University of Cambridge
Faculty of English
Research Centre for English and Applied Linguistics

The interpretation and use of numerically-quantified expressions

Christopher Raymond Cummins

Trinity College

Dissertation submitted for the degree of Doctor of Philosophy

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

No part of this dissertation has been submitted for any other qualification at any University.

This dissertation does not exceed the regulation length, including footnotes, references and appendices but excluding the bibliography[1].

---

[1] Nor does it exceed the regulation length if the bibliography is included, even though the 'but' clause of the above declaration might be assumed to implicate this. This could be analysed as a failure of implicature due to adherence to a priming constraint, viz. the need to quote the above declaration literatim from the Faculty regulations (see section 5.4.2 for a corresponding example in the numerical domain).

ABSTRACT

This thesis presents a novel pragmatic account of the meaning and use of numerically-quantified expressions. It can readily be seen that quantities can typically be described by many semantically truthful expressions – for instance, if 'more than 12' is true of a quantity, so is 'more than 11', 'more than 10', and so on. It is also intuitively clear that some of these expressions are more suitable than others in a given situation, a preference which is not captured by the semantics but appears to rely upon on wider-ranging considerations of communicative effectiveness.

Motivated by these observations, I lay out a set of criteria that are demonstrably relevant to the speaker's choice of utterance in such cases. Observing further that it is typically impossible to satisfy all these criteria with a single utterance, I suggest that the speaker's choice of utterance can be construed as a problem of multiple constraint satisfaction. Using the formalism of Optimality Theory (Prince and Smolensky 1993), I proceed to specify a model of speaker behaviour for this domain of usage.

The model I propose can be used to draw predictions both about the speaker's choice of utterance and the hearer's interpretation of utterances. I discuss the relation between these two aspects of the model, showing how constraints on the speaker's choice of utterance are predicted to make pragmatic enrichments available to the hearer. I then consider applications of this idea to specific issues that have been discussed in the literature. Firstly, with respect to superlative quantifiers, I show how this model provides an alternative account to that of Geurts and Nouwen (2007), building upon that offered by Cummins and Katsos (2010), and I present empirical evidence in its favour. Secondly, I show how this model yields the novel prediction that comparative quantifiers give rise to implicatures that are conditioned both by granularity and by prior mention of the numeral, and demonstrate these implicatures empirically. Finally I discuss the predictions that the model makes about the frequency of quantifiers in corpora, and investigate their validity.

I conclude that the model presented here proves its worth as a source of hypotheses about speaker and hearer behaviour in the numerical domain. In particular, it serves as a way to integrate insights from distinct domains of enquiry including psycholinguistics, theoretical semantics and numerical cognition. I discuss the claim of this model to psychological plausibility, its relation to existing approaches, and its potential utility when applied to broader domains of language use.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

In 1970, Richard Montague advanced the view that formal semantics could be used to analyse natural language. According to this account, the meaning of natural language expressions could adequately be characterised in the same terms used to account for meanings of logical and mathematical expressions, using tools such as set theory, predicate logic and lambda calculus.

Such an approach appears to offer a natural solution to problems concerning the semantic treatment of quantified expressions. In (1)-(4), I illustrate this with reference to simple expressions, presenting glosses of these expressions in set theory, and glosses of these glosses in English[2].

(1)   All Xs are Y.

$[X] \setminus [Y] = \emptyset$

The set of Xs contains no members that are not also members of Y.

(2)   No Xs are Y.

$[X] \cap [Y] = \emptyset$

The set of Xs and the set of entities that are Y have no members in common.

(3)   Two Xs are Y.

$|[X] \cap [Y]| = 2$

The set of Xs and the set of entities that are Y have exactly two members in common.

(4)   More than 2 Xs are Y.

$|[X] \cap [Y]| > 2$

The set of Xs and the set of entities that are Y have more than two members in common.

As observed by Geurts et al. (2010: 131), treatments of this type are prevalent in linguistics (Barwise and Cooper 1981), psycholinguistics (Moxey and Sanford 1993) and the psychology of reasoning (Evans et al. 1993). In recent years, some semantic accounts of

---

[2] English is used here, as is customary, as a metametalanguage (OED 1989, IX: 669, 674) in which to describe logical expressions. This results in the existence of 'false friends' in formal semantics: for instance, the logical operator IF is not adequate for characterising the semantics of English 'if'.

specific quantifying expressions have dissented from the obvious and proposed more nuanced formalisations (e.g. Geurts and Nouwen 2007 for 'at least'/'at most', Hackl 2009 and Solt 2010 for 'most' vs. 'more than half'), but for a core range of expressions this type of semantic account appears broadly adequate.

In this thesis, I focus on the broader question of the interpretation and use of numerically quantified expressions: that is, on the pragmatics as well as the semantics of such expressions. To motivate this enquiry, I examine the limitations of semantic accounts for numerically quantified expressions. I then discuss the role of pragmatic factors in answering two questions about numerical quantification: in a given situation, how do we decide which expression to use? And precisely what information is conveyed by the expression we eventually select?

This is not a trivial problem. An intuitive answer to the first question is that we choose an expression that enables us to make a statement that we believe to be true. However, in semantic terms, this is scarcely an adequate answer when we are talking about numerical quantification. Through the structure of the number system (and its relation to observable reality), numerically quantified expressions participate in extensive entailment relations. 'More than n', for instance, entails 'more than m' for any $m < n$. Thus, typically, there are arbitrarily many truth-conditionally correct statements that can be made in any situation requiring numerical quantification. On purely semantic grounds, none of these could be ruled out of consideration.

For example, suppose you are asked (for the speaker's information) how many times Elizabeth Taylor was married, and you believe that the answer lies in the range 6-8. It seems intuitively plausible that you might reply using any of the following quantifiers, among others.

(5a)   between 6 and 8 (inclusive)

(5b)   6, 7 or 8

(5c)   about 7

(5d)   more than 5

(5e)   not more than 8

(5f)   fewer than 9

(5g)   not fewer than 6

(5h)   at least 6

(5i)   at most 8

Of these, (5d-g) clearly exhibit entailments of the type discussed above: that is to say, the following expressions would also be semantically valid alternatives[3].

(6d)   more than 4, 3, 2...

(6e)   not more than 9, 10, 11…

(6f)   fewer than 10, 11, 12…

(6g)   not fewer than 5, 4, 3…

As (5a) combines the meaning of (5e) and (5g), the following would also be valid alternatives.

(6a)   between 5 and 9, 4 and 9, 5 and 10, 3 and 9…

Thus it appears inevitable that responding to a question of this type involves choosing between a panoply of quantified expressions that are valid on their semantics. A similar multiplicity of options can arise even in situations in which we know the correct answer. Suppose that the question asked is 'How high is Mount Everest?', and we are confident that the answer is 29,028 feet. The following quantifiers could all be legitimate, depending on the specificity requirements of the situation.

(7a)   29,028

(7b)   exactly 29,028

(7c)   29,000

(7d)   about 29,000

(7e)   30,000

(7f)   about 30,000

It may not be necessary or desirable to give the most accurate response (7a-b), as this may be communicatively inefficient. Instead, we may wish to give the answer to the nearest round number, as in (7c) and (7e). However, these are ambiguous, as they could be interpreted as

---

[3] This is debatable in the case of (1h) and (1i); see Geurts and Nouwen (2007), Geurts et al. (2010).

either approximate or precise values, so it may be necessary to add a modifier such as 'about' (7d, 7f)[4].

Indeed, even if we know the precisely correct answer and rounding is not possible, the entailment relations still give rise to infinitely many options. If I am sure that Elizabeth Taylor married exactly eight times, all the responses (5a-i) and (6a,d-g) are still semantically appropriate. Given that the most precise response may not be the most apposite, as seen in the Everest example, it is difficult to dispense with these unwanted options in a principled fashion.

Thus, given the range of possible statements that can be made, it seems natural to ask how we choose among these in actual conversation: how are numerically quantified expressions used? This is a speaker-referring question. We might also ask the corresponding hearer-referring question: how are numerically quantified expressions interpreted? If the speaker's behaviour depends upon pragmatic considerations, then the hearer should be able to draw corresponding pragmatic enrichments from the speaker's choice of utterance. For this reason, an account of the speaker's usage preferences should also make predictions about the hearer's interpretative preferences, a point which will recur throughout this thesis.

Current understanding of the criteria governing the speaker's choice of quantifier does not appear to be particularly advanced. An appropriate starting point might be Grice's maxims. Grice (1975, 1989) attempted to enumerate a set of pragmatic principles that govern conversation, based upon the fundamental assumption that speaker and hearer are engaging in cooperative social behaviour. Applying his maxims, we would expect to find that the speaker's choice of utterance is generally governed by the need to be truthful, informative (to the appropriate degree), brief, relevant and clear. However, with reference to the specific case of numerical quantification, we might also expect certain domain-specific factors to be at play. For example, numbers may differ in their intrinsic salience, and numbers low in salience would presumably be dispreferred for this reason. Similarly, quantifiers may differ in complexity, from a cognitive perspective. Moreover, the salience of a number or

---

[4] It has been claimed that the original measurement of Everest was precisely 29,000 feet, but was reported as 29,002 to avoid the appearance of rounding. In fact, as this measurement relied upon an estimate for the coefficient of refraction (de Graaff-Hunter 1955), it would be more accurate to say that the estimate was made in such a way as to avoid a round outcome.

quantifier may vary according to its status in the discourse: for instance, whether or not it has been mentioned or alluded to before.

For specific types of numerically quantified expressions, linguistic and psycholinguistic studies have investigated the semantic and pragmatic components of meaning in some detail. There has been much debate as to the core meaning of numerals, and specifically whether these are punctual ('exactly *n*') or lower-bounding ('at least *n*'). Bultinck (2005) considers this question in detail by appeal to corpus data, and Breheny (2008) discusses more recent theoretical and empirical developments. Several researchers have addressed the meaning of comparative and superlative quantifiers (those of the form 'more than *n*' and 'at least *n*' respectively): Geurts and Nouwen (2007) advance the idea that superlative quantifiers possess modal semantics, which is supported empirically by Geurts et al. (2010), but alternative accounts are offered by Nouwen (2010) and Cummins and Katsos (2010). However, of these proposals, only the last-mentioned touches upon the question of how the meanings are related to the speaker's choice of a specific expression in a given situation.

A similar picture emerges from the literature on the psychology of number and quantification. Major debates concern the representation of number (Dehaene 1997, Butterworth 1999) and whether quantifier complexity has a neuropsychological correlate (McMillan et al. 2005, McMillan et al. 2006, Szymanik and Zajenkowski 2010). However, little use has been made of these findings in attempting to account for speakers' use of numerical quantification, despite the acknowledged desirability of interdisciplinary approaches to this topic (see for instance Musolino 2004).

The most notable exceptions to this generalisation are studies appealing to the notion of relevance, as articulated by Sperber and Wilson (1986/1995). From a relevance-theoretic perspective, the selected utterance should be that which is optimally relevant given the speaker's circumstances: that is, it will yield maximal cognitive effects for minimal cognitive effort on the part of the hearer.

Research on the reporting of time particularly well exemplifies the relevance theory (RT) approach to numerical communication. In a series of studies, Van der Henst, Carles and Sperber (2002; see also Van der Henst and Sperber 2004) asked passers-by for the time, using three distinct approaches. In the control condition, the enquiry was made neutrally, without any specific reason being given. In the other two conditions, the experimenters expressed a

particular interest in accuracy, telling the participant either that they wanted to reset their watches or that they had an appointment at a particular time. Van der Henst and colleagues observed that the participants rounded the time they reported (e.g. responding 'quarter past two' for 2:13pm) significantly more frequently in the control condition than in the other conditions. This held both for wearers of digital and analogue watches, despite the extra processing load presumed to arise when rounding a digital time. However, wearers of analogue watches did give rounded responses more frequently than wearers of digital watches. This line of enquiry was further developed by Gibbs and Bryant (2008), who replicated the findings of the earlier studies and also analysed the role of procedural cues and filled pauses in the hearer's understanding of the responses.

As discussed by Van der Henst and Sperber (2004), such studies provide strong evidence that the choice of numerical expression in answering time questions involves the interaction of distinct factors. Rounding, they argue, enables speaker and hearer to work with values that are more cognitively salient, i.e. those corresponding to the marks on the analogue clock-face. They further suggest that the use of rounding reduces the speaker's commitment to the accuracy of their utterance, and therefore might be favoured if the speaker doubts the precision of their own information – that is, if they are willing to sacrifice informativeness to retain truthfulness. In their model, these notions are subsumed under the general heading of relevance, as they go to ensuring that the interaction is communicatively effective. Additionally, they tacitly assume that the wearers of digital watches would be naturally disposed to give precise answers, presumably because these answers are literally spelt out on the watch-face. Gibbs and Bryant (2008: 364) are more explicit about this: they argue that the use of rounding in such cases shows that 'answering questions is not guided by a desire to say what is most truthful by giving the exact time or done egocentrically given what is easiest to produce. Instead, people aim to speak in an optimally relevant manner'. According to this account, the use of a precise answer is more accurate than rounding, and – for wearers of digital watches – reading out the precise time is also assumed to be less effortful than rounding. However, considerations of utility to the hearer override these in determining the speaker's choice of answer.

In sum, several distinct factors governing the nature of the response to time questions can be identified in this literature – these include truthfulness, accuracy, cognitive effort, and relevance to the communicative purpose. To this we might add the contextual activation of

the numerals: it is apparent that the presentation of time in numerals on the digital watch primes the speaker to give a precise answer, to a greater extent than the analogue presentation does. Simplicity of expression appears also tacitly to play a role: witness the preference for rounded answers without modification by 'about'. It can convincingly be argued that these factors interact in determining what constitutes the preferred utterance in a given situation.

It is tempting to argue that these accounts are built upon plain common sense, and moreover that they add little to plain common sense. In particular, they are coy about how the contributory factors to relevance interact. The speaker is assumed to produce an optimally relevant utterance by balancing considerations of the effort needed by and the effects available to the hearer, while also factoring in their own abilities and interests (Wilson and Sperber 2002: 257). In the time-reporting example, the speaker is assumed to take into account the informational needs of the hearer, the cognitive effort that the hearer would require to meet these needs from a given utterance, the cognitive effort necessary for the speaker to formulate the utterance, and perhaps the precision and reliability of the information available to the speaker. By weighing up these considerations, the speaker selects the optimal utterance. The question of how the speaker performs this complex task is left less fully explored in this literature.

In this thesis, I attempt to address this limitation by offering a theory of how the speaker selects a numerically-quantified utterance. In doing so, I remain broadly sympathetic to the relevance-based approach, but aim to codify the factors bearing upon the speaker's decision in such a way as to enable testable predictions to be drawn about usage and interpretation. Thus, the remainder of this thesis is organised as follows.

In **chapter 2**, I postulate a framework within which the contributory factors to the speaker's choice of utterance can be treated as constraints, and within which their interaction can be modelled. I then populate this framework with a set of constraints, drawn from the existing literature and (where necessary) supported by novel experimental data.

In **chapter 3**, I explore how testable predictions can be drawn from this framework, and how these predictions are mediated by various possible assumptions as to the behaviour of constraint-based systems.

In **chapter 4**, I turn to the topic of comparative and superlative quantifiers, reviewing the recent literature and considering the advantages of a constraint-based account for these data.

In **chapter 5**, I present novel predictions on the scalar implicatures arising from modified numerals, and new empirical data which bears out these predictions.

In **chapter 6**, I discuss corpus data, and test further predictions of the constraint-based model with respect to corpus frequencies of numerically-quantified expressions.

In **chapter 7**, I conclude by reviewing the previous chapters and discussing the possible future directions of the research programme outlined here. I discuss the nature of quantity representations, possible ways of recasting constraints in more intuitively satisfactory terms, and the extension of this proposal to other domains of language use.

2. CONSTRUCTING A CONSTRAINT-BASED MODEL

2.1. Introduction

The goal of this chapter is to set out a constraint-based model for the selection of numerically quantified expressions. This task comprises two sub-tasks: enumerating the set of relevant constraints, in some principled fashion, and establishing how these constraints interact.

The determination of the constraints on the choice of expression, and the nature of their interaction, is already a central issue in various areas of linguistics and cognitive science. For example, one strand of literature concerns the form of referential expression used by a speaker: that is, whether an entity is picked out by a pronoun, a noun phrase, or a modified noun phrase such as an adjectival or prepositional phrase, or whether the referential expression is not realised at all. This has been shown to depend upon numerous factors, including the referent's status in the common ground, how the entity was previously referred to, the referential context (e.g. the presence or absence of other competitor entities of the same kind), the discourse genre (e.g. narrative vs. instructional), politeness considerations, and perceptual salience (some attributes, such as colour, being referred to more often even if they do not contribute to the unique identification of the referent), among others (e.g. Arnold 2008, Gordon et al. 1993, Gundel et al. 1993, Sedivy 2003; for a review see Almor and Nair 2007).

The problem of formulating an optimal expression can thus be construed as a constraint satisfaction problem. As speakers, we are simultaneously confronted with a number of potentially conflicting demands, which we aim to reconcile as best we can. In the case of referential expressions, researchers have already begun to explore how constraints such as those discussed in the previous paragraph interact, and the precise time-course over which these constraints apply. For instance, the relation between entities' status in the common ground and the interpretation of adjectivally-modified descriptions of these entities is discussed by Hanna et al. (2003), Nadig and Sedivy (2002), Horton and Keysar (1996) and others. Moreover, much recent work is dedicated to integrating such insights from the psychology of language into computational language generation models (see Krahmer et al. 2003, van Deemter 2006, Viethen and Dale 2008).

With specific reference to numerical quantification, previous research (as discussed in the previous chapter) has shown that various considerations can bear upon the speaker's choice

of expression. Some of these factors, arising from Gricean pragmatic considerations, may be presumed general to a wide range of discourse environments (e.g. simplicity and accuracy), while others are specific to the numerical domain (e.g. numeral salience). In any given situation, several of these factors are plausibly relevant to the choice of utterance. It is also clear that we cannot generally satisfy all these considerations simultaneously. Taking the Everest example, repeated below, we are forced to make a trade-off between accuracy and numerical salience, assuming that the precise response (7a) uses a less salient numeral than the imprecise (7c) or (7e). Similarly we balance accuracy (broadly construed) and simplicity (or brevity) when we choose between (7c) and (7d), or between (7e) and (7f): (7c) and (7e) are potentially misleading in that they might be understood to convey precise values, but are more economical than their unambiguous counterparts (7d) and (7f).

(7a)   29,028
(7b)   exactly 29,028
(7c)   29,000
(7d)   about 29,000
(7e)   30,000
(7f)   about 30,000

It seems natural to articulate this intuition in terms of a constraint-based model such as Optimality Theory (OT), as set out by Prince and Smolensky (1993). OT is a paradigm for analysing systems of violable constraints. It provides an algorithmic model for selecting optimal output, given an input and a constraint set. In the following sections I outline the constitution of an OT system, and then consider how to construct a system that is appropriate for the purpose at hand.

2.2 OT modelling of the speaker's choice of utterance

A fully-specified OT system fulfils the function of selecting an optimal output candidate given an input. In phonology, the area for which OT was initially developed, the input is understood as an underlying form, and the output is the corresponding surface form. OT serves the function of mapping between these two levels. In order to address the question that motivates this enquiry – how a speaker selects which utterance to use among a wide range of possible utterances – I propose an account in which the input is the situation (broadly construed) and the output is the speaker's utterance. Technically this will be a

speaker-oriented unidirectional OT account. I discuss its relation to other styles of OT in chapter 3, with particular reference to bidirectional OT, which is used in the pragmatics literature to address a related but distinct set of research questions[5]. According to this model, the speaker's optimal utterance will be that which best satisfies a ranked set of constraints given the situation at hand.

To propose that the speaker optimises the utterance with respect to the situation, we require a working definition of 'situation' for this model. In principle, the 'situation' here might embrace any or all contextual considerations known to the speaker, including the preceding linguistic context (to the extent that the speaker is aware of it) and the speaker's own psychological state, including their knowledge of language and their communicative intention. Notions such as 'question under discussion' might also be considered part of the situation, if they are known to the speaker. Within this model, the speaker may potentially be optimising the choice of utterance with respect to all of these considerations.

The model proposed here is intended to predict the speaker's choice of utterance, given the situation. Therefore, constraints in such a model must refer either to the utterance in isolation, to the situation in isolation, or to the relation between utterance and situation. If we are to take the speaker's perspective, then it is clear that we must restrict the notion of 'situation' to that which is known to the speaker, as logically aspects of the situation that were entirely unknown to the speaker could not be said to condition the speaker's choice of utterance. In effect, then, I will be working with a definition of situation that is restricted to information available to the speaker – this can be read as 'the speaker's representation of the situation'. The question of what constitutes the situation is then simply a restatement of the question of what factors are relevant to the speaker's choice of utterance.

If a model of this type were to prove descriptively adequate, it would specify exactly what categories of contextual information were potentially relevant to the choice of utterance: namely, only those categories of contextual information that are referred to by the constraints in the model. If some category of contextual information is not referred to by any constraint, then this model simply does not make use of that information. However, the adequacy of

---

[5] OT approaches generally offer the notable advantage that the absolute extent of constraint violations is not critical, merely the relative extent of violations by different candidate outputs. This makes them particularly suitable for the type of modelling undertaken here, where the absolute extent of constraint violation would typically be an open empirical question, as we shall see in section 2.4.

such a model would not necessarily suppose that all the sources of information alluded to by its constraints were fundamental to the selection process, because an adequate model might nevertheless contain redundant constraints.

Before going further, let us also consider the hearer's perspective. How can interpretation proceed effectively within such a system? This cannot be a mirror image of the speaker's task, not least because the same utterance may be optimal for distinct situations[6]. Moreover, while the speaker takes the entire situation as input and produces an utterance as output, the hearer – starting from the utterance – does not need to reconstruct the whole situation. Much of this information is already known to the hearer: for instance, which numbers and quantifiers are salient in the preceding discourse.

Instead, I assume that the hearer's goal is to reconstruct the speaker's communicative intention, given knowledge of the speaker's optimal utterance, as well as the hearer's own independent knowledge about the situation. This enables the hearer to draw inferences about the speaker's communicative intention, thus filling in the crucial gap in their knowledge about the situation (from the speaker's perspective).

This approach closely resembles a system such as Dual Optimization (Smolensky 1996). From that perspective, production and comprehension are distinct processes. In production, the interpretation is fixed and the optimal expression is selected given that interpretation. In comprehension, the expression is fixed (by virtue of having been uttered) and the optimal interpretation is selected given that expression.

As this thesis predominantly addresses the problem of usage, I will not attempt to construct a general model of the hearer's interpretation as a counterpart to the model of the speaker's choice of utterance. However, in chapter 5 I discuss and exemplify how a rational hearer should behave under the assumptions of this model, and present experimental data in support of this account of hearer behaviour.

---

[6] Under a definition of 'situation' that includes reference to prior linguistic context, the same situation typically never occurs twice. If we neglect context, I would still argue that the number of distinct communicative intentions that a speaker might hold far outweighs the number of distinct expressions available for conveying these intentions, in which case it follows that the mapping from expression to intention cannot be injective (i.e. it is not generally possible to establish the speaker's intention with certainty given their utterance).

2.3 Constitution of an OT system

Having outlined the type of model being proposed here and briefly considered its consequences for speaker and hearer, I now turn to the technical preliminaries for such a model.

The three major components of a classical OT system are as follows.

- GEN, the candidate generation system. This generates the list of possible outputs.
- CON, a ranked set of constraints.
- EVAL, the evaluation system. This assesses the extent to which each of the candidates violates the constraints.

Given an appropriate input, the system functions as follows.

- GEN generates a set of output candidates.
- For each candidate, EVAL walks the list of ranked constraints from highest- to lowest-ranked. It assigns violations to candidates for each failure to comply with the constraints.
- An optimal candidate is selected by the following procedure.
  - Any candidates incurring more than the minimum number of violations of the top-ranked constraint are excluded.
  - This step is repeated for the next highest-ranked constraint, on the remaining candidates.
  - This continues until only one candidate is left. This is the optimal candidate.

In OT, the set of constraints is usually held to be universal. Languages, and indeed idiolects, are distinguished from one another by their constraint rankings. It is nevertheless an architectural imperative of OT that constraints are never deactivated; however, by being lowly-ranked, they cease to be relevant in a wider range of contexts.

To take a specific example from the domain of phonology, it is posited that there is a constraint against syllabic codas, the NOCODA or -COD constraint (Prince and Smolensky 1993: 93). In some languages, such as Japanese and Hawaiian, this appears to be highly ranked and has the effect of prohibiting all syllabic codas: underlying coda consonants are

resyllabified or omitted in the surface form in order to avoid violations of NOCODA. In other languages, such as English, the constraint is much lower ranked; in particular, it is ranked below MAX, a constraint requiring that all underlying segments should have corresponding realisations in the output. Under this ranking, underlying coda consonants are realised (satisfying MAX) rather than omitted (which would satisfy NOCODA), but are still disfavoured and consequently prone to processes such as resyllabification (which, where possible, satisfies both MAX and NOCODA). This is of course merely a snapshot of part of the system, and other constraints may influence what happens to underlying codas in any specific instance. In any case, the system is deterministic, with the constraint ranking yielding a preferred surface form for any given underlying form.

Constraints in OT may more generally be divided into two classes: faithfulness and markedness constraints. Faithfulness constraints govern the relation between the input and output. A phonological example of this is the constraint DEP, 'don't epenthesise', requiring that segments are not inserted in the output form that are not present in the input (McCarthy 2002: 13). Markedness constraints govern the output, and apply irrespective of the input. NOCODA, discussed above, is a markedness constraint.

From the analyst's point of view, the task of building an OT system consists of establishing the constraint set, which includes defining precisely what constitutes a violation of each constraint. Given the nature of the OT system, this is not a trivial task. The constraints cannot usefully relate to Chomskyan universals, as all OT constraints are violable under some ordering, so nothing that is universally required or prohibited can usefully be encoded in constraint-based terms[7]. However, linguistic typology is a more promising hunting-ground for plausible constraints. In an OT system, constraints may help us to account for why some things occur less frequently than others (e.g. that syllabic codas are less widespread than syllabic onsets).

In order to populate the list of constraints, I follow the line suggested by McCarthy (2002: 41f): 'Take an intuition or observation about language and restate it as a constraint: that is, formulate it as a simple, unadorned demand or prohibition…. [Then] begin studying the

---

[7] In principle, a property mandated by constraint might be universal if it happens that all speakers rank this constraint highly. In section 2.5.1 I consider truthfulness as a possible example of this. However, little is gained by adopting such a constraint: instead, we can think of universal principles acting alongside a constraint-based model to exactly the same effect.

typological and interactional capacities of the hypothesized constraint.' Here I will take this approach to numerically quantified expressions. I will propose constraints based on the previous literature, on introspection and on functional considerations, and attempt to further demonstrate their plausibility by appeal to empirical data. These will include both markedness constraints (which hold certain output forms to be intrinsically unfavourable) and faithfulness constraints (which hold certain output forms to be incompatible with the demands of the situation). In the following chapter I will consider how these constraints interact within the type of system we are describing, and discuss how to test the resulting predictions.

2.4. Proposed constraints and their empirical basis

In the following sections, I propose a set of constraints that affect the choice of numerically quantified expressions. I will discuss these constraints with reference to various kinds of numerically quantified expressions, including bare numerals, simple quantifiers ('all', 'some'), and comparative and superlative quantifiers ('more than', 'at least'). I aim to demonstrate that each of these constraints is individually implicated in the speaker's choice of utterance and, in accordance with the above discussion, that this is also reflected in the hearer's interpretation of utterances.

2.4.1. Informativeness

The need to be informative is clearly a relevant criterion when selecting utterances, whether this is construed in terms of Grice's maxim of quantity (1975/1989), Horn's Q-principle (Horn 1984) or a relevance-based account (Sperber and Wilson, 1986/1995). However, as discussed with reference to Van der Henst and colleagues' work on the reporting of time, not all attested utterances are optimally informative, inasmuch as speakers sometimes refrain from supplying the most detailed information available to them. Therefore, on the grounds that informativeness is preferred but need not be rigorously adhered to, it seems plausible to conceive of it as being governed by a constraint in our system.

For Grice, informativeness is coupled with a notion of relevance, in that Grice's maxims enjoin interlocutors to exchange as much information as they can, to the extent that it would be relevant to the purpose of the conversation. Relevance theory adopts a similar approach. However, it is possible to disentangle informativeness from the other factors that contribute

to relevance. In this section I address informativeness *per se*, returning later to consider the related issues of granularity and truthfulness.

For numerically quantified expressions, a first approach to characterising informativeness could be to work in terms of truth-conditions: an informative expression seems to be one that leaves few possibilities open truth-conditionally. This is compatible with the notion of informativeness used in the literature on scalar implicature and focus (cf. Chierchia 2004, Krifka 1995). On this view, the maximally informative expression is an exact number, which admits only one possibility for the relevant value and excludes all others. Approximations and ranges of possibilities are less favourable with respect to this constraint; single bounds, as encoded by expressions such as 'more than *n*', are typically less informative still. The role of this constraint should then be, in effect, to assign violations for the failure to exclude possibilities that are known to be false.

However, to formalise a constraint along these lines, we need to resolve some issues. First, approximations tend to have fuzzy boundaries, so it is not clear how to quantify the number of possibilities admitted by an expression such as (8a). What precisely does this exclude? Furthermore, as suggested by Dehaene (1997) and Krifka (2009) i.a., round numbers seem to have an approximative semantics, so the same issue arises. For instance, (8b) seems plausibly to be an approximation, while this is intuitively not the case for (8c) and (8d)

(8a)   There were about 100 people there.
(8b)   There were 100 people there.
(8c)   There were exactly 100 people there.
(8d)   There were 98 people there.

The further and much-discussed question arises of whether the unmodified numeral is punctual in its semantics, or whether it is lower-bounding. In the latter case, (8b) and (8d) are both much less informative than (8c), both being compatible with infinitely many distinct possibilities. This model does not adjudicate directly on semantic questions such as this: in fact, under its assumptions, semantic considerations determine whether these candidates are available for selection in a particular situation (see section 2.5.1). For reasons of expediency I will assume where applicable that numerals have punctual rather than lower-bounded semantics in the analyses presented here, following the tacit assumptions of the bidirectional OT work of Krifka (2009). However, if we were to assume instead that numerals possessed

lower-bounded semantics, or were ambiguous between precise and lower-bounded readings, then similar analyses would be tractable, appealing to slightly different sets of constraints. For instance, the numeral priming constraint would be called upon more extensively, as I briefly discuss in section 2.4.6.

A further problem rears its head in the case of scaling. For large numbers, approximations admit more distinct possibilities than for small numbers – is it appropriate to label these as absolutely less informative? For instance, (9a) must be much less informative than (9b) on this measure.

(9a)   A million people live in the city.
(9b)   A thousand people live in the village.

However, although counter-intuitive, this is not a practical problem, because numerals of different orders of magnitude are not competing for selection in a given context. Hence their levels of absolute informativeness need not be comparable. That is, we never have to choose between 'a thousand' and 'a million' on the basis of informativeness, because the ranges of values they express do not overlap, and therefore it is never the case that both could be accurately used to describe a situation. Hence, for any given situation, at least one of these is ruled out on factual grounds.

