Conflicts in Interpretation – A Critical Review

Michael Franke & Maria Aloni

Abstract

This squib gives a critical review of the monograph entitled *Conflicts in Interpretation* (Hendriks et al., 2010) written by Petra Hendriks, Helen de Hoop, Irene Krämer, Henriette de Swart and Joost Zwarts, which was published in 2010 by Equinox Publishing, London. After sketching the relevant background of optimality theoretic approaches to semantics and pragmatics, we give a detailed summary of the contents of this book, discuss its merits and mention a few critical issues that, we feel, future research in this tradition may wish to adress more carefully.

1 Background

The friction between formalist and functionalist views of grammar has triggered lively debates within the wider field of the cognitive sciences for decades. Formal approaches attempt to provide a systematic theory of linguistic phenomena based on the idea that linguistic knowledge is rule-governed, using formal tools in the tradition of Chomsky and Montague. Functional models, instead, attempt to explain linguistic phenomena in terms of general principles of human communication, thereby appealing to both psychological and social factors. Formal models of language start from the assumption that language is primarily an autonomous system. Functionalists, who are traditionally skeptical towards the possibility of a rigorous analysis of linguistic phenomena, take the communicative instrumentality of language as their point of departure. Optimality Theory (oт) tries to reconcile these two traditions by presenting a model of grammar which is not modular, but still maintains the predictive power of the formalist tradition.

In oт well-formed linguistic exchanges are considered to be the result of an optimization process between speakers and hearers. Speakers select the optimal form expressing a given meaning and hearers select the optimal interpretation for a given form. This yields two varieties of *unidirectional optimality*, once for the production perspective, once for the comprehension perspective. But there is also *bidirectional optimality*, first introduced by Blutner (1998, 2000), which combines both the production and the comprehension perspective. There are two notions of bidirectional optimality. *Strong bidirectional optimization* selects those form-meaning associations which are unidirectionally optimal for both speaker and hearer. *Weak bidirectional optimization* selects all strongly optimal associations plus those that become optimal when previously established optimal associations are removed from the competition. *Strong bidirectional optimization* explains blocking phenomena: e.g., if a meaning m can be expressed equally well with two different forms f_1 and f_2, but f_1 would not be

interpreted by the listener as *m*, while f_2 would, then strong bidirectional optimization, taking the hearer's perspective into account, would block the form-meaning association f_1-*m* also for production. *Weak bidirectional optimization* explains a well-studied pattern in language use known as *Horn's division of pragmatic labor* (Horn, 1984) according to which unmarked forms express unmarked meanings and marked forms express marked meanings. All of these optimality notions are formally defined in OT but capture a functionalist pressure on language use to be efficient for the speaker and/or the hearer. It is in this sense that OT integrates both formalist and functionalist tendencies.

To determine what is optimal, OT considers a set of possible input-output pairs, which in applications within semantics or pragmatics are usually form-meaning pairs, and checks which of these least violate a number of ranked constraints. These constraints capture general linguistic preferences or cognitive dispositions. Crucially, these constraints can be conflicting, so that their rank ordering is important. To illustrate this with an example borrowed from Chapter 1 of *Conflicts in Interpretation*, consider the following two constraints:

(1)   a. SUBJECT: All clauses must have a subject.

       b. FULL-INTERPRETATION: All constituents in a sentence must be interpreted.

SUBJECT, penalizing subjectless forms, is an example of a *markedness constraint* which express preferences for some forms (or meanings) over others because they are unmarked, i.e., more frequent, less complex, or easier to interpret. FULL-INTERPRETATION is an example of a *faithfullness constraint*. While markedness constraints take into account only forms (or only meanings), faithfullness constraints associate particular forms with particular meanings. A conflict arises between SUBJECT and FULL-INTERPRETATION whenever there is no semantic need for a subject, for example with verbs like *to rain*. SUBJECT requires to express the subject nevertheless, but if so, FULL-INTERPRETATION would be violated. Different languages solve this conflict in different ways. English allows for a meaningless element to fill the subject position ('it is raining'), while languages like Italian ('piove') don't. Crucially, OT allows constraints to be violable and captures differences between various languages in terms of different rankings of the same set of violable principles: in English, SUBJECT outranks FULL-INTERPRETATION; in Italian, this is the other way around. This way, OT is able to account for both cross-linguistic variability and uniformity.

