducers

## 1. Introduction

Optimality Theory (OT henceforth) has been introduced by Prince and Smolensky (1993) mainly as a model for generative phonology, but in recent years this approach has been applied successfully to a range of syntactic phenomena, and it is currently gaining popularity in semantics and pragmatics as well. It rests on the old conception that the mapping from one level of linguistic representation to another level should be described in terms of rules and filters. Such a distributed description is frequently more concise and elegant than a formulation solely in terms of rules. The novel contribution of OT lies in the idea that filters—or, synonymously, constraints—are ranked and violable. So the result of a certain sequence of rule applications may be licit even if it violates some constraints, provided all alternative derivations lead to more severe constraint violations. Violations of higher ranked constraints count as more severe than violations of lower ranked constraints.

OT is attractive for working linguists mainly for two reasons. First, the ideas of constraint ranking and of different degrees of severity of constraint violations have been part of the linguistic folklore since decades. OT supplies a concise and mathematically clean formalization of these concepts. Furthermore, OT offers an intriguing perspective on language typology on the one hand and language universals on the other hand. Many OT researchers use the working hypothesis that both the underlying rules and the constraints are universal, while languages differ only according to the ranking of the constraints.

In the generative tradition of syntax, phonology and morphology, the focus of interest have been on transformation rules, i.e. mappings from underlying abstract representations to concrete surface representations. OT researchers usually adopt this perspective too; competition takes place between different possible realizations of some underlying form. In other words, OT usually takes the generation perspective. It is a theory about the optimal realization of a given underlying form.

On a somewhat more abstract level, the OT philosophy can be described by the idea that only the most economical candidates of a given candidate set are legitimate linguistic objects; less economical competitors are blocked. Ranked constraints serve to induce an ordering on the candidates that makes optimization possible. The idea of optimization has a long history in semantics and pragmatics too, and it is suggestive to integrate this tradition into the OT framework. Some caution has to be exerted here though. The generation perspective that is prevalent in phonology and morphology has some plausibility when applied to semantics. Here it amounts to saying that a certain verbalization of a given meaning, though licit, might be blocked by a more economical linguistic form expressing the same meaning. Such effects do in fact occur. A case in point is the well-known phenomenon of "conceptual grinding", where the name of an animal kind is used to refer to meat of this animal:

(1) We had chicken for dinner.

However, conceptual grinding is only possible if there is no lexicalized expression for the kind of meat in question:

(2) a. ?We had pig for dinner.
    b. We had pork for dinner

Arguably, using the lexicalized expression *pork* is a more economical way to refer to meat from pigs than using the noun *pig* in its shifted meaning. Thus (2b) blocks (2a).

On the other hand, there is also a considerable tradition in semantics and pragmatics which assumes that a certain interpretation of a given linguistic form may be blocked by a more coherent alternative interpretation of the same form. In other words, the candidate set for optimization in semantics may also be determined by the parsing perspective, where we compare different interpretations of a given surface form. A typical example is the behavior of presupposition accommodation. Consider the following sentence:

(3) If Mary becomes a politician, the president will resign

The consequent of this conditional contains a definite NP and thus a presupposition trigger. The presupposition triggered is *there is a president.* If we assume that the sentence is uttered out of the blue, this presupposition is not entailed by the context and hence must be accommodated. According to the literature on presupposition accommodation, there are three structural options for accommodation in a construction like (3), namely local, intermediate and global accommodation (cf. Heim (1990), van der Sandt (1992)). The resulting readings can be paraphrased as in (4):

(4) a. If Mary becomes a politician, there is [a president]$_i$, and he$_i$ will resign (= local accommodation)

b. If Mary becomes a politician and there is [a president]$_i$, then he$_i$ will resign (= intermediate accommodation)

c. There is [a president]$_i$, and if Mary becomes a politician, he$_i$ will resign (= global accommodation)

There is agreement in the literature that global accommodation is preferred, thus we (correctly) predict (3) to be interpreted as (4c).

Now consider the slightly altered example

(5) If Mary becomes member of [a club]$_i$, its$_i$ president will resign

The definite NP *its$_i$ president* triggers the presupposition *it$_i$ has a president*. The three options for accommodation now come out as

(6) a. If Mary becomes member of [a club]$_i$, it$_i$ has [a president]$_j$, and he$_j$ will resign (= local accommodation)

b. If Mary becomes member of [a club]$_i$ which has [a president]$_j$, he$_j$ will resign (= intermediate accommodation)

c. *It$_i$ has [a president]$_j$, and if Mary becomes member of [a club]$_i$, he$_j$ will resign (= global accommodation)

Now global accommodation is blocked because the pronoun *it* occupies a structural position where it cannot be bound by the indefinite *a club*. In such a configuration, intermediate accommodation becomes the preferred option. Hence (5) is interpreted as (6b).

A concise way to describe this pattern is to assume that the grammar generally admits all three kinds of accommodation, but that global accommodation is more economical than intermediate one (which is in turn more economical than local accommodation). So if a construction structurally admits all readings, global accommodation wins and blocks all competing readings. If global accommodation is blocked, intermediate accommodation wins.

So it seems that the mapping of linguistic forms to interpretations requires optimization both in the parsing and in the generation direction. This insight is not new; some form of bidirectional optimization has been assumed in the pragmatics literature for quite some time (see for instance Horn (1984) and Levinson (1987)). In a series of recent publications, Reinhard Blutner has made the interplay between generation optimization and parsing optimization precise and integrated it into the overall framework of OT (Blutner (1998), Blutner (2000)).

It has frequently been observed that a naive evaluation algorithm for an OT style theory is computationally extremely costly even if the candidate sets involved are finite. One might add that the problem is even more severe if the candidate sets are infinite. Then we cannot be sure whether the set of optimal candidates is recursive, even if all components (the generator relation and the constraints involved) are. The issue of the automata theoretic complexity of OT style theories is currently a topic of active research, and several interesting results have been reported in the literature. The most intriguing one is Frank and Satta (1998). There it is shown that under certain general restrictions, (unidirectional) optimization is a finite state technique. This means that an OT-system can be implemented as a finite state transducers provided the underlying generator relation is a rational relation and all constraints are regular languages. In other words, if all components of an OT-system are finite state objects, the system as a whole is so too.

The plan for the present paper is the following. In the next section, we will have a closer look at Blutner's formalization of bidirectional OT. We will propose a simplified but equivalent definition, and we will investigate some properties of bidirectional OT-systems. Section 3 briefly reviews the basic notions of finite state automata, and it discusses Frank and Satta's construction. In section 4 the complexity of bidirectional OT will be considered. As main result, we show that an analogue of Frank and Satta's result can be obtained for bidirectional optimization as well. Section 5 sums up the findings and lists a couple of open question for future research.

## 2. Bidirectional OT: Z vs. X

The notions of parsing optimization and generation optimization have ancestors in the literature on formal pragmatics from the eighties. There several authors assumed an interplay of the competing forces of speaker economy and hearer economy. A representative of this line of thought are the principles “Q” and “I” proposed in Horn (1984), p. 13:

*Q-principle:* Say as much as you can (given I).

*I-principle:* Say no more than you must (given Q).

In Blutner (1998) and Blutner (2000) this idea is formalized. Following standard practise in OT theories, Blutner assumes that there is a (very general and underspecified) relation **GEN** that relates input to output. In case of the syntax-semantics interface, **GEN** can be identified with the compositional semantics that relates syntactic structures and meanings. Furthermore, Blutner assumes an ordering relation on form-meaning pairs. In OT theories, this ordering is induced by a set of ranked constraints, but this is inessential for the notion of optimization as such. So let us just assume that $<$ is an ordering on **GEN**. We adopt the convention that "$a < b$" is to be understood as "$a$ is more economical than $b$".

Given this, Blutner formalizes Horn's principles as follows:[1]

DEFINITION 1. (Blutner's Bidirectional Optimality).

1. $\langle f, m\rangle$ satisfies the Q-principle iff $\langle f, m\rangle \in$ **GEN** and there is no other pair $\langle f', m\rangle$ satisfying the I-principle such that $\langle f', m\rangle < \langle f, m\rangle$.

2. $\langle f, m\rangle$ satisfies the I-principle iff $\langle f, m\rangle \in$ **GEN** and there is no other pair $\langle f, m'\rangle$ satisfying the Q-principle such that $\langle f, m'\rangle < \langle f, m\rangle$.

3. $\langle f, m\rangle$ is optimal iff it satisfies both the Q-principle and the I-principle.

In contrast, standard (unidirectional) OT boils down to a version of the I-principle; only different outputs for a given input are compared.

DEFINITION 2. (Unidirectional Optimality). $\langle f, m\rangle$ is unidirectionally optimal iff
$\langle f, m\rangle \in$ **GEN** and there is no other pair $\langle f, m'\rangle < \langle f, m\rangle$.

