

Monotonicity and Processing Load

BART GEURTS AND FRANS VAN DER SLIK

University of Nijmegen

Abstract

Starting out from the assumption that monotonicity plays a central role in interpretation and inference, we derive a number of predictions about the complexity of processing quantified sentences. A quantifier may be upward entailing (i.e. license inferences from subsets to supersets) or downward entailing (i.e. license inferences from supersets to subsets). Our main predictions are the following:

- If the monotonicity profiles of two quantifying expressions are the same, they should be equally easy or hard to process, *ceteris paribus*.
- Sentences containing both upward and downward entailing quantifiers are more difficult than sentences with upward entailing quantifiers only.
- Downward-entailing quantifiers built from cardinals, like ‘at most three’, are more difficult than others.
- Inferences from subsets to supersets are easier than inferences in the opposite direction.

We present experimental evidence confirming these predictions.

1 INTRODUCTION

This paper is about quantification, a topic that has been central to natural language semantics since the very inception of the field. Many semantics textbooks (and all the good ones) will have at least one chapter on quantification, and will recount in more or less detail that there is a rich and widely accepted framework for treating quantified expressions, i.e. the theory of generalized quantifiers. Somewhat surprisingly, this framework has almost completely failed to affect the branches of psychology that might have benefited from it, such as the psychologies of language, acquisition, and reasoning. Our aim is to demonstrate that this neglect is not justified. More concretely, it will be shown how the theory of generalized quantifiers can contribute in a novel and non-trivial way to our understanding of cognitive complexity. Some expressions are more difficult to process than others, and in certain cases the difference is due to semantic rather than structural factors. We are concerned with such semantic sources of complexity.

The paper is organized as follows. We start with a quick overview of generalized-quantifier theory, with emphasis on the concept of monotonicity (section 2), and discuss in some detail the linguistic and psychological evidence for the importance of this concept (section 3). After a brief excursus on the interpretation of number words (section 4), we present our predictions (section 5), and report on an experimental study we carried out to test them (section 6).

2 QUANTIFICATION AND MONOTONICITY

In accordance with standard practice in semantics we assume that determiners like ‘all’, ‘some’, ‘most’, and so on, combine with set-denoting expressions to form compound expressions that denote properties of sets. The denotation of ‘all A’, for instance, is the family of sets of which A is a subset, and consequently ‘All A B’ is true iff B is a member of the denotation of ‘all A’; i.e. ‘All A B’ is true iff $A \subseteq B$. Similarly:

- (1) a. ‘No A B’ is true iff $A \cap B = \emptyset$
- b. ‘Some A B’ is true iff $A \cap B \neq \emptyset$
- c. ‘At least five A B’ is true iff $|A \cap B| \geq 5$
- d. ‘Most A B’ is true iff $|A \cap B| > |A - B|$

A quantifier is upward entailing (or monotone increasing) if it licenses inferences from subsets to supersets, and it is downward entailing (or monotone decreasing) if it licenses inferences in the opposite direction. Less informally: ‘Det A’ is upward entailing iff, for any B and C such that $B \subseteq C$, ‘Det A B’ entails ‘Det A C’; and ‘Det A’ is downward entailing iff, for any B and C such that $C \subseteq B$, ‘Det A B’ entails ‘Det A C’. A quantifier is monotonic iff it is either upward or downward entailing; otherwise it is non-monotonic. We will always confine our attention to the second argument of a determiner, and ignore the first one, so no harm will be done if we say that a determiner Det is downward (upward) entailing iff ‘Det A’ is downward (upward) entailing, for any A; similarly for monotonicity and non-monotonicity. Table 1 illustrates the monotonicity properties of some determiners.

The concept of monotonicity is generally applicable to any expression whose denotation is partially ordered. For example, the prepositional object in (2a) is upward entailing, as witness the fact that (2a) entails (2b):

- (2) a. Betty lives in Berlin.
- b. Betty lives in Germany.

Table 1 Monotonicity properties of some determiners

<i>Upward entailing</i>	
all	All dots are scarlet \Rightarrow All dots are red
most	Most dots are scarlet \Rightarrow Most dots are red
some	Some dots are scarlet \Rightarrow Some dots are red
at least n	At least five dots are scarlet \Rightarrow At least five dots are red
<i>Downward entailing</i>	
no	No dots are red \Rightarrow No dots are scarlet
fewer than n	Fewer than five dots are red \Rightarrow Fewer than five dots are scarlet
<i>Non-monotonic</i>	
exactly n	Exactly five dots are scarlet \nRightarrow Exactly five dots are red Exactly five dots are red \nRightarrow Exactly five dots are scarlet

If we negate (2a), the direction of monotonicity is reversed: (3a) entails (3b), rather than the other way round:

- (3) a. Betty doesn't live in Germany.
b. Betty doesn't live in Berlin.

The direction of monotonicity can be reversed again by embedding in yet another negative context:

- (4) a. It's not true that Betty doesn't live in Berlin.
b. It's not true that Betty doesn't live in Germany.

These examples illustrate a point that is of some importance for the objective of this paper, namely, that the monotonicity properties of a position in a sentence may be affected by the interplay between several expressions. In (3) and (4) this was shown by embedding a sentence under negation. In the following examples, different combinations of quantifiers are used with similar effects (we use up and down arrows to indicate upward and downward entailment, respectively):

- (5) a. At least three hunters shot \uparrow [more than five \uparrow [colleagues]].
b. At least three hunters shot \uparrow [fewer than five \uparrow [colleagues]].
c. At most three hunters shot \downarrow [more than five \uparrow [colleagues]].
d. At most three hunters shot \downarrow [fewer than five \uparrow [colleagues]].