More problematic is the case of non-cardinal usage of numerical quantification. If we give a measurement, for instance, using an expression such as (10), this still admits infinitely many real-valued possibilities.

(10)   The car is 4-5 metres long.

For cases such as this, it is clear that we cannot count the number of possibilities admitted, unless we make some kind of stipulation as to which possibilities are allowed. For the moment I lay aside the issue of non-cardinal usage, but return to this issue briefly in Chapter 7, where I discuss how we can revise our notion of informativeness to encompass this category of expressions.

A similar problem arises in the case of single bounds, such as (11a) and (11b). Semantically speaking, these admit infinitely many possibilities, and are therefore equally (un)informative

in these terms. Moreover, they are 'infinitely' less informative than any double bound, no matter how vague, such as (11c).

(11a)   There were more than 10 people present.

(11b)   There were more than 20 people present.

(11c)   There were between 5 and 100 people present.

The intuition we wish to preserve here is that, although (11a) and (11b) are both compatible with infinitely many possibilities (at least at a semantic level), it remains the case that (11b) is more informative than (11a), in that it admits a proper subset of the possibilities admitted by (11a). A convenient way of doing this is to part with the notion of 'absolute' informativeness and instead adopt a notion of 'relative' informativeness, measuring alternative utterances against a notional optimum level. In technical terms, this means that we construe the informativeness constraint as a faithfulness constraint, relating the speaker's intention to their utterance. We can formulate this constraint as follows.

Constraint #1: *Informativeness*. The utterance must convey the strongest numerical information available to the speaker about the topic. Incur a violation for every possibility admitted by the utterance that is known to be impossible by the speaker.

Effectively this corresponds to an appeal to the notion of logical entailment. If an expression P entails an expression Q, P incurs a subset of Q's violations of informativeness. To take a concrete example: suppose that we know the answer to a question is '8'. Saying '8' incurs no violation of informativeness. '8 or 9' would incur one violation, 'between 7 and 9' would incur two violations, 'about 8' would incur perhaps two violations, and so on. However, if our best knowledge about the question extends to 'more than 7', then 'more than 6' incurs one violation, 'more than 5' incurs two violations, and so on.

In principle, this reformulation still leaves a step-change between the single-bounded and double-bounded cases. If the speaker's knowledge extends to 'between n and m', then any single-bounded output might incur infinitely many violations. We can obviate this by being more precise about the definition of 'topic', as referred to in the constraint. We can think of this in terms similar to the Question Under Discussion (QUD) of Roberts (1996). However,

care is needed to avoid circularity in appealing to the notion of QUD within this model[8]. Rather, we can suppose that the context specifies whether the lower bound, upper bound or both are under discussion, and that the violations of the informativeness constraint are assessed with respect to the bound(s) under discussion.

In practical terms, these technical issues are not generally crucial to the working of the constraint. The intuition underlying it is straightforward enough: wantonly underinformative statements are disfavoured for that reason. We understand that 'more than 10' is less informative than 'more than 20', and therefore do not expect to hear the former statement from speakers who know the latter to be true, unless there is a good reason. Possible reasons will be explored in the following sections. The theoretical contortions of the above paragraphs are merely an attempt to formulate a constraint that achieves this effect and that can neatly be slotted into the proposed system. However, the resulting constraint admits subsequent development in order to make it more systematic and more explanatory.

2.4.1.1. Experimental support

There is ample evidence that underinformativeness is generally dispreferred in the literature on the production of referring expressions (see Engelhardt et al. 2006, Davies and Katsos 2010). For instance, not providing a uniquely identifying expression (e.g. referring to 'the ball' in a context where there are two balls) is largely avoided by speakers (and also penalised by hearers). A similar situation obtains in the case of overinformativeness: for instance, research by Davies and Katsos (2010) documents that speakers generally avoid and listeners penalise redundant use of adjectival modification (e.g. referring to 'the big ball' in a context where there is only one ball) unless there are other pragmatic reasons that may favour over-specification of the referent.

Direct evidence in support of the informativeness constraint for quantified expressions comes from the literature on scalar quantifiers such as 'all', 'most' and 'some'. It is widely agreed that 'some' possesses existential semantics, and hence that the use of a proposition with 'some' in a situation in which a proposition with 'all' could be used is a case of underinformativeness rather than semantic falsehood. Katsos and Smith (2010) report that

---

[8] For instance, Zondervan (2007) identifies the QUD as the question that an utterance is supposed to answer, but under such an approach the QUD is determined by the utterance. If we define 'topic' in terms of this construal of QUD, any utterance is automatically 'on topic', which renders this approach useless for the model under discussion here.

when native speakers of English spontaneously describe a situation in which – for example – all the boys are holding balloons, they do not use an underinformative statement such as 'some/many/most of the boys are holding a balloon'. Moreover, they report that adult hearers consistently reject or penalise underinformative descriptions relative to fully informative descriptions, although to a lesser extent than they reject or penalise false descriptions (Katsos and Bishop 2011; see Noveck and Reboul 2008, Katsos and Cummins 2010 for reviews), which suggests that they expect speakers to abide by considerations of informativeness. While there is no obvious evidence in the literature for informativeness as a factor in the use of specifically numerically quantified expressions, the existing data strongly support the expectation that this would be elicited by the appropriate experiment.

2.4.2. Granularity

Closely related to the notion of informativeness (and numeral salience, as discussed in section 2.4.4) is the notion of granularity – that is, the level of precision at which information is conveyed. This is connected to the way in which scales are partitioned up by expressions. Work in the Vagueness, Approximation and Granularity (VAAG) project, led by Manfred Krifka and Uli Sauerland, has focused on the considerations underlying the speaker's selection of an appropriate granularity level. Krifka (2009: 116) specifically discusses the idea that speakers have a bias towards simple representations, and augments this with the notion that coarse-grained representations are typically simpler.

This notion of granularity, or representational simplicity, is necessary in order to account for the validity of approximations that do not cohere with round numbers *per se*. For instance, '18 months' appears to refer to a wider range of ages than '20 months'; similarly, '24 hours' is more approximative than '25 hours'. This is because '18 months' and '24 hours' are especially significant values in their particular systems of measurement, because they correspond to values that can be expressed with a higher denomination of units (1½ years and 1 day respectively). To put it another way, they are values on a scale at a coarse level of granularity. '20 months' and '25 hours' are not significant values in their scales, and are therefore interpreted at a fine level of granularity. In order to interpret numerically quantified expressions of this type correctly, the hearer has to establish the appropriate level of granularity.

I assume that the level of granularity at which an utterance is interpreted is itself determined by the preceding context. From this perspective, the maintenance of granularity could be considered a form of accommodation in conversational scorekeeping in the sense of Lewis (1979: 346f). The speaker must choose an utterance which conveys the appropriate granularity effects in order to be correctly understood, as a mismatch in granularity levels will lead to misunderstanding. Consider, for example, the contrasting interpretations of '4 years old' in (12a) and (12b). In (12a) it appears to impose a maximum age of 4;0 years, whereas in (12b) it conveys a maximum of 4;11 or 5;0 years. If the speaker of (12a) intends to convey that the children's ages ranged up to 5.0 years but did not exceed this (the interpretation of (12b)), (s)he should use a more precise value.

(12a)   The children in the study were between 2 years 11 months and 4 years old.
(12b)   The nursery caters for children who are 3 to 4 years old.

Within this model, we can then construe granularity as a faithfulness constraint. If the context does not determine the level of granularity, either we could suppose that a default level of granularity applies, or we could suppose that any level of granularity is acceptable (i.e. that the constraint cannot be violated by any utterance). Failure to use the appropriate level of granularity (where this is specified) would constitute a violation of this constraint. We could quantify this more precisely with reference to the number of levels' distance between the appropriate level of granularity and that which is actually used: for instance, in the time domain, to give a response to the second when hours are called for might constitute a double violation of this constraint. We can therefore formulate this constraint as follows.

Constraint #2: *Granularity.* The utterance must use the appropriate level of granularity. Incur a violation for each level of mismatch between the granularity set by the context and that used in the utterance.

2.4.2.1. Experimental support

I consider that the evidence discussed by Van der Henst and Sperber (2004), and replicated by Gibbs and Bryant (2008), can be adduced in favour of this constraint. Recall that Van der Henst and colleagues conducted experiments in which they asked passers-by to tell them the time, both under conditions in which precision was required and in which no particular degree of precision was required. In the latter (control) case, subjects tended to round their answers to the nearest major unit. Subjects with both analogue and digital watches responded

in this way, although approximate responses were more frequent among the former group. However, when the question stipulates greater precision (by specifying that the questioner wants to set their watch, or that they are late for a meeting at a specified time), both sets of subjects were significantly more likely to give a precise answer.

This behaviour is consistent with the constraint being proposed here: participants respond at a default level of granularity unless there is a contextual cue for a finer-grained response. We could interpret this as the participants' responses being conditioned by their adherence to a granularity constraint.

It should be stressed that this explanation is not necessarily in opposition to that proposed by Van der Henst and Sperber (2004), despite first appearances. They construe the response pattern as reflecting the speaker's desire to provide information at a level of precision which is tailored to the specific needs of the hearer: that is, they analyse the interaction as reflecting hearer-oriented behaviour. By positing a constraint on granularity, we offer a mechanism by which this can be accomplished in terms of speaker-referring constraints. In this model, one interlocutor (tacitly or explicitly) introduces the requirement for a specific level of granularity into the discourse, and the other interlocutor reacts accordingly. Interestingly, it could be argued that this is potentially more economical, as it does not require explicit perspective-taking on the part of the speaker, but merely supposes that the speaker is adhering to established conversational strategies. However, the claim that perspective-taking is costly in this way remains to be proven.

2.4.3. Quantifier simplicity

Grice's maxim of manner enjoins interlocutors to avoid using marked or prolix expressions. We can attempt to capture this within our model by positing a constraint requiring quantifier simplicity. Quantifiers will violate this constraint if they exhibit prolixity – the use of additional words to express the same concept – or if they are marked for some other reason, for instance as a consequence of additional complexity at some level of representation, as discussed in chapter 4.

However, within the system we are attempting to build, such a constraint needs further development in order to be useful. Specifically, we need to consider how quantifiers compare in complexity. In the case of numerical quantification, there does not appear to be any agreed metric that can be used to measure complexity across the board, which excludes

the possibility of using this constraint to compare all possible utterances. However, we may be able to establish some partial orderings of numerical quantifiers by complexity, which would enable us to use the quantifier simplicity constraint to frame predictions about their comparative usage patterns. With that goal in mind, I consider four distinct classes of numerically quantified statement (precise, approximate, double-bounded and single-bounded), and consider how the various means of expressing these compare in complexity.

Within the 'precise' category, it seems intuitive that the bare numeral is simplest, on account of its brevity. Thus, we could propose that 'exactly n' incurs one violation of the simplicity constraint[9]. This coheres with the intuition that 'exactly' tends not to co-occur with unambiguous, non-round numerals in cardinal contexts.

For the 'approximate' category, we could similarly argue that modifiers of the type 'about', 'around', etc., each incur one violation, without attempting to distinguish between these forms at this point. We might also posit that more prolix forms such as 'give or take' or 'plus or minus m' incur additional violations of the simplicity constraint. However, such forms might be favoured on the grounds of informativeness (indeed, might be double-bounded).

For the 'double-bounded' category, the use of 'inclusive' as a modifier of 'between' might similarly incur an additional violation, but might be motivated by the wish to mention a particular numeral (see sections 2.4.2, 2.4.4 and 2.4.5).

For the 'single-bounded' category, things are potentially more interesting, because there are several productive ways to express this type of quantification. These include 'more/fewer/less than', 'at least/most', and 'no(t) more/fewer/less than'. If, despite their similarity in length, it were to transpire that these forms differed in complexity, then the quantifier simplicity constraint might be able to adjudicate between them. For convenience I shall assume for the time being that 'fewer' and 'less' are adequately distinguished by usage conventions, the former applying to discrete quantities and the latter to continuous quantities, although nothing hangs on this point. In chapter 4, I return to the topic of comparative and superlative quantifiers, and present the case for distinguishing between these on the grounds of complexity.

---

[9] The legitimacy of this argument may depend on the precise semantics of the numeral. However, if we impute a more complex semantics to the numeral, it is difficult to explain how this complexity is dispensed with by the use of a modifier. In the apparent absence of clear theoretical arguments to the contrary, it appears natural to assume that 'exactly n' will still be more complex than 'n' on account of its prolixity if nothing else.

Finally, what about negation? Given that it has to be explicitly expressed in English, there should be no controversy involved in labelling it as marked, at least in our terms, in accordance with a long tradition in linguistics (see Horn 1989 and references therein) and psycholinguistics (Just and Carpenter 1971, i.a.). The quantifier simplicity constraint seems to be the appropriate place to attempt to capture that markedness. For instance, 'not more than three' is (arguably) semantically equivalent to 'at most three', and (for cardinals) 'fewer than four'. It is reasonable to suppose that negation of this type involves a further violation of the simplicity constraint, by comparison with the non-negated form 'more than three'. Whether we consider this to be an issue of prolixity or of representational complexity does not assume particular importance within this model.

In brief, then, our simplicity constraint is a markedness constraint, which we can formulate as follows.

Constraint #3: *Quantifier simplicity*. The utterance must use the simplest quantifier possible. Incur a violation for each degree of complexity exhibited by the quantifier used.

2.4.3.1. Experimental support

Obtaining clear experimental support for this constraint is difficult without making stipulations as to the relative complexity of the quantifiers. However, we can nevertheless obtain evidence in its favour, as discussed in the following paragraphs.

First, we have already considered how round numbers can have approximate interpretations. Therefore, when an unmodified round number is used to convey an approximation, this is intrinsically vague, and could be disambiguated by the addition of an explicit quantifier. A willingness to use the plain numeral, despite its ambiguity, therefore constitutes evidence in favour of our proposed quantifier simplicity constraint. In passing, I should remark that this preference could clearly be attributed to general principles of economy of expression. As the quantifier simplicity constraint is also motivated by considerations of economy, I would argue that there is no important distinction here. A more general account of usage based on similar grounds might, of course, be able to subsume the quantifier simplicity constraints under the more general heading of an economy constraint.

Furthermore, under the assumption of Geurts et al. (2010) that superlative quantifiers are more complex than comparative quantifiers, the reaction-time preference for the latter

documented by Geurts et al.'s (2010) Experiment 3 also constitutes evidence in favour of this constraint. In this case, the issue is complicated somewhat by the change of numeral required to switch between comparative and superlative quantifiers: the question of preferences in the numerical domain will be discussed further in the following sections.

Under the assumption that negation incurs a violation of this constraint, we also obtain evidence in favour of this constraint from the literature on negation. For instance, we can examine whether speakers are reluctant to use explicit negation when a quantifier conveying the same meaning is available. Data on this point have been obtained by Gennari and McDonald (2006). In their first production study, participants were asked to describe situations in which 'some' or 'none' would be appropriate – for instance, in summarising a story in which ducks attempted to cross a river and either some or none of them succeeded. In the 'some' condition, there was a strong preference for positive statements – 61% of their subjects using 'some' and 15% simply replying 'the ducks crossed', while only 22% replied 'some…didn't' and 2% 'not all'. Even more strikingly, in the 'none' condition, 43% of their subjects produced what they termed 'implied negation', in which negation was not explicit (e.g. 'the ducks tried to cross'). Hence, in these apparently neutral contexts, participants seem to disprefer utterances with explicit negation. This also supports our claim that quantifier simplicity is a relevant constraint on the use of quantifying expressions.

2.4.4. Numeral salience

It is obvious that some numbers are used more frequently than others. What are popularly known as 'round numbers' are used especially often and seem to possess a kind of enhanced salience. Dehaene (1997) argues that round numbers are associated with particular levels of a numerical accumulator mechanism, which provides a system for representing analogue quantities, whereas non-round numbers do not correspond to levels in that system.

There are several arguments in favour of the use of round numbers. If Dehaene's claim about their relation to the accumulator system is correct, round numbers immediately convey a sense of magnitude that prefigures numerical calculation. Also, round numbers are expressed in few words (given their magnitude), which makes them especially efficient in oral communication. This is reflected in the fact that (spoken) communication systems for numerals are typically organised around round numbers: many languages, including English, have systems in which all numbers are expressed as concatenations of round numbers. Given

the relevance of finger-counting to the early acquisition of number concepts, it is easy to see how numbers such as 5 and 10 might acquire particular psychological prominence (following Dehaene 1997, Butterworth 1999, i.a.).

For these reasons, I posit a constraint to 'use a round number'. We must adopt some measure of 'roundness' in order to define and quantify violations of this constraint. This cannot depend purely on the simple divisibility properties of the number, as larger numbers can exhibit these divisibility properties without being round. For example, 3180 is a multiple of 20, but is intuitively less round than 20. Therefore, we need a characterisation of roundness that depends also on the magnitude of the number. One such measure arises from Jansen and Pollmann's (2001) work, where they define the notion of 'k-ness', a concept which can be expressed in the formalism (13).

(13) N has k-ness iff $N = 10^b mk$ for $m \in \{1, 2, \ldots, 9\}$, k and b non-negative integers.

To illustrate this, 5, 10, 15, …., 45 possess '5-ness', as do 50, 100, …, 450 and 500, 1000, …, 4500. However, 55, 60, 65, …, 95 do not; neither do 550, 600, …, 950, nor 5500, 6000, …, 9500. '2½-ness' can be defined analogously: 5, 10, 15, 20 and 25 possess this, as do 50, 75, 100, …, 250 and so on. Loosely speaking, k-ness characterises the divisibility properties of numbers in terms of the divisibility properties that numbers of these magnitudes could have, with particular reference to their divisibility by powers of 10.

Jansen and Pollmann (2001) posit that the forms of k-ness that significantly contribute to numeral roundness are those which relate to the base (10) and to notions of doubling and (particularly) halving, which they argue are especially important notions in quantity processing. They demonstrate that 2-ness, 5-ness, 10-ness and 2½-ness are significant predictors of numeral frequency[10]. If we take these as the constituents of roundness, then we can define an 'entirely round' number as one that exhibits 2-ness, 5-ness, 10-ness and 2½-ness, such as 20 or 100. Our constraint is then violated once for each crucial type of k-ness that the number used does not possess. For instance, 40 incurs one violation (for its lack of 2½-ness), 30 two violations (lacking 2½-ness and 2-ness), 12 three violations (5-ness, 10-ness and 2½-ness), and so on.

---

[10] Under the definition of k-ness, non-integers such as 2.5, 7.5, 12.5 etc. also possess 2½-ness. However, this does not matter, as we are currently only interested in the roundness properties of integers. Note that, for each value of b, ten values possess k-ness for any k: for $b=1$ and $k=2.5$, only five of these values are integers. For $b > 1$ all the values possessing 2½-ness are integers, as they are by definition multiples of 25.

This account makes several simplifying assumptions. First it assumes that each type of *k*-ness discussed by Jansen and Pollmann (2001), and only those types of *k*-ness, are relevant considerations for this roundness constraint. It further assumes that each of these types of *k*-ness is equally important. Moreover, it assumes that *k*-ness is the correct formalism, which is almost certainly untrue: there does not appear to be a psychological motivation for the existence of clear cut-off points (45 fully possesses 5-ness, for instance, and 55 fully lacks it) or for where these points are located. We could easily relax these assumptions by changing the definition of the constraint in various ways: assigning violations for the lack of further types of *k*-ness, and assigning multiple violations to more flagrant breaches of roundness. I do not intend to address this question here, not least because refinements of this type go beyond the present needs of the model. However, in principle, the question of which kinds of roundness give rise to violations is an empirical matter.

To summarise, our numeral salience constraint is a markedness constraint, which can be formulated as follows.

Constraint #4: *Numeral salience.* The numeral used in the utterance must be intrinsically salient. Incur a violation for each type of roundness that the numeral fails to exhibit.

2.4.4.1. Experimental support

The preference for round numbers has been discussed by a number of authors (notably including Dehaene 1997). The consensus view appears to be that trends in the frequency of number use in corpora reflect speakers' preferences about number use *per se*, rather than reflecting the frequency of sets of that cardinality in the world. This claim is not entirely obvious upon reflection, and merits brief discussion.

Firstly, this supposes that we are comparing numbers of similar magnitude: specific large values occur less frequently than specific small values in cardinal contexts, which is intuitively obvious. Jansen and Pollmann (2001) back this up by demonstrating that the values n^{-1} and n^{-2} are significantly predictive of numeral frequency: i.e. that magnitude (and its square) are inversely correlated with frequency. This accounts for the underlying trend in numeral frequency, while *k*-ness is used to account for the peaks of frequency.

Secondly, this supposes that the preference for round numbers is not attributable to the existence of more sets of objects of round cardinalities than non-round cardinalities in the

speaker's world. Many cultural artefacts are constructed with reference to round numbers: packages of 10 or 20 items, metric measurements, percentages, and so on. Although this phenomenon may reflect the relevance of numeral salience to human cognition, it is a confounding factor when we consider the behaviour of the individual speaker. Suppose that an individual has no preference for round numbers, and participates in this culture. We would expect a corpus of their speech to reflect a bias towards round numbers: they would still have to ask for 20 cigarettes, learn that 1000 grams equals one kilogram, and be asked for 100% commitment, etc. Our question, then, is whether the bias towards round numbers is greater than can be accounted for by cultural factors alone. For the moment I will lay this question aside and simply appeal to corpus data to demonstrate the importance of roundness. In chapter 6, I will return to this topic and attempt to obtain more satisfactory evidence, taking the above discussion into account.

Jansen and Pollmann (2001) fitted a regression line to corpus data of number usage, and demonstrated that 2-ness, 5-ness, 10-ness and 2½-ness each had significant predictive utility. As remarked above, they also demonstrated significant effects of n^{-1} and n^{-2}. They showed that no other form of k-ness (for integer k) had predictive power in this model. The general trend in the usage data they discussed was one of gradually decreasing frequency as the numerals increase in magnitude, interrupted by sizeable peaks at particular values. Jansen and Pollmann effectively demonstrated that the forms of k-ness they discuss predict the location of these peaks.

In our terms, this finding provides empirical support for the numeral salience constraint we propose above. As we have construed numeral salience purely in terms of k-ness, Jansen and Pollmann's evidence for the predictive utility of k-ness applies to our constraint. The decision to label this as numeral salience arises from the shared intuition that the purpose of k-ness is to index this more fundamental property which underlies the usage preferences.

In addition to any cognitive preference for the use of round numbers, these numbers can also convey approximate meanings (Krifka 2009) which renders them semantically available for use in a wider range of situations. It is arguable whether this constitutes evidence for this constraint within a model, or merely bears upon the question of when this constraint is relevant.

2.4.5. Numeral and quantifier priming

Aside from its intrinsic salience, another reason why a numeral might conceivably be preferred in a given discourse context is because it is already activated in the preceding context. Similarly, quantifiers might be preferred on the grounds of prior activation rather than just because of their simplicity. Artificial examples of this kind of effects are (14) for numerals and (15) for quantifiers.

(14)   A. Will there be 20 people at the meeting?
        B. Yes, at least 20.

(15)   A. If we sell more than 20 tickets, it'll be a success.
        B. I think we'll sell more than 20.

In (14), A asks a question concerned with a particular numerosity, and B answers with reference to that same numerosity. This seems like a plausible response even if B is in a position to make a more precise or informative statement – for example, if B knows that 25 people will be there – on the intuitive basis that this response is addressed directly to A's precise concern. Similarly, in (15), B's utterance seems plausible even if B thinks they will sell more than 25 tickets.

There are several reasons why this kind of priming might occur. First, a recently mentioned numeral or quantifier is likely to be highly activated in the minds of the discourse participants, making it particularly accessible for both the speaker and the hearer, which is communicatively advantageous. Furthermore, a numeral or quantifier used by one discourse participant is likely to cohere closely with notions such as the question under discussion, and therefore is likely to be used again if the question under discussion remains the same. It is even conceivable that the speaker might reuse structures from the previous discourse in constructing the syntactic frames used in their productions, in which case we could think of quantifier priming as a species of *alignment in discourse* (see Pickering and Garrod 2004 for a discussion of interlocutor alignment effects). However, as far as this constraint is concerned, the first of these points suffices: I assume that recently mentioned numbers and quantifiers are highly activated and hence accessible at a semantic level. I follow Pickering and Garrod (2004) in characterising this process as priming, although I do not commit to the view that the motivation for this behaviour is necessarily wholly explicable in terms of low-level automatic processing.

In principle, the constraints governing numeral and quantifier priming could be distinct and ranked separately. However, for ease of exposition, I will discuss both in parallel in the following paragraphs concerning the evaluation of constraint violations, as this will be essentially identical for both constraints.

First, let us suppose that a violation is incurred by the failure to use a quantifier/number that is primed. We will further suppose that a quantifier/number is primed if it occurred in the preceding conversational turn. However, the constraints should allow for numbers or quantifiers not specifically mentioned also to be primed, as in (16). Here we assume that both speakers share the relevant world-knowledge about the number of people required to play bridge.

(16)   A. Will we able to play bridge this evening?
         B. Yes, there will be at least four of us there.

If no quantifier/numeral is primed in a relevant way, then this constraint imposes no requirements on the output: that is, it will not be violated by any candidate output, and therefore will not discriminate between them.

In summary, our numeral/quantifier priming constraints are faithfulness constraints, which can be formulated as follows.

Constraint #5: *Numeral priming*. If a numeral is primed in the preceding context, it must be used in the utterance. Incur a violation if there is a numeral primed in the preceding context but a different numeral is used in the utterance.

Constraint #6: *Quantifier priming*. If a quantifier is primed in the preceding context, it must be used in the utterance. Incur a violation if there is a quantifier primed in the preceding context but a different quantifier is used in the utterance.

I consider numeral and quantifier priming separately, on the grounds that these constraints could conflict with each other. For example, suppose an individual hears (17) but knows that this is underinformative and a stronger statement could be made. The response (18a) would obey quantifier priming but violate numeral priming, whereas (18b) would obey numeral priming but violate quantifier priming.

(17)   There will be at least three of us at dinner.

(18a) No, there will be at least four.

(18b) No, there will be more than three.

The numeral priming constraint becomes particularly relevant within this model if we consider numerals to possess a lower-bounded rather than a precise semantics. Under these conditions, the speaker is not constrained to use a specific unmodified numeral by considerations of truthfulness – if 'ten' would be true, so would 'nine' be, etc. However, the precise cardinality may be more salient than the other competitors, for instance because it equals the number of objects present in the environment. In such a case, numeral priming might mandate the use of the precise value even if this is not semantically obligatory.

Nevertheless, in postulating two constraints it could be argued that this approach multiplies entities beyond necessity. Given the examples above, it might also appear that this machinery is being invoked merely to mop up a few marginal cases of usage, and that these could be treated with reference to alternative approaches, for instance by direct appeal to question under discussion. Within this model it is difficult to argue for the use of specific constraints rather than functionally similar alternatives, so this last objection must be conceded. However, in the following subsection I present some evidence in support of the validity and necessity of priming constraints, and foreshadow a more direct demonstration of their applicability in chapter 5 of this thesis.

2.4.5.1. Experimental support

It is comparatively difficult to look for evidence of numeral and quantifier priming in corpora. In the first place, these types of priming rely on the preceding context, potentially over several conversational turns, which may not be available in corpora. In the second place, they may also rely upon encyclopaedic knowledge, as in (16); in this case, the question arises of precisely what knowledge the participants possess.

However, relevant empirical data can be found in cases where a preceding context has been supplied against which an utterance is selected or interpreted. One such case is presented in chapter 5 (Cummins, Sauerland and Solt submitted), where participants are asked to give a preferred interpretation of numerical quantifiers in contexts where the number has or has not already been mentioned ('primed' and 'unprimed' conditions), such as (19) and (20) respectively.

(19)　　A. This case holds 60 CDs.

　　　　B. I own more than 60 CDs.

(20)　　A. This case holds CDs.

　　　　B. I own more than 60 CDs.

Participants consistently gave a wider range of possible values in cases like (19) than in control conditions, (20), in which the number was not previously mentioned. As I discuss in chapter 5, this implies that (in the hearer's opinion) there are situations in which a primed numeral can be used but in which an unprimed numeral cannot. That is, reuse of a previously-mentioned number can license the use of a particular quantified expression. This supports the presence of the numerical priming constraint within our system.

The relevance of the quantifier priming constraint is documented experimentally in the following experiment.

2.4.5.2. Experiment 1 – Quantifier priming in the Cavegirl experiment

A laptop-based task was administered to assess the comprehension and production of quantified expressions. In this task the experimenter introduces participants to a fictional cartoon character and explains that they should help the character learn to speak their language better. In this particular game, the cartoon character is a female, the Cavegirl, and she is asked 'to say how many boxes have a toy'. If what she says is right, the participant should tell her 'that is right'. If what she says is wrong, the participant should tell her 'that is wrong', and also tell her why it was wrong, in order to help her learn.

In each trial of the experiment, an array of five boxes appears on the screen, along with the corresponding number of instances of an object. The objects are those that young children are familiar with, such as dolls, balls, cars etc. (for a description of the criteria employed for selecting objects see Katsos et al. 2011). Each instance of an object may be inside or outside the corresponding box. An audio recording is played, representing the Cavegirl's description of the situation. The participant gives a verbal response, which is recorded by the experimenter.

The version of the experiment administered to adults involves the expressions 'all', 'all…not', 'none', 'some', 'some…not', 'not all' and 'most'. For 'all' and 'none', there are true and false conditions. For 'all…not', 'some', 'some…not', 'not all' and 'most', there are

true, false and under-informative conditions. The under-informative condition is one for which the expression is logically true, but there exists an alternative expression that would have been more informative. For instance, in the case of 'some', the utterance is 'Some of the boxes have a toy'. The condition in which none of the boxes have a toy is false for this description; that in which exactly two of the boxes have a toy is true; that in which all of the boxes have a toy is under-informative.

For the purposes at hand, we are concerned with the production component of this experiment. According to our quantifier priming hypothesis, we expect participants' productions to exhibit quantifier priming effects. Sentences of the form 'Q of the boxes [do not] have an <object>' are predicted to occur more frequently after sentences involving the specific quantifier Q.

METHOD

Participants

20 adult participants were recruited to take part in this and other unrelated experiments. Ages ranged from 16-45. 16 were female.

Materials and procedure

The full set of conditions and the semantically correct responses are specified in the results section. The experimental task was introduced to participants as described above. Sample materials are presented in Appendix A.

Results

Acceptance rates for the test items were as shown in Table 1.

Table 1: Acceptance rates for test items in Cavegirl experiment

Quantifier	Condition	Acceptance rates (%)
All	2/5 items	3.33
	5/5 items	96.7
All…not	0/5 items	82.2
	2/5 items	22.0

Quantifier	Condition	Acceptance rates (%)
	5/5 items	6.25
Most	2/5 items	7.0
	4/5 items	100
	5/5 items	10.0
None	0/5 items	98.3
	2/5 items	0.0
Not all	0/5 items	19.2
	2/5 items	98.3
	5/5 items	1.69
Some	0/5 items	5.00
	2/5 items	98.3
	5/5 items	16.0
Some…not	0/5 items	18.3
	2/5 items	88.3
	5/5 items	1.69

For the items that were consistently corrected, the corrections issued were as shown in Table 2.