Seen in a procedural way, to check whether a given form-meaning pair $\langle f, m\rangle$ is optimal in Blutner's sense, you have first to check whether it satisfies the I-principle and than whether it satisfies the Q-principle. To do the former, you have to test whether there are alternatives $\langle f', m\rangle < \langle f, m\rangle$ that satisfy the I-principle. To this end, you have to go through competitors $\langle f', m'\rangle < \langle f', m\rangle$ that possibly satisfy the Q-principle etc. The shape of this zigzag pattern (graphically sketched in figure 1) resembles the letter "Z". Therefore I will call Blutner's notion of optimality **Z-optimality**. Taken in isolation, this definition might seem circular, since the Q-principle indirectly occurs in the definiens of this very principle, and likewise for the I-principle. This is not a real problem, however, since we may safely assume that the ordering relation $<$ is well-founded.[2] We will see below that this follows from the fact that $<$ is induced by a system of ranked constraints. Given this, it follows from the General Recursion Theorem that Z-optimality is well-defined. Recall that the General Recursion Theorem says:

[1] We change notation and terminology slightly without touching the content of the definition.

[2] A relation $R$ is well-founded iff there are no infinite descending $R$-chains, i.e. there is no infinite sequence $a_1, a_2, a_3, \ldots$ with $a_{i+1} R a_i$ for all $i \in \mathbb{N}$.

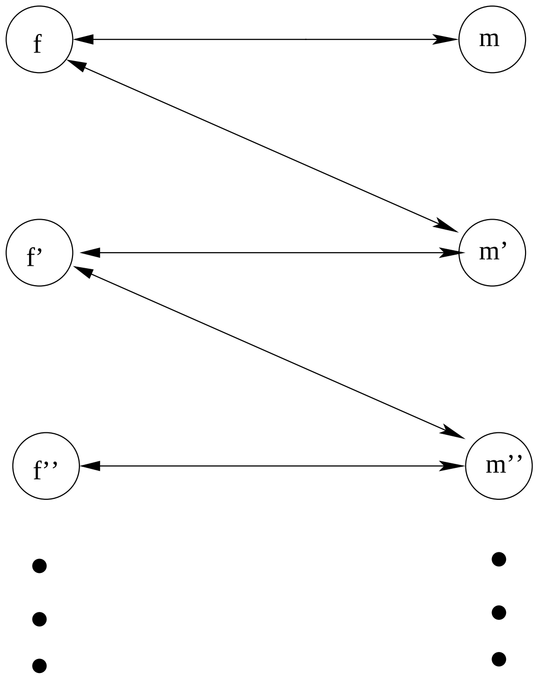


*Figure 1.* Z-Optimality

THEOREM 1. (General Recursion Theorem). Suppose $H$ is a two-place operation and $R$ a locally well-founded relation.[3] Then the equation

$$\forall x[F(x) = H(x, F \restriction \{y|yRx\})]$$

has exactly one solution for $F$.

As an immediate consequence, we get

LEMMA 1. If $<$ is well-founded, z-optimality is uniquely defined by Definition 1.

*Proof.* Let $F_z$ be the function that returns the pair of truth values $\langle q, i\rangle$ for a given input $x$. $q = 1$ iff $x$ is a form-meaning pair $\langle f, m\rangle$ that satisfies the Q-principle, and likewise for $i$. $G$ is assumed to be the characteristic function of the graph of **GEN**, i.e. it returns 1 iff its argument is in **GEN** and 0 otherwise. Given this, we can reformulate the first two clauses of definition 1 as a fixed point equation for $F_z$. (The projection functions $\pi_1, \pi_2$ return the first and the second element

[3] A relation is called *locally* well-founded iff it is well-founded and it holds for each $x$ that the class of $R$-predecessors of $x$ is a set (rather than a proper class). Formally put, this means that $\forall x \exists y. y = \{z|zRx\}$.

respectively of an ordered pair.)

$$\begin{aligned} F_z(x) = \langle & min(G(x), \\ & 1 - max(\{0\} \cup \{\pi_2(F_z(y))|y < x \wedge \pi_2(y) = \pi_2(x)\})), \\ & min(G(x), \\ & 1 - max(\{0\} \cup \{\pi_1(F_z(y))|y < x \wedge \pi_1(y) = \pi_1(x)\}))\rangle \end{aligned}$$

In the right hand side of this equation, $F_z$ is only applied to predecessors of $x$ with respect to $<$, so we may replace $F_z$ there with $F_z \upharpoonright \{y|y < x\}$. Since $<$ is well-founded by assumption, it follows from the General Recursion Theorem that there is a unique solution for $F_z$. Now we reproduce the third clause of Definition 1 as *$x$ is z-optimal iff* $F_z(x) = \langle 1, 1\rangle$. ⊣

In the sequel we will develop a conceptually somewhat different notion of bidirectional optimality, x-optimality, and we will show that under very general conditions, x-optimality and z-optimality coincide.

On a somewhat metaphorical level, the Q-principle above expresses speaker economy. It says: for a given meaning, choose the most economical verbalization you can think of. Symmetrically, the I-principle captures hearer economy. It advises a hearer to pick out the most economical licit interpretation for a given form. Now the main objective of the participants of a conversation should be successful communication, one should think. Economy considerations can only be taken into account if the main objective is granted. The following two definitions capture this intuition.

1. A form-meaning pair $\langle f, m\rangle$ is speaker-optimal iff
   a) $\langle f, m\rangle \in$ **GEN**,
   b) $\langle f, m\rangle$ is hearer-optimal, and
   c) there is no $\langle f', m\rangle \in$ **GEN** that is also hearer-optimal and that is more economical than $\langle f, m\rangle$.
2. A form-meaning pair $\langle f, m\rangle$ is hearer-optimal iff
   a) $\langle f, m\rangle \in$ **GEN**,
   b) $\langle f, m\rangle$ is speaker-optimal, and
   c) there is no $\langle f, m'\rangle \in$ **GEN** that is also speaker-optimal and that is more economical than $\langle f, m\rangle$.

According to these definitions, speaker-optimality entails hearer-optimality and vice versa. Thus these two notions of optimality coincide and we may identify them. So simplified versions of the above definitions run as follows:

1. A form-meaning pair $\langle f, m \rangle$ is optimal iff
   a) $\langle f, m \rangle \in$ **GEN**,
   b) $\langle f, m \rangle$ is optimal, and
   c) there is no $\langle f', m \rangle \in$ **GEN** that is also optimal and that is more economical than $\langle f, m \rangle$.
2. A form-meaning pair $\langle f, m \rangle$ is optimal iff
   a) $\langle f, m \rangle \in$ **GEN**,
   b) $\langle f, m \rangle$ is optimal, and
   c) there is no $\langle f, m' \rangle \in$ **GEN** that is also optimal and that is more economical than $\langle f, m \rangle$.

Now these definitions have the form $\phi \leftrightarrow \psi \wedge \phi \wedge \chi$, which, according to elementary propositional reasoning, is equivalent to $\phi \rightarrow \psi \wedge \chi$. So we can further simplify to

1. A form-meaning pair $\langle f, m \rangle$ is optimal only if
   a) $\langle f, m \rangle \in$ **GEN**,
   b) there is no $\langle f', m \rangle \in$ **GEN** that is also optimal and that is more economical than $\langle f, m \rangle$.
2. A form-meaning pair $\langle f, m \rangle$ is optimal only if
   a) $\langle f, m \rangle \in$ **GEN**,
   b) there is no $\langle f, m' \rangle \in$ **GEN** that is also optimal and that is more economical than $\langle f, m \rangle$.

One more step of propositional reasoning (from $(\phi \rightarrow \psi) \wedge (\phi \rightarrow \chi)$ to $\phi \rightarrow \psi \wedge \chi$) yields

- A form-meaning pair $\langle f, m \rangle$ is optimal only if
  1. $\langle f, m \rangle \in$ **GEN**,
  2. there is no $\langle f', m \rangle \in$ **GEN** that is also optimal and that is more economical than $\langle f, m \rangle$,
  3. there is no $\langle f, m' \rangle \in$ **GEN** that is also optimal and that is more economical than $\langle f, m \rangle$.

