The direction of optimization

In this chapter, some variations of the formal setup defined in chapter 4 are discussed. So far in this book, optimization has been used to find the most harmonic way of expressing some meaning, but we can also use optimization to find the most harmonic interpretation of a given utterance (i.e., a string of words, as we are focusing on the field of syntax). If we call the standard optimization we have looked at so far production-based or expressive optimization, we now get comprehension-based or interpretive optimization.

To illustrate the character of the two difference optimizations, let us look at a very simple example. We assume a $G_{\text{niviol}}$ grammar which leaves open the serialization of arguments (say, of a transitive verb). For simplicity, the verb position is assumed to be fixed: all candidates produced by $G_{\text{niviol}}$ have the verb in final position. So for a transitive verb, $G_{\text{niviol}}$ will produce the orders subject-object-verb or object-subject-verb. Both variants can occur in embedded clauses in German:

(146) (weil) Anna den Film kannte
(because) A. the movie (ACC) knew
‘... because Anna had seen the movie’

(147) (weil) ihn der Geruch störte
(because) him the smell (NOM) bothered
‘... because he didn’t like the smell’

Now let us assume two constraints (for less ad-hoc constraints see the discussion in sec. 5.1.1 below): $\text{SUBJPRECOBJ}—$The subject precedes the object$, and $\text{PRONPRECFULLNP}—$A pronoun precedes a full NP$. In
The direction of optimization

standard production-based/expressive optimization, we can use these constraints to determine the most harmonic realization of an underlying meaning like in (148). (In English, the underlying meaning would be realized as Peter saw her.) The meaning, encoded as an f-structure, determines the candidate set; formally, it is our input. For both of the two orderings of the argument phrases we get a candidate:

\[(148)\]

\[\text{Input F-Structure:} \]

| PRED | 'sehen(x, y)' |
| SUBJ | {PRED 'Peter'} |
| OBJ  | {PRED 'PRO', GEND FEM} |

With the given constraint ranking, the subject-object-verb candidate wins.

In comprehension-based/interpretive optimization, which this chapter adds to the discussion (compare Smolensky (1996)), all candidates have the surface word string in common, but not necessarily the meaning (f-structure). Let us look at the string

\[(149)\] (weil) sie Peter sah

(‘because’ she/her P saw

‘. . . because she saw Peter’

[or in principle: ‘. . . because Peter saw her’]

The two NPs are both ambiguous between nominative and accusative, so each can either be the subject or the object of the sentence. This is reflected in the candidate set, which includes candidates with different f-structures.
The fact that candidate (150a) is the winner tells us that the most harmonic analysis for the given string involves the meaning which has *sie* (‘she’) as the subject and *Peter* as the object. Note that this is different from the meaning underlying all candidates in tableau (148). Candidate (150a) was not in the candidate set of the production-based/expressive optimization (148), only (148b)/(150b) is a shared candidate.

The interpretation in (150a) is indeed the strongly preferred reading for (149). This indicates that it may be possible to capture generalizations by using the same set of constraints in the two types of optimizations.

This may suffice as an initial illustration. Differences and commonalities between the two optimizations, and possible ways of combining them are discussed in the remainder of this chapter. Sec. 5.1 specifies comprehension-based/interpretive optimization in the declarative formal framework introduced in chapter 4 and reviews the motivation for assuming a close formal similarity of the two optimizations; sec. 5.2 discusses some potential counterarguments against the close similarity. Sec. 5.3 discusses various ways of combining the two optimizations in a bidirectional optimization model.
The direction of optimization

5.1 Varying the input to optimization

The declarative character of the OT formalization proposed in chapter 4 makes it very easy to characterize the two types of optimization as variants of the same abstract mechanism.

In the definition of candidate generation \( Gen_{\text{Gen}} \) (94) (repeated below), which was discussed in the previous chapter, the argument that \( Gen_{\text{Gen}} \) takes is a partial LFG analysis (a partial f-structure), which is used as the common part of all LFG analyses in the candidate set.

\[
\text{(94) Definition of Gen} \quad \text{For } \Phi_{\text{in}} \text{ a partial f-structure: } \\
Gen_{\text{Gen}}(\Phi_{\text{in}}) = \{ \langle T, \Phi' \rangle \in L(G_{\text{involt}}) \mid \Phi_{\text{in}} \subseteq \Phi' \}, \text{where } \Phi' \text{ contains no more semantic information than } \Phi_{\text{in}} \}
\]

A formally very similar function can be defined to create the candidate set required for comprehension-based or interpretive optimization. Rather than using the (partial) input f-structure as the common substructure across candidates, we can use a part of c-structure: the string of terminal symbols (i.e., words). The remainder of c-structure and f-structure will vary freely across candidates.

\[
\text{(151) Extension of the definition of Gen} \quad \text{For } w \text{ a word string: } \\
Gen_{\text{Gen}}(w) = \{ \langle T, \Phi \rangle \in L(G_{\text{involt}}) \mid w \text{ is the terminal string/yield of } T \}
\]

Beside these two criteria of specifying the candidate set ((94) and (151)), all kinds of other criteria for specifying the candidate set are conceivable from a purely formal point of view. For example, one might keep some higher-level part of c-structure trees fixed, while the c-structure terminal symbols and f-structure varies freely. However, in sec. 3.3 the observation was made that in order to ensure learnability, the input has to be deducible from the utterance or utterance context (formulated as restriction (49)).

\[
\text{(49) Restriction on the character of the input} \quad \text{All aspects of the OT input must be such that they can be in principle inferred from world knowledge and the general context of utterance.}
\]

128
5.1 Varying the input to optimization

For production-based/expressive optimization, the consequence was that only semantically and pragmatically relevant information could be included in the input (and that at the level of interpretation, all candidates had to be faithful to this information). Now, when we consider a more surface-oriented input, the learnability consideration forces us to assume a fully surface-detectable concept of the input. So we should assume the string of words without any fixed phrase structure representation. Definition (151) is compatible with this restriction.

Inserting the modified definition (151) of the candidate generation function \( Gen \) into the context of the full optimization scheme gives us the diagram in (152)—a modification of (145).

(152) Comprehension-based/interpretive optimization in OT-LFG

Note that the overall declarative system of candidate comparison remains unaffected when the input is changed as just argued. This gives us a perspicuous high-level model of the sound-meaning relation, abstracting away from all processing details. Since the constraint set and
The direction of optimization

the candidate evaluation mechanism is identical for expressive and interpretive optimization, empirical predictions of the two formal devices are also directly linked.

It is worthwhile pointing out that the straightforward reversibility of the architecture is a consequence of the strictly representational nature of the candidate analyses, which was argued for in sec. 3.3.5. The candidate status and the constraint profile of a particular candidate analysis is independent of its “use” in a candidate set; this was illustrated by candidate (148b)/(150b) which showed up in a candidate set for expressive optimization (along with other candidates sharing its meaning representation) and also in a candidate set for interpretive optimization (with other candidates sharing its surface string). If the candidate-internal relation between meaning and form were modelled as a derivational sequence of structural transformations, a reversal of the direction of candidate generation would involve a reverse application of the individual transformational steps. This does not necessarily affect the two optimizations: the proposed reversal remains feasible if declarative OT constraints are applied on a single (the “final”) structure that includes a record of the entire derivational history, in the form of traces or chains. (Note however that in this case the question arises why the definition of candidates is still stated in derivational terms and not declaratively too; compare the discussion of (62) in sec. 3.3.5.) If on the other hand application of the OT constraints is sensitive to the derivational order (for instance, through stepwise filtering of the candidate set as in a “local optimization” approach, e.g. Heck and Müller (2000)), a reversal of the derivational order may change the candidate status of a particular analysis: in one direction the candidate may survive until the final optimization step, while in the other direction it is filtered out at the very beginning. This may have a snowball effect on other candidates, for which the candidate would have been a competitor later on. So, a controlled reversibility of optimization seems to be guaranteed only if we adopt declarative constraints that apply simultaneously (or “globally”) on comprehensive candidate representations.

Comprehension-based optimization has been proposed in various contexts in the literature. For example, Hendriks and de Hoop (2001) use such a comprehension-based optimization model in what they call OT semantics; the winning structure models what native speakers conceive as the correct interpretation in the given context. Comprehension-based optimization is also being applied as a preference mechanism in the large-scale LFG grammars developed in the ParGram project (Kuhn and Rohrer (1997); Frank et al. (1998), 2001).
5.1 Varying the input to optimization

Comprehension-based optimization also plays a role in learning. Tesar and Smolensky (1998, 2000) assume it as robust interpretive parsing. Gibson and Broihier (1998) explore to what degree such an optimization model can be used to derive disambiguation strategies in human sentence processing. 99

Smolensky (1996) proposes to explain the lag of children’s production abilities behind their ability in comprehension without having to assume a special grammar or special processing devices. The lag is predicted by an OT system if one assumes that in comprehension a simple interpretive optimization along the lines of (152) is performed, which permits processing the strings that the child hears with the same constraint ranking that is applied in production. Thus in comprehension, many analyses are accepted that are not grammatical under the child’s current constraint ranking (according to the GenN-based definition of grammaticality). The simple parsing task is liberal enough not to filter them out. However in production, the common underlying structure does determine the candidate set, and the constraints will have a strong filtering effect. The result is a very reduced production ability for the initial constraint ranking.

In the following, I will address the question whether there are any empirical phenomena in adult language indicating that the constraint set used in production-based optimization has also an explanatory impact on comprehension-based optimization. This is done by applying a constraint set motivated independently for a production-based optimization account to the disambiguation task. Indeed, the disambiguation preferences observed for native speakers follow. The presentation follows (Kuhn, 2001c, 4.2). A similar analysis of word-order freezing was proposed independently by Lee (2000, 2001a,b), who goes into much more empirical detail than I do here. Here, the analysis is just used to make a formal point about the OT-LFG architecture.

