

Input-Output Mismatches in OT*

David I. Beaver and Hanjung Lee

1 Introduction

Bidirectional Optimality Theory allows us to see a wide range of problems which would previously have been considered unrelated from a new perspective, the perspective of asymmetric relationships between input and output. For interpretation, the input is a form and the output a meaning, and for production the input is a meaning and the output is a form. A mismatch is any case where there is no isomorphism between the space of meanings and the space of forms, say because one form has no meaning, or multiple meanings, or because a meaning is inexpressible, or may be expressed in multiple ways.

Is there such a thing as a perfect language, one that would lack any mismatch of this sort? Certainly there are subsystems of natural and formal languages that, if taken in isolation, would be perfectly symmetric. For example, the Arabic notation for integers (assuming that initial zeroes are ill-formed) stands in a one to one relationship with the abstract semantic space of integers. But even formal languages are commonly not perfect in this very strong sense. For example, in first order logic there may be multiple constants referring to the same individual, and more generally there are an infinite number of ways of expressing any proposition that can be expressed at all. There may also be objects in the model for which there is no corresponding constant, or facts that are true in a given model or frame and yet inexpressible in first order logic. As far as form-meaning symmetry goes, the only way that first order logic scores qualitatively over natural language is that the former is (when properly notated, and interpreted with respect to a specific model) unambiguous: for any form there is exactly one meaning.

Along with ambiguity, we will be considering optionality, ineffability, uninterpretability, blocking and freezing. All of these involve a mismatch between form and meaning, and we will study how various versions of OT handle these mismatches.

Initially, we will be considering simpler, relatively standard OT architectures. The first two of these are unidirectional. What we will term *naive OT production* is the approach seen in most OT syntax papers, and is close to the model that is used in OT phonology. To recap what we assume is already familiar to most readers of this article, naive OT production starts with some representation of meaning as input, and a set of candidate outputs provided by a function referred to as GEN. A set of linearly ranked constraints is then used to select between candidate surface forms. The second unidirectional approach, not surprisingly,

*To appear: Reinhard Blutner and Henk Zeevat (eds.), *Optimality Theory and Pragmatics*, Palgrave/Macmillan.

works the other way: we will term it *naive OT comprehension*, although Hendriks & de Hoop (2001) term it *OT semantics*. The input is a surface form, GEN offers a set of candidate meanings, and the linearly ranked constraint set is used to find the best meaning for the given form.

In this paper we are not concerned with processing issues, computational complexity or the psychological plausibility of the OT tableau method. Rather, we take an abstract view of the languages that various OT models generate. As a result, and despite the danger of terminological confusion, naive OT production can be considered a theory of both comprehension and production. The same goes for naive OT comprehension. The reason is that both unidirectional accounts ultimately capture a relation between meaning and form, or, equivalently, a set of meaning-form pairs. Thus, naive OT production characterizes a language as the set of pairs of meanings and forms such that for the given meaning, the form is optimal. Likewise, naive OT comprehension characterizes a language as the set of pairs of meanings and forms such that for the given form, the meaning is optimal.

Some OT architectures provide grammars that cannot be reduced to a set of meaning-form pairs. One of these, which we will term *naive back-and-forth OT*, consists of an obvious combination of naive OT production and comprehension: the first is used for production only, and the second for comprehension only, an architecture discussed by Hendriks & de Hoop (2001). Note that even if the constraints used in each direction are the same, this model may not assign a consistent relation between meanings and forms. In particular for some choices of constraints, if you take a meaning, apply naive OT production to get a form, and then apply naive OT comprehension, you may not get back to the original meaning.

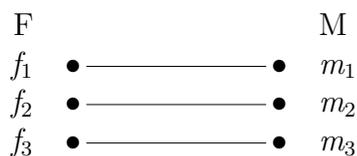
In addition to these three *naive* models, we will also consider four more sophisticated variants, sophisticated in the sense that they have been specifically designed to target some of the mismatch phenomena we will be discussing. The four other models to be studied are the *strong bidirectional* OT and *weak bidirectional* OT of Blutner (2001), and the *asymmetric* OT models of Wilson (2001) and Zeevat (2001). We will introduce these models individually later in the paper.

2 Patterns of Mismatch

In this section we will consider various phenomena involving mismatches between form and meaning, and discuss the significance of these phenomena for naive OT architectures.

Perfect language

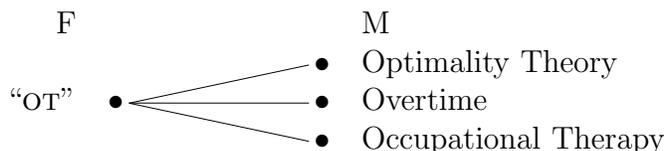
Before considering the ‘imperfections’ of natural languages, let us briefly gaze upon perfection. A perfect language would be one in which there was a one-to-one correspondence between forms and meanings:



As noted, even formal languages usually fail to achieve this level of perfection.

Ambiguity

This is the case of multiple meanings corresponding to a single form.¹ An example is the multiple interpretations of the abbreviated form “OT”:²



As regards unidirectional OT models, ambiguity constitutes a *prima facie* problem for naive OT comprehension, but not for naive OT production.

In principle, a given constraint set may produce multiple outputs for a given input. Thus there is potential for modeling ambiguity in OT comprehension. However, in practice the multiple outputs of a linearly ranked constraint set do not provide a good tool for modeling natural language ambiguity. The problem can be seen as follows: although the constraints are merely preferences, there is no way to distinguish in the output set between winners that result from strong preferences (i.e. highly ranked constraints) and winners that result from weak preferences (low ranked constraints). As a result, interpretations which one might expect to be available, if mildly dis-preferred, end up being ruled out altogether.

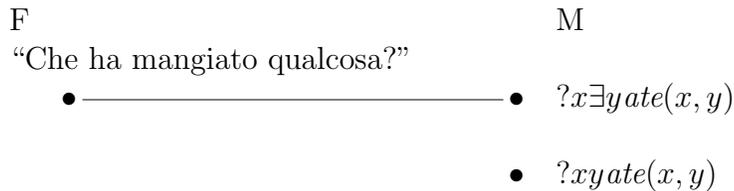
Standard examples are found in phonology. For instance, consider the neutralization between “d” and “t” in standard Dutch and English. In Dutch, “rat” (“*rat*”) and “rad” (“*wheel*”) may be pronounced identically, as discussed by Boersma (1998) and Hale & Reiss (1998) and also by Zeevat (2001), and the same goes for “wader” and “waiter” in many US varieties of English. Suppose we have the spoken Dutch input /rat/. By assumption, there is a faithfulness constraint preferring interpretation via the underlying phonological form [rat] to interpretation via underlying [rad]. If we assume linear ranking of constraints, then this faithfulness constraint is either dominated by a constraint preferring the reverse interpretation, or it is not dominated by such a constraint. Either way, /rat/ comes out unambiguous. Similarly, for US English phonetic-phonological faithfulness would lead us to expect unambiguous interpretation of /wɛɪrə/ as something which wades. But in fact both this and the alternative interpretation, as someone who waits on tables, are available. For other examples of why ambiguity is problematic for unidirectional OT, the reader is referred to Anttila & Fong (2001) and Asudeh (2001).

For naive production, ambiguity presents no obvious problem. While unidirectional OT tends to mitigate against multiple outputs for a given input, it actually favors multiple inputs producing the same output. The /rat/ example could be derived if some constraint favoring devoicing in the given phonological environment outranked the constraint enforcing voicing faithfulness. In that case, both /rat/ and /rad/ would be realized as [rat].

What does naive back-and-forth OT predict? As regards production, the ambiguity is correctly predicted, but comprehension examples like those above are problematic: no ambiguity is predicted.

- (1) a. Who ate what?
 b. *Che ha mangiato che cosa?
 Who has eaten which thing
 ‘Who ate what?’
 c. Che ha mangiato qualcosa?
 Who has eaten something
 ‘Who ate something?’
 ‘*Who ate what?’
 d. $?xyate(x, y)$
 e. $?x\exists yate(x, y)$

In diagrammatic form, the mismatch appears as an unconnected node in the space of meanings:



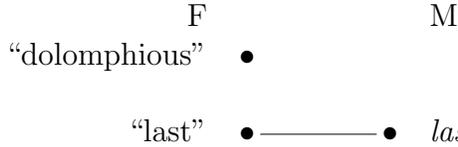
Ineffability presents a problem for naive production OT and naive back-and-forth OT. By assumption, the nature of the input (meaning) should not vary cross-linguistically, so the range of licit inputs is the same for English as for Italian. And in unidirectional OT any input produces some output, so there should be no such thing as ineffability. This is not a problem that could be wriggled out of using clever choices of constraints or a special approach to ranking. No, if naive production OT is to be taken seriously as a model, then the very existence of ineffability would have to be denied. We would have to claim that every input has an output, and perhaps broaden GEN to include multiple sentence outputs combined with appropriate gesture amongst the candidates. This would model an Italian expressing the meaning of a multiple wh-question via a complex discourse and, to use a common stereotype, plenty of hand-waving. We will not pursue this line of thought further here, but assume, in agreement with e.g. Fanslow & Féry (to appear) and Zeevat (2001), that ineffability does occur, and that our model of grammar must account for it.

For naive comprehension OT, ineffability is no problem at all. While every form corresponds to some meaning in this model, there is no reason at all why all meanings should correspond to some form.

Uninterpretability

The inverse of ineffability is uninterpretability, a form with no corresponding meaning.⁵ Thus Chomsky maintains that “colorless green ideas sleep furiously” is grammatically well formed, but lacks any semantic interpretation. Lear’s “run-cible” lacked conventional meaning when he applied it to “spoon”, and still lacks conventional interpretation in its application to “cat”, unless it is a cat that is

curved like a spoon and has three prongs, one with a sharp edge. Lear’s “dolomphious”, an adjective of ducks, still lacks conventionalized meaning. We have the following type of picture:⁶

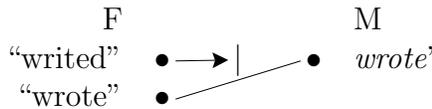


By obvious analogy with the case of ineffability, the existence of uninterpretable strings is problematic for naive comprehension OT and for naive back-and-forth OT, since they will provide an interpretation for any string given as an input. Uninterpretability is unproblematic for naive production OT.

Blocking

Blocking is a process which prevents or removes asymmetries. The most common example cited is that where a given meaning could potentially be realized either by an idiosyncratic irregular form, or by a regular productive morphological process applied to a root. The existence of an irregular form may then be said to block the regular form :

- (2) a. wrote, *writed
- b. sheep [+pl], *sheeps



The existence of a lexical form produced by semi-productive morphology may also block a phrasal form. Poser (1992) and Bresnan (2001) consider English comparative and superlative adjectival inflections: the existence of “cheaper” can be said to block “more cheap” in (3), whereas the absence of “expensiver” means that “more expensive” is available. Note that from a purely logical point of view, we could analyze “more expensive” as blocking “expensiver”, but it is standard to analyze simpler forms (e.g. a single lexeme) as blocking more complex ones rather than the other way around.

- (3) a. cheaper/cheapest, ?more/?most cheap
- b. *expensiver/*expensivest, more/most expensive

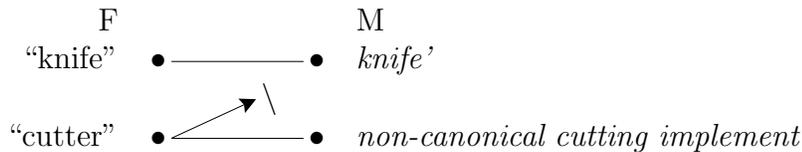
From our birds-eye perspective, we would equally term as *blocking* a case where the existence of a special meaning prevents an otherwise logically possible interpretation. Idiomatic meaning may be of this sort: “Mary kicked the bucket” could mean just that, but is invariably interpreted less fortunately. We can also understand cases involving alternative binding possibilities for pro-forms in terms of blocking of meaning (Levinson, 2000; Huang, 2000, c.f.). For example, in the Marathi case in (4) a preference for more local anaphora resolution prevents resolution outside of the clause:

- (4) Tom_i mhanat hota [ki Sue_j ni *swataahlaa*_{*i/j} maarle]. [Marathi]
 Tom said that Sue ERG ANAPHOR-ACC hit
 ‘Tom said that Sue hit herself/*him.’ (Dalrymple 1993: 19–20)

Note that none of the naive OT models provide any account of blocking, or of the variant *partial blocking* to which we now turn.