This is not a good definition yet since it is an implication rather than a biconditional, and there may be many sub-relations of **GEN** that obey this constraint. In particular, the empty relation would count as an optimality-relation. What is still missing there is the intuition

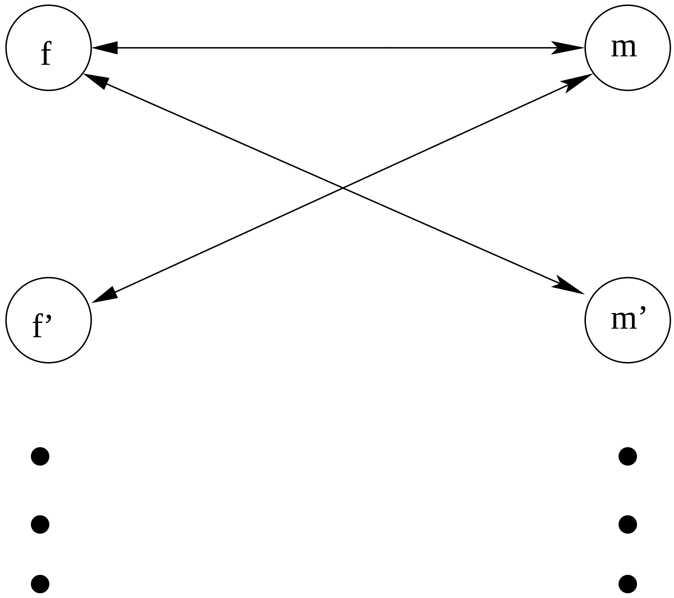


*Figure 2.* X-Optimality

that a given form-meaning pair is optimal if there is no reason to the contrary. So the optimal form-meaning relation we are after should be the largest subrelation of **GEN** that obeys the above constraint. This amounts to turning the implication into a biconditional. For reasons that will become obvious immediately, we call this notion of optimality **x-optimality**.

DEFINITION 3. (X-Optimality). A form-meaning pair $\langle f, m \rangle$ is x-optimal iff

1. $\langle f, m \rangle \in$ **GEN**,

2. there is no x-optimal $\langle f', m \rangle$ such that $\langle f', m \rangle < \langle f, m \rangle$.

3. there is no x-optimal $\langle f, m' \rangle$ such that $\langle f, m' \rangle < \langle f, m \rangle$.

Checking whether a form-meaning pair is x-optimal requires simultaneous evaluation of form alternatives and meaning alternatives of this pair (see figure 2 ). This structure resembles the letter "X"—this motivates the name. Under the proviso that $<$ is well-founded, x-optimality is also well-defined. Furthermore, if we additionally assume $<$ to be transitive, x-optimality coincides with z-optimality.

THEOREM 2. If "$<$" is transitive and well-founded, then

1. there is a unique x-optimality relation

2. $\langle f, m\rangle$ is x-optimal iff it is z-optimal.

*Proof.* The proof of part 1 is analogous to the proof of the corresponding property of z-optimality. Here we rewrite the definition as the fixed point equation

$$\begin{aligned} F_x(x) \;=\; & min(G(x), 1 - max(\{0\} \cup \{F_x(y)|y < x \wedge \\ & (\pi_1(y) = \pi_1(x) \vee \pi_2(y) = \pi_2(x))\})) \end{aligned}$$

A candidate $x$ is x-optimal iff $F_x(x) = 1$ according to the unique solution for $F_x$.

As for part 2, suppose $\langle f, m\rangle$ is x-optimal but not z-optimal. This means that it either violates the I-principle or the Q-principle. Suppose it violates the I-principle. Then there is an $m'$ with $\langle f, m'\rangle < \langle f, m\rangle$ such that $\langle f, m'\rangle$ satisfies the Q-principle. Since $\langle f, m\rangle$ is x-optimal, $\langle f, m'\rangle$ cannot be x-optimal. Thus there is either an x-optimal $\langle f, m''\rangle < \langle f, m'\rangle$ or an x-optimal $\langle f', m'\rangle < \langle f, m'\rangle$. The first option is excluded since if it were the case, by transitivity, $\langle f, m''\rangle < \langle f, m\rangle$, thus contradicting the assumption that $\langle f, m\rangle$ is x-optimal. So there is an x-optimal $\langle f', m'\rangle < \langle f, m'\rangle < \langle f, m\rangle$. Since $\langle f, m'\rangle$ satisfies the Q-principle, $\langle f', m'\rangle$ does not satisfy the I-principle. By repeated application of this argument, we can construct an infinite chain $\ldots < \langle f''', m'''\rangle < \langle f'', m''\rangle < \langle f', m'\rangle < \langle f, m\rangle$, all members being x-optimal and violating the I-principle. This is excluded by the assumption that "<" well-founded, so $\langle f, m\rangle$ cannot violate the I-principle if it is x-optimal. By a symmetric argument, we conclude that it cannot violate the Q-principle either, so it is z-optimal.

As for the other direction, suppose $\langle f, m\rangle$ is z-optimal but not x-optimal. Then there is either an x-optimal $\langle f', m\rangle < \langle f, m\rangle$ or an x-optimal $\langle f, m'\rangle < \langle f, m\rangle$. Suppose the former is the case. From the previous paragraph we know that any x-optimal candidate satisfies the Q-principle, so $\langle f', m\rangle$ satisfies the Q-principle since it is x-optimal. This is excluded though since by assumption, $\langle f, m\rangle$ satisfies the I-principle. By the same kind of reasoning, we also derive a contradiction if $\langle f, m\rangle$ is blocked by some $\langle f, m'\rangle$. ⊣

It remains to be shown that the ordering relation that is induced by a system of ranked constraints in an OT style system is in fact transitive and well-founded. To this end, we have to make precise what an OT style system is. In the general case, it consists of a relation **GEN** and a finite set of constraints that are linearly ordered by some constraint ranking.[4] Constraints may be violated several times. So a constraint

[4] Some authors only require the constraints to be partially ordered. Since a given candidate is optimal according to some partial ordering iff it is optimal according to all total extensions of this partial ordering, the results obtained in this section can easily be extended to this more general setup.

should be construed as a function from **GEN** into the natural numbers. Thus an OT-system assigns every pair in **GEN** a finite sequence of natural numbers. The ordering of the elements of **GEN** that is induced by the OT-system is according to the lexicographic ordering of these sequences. This leads to the following definition:

DEFINITION 4. (OT-System).

1. An OT-system is a pair $\mathcal{O} = \langle \mathbf{GEN}, C \rangle$, where **GEN** is a relation, and $C = \langle c_1, \ldots, c_p \rangle, p \in \mathbb{N}$ is a linearly ordered sequence of functions from **GEN** to $\mathbb{N}$.

2. Let $a, b \in \mathbf{GEN}$. $a <_{\mathcal{O}} b$ iff there is an $i$ with $1 \leq i \leq p$ such that $c_i(a) < c_i(b)$ and for all $j < i : c_j(a) = c_j(b)$.

LEMMA 2. Let $\mathcal{O}$ be an OT-system. Then $<_{\mathcal{O}}$ is transitive and well-founded.

*Proof.* We assign every element of **GEN** an ordinal number by the function $f$ that is defined by

$$f(x) = \sum_{i=1}^{p} \omega^{p-i} \cdot c_i(x)$$

It is easy to see that $x <_{\mathcal{O}} y$ iff $f(x) < f(y)$. Since the ordering of the ordinal numbers is transitive and well-founded, so is $<_{\mathcal{O}}$. ⊣

## 3. OT and finite state techniques: Frank and Satta's result

In most research papers on OT, the candidate sets that are taken under consideration are finite and even fairly small, and the search for the optimal candidate is done manually by comparing the patterns of constraint violations. It has frequently been observed that in realistic applications, candidate sets might be very large, which would render this kind of naive brute force algorithm computationally very expensive. Even worse, if the candidate set may be infinite, there is no guarantee this kind of algorithm terminates. Thus the success of the OT research program crucially hinges on the issue whether there are computationally tractable evaluation algorithms.

It is obvious that the complexity of the task of finding the optimal candidates for a given OT-system heavily depends on the complexities of **GEN** and of the constraints. In the general case, these will provide a lower bound for the complexity of the OT-system as a whole, both in terms of automata theoretic complexity and in terms of resource complexity. The crucial question is whether an OT-system as a whole may have a higher complexity than the most complex of its components. Furthermore, this issue may depend on the mode of evaluation that we choose. For instance, unidirectional OT might be less complex than bidirectional OT.

While these issues are still open in the general case, the literature contains some promising results about the complexity of unidirectional OT in cases where all components of the OT-system are finite state objects. These insights might be of practical importance in phonology and morphology, where finite state techniques are usually sufficiently expressive. In syntax and semantics, this kind of result cannot be employed immediately since it is well-known that more automata-theoretic power is needed here. Nevertheless the finite state case is interesting since it indicates that the OT mechanism as such is not all that powerful after all.

In this section we briefly review some basic properties of finite state objects, and we will discuss the most impressive piece of work on the complexity of OT, Frank and Satta's (1998) construction. This will pave the ground for the extrapolation of Frank and Satta's result to the bidirectional case that is to be presented in the next section.