The syntactic structure of each of these sentences is same: 'Det₁ hunters shot Det₂ colleagues'. Semantically, however, they are rather different. In the first two sentences the object position is upward entailing, because the subject position is occupied by an upward entailing quantifier. In (5a) the

object NP is upward entailing, as well, and therefore the position occupied by the noun ‘colleagues’ is upward entailing, too. In (5b), by contrast, the object NP is downward entailing, and the noun position follows suit. In (5c,d) the subject quantifier makes the object position downward entailing. In combination with an upward entailing object, this makes the position of the noun downward entailing, (5c); when combined with a downward entailing quantifier, the same position is upward entailing, (5d).

Intuitively speaking, the monotonicity profile of (5a) is more harmonic than that of the other sentences in (5), in the sense that (5b–d) contain conflicting clues as to what monotonicity property of the last noun position is. Our main empirical tenet is that harmony, in this sense of the word, is among the factors that determine semantic complexity. A sentence whose parts are upward entailing throughout, for example, is less complex than a sentence whose monotonicity profile is mixed. However, before we can work out this idea in more detail, there are some preliminary issues that need to be addressed.

3 MONOTONICITY MATTERS

It is one thing to show that linguistic expressions have monotonicity properties. It is quite another thing to show that such properties matter to languages and their speakers. (After all, every linguistic expression, and indeed every thing, has an indefinite number of properties that aren’t of interest to anyone.) So let us consider some of the linguistic and psychological evidence for the relevance of monotonicity.

To begin with, monotonicity is crucially involved in the following general constraint on lexicalization:

- (6) Lexical NPs and NPs built from simple determiners are always monotonic, and usually upward entailing (Keenan and Westerståhl 1997).

Lexical NPs are pronouns, demonstratives, proper names, and lexical quantifiers such as ‘everybody’, ‘someone’, and ‘nobody’. NPs built from simple determiners are ‘every aardvark’, ‘no bonsai’, ‘most cassowaries’, and so on. That all these expressions should be monotonic is quite remarkable, for it means that the overwhelming majority of quantifiers expressible in English cannot be expressed either lexically or by NPs built from simple determiners. The constraint entails, for example, that there couldn’t be a lexical NP meaning ‘everyone but myself’ or ‘about five dolphins’.

A possible exception to the constraint in (6) are NPs with bare cardinals, like ‘three elephants’. In many cases, a quantifier of the form

' n A' will be construed as 'no more and no fewer than n A'. For example, the following sentence would normally be taken to mean that the exact number of elephants photographed by Fred was three:

(7) Fred photographed three elephants.

Taken at face value, this observation implies that 'three elephants' is non-monotonic, and therefore an exception to the constraint in (6). However, there are rather good reasons for assuming that the lexical meaning of 'three' is 'at least three', and that the 'exactly' reading of (7) is due to a conversational implicature (Horn 1989). If we accept the standard arguments, quantifiers of the form ' n A' are upward entailing, and conform to the constraint in (6). We will return to the semantics and pragmatics of cardinals in the next section.

All languages have negative polarity items, which are so-called because they can occur within the scope of a negative expression, and are banned from positive environments. English 'any' is a case in point:

- (8) a. Wilma {*has/doesn't have} any luck.
 b. {*Someone/No one} has any luck.

On closer inspection, it turns out that negative polarity items do not necessarily require a negative environment, though there certainly are constraints on where they may occur, as witness:

- (9) a. If Wilma has any luck, she will pass the exam.
 b. *If Wilma passes the exam, she must have any luck.
 (10) a. Everyone who has any luck will pass the exam.
 b. *Everyone who passes the exam must have any luck.

The generalization is that negative polarity items may only occur in downward entailing positions. In effect, a negative polarity item serves to signal that the environment in which it occurs is downward entailing.¹ This is a linguistic fact, but it is a fact about language users, too, for it can hardly be an accident that speakers will produce sentences like (9a) and (10a), and refrain from producing sentences like (9b) or (10b). The most plausible explanation for this behaviour is that speakers know that 'any' may only occur in downward entailing environments, and that they routinely compute the monotonicity

¹ Actually the facts about polarity are rather more complex than this brief discussion might suggest. See (van der Wouden 1997) for an overview of the main issues. One particular problem we finish here is that in fact conditional antecedents may not be downward entailing in a straightforward semantic sense (see Heim 1984).

properties of incoming and outgoing utterances. In short, polarity items provide evidence that monotonicity is relevant to the psychology of production and comprehension.

There are several studies on language acquisition which suggest that already at a very tender age, speakers are attuned to monotonicity properties. For example, in an elicited production study, O'Leary and Crain (1994) showed that even before they reach the age of four children consistently avoided using 'any' in environments like (10b). In this study children were presented with incorrect descriptions of stories previously acted out by the experimenters. Target sentences, which were produced by a puppet, contained the negative polarity item 'anything' and in their corrections children were expected to avoid using this expression:

- (11) a. Puppet: None of the Ninja Turtles got anything from Santa.
Child: No, this one got something from Santa.
b. Puppet: Only one of the reindeer found anything to eat.
Child: No, every reindeer found something to eat.

O'Leary and Crain found that negative polarity items were practically never used in corrections, which indicates rather strongly that young children are aware of the distributional constraints on 'anything'. See Musolino et al. (2000) for further discussion of monotonicity in language acquisition.

Our last illustration of the importance of monotonicity relates to the interpretation of, and inferences licensed by, so-called 'donkey sentences' (Kanazawa 1994, Geurts 2002):

- (12) Every farmer who owns a donkey beats it.