Partial blocking

Blocking can leave a form unemployed, but the unemployed form may soon find a new job, generally expressing something closely related to but subtly different from the canonical interpretation that one might have expected. This is partial blocking: an asymmetry is eliminated, but removal of a link creates a new form-meaning pair. An example from Kiparsky (1983) is the interpretation of “cutter”, a nominalization involving application of a regular and productive rule (“-er” addition). The observation is that when someone refers to “a cutter” they could not ordinarily be referring to an object for which a standard idiosyncratic expression exists, like “scissors” or “a bread knife”. So “a cutter” is interpreted as a non-canonical instrument used for cutting.



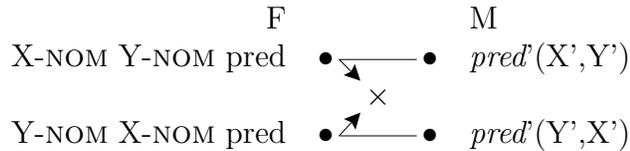
Similarly, it has often been argued that the existence of a lexical item “kill” blocks “cause to die” from having its canonical meaning, i.e. the meaning that would be derived compositionally. “Cause to die” comes to denote a non-canonical killing, for instance one where the chain of causation is unusually long or unforeseeable (McCawley, 1978, c.f.).

There are also cases where a form-meaning pair is blocked because the form has a different interpretation, and so the meaning comes to be expressed in another way. For example, “computer”, “calculator” and “reckoner” are all understood to refer to non-humans, but originally referred to humans who computed, calculated or reckoned. When we wish to refer to a human who performs these tasks, or one who performs them particularly well, we now use terms like “human calculator”, which once would have been tautological.

Freezing

Freezing is a phenomenon which can be seen in terms of a combination of ambiguity and optionality: it may constrain optionality to prevent ambiguity. Above, we mentioned word order freedom for the arguments of canonical Korean transitives. The caveat *canonical* is crucial, since the optionality vanishes for certain classes of verbs, notably a group of psychological predicates. For these predicates the subject and the object have identical case marking, in fact nominative case. This identity of case marking has the potential to create ambiguity, since one cannot tell from the morphological form alone which is the subject and which is the object. For verbs in this class, but for no others, word order is the primary

means used to represent argument structure, with SOV order fixed in most contexts. In this case, if we may speak teleologically, it appears that word order has been frozen in order to prevent ambiguity of argument structure. Graphically, we may represent the situation, in which multiple input-output mismatches are simultaneously blocked, as follows:



As was the case for blocking, freezing phenomena are not modeled by any of the naive OT strategies. However, we will now turn to a more detailed consideration of a class of bidirectional OT models which were originally introduced precisely because they suggested a line of attack for such phenomena.

3 Strong Bidirectional Optimization

Besides the phenomena of form-meaning mismatches we discuss here, arguments for bidirectional optimization have come from various sources. These include the production/comprehension asymmetry in child grammar (Smolensky 1996), decidability in computational processing (Kuhn 2001) and learning algorithms (Jäger, this volume). Given that production-based and interpretation-based optimization are both well motivated, a question immediately arises as to how the two directions of optimization can be combined into a coherent theory of language structure and interpretation.⁷ One option is to combine them conjunctively, producing a model which Blutner (2001) calls the *strong bidirectional* OT model (this will be compared with a *weak* version in section 4). The idea is that in order to be grammatical, a form-meaning pair $\langle f, m \rangle$ has to be optimal in both directions of optimization. That is, a form-meaning pair is strong OT optimal iff the form produces the meaning in Interpretation OT and the meaning produces the form in Production OT. So we arrive at the following definition of bidirectional optimality (The connective “ \succ ” is read as “more harmonic than” or “more economical than”):

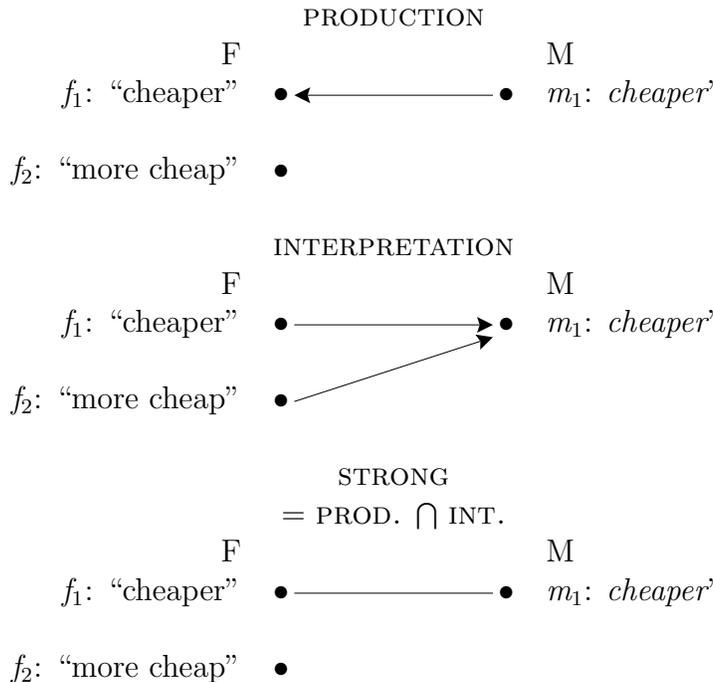
- (5) $\langle f, m \rangle$ is strong OT optimal iff
- a. $\langle f, m \rangle \in \text{GEN}$,
 - b. there is no $\langle f', m \rangle \in \text{GEN}$ such that $\langle f', m \rangle \succ \langle f, m \rangle$, and
 - c. there is no $\langle f, m' \rangle \in \text{GEN}$ such that $\langle f, m' \rangle \succ \langle f, m \rangle$.

For a more detailed discussion of the formal properties of this notion of optimality, the reader is referred to Blutner (2001) and Jäger (2002).

Strong OT removes form-meaning pairs that are only optimal under one direction. In this way, it produces strictly fewer form-meaning pairs than either naive production or interpretation OT would with the same constraint ranking, and consequently it can model both ineffability and uninterpretability. Ineffability results if the optimal realization for m is the surface string f , but in comprehension-based

optimization for f we get a different meaning m' ($m \neq m'$). So, m' blocks m , making m ineffable. Uninterpretability occurs when the interpretation-based winner m for the form f has a different form f' in production-based optimization. See section 4 for a more detailed discussion and illustration.

Strong OT offers a treatment of synonymy blocking, a phenomenon which remains unaccounted for in (unidirectional) interpretation OT. Suppose that we are analyzing two forms f_1 and f_2 which are semantically equivalent and that we have some meaning m_1 that is optimal for both forms. In Interpretation OT the two forms would not belong to the same candidate set and thus would both be grammatical. In the Strong OT model, f_2 , even if optimal in the interpretation-based optimization, may be blocked by the more economical alternative form f_1 . Hence, the form-meaning pair $\langle f_2, m_1 \rangle$ is removed from the set of the language generated by the Strong OT system. We can illustrate this by the following picture:



Strong OT also opens up a simple way of modeling blocking of meaning, a phenomenon which is unaccounted for under unidirectional production OT. Consider the Marathi example from section 2 repeated in (6) below.

- (6) Tom_i mhanat hota [ki Sue_j ni *swataahlaa*_{*i/j} maarle]. [Marathi]
 Tom said that Sue ERG ANAPHOR-ACC hit
 'Tom said that Sue hit herself/*him.'

Example (6) has the form $[A_i \dots [\delta B_j \dots \text{anaphor} \dots]]$, in which A and B are potential antecedents for the anaphor and δ is the domain in which the anaphor must have an antecedent (the minimal finite clause that contains the anaphor). Parsing this sentence will result in two classes of analyses: one in which the binding relation is local (i.e., anaphor = j) and one in which the binding relation is non-local (i.e., anaphor = i). In production-based optimization, the two interpretations do not compete with each other and thus the sentence is grammatical for

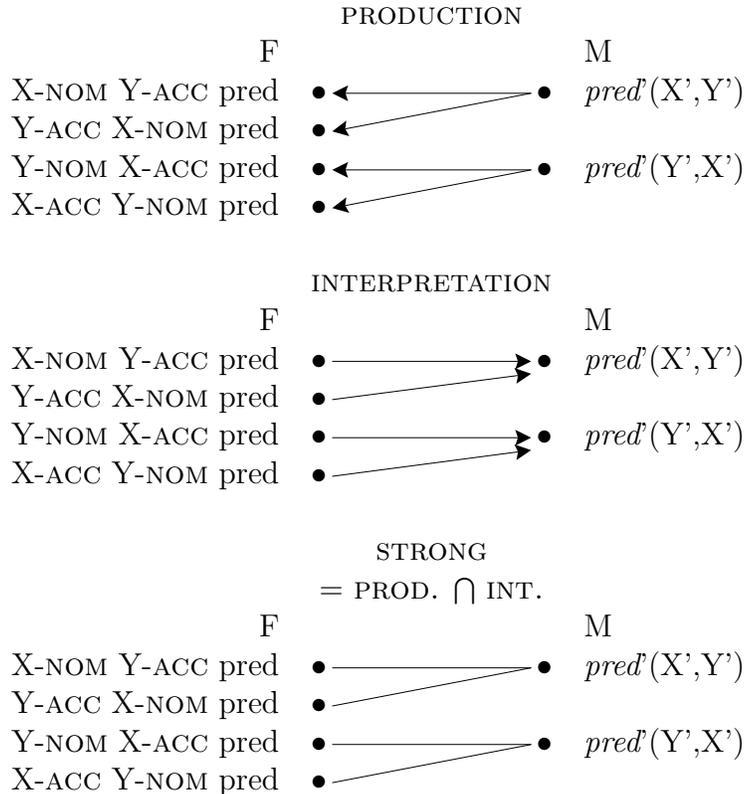
- (9) a. Mary-ka nonmwun-ul sse-ss-ta. [Korean]
 Mary-NOM paper-ACC wrote-PST-DECL
 ‘Mary wrote a paper.’
- b. nonmwun-ul Mary-ka sse-ss-ta.

Lee (2001) assumes two conflicting constraints on word order first proposed by Choi (1999); a canonical word order constraint (10a) and a discourse-based word order constraint(10b):

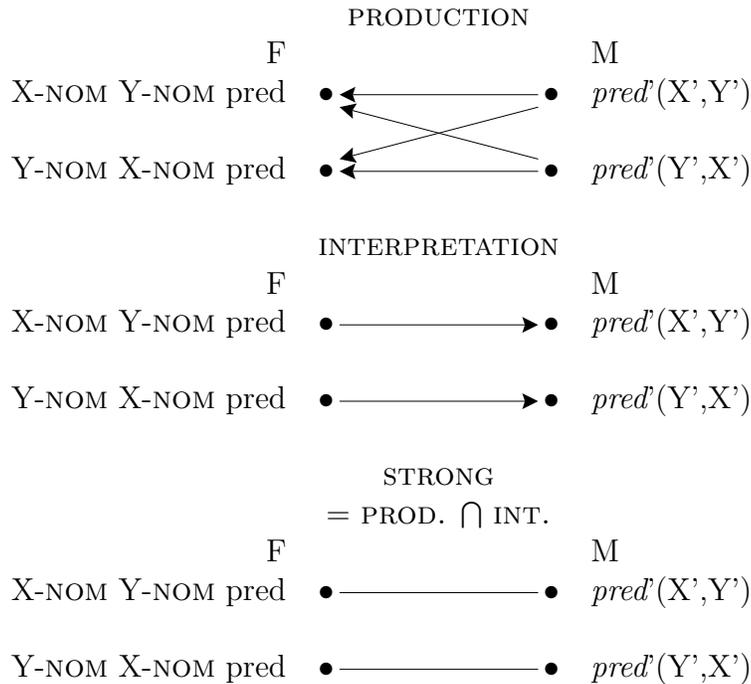
- (10) a. SO: Subject precedes object.
 b. TOPIC: Topic precedes non-topic.