In the subsequent discussion of finite state automata, finite state transducers, regular languages and rational relations, we make heavy use of Roche and Schabes (1997). The interested reader is referred there for further information and references.

We assume that the reader is familiar with the basic concepts of a finite state automaton and a regular language and give the definition here for reference.

DEFINITION 5. (FSA). A finite-state automaton $A$ is a 5-tuple $\langle \Sigma, Q, i, F, E\rangle$, where $\Sigma$ is a finite set called the *alphabet*, $Q$ is a finite set of *states*, $i \in Q$ is the *initial state*, $F \subseteq Q$ is the set of final states, and $E \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ is the set of *edges*.

Following standard practice, we use $\Sigma^*$ to refer to the set of strings over the alphabet $\Sigma$, including the empty string. The letter $\varepsilon$ symbolizes the empty string.

DEFINITION 6. The *extended set of edges* $\hat{E} \subseteq Q \times \Sigma^* \times Q$ is the smallest set such that

1. $\forall q \in Q, \langle q, \varepsilon, q\rangle \in \hat{E}$
2. $\forall w \in \Sigma^*$ and $\forall a \in \Sigma \cup \{\varepsilon\}$, if $\langle q_1, w, q_2\rangle \in \hat{E}$ and $\langle q_2, a, q_3\rangle \in E$, then $\langle q_1, wa, q_3\rangle \in \hat{E}$.

A finite-state automaton $A$ defines the following language $L(A)$:

$$L(A) = \{w \in \Sigma^* | \exists q \in F. \langle i, w, q\rangle \in \hat{E}\}$$

If $\mathcal{L} = L(A)$, we say that the FSA $A$ *recognizes* the language $\mathcal{L}$. The class of *regular languages* is the class of languages that are recognized by some FSA.

A finite state transducer (FST) is a FSA that produces an output. Every edge of the automaton is labeled with an input and an output, where both input and output are strings over the input alphabet and the output alphabet respectively. An FST does not just recognize strings but transforms inputs strings in output strings.

DEFINITION 7. (FST). A *Finite-State Transducer* is a tuple $\langle \Sigma_1, \Sigma_2, Q, i, F, E\rangle$ such that

- $\Sigma_1$ is a finite alphabet, namely the *input alphabet*
- $\Sigma_2$ is a finite alphabet, namely the *output alphabet*
- $Q$ is a finite set of *states*
- $i \in Q$ is the *initial state*
- $F \subseteq Q$ is the set of *final states*
- $E \subseteq Q \times \Sigma_1^* \times \Sigma_2^* \times Q$ is the set of *edges*.

The notion of an extended edge of a FST is analogous to the corresponding concept for FSA.

DEFINITION 8. The *extended set of edges* $\hat{E} \subseteq Q \times \Sigma_1^* \times \Sigma_2^* \times Q$ is the smallest set such that

1. $\forall q \in Q, \langle q, \varepsilon, \varepsilon, q\rangle \in \hat{E}$
2. $\forall v_1, w_1 \in \Sigma_1^*$ and $\forall v_2, w_2 \in \Sigma_2^*$, if $\langle q_1, v_1, v_2, q_2\rangle \in \hat{E}$ and $\langle q_2, w_1, w_2, q_3\rangle \in E$, then $\langle q_1, v_1 w_1, v_2 w_2, q_3\rangle \in \hat{E}$.

A finite-state transducer $T$ defines the following relation between $\Sigma_1^*$ and $\Sigma_2^*$:

$$R(A) = \{\langle v, w\rangle \in \Sigma_1^* \times \Sigma_2^* | \exists q \in F. \langle i, v, w, q\rangle \in \hat{E}\}$$

The class of relations that is defined by some FST is called the class of *rational relations*. A simple FST that implements the rational relation $\{\langle a^n, b^n c^*\rangle | n \in \mathbb{N}\}$ is given in figure 3 for illustration.[5] The classes

[5] Following standard conventions, we mark the initial state with an arrow, and the final states are depicted as squares.

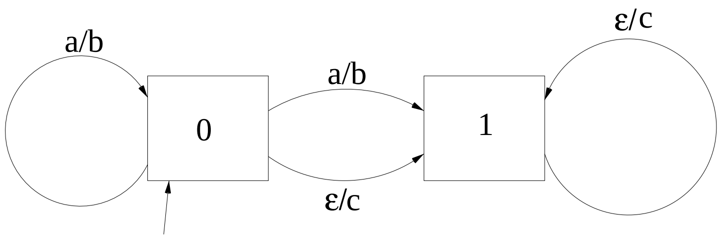


*Figure 3.* FST implementing the rational relation $\{\langle a^n, b^n c^*\rangle | n \in \mathbb{N}\}$

of regular languages and of rational relations are subject to certain *closure properties.* ($R_1 \circ R_2$ is the relation composition of $R_1$ and $R_2$, i.e. $\{\langle v, w\rangle | \exists x(\ vR_1x \wedge xR_2w)\}$. $\mathbf{I}_L$ is the identity relation on $L$, i.e. $\{\langle v, v\rangle | v \in L\}$.)

- Every finite language is regular.
- If $L_1$ and $L_2$ are regular languages, then $L_1 \cap L_2, L_1 \cup L_2, L_1 - L_2$ are also regular languages.
- If $R_1$ and $R_2$ are rational relations, then $R_1 \cup R_2$ and $R_1 \circ R_2$ are also rational relations.
- If $R$ is a rational relation, then $Dom(R)$ and $Rg(R)$ (the domain $\{x | \exists y.xRy\}$ and the range $\{y | \exists x.xRy\}$ of $R$) are regular languages.
- If $L_1$ and $L_2$ are regular languages, then $L_1 \times L_2$ and $\mathbf{I}_{L_1}$ are rational relations.

Roche and Schabes (1997) do not mention the fact that the Cartesian product $L_1 \times L_2$ of two regular languages $L_1$ and $L_2$ is a rational relation. The construction is quite simple. If $L_1$ and $L_2$ are regular languages, there are FSAs $A_1$ and $A_2$ that recognize $L_1$ and $L_2$ respectively. It is straightforward to turn these FSAs into FSTs. To this end, we treat the labels of the edges of $A_1$ as inputs and assume $\varepsilon$ as output of every transition. Likewise, we interpret the transition labels of $A_2$ as output symbols and assume everywhere the input symbol $\varepsilon$. Seen as FSTs, $A_1$ and $A_2$ define the rational relations $R_1 = L_1 \times \{\varepsilon\}$ and $R_2 = \{\varepsilon\} \times L_2$ respectively. Since rational relations are closed under composition, $L_1 \times L_2 = R_1 \circ R_2$ is also rational.

Note that the rational relations are not closed under intersection and complement.

Frank and Satta use these closure properties to show that for a significant class of OT-systems, unidirectional optimization is a rational relation provided all building blocks are rational. They restrict the class of OT-systems in two ways. First, OT constraints in general "count", a given constraint may be violated arbitrarily many times. It goes without saying that this cannot be implemented by a FST. So Frank and Satta restrict attention to binary constraints, i.e. constraints $c$ with the property $Rg(c) = \{0, 1\}$. OT-systems which are not binary but have an upper limit for the number of constraint violations are implicitly covered; a constraint $c$ that can be violated at most $n$ times can be represented by $n$ binary constraints of the form "Violate $c$ less than $i$ times" for $1 \leq i \leq n$. The ranking of these new constraints is inessential for the induced ordering relation.

Second, we may distinguish constraints that evaluate solely the output and constraints that properly evaluate an input-output pair. The former type of constraint is called *markedness constraints* in the literature (see for instance Kager (1999)), while the latter are covered under the term *faithfulness constraint*. Let us make this precise. We use the term "Output Markedness Constraint" since markedness constraint may also evaluate solely the input. Such input constraints have no effect for unidirectional OT, but they become important in the next section when we discuss bidirectionality.

DEFINITION 9. ((Output) Markedness Constraint). Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system. Constraint $c_j$ is an *output markedness constraint* iff

$$\langle i, o\rangle \in \mathbf{GEN} \wedge \langle i', o\rangle \in \mathbf{GEN} \rightarrow c_j(\langle i, o\rangle) = c_j(\langle i', o\rangle))$$

Frank and Satta restrict attention to binary output markedness constraints. Obviously, these can be represented as languages over the output alphabet. The central part of their construction is an operation called *conditional intersection* (Karttunen (1998) calls it *lenient composition*) that combines a relation with a language.