Table 2: Corrections to rejected utterances; frequency quoted as percentage of total responses

Quantifier	Condition	Corrections (frequency)
All	2/5 items	some (52%); two (15%); not all (12%); three out (5%); some out (3%); most out (2%); other (7%)
All…not	0/5 items	none (17%); other (1%)
	2/5 items	some (42%); two (23%); some…not (5%); some out (1%); most (1%); not most (1%)
	5/5 items	all (93%); other (1%)
Most	2/5 items	some (37%); most…not (23%); most…out (22%); two (10%); other (2%)

Quantifier	Condition	Corrections (frequency)
	5/5 items	all (90%)
None	2/5 items	some (61%); two (39%)
Not all	0/5 items	none (73%); all…not (5%); all out (3%)
	2/5 items	some (2%)
	5/5 items	all (98%)
Some	0/5 items	none (90%); all out (5%)
	2/5 items	other (2%)
	5/5 items	all (84%)
Some…not	No items	all…not (39%); none (38%); all out (4%); not all (1%)
	2 items	most…not (12%)
	All items	all (97%); none…not (2%)

Discussion

It was posited that quantifier priming would exert influence on the choice of corrections issued. To support this claim, we must show that participants' choices of correction are systematically influenced by the expression being corrected. In particular, if this relationship is mediated by quantifier priming, we would expect to see that the relation between expression used by the Cavegirl and correction produced by the participant is transparent, and that quantifiers used by the Cavegirl are disproportionately reused by the participant. The data from this experiment broadly meet both conditions. In particular, the data for 'most' show direct influence from the quantifier on the choice of correction. 'Most' is produced (negated) in 45% of adult corrections of 'most' in the two-item case, while occurring in less than 5% of two-item cases of other quantifiers. The corrections for 'some…not' also clearly exhibit a quantifier priming effect in that they disproportionately elicit responses with post-verbal negation. In the no-item cases, 'all…not' is the preferred correction to 'some…not' by adults whereas 'none' is strongly preferred as a correction to the other semantically false prompts.

We interpret these findings as evidence that quantifier priming is exhibited by participants in this experiment. The choice of correction is conditioned by the utterance being corrected, and that this conditioning is – at least in part – transparent. Hence these results support the

decision to posit a specific quantifier priming constraint. However, this is arguably unsurprising given that this experimental methodology induces the use of a quantifier as an explicit correction to a previously-uttered quantifier, which is therefore especially salient in the context. It could be argued that this level of salience would not typically be achieved by a quantifier in a naturalistic setting. Directly demonstrating the relevance of this constraint in a naturalistic setting, however, is a particularly difficult task given the need to control for all other relevant factors.

2.4.6. Interim summary

So far, I have laid out a set of constraints which codify observations about the use of numerically-quantified expressions, and supported each one by appeal to the experimental literature or novel empirical data. Before considering how these constraints would interact within the type of overarching system being proposed here, I will consider some additional constraints for which empirical support cannot readily be adduced.

2.5. Additional constraints

2.5.1. Truthfulness

Given the importance of cooperative behaviour in social contexts (Grice 1975), we might expect truthfulness to be the most important criterion governing the use of numerical quantifiers. In the examples given earlier, this is assumed. It is therefore tempting to posit a constraint stating that the statement uttered must be true: that is, one for which false statements incur a violation.

It is clear that speakers frequently make intentionally false statements. However, even in these cases, the speaker is choosing among a wide selection of possible utterances in order to achieve a certain communicative effect. Therefore, lying ought to fall within the explanatory range of this theory. The same applies to cases in which the speaker is misinformed. Hence we would need to formulate this constraint in terms of faithfulness to the speaker's intention, rather than in terms of the relationship between the statement and observable reality. On this view, the truthfulness constraint could be defined as follows: incur a violation if the truth-conditional content of the utterance fails to match the truth-conditional content intended by the speaker.

However, this constraint appears so universal that it may be a poor candidate for an OT system. This argument follows Grice (1975: 46), who discussed whether his first quality sub-maxim 'Do not say what you believe to be false' was appropriate to his system: 'it might be felt that [its importance] is such that it should not be included in a scheme of the kind I am constructing; other maxims come into operation only on the assumption that this maxim of Quality is satisfied.' Similarly, from the point of view of this model, no purpose is served in including a constraint that cannot be violated: it makes more sense to consider truthfulness as a requirement that serves to filter possible utterances before speakers evaluate them in detail for adherence to the extant constraints.

So, is truthfulness – defined with reference to the speaker's intention – ever violable? Possibly: but examples appear marginal. One possible case would be the use of round numbers as tacit approximations, as in (9a) and (9b). However, it appears more realistic to argue that this usage is intrinsic to the meaning of round numbers, and therefore merely invokes a particular aspect of meaning in a truthful way. Perhaps a more plausible example could run along the lines of (21).

(21)   A. Unless 20 people turn up, my party will be a failure.
        B. There'll be 20 people, no worries.

Here, B's response might be considered as an expression of confidence that A's party will not be a failure, even if B is not confident that A's required number of people will be present. If this is the case, then the truthfulness of B's choice of quantifier has been subjugated to the communicative needs of the situation.

Another possibility is an interaction such as (22), where B's response is false but arguably conveys a more accurate notion of the size of the city than a simple 'yes' answer[11].

(22)   A. Does Glasgow have a population over 100,000?
        B. Over a million.

In sum, I argue that a constraint governing truthfulness could conceivably be part of the system we are building, although the motivation for it does not appear compelling. Such a constraint is likely to be highly-ranked and is seldom violated by optimal outputs. It is a

---

[11] This could be treated as a case of hyperbole, but I do not consider this constraint particularly well-suited for treating instances of non-literal usage in general.

faithfulness constraint, constraining the relation between the situation (specifically the speaker's intention) and the linguistic form uttered. For convenience I assume that it is violated once by unfaithfulness to the speaker's intention, although it would obviously be possible in principle to calibrate the extent of violation more precisely.

There is, of course, little direct evidence available for the operation of this constraint in production data, as such evidence would rely upon knowledge of the speaker's intention. However, I do not consider it a controversial proposition that hearers understand statements to be representative of their speakers' intentions. Therefore the only question to be resolved is whether this is best construed as a violable constraint, or whether we should regard it as an overarching principle.

2.5.2. Communicative intention of the speaker

In discussing the reuse of primed numerals and quantifiers (section 2.4.5), I made no reference to the communicative intention of the speaker, except insofar as it was determined by the preceding context. This is clearly not the whole story, however: it is intuitively obvious that the speaker is at liberty to use a number or quantifier if they wish to do so, regardless of whether or not it appears in the preceding context.

In order to encompass this, we need to expand the system outlined so far in some way. One approach would be to change the definition of 'primed' numerals and quantifiers to include those which are activated in the mind of the speaker as well as those which are present in the preceding discourse. Tacitly this was already done in section 2.4.5, where for instance it was argued that this activation could be modulated by the speaker's encyclopaedic knowledge, as in (16), where mention of a bridge game makes the number 4 salient.

An alternative line of attack would be to add a further constraint or constraints to the system, requiring the speaker to use a number or quantifier that they wish to make salient in the discourse, and incurring a violation if they fail to make such an entity salient. Again, these would be faithfulness constraints, as they would govern the relation between the speaker's intention and the output.

The objection to any such approach is that it could render the whole system vacuous. If a particular instance of usage is not accounted for by any combination of the other constraints, could we not simply say that it corresponds to the speaker's own intention? This would be

unfalsifiable, at least at our current level of understanding. Therefore, if we wish to extend our system in this way, we must lay down clear ground rules about when such an intention may be imputed to the speaker. Specifically, we would have to require that the particular number or quantifier, rather than merely the semantic content of the utterance, is what the speaker wishes to make salient to other conversational participants.

The most obvious instances of this subsystem in action would involve cases in which speakers initiate a discourse by introducing a numeral and/or quantifier that they wish to make particularly salient, as in examples (23a)-(23c).

(23a)   Did you know, there are 2.38 million unemployed now?

(23b)   The US national debt is more than 12 trillion dollars.

(23c)   There are 342 bricks in that wall.

However, it could be argued that, in all these cases, the speaker is merely introducing into the explicitly spoken discourse a numerical concept which is salient in the broader context – either they have read or heard the number, or it is present in the environment in some way. If so, these are not *bona fide* examples of the speaker spontaneously introducing a numerical quantifier, but can be accounted for in terms of faithfulness to a previously activated concept. Viewed this way, the spontaneous emergence of a numerical quantifier in discourse is an extremely rare thing, perhaps restricted to the cases in which an author makes up a sentence such as (23c).

For this reason, I will not attempt to pursue the idea that the communicative relevance of the numeral or quantifier, from this perspective, is governed by the operation of separate constraints. Rather, I will allow a broad definition of 'preceding context' to apply to the previously-discussed constraints favouring the use of contextually-primed numerals and quantifiers.

2.6. Summary

In this chapter, I have outlined a constraint-based model designed to account for the use of numerically-quantified expressions, and discussed the constraints with which to populate this model. The constraints discussed in section 2.4 (informativeness, granularity, quantifier simplicity, numeral salience, quantifier and numeral priming) are supported by empirical considerations, either from the existing literature or from original work. Those mooted in

section 2.5 (truthfulness and speaker's intention) do not admit similar empirical support and are thus not considered to be central to this model. In the following chapter, I move on to consider how the model, populated by the empirically-supported constraints, can be used to generate testable predictions about numerical quantifier usage.

3. DERIVING PREDICTIONS FROM THE CONSTRAINT-BASED ACCOUNT

In the previous chapter I considered up to nine possible constraints. Of these, we are particularly interested in the six which can be argued to be both preferred and non-obligatory: informativeness, quantifier simplicity, numeral salience, granularity, numeral priming and quantifier priming. These could be considered as independent factors that each go to determine what constitutes an optimal utterance. However, in this thesis I explore the possibility that the interaction of these constraints can be formalised in such a way as to yield testable predictions about the usage and interpretation of numerically-quantified expressions.

In this chapter, I discuss how these constraints influence speakers' quantifier selection within an Optimality Theory framework. First I consider how classical OT predicts that these constraints will interact, and consider its implications for the characterisation of individual preferences. Then I look briefly at other types of OT, with particular reference to forms such as bidirectional OT that are current in the pragmatics literature, and consider how the predictions arising from these frameworks would differ from those arising from classical OT. I also touch upon the question of whether these alternatives might be preferable to classical OT on the basis of having a superior claim to psychological plausibility. Finally, operating predominantly within a classical OT framework, I sketch some particular predictions arising from this model.

3.1. Constraint interaction in classical OT

In a classical OT system, the constraints are arranged into a hierarchy of strict domination – a complete ordering – by each speaker. This constraint hierarchy would then deterministically predict the output generated by the speaker given any input, which in this case is the context. That is, although the output of the model has the superficial appearance of being probabilistic, in that truth-conditionally equivalent forms may both surface in the output of a given individual, it is in fact deterministic, because the preferred form is actually always selected. The appearance of probabilistic variation in this model's output arises merely because the competing forms are each optimal in certain contexts.

To illustrate this point, we can consider a simple system containing only three of our constraints, informativeness (INFO), numeral salience (NSAL) and numeral priming (NPRI). First I apply this to an artificial situation in which the speaker wishes to utter a numerically-

quantified expression to describe a value greater than or equal to 22. Possible options include (24) and (25): for ease of exposition I will overlook other possibilities in this example.

(24)   More than 21

(25)   More than 20

Of these, (24) incurs a violation of NSAL because it uses a non-optimally salient numeral. (25) incurs a violation of INFO because it is not maximally informative (as it does not exclude the possibility of 21). In this context, neither violates NPRI as there is no contextually activated numeral. The OT tableau for these possibilities is as shown in Table 3.

Table 3: OT tableau for 'more than 21' vs. 'more than 20', INFO, NSAL & NPRI

	INFO	NSAL	NPRI
More than 21		*	
More than 20	*		

Under the assumptions of classical OT, speakers may differ in their constraint ranking, and each constraint ranking corresponds to a distinct idiolect. In the current example, the precise constraint ranking of the speaker determines which output is to be preferred. For a speaker who ranks INFO higher than NSAL (which I shall write as INFO > NSAL), 'more than 21' is the preferred option, because (unlike 'more than 20') it does not violate INFO. For a speaker who ranks NSAL higher than INFO, 'more than 20' is the preferred option, because (unlike 'more than 21') it does not violate NSAL. The ranking of NPRI is irrelevant to this selection process as it is not violated by either of the candidates we are considering. In each case, classical OT predicts that the speaker will choose consistently – that is, every time a given speaker encounters this situation, that speaker will select the same output, in accordance with their constraint ranking.

By contrast, if we apply this toy model to a situation in which the speaker wishes to express the same value, but in which the number 21 is already contextually primed, the resulting tableau is as shown in Table 4.

Table 4: OT tableau for 'more than 21' vs. 'more than 20', INFO, NSAL & NPRI; 21 primed

	INFO	NSAL	NPRI
More than 21		*	
More than 20	*		*

In this case, a speaker who ranks NSAL above both INFO and NPRI will prefer 'more than 20', but a speaker who does not will prefer 'more than 21'. There are six possible rankings for the three constraints, and the preferences of speakers with each ranking are presented, for both the unprimed and the primed condition, in Table 5.

Table 5: Preferred output for possible constraint rankings in toy INFO/NSAL/NPRI example

Constraint ranking	Unprimed preference	Primed preference
INFO > NSAL > NPRI	More than 21	More than 21
INFO > NPRI > NSAL	More than 21	More than 21
NSAL > INFO > NPRI	More than 20	More than 20
NSAL > NPRI > INFO	More than 20	More than 20
NPRI > INFO > NSAL	More than 21	More than 21
NPRI > NSAL > INFO	More than 20	More than 21

Hence, across the range of possible constraint rankings, there is considerable variability in the output we would expect from individuals given these inputs. Moreover, in the case of a speaker with the (presumptively fixed) ranking NPRI > NSAL > INFO, that speaker's behaviour differs between the unprimed and primed conditions – they prefer 'more than 20' in the unprimed condition but 'more than 21' in the primed condition. If we were to count output tokens from such a speaker in a way that did not respect context, e.g. through a straightforward corpus search, we would find this speaker exhibiting what superficially appeared to be probabilistic behaviour. Considering a system with as many as six constraints, this type of intra-speaker variability will be the rule rather than the exception.

A pertinent question arising from this is whether it is reasonable to expect such wide variability between speakers within a single speech community. Within domains such as phonology, two individuals must have approximately the same constraint ranking in order to be mutually intelligible. However, with respect to pragmatic concerns such as the use of

quantity expressions, there is more scope for individual variation without mutual comprehensibility being jeopardised. Studies in experimental pragmatics typically document behaviours that are non-categorical across a range of participants. For instance, in Van der Henst, Carles and Sperber's (2002) time-reporting study, 57% of the wearers of digital watches gave rounded answers, 43% did not. Similarly, Branigan, Pickering and Cleland (2000) demonstrated differences between speakers as to whether they exhibited syntactic priming effects. In both cases, the input to the decision-making process is the same for different speakers, but the preferred outputs differ. These findings are compatible with a view in which individual differences in constraint rankings underlie individual variation in output.

Even if we allow a broad range of individual differences in constraint rankings, the OT framework still predicts that certain similarities should be apparent in the behaviour of different speakers, at least on the assumption that all individuals have access to the same constraint set[12]. Specifically, OT predicts the emergence of the unmarked (McCarthy and Prince 1994): there will be a general preference for forms which do not violate markedness constraints. For example, our system contains a constraint requiring the use of salient numerals. Given this constraint, we expect salient numerals to surface more often than non-salient numerals *in general*. The argument can be sketched as follows. There is a constraint specifically favouring salient numerals, but no constraint specifically favouring non-salient numerals. For speakers who rank this constraint highly, the effect will be pronounced, as the constraint will presumably often be decisive in selecting the optimal output (in which case a non-salient numeral is rejected and a salient numeral selected). For speakers who rank the constraint lowly, it will take effect only very infrequently, but when it is decisive the effect will be the same (the rejection of a non-salient numeral and the selection of a salient numeral). Hence, across speakers, this constraint biases for the selection of salient numerals. Similarly, all markedness constraints in an OT system create a preference for the use of unmarked forms, even if they are lowly ranked.

Assuming, then, that individual speakers have free rein in how they rank their constraints, it will be difficult to make precise predictions about usage. A set of six constraints admits 720 possible orderings: for nine constraints, this would rise to 362,400 possible orderings. As

---

[12] This assumption is reasonable given that the constraints are underpinned by general, and putatively universal, considerations of cooperative behaviour. I do not wish to commit to the stronger assumption that the constraints are innately given.

even low-ranked constraints can influence the choice of output, it is not a workable proposition to evaluate all the constraints at once. Therefore, following previous applications of OT to the domains of semantics and pragmatics (such as Hendriks and de Hoop 2001 and Krifka 2009), I propose instead to address particular questions within the domain of quantifier usage, considering in particular how the relative ranking of some of the constraints I propose is predicted to influence behaviour in certain contexts. At the same time, given that the constraint rankings are not known *a priori*, we can also attempt to explore whether the observed behaviour of individuals can be accounted for by a constraint-based system of this type.

From both of these perspectives, it may be useful to consider alternatives to the strictures of classical OT. Relaxing the stipulation of strict domination and full constraint ranking might enable us to frame more general predictions, and will certainly provide additional explanatory power, if it should transpire that a classical OT account is insufficiently flexible to accommodate these data. More broadly, it could be argued that certain non-classical OT accounts have the intrinsic advantages of being more psychologically plausible than classical OT. Meanwhile, bidirectional OT has been used to account for aspects of pragmatic meaning in the established literature. Thus, before detailing specific examples of the predictions from constraint interactions, and before considering detailed case studies of quantifier usage from the perspective of this model (chapters 4 and 5), I will pause to review some alternative constraint-based formalisms.

3.2. Alternative formalisms

3.2.1. Stochastic OT

Given that classical OT is strictly deterministic, it is tempting to look at alternative approaches that allow more variation. For instance, it seems to be a reasonable intuition that our adherence to certain constraints may not be robust. Perhaps, given ideal conditions, we are able to formulate the optimal utterance as determined by a classical OT system, but when conditions are non-ideal we formulate a potentially sub-optimal utterance. We could construe this in terms of constraints themselves differing in availability throughout the process of planning an utterance. For instance, the salience of numerals could conceivably be relevant at an early stage of utterance generation, on the grounds that more salient numerals are presumably made more available for use in utterances, while a constraint such as

informativeness might only take effect when resources are available for comparatively high-level planning.

One development which coheres with this notion is stochastic OT (Boersma 1997). In a stochastic OT model, constraints are rated quantitatively on a scale. These ratings are noisy: that is, the precise rating of a constraint at a given point in time is expressed by a probability distribution. However, as in classical OT, at the moment of evaluation, there is a strict ranking of constraints and the candidate outputs are evaluated in accordance with that ranking.

Effectively, this means that the relation between two constraints is potentially inconsistent. If constraint A is rated above constraint B by a sufficiently wide margin, then A will outrank B at 100% of evaluation times. However, if the rating difference is smaller, then B will sometimes outrank A: in principle, a negligible rating difference will result in A being ranked above B only about 50% of the time, with B being ranked above A around 50% of the time.

Given that the constraint ratings are noisy, a stochastic OT model does not deterministically predict the constraint ranking that will take effect in any given scenario. Consequently, such a model is more accommodating than the corresponding classical OT model. However, it has less predictive power, as it is not generally possible to state with certainty that a given output will be generated given any specified input.

In particular, the stochastic OT model is accommodating in that a set of data that would be inconsistent in the classical model (for a specified constraint set) might be consistent for the stochastic model. A set of data can be inconsistent in the classical model if it requires a set of constraint rankings that are mutually irreconcilable: for instance, if it can be explained only by positing constraint rankings A > B, B > C and C > A. By contrast, in a stochastic model, it is entirely possible for such a set of data to arise from a single set of constraints: if A, B and C are similarly rated, then it is quite possible for the pairwise constraint rankings A > B, B > C and C > A to arise at three separate evaluation times. Hence, an individual speaker whose productions were classically inconsistent might still be acting in accordance with a stochastic OT system based upon the same constraint set. This could encompass, for instance, the case of a speaker producing different outputs on different occasions in response to identical experimental stimuli. I discuss the implications of this for profiling individual speakers' constraint rankings in chapter 7.

This additional flexibility represents a disadvantage when it comes to attempting to disprove the validity of a model of this type, given a specified constraint set. For classical OT, a data set which is inconsistent suffices to demonstrate the inadequacy of the model: if we do not allow for error, a set of n data points could in principle be sufficient to prove a model with n constraints inadequate. By contrast, disproving the validity of a stochastic model would require a statistically significant set of data, which in principle might have to be arbitrarily large.

In summary, as stochastic OT is less predictively useful than classical OT, it makes sense first to implement a model using the assumptions of consistent constraint ranking that are built into classical OT. However, stochastic OT provides a standard method of weakening the assumptions of OT in such a way as to account for deviations from the predicted data.

3.2.2. Bidirectional OT

Given the prevalence of bidirectional approaches within the OT pragmatics literature, it is worth considering the motivations for and characteristics of such accounts. The bidirectional OT paradigm, introduced by Dekker and van Rooij (2000), is intended to apply game-theoretic considerations to the study of linguistic phenomena. In bidirectional OT, a linguistic interaction is construed as a strategic situation, and the relation between the utterance and the interpretation is identified with the strategy adopted by the participants. Focusing on both the speaker and the hearer, the bidirectional OT gives as its output a set of form-meaning pairs that is optimally harmonic. Modelling a system in this way gives us a characterisation of the typical meaning of the terms in the system.

Within semantics and pragmatics, OT was initially used to account for hearers' interpretations of utterances, by characterising natural language interpretation as an optimisation problem (Hendriks and de Hoop 2001). In this approach, constraints are proposed as a means of disambiguating utterances, for instance by performing anaphora resolution. Blutner (2000, 2006 i.a.) elaborated upon this approach, arguing for the necessity of speaker-referring constraints in addition to hearer-referring constraints, and thus proposing a bidirectional OT account of various pragmatic phenomena, including embedded implicatures and free choice implicatures. Blutner's bidirectional OT is a constraint-based theory that selects optimal pairs (in the semantic case, form-meaning pairs), using constraints that refer both to the speaker and the hearer. By contrast, traditional 'unidirectional' OT merely selects the optimal candidate at the output level corresponding to a given form at the

input level, and does not guarantee a perfect matching. In the model of Hendriks and de Hoop (2001), this means that multiple forms may share the same meaning.

As a case in point, Krifka (2009) adopts a bidirectional OT analysis of the meaning of round and non-round numerals. He argues for the relevance of two constraints: a speaker-referring preference for simplicity of expression, and a hearer-referring preference for approximate rather than precise interpretations. He then demonstrates how a bidirectional OT model applied to the numerals 'one hundred' and 'one hundred and three' correctly assigns an approximate meaning to the former, and a precise meaning to the latter.

However, although bidirectional OT is potentially effective for characterising the whole system, and thus for establishing the preferred interpretations of forms, it is not as useful for generating psychologically plausible accounts of individual instances of usage. Blutner (2006: 16) draws a distinction between strong bidirectional OT, for which 'unidirectional optimization (either speaker or hearer perspective) is sufficient to calculate the solution pairs' and which can plausibly be used 'to construct cognitively realistic models of online, incremental interpretation', and the weaker form of bidirectional OT that he advocates. The mechanism underlying this latter form is non-local, so 'the proposed algorithms that calculate the super-optimal solutions do not even fit the simplest requirements of psychologically realistic models of online, incremental interpretation' (ibid.). That is, although weak bidirectional OT provides a suitable answer to the questions discussed by Blutner, it does not give rise to a psychologically plausible procedure for selecting a single meaning corresponding optimally to a single form (or vice versa). By contrast, strong bidirectional OT can be implemented by applying unidirectional optimisation, but does not yield satisfactory results, in Blutner's account. However, I agree with Blutner that appeal to unidirectional optimisation might provide a plausible means of constructing a cognitively realistic model of interpretation, and further observe that the same could be said for usage if we adopt a speaker-referring rather than hearer-referring unidirectional account.

In addition to the psychological considerations discussed by Blutner, there are other reasons to suppose that a bidirectional account might be unsuitable as a treatment of numerical quantifier usage and interpretation. For instance, it is not entirely clear how an account such as Krifka's could generalise to a case where there are arbitrarily many distinct forms and meanings in the system, and a counting argument suggests that these cannot be paired off (unless the range of possible meanings is restricted in some way). As a purely factual

generalisation, I also doubt that distinct forms partition up the space of possible meanings anywhere near as efficiently as they might, and therefore suggest that a game-theoretic account of speaker-hearer interaction is more a statement of what might be possible in interaction than one about what actually occurs.

Nevertheless, despite these reservations, we should consider what bidirectional OT adds to the unidirectional system: namely, reference to the hearer. A motivation for Blutner's work on bidirectional OT was the notion that Hendriks and de Hoop's (2001) unidirectional hearer-referring account was inadequate, and that speaker-referring constraints were necessary. Here we are entertaining the notion that a system of speaker-referring constraints may be sufficient, on the principled grounds that utterances arise solely from the speaker and must be attributed to the speaker's mental processes, even if their effect appears to be directed toward the benefit of the hearer.

The question naturally arises of whether we can dispense with hearer-referring constraints in this way. It is not a logical necessity that we must do without them entirely: I would merely argue that any such constraints must be available to the speaker and can be considered to act upon the speaker. Even so, these might invoke theory of mind considerations and thus be addressed specifically to the communicative needs of the hearer. Such an extension of the theory would again represent a weakening, through the addition of extra constraints: and these constraints could be argued to make the theory unfalsifiable, as in the case of the proposed constraint referring to the speaker's intention (section 2.5.2).

Thus, although bidirectional OT is not suitable as a model of numerical quantifier selection, its awareness of the needs of both speaker and hearer provides an indication of a further way in which the unidirectional model under discussion could be weakened, if necessary.

3.2.3. Connectionism and Harmony Theory

The roots of Optimality Theory lie in connectionism, and specifically the development of Harmony Theory (Smolensky 1986). A connectionist model consists of layers of nodes, each of which can possess some level of activation, plus a set of connections between these nodes. When activation is applied to such a system at the input layer, this activation spreads to connected nodes on other layers. The way in which this activation spreads is modulated by the strengths of the connections. In an established network, an activation pattern applied at the input layer will thus correspond systematically with an activation pattern which it induces

at the output layer. Within a phonological setting, we could interpret the input layer as representing the underlying form and the output layer as representing the surface form; in terms of numerical quantification, we could think of the input layer as corresponding to the situation and the output layer as representing the possible utterances. The information content of such a network is encoded in the connection strengths.

Smolensky (1986) defines a notion of harmony on a connectionist network as shown in (26).

(26) $\textbf{Harmony} = \sum_{i,j} u_i w_{ij} v_j^{\square}$

> where u$_i$ is the activation of unit *i* on the input layer,
> v$_j$ is the activation of unit *j* on the output layer,
> w$_{ij}$ is the strength of the connection between input unit *i* and output unit *j*.

We assume here that the network has only two layers, and therefore that input units are directly connected to output units. In such a network, a connection can be thought of as a soft constraint: a connection from unit *i* to unit *j* with a positive weight ($w_{ij} > 0$) effectively states that given an input *i*, an output *j* is to be preferred. If this connection has a negative weight ($w_{ij} < 0$), it states that given an input *i*, an output *j* is to be dispreferred.

The measure of harmony defined in (26) provides a way to evaluate outputs given a particular input. Specifically, the principle of Harmony Maximization (Smolensky 1986) is then argued to apply: the optimal output will be that which maximises harmony. Put simply, the output is required to achieve the best possible satisfaction of all the soft constraints in the network. Harmony Maximization defines what constitutes this optimal outcome, bearing in mind that some constraints are stronger than others (as encoded by connection weights of large magnitude), and also that multiple nodes at the input level may be activated to differing degrees. As McCarthy (2002: 60) puts it, 'a high harmony value means that lots of relatively robust soft constraints are being obeyed throughout the network'.

By contrast with classical OT, this type of model does not guarantee that the strongest single constraint is obeyed. In OT, there is no possibility of lower-ranked constraints 'ganging up' and overruling a higher-ranked constraint[13]. In a harmonic system, we could identify

---

[13] The notion of local conjunction, originally proposed by Smolensky (1995), represents an alternative way of adding this property to the system.

something akin to a highest-ranked constraint, either as the connection with the greatest magnitude of weight w_{ij} or the connection with the greatest magnitude value of $u_i w_{ij}$ for a given input. But in either case, there is no guarantee that the output preferred by this constraint (the corresponding v_j) will actually be selected: other constraints in the system, weaker individually, may collectively militate against its selection.

From a Harmony Theory perspective, OT 'can be viewed as abstracting the core idea of the principle of Harmony Maximization and making it work formally and empirically in a purely symbolic theory of grammar' (Prince and Smolensky 1993: 202). OT replaces the notion of weights with one of strict domination, which cannot readily be captured in any kind of connectionist model.

In this proposal, I proceed with an OT-type model, exploiting the ease of use it provides through its formal clarity. However – as remarked earlier – I am cautious about the plausibility of the notion of strict domination as applied to these constraints, and have no *a priori* reason to suppose that they should interact in this way. In many ways, it seems intuitively more plausible that a minor violation of a highly-ranked constraint might be less significant to a speaker than a gross violation of a lower-ranked constraint, and neither classical nor stochastic OT provides us with any way of modelling this behaviour. There is, therefore, something to be said for back-pedalling through a decade of enhancements and implementing a classic connectionist model of numerical quantification.

Schematically, such a model would appear as follows. On the input layer, the situation is encoded. This would encompass all information of relevance to ascertaining the constraint violations, in our classical model. On the output layer, the possible utterances are encoded. We might suppose the existence of intermediate 'hidden' layers. Acquiring the numerical quantifier system can then be characterised as establishing a set of weights for the connections in the system. Employing the system then requires situational information to be presented at the input level. The system will activate candidate utterances at the output level, and the most highly activated can be selected. Reversing the process, interpretation would involve presenting an activation pattern at the output level: this activation could then cascade to the input level, activating situations consonant with the stimulus.

In detail, such a system would be extremely complicated to implement; however, in a couple of areas, plausible short cuts might be available. For instance, if we adopt the accumulator metaphor proposed by Dehaene (1997), then we might think of all possible numerals as being

encoded on a very small set of input units, exploiting the potential for continuous rather than discrete activation levels.

Furthermore, this account might offer a particularly appealing way of handling contextual activation. In such a model, the activation of the input units need not be categorical, but can range from zero to its maximal level. Consequently, if these input units encode contextual occurrences of material, their activation levels could be presumed to decline gradually from the level achieved immediately after exposure. Thus the maximally harmonic output might differ depending on time after exposure to priming, which is an intuitively appealing consequence of this model.

It should be stressed that we need not consider the competing options of OT and connectionism as being an either/or matter. The important thing to note is that we have proposed a set of functionally motivated constraints. Instead of using these constraints as the content of an OT model, we could equally well use them as the basis for a connectionist model. Recall that our model provides a definition of relevant contextual factors: only those factors which are required in the evaluation of constraint violations can influence the selection of an output. Thus, we could unpack from our model the set of contextual factors and use these as the specification of the input layer of a connectionist model.

