Conflicts in Interpretation

Gerlof Bouma, Petra Hendriks, Helen de Hoop,
Irene Krämer, Henriëtte de Swart, and Joost Zwarts

University of Groningen, Radboud University Nijmegen, Utrecht University

0. Introduction[1]

In this article we take an Optimality Theoretic approach to interpretation which integrates various factors into a set of typically conflicting constraints of varying strengths. The hypothesis that optimization is a leading principle in natural language interpretation strengthens the connection between linguistic theory and other cognitive disciplines. We will provide further support for this view from experimental research on computational and human sentence processing as well as on the acquisition of interpretation. We claim that our approach opens new perspectives for the study of natural language interpretation for several reasons. It has a clear view of the general relation between universal and language-specific properties (the constraints are universal, but the ranking of the constraints is language-specific). It is not modular, which allows us to locate variation in meaning not only in the semantic component proper, but in the syntax-semantics interface and the semantics-pragmatics interface. It can explain cross-linguistic variation in meaning by investigating how languages vary with respect to their weighting (ranking) of the constraints. Finally, the development of a bidirectional Optimality Theory allows us to separate the question of how to (best) formulate what you want to say, from the question of how to (best) interpret something that has been said, and this also provides us with a straightforward explanation of some facts concerning children's interpretations of natural language.

[1] In this article we report some results of the first two years of our project 'Conflicts in Interpretation', funded by the Netherlands Organisation of Scientific Research (NWO), which is gratefully acknowledged.

By considering interpretation as a process of conflict resolution, and the preferred meaning as the optimal meaning for a given form within a specific context, it is possible to account for the influence of context on interpretation. Because the constraints are applied simultaneously, the possibility is opened up for cross-modular constraint interaction. Contextual influence on interpretation is generally considered an interface phenomenon that does not pertain to the grammar itself. However, in section 1 we will show that context also plays an important role in lexical semantics. In Optimality Theory the procedure that provides us with an optimal interpretation of a given word within a certain context can be viewed in two different ways. The first approach combines the view of radical underspecification with a mechanism of contextual enrichment. This approach is taken, for example, in Blutner (2000). The second approach takes the opposite position in crucial respects. Rather than strengthening a weak (underspecified) meaning with contextual knowledge, we may take as our point of departure the strongest possible meaning and have it weakened by contextual information. This is the approach advocated in section 1.

Although semantic interpretation is not a simple process, people seem to accomplish it without any effort. For example, pronouns are extremely underspecified in their lexical semantics but interpreting them – i.e., finding their discourse referent – is hardly ever a problem. The different sorts of information that become available during language comprehension are very diverse, varying from lexical and syntactic information to information from context and world knowledge. In section 2 we will first present a computational model of pronoun resolution, based on Optimality Theoretic semantics. Secondly, we turn to human sentence processing and argue that the optimal interpretation of a sentence is being built up word by word (incrementally). We will show that patterns of constraint violations correspond to differences in cognitive processes as measured in experimental research on human sentence processing.

One of the strong points of Optimality Theory (OT) in phonology and syntax is its ability to account for cross-linguistic variation with respect to linguistic forms. Under a bidirectional view on optimization, differences in form can be associated with differences in meaning in an intuitively plausible way. For instance, in double negation languages it

is important to interpret each negative expression as contributing a semantic negation, whereas in negative concord languages it is more important to mark arguments of a negative chain formally. As a result, negative expressions are interpreted differently in these languages. In section 3 we will argue that these differences do not follow from differences in meaning per se, but arise because the languages rely on a different balance between form and meaning.

If bidirectional optimization is crucial for correctly producing and interpreting linguistic expressions, then it is expected that children will have to acquire this optimization strategy in the course of language acquisition. In section 4 we will argue that when adults interpret a sentence, they do not only take into account the form of the sentence, but also the other possible forms the speaker could have chosen. Children, however, start out by learning to associate unmarked forms and unmarked meanings, they optimize unidirectionally. Only later, the possibility of bidirectional optimization, i.e., of associating marked forms with marked meanings, seems to be acquired.

1. Conflicts in the interpretation of a word

In order to determine the meaning of a given sentence, we first need to determine the meanings of each of the words of the sentence. Not only can the words of a sentence interact and conflict with each other, they also interact with the context in which they are uttered. Two major problems in the study of word meaning are *polysemy* (words have multiple, related senses) and *flexibility* (words adapt their meaning to the context in which they are used). In this section we apply the idea of Optimality Theoretic conflict resolution to the domain of lexical semantics.

1.1 The polysemy of round

The problems of polysemy and flexibility can be illustrated with the preposition and adverb *round* (Hawkins 1984, Schulze 1991,1993, Taylor 1995 and Lindstromberg 1998). Some typical senses are illustrated in (1) and Figure 1 for the preposition *round*

and in (2) for the adverb.

(1) a. The postman ran round the block
b. The burglar drove round the barrier
c. The steeplechaser ran round the corner
d. The captain sailed round the lake
e. The tourist drove round the city centre

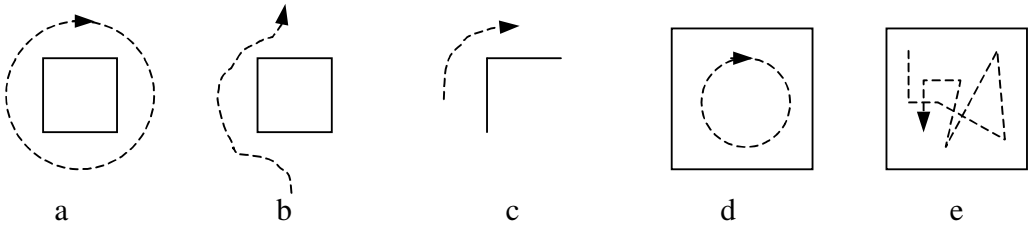

Figure 1: Paths corresponding to the preposition *round*

(2) a. The driver took the long way round (i.e. making a detour)
b. The woman came round again (i.e. back to her point of departure)
c. The picture was turned round (i.e. so as to face the other way)

The first question is how we derive these different senses, the second question is how we get them in the right contexts. We do not want just to list the senses and allow them to occur anywhere, but find out how they hang together and how they depend on the meanings of neighbouring words (like *corner* in (1c)). In the cognitive semantic literature the structure of polysemy has been a central concern (e.g. Lakoff 1987), but the context-dependence has not been studied much. We would like to show here how both aspects can be insightfully modeled by a combination of model-theoretic semantics and Optimality Theory (Blutner 2000, Hendriks and de Hoop 2001, de Hoop and de Swart 2000, Zeevat 2000). The basic idea is that *round* has an underlying prototype meaning (the 'circle') that can be formally modeled in terms of paths. Candidate meanings are generated from this prototype in a systematic way and a system of ranked constraints takes these candidates and selects the best meaning for a particular context. The full story can be found in Zwarts (2004).

1.2 The underlying prototype of round

We assume that the underlying prototype meaning of *round* corresponds to a circle, modeled here as the set of *paths* that describe exactly one perfect circle, with different radii. Let's call this set CIRCLE. Here is an impression of what a circular path looks like:

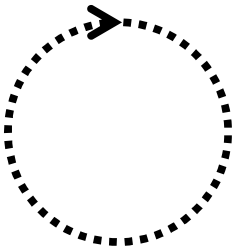

Figure 2: A prototypical *round* path

The path can be a path of motion, with a reference object in the middle (the preposition in (1a)) or without such an object (the adverbial counterpart *The postman ran round*). It can also be used to represent *extension* and *rotation*:

(3) a. Mary has a necklace round her neck
b. John turned the wine glass round in his fingers

In (3a) the necklace is distributed along a circular path round Mary's neck. For (3b) the path can describe the rotation of the wine glass around a vertical axis.

1.3 The candidate meanings of round

The next step is to generate a set of candidate meanings from this underlying prototype meaning. The central assumption here is that every property of the CIRCLE prototype is a candidate meaning. One such property is COMPLETENESS:

(4) COMPLETENESS

There is a point of the path in every direction from the centre.

Not all paths with COMPLETENESS are circles. Spirals and ellipses are not circles, but they do have the property of COMPLETENESS. What they are lacking is another property that circles have:

(5) CONSTANCY
Every point of the path has a constant distance to the centre.

Notice that an arc has CONSTANCY but not COMPLETENESS. Only perfectly circular paths have both COMPLETENESS and CONSTANCY. In the following examples *round* has COMPLETENESS without CONSTANCY:

(6) a. The earth goes round the sun in one year (elliptical path)
b. There is a wall round the garden (rectangular path)
c. The planet spirals round towards its sun (spiral path)
d. The tourist drove round the city centre (crisscross path)

Circular paths also satisfy the following two properties, weaker versions of COMPLETENESS:

(7) INVERSION
There are points of the path at opposite sides of the centre.