The ranking TOPIC \gg SO ensures that object-subject order is optimal, if the object is marked [+TOPIC] in the input. When it is not marked [+TOPIC], however, the TOPIC constraint is vacuously satisfied and the lower-ranked SO constraint becomes active, favoring subject-object order over object-subject order.⁹

What does bidirectional optimization predict for sentences like (9)? In Strong OT the two surface forms that correspond to winners of different production optimizations are evaluated in comprehension optimization. As illustrated in the diagram below, both forms (‘X-NOM Y-ACC pred’ and ‘Y-ACC X-NOM pred’) are interpreted as having the same underlying structure, a structure corresponding to the original input to production. Any alternative interpretation, for example a candidate which interprets an accusative NP as an agent, would violate higher-ranked faithfulness constraints on case interpretation and case markedness constraints, and hence is eliminated from the competition.



However, applying optimization in both directions produces rather surprising results for sentences with arguments that are identically case-marked. For such cases, high-ranking faithfulness constraints on case interpretation and markedness constraints penalizing marked grammatical function/case associations are inapplicable (hence inactive) and low-ranking constraints that prefer canonical word order become decisive. The result is the subject-object interpretation of potentially ambiguous strings. The marked object-subject interpretation is eliminated not because it violates high-ranking faithfulness constraints but because it violates low-ranking alignment constraints. We can illustrate this graphically as follows:



Lee (2001) thus argues that if we define grammaticality in terms of bidirectional optimization, word order freezing within particular languages can be accounted for as an ‘emergence of the unmarked’ (McCarthy and Prince 1994) in interpretation-based optimization, based on the same set of constraints that characterize cross-linguistic variation in case patterns and word order.

In sum, Strong OT offers a unified approach to the problems of ineffability, uninterpretability, total blocking and freezing. However, Strong OT does not help with ambiguity and optionality. Since the set of Strong OT meaning-form pairs is a subset of those provided by naive interpretation for a given constraint set, Strong OT deals with ambiguity as badly as naive interpretation does. And since the set of Strong OT meaning-form pairs is a subset of those provided by naive production, it does not account for optionality either.

A related problem of Strong OT, pointed out by Blutner (2001), is that the blocking effect is so strict. For example, strong OT predicts that “cause to die”, since it is blocked by the lexicalized “kill”, should be uninterpretable. But in fact it is only partially blocked, and comes to have an application in situations where “kill” would be deemed inappropriate. We now turn to Blutner’s proposed solution to this problem.

4 Weak Bidirectional Optimization

Blutner’s *weak* notion of optimality, which we refer to simply as Weak OT, is an iterated variant of Strong OT that produces partial blocking instead of strict blocking. In Weak OT, sub-optimal candidates in a strong bidirectional competition can become winners in a second or later round of optimization. As we will see, in Weak OT, everyone is a winner.

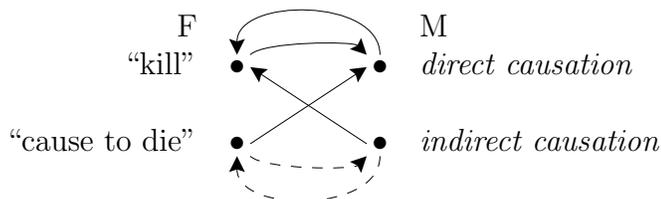
Strong OT picks out a set of form-meaning pairs such that none of them is beaten by any form-meaning pair in GEN in either direction of optimization. Weak OT picks out a larger set of form-meaning pairs such that no member of that set beats any other member of the set in either direction of optimization. Thus some of the Weak OT optimal pairs may be beaten by other pairs in GEN. One may say that some Weakly optimal pairs are sub-optimal. Crucially, these sub-optimal optimal pairs can only be beaten by form-meaning pairs that are themselves blocked. For example, the pair $\langle \text{“cutter”}, \textit{non-canonical cutting implement} \rangle$ could be weakly optimal, even though it might be beaten by the pair $\langle \text{“cutter”}, \textit{knife} \rangle$ in a full competition amongst pairs in GEN. But this is only possible if the latter pair is itself blocked, e.g. beaten by the pair $\langle \text{“knife”}, \textit{knife} \rangle$.

The formal definition of optimality in Weak OT runs along similar lines to the Strong OT definition, but is recursive:

- (11) $\langle f, m \rangle$ is weak OT optimal iff
- a. $\langle f, m \rangle \in \text{GEN}$,
 - b. there is no Weak OT optimal $\langle f', m \rangle \in \text{GEN}$ such that $\langle f', m \rangle \succ \langle f, m \rangle$, and
 - c. there is no Weak OT optimal $\langle f, m' \rangle \in \text{GEN}$ such that $\langle f, m' \rangle \succ \langle f, m \rangle$.

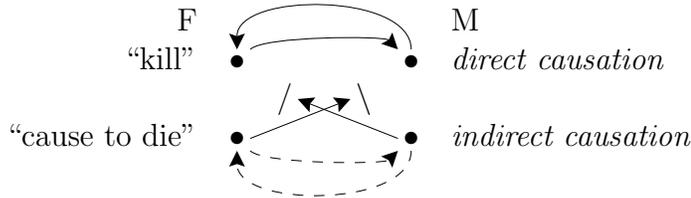
The application of Weak OT, described formally by Blutner (2001), Blutner and Jäger (1999) and Jäger and Blutner (2000), can be thought of as involving repeated pruning and grafting of links between forms and meanings. We illustrate the Weak OT pruning and grafting cycle using the example of lexical and periphrastic causatives “kill” / “cause to die” which we assume are matched on the meaning side by two possible interpretations, direct causation (canonical killing) and indirect causation (non-canonical killing). The following three diagrams, illustrate three phases of weak optimization. In the first diagram, all the unidirectionally optimal links are shown. In addition to the optimal links, two links are shown with dashed lines. Both of these links are unidirectionally sub-optimal at this stage, beaten by other candidates.

PHASE 1 — NAIVE INTERPRETATION AND PRODUCTION:



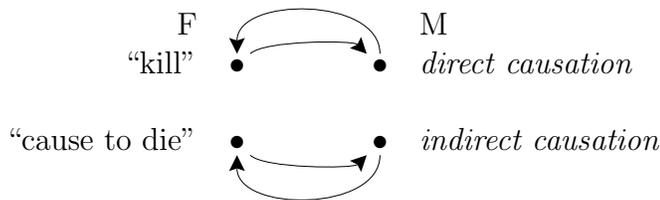
In phase 2 of Weak optimization, two unidirectionally optimal links are blocked, leaving a single bidirectionally optimal link, that between the form “kill” and the meaning corresponding to direct causation.

PHASE 2 — PRUNING:



Now we graft the originally sub-optimal links between “cause to die” and the indirect causation meaning back into the picture, since the candidates which originally beat them have been removed by blocking. This gives us two bidirectionally optimal links. In the resulting happy picture, all the candidate meanings are uniquely expressible and all the candidate forms are uniquely interpretable:

PHASE 3 — GRAFTING:



Blutner (2001) argues that Weak OT captures the essence of the pragmatic generalization that “unmarked forms tend to be used for unmarked situations and marked forms for marked situations” (Horn 1984:26; see also Levinson 2000:136). The concept also seems useful for deriving various alignment scales that are widely used in OT syntax work (e.g., Aissen 1999), suggesting an interesting connection to (psychologically inspired) prototype theory. But there is a dark side to Weak OT.

First, note that Weak OT does not help with ambiguity and optionality.¹⁰ Weak bidirectionality would predict (i) that for a form f , only one meaning is available if one of the meanings in pairs $\langle f, m_1 \rangle$ and $\langle f, m_2 \rangle$ incurs a more serious constraint violation and (ii) that of two forms that are semantically equivalent, only one form is grammatical if one of the forms in $\langle f_1, m \rangle$ and $\langle f_2, m \rangle$ involves a more serious constraint violation. The grafting stage of Weak OT can add links to make an ineffable meaning expressible, or to give meaning to an uninterpretable form. But it cannot add new ways to express a meaning that is already expressible, or add meanings to a form that is already interpretable. So we are stuck with just the same ability to deal with ambiguity and optionality that we had in Strong OT, i.e. probably not enough.

Besides this problem of undergeneration, Weak OT suffers from a more serious problem of overgeneration. Specifically, the process of adding extra links will eventually provide links for every form (if there are at least as many forms as meanings), or every meaning (if there are at least as many meanings as forms). This poses an empirical problem for uninterpretability and ineffability, and indeed also for the blocking phenomena which Weak OT was designed to account for.

The problem of overgeneration becomes intuitively clear when we apply weak bidirectionality to cases involving a fair number of form and meaning alternatives. A good example is the case pattern and relatively free word order in Korean, modeled within OT by Lee (2001, to appear).

In canonical transitives, the case pattern in Korean is nominative-accusative, as seen in the examples in (9) above. The order of nominal arguments of the verb is relatively flexible, except for a strong verb-final restriction. However, as mentioned in section 2, word order in this language is not random. Rather, the varied word orders are motivated by discourse and semantic factors.

Lee (2001, to appear) models the case pattern and word order variation in Korean, assuming competing sets of case markedness constraints and alignment constraints. For our purpose here, it suffices to consider the following five constraints, ranked in the order shown in (12):

- (12) a. *SUBJ/ACC: Subject is not in the accusative case.
 b. HEAD-R: Head aligns right in its projection (e.g., VP) (Grimshaw 1997).
 c. SO: Subject precedes object (Choi 1999).
 d. *SUBJ^{new}: Subject is not discourse-new information.
 e. *OBJ^{given}: Object is not given information.

We now consider how a plausible set of forms and meanings, shown in (13), are evaluated with respect to the constraints in (12) in Weak OT. The six forms differ in argument-case association and the surface order of the head and argument NPs; the five meanings differ in argument-function association and the givenness of arguments.¹¹

forms	meanings
f_1 : X-NOM Y-ACC V	m_1 : $pred^i(X', Y')$, X=S ^{given} , Y=O ^{given}
f_2 : X-ACC Y-NOM V	m_2 : $pred^i(X', Y')$, X=O ^{given} , Y=S ^{given}
f_3 : X-ACC Y-ACC V	m_3 : $pred^i(X', Y')$, X=O ^{new} , Y=S ^{given}
f_4 : X-NOM V Y-ACC	m_4 : $pred^i(X', Y')$, X=S ^{new} , Y=O ^{new}
f_5 : X-ACC V Y-NOM	m_5 : $pred^i(X', Y')$, X=O ^{new} , Y=S ^{new}
f_6 : X-ACC V Y-ACC	

Of 30 possible pairs of forms and meanings in (13), we will consider just the evaluation of 12 pairs in the tableaux that follow. They are shown in Tableau 1. Candidates labeled with the same alphabetical letter share the same meaning and differ only in positioning of the verb. (● indicates a candidate blocked by another candidate with the same form and ○, a candidate blocked by another candidate with the same meaning; ♯ marks a bidirectionally optimal form-meaning pair.) Due to bidirectional optimization, the evaluation procedure is somewhat different from standard OT: checking whether a form-meaning pair is optimal requires simultaneous evaluations of form alternatives and meaning alternatives. Tableau 1 corresponds loosely to phase 1 of the treatment of “kill”/“cause to die” above, showing which candidates are superior in both comprehension and production. Candidate (a), with the given nominative subject and the given accusative object, emerges immediately as a bidirectionally optimal form-meaning pair.