2   Summary

Petra Hendriks, Helen de Hoop, Irene Krämer, Henriette de Swart and Joost Zwarts are established researchers working at different Dutch institutions. Their recent book *Conflicts in Interpretation* (Hendriks et al., 2010) presents the result of a 4-year collaboration in a project funded by the Netherlands Organisation for Scientific Research (NWO). This collaboration resulted in many publications in prestigious international journals. The refined upshot of this research is concisely presented as a unity in this book, thus documenting the great potential and wide applicability of OT within linguistics.

Conflicts in Interpretation introduces the reader to ot techniques and applies ot to a wide variety of phenomena ranging from rhetorical relations over stressed pronouns to spatial prepositions. The book contains a general introduction to ot in the first chapter and a conclusion summarizing the main ideas in Chapter 9. Each of the book's seven core chapters deals with a unique linguistic phenomenon to which ot, especially bidirectional ot, is fruitfully applied. These core chapters address many issues which are crucial to various areas of linguistic research from discourse analysis to morphology, from acquisition, to typological variation and diachronic change. The following summarizes their respective content.

Chapter 2 discusses different rhetorical effects of preposed or postposed *when*-clauses (as in 'they stabbed him when he died' vs. 'when they stabbed him, he died', or 'he began to read with pleasure when he was killed' vs. 'when he began to read with pleasure, he was killed'). Using a wealth of natural examples from novels and other sources, the chapter accounts for these different interpretations in terms of interactions between production and interpretation constraints. It argues in general that speakers (writers) use their implicit linguistic knowledge of how hearers arrive at optimal interpretations as a means of persuasion.

Chapter 3 investigates different word orders in Dutch (*unscrambled*, i.e., adverb-object vs. *scrambled*, i.e., object-adverb) and, focusing on pronominal objects, accounts for the interplay between position, stress and meaning in terms of bidirectional competition. Taking the role of stress into account, the resulting analysis goes beyond the simple picture that marked forms (i.e., unscrambled word order for pronominal objects) go with marked (contrastive) meanings whereas unmarked forms (i.e., scrambled word order for pronominal objects) go with unmarked (anaphoric) meanings. Different forms (scrambled or unscrambled) can come out as unmarked, if unstressed; different forms (scrambled or unscrambled) can come out as marked, if stressed.

Chapter 4 and 5 discuss a number of interesting cases of asymmetry between comprehension and production in child language. A well-known example concerns pronouns. When interpreting a sentence like 'Bert is washing him', adults normally conclude that 'Bert' and 'him' cannot co-refer, whereas children until at least 6;6 years of age do allow co-reference in these cases. This fact is particularly puzzling because children at that age do not seem to have problems in the production of sentences involving (reflexive) pronouns. Hendriks et al. explain the adult behavior in terms of bidirectional blocking (the hearer concludes that 'Bert' and 'him' cannot co-refer because if the speaker had wanted to express co-reference, she would have used the reflexive 'himself'), and propose that it is precisely this bidirectional reasoning that children cannot perform (possibly because of insufficient working memory, lower processing speed or lack of a theory of mind). On this account children do not lack knowledge of the grammar and its constraints, their interpretations differ from those of adults because they use this knowledge in a different way. These chapters extensively discuss two more phenomena where children's production of a given form precedes its comprehension, namely wide scope vs. narrow scope interpretations of (scrambled) indefinites in Dutch (Chapter 4) and wide focus vs. narrow focus interpretation of contrastive stress in English and Dutch (Chapter 5).

Chapter 6 investigates the phenomenon of negation across languages. All languages seem to have an expression for negation. Assuming that negative meanings are marked

(because expressed less frequently), this universal pattern is explained as emerging from evolutionary pressure as an instance of Horn's division of pragmatic labor (cf. Zeevat and Jäger, 2002; Jäger, 2004; Mattausch, 2004, 2007). On the other hand, the large cross-linguistic variation in how human languages integrate the expression of negation in the syntax as pre-verbal, post-verbal or discontinuos negation is explained in terms of re-ranking of a number of partially conflicting constraints, some expressing different strategies on how to arrive at communicative efficiency. From a diachronic perspective it is also shown that stepwise re-ranking of these constraints, modeled in terms of stochastic or, derives the well-known pattern of language change known as *Jespersen's cycle*: languages tend to move from pre-verbal negation to discontinuous negation to post-verbal negation and then back to pre-verbal negation, and so on. In the final part of the chapter, weak bidirectional optimality is shown to account for double negation readings of combinations of negation with a negative indefinite in negative concord languages like French, Afrikaans and Welsh.