Paths with this property will be at least semicircular, at illustrated in the following examples:

(8) a. The burglar drove round the barrier (Fig 1b)
b. The children sat round the television
c. The car turned right round

A still weaker property is ORTHOGONALITY:

(9) ORTHOGONALITY

There are points of the path at perpendicular sides of the reference object.

(10) a. The steeplechaser ran round the corner (Figure 1c)

b. A man put his head round the door

c. John turned round to the woman sitting next to him

In each of the sentences in (10) there is a change of position or direction from one side to an orthogonal side, not the opposite side.

The CIRCLE prototype also implies DETOUR:

(11) DETOUR

The length of the path is longer than the distance between starting point and end point.

This property holds of every path that does not form a straight line between its starting point and end point (example from Schulze 1991):

(12) The bridge is damaged, so you will have to go round by the lower one

LOOP is a property that paths have when their starting point and end point are identical:

(13) LOOP

The starting point and the end point of the path are identical.

Circular paths have this property, but when a woman had been visiting a friend down the road and came back we can also say:

(14) She came round again

This gives us an ordered set of candidate meanings generated for *round* (or rather, that part of the meanings that involves the shape of the path), each line representing a subset or implication relation holding between two meanings:

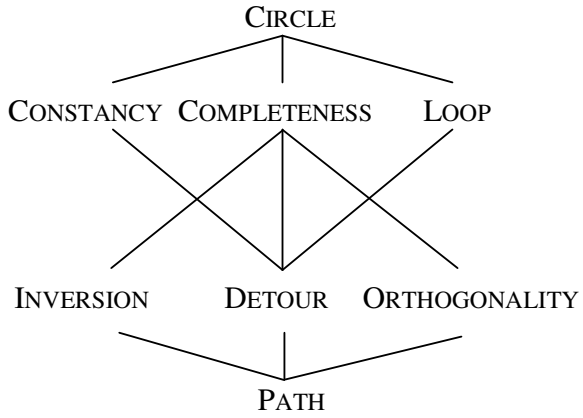

Figure 3: Strength of *round* properties

Also included at the bottom is the most general meaning PATH, represented by the set of all paths. When we would only consider paths with CONSTANCY, then the partial ordering becomes a strict ordering:

(15) CIRCLE/LOOP > COMPLETENESS > INVERSION > ORTHOGONALITY > DETOUR

This orders the meanings on a scale from the strongest meaning (a perfect complete circle) down to the weakest meaning (anything that is not a straight line).

1.4 The flexibility of round

Now that we have some idea of how the meanings of *round* hang together, we can tackle the issue of getting them in the right context, which is where Optimality Theoretic constraint interaction comes into play. The assumption is that lexical semantic flexibility is the result of the interaction between two constraints:

(16) STRENGTH: stronger interpretations are better than weaker interpretations

(17) FIT: interpretations should not conflict with the (linguistic) context.

STRENGTH is a constraint that favours stronger interpretations over weaker interpretations (Blutner 2000, Zeevat 2000), and FIT favours interpretations that do not give rise to a contradictory or unnatural reading (similar constraints are AVOID CONTRADICTION in Hendriks & de Hoop (2001) and CONSISTENCY in Zeevat (2000). If FIT is ranked over STRENGTH, we get the effect that a weaker non-contradictory meaning wins over a stronger contradictory meaning.

The following *tableau* illustrates how this works for the PP *round the door*:

round the door	FIT	STRENGTH
COMPLETENESS	*	
☞INVERSION		*
ORTHOGONALITY		**
DETOUR		***

Table 1: An OT tableau for the interpretation of *round the door*

The upper left corner of the table gives the input. Underneath this input are four possible interpretations of that phrase relevant for the discussion. The two columns to the right show the two constraints on the interpretation of *round the corner* in their ranking and to what extent the candidate interpretations satisfy these constraints. The COMPLETENESS interpretation violates FIT, as indicated by the asterisk under FIT, because the fact that a door is usually connected to a wall makes it impossible to have a complete path round it (see Figure 4). STRENGTH is violated to different degrees by the four candidate interpretations: less asterisks under STRENGTH means a stronger interpretation.

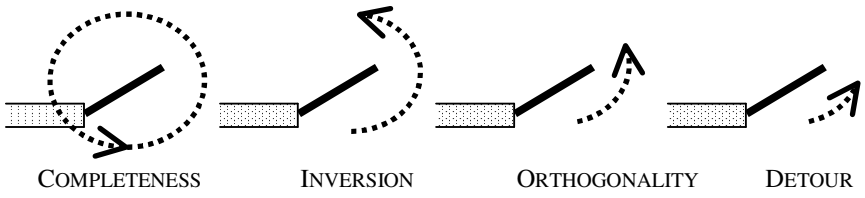

Figure 4: Four of the possible interpretations of *round the door*

The optimal interpretation of *round the door* is, as indicated by the pointing finger in the tableau, the interpretation that best satisfies the two contraints FIT and STRENGTH, namely INVERSION, which is the strongest interpretation that still fits.

To sum up, the preposition *round* all by itself will have the strongest meaning (CIRCLE), adding the object *the door* forces the interpreter to weaken the meaning to INVERSION and putting the PP in a sentential context (*put your head round the door*) further boils down the meaning to ORTHOGONALITY (because the length and flexibility of a human neck does not allow one to move one's head all the way to the other side of the door). ORTHOGONALITY will then be the strongest interpretation that still fits the sentence meaning as a whole, because INVERSION gives a violation asterisk under FIT. See Zwarts (2004) for more discussion about the role of context and for the role of a third constraint favouring weaker meanings, namely VAGUENESS (following Krifka 2002). What we see in general is that more context gives weaker meanings.

2. Conflicts in processing

In the previous section we studied the interaction of word meaning and the linguistic and extra-linguistic context in which the word was uttered. Notoriously, pronouns are by definition underspecified in their lexical meaning and therefore highly dependent on the surrounding context for their interpretation. When we broaden our empirical domain from word to sentence level, we see again overwhelming evidence that the context with its various sources of information (world knowledge, syntactic structure, lexical information, etcetera) plays an important part in determining the optimal interpretation. To capture the role of different types of information in sentence comprehension, we apply principles of OT semantics to natural language processing. The integration of semantic, syntactic, and

contextual information in a system of ranked constraints is proposed to correctly derive the optimal interpretations for pronouns, as discussed in section 2.1, and for transitive sentences in Dutch, as discussed in section 2.2.

2.1 Computational sentence processing: pronoun resolution

The many factors that the resolution of pronouns is sensitive to, suggest that it makes sense to view this as an optimization problem. Clearly, information from the different modules is used, that is syntactic information (such as agreement and configurational binding restrictions), semantic information (such as selectional properties of the verb that takes the pronoun as an argument), pragmatic information (such as the degree of activation or salience of a discourse referent), and world knowledge (which influences the plausibility of a certain reading). Moreover, all this information need not point in the same direction all of the time. Consider the following two sentences:

(18) Fitz killed Gerald. He was arrested near London.
(19) Fitz killed Gerald. He was buried near London.

Depending on the rest of the second sentence, the pronoun *he* in (18) refers to the subject of the first sentence, that is, *Fitz*, but not in (19). In sentence (19) we witness a conflict between a constraint that favours a pronoun to refer to the subject of the preceding sentence, and a constraint that deals with world knowledge (namely, that it is usually the killed man and not the killer that gets buried).

The computational pronoun resolution literature has long since recognized this fact and it is common for a resolution system to combine many factors in the resolution of a pronoun (see Mitkov 2004 for an overview). Such systems pick a referent from a candidate set of referents by scoring them using weighted factors or perhaps some more complicated statistical machinery. Of course, such systems easily deal with conflicting information, because it is the final score of a candidate that determines whether it will be selected, not whether it complies with all of the possible constraints.

A downside to such models is that *on their own* they are linguistically rather opaque. Most theoretical linguistic models do not allow for weighted factors and it is therefore not always clear to see what the implications from computational results should be in theoretical research. Also, a weighted factor model is rather unconstrained. It might for instance predict that if enough lighter factors point to some referent, a stronger factor can be overruled. Whether this is actually needed is an empirical matter; but it should be noted that this adds to the opacity of the resulting system.

Another line of computational research tries to remain closer to theoretical work. They implement rule based systems, typically inspired by Centering Theory (Grosz et al, 1995). However, as has been pointed out in Beaver (2004) and Byron and Gegg-Harrison (2004) it is hard to judge how much of the results depend on the algorithm rather than the principles that are taken from theoretical work. The algorithm dependence makes detailed comparison of systems hard and development time-consuming.

Computational results are interesting for theoretical linguists in particular because of the possibility to test a theory on large amounts of natural data. This has as a positive side effect that aspects of the phenomenon at hand can be studied that will not reveal themselves by introspection. Consider the special status of subjects as antecedents, as indicated above with respect to the preferred interpretation of sentences (18)-(19). It is commonly claimed that a referent recently realized as a subject is preferred as a referent for a following pronoun. However, with constructed examples, this effect is very hard to show convincingly, especially if it is compared with the effects of world knowledge, as discussed above, or agreement, as illustrated below.