(14) Tableau 1. 1st round of optimization (Weak OT)

	*SUBJ/ACC	HEAD-R	SO	*SUBJ ^{new} ,	*OBJ ^{given}
♯ a. S/NOM ^{given} ₁ O/ACC ^{given} ₂ V ($\langle f_1, m_1 \rangle$)					*
b. O/ACC ^{given} ₁ S/NOM ^{given} ₂ V ($\langle f_2, m_2 \rangle$)			*		*
c. O/ACC ^{given} ₂ S/NOM ^{given} ₁ V ($\langle f_2, m_1 \rangle$)			*		*
d. S/NOM ^{new} ₂ O/ACC ^{given} ₁ V • ($\langle f_1, m_3 \rangle$)				*	*
e. S/ACC ^{new} ₁ O/ACC ^{new} ₂ V ($\langle f_3, m_4 \rangle$)	*			*	
f. S/ACC ^{new} ₂ O/ACC ^{new} ₁ V ($\langle f_3, m_5 \rangle$)	*			*	
a'. S/NOM ^{given} ₁ V O/ACC ^{given} ₂ ○ ($\langle f_4, m_1 \rangle$)		*			*
b'. O/ACC ^{given} ₁ V S/NOM ^{given} ₂ ($\langle f_5, m_2 \rangle$)		*	*		*
c'. O/ACC ^{given} ₂ V S/NOM ^{given} ₁ ($\langle f_5, m_1 \rangle$)		*	*		*
d'. S/NOM ^{new} ₂ V O/ACC ^{given} ₁ ($\langle f_4, m_3 \rangle$)		*		*	*
e'. S/ACC ^{new} ₁ V O/ACC ^{new} ₂ ($\langle f_6, m_4 \rangle$)	*	*		*	
f'. S/ACC ^{new} ₂ V O/ACC ^{new} ₁ ($\langle f_6, m_5 \rangle$)	*	*		*	

In the Strong bidirectional model, we would already be finished. But in Weak OT, we have to consider the next best candidates in competitions that do not involve links blocked by the bidirectionally optimal candidate (candidate (a)). Recall that under weak bidirectionality the structures that compete in production-based optimization are constrained by the outcomes of interpretation-based optimization and *vice versa*. Hence candidates (d) and (a'), which lose out to candidate (a) in either direction, are not contained in the candidate set for further optimization procedures. Furthermore, we remove the winning candidate (a) from the tableau: it should not be compared directly with any of the remaining candidate pairs, since it has neither the same form nor the same meaning as any of them. Hence we arrive at the tableau in (15). As can be seen, candidate (b), in which the given accusative object precedes the given nominative subject, is selected as a winner, though it could not win under Strong OT. This is a desirable result, as this candidate, even if it violates the canonical word order requirement, is clearly a grammatical option for Korean.

(15) Tableau 2. 2nd round of optimization (Weak OT)

	*SUBJ/ACC	HEAD-R	SO	*SUBJ ^{new} ,	*OBJ ^{given}
$\frac{1}{2}$ b. O/ACC ^{given} ₁ S/NOM ^{given} ₂ V ($\langle f_2, m_2 \rangle$)			*		*
c. O/ACC ^{given} ₂ S/NOM ^{given} ₁ V ($\langle f_2, m_1 \rangle$) •			*		*
e. S/ACC ^{new} ₁ O/ACC ^{new} ₂ V ($\langle f_3, m_4 \rangle$)	*			*	
f. S/ACC ^{new} ₂ O/ACC ^{new} ₁ V ($\langle f_3, m_5 \rangle$)	*			*	
b'. O/ACC ^{given} ₁ V S/NOM ^{given} ₂ ($\langle f_5, m_2 \rangle$) ○		*	*		*
c'. O/ACC ^{given} ₂ V S/NOM ^{given} ₁ ($\langle f_5, m_1 \rangle$) ○		*	*		*
d'. S/NOM ^{new} ₂ V O/ACC ^{given} ₁ ($\langle f_4, m_3 \rangle$)		*		*	*
e'. S/ACC ^{new} ₁ V O/ACC ^{new} ₂ ($\langle f_6, m_4 \rangle$)	*	*		*	
f'. S/ACC ^{new} ₂ V O/ACC ^{new} ₁ ($\langle f_6, m_5 \rangle$)	*	*		*	

However, the process of recursion continues, and produces unintuitive consequences. Tableaux 3, 4 and 5 below show what happens when we consider next best candidates, even though we already found the best two. What we find is that there are many candidates generated by the Weak OT system that are not grammatical in the language modeled: none of the winners in Tableaux 3, 4 and 5 are acceptable. This shows that the present form of Weak OT is highly problematic as a model of synchronic linguistic competence.

(16) Tableau 3. 3rd round of optimization (Weak OT)

	*SUBJ/ACC	HEAD-R	SO	*SUBJ ^{new} ,	*OBJ ^{given}
e. S/ACC ^{new} ₁ O/ACC ^{new} ₂ V ($\langle f_3, m_4 \rangle$)	*			*	
f. S/ACC ^{new} ₂ O/ACC ^{new} ₁ V ($\langle f_3, m_5 \rangle$)	*			*	
$\frac{1}{2}$ d'. S/NOM ^{new} ₂ V O/ACC ^{given} ₁ ($\langle f_4, m_3 \rangle$)		*		*	*
e'. S/ACC ^{new} ₁ V O/ACC ^{new} ₂ ($\langle f_6, m_4 \rangle$)	*	*		*	
f'. S/ACC ^{new} ₂ V O/ACC ^{new} ₁ ($\langle f_6, m_5 \rangle$)	*	*		*	

(17) Tableau 4. 4th round of optimization (Weak OT)

	*SUBJ/ACC	HEAD-R	SO	*SUBJ ^{new} ,	*OBJ ^{given}
$\frac{1}{2}$ e. S/ACC ^{new} ₁ O/ACC ^{new} ₂ V ($\langle f_3, m_4 \rangle$)	*			*	
$\frac{1}{2}$ f. S/ACC ^{new} ₂ O/ACC ^{new} ₁ V ($\langle f_3, m_5 \rangle$)	*			*	
e'. S/ACC ^{new} ₁ V O/ACC ^{new} ₂ ○ ($\langle f_6, m_4 \rangle$)	*	*		*	
f'. S/ACC ^{new} ₂ V O/ACC ^{new} ₁ ○ ($\langle f_6, m_5 \rangle$)	*	*		*	

(18) Tableau 5. 5th round of optimization (Weak OT)

		*SUBJ/ACC	HEAD-R	SO	*SUBJ ^{new} , *OBJ ^{given}
g'. ... (V O S?)					
h'. ...					

The problem of overgeneration just mentioned obviously affects accounts of phenomena other than Korean word-order freezing. Before closing this section, we discuss its significance for ineffability.

There have been several proposals within standard OT to deal with cases of ineffability. Among these proposals are reference to null parses (Prince and Smolensky 1993), the assumption of LF-unfaithful candidates (Legendre, Smolensky and Wilson 1998), and the postulation of the lexical control component that is imposed on the optimal candidates computed by EVAL (Orgun and Sprouse 1999). The addition of the control component may be called for independently to deal with cases of ineffability which arise from the absence of certain lexical items,¹² whereas the former two amendments of standard OT have been criticized as highly problematic from linguistic and learnability points of view (e.g., Kuhn 2001).

Smolensky, in unpublished work Smolensky (1998), has proposed a solution to language-particular ineffability, based on bidirectional optimization. What we will show is that even though a bidirectional approach may be merited, Weak OT does not fit the bill.

Recall the discussion of multiple *wh*-questions in Italian, illustrated in (1): while English has single clause multiple-*wh* questions, Italian does not. This is because in Italian, a markedness constraint that is violated by multiple *wh*-questions in a single clause (Legendre, Smolensky and Wilson call this *ABSORB) is ranked higher than a faithfulness constraint to the *wh*-feature in the input (PARSE(WH)). According to the analysis of Legendre, Smolensky and Wilson (1998), English resolves the conflict at the cost of violating the markedness constraint. Another option for resolving the conflict is by adjoining both *wh*-phrases in [Spec, CP], as Bulgarian does. However, this option is unavailable in Italian and violates another markedness constraint *ADJOIN, which also dominates PARSE(WH) in Italian.¹³

Let us now look at what is predicted by Weak bidirectional optimization. Here, we will just go through a simplified analysis to illustrate the general effects of Weak OT; the table in (19) shows some sample forms and meanings that are relevant to our discussion:

forms	meanings
f_1 : <i>who ate what</i>	m_1 : $?xyate(x, y)$
f_2 : <i>who ate something</i>	m_2 : $?x\exists yate(x, y)$
f_3 : <i>who what ate</i>	m_3 : $?xate(x, y)$
f_4 : <i>who ate</i>	m_4 : $?xate(x, y), y = familiar$

The bidirectional competition for possible form-meaning pairs is shown in Tableau 6. With the ranking in (20), candidate (b2) is correctly predicted to be the winner in Italian; candidate (b3) is selected also as the winner, but the small set of constraints we use here does not differentiate it from (b2):

(20) Ranking for Italian: *[wh wh], *ADJOIN \gg PARSE(WH) \gg MARK-FAM¹⁴ \gg PARSE

(21) Tableau 6. Multiple *wh*-questions in Italian (Weak OT)

	*[wh wh],	*ADJOIN	PARSE(WH)	FAM-DEF	PARSE
a1. $\langle who\ ate\ what, ?xyate(x, y) \rangle$	*				
b1. $\langle who\ ate\ something, ?xyate(x, y) \rangle \bullet$			*		
c1. $\langle who\ what\ ate, ?xyate(x, y) \rangle$		*			
d1. $\langle who\ ate, ?xyate(x, y) \rangle$			*		*
a2. $\langle who\ ate\ what, ?x\exists yate(x, y) \rangle \circ$	*				
$\cancel{b2}$. $\langle who\ ate\ something, ?x\exists yate(x, y) \rangle$					
c2. $\langle who\ what\ ate, ?x\exists yate(x, y) \rangle \circ$		*			
d2. $\langle who\ ate, ?x\exists yate(x, y) \rangle \circ$					*
a3. $\langle who\ ate\ what, ?xate(x, y) \rangle \circ$	*				
$\cancel{b3}$. $\langle who\ ate\ something, ?xate(x, y) \rangle$					
c3. $\langle who\ what\ ate, ?xate(x, y) \rangle \circ$		*			
d3. $\langle who\ ate, ?xate(x, y) \rangle \circ$					*
a4. $\langle who\ ate\ what, ?xate(x, y), y = familiar \rangle$	*			*	
b4. $\langle who\ ate\ something, ?xate(x, y), y = familiar \rangle \bullet$				*	
c4. $\langle who\ what\ ate, ?xate(x, y), y = familiar \rangle$		*		*	
d4. $\langle who\ ate, ?xate(x, y), y = familiar \rangle$				*	*

After the bidirectionally optimal candidates (b2) and (b3) have been removed from the candidate sets, candidate (d4), which could not win in the first round of optimization, becomes the winner:

(22) Tableau 7. Multiple *wh*-questions in Italian (Weak OT)

	*[wh wh],	*ADJOIN	PARSE(WH)	DEF-FAM	PARSE
a1. $\langle who\ ate\ what, ?xyate(x, y) \rangle$	*				
c1. $\langle who\ what\ ate, ?xyate(x, y) \rangle$		*			
d1. $\langle who\ ate, ?xyate(x, y) \rangle \bullet$			*		*
a4. $\langle who\ ate\ what, ?xyate(x, y), y = familiar \rangle \circ$	*			*	
c4. $\langle who\ what\ ate, ?xyate(x, y), y = familiar \rangle \circ$		*		*	
$\cancel{d4}$. $\langle who\ ate, ?xyate(x, y), y = familiar \rangle \circ$				*	*

The third competition certainly does not give us the correct result for Italian. As Tableau 8 shows, it predicts that multiple *wh*-questions (f_1 and f_3) are the optimal expression for the multiple *wh*-input $?xyate(x, y)$, and $?xyate(x, y)$ is the optimal meaning for the relevant multiple *wh*-question. For Italian, these are unwelcome predictions.