Chapter 7 investigates phenomena of referentiality in human languages. As in the previous chapter, cross-linguistic variation in this domain is explained in terms of a universal set of violable constraints, combined with the possibility of language specific ranking. It seems to be part of human cognition that we use nominals to set up referents in a discourse space, but some languages like English encode this cognitive ability in the functional structure of the nominals (definite/indefinite article, singular/plural distinction), while other languages don't. A well-known example of the latter kind is Chinese which lacks a formal singular/plural distinction and has no system of definite/indefinite articles. Hendriks et al. argue against traditional accounts of this typological variation in terms of covert functional operators or parametrization, and propose that language variation in the expression of reference arises from the interaction of a number of faithfulness constraints, on the one hand, which govern the correspondence between functional structure and discourse meaning (e.g., a principle expressing that a definite article corresponds to a discourse referent with determined meaning), and a general markedness constraint, on the other hand, which blocks functional structure in the nominal domain. Different languages solve the conflict between this economy constraint and the correspondence rules via language-specific ranking. Faithfulness constraints, driving towards the expression of semantic distinctions in the form of the nominal, carry more weight than the markedness constraint in some languages (e.g., English), but not in others (e.g., Chinese). The chapter develops a full typology of bare nominals in or terms including data from Chinese, Georgian, Hebrew, St'át'imcets, English and French. The point of departure is the intuition that bare nominals, lacking functional complexity, are the unmarked case. Even in languages like English where the economy principle is ranked lower than the faithfulness constraints, bare nominals are possible in a range of special constructions like *at school* or *in jail*. In the final part of the chapter, the association of these bare forms with stereotypical interpretations is explained using weak bidirectional optimality as an example of Horn's division of pragmatic labor.

Chapter 8 investigates how the meaning of spatial prepositions is shaped by competition and blocking with neighboring prepositions. For example, the difference in meaning between the English directional prepositions *through*, *into* and *out* are explained by a blocking mechanism similar to the one we find in the derivation of scalar

implicatures in Gricean pragmatics (Grice, 1975; Levinson, 1983). *Through*, *into* and *out* are assigned simple semantic values which are assumed to create an implicational scale, like ⟨*and*, *or*⟩ or ⟨*all*, *some*⟩ in logical semantics. Their difference in meaning is then explained in terms of a specificity constraint expressing a preference for stronger meanings similar to the Gricean Maxim of Quantity. Other constraints are employed to explain the contrasts between *around* and *over*, *off* and *out of*, and between the Dutch 'round' prepositions *om*, *om . . . heen* and *rond(om)*. The latter are assumed to exemplify once again Horn's division of pragmatic labor with *om* as the least marked form of the three covering all three senses expressible by these prepositions (namely a quarter/a half/a full circle around the reference point), and *rond(om)* as the most complex form compatible only with the most marked interpretation, namely the one requiring a full enclosure path. Weak bidirectional optimality is again applied here to predict this iconic alignment of expressive and interpretative complexity.

3   Praise

The general conceptual and empirical merits of OT are well known. On the conceptual level, OT strikes a fine balance between symbolic and sub-symbolic accounts of information processing within the cognitive sciences (Prince and Smolensky, 1997; Smolensky and Legendre, 2006). When applied to linguistics, OT similarly mediates between functionalist and generative approaches. On the empirical level, the idea of violable constraints yields versatile empirical predictions, including, for instance, the possibility of predicting preferred interpretations also for ungrammatical sentences. By allowing violable constraints to be ranked differently, OT succeeds in succinctly and systematically capturing cross-linguistic variability and universality. *Conflicts in Interpretation* is an excellent example of this latter feature, in particular with its contribution on the distribution of post-, pre-verbal or discontinuous negation (Chapter 6) and on the distributive pattern of bare nominals in the languages of the world (Chapter 7).