(20) Ella met Fitz in the park when he was on his daily workout.
(21) Fitz met Gerald in the park when he was on his daily workout.

In (20) it is clear that Fitz was on his daily workout, because *he* can only refer to Fitz, and not to Ella. This would be in accordance with the view that a constraint that requires the pronoun and its antecedent to agree (in gender and number) outranks the constraint that favours the antecedent to be the subject of the preceding sentence. In (21), both readings

of the pronoun are equally possible, despite the claimed subject effect. However, as can be seen in any corpus based evaluation -- for instance for the system described below -- the subject effect clearly shows up in the statistics: a model that includes a preference for subjects as antecedent performs better.

Bouma (2003) shows how a relatively simple OT based pronoun resolution model can be implemented and evaluated on a corpus. Using OT, the disadvantages of computational modelling can partly be dealt with. OT is a competition based model and therefore naturally deals with different information sources and conflicts between these sources. It is however more restricted, in that is explicitly rules out the case in which lower constraints together take on a higher constraint. As a matter of fact, Prince and Smolensky (2004) argue that this is a desired result for a theory of grammar. Another advantage of using OT is that there is a close if not direct connection with theoretical models. There is no a priori reason why constraints and ranking arguments from computational OT work should not carry over to theoretical OT work (and vice versa). An advantage over the rule-based systems is that OT models in principle are completely declarative – after all, OT is a constraint based framework.

Bouma (2003) develops a unidirectional OT semantics model; the input is a pronoun that needs to be resolved; the candidate outputs are possible antecedents. The pronoun is assigned the referent of the optimal antecedent. All other NPs receive a fresh discourse referent. Being a simple model it optimizes over just one pronoun at a time (see Beaver 2004 and Buchwald *et al.* 2002 for proposals of optimization over larger units, and Byron and Gegg-Harrison 2004 for an implementation of the former). If we take the constructed discourse in (22), the tableau for the pronoun in the third sentence might look like the one in Table 2.

(22) The pronoun resolution system moves through the text. New antecedents are picked up and added. It resolves pronouns incrementally.

it	AGREE	SUBJECT	DISTANCE
☞ the pronoun resolution system: *x1*			**
the text: *x2*		*	*
new antecedents: *x3*	*		

Table 2: An OT tableau for the interpretation of *it* in (22)

After this *it:x1* becomes a possible antecedent itself and the system continues through the text. The model is tested on a modest corpus of Dutch newspaper articles. For the implementation several robust computational resources are used, like the EuroWordNet lexical database and the Alpino grammar and parser for Dutch (Bouma et al, 2001).
Bouma shows how adding constraints and trying different rankings influences the performance of the model. For instance, adding first binding and agreement constraints and then a subject constraint leads to increase in pronouns that are connected to a coreferential antecedent. However, a constraint that prefers antecedents in the current sentence over ones in the previous sentence over ones in the sentence before (etc), makes things rather worse than better, mainly indicating that intrasentential antecedents are not necessarily preferred:[2]

1 DISTANCE	26% correct
2 BINDING, AGREEMENT >> DISTANCE	40% correct
3 BINDING, AGREEMENT >> SENTENCE >> SUBJECT >> DISTANCE	50% correct
4 BINDING, AGREEMENT >> SUBJECT >> DISTANCE	55% correct

Bouma (2003) finally reports a score of 68%. Given the relative shallow information that was used in the model, and given that it was fully automatic, this is comparable to other resolution models.

Of course, Bouma's model is very simple and the directions in which it should be

2 The figures should be taken as indications because there was no analysis of statistical significance of the results.

extended for it to be taken seriously as a linguistic model are clear. For example, even if we accept the fact that the model only deals with pronouns, leaving cataphoric pronouns out of the picture is a clear shortcoming. However, given the results of Bouma (2003) there is no reason to believe that the approach taken there could not be extended to include these and other phenomena.

The above model takes natural language interpretation to be incremental, i.e. that the information comes in bit by bit – and it is in fact *strictly* incremental in the sense that earlier choices are never reconsidered. This need not be the case, however. In the next subsection we present a non-strictly incremental model of subject-object disambiguation based again on OT semantics. We will show how the optimal interpretation at a certain point in time is the result of the optimization process until that point, but the next moment the optimal interpretation of the utterance may have changed again because other constraints have been activated that did not apply before. While interpreting a sentence, hearers may 'jump' in the course of time from one optimal interpretation to another.

2.2 Human sentence processing: subject-object disambiguation

In German, not only personal pronouns, but also articles and adjectives are overtly case-marked. If a sentence starts with an accusative case-marked NP, as is illustrated in (23), it will be identified as the object of the sentence. Thus case information helps to determine the interpretation.

(23)	Den Zaun	hat	der Junge	zerbrochen.
	[the fence]$_{ACC,3rd,sg}$	has$_{3rd,sg}$	[the boy]$_{NOM,3rd,sg}$	broken
	"The fence, the boy broke."			

Because Dutch has a relatively free word order and no case-marking on full noun phrases, many sentences are ambiguous. However, when due to morphological poverty of certain phrases a subject-object ambiguity occurs, the interpretation will be based on the strong preference for the canonical word order with its concomitant reading, i.e. the reading in which the subject precedes the object. In a sentence such as in (24), there is a

strong preference for interpreting the first NP as the subject. In other words, in the absence of conflicting information, the preferred interpretation is one in which the subject precedes the object.

(24) De patiënt heeft de arts bezocht.
[the patient]$_{SUBJ/OBJ}$ has [the doctor]$_{SUBJ/OBJ}$ visited
"The patient visited the doctor."

Most experiments that found evidence for the subject before object preference dealt with sentences where the subject and the object both were animate used in combination with agentive transitive verbs. However, several off-line and on-line experiments in Dutch and English showed that besides case-marking and word order, animacy information is an important source of information for the comprehension process as well (cf. Lamers and De Hoop 2005).

Although not as common as an animate subject, it is not only possible that the subject of a sentence is inanimate, it might also be that given an animate and inanimate noun phrase, the inanimate one has to be the subject of the sentence and the animate one the object. It is the verb that imposes selectional restrictions onto the arguments, as is illustrated in (25)-(26). Verbs such as 'bevallen' *please* take an experiencer (animate) object (25), while verbs as *like* take an experiencer (animate) subject (26).

(25) De vakantie beviel de jongen.
the holiday pleased the boy
(26) The boy liked the holiday.

So far, three constraints that play an important role in language processing follow from the above discussion. The first constraint is based on morphological case-marking, the second on a general preference of word order, whereas the third constraint is related to the selectional restrictions of the verb. The constraints can be defined as follows:

(27) CASE: The subject is nominative case-marked, the object is accusative case-marked.

(28) PRECEDENCE: The subject (linearly) precedes the object.

(29) SELECTION: Fit the selectional restrictions of the verb.

Following De Hoop and Lamers (to appear) and Lamers and De Hoop (2005), we assume that during human sentence processing, the optimal interpretation of a sentence (form) is being built up incrementally. Hence, we assume the process of optimization itself to be incremental. That is, the process of optimization of interpretation proceeds while the information comes in word-by-word, or constituent-by-constituent.

Before incremental optimization of interpretation can be used to analyze subject-object disambiguation, the three constraints that were defined above, have to be ranked. We determine the ranking by examining the optimal output interpretation in case of a conflict between constraints. Such a situation was illustrated in example (23) in which a conflict arises between CASE and PRECEDENCE. Since the optimal reading of (23) will be an object-initial reading, we may conclude that CASE outranks PRECEDENCE. Likewise, we conclude from the optimal reading of (30) below that SELECTION outranks PRECEDENCE.

(30) De jongen beviel de vakantie.
the boy pleased the holiday
"The holiday pleased the boy"

Finally, we assume that CASE is the strongest of the three constraints. This can be derived from the fact that we obtain sometimes a pragmatically odd yet optimal interpretation when SELECTION must be violated in order to satisfy CASE, as in a German sentence such as *Der Zaun hat den Jungen zerbrochen* glossed as 'the fence$_{NOM}$ has the boy$_{ACC}$ broken' ("The fence broke the boy"). In this sentence, despite the fact that transitive *break* normally selects an animate subject, the inanimate NP *der Zaun* 'the fence' is interpreted as the subject because of its nominative case. That is, the sentence can only mean that the fence broke the boy. In (31) the ranking of the three constraints is given.

(31) CASE >> SELECTION >> PRECEDENCE

Lamers (2001) investigated sentences such as given in (32), (33) and (34) in two ERP studies.

