

Chapter 9

Implications for future research

9.1 Current results

In the previous chapters we showed that the process of linguistic interpretation can be viewed as a process of optimization. Under this view, the optimal meaning for a given form is determined on the basis of the transparent mechanism of conflict resolution provided by Optimality Theory. Conflicts may arise at different levels of interpretation: the level of the discourse (Chapter 2), the level of the interface between discourse and sentence (Chapters 3, 4 and 5), the level of the sentence (Chapters 6 and 7), and the level of word meaning (Chapter 8). These conflicts are the result of the different, and often incompatible, demands the grammar places on forms, meanings and the relations between them. To decide among the potential meanings for a given form, these potential meanings are evaluated on the basis of a set of ranked violable constraints. The meaning that optimally satisfies these constraints will be selected by the hearer. An optimization approach was already quite successful in the area of phonology and has also been proposed to explain phenomena in the areas of morphology and syntax. By extending this approach to the area of semantics and pragmatics, we are in the strong position of being able to contribute to, and test, a unified theory of language that aims to account for all aspects of the grammar. Such a unified theory allows for a better understanding of issues at the interface between grammatical modules, for example the interaction between word order and rhetorical structure (Chapter 2) or the interaction between word order and sentence stress (Chapter 3). Moreover, it makes it possible to look for commonalities across different modules of the grammar. Do markedness constraints in phonology, syntax and semantics have certain properties in common (see also Section 2 below)? Can we distinguish common stages in the acquisition of phonology, syntax and semantics on the basis of the unified optimization model of language? Can we account for cross-linguistic variation in the domains of phonology, syntax and semantics in a similar way?

In the areas of phonology and syntax, Optimality Theory makes specific claims with respect to language acquisition and language variation. An important assumption is that languages share the same set of constraints but differ in the ranking of these constraints.

Children, when acquiring their native language, have to learn the particular ranking of the constraints in their language. To obtain the adult constraint ranking, they have to re-rank their innately specified constraints on the basis of positive evidence provided by the language input. The constraint re-ranking model of language acquisition has not only been applied to phonology, but also to syntax (e.g., Legendre, Vainikka, Hagstrom, and Todorova, 2002; Legendre, 2006) and semantics (Hendriks, de Hoop, and Lamers, 2005).

In Chapters 4 and 5 we showed that, in addition to constraint re-ranking, the acquisition of meaning involves an extra step. In comprehension, children also have to learn to take into account the forms the speaker could have used but did not use. These alternative, unheard, forms influence the adult meaning of the utterance. Acquiring the ability to take into account the speaker's perspective occurs relatively late in child language development, in general around or even after the age of 7. This accounts for the delays in language acquisition found for the comprehension of indefinite subjects and objects (Chapter 4), contrastive stress (Chapter 5) and other phenomena briefly mentioned in these chapters (such as the comprehension of personal pronouns and scalar implicatures).

Interestingly, speaker influences on comprehension are not only found for phenomena that are traditionally categorized as pragmatic phenomena (such as contrastive stress and scalar implicatures), but also for phenomena traditionally seen as belonging to the grammar (for example, the word order effects discussed in Chapters 3 and 4, and the presence or absence of an article as discussed in Chapter 7) or the lexicon (e.g., the polysemous prepositions studied in Chapter 8). In this book, we showed that these phenomena can all be accounted for in exactly the same way, namely through bidirectional optimization. That is, not only the use of the grammar, but also parts of the grammar and the lexicon themselves, are shaped by the essentially pragmatic principle of bidirectional optimization.

A topic which is still relatively unexplored in semantics is the potential variation of meaning across languages. In Chapters 6 and 7 we showed that languages differ in the way they express and interpret negation and the way they use and interpret articles. Languages use the same set of faithfulness constraints linking forms to meanings, but they resolve the conflicts between these universal faithfulness constraints and the general drive towards speaker economy in a different way. This accounts for the observed cross-linguistic variation, while conforming to the general pattern that unmarked forms pair up with unmarked meanings, and marked forms pair up with marked meanings.