But the contribution of *Conflicts of Interpretation* goes beyond these basic merits of OT. The book shows in a variety of settings that the inclusion of different notions of optimality can do outstanding predictive work in linguistics. Next to the usual unidirectional optimality notions, the book also considers bidirectional optimality, in both its strong and weak variety. As bidirectional optimality brings together the production and the comprehension perspective, it helps explain a number of otherwise puzzling phenomena. This is a recurring theme of the book, but the strongest point in case of bidirectionality is certainly its role in explaining comprehension lags in first language acquisition (Chapters 5 and 6). The book argues that young children display the pattern of gradually evolving grammatical competence *not* because they have not mastered the proper grammar, i.e., constraint ranking, but because they fail to optimize bidirectionally based on the proper grammar. This makes it possible to explain how young children can display almost adult-like production behavior, yet fail miserably on comprehension.

These particular benefits notwithstanding, the book's most convincing argument in favor of OT, from our point of view, is a delicately performed proof-of-concept that OT *can* be applied, as a simple and elegant uniform framework, to many different linguistic

domains, viz., lexical semantics, the syntax-semantics interface, pragmatics, discourse structure. The authors succeed to deliver this demonstration at a very detailed, even meticulous empirical level: the book excels in investigating a manifold of divergent linguistic phenomena in a detailed, systematic and exceptionally illuminating manner. Afloat of this strong empirical contribution the book drives home its main conceptual message:

> [N]ot only the use of the grammar of a language but also important aspects of the grammar itself and of the lexicon seem to be shaped by the essentially pragmatic principle of bidirectional optimization. This suggests that the traditional distinction between the representation of linguistic knowledge and the use of linguistic knowledge may have to be reconsidered.
>
> (Hendriks et al., 2010, p. 197)

4 Critique

Although we do not want to draw into doubt that *Conflicts in Interpretation* is a milestone in optimality-theoretic semantics and pragmatics, there are a number of critical issues that we feel ought to be mentioned. We want to raise three points of critical reflection here, not because we want to claim that these might even partially invalidate the book's positive contribution, but because we believe that future research could benefit from tending to these matters more carefully. The points we raise here, in descending order of felt importance, are: (i) the interpretation of bidirectional optimization, (ii) the possibilities of obtaining quantitative predictions from OT, and (iii) the book's lack of explicit consideration of context within the OT-formalism.

Firstly, we consider it mildly unsatisfactory for the sake of concreteness of predictions that the very notion of bidirectional optimization that is so central in *Conflicts in Interpretation* is conceptually underdetermined in that it allows for a number of reasonable interpretations. The authors themselves notice this and discuss at some depth (mainly Chapter 5) that the definition of bidirectional optimality can be motivated in at least two ways, namely as a synchronic process of online reasoning about language use or as a diachronic process capturing language change.[1] We may think of this as a certain "conceptual abstractness" which is both a virtue and a vice. It is a virtue foremost because it's due to this abstractness that the framework is versatile enough to be applicable to so many different linguistic phenomena from a manifold of linguistic domains. But it is also a vice because it obscures exactly which empirical predictions might be drawn from the framework. For example, when weak optimality is used in Chapter 6 to account for double negation in negative concord languages, or in Chapter 8 to explain the meaning differences between Dutch directional prepositions *om*, *om ... heen* and *rond(om)*, would this or would this not give rise to predictions about the expected acquisition pattern of these constructions similar to the pattern discussed in Chapters 4 and 5? In other words, would we or would we not predict that young children only

---

[1]Another central notion that is also left conceptually underdetermined is "markedness" of forms and/or meanings. Here, too, the book is explicit about this potential conceptual problem, but we believe that conceptual clarity about the nature of bidirectional optimization may force conceptual clarity about markedness as a welcome side effect. (But see also our later remarks about the possible context-dependence of markedness.)

produce the least marked form *om* and interpret each of the three forms as the least marked interpretation "at least a quarter circle around the reference object"? The problem we are raising here is not that these predictions might not be true, but that there is nothing intrinsic to the applied ot-framework that would tell us whether these are the predictions of the approach or not. As a consequence of this, we suggest that future ot-research in the vein of *Conflicts in Interpretation* should be more concerned with its conceptual foundations (cf. Franke and Jäger, 2012). It should back-up the beautiful picture of versatile uniform applicability with an explication of the *differences* between phenomena and domains under scrutiny and an explanation why *nonetheless* we discern an underlying common core principle that can be captured abstractly as bidirectional optimization.