(32) De oude vrouw in the straat verzorgde hem …
The old woman in the street took-care-of him …
"The old woman in the street took care of him …"

(33) Het oude park in the straat verzorgde hij …
The old park in the street took-care-of he …
"He took care of the old park in the street …"

(34) De oude vrouw in the straat verzorgde hij …
The old woman in the street took-care-of he …
"He took care of the old woman in the street …"

When people start to interpret sentences (32)-(34), they interpret the initial noun phrase as the subject (in accordance with PRECEDENCE). However, when the verb is encountered, the inanimate initial noun phrase in (33) can no longer be interpreted as the subject because of the selectional restrictions of the verb. The sentences in (32) and (34), on the other hand, maintain the subject-initial reading as the optimal interpretation until the second noun phrase. However, when the nominative case-marked pronoun is encountered in (34), CASE overrules PRECEDENCE. The object-initial reading becomes the optimal interpretation for (34) then. Only sentence (32) can maintain the preferred subject-initial reading in the course of processing.

Event related brain potentials (ERPs) are small changes in spontaneous electrical activity of the brain that occur in response to certain sensory. They can be recorded by means of electrodes attached to the scalp. Because of the high temporal resolution of the ERP signal with its various dimensions, ERPs have been proven to be particularly useful to investigate the time-course of brain activity related to language processing. Lamers (2001) reports certain ERP effects (*viz.*, early and late positive peaks) at the verb for

sentence (33) starting with the inanimate NP in comparison to sentence (32) as well as at the nominative case-marked pronoun in sentence (34) in comparison to the accusative case-marked pronoun in (32). Strikingly, the effects are similar although different sorts of information are used in the two comparisons (the selectional criteria of the verb and the case-marking of the pronouns, respectively). We claim that this similarity can be explained within the incremental optimization model. The evaluation of these sentences against the set of constraints shows that the pattern of constraint violations is the same in the two different contexts.

First, let us have a look at the constraint violation pattern of sentence (32). In the tableau the interpretations at the three stages are given in the first three columns. The violation pattern given by the right three columns of the tableau reflects the pattern of only the final - third – stage (in the first two stages, CASE is not violated by the object-initial reading yet, only PRECEDENCE is). The subject-initial reading is maintained throughout the three stages of interpreting.

De oude vrouw… 'the old lady'	verzorgde… 'took care of'	hem… 'him'	CASE	SELECTION	PRECEDENCE
☞SI	☞SI	☞SI			
OI	OI	OI	*		*

Table 3: An OT tableau for the incremental interpretation (32)

At this point, consider Tableau 4 below that gives the incremental optimization of the interpretation of sentence (33). Up to the verb SELECTION does not play a role in the parsing process, since no relevant information is available. At the verb, however, it becomes clear that the subject has to be animate. Hence, the optimal interpretation of the initial inanimate NP changes from subject to object. The optimal (object-initial) interpretation clearly violates PRECEDENCE but the stronger constraint SELECTION is satisfied. The incremental interpretation of sentence (33) involves a 'jump' from the subject-initial to the object-initial interpretation. It is at this point in the sentence (i.e., at the verb) that Lamers (2001) found the significant ERP effects (early and late positivities)

for sentence (33) compared to (32).

Het oude park… 'the old park'	verzorgde… 'took care of'	CASE	SELECTION	PRECEDENCE
☞SI	SI		*	
OI	☞OI			*

Table 4: An OT tableau for the incremental interpretation (33)

Finally, Tableau 5 presents a schematic overview of the constraint violations pattern of the crucial words of sentence (34).

De oude vrouw… 'the old lady'	verzorgde… 'took care of'	hij… 'he'	CASE	SELECTION	PRECEDENCE
☞SI	☞SI	SI	*		
OI	OI	☞OI			*

Table 5: An OT tableau for the incremental interpretation (32)

Let us now compare the pattern of constraint violations found at the nominative case-marked pronoun to the pattern observed at the verb of the sentence starting with the inanimate NP. At the nominative case-marked pronoun the object-initial interpretation overrules the subject-initial interpretation, which was optimal until that point. At that time (i.e., at the second noun phrase), we get a similar 'jump' from one interpretation to the other, as we have seen with respect to the incremental interpretation of sentence (33). Again, the object-initial interpretation violates PRECEDENCE but that is necessary in order to satisfy a stronger constraint, in this case CASE. In other words, the resulting pattern is basically the same pattern as the one created at the verb of the inanimate condition (33). And since Lamers found early and late positivities at the case-marked pronoun in (34) as well, we may conclude that similar ERP effects reported in the on-line studies reflect the similarity of the constraint violation patterns in the incremental optimization of interpretation model.

3. Conflicts in the expression and interpretation of negation

So far we have taken a unidirectional optimization view on interpretation, and we just looked at the constraint ranking of one particular language. However, an important insight from phonological and syntactic studies in OT is that this framework lends itself particularly well to an account of typological variation. This raises the question whether we can benefit from the strength of OT in the study of cross-linguistic semantics. The answer is yes, if interpretation includes taking into account the speaker's use of the available forms in a language as well. In this section we study negation from two perspectives: generation (how does a speaker express a negative meaning?) and interpretation (how does the hearer interpret a sentence with a sequence of negative expressions?). In OT syntax, the input is a meaning, the set of candidates generated by GEN is a set of possible forms, and a ranked set of violable constraints selects the optimal form for the given meaning. In OT semantics, the input is a form (a well-formed sentence), the set of candidates is a set of possible meanings (first-order formulas), and a ranked set of violable constraints selects the optimal interpretation for the given form. The constraints are taken to be universal, but the ranking is language specific. Thus a typology of languages arises in terms of the balance between form and meaning.

3.1 Negation and negative indefinites across languages

Languages generally have ways to express negation, i.e. something that corresponds to the first-order logic connective ¬. In English this would be *not*. Many languages also have nominal expressions negating the existence of individuals having a certain property, i.e. something that corresponds to ¬∃x. In English, this would be *nobody, nothing*. If we assume that knowledge of first-order logic is part of human cognition, we would seem to predict that negation and negative quantifiers behave alike across languages. From empirical research by typologists and theoretical linguists, we know that this is not the case (cf. Jespersen 1917, Horn 1989, Ladusaw 1992, Haspelmath 1997, and many others). We argue that a bidirectional Optimality Theoretic analysis of the marking and interpretation of negation allows us to develop a typology that accounts for variation and

underlying similarities.

On the production side, the question is what form a language uses for an existential meaning within the scope of negation. The relevant possibilities are indefinites, negative polarity items and n-words, as in (35):

(35) a. Ik heb daar toen *niet iets* van durven zeggen. [Dutch]
I have there then *not something* of dare say. [indefinite]

b. I did *not* buy *anything*. [negative polarity item]

c. *No* vino *nadie*. [Spanish]
not came nobody. = Nobody came [n-word]

In Dutch, we can use a plain indefinite for an argument interpreted within the scope of negation. In English and Spanish, this option is blocked, for indefinite pronouns like *someone, something* are positive polarity items that are allergic to negative environments. Instead, English uses a negative polarity item (*anything*), and Spanish a so-called n-word (*nadie*). A negative polarity item like *anything* is dependent on some other expression in the sentence. Sentential negation like in (35b) is a good licensor, but weaker expressions that create a downward entailing context would also do. The distinction between negative polarity items and n-words is subtle, but crucial. According to Ladusaw (1992), n-words are 'self-licensing'. Thus we observe in (36a) that the elliptical answer *nada* means 'nothing', rather than 'something'. In (36b), the n-word *nadie* functions once as a licensor, another time as a licensee:

(36) a. A: Qué viste? B: *Nada* [Spanish]
A: What did you see? B: Nothing

b. *Nadie* maraba a *nadie*
Nobody looked at nobody. = Nobody looked at anybody

The *any*-series in English could not be used in the place of the n-words in (36) without the support of a negative expression (Vallduví 1994, Haspelmath 1997). On the basis of these observations, de Swart and Sag (2002) claim that negative polarity items denote

existential quantifiers (∃x), whereas n-words denote negative existential quantifiers (¬∃x). But then, the paradigm in (37) raises questions for the principle of compositionality of meaning.

(37)	a.	Nobody said nothing.	[English]	[¬∃x¬∃y]
	b.	Niemand zei niets.	[Dutch]	[¬∃x¬∃y]
	c.	Nadie miraba a nadie.	[Spanish]	[¬∃x∃y]
	d.	Nessuno ha parlato con nessuno.	[Italian]	[¬∃x∃y]
	e.	Personne n'a rien dit.	[French]	[¬ or ¬¬]

If both negative quantifiers (English *nothing* in (37a) and Dutch *niets* in (37b)) and n-words (Spanish *nadie* in (37c), Italian *nessuno* in (37d), and French *rien* in (37e)) denote ¬∃x, the question arises why similar forms have the same meaning in isolation, but different meanings in context (single vs. double negation).