There is a much-discussed problem with this kind of sentences. It is that native speakers' intuitions about their truth conditions are insecure in certain respects. To explain how, consider the following situation. We have three farmers each of whom owns a single donkey and beats it, and we have a fourth farmer who owns two donkeys, and beats only one of them. Question: Is (12) true in this situation? It turns out that speakers' intuitions diverge: according to some the sentence is false in the given situation, while others judge the sentence to be true or cannot make up their minds. In brief: speakers disagree about the exact truth conditions of sentences like (12). This being so it is somewhat paradoxical that speakers appear to *agree* about certain inferences involving (12). For example, there is a consensus that (12) entails (13a) and is entailed by (13b).

- (13) a. Every farmer who owns a male donkey beats it.
b. Every farmer who owns a donkey beats it with a stick.

Note that both of these inferences are based on the monotonicity properties of ‘every’: given a proposition of the form ‘Every A B’, inferences to subsets are permitted in the A-part, while inferences to supersets are permitted in the B-part. Apparently, monotonicity inferences may go through even if the interpretation of a sentence remains underdetermined or unclear in certain respects.

Not only do these observations confirm the importance of monotonicity, they also suggest a reason why this should be the case. Monotonicity inferences as such are very simple; they are just a matter of replacing one term by another that is either less or more specific. Furthermore, these inferences are shallow in the sense that they can be ‘read off’ a sentence simply by checking for the occurrence of logical words such as quantifiers, negative particles, and so on; they do not require a deep understanding of what a sentence means. This explains why speakers boldly draw monotonicity inferences from sentences whose truth conditions they are uncertain about.

4 CARDINALS

Some quantifiers, we have seen, are upward entailing, while others are downward entailing; ‘at least n A’ is an example of the former, ‘at most n A’ of the latter. Some quantifiers are non-monotonic, i.e. neither upward nor downward entailing; a case in point is ‘exactly n ’ (see Table 1 for examples). What is the monotonicity profile of bare ‘ n ’? The answer that comes to mind first, perhaps, is that ‘ n ’ is synonymous with ‘exactly n ’, and therefore non-monotonic. But things are not as simple as that:

- (14) Fred had (**★**exactly) three oranges and possibly four.
 if not four.
 and for all I know four.

If ‘three’ was synonymous with ‘exactly three’, the two should always be interchangeable, but it is plain that in (14) they are not. It would seem that bare ‘ n ’ is ambiguous between ‘at least n ’ as well as ‘exactly n ’. If so, the ambiguity becomes manifest in cases like the following:

- (15) If you score five points, you get a teddy bear.

The teddy bear may be yours if you score five points or more, or perhaps you’ll win a Barbie doll (instead of a teddy bear) if you score more than five points.

However, it would be unsatisfactory simply to assume that cardinals are ambiguous between an ‘at least’ and an ‘exactly’ reading, and leave

the matter at that. This worry has led neo-Griceans like Horn (1972, 1989) to argue that the basic meaning of ‘*n*’ is ‘at least *n*’, and that, depending on the context, pragmatic reasoning may strengthen this to an ‘exactly’ interpretation. On the neo-Gricean view, the ‘exactly’ interpretation of the number word ‘seventeen’, say, actually consists of two components: the literal meaning of the word (‘at least seventeen’) and a scalar implicature (‘not more than seventeen’).

There are problems with this treatment of cardinals, which, incidentally, led Horn (1992) to throw in the towel after 20 years, leaving Levinson (2000) as the sole defender of what seems to be a lost cause. If it were true that cardinals are scalar expressions, the ‘at least’ interpretation of a cardinal should be more basic than the ‘exactly’ interpretation, and there are various kinds of evidence which suggest that this expectation is not borne out by the facts (Horn 1992; Carston 1998; Geurts 1998). Most importantly, for our present purposes, there are experimental data indicating that young children, who in general seem to find it easier to obtain ‘at least’ interpretations than adults, prefer ‘exactly’ readings for cardinals (Noveck 2001; Papafragou and Musolino 2003; Musolino 2004).

In brief, the semantics (and pragmatics) of number words, simple though they may seem at first, turns out to be remarkably thorny, but in this paper we will not attempt an analysis of our own. We will assume that both the ‘exactly’ and the ‘at least’ readings of cardinals are easily obtained—and this much is uncontroversial. We will not make any assumptions about how the two readings are related. In particular, we will not assume that one is derived from the other. The only thing that matters to our purposes is that both are readily available.

In saying that the ‘at least’ and ‘exactly’ readings of bare ‘*n*’ are readily available we intend to implicate that they are the *only* readings that are readily available—and this *is* a controversial point. It has been argued by Carston (1998), among others, that in addition to the readings discussed so far there is an ‘at most’ reading, which is on a par with the others. Some of Carston’s examples are:

- (16) a. She can have 2000 calories without putting on weight.
 b. The council houses are big enough for families with three kids.
 c. You may attend six courses.

While it is true that the cardinal expressions in these examples can be paraphrased with ‘at most’, this does not show that there is an ‘at most’ reading just as there are ‘at least’ and ‘exactly’ readings. First, once it is

granted that cardinals may have ‘exactly’ readings, it can be argued that the intended readings of (16a–c) result from pragmatic inferences. For example, if (16c) is read as, ‘There are deontically possible worlds in which the addressee attends exactly six courses’, it is natural to infer that there are deontically possible worlds in which the addressee attends fewer than six courses, and no such worlds in which the addressee attends more than six courses. Secondly, even if ‘at most’ readings are not derived in this way, they only come to the fore in very special circumstances. It would be most unusual for a simple sentence like (17), for instance, to give rise to an ‘at most’ construal of ‘fifty’:

(17) Betty has fifty guilders.

On the strength of these observations, we will assume that the ‘at most’ reading of a cardinal, if available at all, is exceptional.