5.1.1 Word order freezing

With the formalization of OT-LFG, we are in a position to address the question whether constraint systems with production-based competi-

99They use special constraints modelling the Minimal attachment and Late closure strategy in an OT framework and come to the conclusion that a strict ranking of constraints is inadequate for this task. Note however that the Minimal attachment and Late closure constraints are distinct from the constraints one would assume in production-based/expressive optimization models of grammaticality. Fanselow et al. (1999)) argue that human sentence processing facts can be derived in an OT framework, based on an incremental application of the standard grammaticality-defining constraint set.
The direction of optimization

tion are necessarily distinct or even incompatible with systems designed for a comprehension-based competition. As was indicated in the introduction to this chapter, there are phenomena where using the same constraints in both directions makes the correct empirical predictions for preference among readings.

Here I will continue to interpret optimality in comprehension-based/interpretive competition as preference of the respective reading of the input string. So we hope to find empirical cases where the comprehension-based winner (based on a general constraint set) coincides with the intuitively preferred reading. It is best to look at an empirical domain which involves a fair amount of realization alternatives and ambiguity. Therefore, the relatively free word order in German is a good example. In this section, I will go beyond the somewhat ad-hoc constraints used in (148) and (150); I will show that the production-based/expressive optimization account proposed by Choi (1999) extends straightforwardly to the comprehension-based/interpretive application, making correct preference predictions.

In the German Mittelfeld (the region between the finite verb in verb second position and the clause-final verb position), nominal arguments of the verb can appear in any order. However, as has been widely observed (cf., e.g., Lenerz (1977), Höhle (1982), Abraham (1986), Uszkoreit (1987)), a certain “canonical” order is less marked than others (cf. also Kuhn (1995)). Deviations from this canonical order are used to mark a special information structure (or topic-focus structure), i.e., these non-canonical orderings are more restricted through context. Sentence (153) reflects the neutral order as it would be uttered in an out-of-the-blue context. Variant (154a) will be used to mark dem Spion as the focus; (154b) furthermore marks den Brief as the topic.

(153) dass der Kurier dem Spion den Brief zustecken
that the courier (Nom) the spy (Dat) the letter (Acc) slip sollte
should

(154) a. dass der Kurier den Brief dem Spion zustecken sollte
b. dass den Brief der Kurier dem Spion zustecken sollte

A different interpretation is the recoverability interpretation: while the preference interpretation assumes that all candidates have undergone production-based/expressive optimization, the recoverability interpretation uses interpretive optimization as the initial filter. A later expressive optimization is performed on the output of this filter (compare Jäger (2002a), and sec. 5.3.1).
5.1 Varying the input to optimization

(Choi, 1999, 150) models these data assuming competing sets of constraints on word order: the canonical constraints, based on a hierarchy of grammatical functions (and, in principle also a hierarchy of thematic roles) (155); and information structuring constraints (distinguishing the contextual dimensions of novelty and prominence, each marked by a binary feature) (156).\(^{101}\)

(155) **CANON** (Choi, 1999, 150)

a. **CN1:**
   SUBJ should be structurally more prominent than (e.g. ‘c-command’) non-SUBJ functions.

b. **CN2:**
   non-SUBJ functions align reversely with the c-structure according to the functional hierarchy.
   (SUBJ > D.OBJ > I.OBJ > OBL > ADJUNCT)

(156) **Information Structuring Constraints:** (Choi, 1999, 150)

a. **NEW:**

b. **PROM:**

Based on an appropriate ranking of these constraints (PROM \(\gg\) CN1 \(\gg\) {NEW, CN2}), Choi can predict the optimal ordering for a given underspecified f-structure (which in this case will also contain a description of the informational status of the verb arguments). When the arguments do not differ in informational status—e.g., everything is new but nothing is prominent—the canonical constraints will take effect, leading to the order in (153), as illustrated in the tableau (157), for a few sample candidates.

\(^{101}\)Following more recent work (in particular by Aissen (1999)) on the OT treatment of alignment across hierarchies, the mechanism of Harmonic alignment could be used to formulate these constraints more precisely.
The direction of optimization

(157) Input F-Structure:

<table>
<thead>
<tr>
<th>PRED</th>
<th>'sollen'(x, y)</th>
<th>PRED</th>
<th>'Kurier'</th>
<th>PROM</th>
<th>NEW</th>
<th>SUBJ</th>
<th>'zustecken'(x, u,v)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJ</td>
<td>'Brief'</td>
<td>OBJ_</td>
<td>'Spion'</td>
<td>PROM</td>
<td>NEW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBJ_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a.  *der Kurier dem Spion den Brief zustecken sollte*

b.  der Kurier den Brief dem Spion zustecken sollte

c.  ☞ den Brief der Kurier dem Spion zustecken sollte

When there are differences in informational status, the unmarked order will however violate information structuring constraints, such that competitors with a different ordering can win out:

(158) Input F-Structure:

<table>
<thead>
<tr>
<th>PRED</th>
<th>'sollen'(x, y)</th>
<th>PRED</th>
<th>'Kurier'</th>
<th>PROM</th>
<th>NEW</th>
<th>SUBJ</th>
<th>'zustecken'(x, u,v)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJ</td>
<td>'Brief'</td>
<td>OBJ_</td>
<td>'Spion'</td>
<td>PROM</td>
<td>NEW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBJ_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a.  *der Kurier dem Spion den Brief zustecken sollte*

b.  *der Kurier den Brief dem Spion zustecken sollte*

c.  ☞ *den Brief der Kurier dem Spion zustecken sollte*

Like the Grimshaw/Bresnan fragment assumed in most examples so far, Choi's assumptions about Gen can be formulated as an LFG grammar \( G_{\text{invio}} \). The tableaux above showed expressive optimization. What are the predictions if the same constraint set is used in interpretive optimization?

For sentence (153) and its ordering variants, an application of
5.1 Varying the input to optimization

comprehension-based/interpretive optimization does not produce any interesting results, since in parsing the NPs can be unambiguously mapped to argument positions. However, if we look at sentences with ambiguous case marking like (159) and (160), the situation changes.

(159) dass Hans Maria den Brief zustecken sollte
that H. (NOM/DAT/ACC) M. (NOM/DAT/ACC) the letter (ACC)

(160) dass Otto Maria Hans vorschlagen sollte
that O. (NOM/DAT/ACC) M. (N/D/A) H. (N/D/A) suggest

Parsing (159) with the appropriate $G_{invot}$-grammar will result in two classes of analyses: one with Hans as the subject, and Maria as the indirect object, and one with the opposite distribution. The latter reading is strongly preferred by speakers of German (i.e., we observe a “freezing effect”). Note that there is no way of avoiding this ambiguity with hard constraints. For (160), even more readings become possible: any of the three NPs can fill any of the three available argument positions. Nevertheless, speakers clearly prefer one reading. If we apply interpretive optimization, Choi’s original constraints will predict exactly these observations: In the comprehension-based/interpretive optimization the string is fixed for all competing candidates; in addition to the string, we have to assume some representation of the context which clarifies the informational status of the referents that the new sentence is about.\footnote{Some discussion of the character of such a context representation will follow in sec. 5.3.4. The representation proposed there does not immediately accommodate Choi’s representation of context, but it is not difficult to reformulate the underlying grammar and constraints in a way that has the intended effect. I do not go into these details here, since I only want to make a point about the general architecture of optimization.} If this informational status is neutral for all referents (e.g., all are known—\emph{not new}—and not \emph{prominent}), the analysis which violates the fewest constraints will be the one which interprets the arguments in such a way that the observed order is in line with the canonical order. For (160), we will get the following competition:
The direction of optimization

Thus, for the constraints that Choi (1999) assumes, the standard OT view of expressive optimization can be generalized to the interpretive scenario, giving rise to predictions of preferred readings that are clearly met by native speakers’ intuitions.

The present discussion leaves open how exactly the interpretive optimization and the expressive optimization are combined. The application of both optimizations is called bidirectional optimization. I will discuss the possibilities for spelling out the combination in a rigorous way in sec. 5.3. Before this, I will address some issues about the character of comprehension-based/interpretive optimization, posed by the parallelism suggested by the formal treatment.

136
5.2 The character of comprehension-based optimization

The character of comprehension-based optimization

The formal similarity of the two optimization concepts and the prospect of deriving intuitive concepts like grammaticality and preference are intriguing. However, when the idea of comprehension-based/interpretive optimization for disambiguation is applied to a non-trivial set of data, a serious issue arises, challenging the parallelism between the two optimizations.

Although the word-order freezing effect observed in sec. 5.1.1 does occur in the absence of contextual or world-knowledge clues, it can be easily overridden by such non-syntactic information. For example, take (162), which like (159) and (160) contains ambiguous case marking.

(162) dass diese Oper Mozart komponiert hat

Here the selectional restrictions of the verb *komponieren* 'compose'—in combination with knowledge about the argument phrases *Mozart* and *diese Oper* 'this opera'—clearly overrule the ordering preferences. The absurd reading of the opera composing Mozart does not occur to a speaker of German (neither does the sentence sound odd), even if the linguistic context of the sentence is neutral. But the straightforward comprehension-based optimization account predicts the sentence to have only the odd reading.\(^{103}\)

As an immediate reaction, one might try to augment the set of constraints by additional constraints taking these extra-syntactic dependencies into account.\(^{104}\) Rather than developing such an account in detail, I will address some general issues raised by this move.

\(^{103}\) A similar influence of context and world knowledge in comprehension-based optimization is observed by Lee (2001a).