9.2 Markedness of meaning

Many of the semantic analyses presented in this book appeal to the concept of markedness. Markedness constraints are constraints which evaluate outputs only, without regard to the input. This contrasts with faithfulness constraints, which evaluate candidates by considering both the input and the output. Although markedness is a central notion in OT, capturing exactly what markedness means is by no means a straightforward task. In general, markedness constraints in OT phonology and OT syntax seem to have effects as different as prohibiting phonetic difficulty (e.g., the constraint NOCODA: ‘A syllable must not have a coda’), prohibiting conceptual difficulty (e.g., Aissen’s (1999) constraint *SUBJECT/PATIENT: ‘Avoid subject patients’), and avoiding infrequent structures (e.g., the constraint ‘Indefinite objects do not scramble’ in Chapter 4). In fact, Haspelmath (2006) argues that the term ‘markedness’ is used in so many senses in the framework of OT, and in linguistics in general, that there is little hope that we will be able to agree on a core sense of markedness. For this reason, he argues that reference to markedness should be replaced by other, more straightforward, terms such as frequency (more marked forms are less frequent), phonetic difficulty and pragmatic inferences.

In this book we have used the term ‘markedness’ first of all as a purely formal term, referring to the output orientation of a particular set of constraints, in accordance with the use of the term in OT phonology and OT syntax. But this should not prevent us from asking whether marked meanings have something more in common than merely the property of violating a markedness constraint on meaning. According to Haspelmath (2006), many instances of markedness can be explained by frequency or rarity in texts. For forms, relative frequencies can easily be determined on the basis of a corpus of text. For meanings, it is a lot more difficult to determine their relative frequencies, because an automatic search in an unannotated corpus only gives us information about the forms and not their meanings. To determine frequencies of meanings, we have to look at semantically annotated corpora and results from comprehension experiments, which unfortunately do not yet cover all areas of interpretation discussed in this book. However, by looking at the markedness constraints on meaning proposed in this book we may estimate whether reducing markedness of meanings to frequency of meanings is a viable option. In Chapter 3 we introduced the markedness constraint on meanings *CONTRAST: ‘Pronouns refer anaphorically and non-contrastively’. Intuitively, the effects of this constraint seem to be compatible with the pattern of relative frequency: Non-anaphorically and contrastively used pronouns seem to be less frequent than

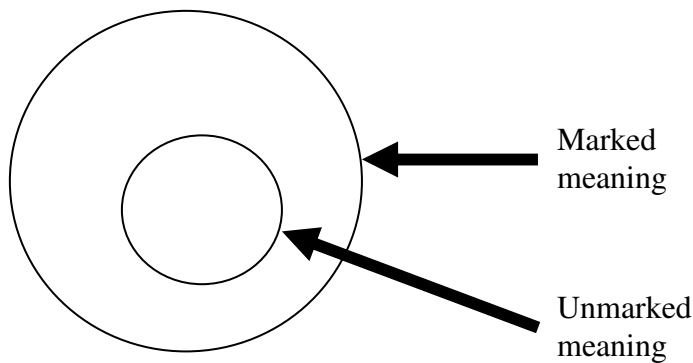
anaphorically and non-contrastively used pronouns. However, we do not have any statistical data yet to support this intuition. The same holds for some of the other markedness constraints on meanings introduced in this book, such as the constraint *NEG ('Avoid negation in the output') in Chapter 7. With respect to the constraint 'Subjects outrank objects in referentiality' discussed in Chapter 4, on the other hand, there is statistical evidence supporting the generalization expressed by the constraint (Dahl & Fraurud, 1996; Zeevat & Jäger, 2002). So marked meanings, as identified by markedness constraints on meaning, indeed seem to be rarer than unmarked meanings, although empirical evidence is not yet available for each case.

If the empirical evidence turns out to support the claim that marked meanings are rarer than unmarked meanings, does this imply that we can do away with markedness constraints on meaning in Optimality Theory, and replace the notion of markedness by the notion of frequency? The answer need not be positive. First of all, if there is a correlation between markedness of meaning and frequency, this does not mean that the observed linguistic pattern is the result of the statistical properties of the language. It may very well be that the statistical pattern of meanings and the observed linguistic pattern are both caused by some third factor (for example, our cognitive architecture which influences the way we perceive and conceptualize the world around us). On the basis of our current knowledge, we cannot distinguish between these two possibilities. Secondly, the markedness constraints are part of a symbolic system of linguistic representations and linguistic constraints on these representations, and as such interact with faithfulness constraints. If we would replace markedness constraints with the subsymbolic notion of frequency, it is unclear how this subsymbolic property would interact with faithfulness constraints at the symbolic level of computation.