5 MONOTONICITY AND INFERENCE

The notion that monotonicity might play an important role in reasoning (especially reasoning with quantifiers) is not new; it can be traced back in part to medieval times and in part even to Aristotle. More recent developments in semantic theory have helped to further develop the idea and incorporate it in the framework of generalized-quantifier theory (Sánchez Valencia 1991). The first sustained attempt at showing that monotonicity is relevant to the *psychology* of reasoning, as well, is due to Geurts (2003a), who proposes a theory of syllogistic reasoning that hinges on the concept of monotonicity.² Geurts presents a processing model in the tradition of ‘mental logic’; that is to say, the core of his system is a collection of weighted inference rules that assigns a complexity index to every argument form within its scope. Methodologically, the strategy pursued in this paper is different: we will for the most part ignore the question how people process inferences, and argue simply that some arguments are bound to be more difficult than others because they involve expressions that, for semantic reasons, are more complex. Hence, the level at which we derive our predictions about human reasoning is different from that of Geurts (2003a). But the key idea is the same; it is that certain patterns of reasoning crucially involve monotonicity inferences. We will now discuss a number of predictions that are derivable from this premiss.

² See Newstead (2003) for critical discussion of Geurts’s theory, and Geurts (2003b) for a reply.

5.1 *Feasible inferences*

Our first prediction is seemingly trivial: it is that certain reasoning tasks should be feasible.

- (18) More than half of the tenors have the flu.
More than half of the tenors are sick.

Intuitively speaking, this inference is valid and obviously so. From a strictly logical point of view this is quite remarkable. On the one hand, it is a familiar fact that first-order logic cannot capture the meaning of the determiner ‘more than half’ (Barwise and Cooper 1981); a more powerful logic is needed for that. On the other hand, to the extent that it is decidable at all, first-order logic is known to be computationally intractable, and therefore a more powerful logic can only be worse. So how do we humans manage to see that arguments like (18) are valid? The answer is fairly obvious: by trading completeness for efficiency. The inference strategies we employ cannot fail to be radically incomplete, but at least they are reasonably efficient.

One of the reasons why monotonicity inferences are important to us is that they are highly efficient. As we noted at the end of section 3, a system for producing monotonicity inferences can be very simple, because it requires only a shallow understanding of the representations it operates on. In the case of (18), for example, such a system would merely need to know that having the flu entails being sick and that ‘more than half’ is upward entailing; the exact meaning of ‘more than half’ is immaterial. That is why (18) is easy.³

Monotonicity inferences are simple because they don’t require a full-blown interpretation; a superficial understanding usually will do. By the same token, we predict that, in so far as human reasoning is based on monotonicity inferences, it will be sensitive only to the logical bare bones of an argument. To illustrate this point, compare (18) with the following argument:

- (19) All the tenors have the flu.
All the tenors are sick.

Logically speaking, the difference between (18) and (19) is profound. For, whereas (18) can be expressed in first-order predicate logic, (19) requires more powerful means (as noted above), and in this sense is more complex. However, ‘more than half’ and ‘all’ are both upward

³ As Larry Horn pointed out to us, the inference (18) may become problematic if it is given that all tenors are sick. More generally, people will tend to avoid endorsing conclusions that are pragmatically infelicitous. See Geurts (2003b) for further discussion of this point.

entailing, so with respect to monotonicity inferences (18) and (19) should be equally easy. As we will see, this prediction turns out to be correct.

5.2 *Harmony*

While the concept of monotonicity may simplify our models of reasoning in certain ways, as we have just seen, it also helps to reveal sources of complexity that would remain hidden without it. To explain how, compare the following sentences:

- (20) a. Some of the sopranos sang with more than three of the tenors.
 b. Some of the sopranos sang with fewer than three of the tenors.

While the surface forms of these sentences are nearly identical, their monotonicity profiles are different, and we believe that the difference matters. Whereas the two quantifiers in (20a) are both upward entailing, their counterparts in (20b) are upward and downward entailing, respectively. Let us say that the monotonicity profile of the former sentence is more harmonic than that of the latter. We predict that harmony, in this sense of the word, is a measure of semantic complexity: all other things being equal, (20a) will be easier to process than (20b), because its monotonicity profile is more harmonic. The rationale behind this prediction is as follows. Both (20a) and (20b) are of the form ‘Some of the sopranos sang with . . .’, where the position marked by the dots is upward entailing. In (20a) this position is occupied by a quantifier that is itself upward entailing, but in (20b) the second quantifier is downward entailing. As a consequence, the latter sentence gives conflicting clues as to what its monotonicity profile is—which should make it harder to process than (20a).

(20b) is less harmonic than (20a), we have said, because its monotonicity profile is mixed: while the two quantifiers in (20a) are both upward entailing, the quantifiers in (20b) point in opposite directions, so to speak. But what if the quantifiers are both downward entailing, as in the following sentence?

- (21) None of the sopranos sang with fewer than three of the tenors.

Even though both quantifiers occurring in (21) are downward entailing, we should expect this sentence to be less harmonic than (20a), because it involves a monotonicity reversal: the position occupied by the noun in ‘fewer than three of the tenors’ is downward entailing, but as the quantifier fills an argument slot that is itself

downward entailing, the ‘tenors’ position becomes upward entailing, as the validity of following argument confirms:

- (22) None of the sopranos sang with fewer than three of the tenors.
None of the sopranos sang with fewer than three of the male singers.

Hence, (22) should be more difficult than (23), which is upward entailing throughout:

- (23) Some of the sopranos sang with more than three of the tenors.
Some of the sopranos sang with more than three of the male singers.

This prediction is intuitively correct (i.e. (23) does seem to be easier than (22)), and is corroborated by experimental data, as we will see, but it is possible that the complexity of (22) is mitigated by an additional factor.

It is a well-established fact that in processing negation, many people replace double negatives with the appropriate positive expression.