DEFINITION 10. (Conditional Intersection). Let $R$ be a relation and $L \subseteq Rg(R)$. The *conditional intersection* $R \uparrow L$ of $R$ with $L$ is defined as

$$R \uparrow L \doteq (R \circ \mathbf{I}_L) \cup (\mathbf{I}_{Dom(R)-Dom(R\ \circ\ \mathbf{I}_L)} \circ R)$$

By applying the definitions, it is easy to see that $\langle x, y\rangle \in R \uparrow L$ iff $xRy$ and either $y \in L$ or there is no $z \in L$ such that $xRz$. In other words, $\{y|\langle x, y\rangle \in R \uparrow L\}$ is the set of $y$s that are related to $x$ by $R$, and that are optimal with respect to the constraint $L$. Furthermore, it follows from the closure properties given above that $R \uparrow L$ is a rational relation provided $R$ is rational and $L$ is a regular language.

Unidirectional optimality can now be implemented in a straightforward way, namely by successively conditionally intersecting the (binary markedness) constraints of an OT-system with **GEN**.

THEOREM 3. (Frank and Satta). Let $\mathcal{O} = \langle \mathbf{GEN}, C \rangle$ with $C = \langle c_1, \ldots, c_p \rangle$ be an OT-system such $C$ solely consists of binary output markedness constraints. Then $\langle i, o \rangle$ is unidirectionally optimal iff $\langle i, o \rangle \in \mathbf{GEN} \uparrow c_1 \cdots \uparrow c_p$.

The proof of this theorem is obvious from the definitions. Crucially, it follows that unidirectional optimality is a rational relation provided **GEN** is rational and all constraints are regular languages.

## 4. Extension to Bidirectionality

In this section we will show that Frank and Satta's construction can be extended to the bidirectional case. To gain an intuitive understanding of the construction that we are going to present, let us consider how bidirectional optimality is evaluated in case of a finite **GEN**. Suppose $\mathbf{GEN} = \{1, 2, 3\} \times \{1, 2, 3\}$, and we have two constraints which both say "Be small!". One of its instances applies to the input and one to the output. Thus formally we have

- $\mathcal{O} = \langle \mathbf{GEN}, C \rangle$
- $\mathbf{GEN} = \{1, 2, 3\} \times \{1, 2, 3\}$
- $C = \langle c_1, c_2 \rangle$
- $c_1(\langle i, o \rangle) = i$
- $c_2(\langle i, o \rangle) = o$

It follows from the way constraints are evaluated that $\langle i_1, o_1 \rangle <_{\mathcal{O}} \langle i_2, o_2 \rangle$ iff $i_1 \leq i_2$, $o_1 \leq o_2$, and $\langle i_1, o_1 \rangle \neq \langle i_2, o_2 \rangle$. Now obviously $\langle 1, 1 \rangle$ is bidirectionally optimal since both its input and its output obey the constraint in an optimal way. Accordingly, $\langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 2, 1 \rangle$ and $\langle 3, 1 \rangle$ are blocked, since they all share a component with a bidirectionally optimal candidate. There are still candidates left which are neither marked as optimal nor as blocked, so we have to repeat this procedure. Among the remaining candidates, $\langle 2, 2 \rangle$ is certainly bidirectionally optimal because all of its competitors in either dimension are known to be blocked. This candidate in turn blocks $\langle 2, 3 \rangle$ and $\langle 3, 2 \rangle$. The only remaining candidate, $\langle 3, 3 \rangle$, is again bidirectionally optimal since all its competitors are blocked.[6] This example illustrates the general strategy for the finite

[6] Bidirectional optimality thus predicts iconicity: the pairing of cheap inputs with cheap outputs is optimal, but also the pairing of expensive inputs with expensive outputs. See Blutner's papers for further discussion of this point.

case: Find the cheapest input-output pairs in the whole of **GEN** and mark them as bidirectionally optimal. Next mark all candidates that share either the input component or the output component (but not both) with one of these bidirectionally optimal candidates as blocked. If there are candidates left that are neither marked as bidirectionally optimal nor as blocked, repeat the procedure until **GEN** is exhausted.

Now let us see how this construction can be extrapolated to an infinite **GEN**. Again we restrict attention to binary markedness constraints. However, for bidirectional optimization competition between different inputs may occur. Thus it makes sense to consider constraints that compare different inputs while ignoring the output.

DEFINITION 11. (Input Markedness Constraint). Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system. Constraint $c_j$ is an *input markedness constraint* iff

$$\langle i, o\rangle \in \mathbf{GEN} \wedge \langle i, o'\rangle \in \mathbf{GEN} \rightarrow c_j(\langle i, o\rangle) = c_j(\langle i, o'\rangle))$$

If we want to conditionally intersect **GEN** with a binary input markedness constraint, we need a mirror image of Frank and Satta's conditional intersection. Thus we define backward conditional intersection as

$$R \downarrow L \doteq (\mathbf{I}_L \circ R) \cup (R \circ \mathbf{I}_{Rg(R) - Rg(\mathbf{I}_L \circ R)})$$

Furthermore, for reasons that will become clear later, in bidirectional optimality it is not sufficient to consider the best outputs for a given input, but we have to look for the best input-output pairs in a global way. Thus we define bidirectional conditional intersection as:

DEFINITION 12. (Bidirectional Conditional Intersection).
Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system and $c_i$ be a binary markedness constraint.

$$R \Uparrow c_i \doteq \begin{cases} R \circ \mathbf{I}_{Rg((\{\varepsilon\} \times Rg(R)) \uparrow c_i)} \\ \text{if } c_i \text{ is an output markedness constraint} \\ \\ \mathbf{I}_{Dom((Dom(R) \times \{\varepsilon\}) \downarrow c_i)} \circ R \\ \text{else} \end{cases}$$

Let us look at this construction in detail. Suppose $c_i$ is an output markedness constraint. $\{\varepsilon\} \times Rg(R)$ is a relation that relates the empty string to any possible output of $R$. Conditionally intersecting this relation with $c_i$ leads to a relation that relates the empty string to those possible outputs of $R$ that are optimal with respect to $c_i$. So if $c_i$ is fulfilled by some output of $R$, this relation is $\{\varepsilon\} \times (Rg(R) \cap c_i)$. If no output of $R$ obeys $c_i$, the relation is just $\{\varepsilon\} \times Rg(R)$. In either way, $Rg((\{\varepsilon\} \times Rg(R)) \uparrow c_i)$ is the set of outputs of $R$ that are optimal with respect to $c_i$. Since $c_i$ only evaluates outputs, $R \Uparrow c_i$ is thus the set of $\langle i, o\rangle \in R$ that are optimal with respect to $c_i$. The same holds *ceteris paribus* if $c_i$ is an input markedness constraint.

Like Frank and Satta's operation, bidirectional conditional intersection only makes use of finite state techniques. It follows directly from the closure properties of regular languages and rational relations that $R \Uparrow c_i$ is a rational relation provided $R$ is rational and $c_i$ is a regular language.

Note that a certain input-output pair may be evaluated as sub-optimal according to this construction even if it neither shares the input component nor the output component with any better candidate. So while Frank and Satta's conditional intersection operates pointwise for each input, bidirectional conditional intersection is global.

LEMMA 3. Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system (with binary markedness constraints only), where $C = \langle c_1, \ldots, c_p\rangle$. Then

$$\langle i, o\rangle \in \mathbf{GEN} \Uparrow c_1 \cdots \Uparrow c_p$$

iff $\langle i, o\rangle \in \mathbf{GEN}$, and there are no $i', o'$ with $\langle i', o'\rangle \in \mathbf{GEN}$ and $\langle i', o'\rangle < \langle i, o\rangle$.

*Proof.* We extend the notion of an OT-system to the degenerate case that $p = 0$, i.e. there are no constraints. In this case, $<$ is the empty relation. Given this, we prove the lemma by induction over $p$, the number of constraints. For the base case $p = 0$, the proof is immediate. So let us assume that the lemma is true for all OT-systems with at most $n - 1$ constraints, and let $\mathcal{O}$ be an OT-system with $n$ constraints. Suppose $\langle i, o\rangle \in \mathbf{GEN} \Uparrow c_1 \cdots \Uparrow c_n$. It is immediate from the definition that $R \Uparrow L \subseteq R$, thus $\langle i, o\rangle \in \mathbf{GEN}$. Now suppose there is an $\langle i', o'\rangle \in \mathbf{GEN}$ with $\langle i', o'\rangle < \langle i, o\rangle$. Then there must be an $m \leq n$ such that $\langle i', o'\rangle$ obeys and $\langle i, o\rangle$ violates $c_m$. By induction hypothesis, $\langle i, o\rangle$ is optimal with respect to $c_1 \ldots c_{n-1}$. If $\langle i', o'\rangle < \langle i, o\rangle$ with respect to $c_1 \ldots c_n$, $\langle i', o'\rangle$ is also optimal with respect to $c_1 \ldots c_{n-1}$. Thus $\langle i, o\rangle, \langle i', o'\rangle \in \mathbf{GEN} \Uparrow c_1 \cdots \Uparrow c_{n-1}$. Thus by induction hypothesis, these two candidates have the same pattern of constraint violations with respect to $c_1 \cdots c_{n-1}$. Hence $m = n$.