(23) Tableau 8. Multiple *wh*-questions in Italian (Weak OT)

		*[wh wh]	*ADJOIN	PARSE(WH)	FAM-DEF	PARSE
♯	a1. $\langle who\ ate\ what, ?xyate(x, y) \rangle$	*				
♯	c1. $\langle who\ what\ ate, ?xyate(x, y) \rangle$		*			

It is not hard to see that ineffability is predicted by Weak OT only if all possible realizations for an input representation are optimal for some other meanings. As Kuhn (2001) points out, however, this does not give us the correct result for Italian, because all strings, including *Che ha mangiato che cuesta*, are predicted to be grammatical for some other meanings. Furthermore, Weak OT does not predict any difference between Italian and English: candidates (a1) and (b2) are predicted to be grammatical in both Italian and English under different rankings.

Although we will not provide detailed analyses, it should be obvious that these same over-generation problems would affect the Weak OT analysis of total blocking. While in the first phase of optimization the successful Strong OT predictions appear to be reproduced, in latter stages peculiar new form-meaning pairs will emerge as winners. Provided the set of candidate meanings is large, Weak OT never predicts total blocking: all blocking is partial. So “writed”, for example, would presumably be the correct expression of some meaning in Strong OT.

There remains one chink of light for Weak OT: word order freezing is still predicted, as in Strong OT, and so, for example, a Korean double nominative construction is predicted to have only a subject-object interpretation. Consider in the abstract the two forms X-NOM Y-NOM pred and Y-NOM X-NOM pred: both of these forms will be paired with meanings in the first phase of Weak optimization, so neither will enter into later competitions, and neither will become associated with incorrect argument mappings.

5 Interpretability as a constraint on production

In this and the following section we consider asymmetric models of bidirectional OT in which interpretation and production optimizations are understood to be applied in sequence, such that the first optimization affects the candidate set for the second.

Wilson (2001) discusses a model in which interpretation precedes production.¹⁵ We refer to this as Asymmetric OT (I(nterpretation)P(roduction)).¹⁶ In more detail, the idea of Asymmetric OT (IP) is as follows: (i) Interpretation: Given any form-meaning pair $\langle f, m \rangle$, find the most harmonic semantic interpretation of f . (ii) Production: Given input meaning m , take as candidate outputs the set of forms f such that $\langle f, m \rangle$ is optimal in stage one, and perform standard OT production optimization with this restricted candidate set. Note that the set of optimal form-meaning pairs in production is a subset of the optimal form-meaning pairs in interpretation. The set of meanings which are in some optimal pair is the same in interpretation and production, although the number of forms

would, for constraint sets which are of interest, be smaller in production than in comprehension. It is the reduced set of forms in production, those which result from the two stage process, which are to be considered grammatical, even though there are others which are interpretable.

Wilson (2001) uses this version of OT to model certain cases of partial blocking. In what follows we briefly review the Asymmetric OT (IP) treatment of partial blocking involving relativized minimality (see example (6)) and referential economy in anaphor binding. An example of a referential economy effect is provided by the following contrast between the Icelandic third-person pronoun *hann* and the anaphor *sig*:

(24) Referential economy in Icelandic (Maling 1984: 212)

- a. Haraldur_i skipaði mér að raka *hann_i/sig_i.
Harold ordered me to shave him/ANAPHOR
'Harold ordered me to shave him.'
- b. Jón_i veit að María elskar hann_i/*sig_i.
Jon knows that Maria loves him/ANAPHOR
'Jon knows that Maria loves him.'

In (24a), the matrix subject *Haraldur* can grammatically bind the anaphor but not the pronoun. In (24b), in contrast, the pronoun is grammatical.

According to Wilson, contrasts like the one in (24) follow from an interaction of two constraints: the LOCAL ANTECEDENT constraint (25a), which is a locality requirement on anaphor binding, and the REFERENTIAL ECONOMY constraint (25b), which requires a bound element to be an anaphor:

- (25) a. LOCAL ANTECEDENT: If a syntactic domain of type δ contains an anaphor α , then it also contains an antecedent for α .
- b. REFERENTIAL ECONOMY: An argument does not have any lexical agreement feature specifications.

The ranking that Wilson assumes for partial blocking in anaphor binding is:

(26) LOCAL ANTECEDENT \gg REFERENTIAL ECONOMY

The main effects of these constraints in anaphor binding are as follows. When a binding relation is sufficiently local (e.g., as in (24a), when it crosses only the boundary of an infinitival clause), an anaphor need not be bound within the infinitival clause that contains it. In such a case, the anaphor, by virtue of being lexically devoid of certain agreement features,¹⁷ is preferred to the pronoun by referential economy. But when the binding relation is nonlocal, as in (24b), the anaphor is excluded by LOCAL ANTECEDENT and the bound element must be realized as a pronoun. However, unidirectional production would predict that the nonlocal bound-variable interpretation is always expressed with an anaphor, since it is less marked than a pronoun in terms of referential economy.

Strong OT suffers from the same problem of strict blocking. The following tableaux will be useful for contrasting the Strong OT analysis and the Asymmetric

OT (IP) treatment of the anaphora data above (to be discussed shortly) more clearly. Consider first the tableaux in (27) and (28), which illustrate interpretation optimizations based on two forms containing bound elements (an anaphor (f_1) and a pronoun (f_2)). There are two potential antecedents one within the minimal finite clause, here labeled δ and one outside that clause. The two candidates we consider are the local binding interpretation (m_1) and the nonlocal binding interpretation (m_2).

For the interpretation optimization in Tableau 9, REFERENTIAL ECONOMY has no effect, since both candidates contain a bound anaphor. Thus, LOCAL ANTECEDENT gives us candidate (a) as the winner.

(27) Tableau 9. Interpretation I (Strong OT)

		REFERENTIAL ECONOMY	LOCAL ANTECEDENT
	Input: [A [δ B ... anaphor]] (f_1)		
☞	a. [A $_i$ [δ B $_j$... anaphor $_j$]] ($\langle f_1, m_1 \rangle$)		
	b. [A $_i$ [δ B $_j$... anaphor $_i$]] ($\langle f_1, m_2 \rangle$)		*

In the interpretation optimization with the string containing a pronoun as the input, both candidates have the same constraint profile for REFERENTIAL ECONOMY and LOCAL ANTECEDENT, so both are selected as winners:

(28) Tableau 10. Interpretation II (Strong OT)

		REFERENTIAL ECONOMY	LOCAL ANTECEDENT
	Input: [A [δ B ... pronoun]] (f_2)		
☞	a. [A $_i$ [δ B $_j$... pronoun $_j$]] ($\langle f_2, m_1 \rangle$)	*	
☞	b. [A $_i$ [δ B $_j$... pronoun $_i$]] ($\langle f_2, m_2 \rangle$)	*	

In production optimizations based on m_1 and m_2 , on the other hand, due to the higher-ranking constraint REFERENTIAL ECONOMY, the same candidate (a) wins for both inputs:

(29) Tableau 11. Production I (Strong OT)

		REFERENTIAL ECONOMY	LOCAL ANTECEDENT
	Input: local binding (m_1)		
☞	a. [A $_i$ [δ B $_j$... anaphor $_j$]] ($\langle f_1, m_1 \rangle$)		
	b. [A $_i$ [δ B $_j$... pronoun $_i$]] ($\langle f_2, m_1 \rangle$)	*	

(30) Tableau 12. Production II (Strong OT)

		REFERENTIAL ECONOMY	LOCAL ANTECEDENT
Input: nonlocal binding (m_2)			
☞	a. $[A_i[\delta B_j \dots \text{anaphor}_i]] (\langle f_1, m_2 \rangle)$		*
	b. $[A_i[\delta B_j \dots \text{pronoun}_i]] (\langle f_2, m_2 \rangle)$	*	

Thus Strong OT produces only one bidirectionally optimal form-meaning pair, i.e., $\langle f_1, m_1 \rangle$, failing to predict partial blocking.

Wilson (2001) offers an Asymmetric OT (IP) account of these facts that overcomes these problems. Crucially, in Wilson's model, interpretation optimization applies first to limit the candidate set for the second, production optimization. To see how the analysis works, compare the tableaux in (29) and (30) with the ones in (31) and (32) below, which correspond to the second stage of optimization in Asymmetric OT (IP).¹⁸ As we noted above, in the Strong OT model, the results of optimization under one direction does not affect which candidates compete under the other direction because the candidate set of both directions of optimization is defined independently. Consequently, all the four form-meaning pairs in the above interpretation tableaux compete under the production optimization also. But in Asymmetric OT (IP), only winning candidates in interpretation enter into the production optimization.

For the anaphora data under discussion here, the consequence of this is as follows: since m_2 loses in the interpretation tableau with input f_1 (Tableau 9), the production competition with m_2 as input no longer includes the candidate f_1 . That is, the original production tableau which took m_2 as input (Tableau 12) must be replaced by Tableau 14, which does not include candidate (a). As a result, candidate (b) wins trivially, and m_2 is predicted to be realized as f_2 . Meanwhile, the production tableau for meaning m_1 (Tableau 11) is unaffected, so m_1 is still realized as f_1 :

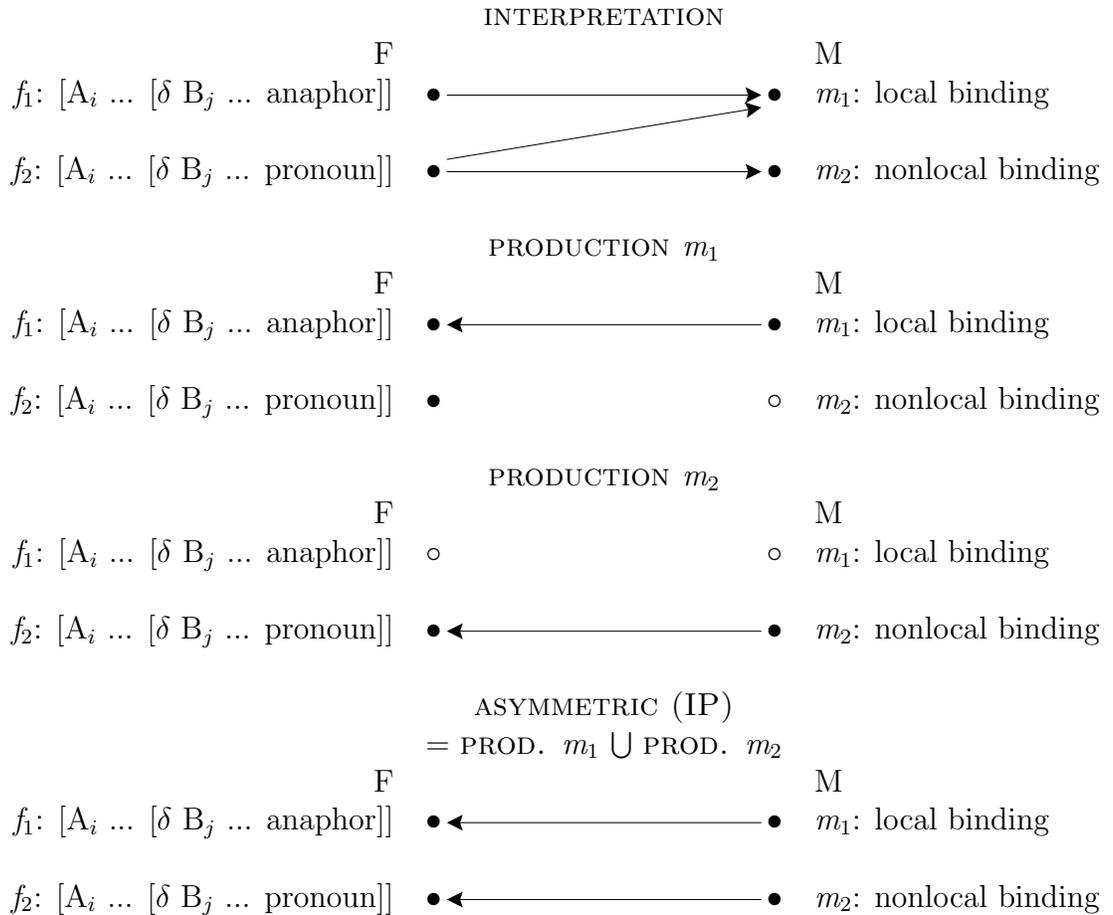
(31) Tableau 13. Production I (Asymmetric OT (IP))