- (24) a. The first number on the list is not odd.
b. The first number on the list is even.

As shown by Clark (1974), if someone has to evaluate (24a), he will often decide to mentally rephrase the sentence as (24b), and evaluate that instead.⁴ This ‘conversion method’, as Clark calls it, applies not only to double negatives, but also to sentences like (21), in which one downward entailing quantifier sits in the scope of another. Now if, in the process of assessing (22), the conversion method is applied, the task is made more manageable, obviously. This is not to say that (22) will be as easy as (23) is; the procedure for deciding (22) will be more complex than for (23), whether or not the conversion method is used. But it is to say that the conversion method will make (22) more manageable than it would be otherwise. Consequently, it may well turn out that (22) is easier than analogous arguments with a mixed monotonicity profile. We are not in a position to predict that it will be easier (or harder, for that matter), because that would require an estimate of the *absolute* complexity index of the procedures and representations involved in solving problems like (22) and (23). But at least we are able to say that certain patterns in the data can be made sense of within the general approach adopted here.

To sum up the foregoing discussion: given a corpus of sentences of the form ‘Det_A A ... Det_B B’, where the latter quantifier is in the scope of the former, and both may be either upward (↑) or downward

⁴ Occasionally, this procedure is overextended and applied in cases in which it is not valid; see Geurts (2002) for an example. The psychological literature on negation is discussed at length and in depth by Horn (1989).

(\downarrow) entailing, we expect to find one of the following patterns (where ' $X < Y$ ' means that X is easier than Y):

$$\begin{array}{c} \uparrow \text{Det}_A \downarrow \text{Det}_B \\ \uparrow \text{Det}_A \uparrow \text{Det}_B < \downarrow \text{Det}_A \uparrow \text{Det}_B \\ \downarrow \text{Det}_A \downarrow \text{Det}_B \\ \\ \uparrow \text{Det}_A \uparrow \text{Det}_B < \downarrow \text{Det}_A \downarrow \text{Det}_B < \begin{array}{l} \uparrow \text{Det}_A \downarrow \text{Det}_B \\ \downarrow \text{Det}_A \uparrow \text{Det}_B \end{array} \end{array}$$

That is, the $\uparrow \text{Det}_A \uparrow \text{Det}_B$ sentences should be easier than all others; the mixed cases should be equally hard; and the $\downarrow \text{Det}_A \downarrow \text{Det}_B$ sentences may be of intermediate complexity.

We argued in section 4 that bare number words are either upward entailing or non-monotonic. If this is so, considerations of harmony yield an interesting prediction: expressions like 'at most five' or 'fewer than five' should be harder to process than others. Hence, (25c) should be more difficult than the other phrases in (25):

- (25) a. some kangaroos
 b. at least five kangaroos
 c. at most five kangaroos
 d. no kangaroos

Semantically, (25a) and (25b) are on a par in that they are upward entailing, while (25c) and (25d) are both downward entailing. But in another respect (25c) is the odd one out: it is a downward entailing phrase that contains an expression that is not downward entailing. None of the other phrases in (25) have this property. In particular, (25b) is no less harmonic than (25a) or (25d), assuming as we do that one of the readily available construals of 'five' is upward entailing. We predict, therefore, that (25c) will be harder to process than the other phrases in (25).⁵

5.3 *Up is easier than down*

To the extent that it relies on monotonicity, reasoning is a matter of interchanging set-denoting representations, going from a set to one of its subsets, or *vice versa*.

⁵ Geurts (2003a) derived a similar prediction by assuming that 'at least' is inherently negative. There are several problems with this; most importantly, it is *ad hoc* and fails to predict a difference between 'at most' and 'no'.

- (26) a. $\frac{\text{Some A are B}}{\frac{\text{All B are C}}{\text{Some A are C}}}$ b. $\frac{\text{Some A are B}}{\frac{\text{All C are B}}{\text{All C are B}}}$

In both of these arguments, the conclusion is obtained by replacing the B-term in the major premiss with the C-term. But (26a) moves upward in the sense that $B \subseteq C$, whereas in (26b) the inference goes in the opposite direction, because $C \subseteq B$. (Note that we are using the up/down terminology in two related but different senses. If we say that expression X is upward or downward entailing, we are describing the inferences *licensed* by X. If, on the other hand, we say that an inference moves upward or downward, we are describing an *actual* inference, be it valid or not.)

There are reasons to expect that downward inferences may be more demanding than upward ones. It is well known that pairs of concepts like big/small, long/short, up/down, above/below, and so forth, which would appear to be symmetric, in fact show a fundamental cognitive asymmetry: by several criteria, the first member of each pair is privileged over the second one.⁶ To begin with, the *words* ‘big’, ‘long’, etc. are unmarked in the linguistic sense of the word. We normally ask how long, not how short, something is, and the name for the measure is ‘length’ rather than ‘shortness’. In many cases markedness is reflected morphologically, as in the pair ‘happy/unhappy’, unmarked words are more frequent than their marked counterparts, and tend to be learned earlier. Furthermore, these regularities hold across languages; so if any language has words that translate into English as ‘big’ and ‘small’, the first one will be the unmarked form—which already indicates that what linguists call markedness does not depend on linguistic convention alone.

Even more compelling evidence for the claim that the asymmetries in question are language-independent is provided by McGonigle and Chalmers’s (1986, 1996) discovery that they are present in non-human primates, too. In one experiment, McGonigle and Chalmers presented squirrel monkeys with series of objects of varying sizes. If all the objects were black, subjects had to select the largest item, if all were white, the smallest. Once their performance was perfect, the monkeys’ decision times were recorded, and it was found that they were consistently faster when they had to find the largest item in a set. This and other results confirm that the asymmetry between large and small is not just a linguistic matter.