Let us assume that $c_n$ is an output markedness constraint. According to the definition of bidirectional conditional intersection, either $o$ obeys $c_n$, or there is no $o_1 \in Rg(\mathbf{GEN} \Uparrow c_1 \cdots \Uparrow c_{n-1})$ that obeys $c_n$. Thus $o$ and $o'$ either both obey or both violate $c_i$. Hence $\langle i', o'\rangle \not< \langle i, o\rangle$, contra assumption. The same argument applied *ceteris paribus* if $c_n$ is an input markedness constraint.

⊣

For simplicity, we will use the notation $R^C$ as shorthand for $R \Uparrow c_1 \cdots \Uparrow c_p$ (where $C = c_1, \ldots, c_p$). Intuitively, this operation picks out the globally optimal set of input-output pairs from **GEN**. Note that $R^C$ is a rational relation if $R$ is rational and all constraints in $C$ are regular languages.

$R^C$ implicitly partitions $R$ into three mutually exclusive subrelations. There is $R^C$ itself—the set of input-output pairs that don't have better alternatives whatsoever. These pairs are certainly optimal. Second, there is the set of input-output pairs that share one component with some element of $R^C$. These pairs are blocked by $R^C$ (where blocking is understood in the sense of x-optimality).

Finally, there is the set of pairs that share neither component with an element of $R^C$. $R^C$ provides no information whether the elements of the third set are optimal or blocked. In analogy to the toy example discussed at the beginning of the section, we have to repeat optimization by applying the operation $(\cdot)^C$ to the third set. This procedure is repeated until the third set is empty.

This idea is formalized by the subsequent definition.

DEFINITION 13. Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system.

$$\begin{aligned} X_0 &= \emptyset \\ X_{\alpha+1} &= X_\alpha \cup (\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_\alpha)} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_\alpha)})^C \\ X_\beta &= \bigcup_{\alpha<\beta} X_\alpha \ (\beta \text{ a limit ordinal}) \\ X &= \bigcup X_\alpha \end{aligned}$$

For every ordinal $\alpha$, $X_{\alpha+1}$ adds those input-output pairs to $X_\alpha$ that are neither elements of $X_\alpha$ nor blocked by an element of $X_\alpha$, and that are minimal in this respect.

$X$ coincides with the set of x-optimal input-output pairs.

LEMMA 4. Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system. Then $\langle i, o\rangle \in X$ iff $\langle i, o\rangle$ is x-optimal.

*Proof.* The strategy of the proof is as follows.

1. We start with defining a monotonic operation $B_{(\cdot)}$ that maps each ordinal $\alpha$ to the set of pairs from **GEN** that are blocked by some elements from $X_\alpha$. It thus holds that $X_\alpha$ and $B_\alpha$ are disjoint for all ordinals $\alpha$.

2. We will show that furthermore, there is an ordinal $\xi$ such that $X_\xi \cup B_\xi = \mathbf{GEN}$.

3. In the next step we prove that for arbitrary ordinals $\alpha$, the relation $X_\alpha$ only contains x-optimal pairs, while $B_\alpha$ contains no x-optimal pairs. Hence if $X_\xi \cup B_\xi = \mathbf{GEN}$, it follows that a pair is x-optimal iff it is in $X_\xi$. Furthermore, since both $X_{(\cdot)}$ and $B_{(\cdot)}$ are monotonic operations, it follows that $X = X_\xi$, since otherwise $X_\beta$ would not be disjoint from $B_\beta$ for some $\beta > \xi$. This completes the proof.

First some notation: We write $a \simeq b$ iff $\pi_i(a) = \pi_i(b)$ for some $i \in \{1, 2\}$, and $a \sqsubset b$ iff $a \simeq b$ and $a < b$.

*First part:* We define the operation $B_{(\cdot)}$ as

$$B_\alpha \;=\; \{a \in \mathbf{GEN} | \exists b \simeq a.b \in X_\alpha\} - X_\alpha$$

We have to show that $B_{(\cdot)}$ is a monotonic operation. We prove that $\alpha \leq \beta \rightarrow B_\alpha \subseteq B_\beta$ by induction over $\beta$. The claim obviously holds for $\beta = 0$, so we next demonstrate that $B_\beta \subseteq B_{\beta+1}$ for arbitrary ordinals by proving that $B_\beta - B_{\beta+1} = \emptyset$. Consider the set

$$B_\beta - B_{\beta+1}$$

Using the definition above, this is equivalent to

$$\{a \in \mathbf{GEN} | \exists b \simeq a.b \in X_\beta\} - X_\beta - (\{a \in \mathbf{GEN} | \exists b \simeq a.b \in X_{\beta+1}\} - X_{\beta+1})$$

By some elementary set theoretic reasoning, this is the same as

$$\{a \in \mathbf{GEN} - X_\beta | \exists b \simeq a.b \in X_\beta \wedge \neg(\exists b \simeq a.b \in X_{\beta+1})\} \cup$$
$$(\{a \in \mathbf{GEN} - X_\beta | \exists b \simeq a.b \in X_\beta\} \cap X_{\beta+1})$$

Since $X_\beta \subseteq X_{\beta+1}$, it holds that

$$\{a \in \mathbf{GEN} - X_\beta | \exists b \simeq a.b \in X_\beta \wedge \neg(\exists b \simeq a.b \in X_{\beta+1})\} = \emptyset$$

Hence

$$B_\beta - B_{\beta+1} = \{a \in \mathbf{GEN} - X_\beta | \exists b \simeq a.b \in X_\beta\} \cap X_{\beta+1}$$

It follows from the definition of $X_{(\cdot)}$ that $a \in X_{\beta+1} \wedge \exists b \simeq a.b \in X_\beta$ entails that $a \in X_\beta$. Therefore $B_\beta - B_{\beta+1} = \emptyset$, and thus $B_\beta \subseteq B_{\beta+1}$. So we have established the induction step for successor ordinals.

Let us consider the case that $\beta$ is a limit ordinal. Then it holds that

$$B_\beta = \{a | \exists b \simeq a.b \in \bigcup_{\gamma<\beta} X_\gamma\} = \bigcup_{\gamma<\beta} \{a | \exists b \simeq a.b \in X_\gamma\} = \bigcup_{\gamma<\beta} B_\gamma$$

So if $\alpha \leq \beta$ and $\forall \gamma < \beta(\alpha \leq \gamma \rightarrow B_\alpha \subseteq B_\gamma)$, then obviously $B_\alpha \subseteq B_\beta$. This completes the proof that $B_{(\cdot)}$ is monotonic. Furthermore, it follows immediately from the definition of $B_{(\cdot)}$ that $X_\alpha \cap B_\alpha = \emptyset$ for arbitrary ordinals $\alpha$.

*Second part:* First observe that the operation $X_{(\cdot)}$ has an upper limit, i.e. there is an ordinal $\xi$ such that $X = X_\xi$. Otherwise we could define an operation from **GEN** onto the class of ordinals. This is impossible since **GEN** is a set. It follows directly from the definition of $X_{(\cdot)}$ that

$$\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_\xi)} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_\xi)} = \emptyset$$

because otherwise $(\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_\xi)} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_\xi)})^C$ would not be empty and thus $X_\xi \subset X_{\xi+1}$. Thus $\forall a \in \mathbf{GEN} \exists b \simeq a.b \in X_\xi$. Therefore

$$X_\xi \cup B_\xi = \mathbf{GEN}$$

*Third part:* We introduce another auxiliary notation. We say that $a \equiv b$ iff $\forall c(c < a \leftrightarrow c < b)$. It follows directly from the definition of $<$ in terms of OT systems that $\forall a, b(a < b \vee a \equiv b \vee a > b)$.

Let $OPT$ be the set of x-optimal elements of **GEN**. We prove that $X_\alpha \subseteq OPT$ and $B_\alpha \cap OPT = \emptyset$ by induction over $\alpha$. The base case for $\alpha = 0$ is obvious. So suppose the claim holds for $\alpha$, and suppose $a \in X_{\alpha+1} - X\alpha$. By the definition of $X_{(\cdot)}$, this means that $a \in (\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_\alpha)} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_\alpha)})^C$. Due to lemma 3 , this entails that $\neg\exists b \simeq a.b \in X_\alpha$ and $\forall c < a \exists d \simeq c.d \in X_\alpha$. Now suppose that $e \sqsubset a$. Then we can infer that $\exists f \simeq e.f \in X_\alpha$. Hence $e \in X_\alpha \cup B_\alpha$. Suppose $e \in X_\alpha$. Since $e \sqsubset a$ and $a \notin X_\alpha$, this would entail that $a \in B_\alpha$, but this is impossible because $B_\alpha \subseteq B_{\alpha+1} \cap X_{\alpha+1} = \emptyset$. Hence $e \in B_\alpha$. By induction hypothesis, $e \notin OPT$. We have thus established that $\forall e \sqsubset a.e \notin OPT$. Hence $a \in OPT$.