De Swart and Sag (2002) solve the compositionality problem in a polyadic quantifier framework, based on Keenan and Westerståhl (1997). All negative quantifiers are collected into an N-store, and are interpreted by means of iteration (double negation) or resumption (negative concord) upon retrieval. This analysis works well for French, in which both interpretations are in principle available (37e). However, many languages are heavily biased towards the use of either iteration ('double negation languages', examples (37a,b)) or resumption ('negative concord languages', examples (37c,d)). The key insight is that languages make use of the same mechanisms, but exploit the relation between form and meaning in different ways. This requires a supplement of the earlier analysis with a bidirectional Optimality Theoretic (OT) component (cf. Blutner 2000, Zeevat 2000).

3.2 A typology of negation within OT

Negative sentences are formally and interpretationally marked with respect to affirmative

sentences. That is, we expect the negative meaning to be reflected in the syntax, and the negative syntax to be reflected in the meaning. The constraint FAITHNEG accounts for this intuition:

(38) FAITHNEG
Reflect the non-affirmative nature of the input in the output.

FAITHNEG is a faithfulness constraint, and aims at a faithful reflection of input features in the output. Since negation is marked in all languages, we take FAITHNEG to be universally ranked at the top. In OT, faithfulness constraints are balanced by markedness constraints, which are output oriented. The markedness constraint that plays a role in negative statements is *NEG:

(39) *NEG
Avoid negation in the output

*NEG is obviously in conflict with FAITHNEG. FAITHNEG and *NEG play a role in OT syntax as well as OT semantics. In addition, we need two maximizing constraints, one aimed at syntax (MAXNEG), and one at semantics (INTNEG):

(40) MAXNEG
Mark 'negative variables' (= arguments in the scope of negation)

(41) INTNEG
Force Iteration (every neg expression in the form contributes a first-order semantic negation in the output)

The functional motivation for the marking of negative variables (Haspelmath 1997, Corblin and Tovena 2003) explains why the use of n-words for arguments that are in the scope of negation is widespread among natural languages. However, n-words are not universal: languages like Dutch, English, Turkish, etc. do not have n-words. This suggests that MAXNEG is not a hard constraint, and its position in the constraint ranking

is not the same for every language. We can account for the difference between languages with and without n-words by varying the position of MAXNEG relative to *NEG. If *NEG is ranked higher than MAXNEG, the optimal way to express the meaning $\neg\exists x_1\exists x_2\ldots\exists x_n$ is by means of indefinite pronouns. If MAXNEG is ranked higher than *NEG, n-words are used to express indefinites under negation. The following OT syntactic tableaux reflect this for the binding of two variables.

$\neg\exists x_1\exists x_2$	FAITHNEG	*NEG	MAXNEG
indef+indef	*		
☞ neg+indef		*	*
neg + neg		**	

Table 6: An OT tableau for the generation of indefinite (for Dutch, Turkish, etc.)

$\neg\exists x_1\exists x_2$	FAITHNEG	MAXNEG	*NEG
indef+indef	*		
neg+indef		*	*
☞ neg + neg			**

Table 7: An OT tableau for the generation of n-word (for Greek, Romance, Slavic, etc.)

The term 'neg' in these tableaux generalizes over both negative quantifiers and n-words, because in isolation, we cannot distinguish them. Because of the top ranking of FAITHNEG, the candidates that we need to compare are those that mark negation somehow in the output. This invariably leads to a violation of *NEG. Two neg expressions are 'worse' than one, so the combination of two neg expressions incurs two violations of *NEG. Tableaux 6 and 7 illustrate that the weight of the constraint, rather than the number of violations is decisive.

As far as generation is concerned, languages that allow indefinites under negation (Dutch, Turkish, etc.), and languages that use n-words (Romance, Slavic, Greek, etc.) differ in their ranking of the two constraints MAXNEG and *NEG. The term 'neg expression' in the tableaux means that we run into the recoverability problem, though. From the candidates

generated, we can derive multiple interpretations, not only the intended one, because in isolation we cannot determine whether a particular neg expression is a negative quantifier or an n-word. Recoverability is assured by the way the generation of negative sentences hangs together with their interpretation. So we need an OT semantic component.

In the interpretive system, FAITHNEG outranks all the other constraints as well. MAXNEG is a purely syntactic constraint that does not play a role in interpretation. So the constraints that need to be ordered are *NEG and INTNEG. If *NEG is ranked higher than INTNEG in the OT semantics, a sequence of multiple neg expressions leads to a single negation meaning by resumption. If INTNEG is ranked higher than *NEG, a series of neg expressions is interpreted as multiple negation by forcing iteration. Tableaux 8 and 9 illustrate this.

neg + neg	FAITHNEG	INTNEG	*NEG
$\exists x_1 \exists x_2$	*	**	
$\neg\exists x_1 \exists x_2$		*	*
☞ $\neg\exists x_1 \neg\exists x_2$			**

Table 8: An OT tableau for double negation (interpretation of Dutch, English, etc.)

neg + neg	FAITHNEG	*NEG	INTNEG
$\exists x_1 \exists x_2$	*		
☞ $\neg\exists x_1 \exists x_2$		*	*
$\neg\exists x_1 \neg\exists x_2$		**	

Table 9: An OT tableau for negative concord (interpretation of Romance, Slavic, Greek, etc.)

The top ranking of FAITHNEG implies that we cannot interpret a statement involving two neg expressions without a reflection of the non-affirmative meaning. So the relevant candidates have at least one negation in the output, and always incur a violation of *NEG. The combination of two neg expressions leads to a double negation reading in languages like Dutch and English, for the constraint INTNEG is ranked higher than *NEG in Tableau

8. Because *NEG outranks INTNEG in Tableau 9, single negation readings win over double negation readings in NC languages such as Spanish, Italian, etc.

3.4 A bidirectional grammar of negation

Collapsing the generation and interpretation perspective, we derive two rankings for negative concord and double negation languages:

(42) Bidirectional grammar
a. Negative concord languages: FAITHNEG >> MAXNEG >> *NEG >> INTNEG
b. Double negation languages: FAITHNEG >> INTNEG >> *NEG >> MAXNEG

Languages strike a balance between the functional desire to express that arguments occur within the scope of negation and the interpretive principle of compositionality of meaning. All languages basically use the same constraints, but they can be ranked in different ways. In sum:

(43) Negative Concord: if you mark 'negative variables' (MAXNEG >> *NEG in syntax), then make sure you do not force Iteration (*NEG >> INTNEG in semantics).
(44) Double Negation: if you force Iteration, (INTNEG >> *NEG in semantics), then make sure you do not mark 'negative variables' (*NEG >> MAXNEG in syntax).

Double negation languages thus sacrifice first-order ('strict') compositionality for functional reasons (cf. Blutner et al. 2003). The typology crucially depends on bidirectionality, for form and meaning go hand in hand in the marking and interpretation of negation.

The two rankings illustrated represent only 2 out of 8 possible rankings of the three relevant constraints. Maximizing both form and meaning (rankings where both MAXNEG and INTNEG are higher than *NEG) or neither (rankings where both MAXNEG and INTNEG are lower than *NEG) are not found in natural language. This finding implies that

languages grammaticalize the basic principle of communication that speakers take the hearer's perspective into account, and hearers take the speaker's perspective into account.

4. Conflicts in children's interpretations

We believe that a bidirectional OT analysis of interpretation is not only the key to cross-linguistic semantics but also clarifies in what sense children's interpretations deviate from the adult interpretations. Before the child will be a competent, adultlike hearer of her language, she must acquire the full process of optimization of interpretation, which crucially involves taking into account the speaker's and the hearer's perspective simultaneously. In this section we present two pieces of evidence for this view, one taken from De Hoop and Krämer (to appear) on children's interpretations of indefinite noun phrases, and one taken from Hendriks and Spenader (2004a,b) on children's interpretations of pronouns and reflexives.

4.1 Acquisition of the interpretation of indefinite subjects and objects

De Hoop and Krämer (to appear) discuss a general, language-independent pattern in child language acquisition in which there is a clear difference between subject and object noun phrases. They explain this pattern within the framework of bidirectional OT.

Consider the Dutch sentences below:

(45) Je mag twee keer een potje omdraaien.
you may two time a pot around-turn
"You may turn a pot around twice."

(46) Je mag een potje twee keer omdraaien.
you may a pot two time around-turn
"You may turn a pot around twice."

In Dutch, the indefinite object noun phrase can either occur to the right of the adverbial phrase *twee keer* 'twice' as in (45), or it can occur to the left of it, as in (46). The left position in (46) is referred to as the scrambled position, the right position in (45) as the unscrambled position. Krämer (2000) tested the interpretation of scrambled and unscrambled indefinite objects in children between 4;0 and 8;0. Children as well as adults get a non-referential (narrow-scope) reading for the unscrambled indefinite. That is, when asked to act out (45) both children and adults turn two pots. For most children below age 7, however, the scrambled indefinites are also interpreted non-referentially, whereas adults always interpret the scrambled indefinites referentially. So, while adults respond to (46) by turning one pot twice, children turn two pots again, just as they did in response to (45).