⁶ The *locus classicus* is Clark (1973). See Horn (1989) for a more recent overview.

In many though not all cases, the dimension to which an opposition pair applies has a natural end point, and if it has the direction away from it is the favoured one. Take size, for example. The extent of an object must be greater than zero, or it wouldn't exist, and accordingly the natural direction is from smaller to greater size. And sure enough, as we noted already, 'big' is the unmarked form *vis à vis* 'small', and in other respects too the concepts related to larger size are privileged compared with concepts related to smallness. Another example is provided by the opposition between up and down. In this case the relevant dimension has no intrinsic end point: there is no contradiction in the notion of something going up or down forever. But, as observed by Clark (1973), for humans and other agents there is a canonical end point, which is the ground on which they stand. If this is correct, it should be the case that the favoured direction is upwards, and this prediction, too, has been confirmed (Clark 1974).

Thus far we have concentrated our attention on pairs of spatial concepts, but what we have said carries over to the temporal domain, for example, and to more abstract opposites like good/bad, clever/dumb, strong/weak, and so on. This being so, we conjecture that there is an asymmetry between upward and downward monotonicity inferences, and if such an asymmetry exists, upward inferences should be easiest, because no set is smaller than the empty set.

6 EXPERIMENTAL EVIDENCE

In the foregoing we have discussed several predictions that follow naturally from the idea that monotonicity inferences play an important part in human reasoning:

- Even if the quantifiers involved are non-trivial, logically speaking, monotonicity inferences may be very easy.
- If the monotonicity profiles of two quantifiers are the same or similar, we expect this to be reflected in experimental data. For example, since they are both upward entailing, monotonicity inferences involving 'all' will be about as difficult or easy as the corresponding inferences with 'most'.
- Complex expressions all of whose parts are upward entailing are easier to process than compounds whose parts have mixed monotonicity profiles. In particular, downward entailing quantifiers built from cardinals ('at most n ', 'fewer than') are more complex than others.
- Upward inferences are easier than downward inferences.

In order to test these predictions, we conducted the following experiment.

6.1 *Materials and procedure*

Forty-five first-year students at the University of Nijmegen received a 24-page booklet with instructions printed on the cover. Each page presented an argument of the following form:

MAJOR PREMISS: Det_A A played against Det_B B.

MINOR PREMISS: EITHER: All B were C

OR: All C were B.

CONCLUSION: Det_A A played against Det_B C.

In each argument Det_A was instantiated with one of the following: ‘every’, ‘most’, ‘some’, ‘at least 3’, ‘at most 3’, ‘no’; Det_B was either ‘more than 2’ or ‘fewer than 2’. A, B, and C were arbitrary substantives like ‘forester’, ‘nurse’, ‘socialist’, etc. Each subject was taken through the individual tasks in a different, randomly determined order. The instructions explained that the arguments referred to an imaginary tennis tournament, in which stewardesses played against foresters, communists against nurses, and so on. Subjects were to decide whether arguments were valid or not, and the notion of validity was summarily elucidated, with due emphasis on the key elements (‘IF the premisses are true, the conclusion MUST be true as well’).

6.2 *Results*

The results of the experiment are presented in Table 2. To analyse these data, a repeated measures analysis of variance was conducted with three within-subject factors: the first determiner of the major premiss (Det_A), the second determiner of the major premiss (Det_B), and the orientation of the minor premiss (Minor), which was either upward or downward. As the sphericity assumption was not satisfied, we used the Huynh-Feldt epsilon. This yielded main effects for all factors (Det_A: $F(4,185) = 12.631, p < 0.001$; Det_B: $F(1,44) = 17.956, p < 0.001$; Minor: $F(1,44) = 25.507, p < 0.001$). There were interactions between Det_A and Det_B ($F(2,88) = 3.267, p < 0.05$) as well as between Det_B and Minor ($F(1,44) = 7.977, p < 0.01$). Other potential interactions failed to reach significance. Pairwise comparisons between the Det_A determiners revealed significant differences between ‘at most’ and all of the others ($p < 0.001$, throughout). Otherwise, differences between Det_A determiners were not significant. With a Bonferroni adjustment for the number of

Table 2 Percentages of correct answers ($n = 45$).

Det _A	Det _B	Minor	% (s)	Det _A	Det _B	Minor	% (s)
every	more than	all BC	91 (29)	most	more than	all BC	91 (29)
		all CB	69 (47)			all CB	67 (48)
	fewer than	all BC	71 (46)		fewer than	all BC	62 (49)
		all CB	58 (50)			all CB	60 (50)
at least	more than	all BC	96 (21)	some	more than	all BC	87 (34)
		all CB	69 (47)			all CB	67 (48)
	fewer than	all BC	53 (51)		fewer than	all BC	60 (50)
		all CB	51 (51)			all CB	62 (49)
at most	more than	all BC	51 (51)	no	more than	all BC	69 (47)
		all CB	38 (49)			all CB	53 (50)
	fewer than	all BC	36 (48)		fewer than	all BC	73 (45)
		all CB	49 (51)			all CB	64 (48)

equations, pairwise comparisons between Det_A/‘more than’ and Det_A/‘fewer than’ showed no significant differences for Det_A = ‘no’ and ‘at most’, significant differences for Det_A = ‘at least’ and ‘most’ ($t = 4.401$, $p < 0.0005$, and $t = 2.874$, $p = 0.003$, respectively), and differences of borderline significance for Det_A = ‘every’ and ‘some’ ($t = 2.319$, $p = 0.012$, and $t = 2.049$, $p = 0.023$, respectively).