Now suppose $a \in B_{\alpha+1} - B_\alpha$. Suppose furthermore that $a \in OPT$. Since $a \in B_{\alpha+1}$, there is a $b \simeq a$ such that $b \in X_{\alpha+1}$. Furthermore $b \notin X_\alpha$, because otherwise $a \in B_\alpha$, contra assumption. By the reasoning from the previous paragraph, it follows from the induction hypothesis that $b \in OPT$. As argued above $a < b \vee a \equiv b \vee a > b$. Since $a \simeq b$ by assumption, $a < b$ entails $a \sqsubset b$, and therefore $b \in OPT$ entails $a \notin OPT$, which is in contradiction to the assumptions. Hence $a \not< b$, and by a similar argument we conclude that $b \not< a$. Therefore $a \equiv b$. Now suppose $c < b$. $b \in X_{\alpha+1} - X_\alpha$, hence $b \in (\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_{\alpha+1})} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_{\alpha+1})})^C$. From this together with lemma 3 we infer that $c \in X_\alpha \cup B_\alpha$. Now suppose $d < a$. Since $a \equiv b$, this entails that $d < b$, and therefore $d \in X_\alpha \cup B_\alpha$. It follows from the definitions that

$$\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_\alpha)} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_\alpha)} = \mathbf{GEN} - (X_\alpha \cup B_\alpha)$$

Since we just established that $\forall d < a.d \in X_\alpha \cup B_\alpha$, and furthermore $a \notin X_\alpha \cup B_\alpha$ by assumption, it follows from lemma 3 that $a \in$

$(\mathbf{I}_{Dom(\mathbf{GEN})-Dom(X_\alpha)} \circ \mathbf{GEN} \circ \mathbf{I}_{Rg(\mathbf{GEN})-Rg(X_\alpha)})^C$ and thus $a \in X_{\alpha+1}$, and this is in contradiction to the assumption that $a \in B_{\alpha+1}$. So the assumption that $a \in OPT$ lead to a contradiction and we thus proved that $a \in B_{\alpha+1} - B_\alpha$ entails that $a \notin OPT$. This completes the induction step for successor ordinals.

Finally, suppose that $\alpha$ is a limit ordinal. Since $X_\beta \subseteq OPT$ for all $b < \alpha$, it follows directly that

$$X_\alpha = \bigcup_{\beta<\alpha} X_\beta \subseteq OPT$$

Likewise, we proved above that $B_\alpha = \bigcup_{\beta<\alpha} B_\beta$ if $\alpha$ is a limit ordinal. So if $B_\beta \cap OPT = \emptyset$ for all $\beta < \alpha$, it also holds that

$$B_\alpha \cap OPT = \emptyset$$

Hence

$$X = X_\xi = OPT$$

⊣

So the operation $X_\alpha$ provides a cumulative definition of the notion of x-optimality. Most importantly for the present purposes, the step from $X_\alpha$ to $X_{\alpha+1}$ makes use only of finite state techniques. In other words, if $X_\alpha$ and **GEN** are rational relations, and all constraints in $C$ are binary markedness constraints that can be represented by regular languages, $X_{\alpha+1}$ is also a rational relation. This follows directly from the closure properties of rational relations and regular languages. $X_0 = \emptyset$ by definition, and since $\emptyset = \emptyset \times \emptyset$ and $\emptyset$ is a finite language, it is a regular language and hence also a rational relation. So it follows by complete induction that $X_n$ is a rational relation for any finite $n$ provided **GEN** is rational and all constraints involved are regular languages. So to show that $X$ is also rational under these conditions, it suffices to demonstrate that $X = X_n$ for some finite $n$.

LEMMA 5. Let $\mathcal{O} = \langle \mathbf{GEN}, C\rangle$ be an OT-system with $C = c_1, \ldots, c_p$, where all $c_i$ are binary markedness constraints. Then $X = X_{2^p-1}$.

*Proof.* We define the *degree* of some $a \in \mathbf{GEN}$ as

$$d(a) \;=\; \bigcup\{d(b)|b < a\} + 1$$

Again it follows from the recursion theorem that this is a valid definition. Intuitively, the ranked constraints of an OT-system partition **GEN** into linearly ranked equivalence classes (where two candidates are equivalent if they have the same patterns of constraint violations), and $d(a)$ measures the rank of the equivalence class of $a$. Put more formally, $a \equiv b$ directly entails $d(a) = d(b)$, and if $a$ and $b$ have the same pattern of constraint violations, $a \equiv b$. Thus in this case $d(a) = d(b)$. If $C$ consists of $p$ constraints, there are finitely many, namely at most $2^p$ possible patterns of constraint violations. Thus $d(a) \leq 2^p$ for arbitrary $a$.

The proof strategy is as follows: (We make once again use of the auxiliary operation $B_{(\cdot)}$ that was introduced in the proof of lemma 4. )

1. We prove that it holds for all natural numbers $n > 0$ and $a \in \mathbf{GEN}$:

$$\begin{aligned} a \in X_{n+1} - X_n &\rightarrow \exists b < a.b \in (X_n \cup B_n) - (X_{n-1} \cup B_{n-1}) \\ a \in B_{n+1} - B_n &\rightarrow \exists b < a.b \in X_{n+1} - X_n \end{aligned}$$

2. Based on this, we demonstrate that for all $n \geq 0$ and $a \in \mathbf{GEN}$:

$$\begin{aligned} a \in X_{n+1} - X_n &\rightarrow d(a) \geq 2n + 1 \\ a \in B_{n+1} - B_n &\rightarrow d(a) \geq 2n + 2 \end{aligned}$$

3. From this we conclude that the smallest number $n$ with $X = X_n$ has the property that $2n \leq 2^p$, hence $X = X_{2^{p-1}}$.

*First part:* Suppose that $n > 0$ and that $a \in X_{n+1} - X_n$. It follows from the definitions together with lemma 3 that for all $b < a$, $b \in X_n \cup B_n$. Now suppose there were an $m < n$ such that for all $b < a$, $b \in X_m \cup B_m$. Then, by the definitions and lemma 3 , $a \in X_{m+1}$, which is in contradiction with the assumptions. Thus $\forall m < n \exists b < a.b \notin X_m \cup B_n$. In particular it thus holds that there is a $b < a$ such that $b \notin X_{n-1} \cup B_{n-1}$. But since $b < a$, it must hold that $b \in X_n \cup B_n$.

Now suppose $a \in B_{n+1} - B_n$. By the definition of $B_{(\cdot)}$, it follows that there is a $b \simeq a$ such that $b \in X_{n+1} - X_n$. We established in the proof of lemma 4 that $a < b \vee a \equiv b \vee a > b$. Suppose $a < b$. Since $b \in X_{n+1}$, it follows from lemma 3 and the definitions that $a \in X_n \cup B_n$, contra assumptions ($a \in X_n$ is directly excluded by the assumptions, and $a \in X_{n+1}$ entails that $a \notin B_n$ because $B_n \cup X_n = \emptyset$ and $X_n \subseteq B_n$, cf. the proof of lemma 4 ). So let us assume that $a \equiv b$. Since $b \in X_n$, it holds for all $c < b$ that $c \in X_{n-1} \cup B_{n-1}$. By the definition of $\equiv$, it thus also holds that for all $c < a$, $c \in X_{n-1} \cup B_{n-1}$. Hence $a \in X_n$, contra assumptions, and thus $b < a$.

*Second part:* We prove these claims by complete induction. Suppose $n = 0$. From the definition of $X_{(\cdot)}$ we infer that $a \in X_1 - X_0$ iff there are no $b < a$. Thus if $a \in X_1 - X_0$, then $d(a) = 1$. If $a \in B_1 - B_0$, there must therefore be a $b$ such that $b < a$. Hence $d(a) \geq 2$.

Now suppose $n > 0$ and the claims hold for $n - 1$. Let us assume that $a \in X_{n+1} - X_n$. By the definitions of $X_{(\cdot)}$ and $B_{(\cdot)}$, $a \notin X_n \cup B_n$. By applying the first result from the first part, we conclude that there is a $b < a$ such that $b \in (X_n \cup B_n) - (X_{n-1} \cup B_{n-1})$. By induction hypothesis, $d(b) \geq 2n$. Since $a > b$, $d(a) \geq 2n + 1$.