The attraction of this approach is that it is minimally stipulative. A general objection to connectionist models (e.g. Gregg 2003: 109-119) is that they are covertly stipulative, in that the choice of input and output units conditions what the network is able to learn, but this choice is typically not made on principled grounds. One response to this is to base the network on neural considerations and not to identify the input and output units with specific concepts (e.g. Garagnani et al. 2008). In a similar spirit, albeit at a higher level of organisation, our approach would take as input units only concepts that are demonstrated to have some kind of independent psychological reality. Such a model would not stipulate rules, but would allow these to emerge as properties of the system in time. Likewise, the output layer need not be pre-specified, its nodes stipulated to correspond to numerically-quantified expressions: rather, any utterance could be represented on this layer, and the particular connections between quantifying contexts and quantifying expressions could be acquired from distributional considerations. However, as such a model is a distant prospect, and far beyond the scope of this thesis, I will set this idea aside in what follows.

3.3. Predictions from constraint interaction

Having briefly considered alternative formalisms to classical OT, we now turn back to the question of how predictions can be derived from the constraint set proposed in chapter 2. For the reasons discussed above, I will focus on predictions within the classical OT model, although the limitations of this will be borne in mind. In the following chapters, I will attempt to apply the model to widely-discussed cases of numerical quantifier usage, but first I shall illustrate the approach to be taken with reference to simpler examples.

3.3.1. Approximations

In a similar spirit to Krifka (2009), we can discuss the use of round numbers as approximations, by appeal to the constraints on numeral salience (NSAL), informativeness (INFO) and quantifier simplicity (QSIMP).

Let us start under the working assumption that the semantics of round and non-round numbers are essentially the same. Given that numerals can be used approximatively, as evident in the acceptability of (27) to describe a situation in which around 100 people are present, we thus assume for the time being that numerals have an approximate semantics[14].

(27)   There are a hundred people in this room.

Now let us consider two situations. In the first case, the speaker wishes to express the notion that 50 people are present. Candidate utterances include (28) and (29).

(28)   There are 50 people here.
(29)   There are 51 people here.

If we further assume that the use of an inexact expression violates informativeness[15], the tableau is as shown in Table 6.

---

[14] At the same time I assume, with Krifka (2009), that numeral semantics is punctual rather than lower-bounding. However, 'punctual' in this sense does not entail that the meaning is a single value, merely that it is double-bounded: so '100' is presumed to mean 'about 100' semantically but not 'at least 100'.

[15] This does not follow from the definition of informativeness proposed in the previous chapter. I return to the question of how to characterise informativeness in a fully satisfactory way in Chapter 7.

Table 6: OT tableau for '50' vs. '51', INFO, NSAL & QSIMP, '50' situation

	INFO	NSAL	QSIMP
50			
51	*	*	

In this case, we see that '50' harmonically bounds '51': that is, it incurs a subset of the latter's violations. This means that '50' is preferred to '51' under any constraint ranking. Therefore, unless we include additional constraints (such as numeral priming), (29) cannot be used in the situation described.

What about the situation in which the speaker wishes to express the notion that 51 people are present? In this case, the tableau is as shown in Table 7.

Table 7: OT tableau for '50' vs. '51', INFO, NSAL & QSIMP, '51' situation

	INFO	NSAL	QSIMP
50	*		
51		*	

This time there is no relation of harmonic bounding. For speakers who rank INFO > NSAL, '51' is preferred. For speakers who rank NSAL > INFO, '50' is preferred.

As a generalisation across speakers, then, this model predicts that non-round numbers cannot be used to express round quantities, whereas round numbers can be used to express both round and non-round quantities. Under the assumptions enumerated above, we thus obtain matching predictions to those of Krifka (2009).

Is it realistic to suppose that the semantics of round and non-round numerals are identical, though? Or is it the case that round numbers possess an approximate semantics and non-round numbers do not? I do not wish to make a dogmatic commitment one way or the other, but this model has potentially interesting implications for questions of this type concerning the semantics-pragmatics interface. Suppose that speakers were to obey a pragmatic system of this type, while possessing a semantics of number which held that both round and non-round numbers were approximative. Their output would then be almost invariably consistent with an alternative hypothesis under which round numbers could be approximative but non-

round numbers could not. Indeed, the only data contradicting the latter hypothesis would arise in cases where the non-round number was contextually activated, among speakers who ranked numeral activation above both numeral salience and informativeness. It seems wholly plausible that learners of language might encounter so few such examples that they could happily internalise a semantics of number in which round numbers could be approximative but non-round numbers could not. Thus, under such a model, a pragmatic preference of this type might become semantic in time.

With this in mind, let us now entertain the assumption that round numbers are potentially approximative and non-round numbers are not. Under this assumption the usage question discussed above becomes trivial, but we can now say something about explicit approximation.

Suppose that the speaker wishes to express the notion that 100 people are present. Candidate utterances include (30)-(33), and their tableau is shown as Table 8.

(30)   There are 100 people here.
(31)   There are exactly 100 people here.
(32)   There are about 100 people here.
(33)   There are about 99 people here.

Table 8: OT tableau for (30)-(33), INFO, NSAL & QSIMP, '100' situation

	INFO	NSAL	QSIMP
100	*		
Exactly 100			*
About 100	*		*
About 99	*	*	*

Given that 100 is ambiguous between precise and imprecise readings, (30) violates informativeness. Nevertheless, it harmonically bounds (32), which violates informativeness as well as quantifier simplicity. (32) in turn harmonically bounds (33), which also violates numeral salience. Therefore the choice is between (30) and (31), with speakers who rank INFO > QSIMP preferring (31) and speakers who rank QSIMP > INFO preferring (30).

63

Similarly, if the speaker wishes to express the notion that about 100 people are present, candidates include (30) and (32). The tableau is as shown in Table 9.

Table 9: OT tableau for (30) and (32), INFO, NSAL & QSIMP, 'about 100' situation

	INFO	NSAL	QSIMP
100	*		
About 100			*

Again, we consider that '100' violates informativeness, as it is ambiguous between approximate and precise readings. Here it is again predicted that speakers who rank INFO > QSIMP prefer (32), whereas speakers who rank QSIMP > INFO prefer (30).

If the speaker intends to say that 99 people are present, the candidate utterances include (34)-(37), and the tableau is shown in Table 10.

(34)   There are 99 people here.
(35)   There are exactly 99 people here.
(36)   There are 100 people here.
(37)   There are about 100 people here.

Table 10: OT tableau for (34)-(37), INFO, NSAL & QSIMP, 'about 99' situation

	INFO	NSAL	QSIMP
99		*	
Exactly 99		*	*
100	**		
About 100	*		*

Here we assume that '99' is not semantically ambiguous, and therefore does not violate INFO. We also assume that '100' is semantically ambiguous, and therefore constitutes a more serious violation of INFO than does 'about 100'. In the resulting tableau, (34) harmonically bounds (35), but any of (34), (36) or (37) might be optimal, depending on the constraint rankings, as follows.

- INFO > NSAL > QSIMP, INFO > QSIMP > NSAL, QSIMP > INFO > NSAL: (34) optimal.

- NSAL > QSIMP > INFO, QSIMP > NSAL > INFO: (36) optimal.

- NSAL > INFO > QSIMP: (37) optimal.

Thus we see how a particular constraint ranking can give rise to a particular signature of preferences over a range of contexts. In principle, this should make it possible to determine whether the behaviour of an individual speaker is consistent with respect to some possible constraint ranking, although it may be difficult to do so, as speaker's intention is problematic to measure. Moreover, it should be possible to make generalisations about the way in which quantifiers of this form are used in contexts, by averaging over possible constraint rankings and appealing to the emergence of the unmarked.

3.3.2. Corrections to underinformative and false statements

In order to illustrate the behaviour of the priming constraints, I will discuss an experimental paradigm designed to elicit corrections of numerically quantified statements. This methodology, similar to that for experiment 1 (section 2.4.5.2), enables context to be manipulated in order to make specific numerals and quantifiers salient. It thus renders the numeral priming (NPRI) and quantifier priming (QPRI) constraints potentially relevant in establishing the optimal outputs.

3.3.2.1. Experiment 2 – Corrections to underinformative and false quantifying statements

In this experiment, participants are exposed to 24 visual displays, each depicting three boxes. In each box there are either n or $n+1$ identical instances of an item ($n = 2$, 3 or 4). The participant hears a description of the display by a cartoon character, 'Mr Caveman'. The participant is then asked whether the description is appropriate, and if not, to say what Mr Caveman should have said.

In each case, Mr Caveman's utterance was 'There are Q Xs in each box', where X denotes the name of the item and Q denotes one of the following quantifiers. Their truth-value and informational status is shown in parentheses.

- more than n-1 (true, informative)
- more than n (false)

- at least n (true, informative)

- at least n-1 (true, underinformative)

- fewer than n+1 (true, informative)

- fewer than n (false)

- at most n+1 (true, informative)

- at most n+2 (true, underinformative).

According to the model proposed in this thesis, the participant's choice of correction (where issued) should be influenced by considerations including quantifier priming (QPRI), numeral priming (NPRI), quantifier simplicity (QSIMP) and informativeness (INFO).

Considering the case of 'more than n' first, possible truthful corrections include (38)-(41)[16]. The tableau is given as Table 11. Here I assume that the superlative quantifiers 'at least/most' are more complex than the corresponding comparative quantifiers 'more/fewer[17] than', an issue discussed at greater length in chapter 4.

(38)   more than n-1

(39)   at least n

(40)   at most n+1

(41)   fewer than n+2

Table 11: OT tableau for (38)-(41), QPRI, NPRI, QSIMP & INFO, 'more than n' situation

	QPRI	NPRI	QSIMP	INFO
More than n-1		*		
At least n	*		*	
At most n+1	*	*	*	
Fewer than n+2	*	*		

In this table, 'at least n' harmonically bounds 'at most n+1', and 'more than n-1' harmonically bounds 'fewer than n+2'. The optimal correction is therefore predicted to be

---

[16] I have omitted less informative options such as 'at least n-1' from consideration here, as these incur a superset of the violations of their more informative counterparts.

[17] I group 'fewer than' and 'less than' together under the former label for convenience of discussion here, as the difference is not theory-critical in this example.

'more than *n*-1' for speakers ranking QPRI > NPRI or QSIMP > NPRI, and 'at least *n*' for speakers ranking NPRI > QPRI and NPRI > QSIMP.

Similarly, for 'at least *n*-1', candidate utterances again include (38)-(41), and the tableau is given as Table 12.

Table 12: OT tableau for (38)-(41), plus test utterance, QPRI, NPRI, QSIMP & INFO, 'at least *n*-1' situation

	QPRI	NPRI	QSIMP	INFO
More than *n*-1	*			
At least *n*		*	*	
At most *n*+1	*	*	*	
Fewer than *n*+2	*	*		

Despite the changes in constraint violations, the relations of harmonic bounding remain the same: 'at least *n*' bounds 'at most *n*+1', and 'more than *n*-1' bounds 'fewer than *n*+2'. Thus, the choice is again predicted to be between (38) and (39). This time, 'more than *n*-1' is preferred if NPRI > QPRI or QSIMP > QPRI, while 'at least *n*' is preferred if QPRI > NPRI and QPRI > QSIMP.

As the above discussion does not rely upon considerations of entailment direction, a similar argument goes through *mutatis mutandis* for the cases of 'fewer than *n*' and 'at most *n*+2'. In the former case, 'fewer than *n*+1' and 'at most *n*' are predicted to be potential corrections under some constraint ranking. In the latter case, 'fewer than *n*+2' and 'at most *n*+1' are predicted to be potential corrections.

Having sketched these predictions, I will now examine the relevant experimental data.

METHOD

The experiment was administered as described above. Sample materials are presented in Appendix B.

Participants

30 participants completed the experiment, all members of the University of Cambridge within the age range 18-40. Of these, two were excluded from the final analysis, one for failing to

respond to sentences judged appropriate, the other for failing to correct sentences judged inappropriate. The discussion below relates to the results from the remaining 28 participants.

Results

The statements labelled 'true and maximally informative' above were accepted in 98% of trials (329/336 cases). Those labelled as 'false' were rejected in 99% of cases (165/168). This indicates that the participants were competent with the task, and that they were satisfied that the semantics of these sentences accorded with the situations presented.

'More than n' was corrected in 81 out of 84 trials (96%). 42 of these corrections were of the form 'at least n' and 32 were of the form 'more than n-1'. In total, 91% of the corrections were of the predicted types. 'Fewer than n' was corrected in all 84 trials: 17 were 'at most n' and 44 were 'fewer than n+1'. In total, 73% of the corrections were of the predicted types.

'At least n-1' was corrected in only 13 out of 84 trials (15%). Each of these 13 corrections used the form 'at least n'. 'At most n+2' was corrected in 59 out of 84 trials (70%), 49 times with 'at most n+1' and 4 with 'fewer than n'. In total, 90% of the corrections were of the predicted types.

Discussion

Broadly these data agree with the predictions of the constraint-based model as to which forms are potentially optimal outputs. For each prompt, the two forms identified above as not harmonically bounded account for a large majority of the preferred corrections. However, several limitations should be noted in this experiment and its results. Harmonically bounded corrections do surface, which may reflect incompleteness in the set of constraints considered in drawing the predictions. Participants are not entirely consistent in their behaviour within and across conditions, although this might be attributable to performance errors. It could be argued the priming effects exerted in this experimental paradigm are stronger than is typical for normal interactions, on account of the proximity of the previously-mentioned quantifier, although this is presumably typical of real-life instances of corrections to false and underinformative utterances. Thus, the experiment is pertinent at least to certain cases of practical usage, and more generally stands as a demonstration of how the ordering of constraints for a given speaker can be explored in an empirical fashion.

3.4. Summary

In this chapter, I have considered how the OT formalism (and its variants) can be used to account for and to predict the selection of numerically quantified expressions. In section 3.1 I exemplified this with reference to a toy example over just three constraints. In section 3.3 I considered two more complex examples, one involving approximation and one involving corrections to numerically quantified expressions. These give some indication of the way in which this model can be used. In the following chapters, I explore this model's relevance to more widely-discussed concerns, first considering its role in accounting for the usage of comparative and superlative quantifiers and then considering its predictions as to the scalar implicatures arising from such expressions.

4. TOWARDS A PRAGMATIC ACCOUNT OF SUPERLATIVE QUANTIFIER USAGE[18]

It has historically been assumed that comparative ('more than', 'fewer/less than') and superlative ('at most', 'at least') quantifiers can be semantically analysed in accordance with their core logical-mathematical properties. However, recent theoretical and experimental work has cast doubt on the validity of this assumption. Geurts and Nouwen (2007) have claimed that superlative quantifiers possess an additional modal component in their semantics that is absent from comparative quantifiers, and that this accounts for the previously neglected differences in usage and interpretation between the two types of quantifier that they identify. Their semantically modal hypothesis has received additional support from empirical investigations. In this chapter, I further corroborate that superlative quantifiers have additional modal interpretations. However, I propose an alternative analysis whereby these quantifiers possess the semantics postulated by the classical model, and that the additional aspects of meaning arise as a consequence of psychological complexity and pragmatic implicature. I explain how this model is consistent with the existing empirical findings, and present the findings of four novel experiments that support this model above the semantically modal account. Finally I explore how these findings can be accommodated within the constraint-based model that is the focus of this thesis, and argue that this can be seen as a generalisation of the pragmatic account.

4.1. Overview

The comparative quantifiers 'more than' and 'fewer/less than' have traditionally been regarded as equivalent to the mathematical symbols $<$ and $>$ respectively. Superlative quantifiers are those of the form 'at most' and 'at least', and have traditionally been regarded as equivalent to the symbols $\leq$ and $\geq$ respectively. Restricting our attention to cardinalities, it follows from this approach that superlative and comparative quantifiers are interdefinable, as in examples (42) and (43).

(42)   John has at most two cars $\Leftrightarrow$ John has fewer than three cars.
(43)   Kelly has at least three children $\Leftrightarrow$ Kelly has more than two children.

This view is challenged by Geurts and Nouwen (2007), who argue that equivalence does not typically hold between comparative and superlative quantifiers. They observe four

---

[18] This chapter incorporates and expands upon work published in Cummins and Katsos (2010).

differences in the usage and interpretation of these types of quantifier. First, they note that superlative quantifiers admit a specific construal, absent from the comparative quantifier. Secondly, they identify differences in the patterns of inference that arise from the putatively equivalent sentences. Thirdly, they observe distributional differences between comparative and superlative forms. And finally, they claim that certain usages of comparative quantifiers give rise to ambiguity that does not follow from the superlative 'equivalent'. To address these issues, they develop a proposal in which comparative and superlative quantifiers differ in modality.

Geurts and Nouwen's proposal gives rise to several empirically testable predictions, notably that the superlative quantifiers will be mastered more slowly by acquirers, that they will be disfavoured in processing, and that they will give rise to different reasoning patterns. These proposals are investigated by Geurts et al. (2010), in a series of experiments. Broadly, the predictions are borne out, and thus Geurts and Nouwen's account is favoured by comparison with the classical approach.

In this chapter, I develop an alternative proposal to that offered by Geurts and Nouwen (2007). Rather than proposing a modal component to the semantics of superlative quantifiers, I propose that there is a fundamental difference in complexity between expressions conveying $<$ and $>$ and those conveying $\leq$ and $\geq$, which I will argue could arise from the disjunctive nature of non-strict comparison. I will first argue, following Büring (2007), that the use of superlative quantifiers triggers an implicature. I show that the classical model of quantifier semantics, augmented with this distinction, gives similar predictions to those made by Geurts and Nouwen, and is consistent with the data that they use to argue against the classical model. I further demonstrate that the data obtained by Geurts et al. (2010) are compatible with this account, and present additional experiments that demonstrate the availability of inferences predicted to be unavailable by Geurts and Nouwen, and the acceptability of statements that they predict to be unacceptable. I show that these data are compatible with the augmented classical model, and argue that this account should be preferred to the semantically modal account on this basis, as well as considerations of parsimony and acquirability. Finally I articulate this intuition in terms of the constraint-based model, which can be thought of as a more general account in the same spirit and which gives rise to the same predictions.

4.2. Problems with the classical view of comparative and superlative quantifiers

As outlined above, the classical treatment of comparative and superlative quantifiers considers them to be interdefinable in a systematic way: 'at most n' $\Leftrightarrow$ 'fewer than $n+1$', 'at least n' $\Leftrightarrow$ 'more than n-1'. In addition to providing us with an elegant formal analysis of these entities, this account conforms with our naïve intuitions about the truth conditions of numerically quantified statements. These intuitions require that 'John has at most two cars' is false only in cases where the cardinality of the set {John's cars} is 3 or more, and 'Kelly has at least three children' is false only in cases where the cardinality of the set {Kelly's children} is 2 or less.

However, Geurts and Nouwen (2007) identify a number of areas in which this account is unsatisfactory. They are suspicious of interdefinability on the grounds that it implies that one set of quantifiers is entirely redundant, given the existence of the other[19]. They also specify a number of additional objections to the analysis, as described in the following paragraphs.

A key objection is that the inference patterns arising from the superlative quantifiers differ from those admitted by the comparative forms. Geurts and Nouwen argue that a sentence such as (44a) gives rise to the inference (44b) but not to the inference (44c), despite (44b) and (44c) being semantically identical on the classical view.

(44a)   Dave had (exactly) three Martinis.
(44b)   Dave had more than two Martinis.
(44c)   Dave had at least three Martinis.

They support their intuitions by verifying this claim experimentally, as will be discussed in section 4.4.1. Hence, it appears that the meanings of comparative and superlative quantifiers systematically differ in some profound way.

In addition to these observations, Geurts and Nouwen discuss examples that suggest that the distribution of comparative and superlative quantifiers differs. They observe that the argument of superlative quantifiers can have a specific construal, which is not licensed by the

---

[19] Arguably the quantifiers are not, even on the classical view, interdefinable on non-discrete sets such as are used in measurements: 'more than 2 metres' cannot be precisely equivalent to 'at least x metres' for any x. This could license the existence of both types of quantifier, although by itself it does not explain why both are actually used in discrete cases. In this chapter, I follow Geurts and Nouwen (2007) in focusing on these discrete cases, but note that this argument has implications for the analysis of the more general continuous case.

theoretically equivalent comparative quantifiers. This provides a referent for the 'namely' clause in (45a), which is unacceptable in (45b).

(45a)   There are at most two people who have that authority, namely the Queen and the Prime Minister.

(45b)   *There are fewer than three people who have that authority, namely the Queen and the Prime Minister.

Geurts and Nouwen (2007: 537f) also note that the superlative quantifiers have a wider range than their comparative counterparts, citing the following examples.

(46a)   Betty had three Martinis at most/*fewer than.
(46b)   At least/*More than, Betty had three Martinis.

They note that some contexts permit the comparative but not the superlative quantifier, though the prohibition is less clear-cut, as in (47).

(47)   Betty didn't have $^?$at least/more than three Martinis.

Finally, they claim that sentences with comparative quantifiers are sometimes ambiguous in a way that those with superlative quantifiers are not. In particular, they contrast (48a) and (48b).

(48a)   You may have at most two beers.
(48b)   You may have fewer than three beers.

Geurts and Nouwen argue that the latter (comparative) has a reading under which it does not necessarily prohibit the addressee from having three or more beers, but merely grants express permission for the addressee to have some smaller number of beers if they wish. Although this reading is sometimes elusive, it represents another difference in meaning between the two sentences that is not captured by the classical account.

In summary, there are reasons arising from both the interpretation and distribution of these quantifiers to support the contention that the classical view of their meaning is inadequate. In the following section, I discuss the specific proposal outlined by Geurts and Nouwen (2007) for dealing with these puzzles.

4.3. The semantically modal account of superlative quantifier meaning

Geurts and Nouwen (2007) propose an account of quantifier meaning in which superlative quantifiers have a modal component of meaning. Specifically, they consider example (49).

(49)   Betty drank at least four highballs.

They ascribe to (49) a semantic formula that can be glossed as 'the speaker is certain that there is a group of four highballs each of which was drunk by Betty, and considers it possible that Betty drank more than four highballs' (p.552). For the corresponding 'at most' sentence, (50), they propose the analysis that 'it grants the possibility that Betty had four highballs, and it excludes the possibility that she had more than four' (ibid.)

(50)   Betty drank at most four highballs.

They also propose that comparative and superlative quantifiers differ in argument type. Their conjecture is that superlative quantifiers accept arguments of any Boolean type: that is, both propositional and predicative arguments. In this way, they analyse (51) as an assertion that 'the speaker is sure it isn't raining, and that he considers it possible that something "better" than non-raining might be the case, as well' (ibid.)

(51)   At least it isn't raining.

Using their account, Geurts and Nouwen are able to give solutions to all the problems they previously discussed. From the assumption that comparative and superlative quantifiers are not interdefinable, it follows that there is no redundancy in the system. The more substantive issues raised in the preceding section may be resolved as follows.

- The inference patterns involving superlative quantifiers are a subset of those arising from comparative quantifiers because of clashes of modality. 'At most n' does not imply 'at most $n+1$', 'exactly n' does not imply 'at most n' or 'at least n' and so on.

- The specific construal of the argument of the superlative quantifier is possible because its argument (e.g. 'two people') may be parsed as an existential quantifier. This is not legitimate in the comparative case because the comparative quantifier does not accept non-predicative arguments.

- The validity of superlative quantifiers in a broader range of contexts than comparative quantifiers stems from their ability to accept additional argument types. At the same

time, the inappropriateness of superlative quantifiers in other contexts – for example, under the scope of negation – reflects their modal semantic content.

- The ambiguity in the fourth case does not arise in the modal case due to factors involving the semantic combination of modal expressions – Geurts and Nouwen refer to this as modal concord. Put simply, their view appears to be that the superlative quantifier enters unambiguously into a concord reading with the preceding modal, as these both express possibility.

In sum, Geurts and Nouwen's modal theory of superlative quantifier meaning accounts well for the observed findings, modulo some concerns about the treatment of superlative quantifiers in conditional environments (to which I return later) and in certain marginal embedded contexts[20]. In addition, their theory gives rise to empirically testable predictions. In the following section, I review the work done on investigating these predictions.

4.4. Empirical investigation of quantifier meaning

Geurts and Nouwen (2007) subject some of their intuitions about the inference patterns arising from these quantifiers to empirical investigation. However, their theory gives rise to a broader range of predictions that are also susceptible to testing by experimental means. Geurts et al. (2010) argue that three particular predictions arise from the modal view of superlative quantifier usage: (i) that superlative quantifiers give rise to different inference patterns than comparative quantifiers; (ii) that superlative quantifiers should be harder to learn than comparatives, on the basis of their additional semantic complexity; and (iii) that superlative quantifiers should be harder to process than comparatives, for the same reason.

In this section I summarise the work done by Geurts et al. (2010) to investigate these predictions, as well as the findings of other research bearing upon these questions. These experiments were conducted in English using native speakers, except where otherwise stated.

4.4.1. Inference patterns arising from comparative and superlative quantifiers

Supporting their intuitions, Geurts and Nouwen (2007) performed a pencil-and-paper experiment in which they asked participants to decide whether certain implications were valid. Broadly, their participants concurred that (in Dutch), 'Beryl had three sherries' implied

---

[20] Although I do not discuss these embedded cases (e.g. "Betty didn't have at least three Martinis", Geurts and Nouwen 2007: 554), they could readily be treated by appeal to priming constraints in this model.

both 'Beryl had more than two sherries' and 'Beryl had fewer than five sherries'. By contrast, 78% of their participants rejected the implication 'Beryl had at most four sherries', and more than half rejected 'Beryl had at least three sherries'.

This technique was employed for a wider range of classically valid premise-conclusion pairs by Geurts et al. (2010), again in Dutch. Instead of using 'fewer than five' and 'at most four' as Geurts and Nouwen did, Geurts et al. used 'fewer than four' (accepted 93% of the time) and 'at most three' (accepted 61% of the time). This controls for informativeness, and thus makes it more valid to compare upward- and downward-entailing quantifiers. They also added six further premise-conclusion pairs, admitting three more comparisons: 'at most two'/'at most three' (14% acceptance) versus 'fewer than three'/'fewer than four' (71% acceptance); 'at least three'/'three' (50% acceptance) versus 'at most three'/'three' (18% acceptance); and 'three or four'/'at least three' (96% acceptance) versus 'two or three'/'at most three' (93% acceptance). As these latter two comparisons do not contrast superlative with comparative quantifiers, but instead explore the effect of entailment direction, I shall not discuss them further here.

In summary, these data constitute additional evidence for the non-equivalence of comparative and superlative quantifiers in a reasoning context. Valid arguments involving comparative quantifiers seem to fail, in the opinion of the majority of untrained participants, when recast using superlative quantifiers that are 'classically' equivalent. Most strikingly, the inference 'at most two' $\implies$ 'at most three' succeeds in only 14% of cases, where the putatively equivalent 'fewer than three' $\implies$ 'fewer than four' achieves 71% acceptance. This concurs with the prediction arising from Geurts and Nouwen's account and thus offers it empirical support.

As Geurts and Nouwen (2007) discuss, there are issues concerning the interpretation of the bare numerals in these items, inasmuch as these might mean 'exactly n' or 'at least n'. This distinction is critical to the validity of several of the inferences under test: for instance, 'at least three' $\implies$ 'three' is false under the first reading of 'three' but is tautologous under the second. Geurts et al. (2010) address this in a follow-up task using 'exactly three', obtaining similar results; however, it could be argued that this modification draws the participants' attention to the underinformativeness of the non-exact statement in the consequent and thus could bias participants towards rejecting the inference. Nevertheless, the distinction between comparative and superlative quantifiers is compellingly supported by this experiment.

4.4.2. Delay in acquisition of superlative quantifiers

Prior to Geurts and Nouwen's (2007) prediction, the acquisition of comparative and superlative quantifiers had already been compared by Musolino (2004: 26-8). In his experiment, participants were given a selection of cards with zero to four objects on them and asked to select those with 'exactly 2', 'at least/most 2' or 'more than 2'. Adults performed at or near ceiling in all conditions. However, while children (aged 4-5) were 100% accurate on 'exactly 2' and 88% accurate on 'more than', they performed at chance on the superlative quantifiers. By asking the child participants about their understanding of these terms, Musolino demonstrated that their poor performance on the superlative quantifiers was rooted in a profound lack of understanding of these quantifiers' meanings.

Geurts et al. (2010) further develop this line of enquiry, using a different experimental protocol. Their experiment involved presenting participants with a set of six boxes, some of which contained a toy of a certain kind. They asked the participants to make the situation match the sentence they were about to hear, either by adding toys, removing toys or leaving the boxes as they were. The sentence was then uttered and the participants' behaviour recorded. The test sentences were of the form 'Q of the boxes have a toy', where Q was a numerical quantifier.

This task was administered to adults and to children aged 11 years. Adults performed at 100% in all conditions, while the children's performance ranged from 97% on 'more than three' to 42% on 'at most three'. 'More than' and 'at least' were easier than 'fewer than' and 'at most', respectively, and comparative quantifiers were privileged over superlative quantifiers.

Assuming that these snapshots of development are a reasonable depiction of stages in the process of acquiring comparative and superlative quantifiers, this experimental evidence also supports Geurts and Nouwen's hypothesis. It appears that children first master 'more than' and much later develop an understanding of the corresponding superlative quantifier 'at least'. Similarly, 'fewer than' is followed at a distance by 'at most', which has typically still not been mastered at the age of 11. Therefore, we can conclude that comparative quantifiers are indeed mastered earlier than superlative quantifiers, as Geurts and Nouwen's theory predicts.

In passing, we note that these experimental data are also coherent with input frequencies, if we take general corpora to be indicative of these. It is clear that comparative quantifiers are used substantially more frequently than superlative quantifiers: for instance, the British National Corpus (BNC) gives the frequency of their occurrence with the numerals 1-20 (in digital form) as follows.

- More than: 1243
- Less than: 698
- Fewer than: 86
- At least: 812
- At most: 44

However, this is clearly not the whole story, as we have to account for these frequency trends themselves. From a modal perspective, these could be argued to arise from the difference in core meaning. Later in this chapter, I will argue that the infrequency of superlative quantifiers stems from their complexity. In either case, I do not commit to a view as to whether the order of acquisition is causally modulated by frequency, although this is not implausible.

It should also be noted that these structures also occur in non-numerical quantificational contexts (e.g. 'Credit cards are accepted at most stores'). These uses need to be considered when we attempt to compare the frequencies of different quantifying expressions. In chapter 6, I discuss corpus data concerning numerical quantifier usage in more detail with a view to resolving this issue.

4.4.3. Delay in processing of superlative quantifiers

Geurts et al. (2010) also test the online processing of comparative and superlative quantifiers by adult participants. Participants were presented with a sentence 'There are Q As' or 'There are Q Bs' (where Q is a quantifier) and then a display in which some number of instances of the letter A or B are present. They were asked to press a button to indicate whether the sentence is true or false of the situation displayed. Reading times and decision times were measured. For the decision times, there were effects both of entailment direction and quantifier type, just as for the acquisition experiment: 'at most' was the slowest to be verified, 'more than' the quickest. For the reading times, there were no significant effects. This was argued to support the hypothesis of complexity; the task involving deeper

processing gave rise to evident delays in the superlative case, as predicted from Geurts and Nouwen's account.