\(^{104}\) Müller (1998) proposes an OT account for deriving markedness judgements regarding certain word order variants in German. Working with the ordinary production-based optimization, he assumes a *subhierarchy* of constraints which is used only for determining markedness, while the matrix hierarchy is used for grammaticality. The constraints in the subhierarchy are in part based on extra-syntactic concepts like definiteness, animacy, and focus; however, no explicit reference to the actual context of an utterance is made. So the empirical predictions can only be checked against intuitions about gradual differences in markedness of isolated sentences. Since to some degree such intuitions depend on the informant’s ability to make up plausible contexts, they have to be regarded with caution. The model that I will argue for in sec. 5.2.2 and demonstrate in sec. 5.3.4, assumes an explicit formal representation of context as part of the input and thus circumvents the vagueness problem in the validation of empirical predictions.
The direction of optimization

One general concern which I will only mention in passing is that clearly the boundaries of what is traditionally considered the scope of syntactic theory are passed when such effects of discourse context and world knowledge are incorporated into a single optimization task. Per se, this is neither completely new (since “pragmatic effects” on syntax have frequently been observed), nor necessarily a bad thing. Ultimately, the parallelism assumption of OT leads us to expect such a global cognitive optimization (“Parallelism: all constraints pertaining to some type of structure interact in a single hierarchy”, (Kager, 1999, 25)). However, so far, reasonable idealizations factorizing out details of the utterance context etc. have been vital for progress in linguistic theory. So one might become skeptical in case there is no way at all of pinning down the linguistic part of the problem by a suitable interface specification (this may indeed be possible as discussed in sec. 5.3.4).

5.2.1 Apparent counterevidence against constraint ranking

Let us assume we want to capture the extra-syntactic factors in disambiguation by additional constraints within comprehension-based optimization. I will concentrate on the subtask of getting the filling of the argument positions correct (this ignores other ambiguities like ambiguous tense/aspect forms, quantifier scope, resolution of definite descriptions etc.). We have to take into account at least the knowledge sources in (163). The inferences, based on linguistic knowledge are informally sketched by implication arrows. Examples are given below. Unless one of the knowledge types leads to an unambiguous result, one typically gets the situation of several knowledge sources interacting to lead the hearer to a particular disambiguation.

(163) Knowledge sources involved in syntactic disambiguation & type of knowledge derivable

- Phonological string
  a. word order ⇒ (information structural status of argument phrases ⇒) grammatical functions of argument phrases ⇒ argument filling
  b. morphological (head or dependent) marking ⇒ grammatical functions ⇒ argument filling
  c. intonational marking (or typographical emphasis) ⇒ information structural status of argument phrases ⇒ grammatical functions ⇒ argument filling

138
5.2 The character of comprehension-based optimization

- Discourse context
  d. information structural status of argument phrases ⇒ grammatical functions ⇒ argument filling
e. anaphoric dependency ⇒ semantic class of (pronominal or polysemous) argument phrases ⇒ argument filling
- Extra-linguistic context
  f. reference of deictic expressions ⇒ semantic class of argument phrases ⇒ argument filling
g. partial disambiguation of argument filling through given situation
- Idiosyncratic/encyclopaedic knowledge
  h. selectional restrictions of the predicate ⇒ argument filling
  i. semantic class of (polysemous) argument phrases ⇒ argument filling
  j. exclusion of particular readings (encyclopaedic knowledge to the contrary)

Languages like English have word order (a.) as a fairly reliable direct source for function specification, which again allows inferences about the argument filling. In languages with a freer word order, like German, word order does not permit hard inferences, but may be an important clue, presumably via inferences about the information-structural status. We have seen case marking as an example for morphological dependent marking of grammatical function (b.), subject-verb agreement is a widespread example for morphological head marking. Intonational marking (c.) in spoken language provides clues about information-structural status (e.g., topic-marking pitch accent), which may for instance explain a certain word order, thus giving additional clues about function specification.\(^{105}\)

Interacting with intonational marking, the discourse context provides clues about the information-structural status of argument phrases (d.). For instance, the second sentence in (164) contains ambiguous case marking of the familiar kind. With the first sentence as context, it is fairly clear that we have two predications about the same topic, Anna Schmidt. The noun phrase die spätere Wahlsiegerin will typically bear a rising accent, and the intonation contour will rise over the rest of the second sentence up to the predicate vorgeschlagen with a falling accent. Furthermore drawing on encyclopaedic knowledge about elections, die

\(^{105}\)Some discussion of such effects can be found in (Kuhn 1996a, 1996b, 1996c).
The direction of optimization

spätere Wahlsiegerin is inferred as the theme argument of vorschlagen 'propose'.

(164) Anna Schmidt stellte sich einer Kampfabstimmung
   A. S. stood-for REFL a competitive election
   Die spätere Wahlsiegerin hatte Otto Müller für die Kandidatur
   The later winner had O. M. for the candidacy
   vorgeschlagen
   proposed

(164) contains an example of discourse-based anaphora resolution illustrating knowledge type (e.) in (163): pronominals and definite noun phrases have to be resolved in order to allow inference about the semantic class of an argument phrase. Likewise, deictic phrases allow the hearer to include information from the extra-linguistic context (f.). Of course, this context may also narrow down the choice of readings (g.) (for example, it may be obvious who is the agent of a certain action).

A fairly reliable source for direct inferences about argument filling (without the reasoning via grammatical functions) is knowledge about selectional restrictions (h.). An example of this was given in (162), where it was clear that compose has to take a human being as its agent. This inference may nevertheless be distorted by polysemy of argument phrases (i.). For instance (165), which normally has a preference for the subject-object reading, may also have a object-subject reading in case Mozart is used for the works of Mozart and diese Oper refers to an opera house. A context for bringing out this reading would be (166).

(165) Mozart hat diese Oper nie aufgeführt
   M. has this opera never performed

(166) Unser Haus ist spezialisiert auf die Italienische Oper der Romantik
   Our house is specialised in the Italian opera of Romanticism

Most of the context-based inferences can also be made without the relevant knowledge being actually introduced in the linguistic context; in this case, the inferences are based purely on encyclopaedic knowledge ((j.) in (163)).

The purpose of presenting this fairly long list of knowledge types involved in disambiguation is not to start the formalization of a particular
5.2 The character of comprehension-based optimization

costRAINT set for comprehension-based optimization. The validation of a non-trivial account of the problem would require a larger-scale computer simulation of learning based on empirical data and is thus clearly beyond the scope of this book. Here, the list of knowledge types should serve as background for considerations about the general character of this comprehension-based optimization.

At first, it may seem as if OT with its soft constraints is ideal for modelling this interaction of uncertain knowledge, which contributes defeasible information. Roughly, we might model each knowledge type as a constraint, introducing violation marks for the marked options. (For example, reverse word order with respect to the functional hierarchy would be marked; likewise type coercions from individuals to their works, as we had it in the Mozart polysemy; for the discourse-context-sensitive knowledge, the constraint system of Beaver (2000) could be applied.) From the different interpretations of a string, the most harmonic one would then be predicted as the preferred reading.

But does constraint interaction under the OT assumptions really capture the way in which the knowledge sources are interrelated in disambiguation? What ranking should we assume for the constraints? Some sources are more reliable than others (in part depending on the language), so these should be high in the constraint hierarchy. But an important knowledge source can be overridden by a sufficient amount of lower-priority clues. This points towards a constraint weighting regime—contrary to OT assumptions.\textsuperscript{106} Within a ranking regime, an additive effect could be reached only if all relevant constraints are unranked with respect to each other (or have a very similar rank in a stochastic OT approach), which again would defeat the idea of modelling reliability of knowledge sources by the rank of the corresponding constraint.

A possible conclusion at this point might be that the two optimizations are not instances of the same formal setup (compare fn. 99 on page 131, pointing to a similar conclusion made in Gibson and Broihier (1998)): constraint ranking has proven a very useful formal restriction on production-based/expressive optimization model of grammaticality, but it may not extend to an optimization model of interpretive preference. Preference might be better captured with a weighting model. If

\textsuperscript{106}Compare also arguments from the modelling of gradient grammaticality judgements Keller (2000), Keller and Asudeh (2001). Keller (2000) provides psycholinguistic results showing that under the assumption of agreed-upon constraint sets, gradience judgements cannot be derived with a pure constraint ranking system.
The direction of optimization
	his conclusion is correct, the formal similarity of the two optimization types in the declarative OT formalization would seem to be of limited empirical value.

However, the issues raised in this section could be interpreted differently, which is attempted in the following section.

5.2.2 Optimization with a fixed context

Due to the “softness” of OT constraints, i.e., the fact that they are violable by the ultimate output analysis, they seem to lend themselves to the modelling of defeasible information used in disambiguation. This was sketched in the previous section: if there is a sufficient amount of defeasible evidence (say, definiteness of argument phrases, intonational marking, plus encyclopaedic likelihood) pointing towards a particular reading of an utterance, this will overwhelm a residual of defeasible evidence in favour of a different reading (say, word order).

However, I conjecture that this interpretation of softness is different in a principled way from the typology-inducing character of constraint softness as used in the OT constraint ranking model of grammar. The typology-inducing constraint ranking model is crucially based on the (idealizing) assumption that there is no uncertainty in the knowledge sources underlying the constraint profile. The fact that a particular markedness constraint (say, Ob-Hd) outranks a particular faithfulness constraint (maybe Dep-IO) in language A has no implications to the effect that in language A, the clues suggesting that Ob-Hd is violated in a given analysis are more reliable than the clues for Dep-IO. Each candidate in a “pure” OT competition is a point in the abstract space of possibilities; the epistemological question of which are the matching abstract candidates for a given utterance lies outside the model. My conjecture is that this matching task is the locus of the uncertainty observed in the previous section. But this uncertainty cannot be modelled with the ranking of constraints too (without leading to an inconsistent overall picture).

In order to accommodate for both aspects of constraint softness, we can assume a two-stage optimization model (where the two stages do not imply any sequentiality, but just different regimes of defeasibility): the (“outer”) context-determining optimization and the (“inner”) optimization relative to a fixed context.
5.2 The character of comprehension-based optimization

The term (formal) context is used here in a technical sense. It comprises all the details that have to be fixed in order to unambiguously determine the full constraint profiles required for the inner optimization; a real-world context of utterance may leave some such decisions open, so several technical contexts are possible (maybe with different probability). While by assumption, the inner optimization follows the typology-inducing ranking approach of OT, the additive effect discussed in the previous section suggests that the outer optimization follows a less restrictive regime. (Only the inner optimization focuses exclusively on linguistic knowledge, the outer system includes the interface to the overall cognitive system; so we do not expect typological predictions to arise in this part. It is a research question whether the encapsulated inner system can be specified in a manner that comprises all and only the predictions pertaining to the typologically relevant linguistic knowledge.)