Secondly, and much in line with the previous point of criticism, we are concerned that the very abstractness of ot also precludes more detailed quantitative predictions. That quantitative predictions are desirable for the kind of phenomena addressed in the book should be obvious. For example, Chapter 2 tries to explain why sentences with preposed *when*-clauses and temporal-overlap meaning are acceptable but *rare* (p. 38); Chapter 3 wants to explain why anaphoric definites scramble *more often than not* (p. 51). Indeed, quantitative predictions of ot are another recurring theme of *Conflicts in Interpretation*, but, unfortunately, the matter is never given the importance we think is due. Different possible scenarios of deriving quantitative predictions from ot are discussed. Chapter 2 discusses the option that frequency of observation of a form-meaning pair is proportional to the number of optimality notions (unidirectionally in production, in comprehension, strong or weak optimality) which that form-meaning pair satisfies. Chapter 3 builds on work by Antilla and Cho (1998) suggesting that frequency of occurrence is derived from the number of times a form-meaning pair comes out as optimal given a set of *variable* constraint rankings. Chapter 6 looks at stochastic ot (Boersma, 1998) to explain how re-ranking of constraints leads to smooth transitions in Jespersen's cycle. Finally, although somewhat indirectly, Chapter 5 suggests in response to criticism towards weak optimality raised by Beaver and Lee (2004) that if weak optimality is calculated with the iterative algorithm of Jäger (2002), then as the number of iterations needed to establish weak optimality is proportional to the cognitive effort that needs to be invested, the number of steps needed would be anti-proportional to the observed frequency. In line with our previous criticism that bidirectional ot might be conceptually too underdetermined, this list of disparate ideas how to obtain quantitative predictions from ot is unsatisfactorily heterogenous. Moreover, it is not clear how some of the mentioned suggestions could be fine-tuned to obtain concrete falsifiable prediction. For example, although stochastic ot, as used in Chapter 6, is the most straightforward way of obtaining qualitative predictions, it is not clear what the mechanism of stochastic change really is that drives the to-be-explained language change, in this case Jespersen's cycle. We are tempted to conclude from this once more that the ot-framework endorsed in *Conflicts in Interpretation* is explanatorily powerful only on an abstract level, with interesting details pushed to the side — details that we feel future work in ot should ideally drag back center-stage into the spotlight of investigation.

Thirdly and finally, we would like to mention what strikes us as an intriguing omission in the way ot is applied to phenomena in semantics and pragmatics in this book.

We believe that the notion of *context* plays a major role in assessing linguistic meaning and it would therefore be expected that, unlike in phonology and maybe even syntax, a formal theory of semantics and pragmatics should include central reference to contextual information and contextual variability. However, there is no place in the OT-formalism that *explicitly* caters for contextual information. Of course, the book frequently discusses constraints that are explicitly or implicitly context-sensitive. An example of such is the constraint dubbed "Don't overlook anaphoric possibilities" (p. 52), where what is an anaphoric possibility is not solely determined by the form-meaning pair in question but also by the context in which it is used. Similarly, but more implicitly, the related constraint "Don't overlook rhetorical possibilities" (p. 32) is also context-sensitive to some extent because, for instance, whether a given rhetorical relation is plausible is very much a question of *contextual* plausibility at the time of utterance. Quite generally, this also applies to the notion of markedness, especially but not exclusively markedness of meaning. What is a marked meaning may depend on what is unexpected *in context*, not only normally or across-the-board. The omission of explicit inclusion of context indeed has peculiar effects. For example, in Chapter 2 example (34) on page 40 is judged to be unidirectionally optimal for both comprehension and production but not as strongly optimal, contrary to what we would expect given that strong optimality is *defined* as the conjunction of the two former notions. But this is because the contextual information whether a particular rhetorical relation is plausible is treated differently in the unidirectional and the bidirectional system. We consider this somewhat odd and conjecture that in general OT-semantics and OT-pragmatics could valuably be extended to include a notion of context explicitly (cf. Aloni, 2007, for an OT-system that does). Bidirectional optimality should possibly not just be defined for form-meaning pairs, but for context-form-meaning *triples*. How exactly, we do not consider this the right place to speculate about. But we do dare to speculate that this inclusion will in turn necessitate a more rigorous stance as to the exact conceptual interpretation of OT— which, from our point of view, would be a very welcome side-effect.