It is also worth remarking upon the correlation between complexity/processing difficulty and age of acquisition for these quantifiers, as demonstrated by the previous experiment. The data from these experiments support a view in which the quantifiers that are acquired later are more difficult to process. The reading times in Geurts et al.'s experiment do not differ significantly, but they pattern numerically with the decision times, raising the question of whether evidence might be obtainable that the additional complexity and later acquisition of superlative quantifiers affects shallower processing as well.

4.4.4. Interim summary

We have seen how Geurts et al.'s (2010) experiments bear out the predictions made by Geurts and Nouwen (2007) and thus constitute evidence in favour of the semantically modal account over the classical model. In the following section, I spell out an alternative proposal that captures the aspects of the interpretation of superlative quantifiers that Geurts and Nouwen highlighted, but posits a different division of labour between semantics and pragmatics than they envisioned. I further examine how this can account for the empirical data discussed above.

4.5. A pragmatic account of superlative quantifier meaning

In discussing the classical model (attributed to Barwise and Cooper 1981), neither Geurts and Nouwen (2007) and Geurts et al. (2010) draw any distinction between the comparative and superlative forms. From a formal point of view, this is correct: the classical view predicts that 'at most n' is true in exactly the same set of circumstances for which 'fewer than $n+1$' is true. In this sense, comparative and superlative quantifiers are also equivalent in mathematical complexity: each corresponds to one symbol, be it $>$, $<$, $\geq$ or $\leq$. This amounts to the observation that each quantifier maps a pair of arguments to a truth-value. However, here I wish to consider the possibility that the operators $>$ and $<$ are not equivalent to the operators $\geq$ or $\leq$ because the latter pair possess additional psychological complexity.

One particular way to flesh out this claim is to propose that the operators $\geq$ or $\leq$ can be regarded as disjunctions, at some level of representation. I suggest that $\geq n$ is represented as '$> n$ or $= n$', and $\leq n$ is represented as '$< n$ or $= n$'. As per the classical model, I propose that $\geq$ or $\leq$ provide the semantics of natural language 'at least' and 'at most', just as the operators

> and < provide the semantics of the comparative quantifiers. Hence, under this proposal, not only ≥ or ≤ but also natural language 'at least' and 'at most' are treated as disjunctions. The latter part of this claim is consistent with Büring's (2007) analysis whereby 'at least n' is interpreted as 'exactly n or more than n', and 'at most n' is interpreted as 'exactly n or fewer than n'. This in turn extends the semantic account proposed by Krifka (1999), in which superlative quantifiers are focus-sensitive operators which presuppose the existence of ordered sets of alternative complements.

This proposal has three important implications. First, in accordance with Büring's analysis, the use of a superlative quantifier, whose meaning is a disjunction, gives rise to a quantity implicature, as a consequence of Grice's (1975) first maxim of quantity. 'At least n' conveys semantically that ≥ n holds and implicates that the speaker does not know (or is not at liberty to say) whether it is the case that > n or = n holds. Similarly, 'at most n' conveys semantically that ≤ n holds and implicates that the speaker does not know (or is not at liberty to say) whether it is the case that < n or = n holds. This is the classical clausal implicature associated with disjunction (e.g. Horn 1972), whereby the speaker's assertion 'p or q' implicates that the speaker is not in a position to make a stronger statement, such as asserting that 'p' or that 'q' (each of which entails 'p or q'). This proposal is distinct from that of Geurts and Nouwen (2007) as no specific notion of modality is stipulated as part of the semantics of superlative quantifiers. Indeed, no semantic difference is proposed between comparative and superlative quantifiers (other than the numerical difference in their arguments); the difference in modality is now captured as a pragmatic inference.

A second consequence of this proposal is that both the natural language expressions 'at least n' and 'at most n' and the logical operators ≥ or ≤ are predicted to be more difficult to process than the natural language expressions 'more than n' and 'fewer than n' and the logical operators > and < respectively. This is because, at a psychological level, the disjunction '> n or = n' is predicted to be more complex than either of the disjuncts '> n' or '= n' on their own. We would expect the difference in psychological complexity to manifest itself in a usage preference for the less complex option, when both are available.

A third consequence of this proposal is that any differences in usage and meaning between superlative and comparative quantifiers are not restricted to the specific forms 'at most' and 'at least' versus the specific forms 'more than' and 'fewer/less than'. Instead, they arise because the underlying meaning of the former category of expressions is represented

disjunctively, and thus involves additional psychological complexity and yields pragmatic implicatures while the underlying meaning of the latter is not. Thus, our account predicts that all forms that are semantically expressible as > and < will pattern together and behave distinctively from the forms expressible as ≥ or ≤. This is coherent with the account offered by Nouwen (2010): his class A modifiers are those traditionally expressible as > or <, and his class B modifiers are those expressible as ≥ or ≤. However, unlike Nouwen, this account does not posit a more substantive semantic difference between these classes of modifier. I follow Nouwen in referring to > and < as *strict comparison* and ≥ and ≤ as *non-strict comparison.*

Thus, I propose that by treating 'at most' and 'at least' as disjunctions, we can identify two grounds on which superlative and comparative quantifiers differ, namely the clausal implicature and the psychological complexity associated with the former. In the following, we shall see how this difference can be invoked to explain extant experimental findings on the difference between comparative and superlative quantifiers. In the remainder of this chapter, I first present an experimental justification of the analysis of superlative quantifiers as psychologically complex. I discuss the explanation of extant experimental findings in terms of this proposal. I then exhibit empirical evidence in support of the classical view of quantifiers, augmented by this disjunctive analysis (and consequent implicatures), against the semantically modal account of Geurts and Nouwen (2007). I then consider how this account can be restated within the constraint-based model, considering non-strict comparison as the locus of additional complexity.

4.6. Demonstrating the complexity of non-strict comparison

To the best of my knowledge, the distinction between strict and non-strict comparison has not been studied in detail within the domain of the psychology of mathematics. However, we posit that non-strict comparison is more complex, and more specifically that it may be disjunctive, on the following grounds.

(a)   Assuming that the basic means of comparison between two quantities are 'more/less than' and 'equal to', > and < each correspond to a single simplex operation of comparison. ≥ and ≤ could be derived from these but would then be secondary and presumably more complex.

(b)   The operators > and < are customarily glossed as 'greater than' and 'less than', while ≥ and ≤ are customarily glossed as 'greater than or equal to' and 'less than or equal to'. The lack of a non-disjunctive expression for these operators in common parlance suggests that they are naturally regarded as complex.

The above observations suggest that the claim that superlative quantifiers are disjunctive might be plausible, but fall far short of making a convincing case for it: for instance, perhaps ≥ and ≤ should be glossed as 'at least' and 'at most', respectively. However, it is also possible to investigate this claim by adapting the technique used by Geurts et al. (2010) for studying the processing preference for superlative versus comparative quantifiers.

As briefly described above, Geurts et al. (2010) presented participants with a visual display of a sentence of the form 'There are Q N Xs', where Q denotes a quantifier of the form 'exactly', 'at least', 'at most', 'more than' or 'fewer than', N denotes a number, and X denotes a letter (either A or B). Participants were instructed to press a key once they had read and understood the sentence. They were then immediately presented with a display consisting of some number of instances of the relevant letter (either A or B, depending which was referred to in the preceding sentence), and instructed to indicate whether the preceding sentence had been true or false of the situation, by pressing the appropriate key. Over a series of 38 trials, the response times for each participant in each condition were measured and analysed.

Geurts et al. (2010) demonstrated a processing preference for comparative over superlative quantifiers, by showing that the former gave rise to shorter response times. If non-strict comparison is processed as a disjunction, we would expect comparisons involving strict comparison to give rise to longer response times than strict comparisons, even in the absence of any linguistically relevant content (such as comparative and superlative quantifiers). By contrast, if strict and non-strict comparisons are both equally demanding, there should be no significant effect of comparison type *per se*.

4.6.1. Experiment 3 – Processing costs of strict and non-strict comparison

In this experiment, I replicated the third experiment of Geurts et al. (2010) as described above, with the following change. In place of sentences of the form 'There are Q N Xs', the participants read statements of the form 'X ? N', where X denotes a letter (either A or B), N denotes a number, and ? denotes a symbol (either =, >, <, ≥ or ≤). Each of the 38 items used

by Geurts et al. (2010) was translated into this form (6 involving equality, and 8 for each of the other symbols).

METHOD

The experiment proceeded precisely as Experiment 3 of Geurts et al. (2010). Participants were given the same instructions, with the single exception that the word 'statement' was used instead of 'sentence' to describe the on-screen display. The full set of test conditions is listed in Appendix C.

Participants

The experiment was administered to 20 subjects, all members of the University of Cambridge. Participants were aged between 20 and 36 years. 16 of the subjects were female. The results of two participants were excluded from the following analysis, one for a high error rate (18/38 = 47%, compared to 30/684 = 4.4% for the 18 participants analysed) and one for a slow mean response time (greater than twice the mean for the 18 participants analysed).

Results

Following Geurts et al. (2010), incorrect responses were excluded from consideration. Also, in order to minimise the effect of outliers arising through lapses in concentration, the data were further trimmed by removing from consideration any responses that exceeded a threshold time (mean + 2 SD for the individual participant). Note that this is a conservative manipulation in that this tends to suppress differences between conditions and thus biases towards the retention of the null hypothesis in which there are no such differences.

The participants' performance across the test conditions was as shown in Table 13.

Table 13: Results of experiment 3 (processing costs of strict and non-strict comparison)

Condition	No. of observations	Mean response time in ms (SD)
=	105	982 (314)
>	132	1007 (369)
<	123	1061 (354)
≥	139	1110 (466)
≤	130	1131 (384)

In order to compare all four test conditions at once, a linear regression was performed in R (R Development Core Team 2008), using two predictor variables – 'direction of entailment', which was set to +0.5 for > and ≥ and -0.5 for < and ≤, and 'comparison type', which was set to +0.5 for > and < and -0.5 for ≥ and ≤. This analysis failed to reach significance for direction of entailment (t=1.08, p = 0.14), but a significant effect was obtained for comparison type (t=2.48, p=0.013). That is, the conditions of non-strict comparison gave rise to significantly longer response times than their strict comparison counterparts.

Further pairwise comparisons were performed between the upward- and downward-entailing conditions, and between those of strict and non-strict comparison. > differed significantly from ≥ (Student's t-test, t=2.01, df=269, p < 0.05 two-tailed), while < versus ≤ narrowly failed to reach significance (t=1.51, df=251, p=0.066 one-tailed).

Discussion

The findings of this experiment closely parallel those of Geurts et al. (2010). Specifically, the condition of equality is numerically the fastest to be verified. Of the remaining four conditions, the ≤ condition is the slowest and > the fastest to be verified. There is a statistically significant main effect of non-strict comparison, supporting the hypothesis that non-strict comparison is more complex than strict comparison.

In this experiment, the effect of entailment direction was marginally non-significant, but the pattern resembled that of Geurts et al. (2010), with > faster than < and ≥ faster than ≤. However, neither this proposal nor the semantically modal account of Geurts and Nouwen (2007) makes clear predictions about the effect of entailment direction, so I leave that aside in what follows.

Overall, the pattern across the five conditions is highly similar to that obtained in Geurts et al.'s (2010) third experiment, suggesting that the substitution of sentences involving comparative and superlative quantifiers with statements involving mathematical operators has not materially influenced the outcome.

This result appears to admit four possible explanations. One is that the delay in processing the mathematical operators of non-strict comparison underpins the delay in processing superlative quantifiers. Another is that the reverse is true. A third is that both processing

delays are caused by processing difficulties with a shared underlying representation of non-strict comparison. A fourth is that the common delay is a coincidence and that both processing paths are entirely independent. I consider the first two explanations unlikely, on the grounds that the response times are sufficiently similar in both experiments to suggest that neither task constitutes a single step in the performance of the other task. Moreover, these explanations would be potentially problematic for Geurts and Nouwen's (2007) account, as the mathematical operators and linguistic expressions are putatively non-equivalent on the basis of the latter's modal content. Of the remaining two proposals, I would argue for the 'shared representations' account as a more parsimonious explanation of the observed data than the attribution of the similarities to coincidence.

For the moment, we do not need to commit to a view on whether the above descriptions correctly characterise the participants' process in performing this task. It suffices for our purposes to draw the conclusion – licensed by the same arguments put forward by Geurts et al. (2010) – that non-strict comparison is more complex than strict comparison. This is compatible with the hypothesis that non-strict comparison is treated as disjunctive, but admits other possible explanations, as we shall see. In what follows, I first explore the consequences of adding this into the classical model of numerical quantifier semantics; subsequently I consider how this can be treated by the constraint-based model.

4.7. Consequence of the complexity of non-strict comparison

Having presented empirical evidence for the claim that non-strict comparison is psychologically more complex than strict comparison, and argued that this is consistent with the idea that forms such as superlative quantifiers are treated as disjunctions, I now consider the consequences of this for the system of numerical quantifiers in general.

One such consequence might be that the reasoning patterns arising from superlative quantifiers are not identical to those arising from comparative quantifiers, in that the superlative quantifiers obstruct the generation of logically valid inferences. This follows because the disjunction gives rise to an implicature, as set out in Büring (2007). Recall that, for instance, 'at most' can be analysed as 'less than or equal to', and hence the use of 'at most' in declarative contexts gives rise to the implicatures that 'less than' and 'equal to' are both possibilities. The latter implicature disrupts the inferential process explored in experiment 1 of Geurts et al. (2010), as described below.

Consider the case discussed in section 4.4.1, where participants are asked whether the implication 'Beryl had at most three sherries' $\implies$ 'Beryl had at most four sherries' is legitimate. According to the classical account, this should be acceptable, but it is rejected by the vast majority of participants. By contrast, under a disjunctive account of 'at most', the consequent 'Beryl had at most four sherries' gives rise to the implicature that it is possible, as far as the speaker is concerned, that Beryl had exactly four sherries. This directly contradicts the antecedent, 'Beryl had at most three sherries'. This contradiction has the potential to block participants' acceptance of the implication under test.

Note that this explanation is similar in character to that proposed by Geurts and Nouwen (2007). In their account, the implication 'at most n' $\implies$ 'at most n-1' fails because of a contradiction at the level of semantics: 'at most n' is held to encompass the explicit possibility of 'exactly n'. Our account differs only in that this explicit possibility arises through pragmatics rather than semantics, and specifically through the implicature derived from the use of the complex disjunctive non-strict comparison.

So far we have seen how the classical account, augmented with a more sophisticated analysis of non-strict comparison, can answer some of the criticisms developed by Geurts and Nouwen (2007) and coheres with their observations about the inference pattern, as supported empirically by Geurts et al. (2010). However, we also need to demonstrate that this account is compatible with the full range of experimental data obtained for comparative and superlative quantifiers.

Geurts et al.'s (2010) second experiment bore out the prediction that the comparative quantifiers are mastered earlier in acquisition than their superlative counterparts. Here, once again, our enhanced version of the classical account makes the same prediction as the semantically modal account of Geurts and Nouwen (2007). As we argue that superlative quantifiers possess a more complex meaning than comparatives, we would also predict that superlative quantifiers are disfavoured in acquisition. Hence, this experiment does not adjudicate between the two competing proposals.

The third experiment conducted by Geurts et al. (2010) shows faster verification times for comparative over superlative quantifiers, as discussed in section 4.6. In that section, I demonstrated that the same applies for strict versus non-strict comparison, in the absence of the specific linguistic constructs under investigation (comparative and superlative quantifiers). Given this finding, we would predict that the same pattern should be replicated

whenever strict and non-strict comparisons are in competition. Therefore, our proposal is compatible with the finding that comparative quantifiers are verified faster than superlative quantifiers. Moreover, I would argue that this proposal is more parsimonious than the semantically modal account of Geurts and Nouwen (2007), in that it explains the outcomes both of Geurts et al.'s (2010) third experiment and of experiment 3 of this thesis (section 4.6.1). Earlier I concluded that the superior performance on strict versus non-strict comparison cannot readily be attributed to a semantic difference between comparative and superlative quantifiers, as participants do not appear to be invoking these quantifiers in their verification process. Hence, with reference to the major lines of argument made by Geurts and Nouwen (2007), and the experimental evidence adduced in support of this proposal by Geurts et al. (2010), we see that the pragmatically augmented version of the classical account is equally as well supported as the semantically modal account. In the following section, I attempt to distinguish between these using new empirical data.

4.8. Experimental evidence in favour of the complexity-driven account of superlative quantifier usage

In this section, I present data from a further set of experiments, first replicating Geurts et al.'s findings on the inference judgement task and then obtaining support for the pragmatic account in preference to Geurts and Nouwen's proposal.

4.8.1. Experiment 4 – Judgements of logical inference patterns

In this experiment, I set out to replicate the findings of Geurts et al. (2010) with respect to the inference judgement task. I performed this experiment in order to verify that the patterns observed for Dutch quantifiers were also to be found in English. I also administered a post-test questionnaire to ascertain whether participants had explicit knowledge of any difference between the two types of quantifier.

METHOD

Participants were presented with a series of pages, each with two sentences written on them. They were instructed to circle the answer 'yes' if the first sentence implied the second and 'no' if it did not. Three sentence pairs were used for each of 12 conditions (36 sentence pairs in all); these included the first eight conditions tested by Geurts et al. (2010) and four additional 'false' conditions as controls. The order of these sentence pairs was randomised for the experiment, but the same order used for each participant. As a post-test, participants

were asked to write a brief explanation of why they had answered the way they did, for the first instance of each of the 12 conditions. Full materials are presented in Appendix D.

Participants

A total of 15 adult participants were recruited, all of whom were students at the University of Cambridge. None had any university-level background in mathematics or logic.

Results

The acceptance rates for the implications under test, and the corresponding figures obtained by Geurts et al. (2010) for the eight conditions common to both studies, are presented in Table 14.

Table 14: Results of experiment 4, and comparison with Geurts et al.'s study

1st quantifier	2nd quantifier	Acceptance (%)	Geurts et al. acceptance (%)
3	At least 3	62	50
3	More than 2	100	100
3	At most 3	42	61
3	Fewer than 4	84	93
At most 2	At most 3	2	14
Fewer than 3	Fewer than 4	64	71
At least 3	3	90	50
At most 3	3	26	18
More than 3	Fewer than 3	9	N/A
Fewer than 3	More than 3	0	N/A
3	Fewer than 3	13	N/A
3	More than 3	0	N/A

The metalinguistic judgements elicited from the post-test questionnaire were generally uninformative. These exhibited a degree of uncertainty about some of the conditions but provided no clear indications of any explicit awareness of modality or other semantic effects. I shall not discuss these further in what follows.

Discussion

The results for the conditions common to both this experiment and Geurts et al.'s (2010) were generally very similar, with the exceptions of the 'three' $\implies$ 'at most three' and, most strikingly, the 'at least three' $\implies$ 'three' pairs. I tentatively attribute this to item effects, arising from the potential ambiguity of the bare numeral. Note that 'at least three' $\implies$ 'three' if and only if the latter 'three' is interpreted as lower-bounding (that is, as existential rather than cardinal) whereupon the relation becomes a tautology. In this experiment, the relevant items had consequents 'Anna wrote 3 letters', 'There are 3 cities on the map' and 'Steve owns 3 suits', all of which could conceivably be existential statements (whereas 'Steve has 3 children' would likely be cardinal and might give rise to a different pattern of responses). Rejection rates for the semantically incorrect control conditions were generally at or near ceiling.

Crucially for our purposes, this experiment constitutes a replication of Geurts et al.'s (2010) study as far as the comparative and superlative quantifiers are concerned. We can see that, in English as in Dutch, performance on comparative quantifiers exceeds that on superlatives – judged by traditional standards of logical correctness – in all the cases for which we have comparable data. Thus I conclude that, as predicted, this is not an effect specific to Dutch.

4.8.2. Experiment 5 – Compatibility judgements on numerically quantified expressions

Following Geurts et al.'s analysis, the rationale for the failure of 'at most two' $\implies$ 'at most three' in experiment 4 was as follows. On the modal view, 'at most three' admits the possibility of '(exactly) three'. 'At most two' uncontroversially excludes the possibility. Therefore, as the possibility of '(exactly) three' cannot follow from 'at most two', the implication fails. Following this line of analysis, it is predicted that two such sentences will be judged as logically contradictory. By contrast, the augmented classical view predicts that the two sentences will be logically compatible but pragmatically infelicitous when juxtaposed. I investigated this issue using a method introduced by Katsos (2007, chapter 3; see Katsos 2008 for a review) with the aim of capturing the difference between logical contradiction and pragmatic infelicity for the case of scalar implicature.

In this paradigm, participants are presented with statements and they are asked to give coherence judgements on a scale. Under this methodology, it is predicted that semantically self-contradictory statements would be judged as incoherent, while statements that are

pragmatically self-contradictory (i.e. in which an implicature is explicitly revised) would be judged more coherent, and statements with neither type of self-contradiction would be judged more coherent still. The modal view of superlative quantifiers, holding that 'at most n' semantically conveys the possibility of 'exactly n', predicts that a statement containing 'at most n' and 'exactly n-1' (with reference to the same entities) should pattern with the semantically self-contradictory statements. The pragmatic alternative proposal put forward in this chapter predicts instead that a statement containing 'at most n' and 'exactly n-1' should pattern either with the pragmatically self-contradictory statements (if an implicature is generated) or with the non-self-contradictory statements (if it is not).

METHOD

Participants were presented with a pair of sentences linked by the word 'specifically', such as in (52), where Q denotes a quantifier and n and m denote numbers.

(52)   Jean has Q n houses. Specifically, she has exactly m houses.

They were asked to give a judgement on the coherence of the utterance, rating it on a Likert scale ranging from 5 ('coherent') to -5 ('incoherent').

Two types of control items were included. One category of control items used 'in fact' rather than 'specifically', partly to disguise the goal of the experiment and partly to test whether participants' judgements would differ if the second sentence could be interpreted as a weakening of the speaker's commitment to the proposition originally expressed. A second category of control items were those in which the quantifier and numeral in the first sentence were replaced by 'some', and the numeral in the second sentence was replaced with 'none of', 'half of' or 'all of'. Sentences were chosen in such a way as to license this partitive usage. This tested the participants' response to violations of both semantic and pragmatic contradiction, as discussed above.

78 items were used in total. The value of n was varied over the range 3-5. A full list of materials is provided in Appendix E.

Participants

A total of 20 participants were recruited, all members of the University of Cambridge, in the age range 20-36 years. 14 were female.

Results

Table 15 shows the mean ratings for coherence, and the corresponding standard deviations (SDs), in each of the experimental conditions.

Table 15: Results of experiment 5

Quantifier in first sentence	Quantifier in second sentence	'Specifically' condition		'In fact' condition		Coherent?
		Mean	SD	Mean	SD	
At most *n*	Exactly *n*-1	1.58	2.57	1.87	2.53	?
At most *n*	Exactly *n*	1.90	2.31	1.25	2.60	?
At most *n*	Exactly *n*+1	-4.08	2.34	-4.05	2.10	No
At least *n*	Exactly *n*-1	-4.48	1.50	-4.27	1.88	No
At least *n*	Exactly *n*	1.28	2.50	1.33	2.56	?
At least *n*	Exactly *n*+1	1.95	2.53	2.55	2.16	?
More than *n*	Exactly *n*-1	-4.70	0.93	-4.28	1.92	No
More than *n*	Exactly *n*+1	3.10	2.18	3.20	1.90	Yes
Fewer than *n*	Exactly *n*-1	3.08	1.80	3.13	1.93	Yes
Fewer than *n*	Exactly *n*+1	-4.75	0.73	-4.23	2.05	No
Some	None	-4.60	1.14	-4.07	1.97	No
Some	Half	3.08	2.23	3.35	2.00	Yes
Some	All	-1.08	3.13	0.22	3.10	?

The global mean across all conditions was -0.64 and the SD 3.88.

The semantically self-contradictory items achieved low ratings in both conditions. The semantically (uncontroversially) non-contradictory items (those marked as coherent in Table 15) achieved high ratings. The 'in fact' condition was regarded as more coherent than the 'specifically' condition for 12 out of the 13 cases (significant by the sign test, $p < 0.01$). I will discuss the 'specifically' condition for the critical cases (those marked as '?' for coherence in Table 15).

The first critical case was 'at most *n*…exactly *n*-1'. Comparing this with 'at most *n*…exactly *n*', a Student's t-test gives t = 0.72, df = 118, $p > 0.1$, indicating that there is no significant

difference between these conditions. Comparing 'at most *n*…exactly *n*-1' with the semantically false control 'at most *n*…exactly *n*+1', we obtain t = 12.6, df = 118, p < 0.01. Thus, this condition is highly significantly more acceptable than the relevant semantically false control.

For 'at least', we can compare 'at least *n*...exactly *n*' with 'at least *n*…exactly *n*+1': in this case, we obtain t = 1.46, df = 118, p > 0.1. Again, there is no significant difference between these cases. Both are highly significantly (p < 0.01) more acceptable than the semantically false control condition 'at least *n*…exactly *n*-1'.

Comparing the superlative with the comparative quantifiers, we find that the semantically true comparative conditions ('more than *n*…exactly *n*+1' and 'fewer than *n*…exactly *n*-1') obtain significantly higher acceptability ratings than the corresponding superlative conditions (again, p < 0.01 for all comparisons).

For the control 'some' conditions, further pairwise comparisons reveal a significant preference for 'some…half' over 'some…all' and 'some…none' (p < 0.01 for each). 'Some…half' also significantly outperforms all four critical conditions with superlative quantifiers, which in turn outperform 'some…all' (p < 0.01 for all comparisons).

Discussion

Broadly, the trends in both the 'specifically' and 'in fact' conditions are clear. The statements in which pairs of sentences are semantically compatible, on a classical view, are systematically judged to be coherent. Those in which the sentences are semantically contradictory are judged incoherent. In accordance with the intuitions discussed above, participants appear to show slightly more leniency to self-contradictory utterances in the 'in fact' case than in the 'specifically' case. The results for the comparative cases license the assumption that this test is diagnostic for semantic coherence.

For our present purposes, the critical data are those in which the first sentence contains a superlative quantifier. In these cases, there is once again a clear-cut division between the cases that are semantically self-contradictory and those that are not. Crucially, given a first sentence containing 'at most *n*', participants accept both the continuations 'exactly *n*' and 'exactly *n*-1', with no statistically significant preference for one over the other.

This fails to confirm the predictions arising from the semantically modal account of superlative quantifiers. On this account, the superlative quantifier encodes the possibility of exact equality, a possibility which in our critical cases is then denied by the second sentence. Recall that the failure of the entailment from 'at most n' to 'at most n-1' was attributed to the consequent denying the possibility of 'exactly n'. Nevertheless, in this experiment, the participants accept the relevant utterances as clearly coherent rather than incoherent. Under the hypothesis of modality in the semantics, the lack of significant differences between the conditions 'at most n…exactly n-1' and 'at most n…exactly n' indicates that revising possibility to certainty is just as incoherent as revising possibility to impossibility, which appears implausible. Notably, revising possibility to impossibility (under this hypothesis) is judged significantly more coherent than revising 'some' to 'all', which is a case of pragmatic self-contradiction. This conflicts with the assumption that revising semantic content is less acceptable in this experimental paradigm than revising pragmatic content. I therefore consider these data incompatible with the hypothesis that 'at most' semantically encodes the possibility of equality.

According to the competing pragmatic proposal, the superlative quantifiers give rise to modal interpretations due to pragmatic implicature. Specifically, I propose that the 'at most n' first sentence gives rise to an implicature that 'exactly n' is possible. Therefore it should follow that the 'at most n...exactly n-1' case is comparable to the 'some…all' case, where incoherence arises from the second sentence contradicting an implicature of the first, rather than contradicting its semantics. This prediction is borne out in the experimental data: 'some…all' is judged significantly more acceptable than 'some…none' and the other contradictory cases, as does 'at most n…exactly n-1'. Both are judged less acceptable than the non-contradictory cases with the comparative quantifiers (that do not trigger an implicature), as predicted.

Furthermore, this account also suggests that 'at most n' gives rise to an implicature that 'fewer than n' is possible, which in turn predicts that the condition 'at most n…exactly n' should yield lower ratings than the comparative conditions and similar ratings to those elicited by 'at most n…exactly n-1'. The results of this experiment suggest that this is indeed the case.

Again, we note that these forms are judged to be significantly more coherent than the 'some…all' case, even though we are arguing that these are both self-contradictory at a

pragmatic level. However, there is a sizeable gulf between these and the control cases that are semantically self-contradictory. Put simply, the non-theory-critical cases present a clear pattern: the semantically self-contradictory cases yield low ratings, the pragmatically self-contradictory cases inhabit the middle of the scale, and the non-self-contradictory cases are rated higher. If we measure the superlative quantifier cases against this scale, we see them clearly indicated as pragmatically self-contradictory or fully non-self-contradictory. Either way, this contradicts the semantically modal account. Within the pragmatic account, we would suggest that the intermediate status of these results indicates implicature being generated, but less reliably or robustly than in the paradigm case of 'some' ('but not all'). This resonates particularly with the predictions arising from the constraint-based account, as discussed later in this chapter.

In summary, the coherence judgements elicited in this experiment point to a clear division between semantically self-contradictory and non-self-contradictory utterances, with pragmatically self-contradictory utterances occupying the middle ground. The case of 'at most n…exactly n-1', which should be semantically self-contradictory on the semantically modal account, appears to be either pragmatically self-contradictory or fully coherent, in accordance with the pragmatically augmented classical account. In fact, as predicted by the latter account, it patterns similarly to 'at most n…exactly n'. Therefore, I conclude that this experiment favours the pragmatic over the semantically modal account.

4.8.3. Experiment 6 – Inference patterns in a conditional context

The results of experiment 5 raise the question of whether the inference 'at most two' $\implies$ 'at most three', rejected in experiment 4, is available under the right conditions. In experiment 6, I aim to elicit this inference, using a conditional context.

Intuitively, it seems that the superlative quantifiers behave like comparative quantifiers under the scope of conditionals (e.g. 'If Berta has had at least/most three drinks…'). It was noted by Geurts and Nouwen (2007) that their semantically modal account is unsatisfactory with regard to these and certain other contexts. In this experiment, we investigate whether hearers interpret an utterance such as (53) as a commitment on the part of the speaker to the corresponding condition (54).

(53)   If Berta has had at most three drinks, she is fit to drive. Berta has had at most two drinks.

(54)   Berta is fit to drive.

This conclusion appears to be licensed by the inference from 'Berta has had at most two drinks' to 'Berta has had at most three drinks'. If the semantically modal account is correct, this inference should not be available and the conclusion cannot be drawn. If the classical account is correct, this inference is available and the conclusion can be drawn unless it is blocked by an implicature.

METHOD

A questionnaire was administered consisting of 14 items, three instances of the critical 'at most two'/ 'at most three' case and one instance of each of the other 11 conditions used in experiment 4. Each item consisted of an utterance patterned after (53), for which participants were asked whether the speaker believes that the corresponding consequent (patterned after (54)) holds. Participants were invited to respond 'yes', 'no' or 'don't know'. 'No' and 'don't know' responses were both treated as negative with regard to the implication being investigated. A full list of materials is provided in Appendix F.