Why did the issue of the two types of constraint defeasibility arise only when we started looking at preferences in the disambiguation task (as modelled by comprehension-based/interpretive optimization)? When we are interested in native speakers’ (or rather hearers’) preferences for a particular reading of a sentence, it is next to impossible to abstract away from the outer optimization task. Too many information sources are effectively underspecified—especially under the standard linguistic methodology where examples are presented with no or very little linguistic context and in written form.

For modelling grammaticality in an expressive optimization framework, it is less unnatural to assume the idealization that a speaker has direct access to the fully specified contexts of the inner optimization. The input is not just the surface part of an utterance made by some other speaker, but it is the meaning that the speaker wants to convey.
The direction of optimization

in the actual context. The only way the speaker can go wrong in making a choice in the outer optimization is by making wrong assumptions about the common ground shared with the hearer.

So if this two-stage optimization model is indeed adequate, we actually expect that superficially, a simplified model using the ranking architecture of the inner optimization for the entire problem makes (almost) correct predictions for expressive optimization, while it fails for interpretive optimization (where a less restrictive model, i.e., a constraint weighting approach seems to be required). However, for an approach taking into account both directions, the two-stage architecture captures the relation between the two directions, in particular in the inner, constraint-ranking based optimization. The effects of this ranking in interpretive optimization may be generally masked by the outer optimization, but one may nevertheless be able to prompt them by narrowing down the contextual choices.

In this section, I argued that apparent counterevidence against a formal parallelism of the two directions of optimization can be accommodated in a parallel architecture that is not necessarily more complicated than an the asymmetrical model one would be forced to assume otherwise.

5.3 Bidirectional optimization

So far in this book, production-based optimization and comprehension-based optimization have been discussed in isolation (in chapters 2–4 and in sec. 5.1, respectively) or compared (sec. 5.2), but quite obviously the question arises of how the two optimizations relate to each other.

If one of them models grammaticality and the other some concept of preference, we expect that they have to make some reference to each other. A model taking into account both directions of optimization is called a bidirectional optimization model. (Recall that the term direction should not be taken literally, which would suggest a procedural/derivational definition; the formalization in chapter 4 and the extension in sec. 5.1 showed that a strictly non-derivational/declarative formulation is possible.)

Bidirectional optimization has been argued for variously in the theoretical OT literature, on empirical and conceptual grounds (see, e.g., Wilson (2001), Boersma (1998), Smolensky (1998), Lee (2001a), 144
5.3 Bidirectional optimization

Kuhn (2001c), Blutner (2000)). There are surprisingly many alternative formal options of combining the two unidirectional optimizations. Here, I will point out the most important options and discuss the general consequences for the character of the different resulting combined models (for a comparison with a similar goal, compare Beaver (2003)). Since there are various open issues about the individual optimization models (in particular the comprehension-based one, as discussed in sec. 5.2), it is certainly too early to draw any final conclusions.

5.3.1 Sequential bidirectional optimization models

There are two conceptually simple ways of combining the two unidirectional optimizations: one could run one after the other (resulting in a choice of which comes first), or one could run both simultaneously. The first variant is discussed in this section, the other one in sec. 5.3.2. Note that the terms “sequential” or “simultaneous” are used to refer to the logical sequence between the optimization relations, they are not meant to model any cognitively relevant order of processing. Procedures for computing these abstract relations will be discussed in chapter 6.

Recall the definition of the language generated by an OT system (based on the original concept of $Gen_{\text{inv} \circ \text{eval}}$, which was defined for production-based optimization—the essential detail is that the candidate set is defined by a common input f-structure $\Phi_{\text{in}}$):

\[
\text{(87) Definition of the language generated by an OT-LFG system } \mathcal{O} = \langle \text{inv}, \langle \text{eval}, \mathcal{L} \rangle \rangle
\]

\[
L(\mathcal{O}) = \{ \langle T_j, \Phi_j \rangle \in L(\text{eval}) \mid \exists \Phi_{\text{in}} : \langle T_j, \Phi_j \rangle \in \text{Eval}(\langle \text{eval}, \mathcal{L} \rangle)(Gen_{\text{inv} \circ \text{eval}}(\Phi_{\text{in}})) \}
\]

The grammaticality/preference model

We may base the construction of a bidirectional model on the linguistic intuitions we have about the concepts being modelled by the two optimization tasks. The original production-based optimization models grammaticality, while comprehension-based optimization models preference of readings. If we would like to give preference only a secondary status with regard to grammaticality, the production-based grammaticality model has to “feed” the comprehension-based preference model. That means that losers in the first optimization are discarded as ungrammatical, while the losers in the second optimization represent dispreferred readings of a string, but are nevertheless grammatical.
The direction of optimization

To illustrate the intuitions behind the second optimization step as modelling the preferred reading of a string, let us use a very simple concrete example. Let us assume a constraint FIRST_LASTNAME that is satisfied when a sequence of a first name and a last name (like John Smith) is interpreted as forming a single full name. It is violated when such a sequence is split into two nominal phrases. (There are various ways of implementing this constraint technically, which is not relevant here.) In comprehension-based optimization, the string is fixed across candidates, while the syntactic structure and interpretation (c-structure and f-structure) may vary. Assume the sentence Today, Sue called John Smith as the input string (as in tableau (168)). The FIRST_LASTNAME will be satisfied in the reading which we may paraphrase as Today, Sue gave Mr. John Smith a call, but it is violated in the reading Today, Sue called John by the name “Smith”. So, interpreting optimality as hearer preference, our tiny OT system will predict the former reading to be the preferred interpretation of the sentence.

(168) Comprehension-based optimization

As pointed out, in this scenario comprehension-based optimization has only a secondary status. Prior to it, we may assume the original production-based optimization model of grammaticality. Both of the two candidate analyses of the string are actually grammatical for English, so we model both of them as winners of two different initial production-based optimizations, based on their respective input f-structures. So, candidate (168a) is grammatical, but not the preferred reading of the given string. This distinguishes (168a) from losers in the initial optimizations, like Today, Sue John Smith called.

146
5.3 Bidirectional optimization

Formally, this means that in the sequential grammaticality/preference model, the definition of the language (=the set of grammatical analyses) generated by an OT system is not changed. But we are adding an additional notion, the notion of the preferred reading of a string on top of this.

\[(169) \text{Preferred analyses in the grammaticality/preference model}
\]
\[
(T_j, \Phi_j) \text{ is among the preferred analyses of a string } w \text{ iff }
\]
\[
(T_j, \Phi_j) \in \text{Eval}_{\mathcal{C}, \mathcal{R}, \mathcal{Z}}(\mathcal{L}(\mathcal{O}) \cap \{ (T', \Phi') \in \mathcal{L}(\mathcal{G}_{\mathcal{O}}) \mid w \text{ is the terminal string of } T' \})
\]

Note that here the set of analyses that is evaluated—the argument of \(\text{Eval}_{\mathcal{C}, \mathcal{R}, \mathcal{Z}}\)—is not defined by \(\mathcal{G}_{\mathcal{O}}\), but it is a subset of the language defined by the (production-based) OT-LFG system. The interactions are visualized in the schematic illustration in (170) on page 148.

The upper half reflects the production-based optimization underlying the definition of the language \(\mathcal{L}(\mathcal{O})\). Only the winners which contain the relevant input string are considered further on; this is indicated by the broken lines from the fixed input string on the right-hand side in the middle. The lower half then shows how from this (infinite) set of analyses with a particular terminal string, the preferred analysis is determined by optimization.
The direction of optimization

(170) The grammaticality/preference model
5.3 Bidirectional optimization

The comprehension-based/production-based sequence

Besides the grammaticality/preference sequence, which leaves the concept of grammaticality untouched, it is conceivable to adopt the opposite order of the two optimizations. An initial step of comprehension-based optimization filters out certain readings of a string, and only the winners are candidates for the final production-based optimization. Such a model is assumed by Jäger (2002a). Intuitively, the initial step filters out readings that would be irrecoverable for the hearer; grammaticality is defined through the final optimization. Note that this opposite order leads to different empirical predictions (if the constraint set is left unchanged). Candidate (168a) would no longer be called grammatical: it would have been filtered out in the first step since it loses against (168b).

The comprehension-based/production-based sequence of optimization can be useful in predicting certain ways of expressing a thought because another alternative would be hard to interpret for the hearer. Take the string Mary told John Smith left. If we split John from Smith, we get a reading that should be perfectly acceptable. But unless we get very clear prosodic clues that disambiguate the sentence in this direction, the string is odd. A speaker would say Mary told John that Smith left. This rephrasing is not predicted by the grammaticality/preference model (169), since preference is only modelled over identical strings.

In the comprehension-based/production-based sequence of optimization, the winner of the initial comprehension-based optimization would be an analysis with the bracketing Mary told [xp John Smith] left, as shown in (171).\textsuperscript{107} The verb left has a null subject in this analysis. (We assume here that the constraint FIRSTLASTNAME is highly ranked.) However, in the subsequent production-based optimization starting from (171b)’s f-structure, it will turn out that the winner of the first optimization does not survive as a grammatical analysis for English: The analysis with the original string loses against Mary told John Smith she left (compare (40) on page 37). Therefore, the initial string Mary told John Smith left is predicted to be unacceptable. If we rephrase it as Mary told John that Smith left, the constraint FIRSTLASTNAME will no longer filter out the intended reading, so grammaticality follows.