These Dutch data are in accordance with data from French, English and the Dravidian language Kannada (Boysson-Bardiès and Bacri 1977, Foley et al. 2000, Lidz and Musolino 2002). Cross-linguistically, children between roughly 4 and 6 years old prefer to interpret indefinite object noun phrases non-referentially, even in situations when adults interpret them referentially.

On the other hand, for the interpretation of indefinite *subjects*, the picture is completely different. In a number of experiments, children, just like adults, provided nearly exclusively referential interpretations of indefinite subject noun phrases (Musolino 1998). For Dutch, Bergsma-Klein (1996) found that children correctly assign a referential (wide-scope) reading to indefinite subjects as in (47).[3]

(47) Een meisje gleed twee keer uit.
A girl slipped two time $out_{PARTICLE}$
"A girl slipped twice."

Strikingly, when adults prefer a non-referential interpretation for indefinite subjects, most children only allow the referential interpretation, as shown by by Termeer (2002).
That is, 68% of the children between age 8;7 and 10;4 rejected the adult-like non-

[3] Note that there are exceptions, as in one experiment in Krämer (2000), and some exceptional responses in Bergsma-Klein (1996). The tendency, however, is clear.

referential reading for the embedded indefinite subject in (48).

(48) Er ging twee keer een jongen van de glijbaan af.
There went two time a boy of the slide off
"Twice, there went a boy down the slide."

In conclusion, children are adult-like in their interpretation of referential indefinite subjects and in their interpretation of non-referential indefinite objects. They differ from adults when they have to interpret non-referential indefinite subjects and when they have to interpret referential indefinite objects. How can we explain this pattern?

Note that, cross-linguistically, subjects outrank objects in referentiality. It is a well-known typological generalization, supported by statistical evidence, that subjects tend to be referential, definite, topical, animate, high-prominent in the discourse, among other notions, while objects tend to be non-referential, indefinite, inanimate, low-prominent in the discourse, instead (Aissen, 2003; Comrie, 1989; Lee, 2003). Children seem to behave in accordance with this generalization, that is, they assign a referential interpretation to subjects and a non-referential interpretation to objects. Adults can depart from this pattern when required, but the children's non-adultlike interpretations can be characterized as a failure to depart from the general pattern. Why do children fail in this respect? De Hoop and Krämer (2004) provide a bidirectional Optimality Theoretic account of the adult data, which allows a straightforward explanation of why children deviate from the adult pattern in exactly the way they do.

De Hoop and Krämer use the following constraints in their analysis:

(49) M1: Subjects outrank objects in referentiality, i.e., subjects get a referential interpretation, while objects get a non-referential interpretation.
(50) M2: Indefinite noun phrases get a non-referential interpretation.
(51) F1: Indefinite objects do not scramble.

(52) F2: Subjects are in standard subject position, referred to as [Spec,IP].

These four constraints will give us the unmarked meanings of indefinite subjects and objects as the optimal candidates from an interpretive point of view, and the unmarked forms from an expressive point of view. These constraints, however, cannot account for the marked meanings or the marked forms. How, then, are these obtained?

When the unmarked form is the only form available, as is the case for indefinite objects in a non-scrambling language like English, the marked reading can only be the optimal reading within a certain context. When both a marked and an unmarked form are available, as is the case for indefinite subjects and objects in Dutch and indefinite subjects in English, the marked reading emerges whenever a marked form is used, irrespective of the context. Bidirectional OT (Blutner, 2000) provides us with a straightforward explanation of how these unmarked and marked form-meaning pairs arise. Let us now give a bidirectional OT analysis of the data under discussion in this section.

[f, m] indefinite object f: 1. unscrambled; 2. scrambled m: 1. non-referential (type <*e,t*>); 2. referential (type *e*)	M1	M2	F1
✌ [- scrambling, <e,t>]	✓	✓	✓
[- scrambling, e]	*	*	✓
[+ scrambling, <e,t>]	✓	✓	*
✌ [+ scrambling, e]	*	*	*

Table 10: A bidirectional OT tableau for indefinite objects

In the above tableau we see that the combination of a referential meaning with a scrambled word order for the indefinite object constitutes a super-optimal pair, even though this pair violates all relevant constraints. For the pair [-scrambling, e], a more harmonic meaning is available, while for the pair [+scrambling, <e,t>], a more harmonic form is available. Thus, the bidirectional OT approach straightforwardly accounts for the scrambling phenomenon of indefinite objects in Dutch.

A similar analysis can be provided for the possible forms and meanings of indefinite subjects, as illustrated below.

[f, m] indefinite subject f: 1. [Spec, IP] (standard); 2. [Spec, VP] (embedded) m: 1. referential (type *e*); 2. non-referential (type <*e,t*>)	M1	M2	F1
✌ [[Spec, IP], e]	✓	*	✓
[[Spec, IP], <e,t>]	*	✓	✓
[[Spec, VP], e]	✓	*	*
✌ [[Spec, VP], <e,t>]	*	✓	*

Table 11: A bidirectional OT tableau for indefinite subjects

One super-optimal pair links the unmarked (referential) meaning to the unmarked position (the standard subject position) while the other super-optimal pair links the marked (non-referential) meaning to the marked position (the embedded subject position).

We thus find that a bidirectional OT analysis straightforwardly explains the adult pattern of the interpretation of both indefinite objects and indefinite subjects. Adults are able to evaluate form-meaning pairs. This means that they cannot only find the optimal form for a certain meaning or the optimal meaning for a certain form, they are also capable of determining as a super-optimal pair the combination of a sub-optimal form and a sub-optimal meaning.

We would now like to use the bidirectional OT framework for our explanation of the children's pattern of interpreting indefinite subjects and objects. As soon as children have acquired the relevant constraints and their ranking, they will assign a non-referential reading to indefinite objects and a referential reading to indefinite subjects, independent of the position these noun phrases occupy, whichever the language they may be learning. This is exactly in accordance with what has been attested in the experiments, as discussed above. We have seen above that two factors play a role in making the adult language user depart from the unmarked meanings: contextual demands, and the availability of marked

forms. These factors, then, must have a different effect in children than adults. Indeed, for English-speaking children, it seems that children have a lesser sensitivity to the contextual factors. When these are modulated (Gualmini 2002, Miller and Schmitt 2004), children can obtain the same marked readings as adults. As to the second factor, the availability of marked forms, the Dutch data show that there is also no effect of the marked forms. It seems that children optimize the interpretation of the marked form unidirectionally instead of bidirectionally. Thus, children's optimal interpretation of a marked form will be the same as their optimal interpretation of an unmarked form in the same context.

What is lacking, therefore, is the following 'reasoning' by the child: I can find the optimal interpretation for this form, but I notice that the form is sub-optimal; the speaker would have used the optimal form for the optimal meaning, therefore, the intended meaning for this sub-optimal form must be the sub-optimal meaning.

4.2 Acquisition of the interpretation of pronouns

In the previous section, we presented an account of children's acquisition of the interpretation of indefinite subjects and objects in Dutch. In this section, we argue that a similar developmental process occurs in connection with the acquisition of pronominal interpretations in English.

One of the core phenomena in syntactic theory is binding. In its standard formulation (cf. Chomsky, 1981), Binding Theory consists of three principles, of which we omit the third one because of its irrelevance to the present discussion:

(53) Binding Theory
Principle A: A reflexive must be bound locally.
Principle B: A pronoun must be free locally.

According to the standard formulation of Binding Theory, Principle A and Principle B

entail complementarity between reflexives and pronouns. Although there are a number of contexts where this complementarity breaks down, in this section we will be concerned with the basic pattern only (but see Hendriks and Spenader 2004b for a discussion of exceptions to this pattern).

There appears to be a clear asymmetry in children's pattern of acquisition of the binding principles A and B. Children correctly interpret reflexives from the age of 3;0, assuming coreference between *himself* and *Bert* in (54) in 95% of the time according to some studies. However, they continue to perform poorly on the interpretation of pronouns even up to the age of 6;6 (Grimshaw & Rosen, 1990; Chien & Wexler, 1985). In sentences such as (55), these children misinterpret the pronoun *him* as coreferring with the subject about half the time, which seems to be the result of chance performance.

(54) Bert saw himself.
(55) Bert saw him.

This pattern is commonly referred to as the Pronoun Interpretation Problem, or the Delay of Principle B Effect. For the pattern, a good explanation has yet to be given. Reinhart (1983) and Chien and Wexler (1990) revise Principle B so that (55) is no longer governed by it, making a distinction between syntactic coindexing and pragmatic coreference. As a result, another explanation has to be found for the interpretation of the pronoun in (55). An alternative explanation, put forward by Grimshaw and Rosen (1990), is that (55) is governed by Principle B but that children do not always obey this principle in an experimental setting. However, this fails to explain why children behave so differently with respect to reflexives and pronouns.