Participants

8 adult participants were recruited and responded to the questionnaire by email.

Results

23 of 24 (96%) responses to the 'at most two'/'at most three' items were 'yes', agreeing that the inference was valid. Responses in the other conditions also patterned with the logically expected outcomes.

Discussion

Under the conditions of this experiment, it does appear that the inference from 'at most two' to 'at most three' (or the inference from 'if…at most three' to 'if…at most two') goes through, contrary to the predictions we might expect to draw from the modal hypothesis. This is, however, consistent with the classical account.

This outcome is also consistent with the pragmatically augmented classical account in which 'at most' is considered disjunctive ('less than or equal to') because the implicature arising from this is clearly different in a conditional environment to that arising in a declarative

environment. For instance, the utterance (55) does not give rise to the pragmatic interpretation (56).

(55)   If Berta has had at most three drinks, she is fit to drive.

(56)   If it is possible that Berta has had at most three drinks and certain that she has had no more than three drinks, then she is fit to drive.

Why not? Because (56) is a weakening of the original semantics of the utterance, namely (57).

(57)   If it is certain that Berta has had no more than three drinks, she is fit to drive.

Therefore, deriving the implicature gives the hearer no additional information. Even if the hearer derives the implicature, this will not supersede the existing semantic content but stands alongside it. Therefore, the pragmatic interpretation of the utterance remains (57), rather than the weaker (56). Rather than triggering a pragmatic enrichment, it seems intuitive to propose that the signal sent by using the 'at most' formulation in this case is to draw attention to the upper bound (see Nouwen 2010 for a semantically oriented account along similar lines).

To verify this observation, we can check the parallelism with the paradigmatic case of a possibility implicature arising from disjunction. The declarative (58) appears to convey the implicature that either (beef or pork) is possible; by contrast, this is not available in a conditional context such as (59).

(58)   There is beef or pork on the menu.

(59)   If there is beef or pork on the menu, Max will be happy.

This is presumably because the pragmatic 'enrichment' of possibility, when applied to the latter utterance, in fact makes it less informative, in that it imposes an extra condition (the possibility of each conjunct) that would have to be satisfied before the conclusion could be drawn. Hence, this implicature does not go through under the scope of a conditional.

It might be argued that the materials used in this experiment are slightly awkward from a pragmatic standpoint. A more naturalistic formulation would be along the lines of (60).

(60)   Anyone who has had at most three drinks is fit to drive. Berta has had at most two drinks.

However, in this case, it might conceivably be argued that the inferential process is not clear, as it requires an additional step. The additional step might either be (61) or (62).

(61)   Berta has had at most three drinks.

(62)   Anyone who has had at most two drinks is fit to drive.

I do not believe that anything hinges upon this distinction, but elected to use the methodology described above in order to allay any concerns arising from this uncertainty.

The result of this experiment is not entirely surprising in the light of Geurts and Nouwen's (2007) disclaimer about the legitimacy of the modal interpretation of superlative quantifiers. The results can be explained by assuming that the superlative quantifier under the scope of the conditional has a purely classical meaning: it is not necessary to assume further that the superlative quantifier in the antecedent ('Berta has had at most two drinks') also lacks modality.

However, there does not appear to be any specific proposal to account for the absence of this type of semantic modal meaning within the conditional environment, apart from the classical view, which denies the existence of the modal meaning *in toto*. It is not impossible to condition on modals, although it is rarely necessary and examples of it are consequently somewhat contrived, as in (63).

(63)   If I think the project *might* succeed, I'll fund it for a trial period only.

For this reason, I consider that the outcome of this experiment lends support to a pragmatically oriented classical account, in which superlative quantifiers are analysed as disjunctions, over the Geurts and Nouwen (2007) account in which these quantifiers possess modal semantics. These results also motivate us to consider whether the same kind of inference might be available in non-conditional contexts, for which Geurts and Nouwen do not consider that a problem arises in their account. This is explored in the following task.

4.8.4. Experiment 7 – Judgements of logical inference patterns in felicitous contexts

Experiment 4 demonstrated the unavailability of certain classical correct inference patterns involving superlative quantifiers. Under the semantically modal account, this arises for semantic reasons. Under the pragmatically augmented classical account, I have argued that

97

this stems from the implicatures that arise from consequents containing superlative quantifiers.

In experiment 7, I attempt to differentiate the predictions of the two accounts by embedding experiment 4's reasoning task into a theoretically more complex scenario in which the superlative quantifiers are licensed by the context. For example, I ask participants to judge whether (64) implies (65).

(64)   Anne has three children but Brian has at most two children.

(65)   Anne and Brian each have at most three children.

Assuming that (65) may correctly be analysed as predicating 'having at most three children' of the two individuals Anne and Brian, I argue that acceptance of this implication requires acceptance that 'at most two' implies 'at most three' (in addition to acceptance that 'three' implies 'at most three'). Recall that on the semantically modal account this is not a legitimate inference, so (64) should not be judged to imply (65). On the pragmatic account, this inference is legitimate, so the inference from (64) to (65) should go through (unless it is blocked by implicature or other considerations).

METHOD

The methodology of Experiment 4 (section 4.8.1) was replicated using 32 pairs of sentences patterned after (64) and (65), using a range of numerical quantifiers. The numerals and other aspects of the sentential content were varied between conditions. A full set of materials is provided in Appendix G.

Participants

20 participants were recruited, all members of the University of Cambridge, aged between 20 and 36 years. 14 were female.

Results

Acceptance rates for the test conditions (for which multiple items were tested) were as shown in Table 16. In addition, results for the filler items were in accordance with semantic expectations.

Table 16: Results of Experiment 7

1st sentence	2nd sentence	Acceptance (%)
n…at least $n+1$	at least $n+1$	0
n…at least $n+1$	at least n	88
n…at most n-1	at most n	68
n…at most n-1	at most n-1	17
n…fewer than n	fewer than n	3
n…fewer than n	fewer than $n+1$	100
n…more than n	more than n	3
n…more than n	more than n-1	95

Across all semantically uncontroversial conditions (i.e. those that are not theory-critical for either approach under discussion), participants were correct on 488 out of 500 items (97.6%). Crucially, the acceptance rate for the theory-critical implication 'n…at most n-1' $\implies$ 'at most n' items was significantly above 50% (41/60, p < 0.01 binomial). The implication 'n…at least $n+1$' $\implies$ 'at least n' is accepted at near-ceiling rates.

Discussion

These findings contradict the prediction of the semantically modal account of superlative quantifier meaning. According to the modal view, as expounded in the interpretation of experiment 1 of Geurts et al. (2010), the implication 'at most n-1' $\implies$ 'at most n' is unavailable. By contrast, the results of this experiment indicate that this implication is available in certain declarative contexts, albeit less reliably than the corresponding implication with comparative quantifiers. Similarly, the implication 'at least $n+1$' $\implies$ 'at least n' is predicted to be unavailable under the modal account, but this is widely accepted in this experiment.

Contrastingly, these data are compatible with the pragmatic alternative proposal in which superlative quantifiers can be analysed as disjunctions and give rise to possible implicatures. According to this proposal, 'at most three' serves the function of drawing particular attention to the possibility that 'exactly three' holds and has the potential to give rise to an implicature to that effect. However, in this experimental situation, the implicature is not entirely inappropriate because the possibility of equality holds for one of the conjuncts. If we

consider (64) and (65), we can think of (65) as giving rise to an implicature that it is possible that either Anne or Brian has exactly three children, which is true according to (64), as Anne does have exactly three children. By contrast, according to the semantically modal account, this possibility of equality is part of the semantics of 'at most three' and therefore should be required of both conjuncts. On this account, (65) states that it is possible that Anne has exactly three children and that it is possible that Brian has exactly three children. This is false according to (64), which specifies that Brian cannot possibly have exactly three children.

The results of this experiment suggest that the majority of participants were able to reason from 'at most two' to 'at most three' when the implicatures arising from the use of 'at most' could be accommodated without a contradiction arising. By contrast, on the semantically modal account, this should not be possible at all, as the modality renders 'at most three' inherently contradictory of 'at most two'.

Is there a way to reconcile these data with the modal hypothesis? To do this, we would have to accept that the modal semantics of 'at most three' does not apply to the individual conjuncts of (65), repeated below.

(65)   Anne and Brian each have at most three children.

Instead, we would have to hold that the modal possibility applies only to the set {Anne, Brian}, and is not binding upon the other. This amounts to saying that (17) is true of a situation in which Anne might have three children, even if Brian cannot possibly have that many.

While this analysis is theoretically viable, it is intuitively unsatisfactory. If this kind of accommodation of the modal semantics is available to speakers, they should also attest the truth of statements such as (66a), or under the right circumstances even (66b), (66c) and similar.

(66a)   Anne and Brian each might have three children.
(66b)   Anne and Brian each might be a young woman.
(66c)   Anne and Brian each might be the only woman named Anne in the village.

As these are apparently false statements, I submit that (65) does not admit this analysis and therefore that experiment 7 gives evidence against the semantically modal account and in favour of a pragmatically augmented classical account.

4.8.5. General discussion of experimental data

Geurts et al. (2010) demonstrate empirical support for the hypothesis that the superlative quantifiers possess a modal component to their semantics, as proposed by Geurts and Nouwen (2007). Through the above experiments, I have corroborated the finding that superlative quantifiers convey modal meanings. I have further proposed that the locus of this modality is in fact the pragmatics of superlative quantifiers, and that the differences between superlative and comparative quantifiers can be attributed to implicature, which in turn can be attributed to the psychological complexity of non-strict comparison. This is empirically supported by the findings from experiment 3. I argued that the classical account, augmented with this notion of complexity, is compatible with the empirical data gathered by Geurts et al. (2010). Experiments 4-7 then serve to suggest that this pragmatic proposal is in fact more satisfactory than the semantically modal account.

Again I should mention that Geurts and Nouwen (2007) are aware of the theory's limitations in its present form, noting in particular the difficulties posed by the conditional context. In experiment 6 I provide a practical demonstration of this lacuna and show how the non-modal superlative quantifier enters into logical relations. However, experiments 5 and 7 go further. Experiment 5 shows that participants treat the 'exactly n is possible' meaning of superlative quantifiers as a pragmatic and readily revisable inference, rather than as part of the logical meaning of the expression. Experiment 7 shows that the superlative quantifiers systematically appear to lack modal semantics, even in declarative contexts, notwithstanding that they typically seem to convey modal meaning when used declaratively.

I contend that the pattern of results observed across these experiments coheres with the notion that the semantic meaning of superlative quantifiers is fundamentally the classical meaning. The 'modal' meaning can instead be analysed as a possibility implicature arising from the use of a superlative quantifier on account of its disjunctive nature. This implies that there are indeed differences in the inferences that are licensed by superlative and comparative quantifiers, but they stem from the presence or absence of this implicature. Moreover, the demonstration that the on-line performance preference for comparative over superlative

quantifiers is matched by the preference for strict over non-strict comparison coheres closely with the claim that non-strict comparison is more complex than strict comparison.

Under this analysis, assuming a disjunctive representation for superlative quantifiers, the use of 'at most n' emphasises the number n and the possibility of equality, neither of which are activated by the comparative alternative 'fewer than $n+1$'. In declarative contexts, the implicature that the superlative quantifier conveys would thus be very similar to the semantic meaning proposed by Geurts and Nouwen (2007). In downward-entailing contexts, by contrast, the implicature does not arise for standard pragmatic reasons, whereas the semantic meaning proposed by Geurts and Nouwen cannot be explained away in such a principled fashion.

This account also goes some way towards accounting for the patterns in corpus data. For instance, in the BNC, tokens of 'at least 20' vastly outnumber 'more than 19' (110 to 6), while 'more than 20' vastly outnumbers 'at least 21' (357 to 23). Similar patterns occur for other round numbers, as discussed in more detail in chapter 6. From a modal semantic perspective, this is surprising, as it appears to suggest that the use of superlative quantifiers is sometimes motivated by something other than the wish to express modality. To put it another way, the decision to use a superlative quantifier seems to be tied up with the decision to use a particular number, which is surprising if we suppose that the comparative and superlative quantifiers encode semantically highly distinctive meanings. Under the pragmatic account, this makes more sense: superlative quantifiers call attention to the numbers they take as arguments, and round numbers are particularly likely to be of interest to the hearer. By contrast, imputing modality to superlative quantifiers curtails the expressive power of the system, by constraining the choice of number unless the speaker is indifferent about whether or not they use a modal expression.

This would explain why the logical implication 'at most two' $\implies$ 'at most three' is accepted in experiment 7 while it was rejected in experiment 4. In experiment 7, 'at most' is used (in the consequent) with a salient number, as determined by the preceding utterance. Under these circumstances, the superlative quantifier in the consequent is licensed and no implicature arises. In the absence of this implicature, the superlative quantifier possessed purely classical meaning and the logically correct inference is drawn.

Such a proposal has implications for the analysis of other expressions. Consider, for instance, the distribution of 'not more than' and 'not fewer/less than'. It could be argued that

these are modal, but it seems more plausible, compositionally speaking, to count them as classically semantic comparative quantifiers. The relevant semantic difference between these and their non-negated counterparts is that while 'more/fewer than n' excludes the possibility of n, 'not more/fewer than n' admits this possibility. It is simply this difference that appears to underlie the following differences in distribution.

(67a)   *Fewer than three people have that authority, namely the Queen and the Prime Minister.

(67b) $^?$Not fewer than two people have that authority, namely the Queen and the Prime Minister.

(67c)   Not more than two people have that authority, namely the Queen and the Prime Minister.

(68a)   *Wilma danced with fewer/less than every second man who asked her.

(68b) $^?$Wilma danced with not more/less than ever second man who asked her.

In (67b) and (67c), explicit reference to 'two people' licenses the 'namely' continuation; in (68b), the possibility of equality seems to license the use of the comparative quantifier in a context in which it is otherwise understood to be forbidden. In both cases, the difference resides in classical semantic properties of the quantifiers.

In sum, there are good empirical reasons to suppose that a pragmatic account of superlative quantifier meaning is preferable to a fully semantic account. However, the above discussion leaves several issues open. The notion of what it means for an expression to draw attention to the notion of equality is left open, and there is no specific evidence for the claim that superlative quantifiers are specifically disjunctive, rather than merely being additionally complex. In the following section, I restate the complexity-driven pragmatic account of superlative quantifier meaning in terms of the constraint-based model introduced in this thesis. I show that this is also able to account for the observed data, and provides a more specific proposal as to the precise nature of the pragmatic enrichments arising from superlative quantifiers.

4.9. A constraint-based account of superlative quantifiers

The constraint-based model proposed in this thesis offers an alternative way to characterise the complexity of superlative quantifiers. Specifically, we can interpret the findings of Geurts et al.'s (2010) third experiment, and experiment 3 in this thesis (section 4.6.1), as supportive of the idea that superlative quantifiers incur an additional violation of the quantifier simplicity constraint QSIMP (defined in section 2.4.3), over and above those incurred by comparative quantifiers. This is distinctive from the proposal discussed above in that it does not suppose that superlative quantifiers are disjunctive at any level of representation, but relies purely upon the complexity data that have been demonstrated directly. We can discuss the implications of this for quantifier usage by building upon the toy example introduced in section 3.1.

Revisiting first the situation in which a speaker wishes to describe a value greater than or equal to 21, possible options are (69)-(71). Their tableau, with respect to the constraints on quantifier simplicity (QSIMP), informativeness (INFO), numeral salience (NSAL) and numeral priming (NPRI), is given as Table 17. I assume here for ease of illustration that the comparative quantifier incurs no violations of QSIMP.

(69)   More than 20
(70)   At least 20
(71)   At least 21

Table 17: OT tableau for (69)-(71), QSIMP, INFO, NSAL and NPRI

	QSIMP	INFO	NSAL	NPRI
More than 20				
At least 20	*	*		
At least 21	*		*	

In this case, (69) incurs no violations and therefore harmonically bounds the alternatives; this model predicts it will be preferred under all constraint rankings.

In the corresponding situation for which the speaker wishes to express a value greater than or equal to 22, alternatives include (72)-(76), and the tableau is given as Table 18.

(72)   More than 20

(73)   At least 20

(74)   More than 21

(75)   At least 21

(76)   At least 22

Table 18: OT tableau for (72)-(76), QSIMP, INFO, NSAL and NPRI

	QSIMP	INFO	NSAL	NPRI
More than 20		*		
At least 20	*	**		
More than 21			*	
At least 21	*	*	*	
At least 22	*		*	

Here the situation is similar to that in Table 17; (74) harmonically bounds (75) and (76), while (72) harmonically bounds (73). So the use of a comparative quantifier is predicted to be obligatory in this situation, under any constraint ranking.

However, superlative quantifiers can surface under certain conditions, in this model. Consider a situation in which the speaker wishes to express a value greater than or equal to 20. Alternatives include (77)-(79), and their tableau is given as Table 19.

(77)   More than 19

(78)   At least 19

(79)   At least 20

Table 19: OT tableau for (77)-(79), QSIMP, INFO, NSAL and NPRI

	QSIMP	INFO	NSAL	NPRI
More than 19			*	
At least 19	*	*	*	
At least 20	*			

Here (78) is harmonically bounded by both (77) and (79), but either of these could surface. The comparative (77) is preferred if QSIMP > NSAL, and the superlative (79) if NSAL > QSIMP.

Similar considerations apply if we consider the effect of numeral priming. Suppose that the speaker wishes to express a value greater than or equal to 21, where 21 is primed in the discourse. Options include (69)-(71), but this time their tableau is as shown in Table 20.

Table 20: OT tableau for (69)-(71), QSIMP, INFO, NSAL and NPRI; 21 primed

	QSIMP	INFO	NSAL	NPRI
More than 20				*
At least 20	*	*		*
At least 21	*		*	

Here there is no relation of harmonic bounding between (69) and (71). The comparative quantifier is preferred if QSIMP > NPRI or NSAL > NPRI; the superlative quantifier is preferred if NPRI > QSIMP and NPRI > NSAL.

Generalising over these and similar cases, 'at least' is harmonically bounded by 'more than', except where the underlying lower bound happens to be a round number, or where the lower bound is primed. In these exceptional cases, 'at least' is preferred under certain constraint rankings. The same applies, *mutatis mutandis*, for 'less/fewer than' versus 'at most' in the parallel upper bound case.

Consequently, this model makes predictions as to the distribution of superlative quantifiers which closely match those obtained either under the semantically modal account of Geurts and Nouwen (2007) or those under the pragmatic account advanced in this chapter. Specifically, it predicts that 'at least/most n' will be used only if n is a round number[21] and the possibility of equality holds, or if n is contextually activated.

Moreover, under the assumptions discussed earlier, we can also look at this from the hearer's perspective and obtain predictions about their interpretative preferences. We assume that the hearer is tacitly aware of the motivations for the speaker's choice. Then, upon encountering

---

[21] In this case, 'round number' includes 1-9, as per Jansen and Pollmann's (2001) formalism.

'at least/most *n*', the hearer is aware that either *n* is contextually activated or that the possibility of equality holds. If the hearer also knows that *n* is not contextually activated, they can infer that the possibility of equality holds. That is, in the absence of contextual activation, this account predicts that the use of 'at least/most *n*' will be understood to convey the possibility of equality. This argument applies irrespective of the type of selection procedure employed by the speaker.

Across the experimental conditions discussed in this chapter, this account predicts the following responses.

- In Experiment 4 (section 4.8.1), the declarative sentence 'at most three' occurs in the absence of contextual activation of 'three'. Thus it conveys the implicature that three is a possibility. This implicature is predicted to cause the inference 'at most two' $\implies$ 'at most three' to fail.

- In Experiment 5 (section 4.8.2), utterances juxtaposing 'at least *n*' and 'exactly *n*+1' are predicted to be pragmatically, but not semantically, anomalous.

- In Experiment 6 (section 4.8.3), the inference is predicted to be available, as the implicature that 'exactly *n*' is a possibility should not arise in the conditional context.

- In Experiment 7 (section 4.8.4), the use of 'at most *n*' in the consequent expression is licensed by the presence of the numeral *n* in the preceding context. As a result, this account predicts that the implicature that *n* is possible fails, as this only arises in situations where the numeral is not contextually activated. Therefore the inference that was unavailable in Experiment 4 should be available in this case.

In short, the constraint-based account makes the same predictions as the pragmatically-augmented classical account outlined earlier in this chapter. It could also be argued to provide a more coherent account of examples such as (67), where the acceptability of the sentence seems to depend on whether the numeral explicitly mentioned is contextually salient. This is accomplished without stipulating any specific internal structure for superlative quantification: 'at least' and 'at most' are not posited necessarily to be disjunctive at any level of representation. Instead, the constraint-based model comes equipped with its own mechanism for translating the difference in representational complexity into explicit pragmatic enrichments.

Thus, the constraint-based account appears to match the predictions of the pragmatic account discussed in this chapter for the case of comparative and superlative quantifiers. Compared to the pragmatic account, it also has the potential advantage of not assuming the hitherto unproven claim that the superlative quantifier has a disjunctive internal structure. Moreover, the constraint-based account generalises more broadly to the domain of quantification, rather than being restricted to the analysis of comparative and superlative quantifiers.

4.10. Summary

In this chapter, I discuss the meaning of comparative and superlative quantifiers, endorsing the view of Geurts and Nouwen (2007) that the latter possess modal meaning, but attributing this to pragmatic considerations. First I offer a pragmatic counter-proposal in which the modal meaning is attributed to implicature arising from the complex nature of non-strict comparison, as manifest in its linguistic forms such as superlative quantifiers. I support this with an empirical verification that non-strict comparison is indeed dispreferred to strict comparison in processing. Then I present a series of experimental demonstrations that the pragmatic account is preferable to the existing semantic proposal. Finally, I turn once again to the constraint-based model introduced in this thesis, and show that (given the complexity of superlative quantifiers) this generates the same predictions as the pragmatic account, but without any additional stipulations as to the structure of superlative quantification or non-strict comparison. Thus, the observations captured in this chapter concerning the relation of comparative and superlative quantifiers can be naturally and economically captured within the constraint-based model. In the following chapter, I consider a different application of the constraint-based model, in which, rather than accommodating existing experimental findings, novel predictions are drawn from this model and these predictions are tested experimentally.

5. SCALAR IMPLICATURES FROM NUMERICALLY QUANTIFIED EXPRESSIONS[22]

Certain categories of numerically quantified expressions, including comparative and superlative quantifiers, have been argued in the literature not to give rise to scalar implicatures. By contrast, other expressions apparently do give rise to pragmatic upper and lower bounds, including (on some accounts) bare numerals. Various theoretical attempts have been made to account for this distinction. In this chapter I take a different tack, first considering how numerical expressions are generally considered to admit pragmatic enrichments, and then arguing that comparative and superlative quantifiers behave similarly and do in fact give rise to pragmatic bounds. I discuss how the constraint-based model predicts that these inferences arise, and how they will be modulated by considerations of granularity or numeral salience. Then I present novel empirical data that bears out this prediction. I then derive a further prediction from the constraint-based model, concerning the effect of prior numeral activation on the strength of the scalar implicature, and verify this prediction experimentally.

5.1. Pragmatic enrichments of bare numerals

Unmodified natural numbers – 'bare numerals' – admit several distinct interpretations, which we might refer to as the exact, lower bound and upper bound readings. (80)-(82) illustrate these readings in turn.

(80)   Tom has three children.
(81)   We need to sell ten tickets to make a profit.
(82)   (Carston 1998) She can have 2000 calories without putting on weight.

This raises the question of whether numerals are semantically ambiguous between these meanings, or whether the semantic meaning is of one of these types and the other meanings are derived pragmatically. In the latter case, we would wish to establish how these meanings are obtained by hearers.

For instance, if we assume that the semantic meaning of n is 'exactly n', obtaining the reading of examples like (81) is reasonably straightforward. Specifically, we can appeal to the process of existential closure, as described by Geurts (2006), after Partee (1986).

---

[22] This chapter incorporates work presented in Cummins, Sauerland and Solt (submitted).

Considering an example such as (80), we can characterise its truth-conditions as those of (83).

(83)   There exist three distinct individuals such that each individual belongs to the set {Tom's children}.

From this it follows that the cardinality of the entire set {Tom's children} is greater than or equal to 3. In this way we derive the 'at least' reading: in order for (80) to be true, it must be the case that the whole set {Tom's children} has cardinality at least 3.

We cannot account for the 'at most' reading of n by a similar argument. However, this may not be a serious problem, as examples of this reading are somewhat tenuous. Considering (82), I would argue that the numeral itself ('2000') has an 'at least' reading, contrary to our immediate intuitions. Indeed, if the numeral possessed a strict 'at most' reading, (82) would imply (84).

(84)   If she has 2001 calories, she will put on weight.

This violates our intuitions: (82) appears to commit to a view that a particular number, 2000, falls within the safe range for calorie intake without weight gain. We might infer that the speaker, being cooperative, would have quoted a larger number if it were possible, as in this case a larger number happens to be more permissive and therefore potentially useful. This appears to be the mechanism by which we derive the 'at most' meaning. But this is clearly pragmatic, and relies on encyclopaedic knowledge as to the likely effect of an increased calorie intake. Thus, I follow Geurts (2006) in considering the 'at most' reading to be readily explicable in terms of pragmatic inference, broadly construed, and therefore will not consider it further at this point.

Another approach discussed in the literature takes the semantics of n to be 'at least n'. In this case, the question remains of how the 'exactly n' reading can be obtained. Levinson (1983: 115), following Horn (1972), argues that the precise reading of numerals is a 'straightforward Quantity implicature' that is 'easily defeasible'. Levinson notes that 'John has three cows' entails 'John has two cows', and claims that it implicates that 'John has exactly three cows'. He then observes that an if-clause suspends the implicature but cannot suspend the entailment, as shown by comparing (85a) and (85b).

(85a)   John has three cows, if not more.

(85b)   ?John has three cows, if not two.

Levinson also argues that 'implicatures are directly and overtly deniable without a sense of contradiction', citing the example (85c), although such examples do appear to be self-contradictory if not furnished with the appropriate context.

(85c)   John has three cows, in fact ten

Horn (1985: 139) further elucidates the nature of this implicature. He places numerals in a class of scalar operators (along with *some*, *possible* and *like*), which he takes to be 'lower-bounded by their truth-conditional semantics'[23] and which he says 'may be upper-bounded (context permitting) by conversational implicature, triggered by Grice's maxim of Quantity'. For Horn, the negation in (86) 'does not negate the PROPOSITION that Max has three children; rather, it operates on a metalinguistic level to reject the IMPLICATUM that may be associated with the assertion of that proposition' (his caps).

(86)   Max doesn't have three children – (*but) he has four

Horn (1985: 146) supports this argument with a useful observation about the class of examples that we use to discuss these questions of meaning. He admits that some speakers interpret the above example as equivalent to (87).

(87)   It's not true that Max has THREE children – he has FOUR.

This appears to argue against the claim that the proposition itself is not being negated. Still, Horn argues that 'the distribution of the English expressions *It is true that*...etc....is a poor guide at best as to where the LOGICAL predicate TRUE is to be applied in the simplest, most elegant semantic/pragmatic theory...We often say that something isn't true, meaning that it isn't assertable' (ibid.). However, Horn's theory of metalinguistic negation in general has been extensively criticised, and Horn himself appears to have retreated from his commitment to it (Geurts 2009: 73f).

In summary, then, semantic and pragmatic analyses of bare numerals disagree as to whether the exact or the lower-bound meaning is taken to be the core semantic meaning. However,

---

[23] In the language of the preceding section, Horn takes the view that only the existential reading is truth-conditional.

they concur that numerical expressions of this type admit pragmatic enrichments. By contrast, other categories of numerically quantified expression, notably comparative and superlative quantifiers, have been argued not to yield implicatures. In the following section, I briefly review the literature on this topic, before proceeding to offer evidence that this generalisation is incorrect and pragmatic enrichments are available, albeit constrained in principled ways.

5.2. Failure of implicature for comparative and superlative quantifiers

Krifka (1999) observes that superlative quantifiers seem systematically to fail to give rise to the pragmatically expected scalar implicature. Specifically, he observes that (88) fails to yield the implicature that negates the stronger statement (89). If this implicature went through, (88) would be pragmatically understood as stating that John has exactly three children, which is clearly intuitively incorrect (and uniformly rejected by experimental participants, as shown by Geurts et al. (2010: 138)).

(88)   John has at least three children.

(89)   John has at least four children.

In attempting to account for this pattern, Krifka considers the possibility that modified numerals of the type 'at least *n*' do not participate in Horn scales, the ordered sets of propositions within which scalar implicatures are generated (Horn 1972), while bare numerals do. Although this solves the problem, it appears to be an arbitrary and unprincipled distinction: the requirements for a Horn scale are that the terms are equally lexicalised, from the same semantic field, and in the same register. This appears to apply just as much to the scale <*at least three*, *at least four*> as to the scale <*three*, *four*>. Krifka (1999: 260) goes further, asserting that '[i]f number words form Horn scales, they should do so in any context in which they appear'.

Krifka (1999: 260) also discusses the idea that 'at least' signals the speaker's unwillingness or inability to give a precise answer. This precludes a scalar implicature from arising, as the epistemic condition on the speaker is not met. Krifka's suggestion is that the notion of the speaker's uncertainty, or reticence, is pragmatically derived from the choice of 'at least *n*' rather than the bare numeral *n*, because the latter would carry the implicature of certainty ('exactly *n*').

This analysis, however, does not appear to generalise to the equally problematic case of 'more than'. Fox and Hackl (2006: 540) assert that 'more than *n*' also systematically fails to admit scalar implicature, an observation they also attribute to Krifka (1999), even though Krifka does not appear to address this case. The relevant observation is that, for example, uttering (90) does not trigger the scalar implicature that (91) does not hold.

(90)   John has more than three children.

(91)   John has more than four children.

Nevertheless, it is intuitively plausible that an informed and cooperative speaker could say something like (90), for example in the context of establishing whether John is eligible for certain benefits, or whether he needs a bigger car, etc. Under such circumstances, the implicature that John has exactly four children does indeed seem not to arise.

Moreover, if we turn our attention to larger numbers, we can observe clear examples of implicatures failing to arise from expressions using 'more than' and 'fewer than'. Consider, for instance, (92) and (93), which might be naturally uttered by an informed speaker but which do not convey the corresponding implicatures (94) and (95).

(92)   More than 100 people got married today.

(93)   Fewer than 20 people have ever walked on the Moon.

(94)   The speaker considers it possible that exactly 101 people got married today.

(95)   The speaker considers it possible that exactly 19 people have walked on the Moon.

The need to account for this apparently anomalous behaviour of modified numerals has been a partial motivation for various semantic accounts of numerical quantifiers. For Krifka (1999), this and other factors motivate a rejection of the generalised quantifier account of Barwise and Cooper (1981) for expressions such as 'at least *n*'. Krifka's account of the way in which scales are built up in such cases feeds into Geurts and Nouwen's (2007) proposal concerning the modal semantics of superlative quantifiers, as discussed in chapter 4. Fox and Hackl (2006) use the absence of implicature as a motivation for their proposal concerning the

Universal Density of Measurements, in which they argue that measurement scales for natural language semantics are necessarily always dense[24].