\textsuperscript{107}For the SUBJECT constraint, which is just used as an example for further lower-ranking constraints, compare (40) on page 37.
The direction of optimization

(171) Comprehension-based optimization

[Diagram of a sentence structure with labels: Subject, Pred, Subj, Pred, Obj, Pred, Comp, Pred.
Input String:
Mary told John Smith left

5.3.2 Strong bidirectional optimization

As an alternative to the sequential bidirectional models, we could assume that both directions are equally important and have to apply conjunctively. In a successful candidate, the string must be optimal among all strings for the underlying meaning and the meaning must be optimal among all possible meanings of the string. This is what Blutner (2000) calls the strong bidirectional optimization model (the corresponding weak model will be discussed in sec. 5.3.3).

The crucial difference between the strong bidirectional model and the sequential comprehension-based/production-based model is the following: in the strong bidirectional model, the candidate sets for both optimizations are independent, while in the sequential model, the candidate set in production-based optimization (the second optimization) consists only of comprehension-based winners. This may result in different predictions. In strong bidirection it may happen that the two optimizations do not agree on a candidate as the winner. This cannot happen in the sequential model, since the second optimization will only have initial winners as candidates.

For concreteness, let us add a very simple Economy-of-expression constraint *STRUCT to our sample OT system, which is violated by any c-structure node. With this constraint, an independent production-based optimization for the input f-structure (172) will produce the
5.3 Bidirectional optimization

string *Mary told John Smith left* as the winner, since it violates *STRUCT fewer times than *Mary told John that Smith left.*

(172) Production-based optimization

The comprehension-based optimization works exactly as shown in (171). Note that (171a) and (172a) are identical LFG analyses: this candidate shows up in both candidate sets. However, it wins only in (172), while losing in (171). Similarly, (172b) is the winner for its comprehension-based optimization (not shown here), but a loser for its production-based optimization (namely (172)).

The two directions do not agree on a winner. In strong bidirectional optimization this means that none of the mentioned candidate analyses is in the language generated by the system. This includes candidate (172b) which would be a final winner in the sequential system, but which is here blocked by the simpler (though ungrammatical) f-structure-realization (172a). This blocking effect can be evoked fairly easily in a strong bidirectional system.

The definition of strong bidirectionality and its consequences

In sec. 5.1, an extended definition of *Gen* was considered which generalizes to the application for both optimization directions (see (151)). So we can define a language consisting of those analyses that are optimal for both directions: the formal definition is given in (174) on page 153. Illustrating the overall system again graphically, we get the scheme (173) on page 152. Note the independence and symmetry of the two optimizations.

| a. Mary told John Smith left | *...* |
| b. Mary told John that Smith left | *...**! |

The definition of strong bidirectionality and its consequences

In sec. 5.1, an extended definition of *Gen* was considered which generalizes to the application for both optimization directions (see (151)). So we can define a language consisting of those analyses that are optimal for both directions: the formal definition is given in (174) on page 153. Illustrating the overall system again graphically, we get the scheme (173) on page 152. Note the independence and symmetry of the two optimizations.
The direction of optimization

(173) Strong bidirectional optimization in OT-LFG

input/Index:
partial f-structure

Subsumption

\[ \phi \circ f \text{-str.} \]

\[ \lambda \circ l \text{-str.} \]

\[ \text{cand}_1 \]

\[ \text{cand}_2 \]

\[ \text{cand}_3 \]

\[ \langle n_1, n_1', n_2', \ldots, n_k' \rangle \]

\[ \langle n_1', n_2', n_3', \ldots, n_k' \rangle \]

\[ \langle n_1, n_2, n_3, \ldots, n_k \rangle \]

Eval \[ (\text{cand} \supseteq \text{C}) \]

Optimal

Terminal string

input/Index:
word string
5.3 Bidirectional optimization

(174) Language generated by a strong bidirectional OT system
\[ O = \langle G_{\text{invio}}, \langle C, \rightarrow L \rangle \rangle \]
\[ \mathcal{L}_{\text{strong}}(O) = \{ \langle T_j, \Phi_j \rangle \in L(G_{\text{invio}}) \mid \exists \Phi_{in} : \langle T_j, \Phi_j \rangle \in \text{Eval}_{\langle C, \rightarrow L \rangle}(\text{Gen}_{G_{\text{invio}}} (\Phi_{in})) \]
\[ \text{and} \]
\[ \exists w : \langle T_j, \Phi_j \rangle \in \text{Eval}_{\langle C, \rightarrow L \rangle}(\text{Gen}_{G_{\text{invio}}} (w)) \} \]

The independence of the two optimizations in the strong bidirectional model has the advantage that the system is conceptually rather simple (this goes along with certain computational advantages, as we will see in sec. 6.3). With the \textit{blocking} technique, strong bidirectionality also opens up a simple way of deriving Language-Particular Ineffability, a phenomenon discussed in sec. 3.3.3 which poses a problem for standard unidirectional OT (recall that the assumption of LF-unfaithful candidates made by Legendre et al. (1998) is highly problematic from a learnability point of view). In strong bidirectional OT, all those underlying forms are predicted to be ineffable for which the production-based winner is suboptimal in the comprehension-based optimization (like in the above example where the input f-structure of (172) is effectively ineffable). See sec. 5.3.5 for a more detailed discussion and illustration.

The simplicity of an Ineffability analysis in the strong bidirectional model obviously excludes the danger of overgeneration that one may see for a standard production-based OT system. It is not so clear however whether this effect is reached at the price of undergeneration. Here, a problem of the independence of the two optimizations may become relevant: for every well-formed analysis one has to ensure that it comes out optimal in both directions. Without further adjustments (such as context-sensitive constraints), the strong bidirectional system predicts that ambiguity of strings is practically nonexistent. This \textit{ambiguity problem} was first pointed out by Hale and Reiss (1998). But even a system with context-sensitive constraints, which could in principle resolve the ambiguity problem, requires much further research. With the enforced context-dependence of the comprehension-based optimization discussed in sec. 5.2, there is at least a practical problem for strong bidirectional OT: the constraint system taking context-effects into account must be extensive for a non-trivial grammar fragment.

A related potential disadvantage of strong bidirectionality is that due to mutual independence, no interaction between the optimizations can be exploited for explanatory purposes. Thus, the inherent idea of con-
The direction of optimization

Constraint interaction as the main explanatory device is defeated at the interface of the two directions. If interaction effects across the two optimization directions are allowed, this can simplify the constraint systems considerably, as will be discussed in the following subsection.

5.3.3 Weak bidirectional optimization

Perhaps most problematic about the strong concept of bidirectionality is the strict blocking effect that any analysis exerts which is optimal in one of the directions. Competitors that are less harmonic cannot play any further role at all. In his OT system for lexical pragmatics, Blutner (2000) proposes a concept of weak bidirection that remedies this problem. Like strong bidirection, weak bidirection is symmetrical—none of the two directions is prior to the other. However, the candidate sets in the two directions are not independent of each other, but highly interrelated—based on a recursive definition: only the ultimate winners are allowed as candidates. In order to visualize the effect of such a recursive optimality relation, one of course has to work with preliminary candidate sets containing analyses that have to be eliminated later on. When the weakly bidirectional winners are computed this way, analyses that would be a loser in strong bidirection “get a second change” because a blocking competitor gets removed.

This is best seen in a small abstract example. Assume we have two strings \(s_1\) and \(s_2\), both of which could have either meaning \(m_1\) or \(m_2\) according to \(G_{inviol}\). So we have four possible candidates \(\langle s_1, m_1 \rangle\), \(\langle s_2, m_1 \rangle\), \(\langle s_1, m_2 \rangle\), \(\langle s_2, m_2 \rangle\). Let us assume that \(\langle s_1, m_1 \rangle\) has the most harmonic constraint profile, and \(\langle s_2, m_2 \rangle\) the least harmonic profile (the other two candidates are in between).

So if we do a comprehension-based optimization for input \(s_1\), candidate \(\langle s_1, m_1 \rangle\) will win. For input \(s_2\), candidate \(\langle s_2, m_1 \rangle\) wins. Similarly for production-based optimization. In strong bidirection, the analysis \(\langle s_1, m_1 \rangle\) is the only acceptable candidate: the \(s_2\) and \(m_2\) candidates are blocked by this candidate.
5.3 Bidirectional optimization

In weak bidirection, only the ultimate winners (“superoptimal candidates”) are actually defined to be candidates in a competition. So, for instance, we cannot tell ahead of time which are the candidates of the comprehension-based optimization for string $s_2$. We have to compute the correct candidate sets (=the superoptimal candidates) recursively. In the recursion, we have to start at the base case. The base case are those analyses for which we know that no other (potential) candidate can be more optimal in either direction. These are exactly the “old” strong bidirectional winners—in our case only $\langle s_1, m_1 \rangle$.

It is the winner for the comprehension-based optimization starting from $s_1$ and for the production-based optimization starting from $m_1$. But since in a single optimization, there can only be one winner (we know there are no ties), we now know that $\langle s_2, m_1 \rangle$ cannot be a superoptimal candidate! Likewise for $\langle s_1, m_2 \rangle$. They are both blocked by $\langle s_1, m_1 \rangle$ in their respective competitions. So the following picture arises:

\[
\begin{array}{c}
\bullet \\
\downarrow \\
\bullet \\
\bullet \\
\bullet \\
\end{array}
\]

Since two potential candidates turned out not to be actual candidates (because they fail to be superoptimal), we realize that there are no competitors left in the optimizations based on $s_2$ and $m_2$ that would block candidate $\langle s_2, m_2 \rangle$. Hence it is superoptimal itself. So, the language includes two analyses: $\langle s_1, m_1 \rangle$ and $\langle s_2, m_2 \rangle$. (In the end we know the exact candidate sets for the optimizations: they are the singleton sets of the respective winner: $\{\langle s_1, m_1 \rangle\}$ and $\{\langle s_2, m_2 \rangle\}$, respectively.)