Children's production data complicate the picture. Bloom, Barss, Nicol, and Conway (1994) studied the spontaneous production of the English pronoun *me* and the reflexive *myself* in data from the CHILDES database. By age 2;3-3;1, the children that were studied consistently used the pronoun to express a disjoint meaning (99.8% correct), while they used the reflexive to express a coreferential meaning (93.5% correct). In addition, there is also various anecdotal evidence suggesting that children do not have

any problems with the correct production of pronouns (see, e.g., Chien & Wexler 1990: 253, Grimshaw & Rosen 1990: 188-9).

Many studies have ignored these production data because there does not seem to be any obvious way to reconcile these data with the comprehension data. Hendriks and Spenader (2004a,b), however, argue that this pattern can be accounted for within the framework of OT. In particular, they claim that children's lag in pronoun comprehension is due to the late acquisition of the ability to reason bidirectionally. Their analysis is based on the following two constraints:

(56) PRINCIPLE A: A reflexive must be bound locally.
(57) REFERENTIAL ECONOMY: Avoid R-expressions » Avoid pronouns » Avoid reflexives (cf. Burzio, 1998)

The first constraint is the soft-constraint version of the well-known Principle A of Binding Theory. This constraint establishes a relation between a specific form (a reflexive) and a specific interpretation (a coreferential meaning). The second constraint, REFERENTIAL ECONOMY, reflects the view that expressions with less referential content are preferred over expressions with more referential content. In effect, reflexives are preferred to pronouns, and pronouns are preferred to R-expressions. Under the formulation in (57), REFERENTIAL ECONOMY applies to the form of an expression only. For simplicity, we will assume any occurrence of a reflexive to satisfy this constraint, and any occurrence of a pronoun to violate this constraint.

PRINCIPLE A must be stronger than REFERENTIAL ECONOMY because a reflexive is used only if the speaker intends to express a coreferential meaning. In all other cases, a pronoun or R-expression must be used. If PRINCIPLE A were weaker than REFERENTIAL ECONOMY, the only NPs occurring would be reflexives.

The interaction between PRINCIPLE A and REFERENTIAL ECONOMY is able to explain the child language data discussed in the beginning of this section. Tableaux 12 and 13 give the results of production. OT predicts that a reflexive is preferred for expressing a

coreferential meaning in sentences such as (54) and (55) (i.e., in sentences where the anaphoric expression and its antecedent occur within the same local domain), because a reflexive satisfies REFERENTAL ECONOMY, whereas a pronoun does not:

Coreferential meaning	PRINCIPLE A	REFERENTIAL ECONOMY
☞ reflexive form	✓	✓
pronominal form	✓	*

Table 12: An OT tableau for producing a coreferential meaning

For a disjoint meaning, a pronoun is preferred over a reflexive, which violates PRINCIPLE A:

Disjoint meaning	PRINCIPLE A	REFERENTIAL ECONOMY
reflexive form	*	✓
☞ pronominal form	✓	*

Table 13: An OT tableau for producing a disjoint meaning

Tableaux 14 and 15 give the results of interpretation. Because REFERENTIAL ECONOMY is a constraint on forms, it does not have any effect here. Thus based on PRINCIPLE A, it is predicted that the optimal interpretation of a reflexive is a coreferential interpretation.

Reflexive form	PRINCIPLE A	REFERENTIAL ECONOMY
☞ coreferential meaning	✓	
disjoint meaning	*	

Table 14: An OT tableau for interpreting a reflexive form

Because PRINCIPLE A only has an effect when a reflexive is present (i.e., as the input or as a candidate output), it is not relevant when the input form is a pronoun. The result of optimizing over the potential meanings for a pronoun is thus that both meanings are equally preferred. This accounts for the observation that children perform at chance level in comprehension experiments.

Pronominal form	PRINCIPLE A	REFERENTIAL ECONOMY
☞ coreferential meaning	✓	
☞ disjoint meaning	✓	

Table 15: An OT tableau for interpreting a pronominal form

So optimization from meaning to form explains children's production of pronouns and reflexives, and optimization from form to meaning explains their interpretations. Because the optimal forms are the correct adult forms for the given meanings, children are predicted to perform correctly with respect to the production of reflexives (tableau 12) as well as of pronouns (tableau 13). Furthermore, reflexives are predicted to receive a coreferential interpretation (tableau 14). This corresponds to children's as well as adult's interpretations. Pronouns, finally, are predicted to be ambiguous between a coreferential and a disjoint interpretation (tableau 15). This corresponds to children's chance performance on sentences with pronouns.

But if pronouns are ambiguous for children, why isn't this true for adults as well? Hendriks and Spenader (2004a,b) argue that this is because adults optimize bidirectionally (cf. Blutner, 2000). Because the pair [reflexive, coreferential] satisfies both constraints, this pair is superoptimal. The coreferential interpretation will now be blocked for the pronoun because a more harmonic form is available for this meaning, namely a reflexive. As a result, the pair [pronoun, disjoint] will be identified as the second super-optimal pair.

	PRINCIPLE A	REFERENTIAL ECONOMY
✌ [reflexive, coreferential]	✓	✓
[reflexive, disjoint]	*	✓
[pronoun, coreferential]	✓	*
✌ [pronoun, disjoint]	✓	*

Table 16: A bidirectional OT tableau for reflexives and pronouns

Thus bidirectional OT predicts the adult usage of pronouns and reflexives. This suggests

that children begin with unidirectional optimization (from form to meaning, or from meaning to form), and only later acquire the ability to optimize bidirectionally. A child must, when hearing a pronoun, reason about what other non-expressed forms are associated with the potential interpretations of pronouns, realize that a coreferential meaning is better expressed with a reflexive, and then by a process of elimination realize that the pronoun should be interpreted as disjoint.

Hendriks and Spenader's explanation of the Pronoun Interpretation Problem is compatible with ideas in Grodzinsky and Reinhart (1993) and Reinhart (to appear a,b). However, rather than postulating that the blocking effects are the result of the parser preferring the most economical derivation, Hendriks and Spenader derive these Principle B effects from Principle A and the grammatical mechanism of bidirectional optimization. Their analysis thus parallels de Hoop and Krämer's (to appear) analysis of children's acquisition of the interpretation of indefinites in Dutch discussed in the previous section. According to both analyses, children's forms and meanings are the result of unidirectional optimization, whereas adult's combinations of marked forms and marked meanings are the result of bidirectional optimization. In addition, the analysis of pronoun acquisition discussed in this section shows that comprehension lags behind production in those cases where comprehension involves reasoning about alternatives not present in the current situation. It is this bidirectional optimization, and not the grammatical principles themselves, that seems to be acquired late.

5. Conclusion

In our view natural language interpretation can successfully be characterized as an optimization process. Optimality Theory provides us with a cross-modular approach to interpretation which integrates various factors into a set of typically conflicting constraints of varying strengths. Within the domain of lexical semantics, this allows us to account for the influence of context on interpretation. The interpretation of a lexical item within a certain context reflects the process of conflict resolution between a faithfulness constraint that requires the lexical item to get its basic (strongest) meaning, and

contextual constraints that in fact weaken the strongest possible meaning.

The same principles work at sentence level where we also find that various sources of information (world knowledge, syntactic structure, lexical information, etcetera) can be in conflict and each play a part in determining the interpretation of a structure. One of the main advantages of using OT in theories of natural language interpretation is that we can establish a straightforward link between language theory and language processing models. In this article we have shown that both computational as well human sentence processing are adequately analysed in terms of (incremental) conflict resolution.

Natural language interpretation also involves taking into account the alternative forms available for a certain meaning, that is, the speaker's perspective. A bidirectional approach integrates the speaker's and hearer's perspective, and two possible rankings are proposed to explain the characteristics of negative concord versus double negation languages. Alternative rankings of the relevant constraints do not yield existing systems of negation marking, since they are not balanced between the speaker's and hearer's perspective. Thus, languages of the world seem to grammaticalize the basic principle of communication, which require the speaker and the hearer to take into account each other's direction of optimization.

Not only is the bidirectional perspective a useful tool for explaining typological generalizations of language, it can also explain a striking asymmetry in children's comprehension and production of certain form-meaning mappings. We have shown that in these cases it can be argued that in fact the process of bidirectional optimization itself, and not conflict resolution among various constraints, is acquired late.

By applying the idea of (bidirectional) optimization to puzzles of natural language interpretation, we have gained more insight not only in the phenomena themselves, but also in the cognitive foundations of interpretation. OT seems to ground semantic processes firmly in our cognitive system.

References

Aissen, J. 2003. Differential Object Marking: Iconicity vs. Economy. Natural Language and Linguistic Theory 21: 435-483.

Grosz, B., A. Joshi and S. Weinstein 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. Computational Linguistics 2 (21): 203-225.

Beaver, D. 2004. The Optimization of Discourse Anaphora. Linguistics and Philosophy 27 (1): 3-56

Bergsma-Klein, W. 1996. *Specificity in Child Dutch: An Experimental Study*. MA-thesis, Utrecht University.