Besides their possible correlation with rarity, there seems to be another property that many (though not all) marked meanings seem to have in common. As was pointed out in Chapter 8, in many cases the unmarked meaning is the strongest meaning possible (often the stereotypical or prototypical meaning). For example, the strongest meaning of the word *through* is the meaning according to which every point of the path is in the object referred to by the prepositional object. In Chapter 8, it was already noted that this seems to be a specific instance of the general relation of implication between meanings which determines Horn scales (Horn, 1972) underlying scalar implicatures. The strongest (most specific) meaning of *through* implies all weaker (more general) meanings of *through*. Similarly, the stronger meaning of *all* on the Horn scale $\langle all, some \rangle$ implies the weaker meaning of *some* ('at least

one and possibly all'). In general, many unmarked meanings stand in a relation of implication to the corresponding marked meaning:

Figure 1: Unmarked meanings imply marked meanings



Unmarked, stronger, meanings tend to imply marked, weaker, meanings. In other words: An unmarked, stronger, meaning excludes more possible situations than a marked, weaker meaning does. If the stronger meaning is available, the weaker meaning is blocked. As a result, the weaker meaning can only be used elsewhere. However, in Chapter 8 also a counterexample was given to this general pattern: With respect to spatial prepositions, the weakest meaning (referring to a quarter of a circle) seems to be the unmarked meaning associating with the unmarked form *om*, whereas the strongest meaning (referring to a full circle) is the marked meaning associating with the marked form *rond(om)*. So how general is this relation between markedness and strength of meaning? And what is the relation between strength of meaning and stereotypicality or prototypicality? Can stereotypicality and prototypicality be viewed as implication relations between meanings as well? Or are stereotypicality and prototypicality alternative ways to determine the strongest meaning? In the latter case, there would not be one particular notion of markedness of meaning, but rather several ones.

Also, relating the issue of implications between meanings to the previous issue of frequency of meanings, are stronger (unmarked) meanings more frequent than weaker (marked) meanings? For example, is the exhaustive meaning expressed by *all* more frequent than the literal meaning of *some* ('at least one and possibly all')? Obviously, this is not the case because the set representing the meaning of *all* is a subset of the set representing the literal meaning of *some* (cf. Figure 1). Or do we have to compare the frequency of occurrence of the meaning of *all* to the frequency of occurrence of the bidirectionally strengthened

meaning of *some* ('at least one but not all'), which is the complement of the meaning of *all* obtained by subtracting the meaning of *all* from the literal meaning of *some*? So which frequencies should be compared? This complication arises because competing meanings can, and in this case do, overlap. But even if we compare the frequency of the exhaustive meaning 'all' to the frequency of the bidirectionally strengthened meaning 'at least one but not all', it is not obvious that the exhaustive meaning is the more frequent meaning. Similarly, stereotypical or prototypical meanings may be less frequent than non-stereotypical or non-prototypical meanings. So when looking at potential meanings, we may need pragmatic implicature as an independent factor, resulting in the strengthening or enrichment of meaning. Whether markedness of meaning can be fully explained by strength of meaning, perhaps in combination with frequency of meaning, remains to be seen.

Summarizing, it is not easy to define what exactly must be understood by a marked meaning. Perhaps markedness of forms can be reduced to frequency of forms, but it is not immediately obvious how markedness of meanings can be reduced to frequency of meanings. In this section, we discussed the relation between markedness of meaning, frequency of meaning, and implications between meanings, and pointed out possible connections and potential differences. In the next section, we focus on the relation between forms and meanings. Here, we discuss the different notions of bidirectionality employed in the book and the questions they evoke.