However, despite the absence of implicatures from expressions such as (92) and (93), there do appear to be some pragmatic enrichments available from utterances using 'more/fewer than *n*', and perhaps 'at least/most *n*', that have not been satisfactorily accounted for. Consider an utterance such as (94).

(94)   London has more than 1000 inhabitants.

This utterance appears to be wilfully misleading, unless it is made in response to an utterance such as 'Give an example of a settlement with more than 1000 inhabitants'. The misleading nature of this utterance intuitively seems to stem from the fact that the description 'more than 1000' conveys a quantity that is appreciably less than the actual population of London. However, semantically, (94) is true. Furthermore, the precise nature of the inference is very difficult to determine introspectively. (94) certainly does not seem to convey the classical scalar implicature (95a), and in this sense it patterns with (88), (90), (92) and (93). However, it does seem to convey an implicature like (95b), if not a stronger one.

(95a)   London has exactly 1001 inhabitants.

(95b)   London has fewer than a million inhabitants.

In this chapter, I will argue that 'more than *n*' and 'fewer than *n*' do in fact give rise to scalar implicatures, but that these are restricted by granularity considerations. However, before turning to the empirical validation of this claim, I consider how these implicatures are predicted to arise in terms of the constraint-based model introduced in this thesis.

5.3. Implicatures predicted by the constraint-based account

The failure of (94) to give rise to the implicature (95a) can be accounted for by considerations of granularity, as defined in section 2.4.2 of this thesis, and numeral salience, as defined in section 2.4.4. Recall that the implicature relies upon the speaker declining to make the stronger statement (96), despite being in a position to do so.

---

[24] This could perhaps be regarded as the opposite of the granularity-based proposal discussed here, in that Fox and Hackl's argument assigns a role to non-integer values in accounting for the infelicity of cardinal statements such as 'John only has more than 3 children', while granularity considerations force our analysis to remain within a subset of integer values when discussing cardinalities.

(96)   London has more than 1001 inhabitants.

However, according to the constraint-based account, there are perfectly good reasons for the speaker to avoid saying (96), even if they know this to be true. 1001 is clearly not a salient numeral, and its use is thus predicted to be disfavoured. Consequently, (96) incurs an additional violation of the numeral salience constraint (NSAL). Another way of looking at this is that the numerals 1000 and 1001 share only the finest granularity level, and therefore 1001 is likely to violate the granularity constraint in circumstances where 1000 is appropriate.

For these reasons, the constraint-based proposal predicts that a speaker who could say (96) will say (94) instead. From the hearer's perspective, then, the utterance of (94) cannot reliably convey the strong implicature that (96) does not hold. Hence, under this account, we do not expect (94) to give rise to the implicature (95a).

If we consider a situation where two utterances are matched in granularity and numeral salience, then the constraint-based model predicts that an implicature of this kind can arise. For instance, we can compare (97) and (98) in this respect.

(97)   More than 70 people got married today.

(98)   More than 80 people got married today.

On the basis of numeral salience, there is no reason to prefer (97) to (98). Likewise, from a granularity perspective, these are matched, so (98) will not incur any additional violations to those incurred by (97). Moreover, (98) is more informative than (97). With respect to all these considerations, (98) is preferred, and therefore the speaker's decision to utter (97) rather than (98) is predicted to have pragmatic consequences: that is, assuming they are knowledgeable, to implicate the falsity of the stronger statement.

The single constraint that could be violated by (98) and not by (97) is that of numeral priming (NPRI), as introduced in section 2.4.5. If the numeral 70 is highly activated in the preceding context, then a speaker who ranks NPRI highly may prefer (97) over (98) on this basis. More generally, in a non-classical OT setting such as those discussed in section 3.2, sufficiently strong activation of the preceding numeral, coupled with a moderate weighting to the NPRI constraint, would achieve the same effect.

From the hearer's perspective, then, the use of (97) can be attributed to either contextual activation of the numeral 70 (coupled with certain assumptions about the speaker's constraint

ranking) or to the speaker's inability to utter (98). If the hearer is aware that the numeral is not activated in the context – or, to be more precise, not aware that the numeral is activated – the model predicts that they infer that the speaker could not utter (98). Under the usual conditions on the speaker's epistemic state and cooperativeness, the hearer is then entitled to derive the strong scalar implicature that (98) does not hold.

However, if the hearer knows that the numeral is activated, the above line of reasoning cannot be undertaken with certainty. In this case, the speaker's choice might reflect adherence to the numeral priming constraint rather than an inability to make the stronger statement. Now, even if the numeral is contextually activated, it is still quite possible within this model for the speaker to make the stronger statement, if their constraint ranking permits. Informally speaking, this would occur when the benefits of making the stronger statement outweigh the costs of failing to use the activated numeral. Therefore, the decision to make the weaker statement still has some potential pragmatic significance. Even though the hearer cannot be as confident in the implicature as they were in the no-activation case, nevertheless it is predicted that the implicature can still be generated, but that it is attenuated in some way.

In what way might this implicature be attenuated? Well, it might, for instance, give rise to a more distant upper bound. Consider (97) and (98) versus (92). These might give rise to a tableau which we can schematise as shown in Table 21.

Table 21: OT tableau for (92), (97), (98); INFO, NSAL, NPRI; 70 contextually activated

	INFO	NSAL	NPRI
More than 100			*
More than 70	**	*	
More than 80	*	*	*

Suppose a speaker has a constraint ranking NSAL > NPRI > INFO. Under this ranking, (97) will be preferred to (98) where both are true, as numeral priming outranks informativeness. That is, the speaker prefers to describe situations in which 'more than 80' is true and 70 is contextually activated by saying 'more than 70', rather than 'more than 80'. However, (92) will be preferred to both, when all three are true, due to 100's salience. That is, the same speaker prefers to describe situations in which 'more than 100' is true and 70 is contextually activated by saying 'more than 100', rather than 'more than 70' or 'more than 80'. For a

116

speaker with this constraint ranking, the utterance of 'more than 70' would not implicate that 'more than 80' did not hold, but it would implicate that 'more than 100' did not. By contrast, if 70 is not contextually activated, then the same speaker is predicted to prefer 'more than 80' to 'more than 70' for situations in which both hold and 'more than 100' does not. Therefore, in the no-activation case, this hypothetical speaker's utterance of 'more than 70' implicates 'not more than 80', but in the activation case it merely implicates the weaker 'not more than 100'.

This exemplifies the claim that implicatures arise from what the speaker chooses not to say, and the extent to which the speaker had a choice is manifest in the strength of the resulting implicature. Just as the classical scalar implicature requires the hearer to make assumptions about the speaker's epistemic state, so this type of inference requires the hearer to make assumptions about the speaker's constraint ranking – i.e. the strategy the speaker uses to select numerically quantified statements.

In summary, we can derive two novel predictions about the implicatures from comparative (and, *mutatis mutandis*, superlative) quantifiers from the constraint-based account. First, we predict that these expressions do give rise to scalar implicatures, at the appropriate granularity level. Under the standard conditions of speaker informativeness and cooperativity, and in simple declarative contexts, 'more than n' gives rise to the implicature that 'more than m' does not hold, where m is any numeral such that $m > n$ and the coarsest granularity level expressed by m is at least as coarse as than that expressed by n. Secondly, we predict that these resulting scalar implicatures are attenuated by prior mention of the numeral. In the following section I present empirical verifications of these predictions.

5.4. Experimental tests of scalar implicatures from comparative and superlative quantifiers

As discussed above, the constraint-based model predicts that comparative and superlative quantifiers will give rise to implicatures that limit their range of interpretation. Note that the stronger the implicature, the more limited the range of interpretation: that is, if a quantifier gives rise to a strong implicature of this type, it admits only a narrow range of interpretation. The following experiments use this approach to ascertain the extent to which implicatures of this type arise.

117

5.4.1. Experiment 8 – Range of interpretation of comparative and superlative quantifiers

Experiment 8 was designed to test the prediction that numerical quantifiers presented without a preceding context convey scalar implicatures that are restricted by granularity. In this experiment, participants were presented with a numerically quantified expression and asked to specify either the range of values they felt the expression conveyed, or the most likely single value given the expression that was used. The roundness of the numeral and quantifier type were manipulated, as discussed below.

METHOD

The experiment was conducted online using the Amazon Mechanical Turk (MTurk) platform, the use of which for linguistic experiments is discussed by Sprouse (2011). The experimental materials were uploaded and participants self-selected for participation, and received a small financial incentive for doing so. In the experiment, participants were shown the following stimulus, consisting of a statement including a modified numeral, and were asked to provide estimates of the number(s) in question, as shown below.

Information: A newspaper reported the following.

'[Numerical expression] people attended the public meeting about the new highway construction project.'

Question: Based on reading this, how many people do you think attended the meeting?

Between ____ and ____ people attended [range condition]
____ people attended [single number condition]

Participants were also given an opportunity to write a comment explaining why they answered the way they did.

These materials were used across 12 conditions, constituting a 2x3x2 design in which the following parameters were crossed.

- Type of numerical expression ('more than *n*', 'at least *n*')

- Roundness level (hundreds (*n*=100), tens (*n*=110), units (*n*=93))

- Condition (range, single number)

The 12 conditions were fielded online over the course of roughly 20 days in December 2009 and January 2010, each condition being fielded separately in order to reduce the likelihood of individual participants completing multiple versions of the task.

Participants

A total of 1200 participants were recruited (100 per condition, as described below). The only inclusion criterion was an acceptance rate of over 95% on prior MTurk tasks. The subjects remained anonymous but reported basic demographic facts: the gender split was 51% female to 49% male. Language background was not tested. Subjects were paid $0.02 for participation.

Results

Prior to analysis, the following categories of anomalous responses were removed.

- Responses not consisting of a single numeral, including non-numerical responses (e.g. 'many' or 'infinity') and those expressing ranges ('more than 110').

- Responses more than one order of magnitude greater than *n*.

- Responses inconsistent with the truth conditions of the original statement. Participants' comments suggested that, in some cases, they had not assumed truthfulness of the original statement (e.g. 'newspapers usually like to exaggerate'), and therefore their interpretation could not be considered to relate to the quantified expression under test.

The number of each category of responses removed was as shown in Table 22. Note that very few responses were outliers, as defined above. This is crucial as the exclusion of these responses could otherwise exert bias on the data, by leaving in responses that clearly express implicatures but removing those that do not.

Table 22: Responses removed from analysis of experiment 8

	Range condition				Single number condition			
	Not single numeral	Outlier	Inconsistent with truth conditions	**Total**	Not single numeral	Outlier	Inconsistent with truth conditions	**Total**
More than n								
n = 100	3	0	15	**18**	6	1	6	**13**
n = 110	2	0	16	**18**	5	0	9	**14**
n = 93	3	1	28	**32**	3	1	8	**12**
At least n								
n = 100	4	2	25	**31**	0	0	11	**11**
n = 110	2	2	39	**43**	0	1	22	**23**
n = 93	2	1	35	**38**	1	0	15	**16**

For each condition, mean, SD and median values were calculated from the cleaned data. To facilitate comparison between conditions, mean and median values are restated in terms of *n* in what follows (e.g. a response of 140 in the 'more than 100' condition represents *n*+40). Full results are presented in Table 23. Figure 1 shows the difference between median estimates and *n*.

Table 23: Results of Experiment 8

		Range Condition		Single Number
		Low	High	Condition
More than *n*				
n = 100	Mean	104.7	219.0	118.9
	(SD)	(14.0)	(229.0)	(44.2)
	Median	101	150	110
n = 110		110.8	167.8	122.4
		(1.9)	(124.0)	(30.9)
		110	130	113
n = 93		93.9	127.5	95.9
		(1.6)	(99.1)	(6.4)
		94	100	94
At least *n*				
n = 100		100.4	153.2	106.0
		(1.7)	(108.4)	(8.2)
		100	126	103

	Range Condition		Single Number Condition
	Low	High	
n = 110	110.4	158.4	115.4
	(1.6)	(94.4)	(9.7)
	110	135	110
n = 93	93.4	115.7	95.9
	(1.5)	(28.1)	(4.2)
	93	100	94

Figure 1: Results of Experiment 8 (medians)
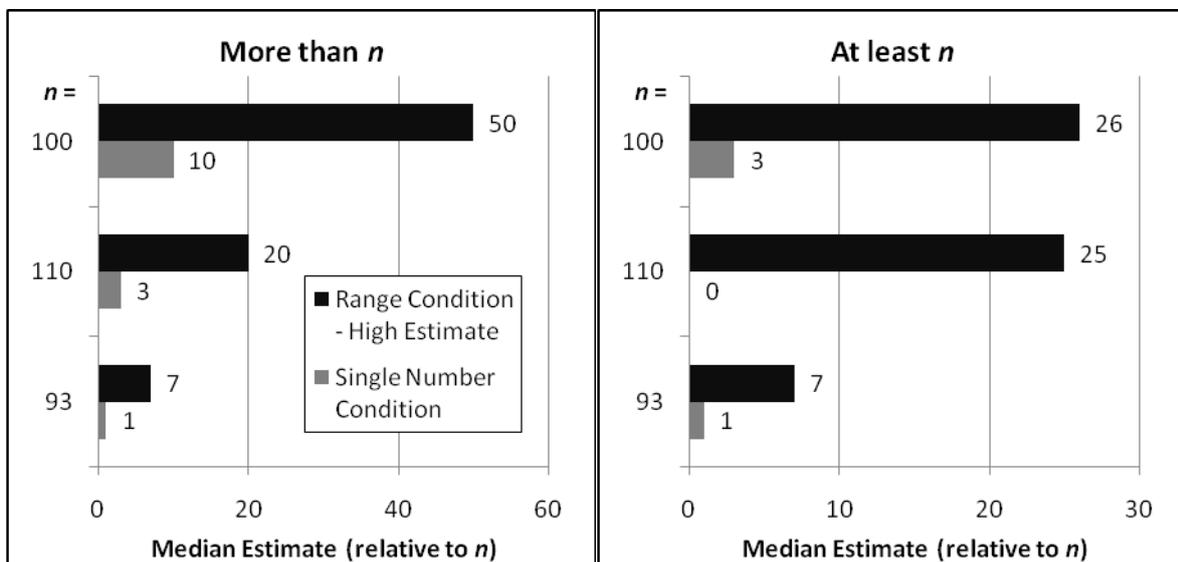
With regard to these numerical distances, ANOVAs show significant effects of granularity level. For the quantifier 'more than n', in the range condition, granularity effects are significant in determining the upper end of the range estimated ($F=5.44$, $p<0.01$). Post hoc testing (Tukey) shows a significant difference between the upper bounds for $n=100$ and $n=93$ ($p < 0.01$) and between those for $n=100$ and $n=110$ ($p < 0.05$). An overall granularity effect is also found in the single number condition ($F=5.81$, $p < 0.01$), with post hoc testing showing a significant difference between the preferred values for $n=100$ and $n=93$ ($p < 0.01$). For the quantifier 'at least n', in the range condition, there is a trend towards granularity effects ($F=2.34$, $p=0.099$). In the single number condition there is a significant effect of granularity level ($F=3.99$, $p < 0.05$), with post hoc testing again showing a significant difference between the preferred values for $n=100$ and $n=93$ ($p < 0.05$).

Discussion

The results of Experiment 8 demonstrate that effects connected with scale granularity or numeral salience, and thus relevant to the corresponding constraints proposed in sections 2.4.2 and 2.4.4 respectively, play a role in the interpretation of numerically quantified expressions of the form 'more than *n*' and 'at least *n*'. Rounder values of *n* yield wider ranges of interpretation. To be specific, looking at the median responses, we see that the range of interpretation for 'more than 100' typically extends to 150, while that for 'more than 110' typically extends to 130 (lower than the upper limit for 'more than 100', despite the numeral being larger), and that for 'more than 93' typically does not extend past 100. Similar patterns are observed in the case of 'at least *n*', although in this case the differences are less pronounced.

These patterns are suggestive of scalar implicatures conditioned by granularity. The majority of participants gave upper bounds of interpretation of order of magnitude *n*, as shown by the very small number of outliers (Table 22). This is not predictable on semantic grounds, and indicates that some form of pragmatic enrichment is available from our prompt. A typical response in each case is that 'more than 100' conveys 'not more than 150', 'more than 110' conveys 'not more than 120' or 'not more than 150', and 'more than 93' conveys 'not more than 100'. Such responses appear to be reliant on the type of inference posited earlier in this chapter: the speaker's decision not to refer to a higher value, on a scale of the same or a coarser granularity level, is interpreted as an indication that (as far as the speaker is concerned) the corresponding statement with this higher value does not hold. That is, when presented with 'more than *n*', the hearer computes the implicature that 'not more than *m*', where *m* is the next highest point on the appropriate scale.

Even more direct evidence in support of this claim comes from the responses of some of the participants to the optional supplementary question, where participants were asked to explain their responses. In the 'more than 100' condition, responses included 'I feel that if there was more than 150, the newspaper would say more than 150', 'I chose the above number because I felt had the numbers been higher the paper would have said more than 200', and 'I think 125 would be the next increment worthy of mentioning'. These responses all reflect awareness of granularity considerations. Moreover, they appear to reflect different interpretations as to the level of granularity that the numeral 100 in the prompt was understood to convey.

The results for 'at least *n*' tend to follow the pattern of 'more than *n*'. The primary difference seems to be that the estimates are typically higher for 'more than *n*': that is, in our terms, 'at least *n*' gives rise to tighter upper bounds. This difference is potentially due to the distinct pragmatic consequences of using 'at least *n*', discussed at length in the preceding chapter. Broadly, the use of 'at least *n*' in preference to 'more than *n*' seems to implicate the possibility that 'exactly *n*' might be the case, and it seems plausible that this would serve to lower hearers' expectations as to the possible range of values that *n* might take. However, as 'at least *n*' does appear to conform to the predicted pattern of responses in general, I set this issue aside in the remainder of this chapter and concentrate on the more straightforward case of comparative quantifiers.

In the case of 'more than 93', we might further ask why participants do not draw the inference that 'exactly 94' is the case, given that 94 is the natural bound in terms of granularity (the next highest numeral at the unit granularity level). Clearly, as Fox and Hackl (2006) pointed out, this interpretation would be futile as it would make 'more than 93' practically synonymous with '94' with respect to cardinalities. It is evident, then, that behaviour at the finest level of granularity is at variance with the initial hypothesis proposed in this chapter. This is an issue that merits further consideration, especially given that the canonical examples in the literature that are used to argue that comparative quantifiers lack scalar implicatures ('John has more than 3 children', etc.) frequently exhibit this fine level of granularity. I return to this topic in the general discussion, where I consider the circumstances under which quantifiers such as 'more than 93' may felicitously be employed.

However, broadly, Experiment 8 supports the main hypothesis of this chapter. According to these findings, scalar implicatures are available from expressions such as 'more than *n*', for which they furnish pragmatic upper bounds of interpretation; and these implicatures are constrained by considerations of granularity or numeral salience. In this respect, the findings from this experiment support the predictions of the constraint-based model.

5.4.2. Experiment 9 – Attenuation of pragmatic bounds through numeral priming

In Experiment 9, I test the hypothesis that the implicatures documented in Experiment 8 will be attenuated if the numeral is contextually activated. Also, given that Experiment 8 used a methodology that precludes close control of the participants' linguistic and cognitive abilities, this experiment aims to replicate those findings in a more conventional experimental setting. In this experiment, quantified expressions are presented to participants in two contexts, one in

which the numeral is previously mentioned and one in which it is not, and participants are asked to interpret these expressions. In this way we determine whether pragmatic bounds are derived, and whether prior mention of the numeral influences the strength of these bounds, as predicted by the constraint-based model. This experiment also extends the domain of enquiry by exploring the opposite entailment direction ('fewer than n').

METHOD

Participants completed a questionnaire consisting of 16 items in a laboratory setting. Each item consisted of a short dialogue transcript, the second utterance of which contained a numerically quantified expression. The first utterance of the dialogue either mentioned the numeral present in the second utterance (primed condition) or did not (unprimed condition). The precise instructions and an example item from the questionnaire, in both conditions, are shown below.

> Please read the following short dialogues, and answer the questions by filling in a value for each blank space, according to your opinion. Consider each dialogue separately. Assume that participant B is well-informed, telling the truth, and being co-operative in each case.

> Primed
> A: We need to sell 60 tickets to cover our costs. How are the ticket sales going?
> B: So far, we've sold fewer than 60 tickets.

> How many tickets have been sold? From …… to ……, most likely …….

> Unprimed
> A: We need to sell tickets to cover our costs. How are the ticket sales going?
> B: So far, we've sold fewer than 60 tickets.

> How many tickets have been sold? From …… to ……, most likely …….

12 test items were used, constituting a 2x3x2 design in which the following parameters were crossed.

- Type of numerical expression ('more than n', 'fewer than n')

- Roundness level (hundreds (n=100, 200), tens (n=60, 80), units (n=77, 93))

- Condition (primed, unprimed)

An additional 4 filler items used the quantifier 'about n', for a total of 16 items. Each item was constructed in such a way as to make the primed and unprimed versions identical but for the presence or absence of the numeral.

Two versions of the questionnaire were constructed, each containing the 16 items, half in the unprimed condition and half in the primed condition, such that each numerical expression and roundness level was represented once in each condition. The versions of the questionnaire differed in each item's condition: every item primed in version 1 was unprimed in version 2, and vice versa. Each participant therefore responded to both quantifiers at every roundness level in both conditions. Full materials are provided in Appendix H.

Participants

45 native adult English-speaking participants were recruited and randomly allocated to one of the versions of the task described below. Their average age was 21.3 years (SD 5.5 years, range 17-46). 27 were female.

Results

Data from three participants were removed from the analysis due to missing values. Participants' mean 'most likely' estimates, expressed in terms of distance from n, are as shown in Table 24, and presented graphically in Figure 2.

Table 24: Results of Experiment 9; mean 'most likely' results, quoted as distance from *n*

	Granularity			
	Fine	**Medium**	**Coarse**	**Total**
Priming				
Unprimed	6.5	11.9	26.3	**14.9**
Primed	11.1	16.4	27.1	**18.2**
Quantifier				
More than *n*	5.8	12.9	22.9	**13.9**
Fewer than *n*	11.8	15.3	30.6	**19.2**
Total	**8.8**	**14.2**	**26.7**	**16.6**

Figure 2: Results of Experiment 9 (mean 'most likely' results, quoted as distance from *n*)
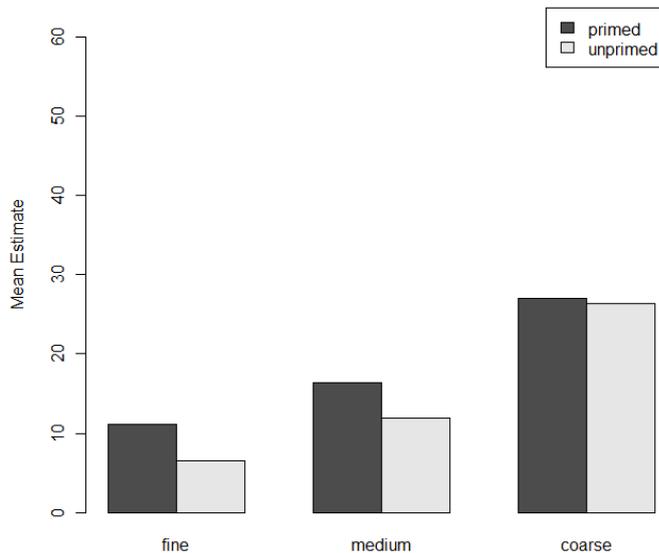
As in Experiment 8, subjects' estimates were more distant from *n* at coarser levels of granularity. Additionally, more distant estimates were observed in the primed versus unprimed condition, and for 'fewer than *n*' versus 'more than *n*'.

The results of a 2x2x3 (priming x quantifier x roundness) within-subjects ANOVA show main effects of roundness ($F_{(2,82)}=47.14$, $p < 0.001$), priming ($F_{(1,41)}=8.06$, $p < 0.01$) and quantifier ($F_{(1,41)}=8.86$, $p < 0.01$). There were no significant interactions. Post hoc testing using paired t-tests identified significant pairwise differences between the three roundness levels (100s vs. units, $p < 0.001$; 100s vs. 10s, $p < 0.001$; 10s vs. units, $p < 0.01$).

Participants' responses for the bound in each condition, in terms of distance from *n*, are shown in Table 25, and presented graphically in Figure 3. This refers to the highest possible value in the case of 'more than *n*' and the lowest possible value in the case of 'fewer than *n*'. Data for 7 participants were excluded from this analysis due to missing values.

Table 25: Results of Experiment 9; mean bound results, quoted as distance from *n*

	Granularity			
	Fine	**Medium**	**Coarse**	**Total**
Priming				
Unprimed	30.1	30.9	53.1	**38.0**
Primed	23.6	37.5	53.3	**38.1**
Quantifier				
More than *n*	26.0	35.6	45.4	**35.7**
Fewer than *n*	27.8	32.7	61.0	**40.5**
Total	**26.9**	**34.2**	**53.2**	**38.1**

Figure 3: Results of Experiment 9 (mean bounds, quoted as distance from *n*)
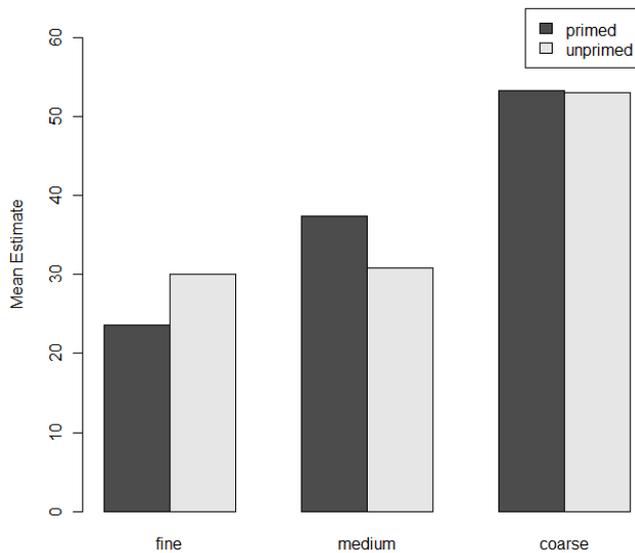
The results of a 2x2x3 within-subjects ANOVA for these data show a significant main effect of roundness ($F_{(2,74)}=10.04$, $p < 0.001$), and no other significant effects or interactions. Post hoc testing using paired t-tests showed significant differences between the 100s and units levels and between the 100s and 10s levels (both $p < 0.001$).

With regard to the bounding values, the findings are affected to some extent by exceptionally high or low values provided by some participants, which could be thought of as outliers. For instance, one participant gave 1000 as the upper limit for 'more than 77'. To address this, responses falling more than 3 standard deviations from the mean for the corresponding roundness level were removed, resulting in the exclusion of data for three additional subjects. Means for the remaining data were as shown in Table 26.

Table 26: Results of Experiment 9 (mean bounds, quoted as distance from *n*) after removal of 'outliers'

	Granularity			
	Fine	**Medium**	**Coarse**	**Total**
Priming				
Primed	23.2	33.2	50.2	**35.5**
Unprimed	17.1	22.6	53.1	**30.9**
Quantifier				
More than *n*	14.0	25.1	44.2	**27.8**
Fewer than *n*	26.2	30.4	59.0	**38.5**
Total	**20.2**	**27.9**	**51.7**	**33.2**

A 2x2x3 within-subjects ANOVA on the 'cleaned' data presented in Table 26 showed main effects of roundness ($F(2,68)=51.34$, $p < 0.001$), priming ($F(1,34)=5.17$, $p < 0.05$) and quantifier ($F(1,34)=7.99$, $p < 0.001$).

Discussion

These data replicate the findings from Experiment 8 concerning the first prediction – once again, there is evidence of implicated bounds, conditioned by roundness or granularity. The range of values ascribed to the expressions does appear typically to be limited by the locations of numerals that match the utterance numeral in granularity level but which would be more informative if used. As a result, participants give higher estimates (relative to *n*) for rounder values of *n*.

In addition to validating the previous experiment's findings with a more carefully verified pool of participants, this experiment also extends it in two ways. First, it shows that granularity-based inferences are available not only for monotone increasing quantifiers such as 'more than *n*' and 'at least *n*', but also for the decreasing 'fewer than *n*'. Here, an

additional finding is that estimates are further from n for 'fewer' than for 'more': this is not predicted by the mechanisms so far postulated as playing a role in quantifier choice, and may therefore require further investigation.

In addition, these data furnish support for the second prediction, namely that the inference should be attenuated by prior activation of the numeral. With respect to the 'most likely' values, participants gave estimates significantly more distant from n in the primed condition (previous mention of n) than in the unprimed condition. With respect to the bounds, similar effects are apparent once the outliers are removed. This pattern appears to indicate that, when there is a contextually determined reason for the speaker to use the numeral, hearers are less likely to draw inferences based on the non-use of the next highest value on the scale of the appropriate granularity level.

The effects observed for priming are, however, less robust than might have been expected if the argument adduced earlier goes through. Two methodological reasons might underpin this observation. First, in the design of this experiment, participants saw both primed and unprimed test items. It is possible that the exposure to priming influenced participants to infer that the numeral presented in the unprimed conditions was nevertheless contextually relevant for some reason, thereby blurring the distinction between primed and unprimed conditions. Second, each of the 12 test conditions (2 quantifiers x 3 roundness levels x 2 priming levels) was represented by a distinct test item, and it is therefore possible that item effects might have contributed to the lack of a stronger priming effect. It is also possible that item effects might underlie the difference between the quantifiers 'more than' and 'fewer than'. These possibilities are investigated in Experiment 10.

5.4.3. Experiment 10 – Direct investigation of the numeral priming effect

In Experiment 9, participants assigned wider ranges of interpretation to numerical quantifiers when the numeral was primed than when it was not. There was also an unexpected difference between behaviour on the quantifiers 'more than' and 'fewer than'. Experiment 10 aims to replicate these findings using a method controlled for item effects.

METHOD

The experiment was implemented using Amazon MTurk as in Experiment 8 (section 5.4.1). Participants were shown a transcript of the following dialogue, either with or without priming, and were asked to estimate the number in question (range and most likely value):

Primed:

Salesperson: This storage unit holds 60 CDs. How many CDs do you own?

Customer: I have more than/fewer than 60 CDs.

Unprimed:

Salesperson: This storage unit holds CDs. How many CDs do you own?

Customer: I have more than/fewer than 60 CDs.

Participants also had the opportunity to comment on their answers.

The two 'more than' conditions were fielded in June 2010; the 'fewer than' conditions were fielded in December 2010. In both cases, fielding of the primed and unprimed conditions was separated by at least a day in order to minimise overlap between the groups of participants.

Participants

A total of 400 participants (100 per condition) were recruited via Amazon MTurk, using the same screening criteria as in Experiment 8.