Suppose $a \in B_{n+1} - B_n$. According to the second result from the first part, there is a $b < a$ with $b \in X_{n+1} - X_n$. By induction hypothesis, $d(b) \geq 2n + 1$. Hence $d(a) \geq 2n + 2$.

*Third part:* Suppose $X = X_n$. Then for all $a \in \mathbf{GEN}$, there is an $m < n$ such that $a \in (X_{m+1} \cup B_{m+1}) - (X_m \cup B_m)$. Due to the results from the second part, it thus holds for all $a \in \mathbf{GEN}$ that there is an $m < n$ with $d(a) \geq 2m + 2$. Hence $d(a) \geq 2n$. Since $d(a) \leq 2^p$ for all $a$, we conclude that $2n \leq 2^p$, hence $n \leq 2^{p-1}$. Furthermore, if $X = X_n$, then $X = X_m$ for all $m \geq n$. Therefore $X = X_{2^{p-1}}$.

⊣

This leads us directly to the main result of this section.

THEOREM 4. Let $\mathcal{O} = \langle \mathbf{GEN}, C \rangle$ be an OT-system with $C = \langle c_1, \ldots, c_p \rangle$, where all $c_i$ are binary markedness constraints. Furthermore, let **GEN** be a rational relation and let all $c_i$ be regular languages. Then the set of x-optimal elements of **GEN** is a rational relation.

*Proof.* Immediately from the lemmas 4, 5 , and the closure conditions of regular languages and rational relations. ⊣

Note that the proof is constructive. So if the components of an OT-system with the described properties are given as finite state automata, the proof provides an algorithm for constructing a finite state transducer that implements bidirectional optimization of this OT-system.

## 5. Conclusion and open ends

In this paper, we investigated some meta-theoretic properties of the model of bidirectional Optimality Theory that was developed in Blutner (2000). We obtained three main results:

1. We developed a conceptually simpler definition of bidirectionality (definition 3 on page 10 ) and proved its equivalence with Blutner's definition under very general conditions.

| **Relations** | | **Languages** |
|---|---|---|
| $\cup, \circ$ | $\overset{Dom,\ Rg}{\longrightarrow}$ <br> $\overset{\times, \mathbf{I}}{\longleftarrow}$ | $\cap, \cup, -$ |

*Figure 4.* Closure conditions needed for x-optimality

2. For a substantial class of OT-systems (those where only non-counting markedness constraints are involved), we gave a cumulative definition of bidirectional optimality that is more constructive than the previous definitions.

3. Inspired by Frank and Satta (1998), we showed that for the mentioned class of OT-systems, the relation of bidirectional optimality between input and output can be modeled by a finite state transducer provided the generator and the constraints can be modeled by such means.

While modeling of optimization with finite state techniques might be of practical importance in computational phonology, there are no obvious applications of such methods in syntax, semantics and pragmatics. Since bidirectional OT is used mainly in these areas of linguistics, the investigations that were described in the last section are of a very theoretical interest only.

The techniques that were developed there can be extrapolated to more interesting classes of languages and relations though. In the proof of theorem 4 , we ignored the specific properties of regular languages and rational relations but we only used their closure properties. As an immediate consequence, bidirectional optimization stays within reach of any class of languages/relations that is closed under the same operations—provided the OT-system in question only has binary markedness constraints. These closure conditions are summarized in figure 4.

So the linguistic relevance of our results largely depends on whether there are automata theoretic complexity classes which are (a) inclusive enough to model natural language syntax, semantics, and the map between the two, and (b) have the mentioned closure properties. At the present point, I have to leave this issue open, but some initial considerations are possible. It has been pointed out at several places—see for instance Morawietz and Cornell (1997) and Wartena (2000)—that tree automata (in the sense of Gécseg and Steinby (1997)) are a promising formal tool to model syntactic operations and constraints, and I dare the conjeture that standard compositional translations from syntactic to semantic representations can be implemented by these means as well. It is noteworthy in the present context that many properties of the class of regular string languages and rational relations carry over to the regular tree languages, paired with the class of linear frontier-to-root tree transductions ($\mathcal{LF}$-transductions). In fact, these classes are closed under all operations from figure 4 (cf. Gécseg and Steinby (1997)), with the single exception that the Cartesian product of two regular tree languages is not necessarily an $\mathcal{LF}$-transduction. The Cartesian product is not needed in Frank and Satta's construction, and therefore their result directly carries over to this class of tree languages/relations, as Wartena (2000) points out. The Cartesian product is needed though in the definition of bidirectional conditional intersection, and it remains to be seen whether it can be redefined in a way that keeps it within the realm of $\mathcal{LF}$-transductions.

It is known, however, that the class of regular tree languages corresponds to the context free string languages and is thus somewhat too weak to model natural language syntax (cf. Shieber (1985)). So it is open whether there are classes of tree languages and relations that correspond to the mildly context-sensitive (cf. Joshi (1985)) string languages and are still well-behaved with regard to their closure properties.

Another open issue is the implementation of faithfulness constraints. In the general case, an extrapolation of conditional intersection to these constraints would require closure of the relevant class of relations under intersection. To my knowledge, there is no interesting class of relations with this property. Thus it seems to be more promising to search for restrictions on faithfulness constraints that make them computationally tractable.

Last but not least, our constructive redefinition of bidirectional optimality rests on the assumptions that all constraints are binary. To work with counting constraints, we will need a more elaborate definition of $R \Uparrow S$. As an additional complication, there is no guarantee anymore that $X = X_n$ for some finite $n$ in the general case. So future research has to show what a constructive reformulation of bidirectional optimization with counting constraints might look like.

## Acknowledgements

Parts of the material laid down here were presented in talks given at the University of Amsterdam, the Heinrich Heine University at Düsseldorf, and the University of California at Los Angeles. I would like to thank the audiences of these talks for valuable feedback. Furthermore I am indebted to Anton Benz, Dejuan Wang, and an anonymous reviewer of the *Journal of Logic, Language and Information* for helpful comments on a previous version of this paper.

## References

Blutner, R.: 1998, 'Lexical Pragmatics'. *Journal of Semantics* **15**(2), 115–162.

Blutner, R.: 2000, 'Some Aspects of Optimality in Natural Language Interpretation'. to appear in *Journal of Semantics*.

Frank, R. and G. Satta: 1998, 'Optimality Theory and the Generative Complexity of Constraint Violability'. *Computational Linguistics* **24**(1), 307–315.

Gécseg, F. and M. Steinby: 1997, 'Tree Languages'. In: G. Rozenberg and A. Salomaa (eds.): *Handbook of Formal Languages*, Vol. III. Berlin/Heidelberg: Springer, pp. 1–68.

Heim, I.: 1990, 'On the Projection Problem for Presuppositions'. In: S. Davis (ed.): *Pragmatics*. Oxford: Oxford University Press, pp. 397–405.

Horn, L.: 1984, 'Towards a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicatures'. In: D. Schiffrin (ed.): *Meaning, Form, and Use in Context*. Washington: Georgetown University Press, pp. 11–42.

Joshi, A.: 1985, 'How much context-sensitivity is necessary for characterizing structural descriptions — tree adjoining grammars'. In: D. Dowty, L. Karttunen, and A. Zwicky (eds.): *Natural Language Processing. Theoretical, Computational and Psychological Perspectives*. Cambridge (UK): Cambridge University Press.

Kager, R.: 1999, *Optimality Theory*. Cambridge: Cambridge University Press.

Karttunen, L.: 1998, 'The Proper Treatment of Optimality in Computational Phonology'. manuscript. Xerox Research Centre Europe.

Levinson, S. C.: 1987, 'Pragmatics and the Grammar of Anaphora'. *Journal of Linguistics* **23**, 379–434.

Morawietz, F. and T. Cornell: 1997, 'Representing constraints with automata'. In: *35th Annual Meeting of the ACL*. Madrid, Spain.

Prince, A. and P. Smolensky: 1993, 'Optimality Theory: Constraint Interaction in Generative Grammar'. Technical Report TR-2, Rutgers University Cognitive Science Center, New Brunswick, NJ.

Roche, E. and Y. Schabes: 1997, 'Introduction'. In: E. Roche and Y. Schabes (eds.): *Finite-State Language Processing*. Cambridge (Mass.): MIT Press, Chapt. 1, pp. 1–65.

Shieber, S.: 1985, 'Evidence against the context-freeness of natural language'. *Linguistics and Philosphy* **8**, 333–343.

van der Sandt, R.: 1992, 'Presupposition Projection as Anaphora Resolution'. *Journal of Semantics* **9**, 333–377.

Wartena, C.: 2000, 'A Note on the Complexity of Optimality Systems'. In: R. Blutner and G. Jäger (eds.): *Studies in Optimality Theory*. Universität Potsdam, pp. 64–72. also available at Rutgers Optimality Archive as ROA 385-03100.