Bloom, P, A. Barss, J. Nicol and L. Conway 1994. Children's Knowledge of Binding and Coreference: Evidence from Spontaneous Speech. *Language* 70, 53-71.

Blutner, R. 2000 . Some Aspects of Optimality in Natural Language Interpretation. Journal of Semantics 17: 189-216.

Blutner, R., P. Hendriks and H. de Hoop 2003. A New Hypothesis on Compositionality. In *Proceedings of ICCS 2003* Sidney.

Bouma, G. 2003. Doing Dutch Pronouns Automatically in Optimality Theory. In *Proceedings of the EACL workshop on the computational treatment of anaphora*, Budapest.

Bouma, G., G. van Noord and R. Malouf 2001. Alpino: Wide-coverage Computational Analysis of Dutch. In *Proceedings of CLIN 2000.*

Buchwald, A., O. Schwartz, A. Seidl and P. Smolensky 2002. Recoverability Optimality Theory: Discourse Anaphora and Bidirectional Optimization. In Bos, Foster & Matheson (eds.) *Proceedings of EDILOG 2002*: 7-44.

Burzio, L. 1998. Anaphora and Soft Constraints. In P. Barbosa et al. (eds.) *Is the Best Good Enough? Optimality and Competition in Syntax* Cambridge, MA: MIT Press.

Byron, D. and W. Gegg-Harrison 2004. Evaluating Optimality Theory for Pronoun Resolution Algorithm Specification. In *Proceedings of DAARC2004*: 27-32.

Chien, Y., and K. Wexler 1990. Children's Knowledge of Locality Conditions on Binding as Evidence for the Modularity of Syntax and Pragmatics. Language Acquisition 13: 225-295.

Chomsky, N. 1981. *Lectures on Government and Binding* Dordrecht: Foris.

Comrie, B. 1989 . *Language Universals and Linguistic Typology* Chicago: University of Chicago Press.

Corblin, F. and L. Tovena 2003. L'expression de la Négation dans les Languages Romanes. In D. Godard (ed.) *Les Langues Romanes: Problème de la Phrase Simple* Paris: CNRS Editions: 279-342.

de Boysson-Bardiès, B. and N. Bacri 1977. The Interpretation of Negative Sentences. International Journal of Psycholinguistics 4: 73-81.

Foley, C., B. Lust, D. Battin, A. Koehne and K. White. 2000. On the Acquisition of an Indefinite Determiner: Evidence for Unselective Binding. In *Proceedings of the 24th Annual Boston University Conference on Language Development.* S. C. Howell, S. A. Fish and T. Keith-Lucas (eds). Somerville, Massachusetts: Cascadilla Press. 1**:** 286-298.

Grimshaw, J. and S. Thomas Rosen 1990. Knowledge and Obedience: The

Developmental Status of the Binding Theory. Linguistic Inquiry 21: 187-222.
Grodzinsky, Y., and T. Reinhart 1993. The Innateness of Binding and the Development of Coreference. Linguistic Inquiry 24: 69-101.
Gualmini, A. 2002. Children Do not Lack *Some* Knowledge. University of Maryland Working Papers in Linguistics, University of Maryland at College Park 12: 49-74.
Haspelmath, M. 1997. *Indefinite pronouns* Oxford: Clarendon Press.
Hawkins, B. 1984. *The Semantics of English Spatial Prepositions*. Ph.D. thesis. University of California at San Diego.
Hendriks, P. & H. de Hoop 2001. Optimality Theoretic Semantics. Linguistics and Philosophy 24:1-32.
Hendriks, P. and J. Spenader 2004a. A Bidirectional Explanation of the Pronoun Interpretation Problem. In P. Schlenker and E. Keenan (eds.) *Proceedings of the ESSLLI'04 Workshop on Semantic Approaches to Binding Theory*. Nancy, France.
Hendriks, P. and J. Spenader 2004b. When Production Precedes Comprehension: An Optimization Approach to the Acquisition of Pronouns. Ms., University of Groningen.
Hoop, H. de & H. de Swart 2000. Temporal Adjunct Clauses in Optimality Theory. Rivista di Linguistica 12 (1): 107-127.
Hoop, H. de and I. Krämer. To appear. Children's Optimal Interpretations of Indefinite Subjects and Objects. Language Acquisition.
Hoop, H. de and M. Lamers. To appear. Incremental distinguishability of subject and object. *Case, Valency, and Transitivity*. Ed. by L. Kulikov, A. Malchukov and P. de Swart. Amsterdam, John Benjamins.
Horn, L. 1989. *A Natural History of Negation* Chicago: University of Chicago Press.
Jespersen, O. 1917. *Negation in English and Other Languages* Copenhagen: A.F. Høst. Reprinted in *Selected writings of Otto Jespersen.* 1962. London: George Allen and Unwin: 3-151.
Keenan, E. and D. Westerståhl 1997. Generalized Quantifiers in Linguistics and Logic. In J.van Benthem and A. ter Meulen (eds.). *Handbook of Logic & Language* Amsterdam: Elsevier: 837-893.
Krämer, I. M. 2000. *Interpreting Indefinites: An Experimental Study of Children's Language Comprehension*, PhD dissertation Utrecht University, Max Planck Institute series in Psycholinguistics 15.
Krifka, M. 2002. Be Brief and Vague! And How Bidirectional Optimality Theory Allows for Verbosity and Precision. In D. Restle & D. Zaefferer (eds.) *Sounds and Systems: Studies in Structure and Change. A Festschrift for Theo Vennemann* Berlin: Mouton de Gruyter: 439-458.
Ladusaw, W. 1992. Expressing Negation. In *Proceedings of SALT 2*. Ohio State University: 237-259.
Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.
Lamers, M.J.A. 2001. *Sentence processing: using syntactic, semantic, and thematic information*. PhD dissertation University of Groningen
Lamers, M. and H. de Hoop 2005. Animacy Information in Human Sentence Processing: An Incremental Optimization of Interpretation Approach. In H. Christiansen *et al.* (eds.), *CSLP 2004*. Berlin/Heidelberg: Springer-Verlag: 158-171.
Lee, H. 2003. Parallel Optimization in Case Systems. In M. Butt and T. King (eds.) *Nominals: Inside and Out* Stanford: CSLI.

Lidz, J. and J. Musolino 2002. Children's Command of Quantification. Cognition 84: 113-154.
Lindstromberg, S. 1998. *English Prepositions Explained*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Miller, K. and C. Schmitt 2003. Wide-scope Indefinites in English Child Language. In *Proceedings of GALA 2003* Utrecht: Uil-OTS working papers.
Mitkov, R. 2002. *Anaphora Resolution*. London: Longman.
Prince, A. and P. Smolensky 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. London: Blackwell
Reinhart, T. 1983. Coreference and Bound Anaphora: A Restatement of the Anaphora Questions. Linguistics and Philosophy 6: 47-88.
Reinhart, T. To appear a. The Processing Cost of Reference-Set Computation: Acquisition of Stress Shift and Focus. Language Acquisition.
Reinhart, T. To appear b. Processing or Pragmatics? Explaining the Coreference Delay. In T. Gibson and N. Pearlmutter (eds.) *The Processing and Acquisition of Reference* Cambridge, MA: MIT Press.
Schulze, R. 1991. Getting Round to *(A)round*: Towards a Description and Analysis of a Spatial Predicate. In G. Rauh (ed.) *Approaches to Prepositions* Tübingen: Gunter Narr Verlag: 251-274.
Schulze, R. 1993. The Meaning of *(A)round*: A study of an English preposition. In R.A. Geiger & B. Rudzka-Ostyn (eds.) *Conceptualizations and Mental Processing in Language* Berlin/New York: Mouton de Gruyter: 399-431.
Swart, H. de 2005. Marking and Interpretation of Negation: a Bi-directional OT approach. In H. Campós et al. (eds.) *Negation, Tense and Clausal Architecture: Cross-linguistic investigations* Georgetown University Press.
Swart, H. de and I. Sag 2002 . Negation and Negative Concord in Romance. Linguistics and Philosophy 25: 373-417.
Taylor, J.R. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press. (Second edition)
Termeer, M. 2002. *Een Meisje Ging Twee Keer van de Glijbaan. A Study of Indefinite Subject NPs in Child Language.* MA-thesis, Utrecht University.
Vallduví, E. 1994. Polarity Items, N-words and Minimizers in Catalan and Spanish, Probus 6: 263-294.
Zanuttini, R. 1991. *Syntactic Properties of Sentential Negation*, PhD. Dissertation, University of Pennsylvania.
Zeevat, H. 2000. The Asymmetry of Optimality Theoretic Syntax and Semantics. Journal of Semantics 17: 243-262.
Zwarts, J. 2004. Competition Between Word Meanings: The Polysemy of *(A)round*. In C. Meier en M. Weisgerber (eds.) *Proceedings of the Conference sub8-Sinn und Bedeutung*. Konstanz: University of Konstanz Linguistics Working Papers: 349-360.