9.3 Types of bidirectionality

One of the main results of this book is that speaking and understanding often require the language user to take into account the opposite perspective as well. Hearers take into account the alternative forms the speaker could have used but did not use. Similarly, speakers may take into account the way a hearer will interpret potential forms. This bidirectional perspective restricts the possible forms and meanings of a language, and was shown to influence the adult pattern of forms and meanings within a particular language as well as semantic variation among languages and the organization of forms and meanings in the mental lexicon.

In different chapters, we exploited different versions of bidirectional optimization. Strong bidirectional optimization (Blutner, 2000), which is the non-recursive variant allowing for total blocking only, is used to account for children's acquisition pattern of contrastive

stress discussed in Chapter 5. Weak bidirectional optimization (Blutner, 2000), which is the recursive variant giving rise to partial blocking and Horn's division of pragmatic labour, is shown to account for *when*-clauses in Chapter 2, scrambling of Dutch pronouns in Chapter 3, Dutch children's use of indefinite noun phrases in Chapter 4, bare singular nouns in Chapter 7, and prepositional patterns in Chapter 8. The adult patterns in these cases all conform to Horn's division of pragmatic labour according to which unmarked forms are used for unmarked meanings, and marked forms are used for marked meanings. A version of bidirectional optimization which is more or less in between strong and weak bidirectional optimization is the evolutionary bidirectional learning algorithm of Jäger (2004). In this asymmetric version, forms are disqualified as candidates if they are not recoverable as the intended meaning *and* at least one other form is. This type of conditional bidirectional optimization is employed in Chapter 6 to account for the cross-linguistic pattern of negation. Thus the book shows applications of various types of bidirectional optimization.

Now which of these types of bidirectionality adequately models our linguistic knowledge and linguistic behaviour: strong bidirectionality, weak bidirectionality, or Jäger's bidirectionality? Blutner and Zeevat (2004) argue that strong bidirectional optimization corresponds to the synchronic process of language acquisition (in particular, it "corresponds to the equilibrium established by the OT-learning algorithm" (p. 15)), whereas weak bidirectional optimization describes the direction of the diachronic process of language change. One way to implement this view is to assume that adults, using their grammar as a basis, consciously reason about alternative forms and meanings. When a certain weakly optimal form-meaning pattern becomes firmly established in the adult language, this pattern could then become a target for children's language acquisition. If children succeed in learning this pattern, this initially pragmatic pattern has become a semantic pattern that is 'wired' into the language. From this point on, we would expect this pattern to be acquired through the regular process of language acquisition, which can be modeled in OT by means of constraint re-ranking in combination with robust interpretive parsing (Tesar and Smolensky, 1998). This learning algorithm is also in a sense bidirectional, since it combines productive parsing with interpretive parsing. The algorithm produces a surface form by means of productive parsing, and then takes this form and optimizes over all structural descriptions ('underlying forms') yielding this surface form. If the outputs of productive parsing and interpretive parsing are different, constraint demotion takes place. The result of this regular process of language acquisition is, according to Blutner and Zeevat, symmetrical and describable in terms of strong bidirectional optimization.

However, there are two potential problems with this view of the relation between bidirectionality and language acquisition. First, OT learning on the basis of constraint re-ranking (such as in the OT-learning algorithms proposed by Boersma and Hayes (2001) and Tesar and Smolensky (1998)) does not always result in a symmetrical pattern describable in terms of strong bidirectional optimization. An example can be found with children's acquisition of personal pronouns such as *him* and *her*. If children below age 7 encounter a pronoun, they seem to guess between a coreferential and a disjoint interpretation. For adults, on the other hand, an object pronoun must be disjoint to the subject of the same clause. In the sentence *Ernie hit him*, the pronoun *him* cannot refer to the subject *Ernie*. Although children seem to guess when trying to determine the meaning of a pronoun, their *production* of pronouns is adultlike. Hendriks, van Rijn, and Valkenier (in press) point out that in an OT model describing this pattern, the two candidate meanings (the coreferential interpretation and the disjoint interpretation) must have the same constraint profile and must violate and satisfy the same constraints. But if the constraint profile is the same, re-ranking the constraints does not yield a different output. As a result, constraint re-ranking is not sufficient to explain the acquisition of the adult pattern on the basis of the children's pattern. Even under the adult constraint ranking, production and comprehension may yield different results. The same type of reasoning can be applied to other language phenomena where children show a guessing pattern in comprehension but not in production, such as the interpretation of contrastive stress discussed in Chapter 5. Thus the task of a child learning her grammar not only involves learning the adult constraint ranking, but also involves learning to take into account the opposite perspective. That is, the child must also learn to optimize bidirectionally. Hendriks et al. model bidirectional optimization as a step of optimization from form to meaning followed by a step of optimization from meaning to form. The results of their computational simulation suggest that children already know how to optimize bidirectionally, but simply lack sufficient processing speed to perform both steps within a reasonable amount of time. Under this account, children do not have to 'learn' to optimize bidirectionally, but merely have to gain sufficient processing speed to perform bidirectional optimization. But whether this is a correct characterization of the course of language acquisition or not, children's acquisition pattern with personal pronouns and contrastive stress shows that bidirectional learning should be distinguished from bidirectional optimization. Contra Blutner and Zeevat, we argue that strong bidirectional optimization is not merely a formalization of the result of OT learning.