		REFERENTIAL ECONOMY	LOCAL ANTECEDENT
Input: local binding (m_1)			
☞	a. $[A_i[\delta B_j \dots \text{anaphor}_j]] (\langle f_1, m_1 \rangle)$		
	b. $[A_i[\delta B_j \dots \text{pronoun}_i]] (\langle f_2, m_1 \rangle)$	*	

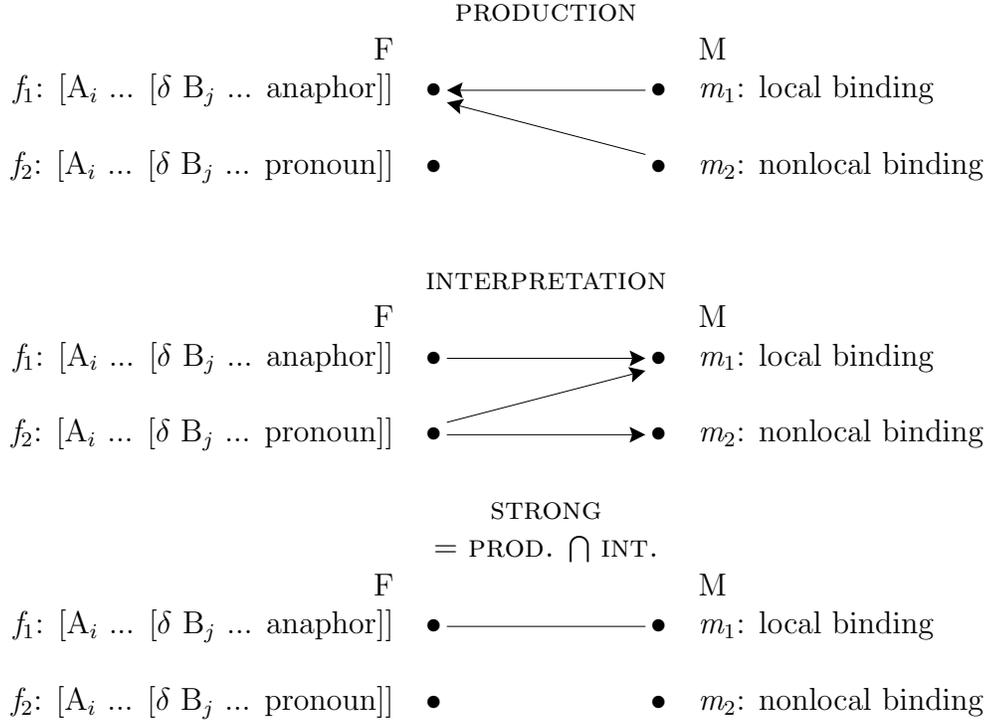
(32) Tableau 14. Production II (Asymmetric OT (IP))

Input: nonlocal binding (m_2)	REFERENTIAL ECONOMY	LOCAL ANTECEDENT
☞ b. $[A_i [\delta B_j \dots \text{pronoun}_i]] (\langle f_2, m_2 \rangle)$	*	

The process Wilson describes is pictured in the following diagram, where candidates are marked using “o” for those competitions where they are not participants:



We may compare Wilson’s successful account of referential economy with the results that would be obtained in Blutner’s models. Whereas Weak OT, which deals quite effectively with partial blocking, would successfully predict the Icelandic data, Strong OT would be less successful. As the following diagram shows, under the constraints assumed, Strong OT incorrectly predicts that Icelandic pronouns are uninterpretable in the given configuration, and that there is no way of expressing nonlocal binding:



So far we have looked at the Asymmetric OT (IP) analysis of partial blocking in anaphor binding. What of the standard cases of partial blocking we considered earlier? Can they be modeled in Asymmetric OT (IP)? It is interesting to note that all cases of partial blocking are subject to two similar kinds of constraints: one that favors a less marked form and the other that favors a less marked meaning. In the case of Icelandic anaphor binding, REFERENTIAL ECONOMY concerns formal markedness, and LOCAL ANTECEDENT concerns semantic markedness; in the example of causatives discussed in the previous section, the formal markedness constraint was a preference for short forms, and the semantic markedness constraint was a preference for the canonical mode of causation.

Yet there is an important difference between the phenomena Wilson models and the partial blocking cases considered earlier. What distinguishes Wilson's anaphora data is that the pair of a marked form and an unmarked meaning ($\langle f_2, m_1 \rangle$ in the above tableaux) and the pair of a marked form and a marked meaning ($\langle f_2, m_2 \rangle$ in the above tableaux) have the same constraint profile for the constraint favoring a less marked meaning (see Tableaux 9 and 10 above; see also Wilson (2001: 496–498) for a detailed discussion). As noted above, the LOCAL ANTECEDENT constraint, preferring local binding over nonlocal binding, targets only an anaphor (f_1) but not a pronoun (f_2). As a result, the pairs $\langle f_2, m_1 \rangle$ and $\langle f_2, m_2 \rangle$ both survive in interpretation. Now when we come to realize m_1 , we don't choose f_2 but instead choose f_1 . In other words, in production, as illustrated in Tableaux 13 and 14, the pair $\langle f_1, m_1 \rangle$ blocks $\langle f_2, m_1 \rangle$, making $\langle f_2, m_2 \rangle$ available.

The standard cases of partial blocking differ in that the two pairs $\langle \textit{marked form}, \textit{unmarked meaning} \rangle$ and $\langle \textit{marked form}, \textit{marked meaning} \rangle$ do not have the same constraint profile. This is illustrated in (33):

(33) Tableau 15. Interpretation

Input: <i>cause to die</i>	ECONOMY	CANON
☞ a. $\langle \textit{cause to die, direct causation} \rangle$	*	
b. $\langle \textit{cause to die, indirect causation} \rangle$	*	*

Asymmetric OT (IP) fails to predict the full “division of pragmatic labor” whereby more marked forms are associated with more marked meanings. The constraints above yield a preferred interpretation of “cause to die” as involving canonical direct causation. Therefore, in the production competition with indirectly caused death as input meaning, “cause to die” is not even amongst the candidate outputs, and cannot be the winner. Presumably the winner would be some even more periphrastic alternative such as “indirectly cause to die”.

We can see the difference between the two cases, and how they are treated, graphically. Diagrams (i–v), below, show both production and interpretation relations. The first two diagrams represent direct applications of naive back-and-forth OT. the first illustrates standard partial blocking cases yielding marked meanings for marked forms such as “cutter” and “cause to die”. The second diagram represents the situation Wilson describes for Icelandic anaphora. The only difference is an extra arrow from the marked form to the marked meaning in the second diagram.

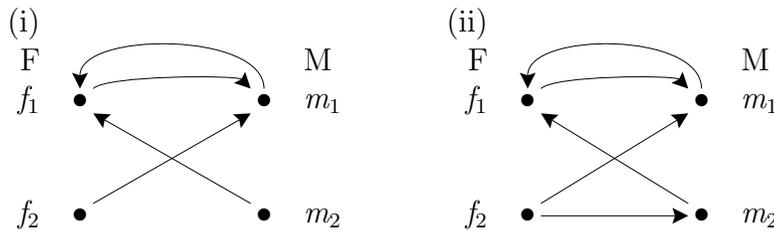
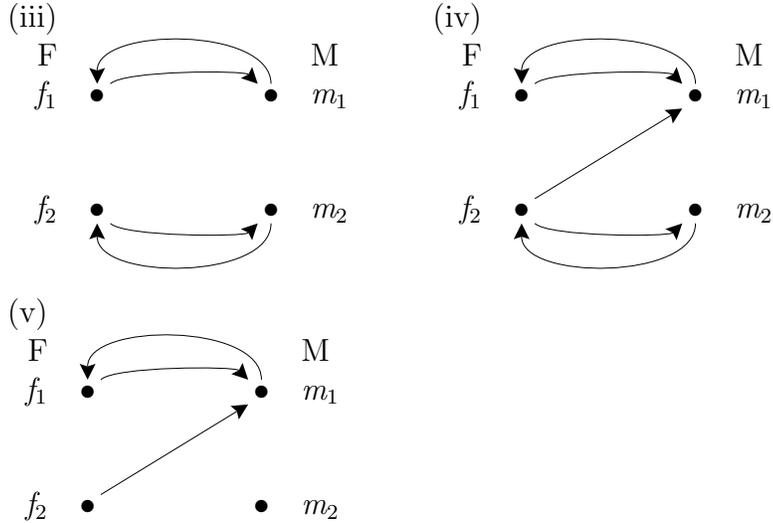


Diagram (iii) shows the results of applying Weak OT to either the situation in (i) or that in (ii): the marked form becomes uniquely associated with the marked meaning in both directions of optimization, while the unmarked form and unmarked meaning continue to be a bidirectionally optimal pair as they were in the original cases. Asymmetric OT (IP) does not achieve the harmonious situation depicted in (iii) for either of the situations given by (i) and (ii). What it does achieve is represented in (iv) and (v). Diagram (iv) shows the results of applying Asymmetric OT (IP) to the Icelandic anaphora case in (ii). here we see that the division of labor depicted in (iii) is almost achieved, except that there remains the possibility of interpreting the marked form as the unmarked meaning. This is a result of the fact that Wilson’s proposal does not innovate above naive back-and-forth OT as regards interpretation. When Asymmetric OT (IP) is applied to the classic “cause to die” situation in (i), what results is (v). Wilson’s system does not succeed in creating any link between the unmarked form and the unmarked meaning, so we can see that it does not provide a very general model of partial blocking. In these cases we might better describe what it does as “almost blocking”.



Asymmetric OT (IP) has an interesting range of strengths and weaknesses. We have just seen that it produces mixed results with respect to partial blocking. It does not help with ambiguity and optionality, since it does not provide new meanings for a form already contained in the set of winners in interpretation, or provide new ways to express a meaning that is already in the set of winners in interpretation. It also does not predict uninterpretability, since interpretation is naive. On the other hand, Wilson’s system can help with total blocking and freezing. Consider, for example, the two Korean double nominative forms X-NOM Y-NOM pred and Y-NOM X-NOM pred: both of these forms will be paired with the subject-object interpretation in the first, interpretation stage of optimization. So the pairs of these forms and the object-subject interpretation will not be included in the legitimate candidate set for the second, production optimization, and we derive the effect of freezing. Ineffability is predicted in some cases. Suppose a meaning is highly marked, such that no form is interpreted as having that meaning. In this case Asymmetric OT (IP) predicts that with this form as input, there will be no output (since there will be no candidates at all in the second stage of the production optimization). But it is not obvious whether this is sufficient to account, for example, for the ineffability of multiple questions in Italian.

6 Reproducibility as a constraint on interpretation

Zeevat (2001), like Wilson (2001), suggests using entirely different architectures for production and interpretation. What is striking is that Zeevat and Wilson choose precisely opposite architectures. Wilson keeps the standard unidirectional OT model of interpretation, but restricts the candidate set for production using the results of interpretation. Zeevat keeps the standard unidirectional OT model of production, but restricts the candidate set for interpretation using the results of production.

Zeevat bases his argument for what we will term Asymmetric OT(PI) in large part on two phenomena we have been discussing in this paper, ambiguity and ineffability. As regards ambiguity, we can gloss the idea as follows: since naive OT production has no problem with ambiguity, we should use the production

architecture as the basis of comprehension, and add further interpretational bells and whistles only as necessary.

In more detail, Zeevat's model starts by assuming that production uses a standard OT syntax set of constraints that we will term PROD. Comprehension is a more involved two-stage process involving both PROD and an additional set of constraints to select between alternative meanings: we will refer to this second set as PRAG. Zeevat's use of two distinct constraint sets for interpretation and production amounts to a significant difference from both Wilson's proposal and the other bidirectional architectures we have discussed, although Hendriks & de Hoop (2001) also advocate such a split.