6.3 Discussion

The results generally accord with our predictions. To begin with, it is evident that this *type* of task need not be particularly difficult, as we predicted. Some of the arguments are correctly assessed in over 90% of the cases, even if they involve compound quantifiers like ‘at least 2 nurses’ or quantifiers that are not first-order definable, like ‘most communists’. Secondly, the response patterns associated with individual determiners are in line with our predictions, too: while the upward entailing determiners ‘all’, ‘most’, ‘some’, and ‘at least’ evoke very similar response patterns, the downward entailing determiners are clearly different. Thirdly, as predicted, arguments with ‘at most’ are exceptionally difficult: their scores are at or below chance level in all conditions.

Note, in particular, that ‘at most’ deviates from ‘no’, even though both determiners are downward entailing, and that the contrast between ‘at most’ and ‘no’ is not mirrored by a contrast in the counterpart pair ‘at least’/‘some’, which suggests very strongly that quantifiers with ‘at most’ are difficult not for structural but for semantic reasons. Hence, the data confirm what is perhaps the most striking

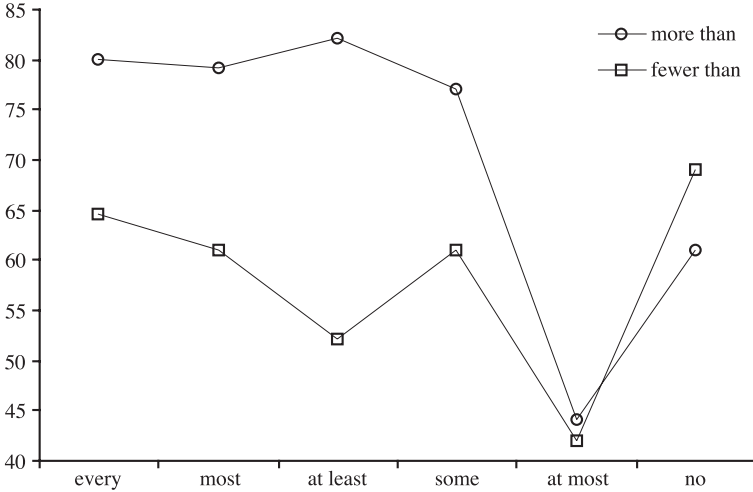


Figure 1 Interaction between Det_A and Det_B (percentages of correct answers).

prediction of our theory, viz. that downward-entailing quantifiers built from cardinals deviate from all others.

The interaction between Det_A and Det_B confirms our expectations, too (see Figure 1). The easiest arguments are those in which Det_A and Det_B are both upward entailing. With Det_A upward and Det_B downward entailing, an argument is harder to process. With $\text{Det}_A = \text{'no'}$ the pattern is reversed, and arguments with 'at most' are so complex, apparently, that the distinction between upward and downward entailing Det_B doesn't register anymore.

One notable finding is that overall arguments with 'no' do not appear to be more complex than arguments with upward entailing determiners, which goes against the general trend, in the experimental literature, for positive expressions to be easier than their negated counterparts (Clark 1974, Horn 1989). A closer look at our data reveals that the explanation for this apparent discrepancy may be in line with our speculations in section 5.2. To explain this, we set aside 'at most' and compare the arguments with $\text{Det}_A = \text{'no'}$ to the arguments in which Det_A is upward entailing. In this group, the 'no' arguments score worse than the others with $\text{Det}_B = \text{'more than'}$ and better than all the others with $\text{Det}_B = \text{'fewer than'}$. More accurately, whereas arguments with upward entailing Det_A are, on the whole, significantly easier with $\text{Det}_B = \text{'more than'}$ than with $\text{Det}_B = \text{'fewer than'}$, the difference is obliterated with $\text{Det}_A = \text{'no'}$. Hence, 'no' arguments confirm our hypothesis that sentences with mixed monotonicity profiles are more complex than others. Furthermore, it appears that the initial impression

that subjects perform surprisingly well with ‘no’ is misleading: they do well on one particular class of ‘no’ arguments, i.e. the ones in which $\text{Det}_B = \text{‘fewer than’}$. That is to say, the positive effect is due entirely to those arguments whose major premiss contains two downward-entailing determiners. This makes sense if we assume, for reasons discussed in section 5.2, that subjects tend to use the conversion method, and before making monotonicity inferences translate ‘No A played against fewer than 2 B’ into ‘All A played against at least 2 B’.

However, even if the conversion method is used to simplify an argument, the procedure for solving experimental tasks with ‘no’ sentences is more complex than for tasks with ‘most’, for example. Why isn’t this reflected in our data? First, it isn’t quite correct to say that the difference isn’t reflected in the data, for in absolute terms subjects perform worse on tasks with ‘no’ than on tasks with upward entailing determiners: with ‘no’ 64.75% of the answers are correct, while the upward entailing determiners range between 67.75% and 72.25% (for ‘at least’ and ‘every’, respectively) with a mean of 69.6%. These differences are not statistically significant, to be sure, but it is at least possible that a weak effect would be found for a larger n . Secondly, it should be noted that the experimental task in our study is different from the experimental tasks that have been used to argue that negative expressions are more complex than positive ones. The cognitive processes studied here are slow and long-winded in comparison to, for instance, Clark’s (1974) reaction time experiments, which makes it less surprising that effects demonstrated in one paradigm fail to appear in another.

The observed interaction between Det_B and Minor seems a bit puzzling at first. As it turns out, if Det_B is upward entailing the orientation of the minor premiss is significant: if it goes upwards, the average rate of correct responses is 0.81 and this figure drops to 0.61 if the direction of the argument is reversed. For downward entailing Det_B , by contrast, the two rates nearly coincide: 0.59 and 0.57, respectively. Why is it that the orientation of the minor premiss makes a difference only in the case of an upward entailing Det_B ? The answer to this question, we suspect, is simply that this is a floor effect: subjects’ performance approximates chance level already with downward entailing Det_B and upward Minor, so the effect of choosing a downward Minor instead is bound to be limited.