Thus strong bidirectional optimization seems necessary for describing certain patterns in adult language. But is strong bidirectional optimization also sufficient for accounting for

the synchronic patterns found in language? A second problem with Blutner and Zeevat's view is that, although strong bidirectional optimization is able to account for the acquisition of pronominal meanings and the acquisition of contrastive stress discussed in Chapter 5, there may also be semantic phenomena which require the recursive version, weak bidirectional optimization. The acquisition of scrambled indefinite noun phrases in Dutch was analyzed in Chapter 4 in terms of weak bidirectional optimization. It remains to be seen whether it is possible to reformulate this analysis in terms of strong bidirectional optimization. If not, then a mechanism as powerful as recursive weak bidirectional optimization is required for the acquisition of the grammar.

A third version of bidirectional optimization employed in this book is Jäger's version of bidirectional optimization. According to this version, the grammar is asymmetrical: The speaker takes into account the hearer, but the hearer does not take into account the speaker (see also Buchwald, Schwartz, Seidl, and Smolensky, 2002; Wilson, 2001). Strong and weak bidirectional OT, on the other hand, are symmetrical: The speaker takes into account the hearer, and the hearer likewise takes into account the speaker. The position that hearers take into account speakers is in line with previous studies that have pointed out that adopting a symmetrical version of bidirectional OT allows for a principled explanation of scalar implicatures and other phenomena requiring Gricean reasoning (Blutner, 2000). Furthermore, this position is corroborated by the analyses discussed in this book. We presented several examples from different areas of the grammar that we argued to require that the hearer take into account the speaker when interpreting the sentence. But note that Jäger's version of bidirectional optimization was specifically designed to account for language change. It is conceivable that the factors influencing language change are different from the factors influencing language use and language acquisition. In particular, it may be the case that speakers actively change the language, whereas hearers merely adapt to language changes. This, however, requires further investigation.

After this discussion of the architecture of the bidirectional optimization model we return to another fundamental issue in Optimality Theory, namely the assumption of strict domination among constraints.

9.4 Cumulativity

In Optimality Theory, the output candidate that violates the lowest ranked constraints is favored over output candidates that violate higher ranked constraints. Violations do not add up. That is, more violations of lower ranked constraints cannot be stronger than one violation of a higher ranked constraint. Also, if a lower ranked constraint is violated more often, it cannot overrule the violation of a higher ranked constraint. Both principles are illustrated in the abstract tableau below:

Tableau 1: OT Example

Input	Constraint 1	Constraint 2	Constraint 3
Candidate a	*!		
Candidate b		*!	*
☞ Candidate c			***

So candidate (c) is the optimal candidate, although it violates constraint 3 three times, because this is better than violating constraint 1 once (as candidate (a) does) and better than violating both constraints 2 and 3 (as candidate (b) does).