The first stage of comprehension of a form \mathcal{F} consists in determining the set \mathcal{M} of meaning inputs which give \mathcal{F} as output using the constraints PROD. The second stage consists in using a standard OT semantics form-to-meaning optimization with the form \mathcal{F} as input, except that rather than using GEN to give candidate outputs, the set \mathcal{M} is used.

As is the case for Wilson's model, the form-meaning relation defined for production in Zeevat's proposal is different than that for comprehension. For Zeevat, the set of form-meaning pairs in comprehension is a subset of those in production. So a first observation on the proposal is that it predicts the existence of cases of *guaranteed misinterpretation*, that is, cases where a given meaning is expressed in a way that would be understood as having an interpretation other than the original meaning. Indeed, the proposal would seem to stand or fall on the existence of such cases, since without them the grounds for introducing a radical difference between production and comprehension are weak.

Zeevat does not cite any cases of guaranteed misinterpretation: the data he gives concerns the form-meaning relationship in the abstract, not differences between the form-meaning relationship provided by the production component of his system and the form-meaning relationship given by the comprehension system. In other words, his data involves form-meaning mismatches, like ambiguity and ineffability, not comprehension-production mismatches involving guaranteed misinterpretation.

Furthermore, while Zeevat describes the constraint set in PRAG, he does not describe PROD, so it is hard to be sure what the range of cases is where he predicts a mismatch between production and comprehension. None the less, we can exemplify the type of comprehension-production mismatch Zeevat predicts. The conjunction in (34a) involves two occurrences of an expression presupposing that there was a mosquito. A natural interpretation would involve only one mosquito, in which case the discourse might be continued with (34b), but it is also possible (if strained) to continue the discourse as in (34c), a two mosquito interpretation.

- (34) a. Hanjung realized that there was a mosquito and David realized that there was a mosquito.
- b. The mosquito was hungry.
- c. Both mosquitos were hungry.

Although we do not know what constraints are in PROD, we can speculate that (34a) might be generated in either the one mosquito or the two mosquito model.

However, Zeevat postulates a constraint *ACCOMMODATE in PRAG, a constraint which would prevent accommodation of presuppositions when the presuppositions are already satisfied in the discourse context. In this case, when the interpreter arrives at the second clause of (34a), a discourse referent for a mosquito has already been established, so there is no need to accommodate an extra mosquito in order to process the presupposition of the second clause. Thus Zeevat predicts that only the one mosquito interpretation should be available. So this may be a case where Zeevat predicts guaranteed misinterpretation. A speaker wanting to express that Hanjung and David have realized that separate mosquitos exist may optimally report this as in (34a), but in this case will be understood to mean that Hanjung and David have both developed existential knowledge about the same mosquito.

As regards (34a), the data is murky, since there is a slight awkwardness to the continuation in (34c). Our point is not to use this case to attack or defend Zeevat’s account, but rather to bring out more clearly the type of prediction that would provide a test for the proposed architecture. Detailed consideration of the predictions would have to wait until we know more about PROD.

As noted, ambiguity is one of the main motivations claimed for Asymmetric OT (PI): Zeevat analyzes the *rat/rad* (rat/wheel) problem at length. In interpretation, it is unproblematic for both the meanings *rat* and *wheel* to be selected in the first stage of comprehension (the reverse production stage), and there is no reason to expect PRAG to produce any preference between them, so ambiguity is predicted. However, there is an important class of examples for which Zeevat’s system incorrectly eliminates ambiguity. The problem is that PRAG includes a constraint STRENGTH which prefers logically stronger interpretations to weaker ones, so that Zeevat’s asymmetric model never predicts that one reading of an ambiguous sentence will entail another.

Consider (35a), which by virtue of a standard quantificational scope ambiguity has the two readings in (35b,c).

- (35) a. Every child liked one toy,
 b. $\forall x \text{child}(x) \rightarrow (\exists y \text{toy}(y) \wedge \text{liked}(x, y))$
 c. $\exists y \text{toy}(y) \wedge (\forall x \text{child}(x) \rightarrow \text{liked}(x, y))$

Here (35c) entails (35b), so Asymmetric OT (PI) incorrectly predicts that only the former is available. Furthermore, note that cases in which one reading entails another are common. Apart from scope ambiguities, this situation often arises when one meaning of a polysemous word has a strictly greater extension than another, as in “finger” (all digits on a hand, or all but the thumb), “gay” (homosexual, or homosexual male), and “New York” (the city or the state containing the city). Thus Asymmetric OT (PI) would predict that “There are rats in New York” can only mean that there are rats in New York City (and hence also in the state), while “There are no rats in New York” can only mean that there are no rats in the state (and hence none in the city either). We can conclude that while the architecture Zeevat proposes can successfully model ambiguity phenomena, the specific constraints he uses are problematic.¹⁹

Another claim of Zeevat’s is that Asymmetric OT (PI) successfully handles ineffability. However, we find that this claim is not yet fully substantiated. Note

that Zeevat’s claim is based on interpretation. But if ineffability consists in the existence of meanings which cannot be realized in production, then Zeevat’s model does not predict any ineffability, since from a production perspective, any meaning will give some winning form. So an Italian wanting to express the multiple wh-question ‘Who ate what?’ would be predicted to produce some utterance, and it is not obvious why this Italian would not imagine he or she had successfully expressed exactly what he or she intended.

Zeevat’s point, then, is more limited: in his system there may be no Italian form that would be understood as “Who ate what?” First, consider the infelicitous “Che ha mangiato che cuesta?” (“*Who ate which thing?*”). Zeevat would analyze this as uninterpretable because the PROD constraints prevent any meaning from being expressed that way. This seems reasonable. So we need to consider which string would be the output for the input $?xyate(x, y)$. Zeevat supposes this to be “Che ha mangiato qualcosa?” (literally, “*Who ate something?*”). The question is then why this string is not interpreted as $?xyate(x, y)$, but instead as $?x\exists yate(x, y)$.

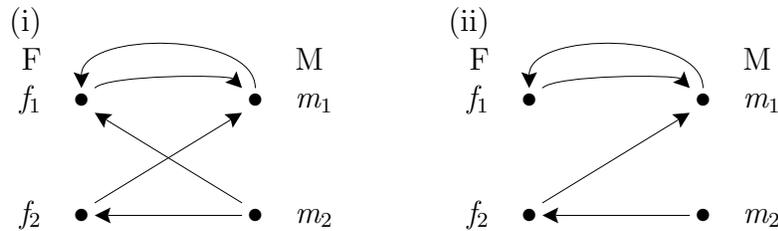
Given the premise that in the first stage of comprehension for the form “Che ha mangiato qualcosa?” both these meanings are found, selection between them is left to PRAG. Zeevat assumes that the crucial constraint will be one he terms *INVENT, that will disallow a mismatch between the numbers of question variables and existentials in the meaning and the numbers of corresponding expressions in the form.

How could *INVENT achieve such careful accounting of the differences between form and meaning? One possibility would be that *INVENT incorporated many or all of the constraints in PROD, but this would call into question the basic premise that PROD and PRAG are independent constraint sets with quite different functions. Zeevat (p.c.) has suggested instead that *INVENT is defined purely on meanings, not making any reference to forms. All it is supposed to do is prefer minimal meanings, e.g. in the sense of requiring less structure in a DRS. On this basis, *INVENT could prevent ‘Che ha mangiato qualcosa?’ from being interpreted as $?xyate(x, y)$, but only if there was some well-defined sense in which this meaning was less minimal than the alternative $?x\exists yate(x, y)$. We see no *a priori* reason why a meaning with two question variables should be less minimal than a meaning with one question variable and one existential variable, but this is perhaps not a major drawback of Zeevat’s proposal. What is clear is that, in principle, the architecture Zeevat advocates is capable of partially accounting for cases of ineffability like multiple questions in Italian. We say “partially” because, as pointed out above, Asymmetric OT (PI) fails to account for why speakers do not produce forms with the intention of expressing a multiple question: it can only account for why the forms they produce are misunderstood.

We have looked at the Asymmetric OT (PI) treatment of ambiguity and ineffability: what of uninterpretability, optionality, blocking and freezing? As with all the other accounts we have considered, Zeevat’s proposal has both strengths and weaknesses.

Regarding optionality, Asymmetric OT (PI) introduces no new insights above naive production OT: typically there will be a single winning form. Also, with regard to freezing, Zeevat’s model does not seem to provide a solution. In the case of Korean psychological verbs, for example, there is nothing to stop produc-

tion of both SOV and OSV word orders. With regard to partial blocking, and by analogy with Wilson’s system Zeevat’s proposal offers at best a partial solution. In particular, it is easily verified that Asymmetric OT (PI) makes incorrect predictions for both “cause to die” type examples and cases with the same structure as found with Icelandic anaphora. On the other hand, we can easily identify the abstract structure of two cases for which Zeevat’s system would successfully isolate two form-meaning pairs from each other. Diagrams (i) and (ii) show naive back-and-forth OT structures which under Asymmetric OT (PI) would yield two bidirectional links, one between f_1 and m_1 , and the other between f_2 and m_2 . Identifying linguistic phenomena to which these two diagrams correspond might provide further insight into the significance of Zeevat’s model, but we leave this task to future research.



Let us briefly consider the option of treating partial blocking and freezing by combining Zeevat’s model with Blutner’s Strong or Weak optimality. For example, we might define Strong Asymmetric OT (PI) as having the form-meaning relationship defined by the intersection of Zeevat’s production and comprehension mechanisms. However, we already noted that the set of form-meaning pairs in Zeevat’s production model is a superset of the form-meaning pairs in his comprehension model. So taking the intersection of the two would amount to using the comprehension model for both comprehension and production. Given the philosophical position taken in Zeevat (2001), and the many arguments he gives for an asymmetry between comprehension and production, a move to Strong Asymmetric OT (PI) would amount to something of a retreat, even if the result successfully modeled freezing. Still, we think it worth noting the possibility of such a model, as one of many directions that Zeevat’s model may be extended, and one of many possibilities in the space of bidirectional OT architectures.

Where Asymmetric OT (PI) certainly does have something to offer is with respect to uninterpretability and total blocking. Regarding uninterpretability, observe that since the first stage of interpretation is identical to naive production, there will in general be many strings which are not produced for any meaning input. All these strings are uninterpretable in Asymmetric OT(PI). For example, if “Colorless green ideas sleep furiously” is the form, we would first consider the set of meanings that would generate it. If we allow Chomsky’s premise that the string is meaningless, we would find no such meaning, and hence the model correctly predicts uninterpretability (due to complete absence of any candidates in the final stage of the interpretation competition).

Last, we consider total blocking. It is easy to see that Asymmetric OT(PI) can model this phenomenon, the analysis being parallel to that of uninterpretability. Consider a standard case:

So we accept that there is room for disagreement about how significant each of the seven phenomena is. Yet we also believe that a strong argument could be made for not restricting ourselves to grammar architectures that makes description of these phenomena impossible. The table above shows that bidirectional OT architectures from the literature are too restrictive: there are many patterns of relation between form and meaning which they cannot describe effectively, regardless of the particular constraints that are used and the ranking between those constraints. Even the account which (narrowly) fares best by our criteria, Strong OT, fails to contribute to our understanding of three of the seven phenomena, ambiguity, optionality, and partial blocking.

In this paper we have not attempted to present an approach which betters existing proposals. However, there is no shortage of directions in which these existing proposals could be developed. Consider first partial blocking. Only one of the proposals discussed, Weak OT, deals with the classic cases of partial blocking described in section 2. Yet Weak OT suffers from severe problems, most notably considerable over-generation. Could a variant of Weak OT maintain the analysis of partial blocking without this leading to such great over-generation? One possibility to consider is the variant of Weak OT discussed by Beaver (to appear). This variant system performs only one iteration of the Weak OT process, pruning once and grafting once. As a result it maintains some of the properties of Weak OT, but lacks Weak OT's "everyone's a winner" profligacy.