Blutner (2000) provides examples indicating that weak bidirection is useful for deriving partial blocking effects in lexical pragmatics. For example, the existence of a lexeme *kill* in the lexicon leads to a special interpretation of the expression *cause to die*: it is usually interpreted as involving an indirect chain of causation (e.g.: ‘Black Bill caused the sheriff to die: he caused the sheriff’s gun to backfire by stuffing it with cotton’). *kill* corresponds to $s_1$, *cause to die* is the less economical expression $s_2$. *kill* is interpreted as a direct causation ($m_1$), so the more complex indirect-causation interpretation ($m_2$) is paired up with *cause to die.*
The direction of optimization

Strong bidirection would only predict strict blocking, i.e., one would expect that *cause to die* could not be used as it is blocked by *kill*.

The definition of weak bidirectionality

Transferred to the terminology of this book, we get the following definition of a weak bidirectional OT system.\(^{108}\)

$$\begin{align*}
\text{(177) The weak bidirectional OT system} \\
O_{\text{prod}}(\mathcal{O}) &= \{ (T_j, \Phi_j) \in L(G_{\text{inv}}) \mid \\
& \exists \Phi_{\text{in}} : (T_j, \Phi_j) \in \text{Eval}_{\{C, \Rightarrow C\}}(O_{\text{compr}}(\mathcal{O})) \\
& \cap \{ (T', \Phi') \in L(G_{\text{inv}}) \mid \Phi_{\text{in}} \subseteq \Phi' \} \} \\
O_{\text{compr}}(\mathcal{O}) &= \{ (T_j, \Phi_j) \in L(G_{\text{inv}}) \mid \\
& \exists w : (T_j, \Phi_j) \in \text{Eval}_{\{C, \Rightarrow C\}}(O_{\text{prod}}(\mathcal{O})) \\
& \cap \{ (T', \Phi') \in L(G_{\text{inv}}) \mid w \text{ is the terminal string of } T' \} \} \\
L_{\text{weak}}(\mathcal{O}) &= O_{\text{prod}}(\mathcal{O}) \cap O_{\text{compr}}(\mathcal{O})
\end{align*}$$

\(O_{\text{prod}}\) and \(O_{\text{compr}}\) are auxiliary languages specifying the production-based optima and the comprehension-based optima. What is crucial is the interdependence of the candidate sets: optimization in the production-based direction is performed on comprehension-based optima which have the same input f-structure; and vice versa, comprehension-based optimization is performed on production-based optima with the same string. Ultimately, the language generated by a weak bidirectional system is the intersection of the two auxiliary notions.\(^{109}\)

\(^{108}\) Without making the system sensitive to context, as discussed in sec. 5.2.2, the application of such a system to the syntactic domain is not very satisfactory. For comparability, I nevertheless keep to the format of the definitions given so far, i.e., ignoring context-dependence of the analyses.

\(^{109}\) Concerns that the notion of weak bidirectionality may not be well-defined, due to the apparent circularity, are refuted by Jäger (2002c). A recursive definition is possible, since the harmony relation, on which the application of Eval is based, is transitive and well-founded (Jäger, 2002c, sec. 2). Intuitively, the recursive computation has to start at the “base cases”—the most harmonic candidates, and OT’s harmony relation guarantees that we can identify them (they are the winners of strong bidirection).

Jäger also proves that the results of Frank and Satta (1998), Karttunen (1998)—that in regular-language based OT systems, the entire system is a rational relation—carry over to weak bidirectionality.
5.3 Bidirectional optimization

So far, there has been little empirically oriented work elaborating the details of such an account. But the concept seems very appealing for the derivation of various alignment scales (cf. also Beaver (2000)). On the other hand, as Beaver (2003) argue, there is an overgeneration problem with weak bidirectionality: at least in the finite case, every form is ultimately paired up with some meaning. So it is unclear how ungrammaticality can be modelled in this architecture.

In the following I discuss null subjects as an example from syntax making use of elements of weak bidirectionality. The analysis could not be modelled in a grammaticality/preference model or in strong bidirectionality. (It is however possible to model it using the comprehension-based/production-based sequence of optimization, which does not suffer from the overgeneration problem just mentioned.)

5.3.4 Deriving recoverability through bidirectional optimization

Null subjects

Recall from sec. 3.2.3 the discussion of null subjects in Italian, based on the OT analysis of Grimshaw and Samek-Lodovici (1998) and Samek-Lodovici (1996). As the example in (178) illustrates in some more detail, the subject is dropped in Italian when it is coreferent with the topic of the preceding sentence: in (178a), Gianni is the subject and thus the topic, so in (178b), no pronoun appears. In (179) however, Gianni is the focus or within the focus of the first sentence (the topic being la mostra ‘the exhibition’). So (179b) is ill-formed with a null subject.

(178) a. Questa mattina, Gianni ha visitato la mostra.
    this morning G. has visited the exhibition

   b. Più tardi, egli/egli lui ha visitato l’università.
    more late (he)/he has visited the university

(179) a. Questa mattina, la mostra è stata visitata da Gianni.
    this morning the exhibition was visited by G.

   b. Più tardi, egli lui ha visitato l’università.
    more late (he)/he has visited the university

Grimshaw and Samek-Lodovici (1998) derive these facts with the three constraints (180) in an ordinary comprehension-based optimization. Parse is a different formulation for our constraint MAX-IO. To avoid terminological confusion, I will keep to the terminology of this book, calling the constraint MAX-IO.

157
The direction of optimization

(180) (Grimshaw and Samek-Lodovici, 1998, 194)

a. DROP\textsc{Topic}
   Leave arguments coreferent with the topic structurally unrealized. Failed by overt constituents which are coreferential with the topic.

b. SUBJECT
   The highest A-specifier in an extended projection must be filled. Failed by clauses without a subject in the canonical position.

c. \textsc{Parse} [MAX-IO]
   Parse input constituents. Failed by unparsed elements in the input.

The ranking Grimshaw and Samek-Lodovici (1998) assume for Italian is the following:

(181) \textsc{DropTopic} \gg \textsc{MAX-IO} [or \textsc{Parse}] \gg \textsc{Subject}

Furthermore assuming that the topicality status of the pronoun’s antecedent is encoded in the input, we get the following tableau for ha cantato (‘he has sung’) —as a simplified version of the second sentence in (178) (candidate b., which is maximally MAX-IO-unfaithful to the input is called the null parse).

(182) (Grimshaw and Samek-Lodovici, 1998, 202)

<table>
<thead>
<tr>
<th>Input: \langle \text{cantare}(x), x = \text{topic}, x = \text{lui} \rangle</th>
<th>\textsc{DropTopic}</th>
<th>\textsc{MAX-IO} [or \textsc{Parse}]</th>
<th>\textsc{Subject}</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. \text{Es}° ha cantato</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>°!°</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. lui ha cantato</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ha cantato lui</td>
<td>°!</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Candidates c. and d. violate DROP\textsc{Topic}, since the subject pronoun is coreferent with the topic, but nevertheless realized in these candidates. Since the DROP\textsc{Topic} outranks the faithfulness constraint MAX-IO, candidate a. with the unexpressed subject pronoun is thus most harmonic.

158
5.3 Bidirectional optimization

When the subject pronoun is not coreferent with the topic (as in (179)), candidates c. and d. satisfy DROPTOPIC, so their faithfulness to the input pays off this time, and c. becomes the winner:

(183) (Grimshaw and Samek-Lodovici, 1998, 203)

\[
\begin{array}{|c|c|c|}
\hline
\text{Input: } (\text{cantare}(x), x = \text{lui}) & \text{DROPTOPIC} & \text{MAX-IO} \\
\hline
\text{a. } \text{ha cantato} & * & * \\
\text{b. } & * & * \\
\text{c. } \text{lui ha cantato} & * & * \\
\text{d. } \text{ha cantato lui} & * & * \\
\hline
\end{array}
\]

For English, Grimshaw and Samek-Lodovici (1998) assume the ranking MAX-IO $\gg$ DROPTOPIC $\gg$ SUBJECT. Hence, the effect of DROPTOPIC is neutralized by the faithfulness requirement, and no null subjects are predicted.

Deriving recoverability from constraint interaction

The constraint DROPTOPIC that Grimshaw and Samek-Lodovici (1998) assume is a very effective constraint, but its formulation is also a fairly complex. It contains a condition checking the discourse status of a pronoun’s antecedent—within the antecedent’s sentence; and sensitive to this condition the non-realization of the argument is rewarded.

Following the discussion in chapter 2 and sec. 3.1, such a constraint poses two questions: (i) is there an independent motivation and (ii) is the constraint really primitive, given its complex logical structure? (Cf. also the commitment to logically simple constraints in Grimshaw (1998).) To answer question (i): What could be a functional motivation for leaving underlying material unexpressed? At first this seems to thwart the most basic communicative principles. But it can be reconstructed straightforwardly from economy principles. To keep utterances brief, it is essential to avoid unnecessary structure. Of course, economy can only be complied with to the degree that the content of the utterance can still be conveyed. In other words, only material that

---

110 Candidate d. may only win if an additional constraint ALIGNFOCUS is assumed, ranked between MAX-IO and SUBJECT. With the subject marked as focus, only d. will satisfy this constraint and thus win.
The direction of optimization

can be easily reconstructed by the hearer should be dropped. This is exactly what the constraint DROP_TOPIC specializes on: avoid the repetition of (a certain type of) material that is already contextually given. So there is independent motivation. But, addressing question (ii), is the constraint really a primitive one?

I would like to argue that in a bidirectional optimization model, the interaction of communicative forces just discussed as a way of motivating DROP_TOPIC can be captured more perspicuously as an instance of constraint interaction, based on simpler constraints. As an additional point, this derivation of the effect of DROP_TOPIC is to demonstrate the independence of this interaction from the faithfulness constraint MAX-IO (or PARSE). Even with a low-ranking MAX-IO, we can get a language that does not allow null subjects. This is important since otherwise, we would have very limited possibilities of explaining ellipsis phenomena in non-pro-drop-languages as an effect of constraint interaction (involving MAX-IO too).