Cumulativity in standard OT is thus rejected. Jäger and Rosenbach (2006), however, argue that cumulativity is necessary to account for probabilistic variation found in actual language use. They distinguish two types of cumulativity. Two weak constraints can gang up to jointly beat a stronger constraint (“ganging-up” cumulativity) and one violation of a stronger constraint can be less severe than more violations of a weaker constraint (“counting cumulativity”). All three candidates could come out as the winner of the competition. Candidate (b) could be the winner if violating constraints 2 and 3 each once is less severe than violating the stronger constraint 1 once, but also less severe than violating constraint 3 three times (counting cumulativity). Candidate (a) could be the winner if both violating constraints 2 and 3 and violating constraint 3 three times is worse than violating the stronger constraint 1 once (ganging-up cumulativity). Several experimental studies pertaining to the grammatical variation of genitive constructions in English reveal that both types of cumulativity occur. They argue that maximum entropy models (or, log-linear models) (e.g., Abney 1997) are superior to Boersma’s (1998) version of stochastic OT in dealing with cumulativity.

Jäger and Rosenbach (2006) acknowledge that the predecessor of OT, Harmonic Grammar, has a very similar mathematical setup to maximum entropy models (Legendre et al. 1990). Harmonic Grammar is the numerical predecessor of OT in which two constraints combined can override one stronger constraint. And if two constraints compete and one is weaker than the other but activated to a higher degree, this one can still win. The shift from Harmonic Grammar to Optimality Theory was mainly empirically motivated by phonological data, while Harmonic Grammar was only applied once, in the domain of syntax and semantics. The input-output mappings in phonology are different from those in syntax and semantics. In phonology, on the one hand, the input is an underlying form (stored in the lexicon), while the optimal output is also a form, the one that is actually produced. The constraints either restrict the complexity of the outputs (markedness constraints) or they restrict the distance between the two forms (faithfulness constraints). In syntax and semantics, on the other hand, a relation between form and meaning is established, and we expect more variation both in grammaticality and in interpretation, since possible forms (sentences) and their meanings are not stored in the lexicon, as phonological words are. The question is whether Harmonic Grammar or maximum entropy models are also superior in explaining other empirical data (that is, apart from the variation in genitive constructions in English). We expect this to be the case, especially in the domain of syntax, semantics, and pragmatics. Smolensky and Legendre (2006) discuss the pros and cons of Harmonic Grammar versus OT and make a difference between constraints that are strictly grammatical and constraints that are more pragmatically based. The former would be better modeled in terms of strict domination, while the latter would interact in a less restricted manner, via arbitrarily weighted constraints. It is still an open issue to us whether a model that can handle cumulativity, such as Harmonic Grammar, is to be preferred over the OT perspective of strict domination of constraints, but at least we believe Jäger and Rosenbach (2006) made a strong case for the existence of cumulativity in natural language variation.

In the next two sections, we return to the relation between Optimality Theory and the field of cognitive science. In Section 5 we relate the acquisition of bidirectional optimization to the acquisition of social cognition. In Section 6 we discuss the possibility of grounding meaning distinctions in human cognition.

9.5 Theory of Mind

For a language user, applying bidirectional optimization requires an insight in the intentions and decisions of the other participant to the conversation, and therefore, a Theory of Mind. Not only should the language user know the expressive options of the language, but she should also understand why the speaker (or hearer) decides for one, rather than another option. Thus, assuming that bidirectionality is implicated in language change (Blutner & Zeevat, 2004), means assuming that social cognition is partly responsible for the state of the grammar. In this book, we have proposed that a failure to apply bidirectional optimization lies at the root of several delays in language acquisition. Bidirectionality then becomes a key notion in explaining the development of grammar, both in the individual who still needs to acquire bidirectional optimization, as well as the development of a grammar in an entire language community.

Theory of Mind has enjoyed the attention of students of language acquisition for some time. Bloom (2000) and Tomasello (2003) point out how social cognition may be instrumental in, or indeed largely responsible for, the acquisition of language, in particular reference and labelling. We add to this a different proposal: That social cognition plays a key role in determining the grammar itself.

When considering the relation between language and Theory of Mind, we must distinguish two ways in which Theory of Mind can be approached. On the first approach, Theory of Mind is undivided and generalizable: One either has a Theory of Mind, or one does not. For instance, an individual is said to have a Theory of Mind if she passes a False Belief test, for which an understanding of others' knowledge or views is required. On such a view of Theory of Mind, one might argue that Theory of Mind development is not likely to be implicated in the various language acquisition delays discussed in this book. These delays disappear at different ages, in many cases well after the age at which children regularly pass False Belief tasks, hence "acquisition" of a Theory of Mind cannot be the cause of acquisition of these grammatical phenomena.