There are also several approaches that could be combined with the proposals discussed here so as to account for optionality and ambiguity. Partial ranking of constraints Anttila & Fong (2001), and stochastic ranking of constraints Boersma & Hayes (2001); Asudeh (2001); Bresnan & Deo (2001) are techniques that allow multiple winners to appear in competitions that might only produce a single winner using linear constraint ranking. Another issue that is very relevant to ambiguity and optionality is the role played by context. For example, so-called *optionality* of Korean transitive word order can also be seen as context-dependence of Korean word-order: in specific discourse contexts where one argument is more prominent than the other, there may be no word-order freedom at all. So it is natural to move from simple form-meaning or meaning-form optimization to optimizations that include three parameters: form, meaning and context. This is exactly what Blutner (2001) proposes, although his main use of context involves presupposition resolution rather than ambiguity resolution or what we might analogously term *optionality resolution* — the context-dependent choice of a particular form from amongst a range of possibilities.

We have shown that existing bidirectional OT systems suffer from serious problems in their treatment of form-meaning asymmetries. But our paper is intended in a constructive spirit. We have laid out a set of issues which we hope developers of bidirectional approaches will tackle in future research.

Notes

¹For a recent discussion of the many aspects of ambiguity and why they constitute a puzzle for linguistics, see Wasow et al. (to appear).

²An occupational therapy web-site (www.otworks.com) reports that: “O.T. stands for... Occupational Therapy, or Over Time Ol’ Timer Original Thinkers Overly Timid Old Testament Over Taxed E.T’s sibling.” O.T. is also used to mean “Off Topic”.

³See Müller (1999) for an overview of approaches to optionality within the standard OT framework whose constraint set forms a total order.

⁴English complementizer drop has been analyzed within OT by Grimshaw (1997) and Baković & Keer (2001).

⁵Note that the meaning de Hoop (2001) gives to the term *unintelligibility* seems, from her examples, to be distinct from our notion of *uninterpretability*. The examples de Hoop considers involve utterances which have (only) a contradictory interpretation, whereas we consider cases in which one cannot determine any proposition expressed by the utterance.

⁶We choose “last” in the diagram as an arbitrary highly unmarked adjective, at least in terms of having higher frequency than any other adjective in the British National Corpus. If this can be taken to indicate that the meaning is less marked than other adjectival meanings, then OT grammars might be expected to interpret “dolomphious” as having the same meaning as “last”.

⁷Some discussion of different options of combining two optimization perspectives and the general consequences for the resulting bidirectional models can be found in Kuhn (2001).

⁸See also Kuhn (2001) and Vogel (this volume).

⁹In this way, unidirectional production OT can produce apparent optionality, based on different inputs. This approach to optionality, which Müller (1999) terms “the pseudo-optionality approach”, predicts cases of optionality that correlate with differences in information status but does not produce multiple outputs for the same input.

¹⁰These problems of Weak OT are also discussed by Gärtner, in this volume.

¹¹Though information about argument-function mappings is represented as part of “meanings” in (13), we do not assume that this information is part of OT input. Rather association of the arguments in the input to a particular grammatical function results from constraint interaction.

¹²Some discussion of a typology of ineffabilities can be found in Fanselow and Féry (2002).

¹³Yet another option which we consider as a candidate for realizing the multiple wh-question in Italian is the ellided form “Che ha mangiato?” (“Who has eaten?”). The elliptical form, however, would express the multiple wh-question only at the cost of violating a faithfulness constraint PARSE, which requires input elements to have an overt correspondent in the output (Grimshaw and Samek-Lodocivi (1998)). Clearly, in reality, this cost is too high.

¹⁴MARK-FAM requires that familiar objects are realized as definites. This is a counterpart to the constraint FAM-DEF in Beaver (to appear) which requires that definites should be familiar. Note that MARK-FAM can penalize indefinites, whereas FAM-DEF can only penalize definites.

¹⁵Our understanding of Wilson’s model is considerably influenced by recent unpublished work of Judith Aissen.

¹⁶Vogel (this volume) develops a bidirectional OT model in which production-based optimization is accompanied by a second step that checks the recoverability of an underlying form. We defer discussion of this model to a later occasion.

¹⁷For example, *sig* is unmarked for gender and number, and *hann* is a masculine and singular form.

¹⁸Wilson (2001) makes two assumptions regarding representations of inputs and candidate structures in his analysis. The first is that the input for interpretation and production is the same and only the candidate set varies. More specifically, for both optimizations, he assumes a highly abstract input consisting of a surface string plus an abstract syntactic structure (i.e., LF) and a semantic representation. In interpretation, the morphosyntactic component of the input is held fixed across the candidate set; in production, the semantic component is fixed. Second, Wilson assumes that binding relations are specified in the input semantic representation and that in interpretation candidates may diverge from the input with respect to binding relations. Relativized minimality in interpretation and referential economy in production then are both accounted for in terms of faithfulness violations. In this discussion, we abstract away from details of representational assumptions that Wilson makes and continue to assume that for interpretation, the input is a form and the output is a meaning, and for production, the input is a meaning and the output is a form. As far as we can tell, this does not affect the overall results of Asymmetric OT (IP).

¹⁹Observe that if Asymmetric OT (PI) can model ambiguity, one might expect by symmetry considerations that Asymmetric OT (IP) would model optionality. But here the fact that Zeevat uses two distinct constraint sets while Wilson uses only one comes into play. It is because Zeevat proposes that the set of interpretation constraints is very limited that his system can model ambiguity. by contrast, Wilson uses the full constraint set in the second phase of production, and this will typically weed out all but one candidate. An architecture like that of Asymmetric OT (IP) would model optionality provided it used only a very limited constraint set in the second stage of production, and kept the bulk of constraints for the first stage of production and for interpretation.

References

- Aissen, J. (1999). Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory*, 17, 673–711.
- Anttila, A. & Fong, V. (2001). The partitive constraint in Optimality Theory. *Journal of Semantics*, 17, 281–314.

- Asudeh, A. (2001). Linking, optionality and ambiguity in Marathi: An Optimality Theory analysis. In P. Sells (Ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, (pp. 257–312). Stanford: CSLI Publications.
- Baković, E. & Keer, E. (2001). Optionality and ineffability. In G. Legendre, J. Grimshaw, & S. Vikner (Eds.), *Optimality-Theoretic Syntax*, (pp. 97–112). Cambridge: MIT Press.
- Beaver, D. I. (to appear). The optimization of discourse anaphora. *Linguistics and Philosophy*.
- Blutner, R. (2001). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17, 189–216.
- Blutner, R. & Jäger, G. (1999). Competition and interpretation: The German adverbs of repetition. Available at <http://www2.hu-berlin.de/asg/blutner/>.
- Boersma, P. (1998). *Functional Phonology*. The Hague: Holland Academic Graphics.
- Boersma, P. & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32, 45–86.
- Bresnan, J. (2001). Explaining morphosyntactic competition. In M. Baltin & C. Collins (Eds.), *Handbook of Contemporary Syntactic Theory*, (pp. 11–44). Oxford: Blackwell.
- Bresnan, J. & Deo, A. (2001). Grammatical constraints on variation: ‘Be’ in the *survey of english dialects* and (stochastic) Optimality Theory. MS., Stanford University. Available online: <http://www-lfg.stanford.edu/bresnan/download.html>.
- Choi, H.-W. (1999). *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford: CSLI Publications.
- Dalrymple, M. E. (1993). *The Syntax of Anaphoric Binding*. Stanford: CSLI Publications.
- de Hoop, H. (2001). Making sense: The problem of unintelligibility. In *GAGL 44 (Making Sense: from lexeme to discourse)*, (pp. 187–194). Department of Linguistics, University of Groningen.
- Fanslow, G. & Féry, C. (to appear). Ineffability in grammar. *Linguistische Berichte*, special issue: Resolving Conflicts in Grammars: Optimality Theory in Syntax, Morphology and Phonology.
- Gärtner, H.-M. (this volume). On the OT-status of “unambiguous encoding”.
- Grimshaw, J. (1997). Projection, heads, and optimality. *Linguistic Inquiry*, 28, 373–422.

- Grimshaw, J. & Samek-Lodovici, V. (1998). Optimal subjects and subject universals. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, (pp. 193–219). Cambridge: MIT Press.
- Hale, M. & Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, 29, 567–615.
- Hendriks, P. & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy*, 24, 1–32.
- Horn, L. (1984). Toward a new taxonomy of pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, Form, and Use in Context*, (pp. 11–42). Washington, DC: Georgetown University Press.
- Huang, Y. (2000). *Anaphora — a cross-linguistic study*. Oxford University Press.
- Jäger, G. (2002). Some notes on the formal properties of bidirectional optimality theory. *Journal of Logic, Language and Information*, 11, 427–451.
- Jäger, G. (this volume). Learning constraint sub-hierarchies. the bidirectional gradual learning algorithm.
- Jäger, G. & Blutner, R. (2000). Against lexical decomposition in syntax. In A. Z. Wyner (Ed.), *The Israeli Association for Theoretical Linguistics. The Proceedings of the Fifteenth Annual Conference*, (pp. 113–137).
- Kiparsky, P. (1983). Word-formation and the lexicon. In F. Ingeman (Ed.), *Proceedings of the 1982 Mid-America Linguistic Conference*.
- Kuhn, J. (2001). *Formal and Computational Aspects of Optimality-theoretic Syntax*. Ph.D. thesis, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Lee, H. (2001). *Optimization in Argument Expression and Interpretation: A Unified Approach*. Ph.D. thesis, Stanford University.
- Lee, H. (to appear). Prominence mismatch and markedness reduction in word order. *Natural Language and Linguistic Theory*.
- Legendre, G., Smolensky, P., & Wilson, C. (1998). When is less more?: Faithfulness and minimal links in *wh*-chains. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is Best Good Enough? Optimality and Competition in Syntax*, (pp. 249–289). Cambridge: MIT Press.
- Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conventional Implicature*. Cambridge: MIT Press.
- Maling, J. (1984). Non-clause-bounded reflexives in Modern Icelandic. *Linguistics and Philosophy*, 7, 211–241.
- McCarthy, J. & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In M. González (Ed.), *Proceedings of NELS 24*, (pp. 333–379). Amherst: GLSA, University of Massachusetts.

- McCawley, J. (1978). Conversational implicature and the lexicon. In P. Cole (Ed.), *Syntax and Semantics 9: Pragmatics*, (pp. 245–259). New York: Academic Press.
- Müller, G. (1999). Optionality in Optimality-Theoretic syntax. *Glott International*, 4, 3–8.
- Orgun, C. O. & Sprouse, R. L. (1999). From MPARSE to CONTROL: Deriving ungrammaticality. *Phonology*, 16, 191–224.
- Poser, W. (1992). Blocking of phrasal constructions by lexical items. In I. Sag & A. Szabolcsi (Eds.), *Lexical Matters*, (pp. 111–130). Stanford: CSLI Publications.
- Prince, A. & Smolensky, P. (1993). Optimality Theory: Constraint Interaction in Generative Grammar. Technical Report RuCCS Technical Report #2, Center for Cognitive Science, Rutgers University, Piscataway.
- Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 27, 720–731.
- Smolensky, P. (1998). Why syntax is different (but not really): Ineffability, violability and recoverability in syntax and phonology. Handout of the talk given at the Stanford/CSLI Workshop: Is Syntax Different? Common Cognitive Structures for Syntax and Phonology in Optimality Theory. Stanford University, December 12–13, 1998.
- Vogel, R. (this volume). Remarks on the architecture of OT syntax grammars.
- Wasow, T., Perfors, A., & Beaver, D. I. (to appear). The puzzle of ambiguity. In P. Sells (Ed.), *Essays in Honor of Steve Lapointe*. Stanford, CA: CSLI Publications.
- Wilson, C. (2001). Bidirectional optimization and the theory of anaphora. In G. Legendre, J. Grimshaw, & S. Vikner (Eds.), *Optimality-Theoretic Syntax*, (pp. 465–507). Cambridge: MIT Press.
- Zeevat, H. (2001). The asymmetry of Optimality Theoretic syntax and semantics. *Journal of Semantics*, 17, 243–262.