In summary, *Conflicts in Interpretation* makes a convincing case for using OT in semantics and pragmatics. The research summarized here may certainly be considered the state-of-the-art in OT-applications to semantics and pragmatics. As such, the book is a must-read for anybody working in this particular field, but it will be highly relevant for a much wider audience. Linguists with a general interest in semantics and pragmatics will particularly appreciate the broader picture of this approach that incorporates a typological perspective and insights from cognitive science into formal linguistic theory. Due to its accessible writing style, the book recommends itself not only to experts, but also to advanced students and interested scholars from other fields.

References

Aloni, Maria (2007). "Expressing Ignorance or Indifference. Modal Implicatures in Bi-Directional Optimality Theory". In: *Logic, Language and Computation: Papers from the 6th International Tbilisi Symposium*. Ed. by Balder ten Cate and Henk Zeevat. Vol. 4363. Berlin: Springer Verlag, pp. 1–20.

Antilla, Arto and Young-mee Yu Cho (1998). "Variation and Change in Optimality Theory". In: *Lingua* 104.1–2, pp. 31–56.

Beaver, David and Hanjung Lee (2004). "Input-Output Mismatches in Optimality Theory". In: *Optimality Theory and Pragmatics*. Ed. by Reinhard Blutner and Henk Zeevat. Palgrave MacMillan. Chap. 6, pp. 112–153.

Blutner, Reinhard (1998). "Lexical Pragmatics". In: *Journal of Semantics* 15, pp. 115–162.

— (2000). "Some Aspects of Optimality in Natural Language Interpretation". In: *Journal of Semantics* 17, pp. 189–216.

Boersma, Paul (1998). *Functional Phonology*. The Hague: Holland Academic Graphics.

Franke, Michael and Gerhard Jäger (2012). "Bidirectional Optimization from Reasoning and Learning in Games". In: *Journal of Logic, Language and Information* 21.1, pp. 117–139.

Grice, Paul Herbert (1975). "Logic and Conversation". In: *Syntax and Semantics, Vol. 3, Speech Acts*. Ed. by Peter Cole and Jerry L. Morgan. Academic Press, pp. 41–58.

Hendriks, Petra et al. (2010). *Conflicts in Interpretation*. London: Equinox Publishing.

Horn, Laurence R. (1984). "Towards a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature". In: *Meaning, Form, and Use in Context*. Ed. by Deborah Shiffrin. Washington: Georgetown University Press, pp. 11–42.

Jäger, Gerhard (2002). "Some Notes on the Formal Properties of Bidirectional Optimality Theory". In: *Journal of Logic, Language and Information* 11.4, pp. 427–451.

— (2004). "Learning Constraing Sub-Hierarchies". In: *Optimality Theory and Pragmatics*. Ed. by Reinhard Blutner and Henk Zeevat. Houndmills: Palgrave MacMillan, pp. 251–287.

Levinson, Stephen C. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.

Mattausch, Jason (2004). "On the Optimization and Grammaticalization of Anaphora". PhD thesis. Humboldt University Berlin.

— (2007). "Optimality, Bidirectionality & the Evolution of Binding Phenomena". In: *Research on Language and Computation* 5.1, pp. 103–131.

Prince, Alan and Paul Smolensky (1997). "Optimality: From Neural Networks to Universal Grammar". In: *Science* 275, pp. 1604–1610.

Smolensky, Paul and G. Legendre (2006). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, Massachusetts: Cambridge University Press.

Zeevat, Henk and Gerhard Jäger (2002). "A Reinterpretation of Syntactic Alignment". In: *Proceedings of the 3rd and 4th International Symposium on Language, Logic and Computation*. Ed. by Dick de Jongh et al.