The economy of expression principle can be formulated as a constraint *XP (184). With this constraint, the basic pattern favouring null subjects is very simple (185) (note the low ranking of MAX-IO).

\[ (184) \quad *XP \\
\quad \text{Avoid structure. Violated by maximal projections.} \]

\[ (185) \\
\quad \text{Input: } \langle \text{cantare}(x), x = \text{lui} \rangle \\
\quad \text{a. } [\text{IP } \text{ha [VP cantato]}] \\
\quad \text{b. } [\text{IP [NP lui] ha [VP cantato]}] \\
\quad \text{Candidate a. wins over c. simply because it contains less structure. For English, we would have } \text{SUBJ} \gg *XP, \text{so c. would win. This pattern alone certainly does not give us the correct results. It predicts that the \null parse (182b) is even more harmonic than a., and it does not predict any difference between the topic context (178) and the non-topic context (179).} \\
\quad \text{The intuition is that material can only be dropped as long as it is contextually recoverable. One way of formalizing this intuition would be to limit the candidate generation function } Gen_{\text{cl_wnt}} \text{ accordingly, assuming a recoverability principle. Then the null parse would no longer} \]

160
5.3 Bidirectional optimization

appear in the tableaux, and the two tableaux (182) and (183) would differ in that a. would be missing from the latter one.

However, the recoverability effects can be explained by assuming a comprehension-based optimization which complements the production-based part just sketched. (Subsequently, I will focus on the problem of topic/non-topic context, leaving aside the null parse problem; once the context problem is solved, a generalization to the null parse problem follows.)

To make the bidirectional optimization explicit, we need an architecture with a representation of context (compare the diagram in (186)).

(186) Architecture for a simple context-dependent OT model

For making the case, I will here assume a very simple context representation: a set of the f-structures of the sentences in the salient context. Of course, what is actually required is some discourse semantic representation, for instance a discourse representation structure (cf. Kamp 161
The direction of optimization

(1981, 1990); Kamp and Reyle (1993)). It is essential that we have discourse referents for the individuals that can be referred to, rather than syntactic representations of surface phrases. But to avoid notational overhead, I will pretend that we can use the PRED value ‘Gianni’ in the f-structure as a discourse referent that is anchored to the individual the speaker and hearer know by the name Gianni.

As a further simplification I will assume that the feature TOPIC in the f-structure represents exactly the relevant information-structural notion (which is not quite true for LFG’s actual discourse function TOPIC—this is the grammaticalized version of the information-structural notion).

A discourse coherence constraint DISCOCOHER checks the f-structure of the currently uttered sentence against the salient context. Finding a good formulation for this constraint is not trivial, but for our purposes, the very simple formulation in (187) one will do.111 In Italian, DISCOCOHER is ranked above *XP.112

(187) DISCOCOHER

For atomic f-structure values in the current f-structure, the current f-structure and an f-structure from the salient context set contain the same feature path.

\[
\text{atomic-f-str} (\star) \rightarrow \exists f, P [f \text{-str}(f) \land (f \text{ CURRENT } P) = \star \land (f \text{ CONTEXT } P) = \star]
\]

Coreference with topic: null subject

Consider the following simplified version of dialogue (178):

(188) a. Gianni è venuto.
\hspace{1cm}G. \hspace{0.5cm} has come

b. Ha cantato
\hspace{1cm}has sung

111 Just like the faithfulness constraints MAX-IO and DEP-IO, DISCOCOHER should probably be seen as a family of constraints, parametrized for particular features.

The formula in the constraint description language assumes that the current f-structure is represented under a feature CURRENT and the set of contextual f-structures under CONTEXT.

112 Presumably this ranking must hold universally; but it should not be required to stipulate this fact; a re-ranking would lead to an absurd system, which should be precluded by the learning scheme. ‘\%’ is used to reference one set member (rather than distributing the information over the entire set).
5.3 Bidirectional optimization

We want to derive that in this context, the null subject in the second sentence is correct. In a symmetrical bidirectional system, we have to check the following: is *ha cantato* the optimal way of expressing the underlying content *sing(Gianni)*? And vice versa, is *sing(Gianni)* the optimal interpretation of the string *ha cantato*? Ignoring the null parse at this point, this turns out to be straightforward in the given context. The relevant optimizations are shown in (189) and (190) below.

(189) Production-based optimization

For the production-based optimization (189), DiscCoher takes no effect, since both the f-structure input and the context are fixed, and so all candidates incur the same DiscCoher violations (every informative utterance will violate discourse coherence to a certain degree). Thus, as in (185), economy gives us *ha cantato* as the winner.

In comprehension-based optimization (190), different interpretations of the string *ha cantato* compete. For the syntactic markedness constraints (*XP and Subject), these candidates will all have the same constraint profile, since the string is fixed.\(^{113}\) We do get differences in the Max-IO violations, as candidate c. shows: this candidate assumes that the string is elliptical for *Maria claimed that Gianni has sung*. So the amount of material left unrealized in the c-structure/l-structure varies with the interpretation assumed as underlying in the candidates. Note however that the number of violations of DiscCoher is decisive for the outcome. Since *Gianni* is the topic in the context, candidate a. incurs fewest violations of DiscCoher and is thus the winner in (190).\(^{114}\)

\(^{113}\)Generally, of course different syntactic analyses are possible for a given string, but here we may assume that all analyses other than [\(IP\ ha [VP\ cantato]\)] are wildly more marked.

\(^{114}\)Strictly speaking, the candidate set contains also candidates with a more heavily unfaithful c- and l-structure. This way *ha cantato* can in fact be combined with an f-structure that is 100% coherent with respect to the context. Given the high ranking
The direction of optimization

(190) Comprehension-based optimization

<table>
<thead>
<tr>
<th>Context</th>
<th>Input String</th>
<th>DISCOHER</th>
<th>XP</th>
<th>SUBJECT</th>
<th>MAX-IO</th>
</tr>
</thead>
</table>
| a. (188) | Topic
SUBJ, PRED 'GIANNI'
TNS, PRES
ASP, PERF | ha cantato | * | * * * | * |
| b. | Topic
SUBJ, PRED 'MARIA'
TNS, PRES
ASP, PERF | | **! | * * * | * |
| c. | Subj
PRED 'MARIA'
TNS, PRES
ASP, PERF | | **! | * * * | ** |

Taking together the two directions of optimization (189) and (190), we get the clear prediction that ha cantato is the correct form in the context under discussion.

Coreference with non-topic: overt subject

Let us now investigate the alternative context, with Gianni not being the topic in the previous sentence. We want to predict that a null subject is unacceptable in that context.

(191) a. Questa mattina, la mostra è stata visitata da Gianni.
this morning the exhibition was visited by G.

b. *e/lui ha cantato
(he)/he has sung

of DISCOHER such a candidate is the actual winner of this optimization. However, as soon as the opposite optimization direction is considered, the irrecoverability of them becomes obvious. I ignore these candidates in this exposition, since it is more illustrative to demonstrate the irrecoverability with more intuitive examples.

164
5.3 Bidirectional optimization

We get (roughly) the following context representation in (192). Note that *la mostra* 'the exhibition' is marked as the topic.

\[(192)\]

This time, the assumption of a weak bidirectional model proves essential. Recall that the candidate sets in weak bidirectionality are defined by mutual reference to the opposite concept of optimality. Abstractly, one should think of all dependencies as applying simultaneously. But if we want to verify the predictions we are forced to proceed in some sequence. This means that initially, we can only work with preliminary candidate sets, which are potentially too large since some candidates may not actually be optimal in the opposite direction (in fact, very many of them will not).

Let us start with comprehension-based optimization for the string *lui ha cantato*:

\[(193)\] Comprehension-based optimization—preliminary candidate set

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Context} & \text{Input String} & \text{DISC/HER} & \text{*XP} \\
\hline
\{ \text{TOPIC} & \text{SUBJ} & \text{PRED} & \text{OBL} \\
\{ \text{PRED} & \text{‘EXHIBITION’} \} & \text{‘visit(x, y)’} & \text{‘GIANNI’} \} & \text{TNS} \\
\{ \text{ASPB} & \text{PERF} & \text{PREP} & \text{‘GIANI’} \} & \text{\ldots} \\
\hline
\text{DISC/HER} & \text{*XP} & \text{SUBJ} & \text{MAX-JO} \\
\text{lui ha cantato} & \text{\ldots} & \text{\ldots} & \text{\ldots} \\
\hline
\hline
\text{a.} & \text{TOPIC} & \text{SUBJ} & \text{PRED} & \text{OBL} \\
\{ \text{PRED} & \text{‘GIANNI’} \} & \text{‘sing(x)’} & \text{TNS} \\
\{ \text{ASPB} & \text{PERF} & \text{PREP} & \text{‘GIANI’} \} & \text{\ldots} \\
\hline
\text{b.} & \text{TOPIC} & \text{SUBJ} & \text{PRED} & \text{OBL} \\
\{ \text{PRED} & \text{‘MARIA’} \} & \text{‘claim(u, v)’} & \text{TNS} \\
\{ \text{PREP} & \text{‘GIANI’} \} & \text{‘sing(x)’} & \text{ASPB} \\
\{ \text{PREP} & \text{‘GIANI’} \} & \text{PREP} & \text{‘GIANI’} \\
\text{\ldots} & \text{\ldots} & \text{\ldots} & \text{\ldots} \\
\hline
\end{array}
\]
The direction of optimization

Since the subject is realized, there is a smaller space of possibilities for filling out the missing part (the coreferent element has to be third person masculine). And although due to the different context, the discourse coherence is now lower than it was in (190), we get candidate (193a)—*sing(Gianni)—as a clear winner.\textsuperscript{115}

We might now expect that the production-based optimization will confirm *lui ha cantato* as optimal for *sing(Gianni)*. This is not the case however. Without the specially tailored DROP\textsc{Top}ic constraint, there is nothing substantial that makes competition (194) (for the non-topic context) different from (189) above (for the topic context).