7 CONCLUDING REMARKS

Natural language semantics is not a branch of psychology. Rather, we view linguistic meaning as a topic lying at the crossroads of several

disciplines, of which psychology is one. This being so, it is to be expected that semantics is directly relevant to human psychology. The results presented in the foregoing illustrate how semantic theory can be brought to bear on behavioural data.

We have proposed a (partial) semantic measure of linguistic complexity, and derived from it predictions about the behaviour of experimental subjects, without using a processing model. The predictions made by our account are based on a competence theory of quantification, which is motivated mainly, though not exclusively, by semantic evidence. Therefore, in some quarters of psychology, our account would not qualify as psychological. Most theories in the psychology of reasoning, for instance, involve a detailed architecture of the reasoning faculty, and it is often supposed, overtly or tacitly, that this is a *sine qua non* for psychological theorizing.

In our opinion, such views are seriously mistaken. Like all natural phenomena, behaviour can be described and explained on many levels, and we see no reason to believe that any one of these levels is more pertinent to psychology than the others. Part of the game of science is to find the level at which a given phenomenon is best explained, and for certain aspects of human reasoning, we maintain, that level is semantical.

Acknowledgements

We would like to thank Larry Horn, Keith Stenning, and the anonymous reviewers for the *Journal of Semantics* for their helpful comments on earlier versions of this paper.

BART GEURTS AND FRANS VAN DER SLIK

Received: 28.05.04

Philosophy Department/Department of Linguistics

Final version received: 12.10.04

University of Nijmegen

P. O. Box 9103

NL-6500 HD Nijmegen

The Netherlands

e-mail: bart.geurts@phil.ru.nl

REFERENCES

- Barwise, J. & R. Cooper (1981) 'Generalized quantifiers and natural language'. *Linguistics and Philosophy* 4:159–219.
- Carston, R. (1998) 'Informativeness, relevance and scalar implicature'. In R. Carston and S. Uchida (eds). *Relevance Theory: Applications and Implications*. 179–236.
- Clark, H. H. (1973) 'Space, time, semantics, and the child'. In T. Moore

- (ed.). *Cognitive Development and the Acquisition of Language*. Academic Press. New York, 27–63.
- Clark, H. H. (1974) 'Semantics and comprehension'. In T. Sebeok (ed.). *Current Trends in Linguistics*, volume 12. Mouton. The Hague, 1291–1428. Also published separately under the same title: 1976, Mouton, The Hague.
- Geurts, B. (1998) 'Scalars'. In P. Ludewig & B. Geurts (eds). *Lexikalische Semantik aus kognitiver Sicht*. Gunter Narr. Tübingen, 85–117.
- Geurts, B. (2002) 'Donkey business'. *Linguistics and Philosophy* 25:129–156.
- Geurts, B. (2003a) 'Reasoning with quantifiers'. *Cognition* 86:223–251.
- Geurts, B. (2003b) 'Monotonicity and syllogistic inference: a reply to Newsstead'. *Cognition* 90:201–204.
- Heim, I. (1984) 'A note on negative polarity and downward entailingness'. In *Proceedings of NELS 14*, pp. 98–107.
- Horn, L. R. (1972) On the semantic properties of the logical operators in English. Ph. D. thesis, University of California at Los Angeles.
- Horn, L. R. (1989) *A Natural History of Negation*. Chicago University Press. Chicago.
- Horn, L. R. (1992) 'The said and the unsaid'. In C. Barker & D. Dowty (eds). *Proceedings of SALT 2*, pp. 163–192.
- Kanazawa, M. (1994) 'Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting'. *Linguistics and Philosophy* 17:109–158.
- Keenan, E. L. & D. Westerståhl (1997) 'Generalized quantifiers in linguistics and logic'. In J. van Benthem & A. ter Meulen (eds). *Handbook of Logic and Language*. Elsevier/MIT Press. Amsterdam/Cambridge, MA, pp. 837–893.
- Levinson, S. C. (2000) *Presumptive Meanings*. MIT Press. Cambridge, MA.
- McGonigle, B. & M. Chalmers (1986) 'Representations and strategies during inference'. In T. Myers, K. Brown, & B. McGonigle (eds). *Reasoning and Discourse Processes*. Academic Press. London, 141–164.
- McGonigle, B. & M. Chalmers (1996) 'The ontology of order'. In L. Smith (ed.). *Critical Readings on Piaget*. Routledge. London, 279–311.
- Musolino, J. (2004) 'The semantics and acquisition of number words: integrating linguistic and developmental perspectives'. *Cognition* 93: 1–41.
- Musolino, J., S. Crain, & R. Thornton (2000) 'Navigating negative quantificational space'. *Linguistics* 38:1–32.
- Newstead, S. E. (2003) 'Can natural language semantics explain syllogistic reasoning?' *Cognition* 90:193–199.
- Noveck, I. (2001) 'When children are more logical than adults: experimental investigations of scalar implicature'. *Cognition* 78:165–188.
- O'Leary, C. & S. Crain (1994) 'Negative polarity items (a positive result), positive polarity items (a negative result)'. Paper presented at the 19th Boston University Conference on Language Development.
- Papafragou, A. & J. Musolino (2003) 'Scalar implicatures: experiments at the syntax-semantics interface'. *Cognition* 86:253–282.
- Sánchez Valencia, V. (1991) Studies on natural logic and categorial grammar. Doctoral dissertation, University of Amsterdam.
- van der Wouden, T. (1997) *Negative Contexts: Collocation, Polarity and Multiple Negation*. Routledge. London.