It is becoming increasingly clear, however, that this notion of Theory of Mind is too simplistic. Whether Theory of Mind is displayed strongly depends on the nature of the task. For instance, it has been fairly standardly assumed since 1983 that the False Belief task (Wimmer & Perner, 1983) indexes Theory of Mind. In this task, an individual needs to decide where someone will, mistakenly, look for a hidden object. Recently, however, it was found

that 15-month old children already show an awareness of Theory of Mind. In this task (Onishi & Baillargeon, 2005) children did not need to make a decision, but rather their looking times were measured, showing that the children distinguished between a person looking in a place where the child knows the object is hidden, or in a place where the person in question believes the object to be hidden. Another striking discrepancy between two tasks was found by Flobbe (2006). She showed that children who successfully passed a second order False Belief task (“she knows that he believes...”) are for the most part unable to apply second order Theory of Mind in a strategic game (“my opponent will not make a certain move, because he knows that I will then make a move which will curb his plans”). In fact, many of these children were unable even to apply first order Theory of Mind in the game (“my opponent will make a certain move, because it is advantageous to him”). These findings show the relevance of the following question: How is Theory of Mind integrated with other areas of knowledge, so as to meet different task demands? Language provides an excellent arena for studying this question. Further investigation of the relation between social cognition and bidirectional optimization will contribute both to our understanding of language, and the nature of Theory of Mind.

9.6 Cognitive plausibility

Optimality Theory was conceived as a way to breach the gap between two fundamental paradigms in the study of cognition: the symbolic approach (in which cognitive computation consists of the manipulation of symbolic representations), and the connectionist approach (which replaces this with spreading activation in a neural network). The symbolic level is grounded in a subsymbolic level of neural associations, making the symbols and their manipulations meaningful and allowing for a notion of optimality that is rooted in the connectionist property of harmony (Prince and Smolensky 1997). Nevertheless, within this attractive overall perspective, important questions remain about the specific grounding of linguistic distinctions, constraints, and generalizations. The cognitive plausibility of a linguistic theory depends on its potential to link, where possible, the existence and nature of its theoretical entities to language-independent facts about human perception and communication. It has been argued that basic aspects of the phonological system follow from phonetic properties of articulation and perception (Boersma 1998, Hayes 1999). When dealing with meaning, we also need to consider what parts of the semantic system can be independently grounded.

There is much work about grounding in the study of language and cognition, but we will point out one connection here that we believe is especially interesting, because it resonates at a deep level with the central concerns of Optimality Theory about cognitive architecture, optimality, learnability, and variation. This is the connection with the theory of Conceptual Spaces of Gärdenfors (2000). Gärdenfors proposes to bridge the gap between the neural subsymbolic level and the higher symbolic level through a geometrical level of conceptual spaces. A conceptual space defines a similarity relation between entities with respect to a number of quality dimensions, such as colours in the colour space (with its dimensions of hue, saturation, and brightness), or vowels in the two-dimensional vowel space. Such conceptual spaces are deeply rooted in our cognitive apparatus, and therefore provide the necessary grounding of categories and concepts defined over them. These concepts are modeled as regions in a conceptual space, subject to a very strong constraint of convexity. A concept is convex if and only if for every two points P and Q in the concept, every point between P and Q is also in the concept. This property is very important for concept formation, induction and learning, as Gärdenfors argues and it is part of an optimal partition of meaning spaces into natural linguistic categories, as he illustrates for the domain of colour terminology.

In recent work, Jäger and Van Rooij (2006) have demonstrated that this convexity property can be explained in an evolutionary game-theoretical setting, as resulting from an evolutionary stable state in the dynamics between speaker and hearer, given the pressure to minimize the semantic distance between what the speaker intends and what the hearer understands. And this brings us back to the bidirectional dialectic between speaker and hearer that plays such an important role in this book. The connections between Optimality Theory, Game Theory and Conceptual Spaces open up new avenues for linguists, philosophers and cognitive scientists to understand more about how the relation between linguistic form and conceptual meaning is shaped through the communicative function of language.