(194) Production-based optimization—preliminary candidate set

<table>
<thead>
<tr>
<th>Context</th>
<th>Input F-Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOPIC</strong></td>
<td><strong>TOPIC</strong></td>
</tr>
<tr>
<td><strong>SUBJ</strong></td>
<td><strong>SUBJ</strong></td>
</tr>
<tr>
<td><strong>PRED</strong> <em>visit</em>(x, y)*</td>
<td><strong>PRED</strong> <em>sing</em>[x]</td>
</tr>
<tr>
<td><strong>OBL</strong> [PRED `GIANNI']</td>
<td><strong>OBL</strong> [PRED `GIANNI']</td>
</tr>
<tr>
<td><strong>TNS</strong> PRES</td>
<td><strong>TNS</strong> PRES</td>
</tr>
<tr>
<td><strong>ASP</strong> PERF</td>
<td><strong>ASP</strong> PERF</td>
</tr>
<tr>
<td><strong>a. [IP ha [VP cantato]]</strong></td>
<td><strong>b. [IP [NP lui] ha [VP cantato]]</strong></td>
</tr>
</tbody>
</table>

Had we assumed a strong bidirectional model, this would be a problem. Since the two optimizations do not agree on a single f-structure/string pair, the underlying input f-structure would be predicted to be ineffable under the ranking for Italian.

In a weak bidirectional model we are not finished yet however. We do not know yet whether we were initially using the correct candidate sets for the two optimizations. So we have to check whether (194a) really should have been in the candidate set for the respective underlying input f-structure (i.e., *sing(Gianni)*). For this to be the case, *sing(Gianni)* would have to be optimal for the string of (194a)—*ha cantato—in the present context. This comprehension-based competition is checked in (195).

\textsuperscript{115}The same proviso as in footnote 114 on page 164 applies.
5.3 Bidirectional optimization

(195) Comprehension-based optimization—preliminary candidate set

This competition starts out similar to (190), but note the difference in the context. For this reason, Gianni is no longer the most coherent choice for filling out the subject. It is more coherent to use the topic of the previous sentence as the subject, i.e., the exhibition. So we have found an instance of irrecoverability: in the present context, sing(exhibition) is optimal for ha cantato and vice versa. This shows that (194a) is not contained in the true candidate set for (194), according to weak bidirectional optimization and we get b. as the winner of the corrected tableau:

---

This means that the actual prediction for the data in (191) and (179) is the following: the null subject version of the second sentence (in both examples) is not syntactically ungrammatical, but it has the semantically anomalous reading in which the exhibition is the agent of an action. This seems a reasonable account of the data.
Hence we have a match between the two directions of optimization (193) and (196): in the non-topic context, *lui ha cantato* is predicted to be the correct form for *sing(Gianni)*.

Note finally, that although the null subject analysis incurs a MAX-IO violation, this constraint is neither involved in deriving the right context effects, nor in the cross-linguistic contrast with non-pro-drop-languages like English (English follows if *subject outranks *XP). This is desirable since it leaves space for ellipsis analyses in non-pro-drop-languages.

The argumentation in this section showed several things:

- the constraint set assumed by Grimshaw and Samek-Lodovici (1998) for the analysis of null subjects can indeed be simplified by moving to a bidirectional optimization model: the complex formulation of the DROP TOPIC constraint is no longer required;
- in a bidirectional optimization account, the interaction between constraints and the directions of optimization derive the concept of recoverability, which is usually assumed as a meta-restriction on syntactic analyses;
- optimization based on a fixed context permits a fine-grained account of the empirical data, while at the same time working with highly general constraints (using the same constraint set for both directions);
- finally, the notion of weak bidirectionality (as opposed to the straightforward strong model) is crucial for such an account. \(^{117}\)

\(^{117}\)As pointed out at the end of sec. 5.3.3, it is however possible to model the facts using the comprehension-based/production-based sequence of optimization.

168
5.3 Bidirectional optimization

5.3.5 Ineffability and strong vs. weak bidirectionality

As pointed out briefly in sec. 5.3.2, a bidirectional OT model provides a solution to the puzzle that Language-Particular Ineffability poses to the OT approach (cf. sec. 3.3.3) without recurring to the assumption of LF-unfaithful candidates (as Legendre et al. (1998) do). Recall that LF-unfaithful candidates are problematic from the point of view of learnability (sec. 3.3.4).

Ineffability and strong bidirectionality

With strong bidirectionality, the mechanism that makes a certain logical form ineffable in a particular language is quite straightforward: only meaning-form pairs that are optimal in both directions are defined to be included in the language of the OT system. Now, suppose the optimal candidate for a certain underlying representation $\Phi$ has surface string $w$. But in the comprehension-based optimization of the alternative candidate analyses with string $w$, we get a different winner: some candidate with the underlying form $\Phi'$ ($\Phi' \neq \Phi$). This means that $\Phi$ is ineffable in that language. One may say that this meaning is irrecoverable for the hearer since she must always assume that when a speaker utters $w$ he means $\Phi'$. In other words, $\Phi'$ blocks $\Phi$, making $\Phi$ ineffable. We can illustrate this by the following schematic picture (following a visualization scheme of Johnson (1998)); I use the * on an f-structure representation to mark ineffability and $\prec$ for the less-harmonic-than relation:

(197) $^*\Phi \quad \Phi' $

Note that ungrammaticality and ineffability are exact mirror images in a strong bidirectional model: Ungrammaticality of a string $w$ results if the optimal among all analyses for a string $w$ has the underlying form $\Phi$, but in production-based optimization for $\Phi$ we get an optimal candidate with a different surface string $w'$ ($w' \neq w$). So, $w'$ blocks $w$, making $w$ ungrammatical.

---

118 This application of bidirectional optimization was brought to my attention by Joan Bresnan (p.c., March 2000). The same general idea underlies the proposal by Smolensky (1998) (see also the application of bidirectional OT-LFG in Lee (2001a)).
The direction of optimization

(198)

With the simplicity of blocking configurations, it is easy to see that we can have a situation where an underlying form $\Phi$ is blocked by an alternative reading $\Phi'$ of string $w$, which is itself ungrammatical due to another string $w'$ for $\Phi'$. String $w$ may for example be a multiple $wh$-question like *Who ate what*, which is ungrammatical in Italian, with the underlying multiple $wh$-question ($\Phi$) being ineffable. An alternative interpretation of *Who ate what*, with an indefinite object ($w'$), blocks $\Phi$. The string $w$ itself is blocked by a different string $w'$—the Italian equivalent of *Who ate anything*.

(199)

For the present purposes it is of subordinate relevance what exact constraints are involved in giving rise to the sketched harmony relation among the candidates. The following ad hoc interpretive constraints under ranking $C^1 \gg_{\text{Italian}} C^2$ would for instance have the intended effect of making $\Phi'$ more harmonic than $\Phi$:

(200)  
   a. $C^1$: non-topicalized pronouns get an indefinite interpretation
   b. $C^2$: $what$ is interpreted as a $wh$-word

In English, the opposite constraint ranking would lead to a different picture, with both the form-meaning pair $\langle w, \Phi \rangle$ and $\langle w', \Phi' \rangle$ being well-formed:

(201)
Ineffability and weak bidirectionality

How can ineffability be modelled in the weak bidirectional model discussed in sec. 5.3.3 and 5.3.4? The direct blocking account of the strong model does not carry over, since with weak bidirectionality the unoptimal candidates of the first round may “get a second chance”. As the bidirectional winners are removed from the preliminary candidate sets, losers of a strong bidirectional competition may still become winners under the weaker regime. For the picture in (199), we would for instance exclude the $(w, \Phi)$ candidate from the preliminary candidate set of $w$: since we know that the optimal form for $\Phi$, we can be sure that $w$ cannot be its optimal form too.

(202)

But this would change the situation in such a way that $(w, \Phi)$ is predicted to be bidirectionally optimal too—both in English and Italian. This shows that with weak bidirectionality a more sophisticated account of ineffability is required. An underlying representation will be predicted to be ineffable only if all possible realizations are bidirectionally optimal for some other representation, as sketched in the schematic picture in (203) (with preliminary candidate sets on the left and the final candidate sets—singletons—on the right). The competition for the ineffable meaning $\Phi$ ends up with an empty candidate set since all preliminary candidates turned out to be optimal for some other meaning.

(203)

Note however that all strings in the original preliminary candidate set for $\Phi$ are predicted to be grammatical (with some other meaning). This poses an empirical problem for ineffability examples like the Italian *Who ate what. Of course, the context-sensitive set-up proposed in sec. 5.3.4 provides further means to derive the unacceptability of
**The direction of optimization**

examples—all possible contexts may be semantically anomalous, but we have to note that the strong bidirectional model allows for a more elegant account of Language-Particular Ineffability. Hence, both the strong and the weak model have conceptual and empirical advantages and disadvantages; clearly more research is required before one can decide which one should be favoured—or how a third alternative should look like, combining the advantages of both.

Interestingly, some of the computational considerations in the following chapter will depend on design decisions in the bidirectionality question too.

### 5.4 Summary

In this chapter I discussed various issues around the directionality question in a formal OT model of syntax. At this point it would be premature to draw any definite conclusions. A striking observation is that although the formal means make it straightforward to define a comprehension-based optimization in analogy to the standard production-based model, it is fairly intricate to judge whether the behaviour of such a system is in line with linguistic intuitions. Future work has to clarify the character of the individual optimizations and their combination further.

Sec. 5.3.4 showed however that the context-oriented application of a weak bidirectional model gives rise to a very promising account of syntactic phenomena in context. The analysis uses a set of well-motivated constraints with a simple structure, deriving the recoverability condition on syntactic structures as an effect of constraint interaction. On the other hand, sec. 5.3.5 revealed advantages of a strong bidirectional OT model in the derivation of Language-Particular Ineffability. The decision between the two types of bidirectionality has to be considered open and will require further research.