Results

As for Experiment 8, the data were cleaned prior to analysis by the removal of non-numerical responses and truth-conditionally incorrect responses, as well as a single outlier (a response of 10^{23} in the 'more than' primed condition). In addition, some participants (between 4 and 12 per condition) seemingly misinterpreted the instructions as calling for a single digit to be entered in each space (resulting, for instance, in 'between 1 and 9' for a 'more than' condition). The remaining results (81 in the 'more than' primed condition, 81 in the 'more than' unprimed condition, 85 in the 'fewer than' primed condition and 85 in the 'fewer than' unprimed condition) were analysed. Table 27 summarises these results, again presenting these in terms of distance from n, which is 60 in this experiment.

Table 27: Results for Experiment 10 (quoted as distance from *n*(=60))

		Boundary value	Most likely value
More than *n*			
Unprimed			
	Mean	97.4	75.3
	(SD)	(38.1)	(15.4)
	Median	85	70
Primed			
	Mean	190.2	97.0
	(SD)	(238.9)	(37.5)
	Median	100	80
Fewer than *n*			
Unprimed			
	Mean	31.2	51.5
	(SD)	(21.9)	(9.3)
	Median	40	55
Primed			
	Mean	24.6	43.5
	(SD)	(18.7)	(11.3)
	Median	30	45

Numerically, participants' estimates are more distant from the value 60 in the primed than in the unprimed conditions. In contrast to the findings from Experiment 9, here the values are more distant from 60 in the case of 'more than' than 'fewer than'. A 2x2 (priming x quantifier) ANOVA on the most likely estimate shows main effects of priming $(F_{(1,328)}=39.20, p < 0.001)$ and quantifier $(F_{(1,328)}=33.86, p < 0.001)$, and an interaction of quantifier and priming $(F_{(1,328)}=8.49, p < 0.01)$. Post hoc testing (Tukey) shows a significant difference between primed and unprimed conditions for 'more than' $(p < 0.001)$, and a marginally significant difference between primed and unprimed conditions for 'fewer than' (p=0.07).

A comparable 2x2 ANOVA on the (upper or lower) bound similarly shows main effects of priming $(F_{(1,328)}=13.56, p < 0.001)$ and quantifier $(F_{(1,328)}=15.30, p < 0.001)$, and an interaction of priming and quantifier $(F_{(1,328)}=10.65, p < 0.01)$. Post hoc testing (Tukey) shows a significant difference between primed and unprimed conditions for 'more than' (p <

0.001); in the case of 'fewer than', the difference between primed and unprimed conditions is not significant.

We can gain a broader insight into the participants' behaviour by examining the response patterns more closely. Figures 4 and 5 show the distribution of values given as the upper bound for 'more than 60' and lower bound for 'fewer than 60' respectively.

Figure 4: Distribution of upper-bound responses to 'more than 60' in Experiment 10
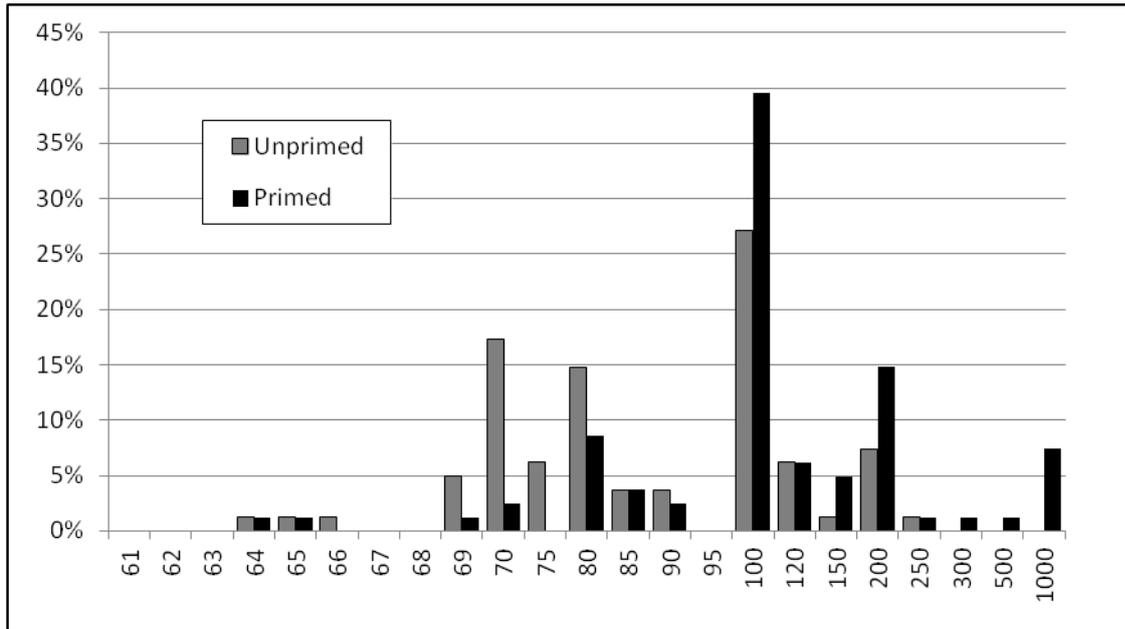
Figure 5: Distribution of lower-bound responses to 'fewer than 60' in Experiment 10
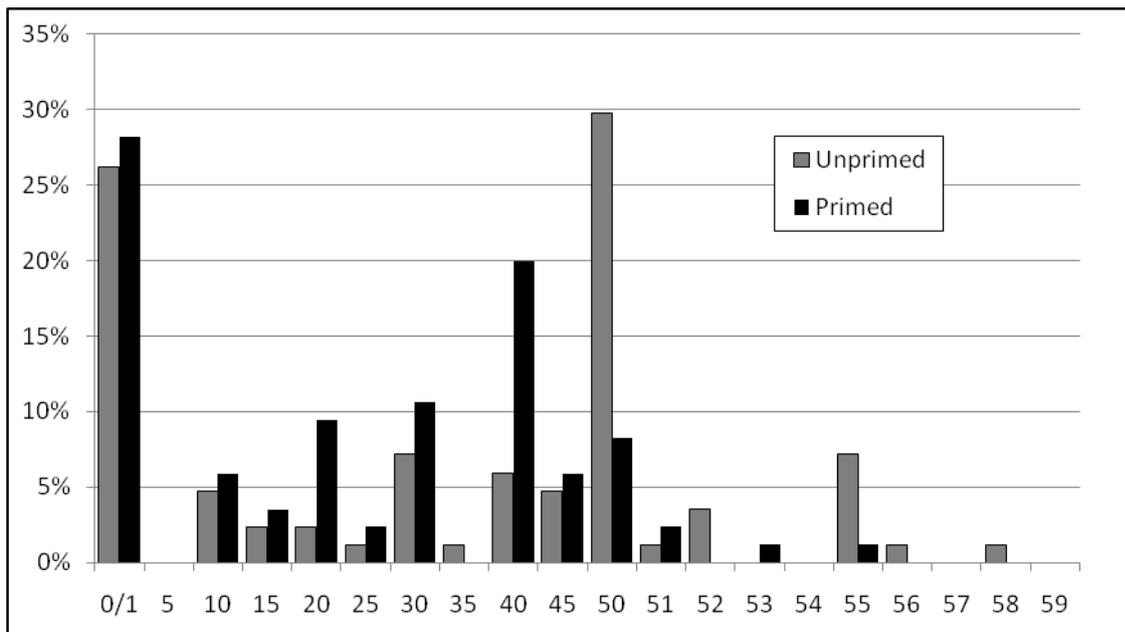
Looking first at the results for 'more than' in Figure 5, the pattern of responses is markedly different between the two conditions. In the unprimed condition, there are three main frequency peaks, at 70, 80 and 100. In the primed condition, there is a single primary peak at 100; responses of 70 and 80 are much less frequent, while higher responses (200, 1000) also occur. A χ^2 test, comparing responses of 70, 80, 100, 200 and 1000 versus all other responses, shows the difference in distribution is significant ($\chi^2 = 20.68$, df=5, p<0.001).

For 'fewer than', an appreciable proportion of subjects gave a response of 0 or 1 for the lower end of the range. We might term these 'truth conditional' responses, in that the range of interpretation is not further restricted by any sort of pragmatic strengthening. Other than these responses, there is again an observable difference between the primed and unprimed conditions: the unprimed condition has a clear primary peak at 50, whereas the primed condition yields a more diffuse set of responses with a smaller primary peak at 40.

Discussion

The results of this experiment confirm that the interpretation of the quantifiers 'more than *n*' is shaped by the contextual status of the numeral *n*. When *n* is not salient in the context, the interpretation assigned to 'more than *n*' is constrained by implicature, as discussed earlier. However, when *n* is salient in the preceding context, this effect is attenuated: hearers allow a wider range of interpretation, indicating that the implicatures are weaker, in accordance with the second prediction made in this chapter.

This observation supports the constraint-based model insofar as these predictions are drawn from that model. However, with respect to the distribution of responses, we can go further. As remarked above, the frequency peaks in responses for 'more than 60' in the unprimed condition are at 70, 80 and 100; in the unprimed condition, there is a single major peak at 100. This accords with the detailed analysis put forward in section 5.3. Recall that, according to the constraint-based account, in the absence of prior activation of a numeral:

- 'more than 80' harmonically bounds 'more than 60' as an expression of quantities over 80

- 'more than 70' is preferred to 'more than 60' as an expression of quantities over 70, except by speakers who rank numeral salience (NSAL) highly.

By contrast, in the presence of prior activation of the numeral 60:

- 'more than 70' and 'more than 80' are only preferred to 'more than 60', for any values, by speakers who rank informativeness (INFO) highly

- 'more than 100' is preferred to 'more than 60' for values over 100 by people who rank either NSAL or INFO highly.

Unpacking this from a hearer's perspective, unprimed 'more than 60' is predicted typically to implicate either 'not more than 70' or 'not more than 80', depending upon the speaker's constraint ranking. It is therefore predicted that hearers will tend to give either 70 or 80 as a pragmatic upper bound. Primed 'more than 60' is predicted typically to implicate only 'not more than 100', so hearers are predicted to tend to give 100 as an upper bound.

The detailed results bear out these predictions. The only clear pattern in the data not accounted for by this model is, then, the use of 100 as an upper bound for the unprimed case. However, this can again be explained if some participants are assuming that the numeral is relevant to the discourse, even if they are not privy to the explicit mention of the numeral that they suppose has taken place. In this case we would expect it to behave like the primed case, which is predicted to yield 100 as an upper bound, as discussed above.

In the case of 'fewer than n', the overall picture is similar, with the important difference that there is a firm semantic/logical lower bound for the possible interpretations that can be assigned, namely zero. In Experiment 10, a considerable proportion of respondents in both primed and unprimed conditions gave 0 or 1 as their lower bound, eschewing any kind of pragmatic enrichment of the numerical expression given. This also tends to blur the difference between the primed and unprimed conditions. As there is no logically necessary upper bound in the case of 'more than n', subjects' estimates can exceed n by an arbitrary amount. It is therefore not entirely surprising that, in this experiment, estimates for 'more than n' were numerically more distant from n than were the estimates for 'fewer than n'. In Experiment 9, the reverse was true, but in that case the values of n tested were higher, and zero responses were consequently more distant (in purely numerical terms) from n than zero responses were in Experiment 10.

5.5. Discussion and conclusions

The experiments presented in this chapter demonstrate the availability of scalar implicatures from numerically quantified expressions such as 'more than n' and 'fewer than n', contrary to

some previous proposals. However, this implicature is seen to be influenced both by numeral salience or granularity considerations, and by preceding context.

Although these findings are free-standing, insofar as the empirical methodology makes no assumptions about the underlying process by which numerically quantified expressions are selected, this account is particularly consonant with the constraint-based proposal that is the central topic of this thesis. Not only does this proposal correctly predict the presence of the implicatures documented, in the face of an established theoretical literature to the contrary, but it also makes reasonably accurate predictions as to the precise strength of the pragmatic bounds that hearers infer.

The constraint-based model can thus be seen as an attempt to codify a fairly clear and uncontroversial intuition about the nature of speaker-hearer interaction. When the hearer knows that the speaker has a perfectly good, relevance-driven reason to use the utterance they did, the hearer is precluded from drawing a scalar inference. From this perspective, each constraint in the model can be seen as a possible obstacle to implicature. Whenever an utterance U_1 obeys a constraint C, then the implicature that some other utterance U_2 could not have been truthfully uttered depends in some measure on whether this other candidate also obeys C. It is therefore not generally possible to draw inferences that statements do not hold if the utterance of those statements would incur extensive constraint violations. Thus, each constraint has an inhibitory effect on inference.

The remaining issue arising from the above experiments concerns the nature of the implicatures from non-round numerals. In Experiment 8, this failed to conform with the initial predictions: 'more than 93' gave rise to the implicature 'not more than 100' rather than 'not more than 94'. Fox and Hackl (2006) observed that implicatures of this latter type would be communicatively inefficient, as '(exactly) 94' would economically convey precisely the same meaning. A speaker aware that 'exactly 94' is the case should therefore say 'exactly 94' rather than 'more than 93'. However, a speaker who is not aware that 'exactly 94' is the case also should not say 'more than 93', as in this case the implicature might be conveyed but might be false, and hence render the utterance infelicitous.

This argument dispenses with the unwanted implicature, but at the cost of excluding the possibility of 'more than 93' ever being uttered at all. Given the perfectly acceptable examples of this that have been discussed, this appears to be throwing the baby out with the bathwater. However, appeal to the numeral priming constraint can answer this objection.

The presence of this constraint in the system signifies that the prior use of a numeral may license its reuse, even in contexts where other expressions would otherwise be preferred. We have also seen how this activation suppresses the implicature. Therefore, if expressions such as 'more than 93' occur in such contexts, the strong implicature 'not more than 94' is predicted not to arise, while a weaker implicature such as 'not more than 100' may be conveyed.

If it is the case that expressions of this type only surface in contexts in which the numeral is activated, the above explanation is sufficient to account for the lack of implicature. The examples discussed in the literature do indeed appear to meet this condition. For instance, consider (89), repeated below for convenience.

(89)   John has more than three children.

The utterance of (89) naturally seems to suppose the existence of some context within which John's having three children is somehow critical, perhaps because it is a threshold for benefits, it is as many as will fit in the back of the car, or because it signifies that John has more children than some other individual who is known to have exactly three. In all these cases, it could be argued that the numeral is salient to the speaker. Another possibility is that the speaker has previously been told precisely that (89) holds, in which case the numeral is clearly activated by prior mention.

Intuitively, it is more natural for examples using non-round numbers, such as 93, to surface in contexts of prior activation. There are good reasons for this. If 'more than 93' does not reuse a number and does not directly address a question under discussion, then both speaker and hearer are forced to process a number that is not salient or highly activated. By comparison with an expression such as 'more than 90', this generates additional processing requirements without achieving much in the way of additional cognitive effects: i.e. it is less relevant. Note that this argument does not rely on the infelicity of the implicature that the expression would convey, but merely on the relevance of the utterance itself.

I would go further and suggest that hearers, when confronted with a quantifier such as 'more than 93', tend to posit that there is some reason for this particular numeral to be mentioned. Once it is supposed that there is some such reason, there will be no implicature, because the choice of utterance is to some extent forced. Or, to be more accurate, to the extent that the hearer supposes that there is a reason for using that number, the hearer does not draw an

inference. It is easy to locate examples of this type, where no implicature arises as a consequence: (99) refers to the number of seats required to command a majority in the UK House of Commons, (100) to the 'perfect' score for an over in cricket, and (101) to the size of a standard pack of cards.

(99)   Maybe fewer than 25% think the unthinkable – that the Tories will obtain fewer than 326 seats[25].

(100)   It is also possible to get more than 36 runs in an over, but it has never happened before[26].

(101)   [M]ost decks used in casinos for poker have more than 52 cards so people don't cheat[27].

This effect should also influence hearers' interpretations at a coarser granularity level, albeit to a lesser extent. When a hearer interprets 'more than 100', they are entitled to question whether there is a specific reason for 100 to be referred to; and if so, the inference should be attenuated. This might further account for the variability exhibited by participants in these experiments: a competent user of language should interpret 'more than 100' without an implicature if the speaker's decision to mention 100 is contextually forced. In this way we, as hearers, can 'repair' infelicitous utterances by positing a context against which they can be interpreted without implicature.

This proposal appears to account for the distribution of responses to quantifiers at the finest granularity level. However, further experimental investigation would be required to ascertain whether hearers are indeed reasoning this way. If so, it further remains to be shown whether the bounds offered by participants in these experiments are derived by scalar implicature or some other means, and how the degree of likelihood hearers attach to the implicature influences the judgements they obtain.

In passing, we might also consider the role of granularity in accounting for the pragmatic enrichments associated with examples such as (82), repeated below.

---

[25]   http://www2.politicalbetting.com/index.php/archives/2010/01/17/do-these-1992-approval-ratings-hold-the-key/, retrieved 2nd August 2010.

[26]   http://www.sports1234.com/cricket/1652-2-cricket.html, retrieved 2nd August 2010.

[27]   http://answers.google.com/answers/threadview/id/100828.html, retrieved 2nd August 2010.

(82)   She can have 2000 calories without putting on weight.

Carston (1998) discusses this as a case of a numeral with an upper-bounded meaning: that is, (82) is interpreted as a specification of the maximal calorie intake permitted. This is counterintuitive on closer inspection. It seems uncontroversial that (82) permits the referent to consume (precisely) 2000 calories without putting on weight, but it does not appear to follow that 2001 calories would cause the referent to put on weight. By contrast, (82a) would have this interpretation. The difference suggests that the numeral in (82) does not convey an 'at most' meaning in its own right.

(82a)   She can have at most 2000 calories without putting on weight.

However, we may be able to salvage an 'at most' interpretation for (82) by appeal to granularity. On this account, (82) would assert that 2000 is the largest number at the appropriate granularity level such that the referent may consume that number of calories without putting on weight. Intuitively this appears to give a more satisfactory explanation of (82)'s meaning than Carston's initial account does.

More broadly, we could then sketch the behaviour of this utterance as follows from a constraint-based perspective. Let k be the actual limit on the referent's calorie intake. By appeal to our encyclopaedic knowledge about dieting, it is true to say (82b) for any $c \leq k$.

(82b)   She can have c calories without putting on weight[28].

For any $k > 2000$, (82) is therefore an admissible alternative ($c = 2000$). If the speaker aims to select an optimal utterance, (82) has the advantage of using a highly salient numeral, and will therefore be preferred under constraint rankings that rank NSAL highly. (82) will in fact be optimal in this case for k up to 2100 and perhaps higher, despite its relative lack of informativeness in such cases. A hearer, then, is predicted to respond to an utterance of the form (82b) by drawing an implicated upper-bound of the type discussed at length in this chapter, conditioned by the salience or granularity of c. Applied to (82), this provides the interpretation I argue for above. Notably it does so without any weakening of the assumption of punctual numeral semantics (not even the reasonable assumption that 2000 could be approximative).

---

[28] By contrast, it will not be truthful to say 'at most c' for $c < k$, so I dismiss these alternatives from consideration here.

In summary, then, the constraint-based mechanism appears able successfully to predict participants' behaviour in the experiments presented in this chapter, and account for our intuitions about related example sentences. This suggests that no additional theoretical machinery need be posited to account for the seemingly deviant pragmatic behaviour of comparative quantifiers. Numeral salience and numeral priming considerations are necessary in order to account for the nature both of the inferences and of their attenuation by prior mention of the numerals. Thus these experiments serve to support the constraint-based mechanism in general and the presence of these constraints in the system in particular. In the following chapter, I consider the role of corpus data in providing further support for the constraint-based model.

6. CORPUS EVIDENCE FOR CONSTRAINTS ON NUMERICAL EXPRESSIONS

In addition to generating experimentally testable hypotheses, the constraint-based model introduced in this thesis makes predictions as to the distribution of numerically-quantified expressions in corpora. In this chapter, I examine the extent to which the constraints should be manifest in corpora, and enumerate the predictions that the constraint-based model makes. I then discuss some relevant methodological issues concerning numerals in corpora, and evaluate the corpus evidence for constraints in the light of these concerns. I expect to show that corpora provide additional support for the constraint-based model, beyond that offered by the existing experimental literature and the new empirical investigations reported so far.

6.1. Constraints and corpus frequencies

From an OT perspective, it is expected that unmarked forms – i.e. those which do not incur violations of markedness constraints – will surface preferentially to marked forms. This arises due in part to the architectural imperative of OT that has been labelled 'the emergence of the unmarked' (McCarthy and Prince 1994). As OT constraints are never deactivated, they may in principle play a decisive role in selecting an optimal output candidate even when they are lowly-ranked (and therefore widely violated in the language as a whole). As discussed in section 2.3, the NOCODA constraint, violated by the presence of syntactic codas, is argued to be highly ranked for languages such as Hawaiian and lowly ranked in languages such as English. In the former case, this has the effect of prohibiting codas across a wide range of environments; in the latter, it is not sufficiently important to prohibit codas in general, but is satisfied whenever possible, for instance when underlying coda material can be resyllabified in order to surface as an onset. Through the emergence of the unmarked, OT recapitulates the observation (attributed to Givón 1979) that 'categorical phenomena in certain languages are mirrored by frequentistic phenomena in others' (Bresnan and Aissen 2002: 88), an observation extensively studied in the literature (e.g. Hawkins 2004).

Under an appropriate set of assumptions, we would expect the markedness constraints in our model to influence the frequencies of surface forms, as evident in corpora, in the same way. Despite the undeniable fact that all these markedness constraints are widely violated, we expect a preference for unmarked forms to show through. This is, loosely speaking, because each markedness constraint is sometimes the critical factor, whatever the constraint ranking; and whenever it is the critical factor, it unequivocally points to the selection of the unmarked option.

140

It is still conceivable within an OT system that no preference for unmarked forms would be manifest in corpora. This would arise if faithfulness constraints happen typically to favour marked forms, thus counterbalancing the effect of the markedness constraints. To take a concrete example, I have argued that 20 is a more salient numeral than 19, and therefore its use is favoured by numeral salience. Suppose that for some reason cigarettes started to be sold in packs of 19 rather than 20: then the numeral 19 would frequently be contextually activated for a speaker who regularly bought cigarettes, and numeral priming would militate in favour of its regular use. This might result in an overall preference in frequency for 19 over 20 in that speaker's production, even under the assumption that 20 remained more salient.

Assuming, however, that no such exotic distribution of contexts is in effect, we expect to be able to draw corpus predictions from markedness constraints. In the model discussed so far, these are quantifier simplicity and numeral salience.

With respect to faithfulness constraints, the situation is more problematic. As these constraints rely for their effect on the situation in which the utterance occurs, we cannot generalise about the types of utterance that satisfy these constraints, other than by generalising about the situations themselves. For instance, any given statement may or may not violate informativeness, depending upon the state of knowledge of the speaker. To draw a prediction about corpus frequencies based on informativeness, we would need first to establish whether there were knowledge states concerning numerical quantities that were especially common, a task beyond the scope of this enquiry (although the nature of these knowledge states is touched upon in the concluding discussion). Furthermore, any trends in the situations that are encountered – if, for instance, there were a general preference for the use of a certain class of quantifier – could more constructively be handled by the addition of markedness constraints to our model.

In short, very little can be said about the predictions that the faithfulness constraints (numeral granularity, numeral and quantifier priming, and informativeness) make as regards corpora. However, we may be able to ascertain whether a particular instance of usage from a corpus does adhere to priming constraints, a point to which I return later.

Hence, the bulk of this section will deal with the predictions arising from the markedness constraints posited as part of this model. These are outlined in the following section.

141

6.2. Predictions arising from markedness constraints

In the following subsections I spell out the predictions arising from the markedness constraints in this model, considered individually and with reference to their interactions. Subsequently I turn my attention to the methodological details of how to test these predictions.

6.2.1. The preference for simple quantifiers

The quantifier simplicity constraint, introduced in section 2.4.3, is a markedness constraint defined as follows: 'The utterance must use the simplest quantifier possible. Incur a violation for each level of complexity exhibited by the quantifier used.' Based on the above argumentation, we expect quantifiers to be dispreferred in corpora according to the extent to which they violate this constraint.

To flesh out this prediction, we need to ascertain the extent to which individual quantifiers incur these violations. However, as discussed in chapter 2, we do not have a convenient global metric for assessing the complexity of quantified expressions. Furthermore, if we are proposing hypotheses to be tested against corpus frequencies, clearly we cannot beg the question by using corpus data as a guide to quantifier complexity. Therefore, we are for the moment restricted to complexity-based predictions that are grounded directly in philosophical and psychological considerations.

In particular, we will focus on two predictions that are foreshadowed in the discussion in the preceding chapter. First, with respect to comparative and superlative quantifiers, we argue that superlative quantifiers are more complex for the reasons discussed at length in chapter 4. These should therefore be disadvantaged in corpora. Second, with respect to negation, we expect that negated quantifiers are more complex than their non-negated counterparts, and therefore should occur less frequently. Our predictions driven by complexity, then, are as follows.

Prediction #1: Superlative quantifiers will occur less frequently than the corresponding comparative quantifiers.

Prediction #2: Negated quantifiers will occur less frequently than the corresponding non-negated quantifiers.

6.2.2. The preference for round numerals

The numeral salience constraint is a markedness constraint specifying that numerals that lack certain types of roundness incur violations. Therefore, we can derive the prediction that round numbers will be preferred in corpora to non-round numbers of comparable magnitudes. On the surface, this is a trivial prediction, in that the definition of the numeral salience constraint in section 2.4.4 co-opted a definition of roundness from Jansen and Pollmann (2001), whose work drew upon corpus data in the first place. However, as I discuss later in this chapter, Jansen and Pollman considered all numeral uses, not merely cardinalities. Therefore the prediction that this preference will continue to be manifest if we restrict our attention to cardinalities is not an immediate corollary of their findings. I state it as follows.

Prediction #3: With reference to cardinal usages, round numbers are more frequently used than non-round numbers of comparable magnitudes.

6.2.3. Interaction between quantifier complexity and numeral salience

Following the discussions of constraint interaction effects in the previous chapters, we can also consider how quantifier complexity and numeral salience are predicted to interact in terms of corpus data.

First, we expect quantifiers such as 'more than' to exhibit peaks in usage when the complement numeral is round, compared to when it is not. As we are making no assumptions about the typical knowledge states of speakers, the informativeness constraint is neutral as to whether the numeral complement of 'more than' should be round or non-round. Numeral salience therefore mandates that the complement is preferentially round (by the emergence of the unmarked). To put it another way, making no assumptions about knowledge states, 'more than 20' is optimal in informativeness just as often as 'more than 21' is; however, sometimes the former is predicted to be preferred on the basis of numeral salience even when the latter is optimal in informativeness. Hence, 'more than n' tends to prevail over 'more than $n+1$' for round n and non-round $n+1$.

We might in fact go further and predict that the preference for round numbers as complements for quantifiers such as 'more than' is more pronounced than is the preference for round numbers in the unmodified case. 'More than 20', for instance, competes with 'more than 21, …, 24', and is predicted to be generally favoured when numeral salience

outranks informativeness. By contrast, the plain numeral '20' competes with a similar range of options when numeral salience outranks informativeness, but is itself dominated by 'about 20' or 'exactly 20' except when quantifier simplicity also outranks informativeness. The net effect of this is that round numbers are more effective 'attractors' when they are the complements of 'more than' than when they appear in isolation. From this we derive the stated prediction.

In the case of 'at least/most', the argument is slightly different. According to the above account, 'at least 20' is used in some situations where 'more than 19' would be an alternative. We can think of this as the round numeral drawing some instances of quantification that would otherwise be realised using 'more than' into being expressed by 'at least'. When the numeral is not round, barring influence from numeral priming, there is no reason to prefer 'at least' over 'more than'. Therefore we predict that the preference for round numbers as complements for 'at least' is also more pronounced than the preference for round numbers in isolation.

Hence, our predictions concerning the interaction of quantifier simplicity and numeral salience are as follows.

Prediction #4: The numeral complements of numerically-quantified expressions are preferentially round.

Prediction #5: The peaks of numeral usage at round numbers are higher for comparative quantifiers than for unmodified numerals.

Prediction #6: The peaks of numeral usage at round numbers are higher for superlative quantifiers than for unmodified numerals.

6.3. Some methodological issues in corpus research on numerically-quantified expressions

Before we evaluate the six hypotheses enumerated above, we need to identify the precise data points we are interested in from corpora, and establish how we can restrict our attention to these data. Specifically, the data we are interested in involve cardinal uses of number, and to obtain these we need either to adopt a specific search strategy or demonstrate that the broader data are themselves representative. This is a more stringent approach than that of Jansen and Pollmann (2001). In this section I review their research in order to examine the motivation

144

for restricting attention to cardinal number usages, and to consider the consequences of their decision not to do so.

First, I note that Jansen and Pollmann's (2001) methodology does distinguish between quantitative and labelling uses of number. In the first instance, they restrict their attention to numbers that are modified by 'about', which only makes sense in a quantitative environment. This is appropriate because ordinal uses of number, such as street numbers, do not necessarily respect considerations of numeral salience: they may be assigned in such a way as to ensure that, for instance, to every 'number 25' there is a corresponding 'number 23'. Hence no salience-based pattern is expected in such cases[29].

Within the domain of quantification, however, Jansen and Pollmann are not selective. This is potentially problematic because numbers are used extensively in non-cardinal measurements: 5 metres, 10 minutes, £15, 20%, and so on. It is evident that quantities of this type, which are either continuous or discrete with a step less than one unit, do not license the same inferences as cardinal quantities. 'About 9 minutes' is entirely felicitous, for instance, and seems naturally to refer to a range of times from some seconds below to some seconds above 9 minutes; 'about 9 people' is less felicitous because to gloss it as 'in the range from 8.7 to 9.3 people' is nonsensical (if we are talking about 9 specific people, rather than an average 9 people from a larger sample, as in 'about 9 out of every 10 people voted in the election'). The objective of Jansen and Pollmann's enquiry is to validate the claim of Dehaene and colleagues that round numbers convey approximate meaning: however, if you consider non-cardinal measurements, any integer can convey approximate meaning. Therefore by including these instances they weaken their evidence in favour of Dehaene's claim. For our purposes, it will also be important to distinguish cardinal quantities when we discuss structures such as 'more than n'. I argued in chapter 4 that, semantically, 'more than n' was equivalent to 'at least $n+1$', but that clearly only holds when n is cardinal. 'More than 3 hours' is transparently not the same as 'at least 4 hours'.

A further problem arises from Jansen and Pollmann's lack of selectivity in the case of percentages. By definition, this system expresses fractions of quantities by adopting the reference point 100% = one whole quantity. Consequently, major subdivisions of 100, which

---

[29] Patterns that still apply to sets of numbers of this type include Benford's law (Newcomb 1881). Benford's law states that the leading digit of numbers from random data sources is typically distributed in a non-uniform way: more specifically, that the first digit is 1 about 30% of the time, 2 about 17% of the time, and so on.

correspond to salient parts of the whole (e.g. ¼, ½, ¾) are widely used as percentage quantities (25%, 50% and 75% respectively). These constitute part of the data set upon which Jansen and Pollmann (2001) perform their regression analysis. Their analysis shows, among other things, that 10-ness, 5-ness and 2.5-ness are significant predictors of frequency, which they interpret as demonstrating that powers of 10 enjoy a privileged status, and (p.201) 'there is a 'natural' propensity for halving or doubling quantities'. However, *to the extent that this relies on percentages*, this is circular reasoning. The percentage system appears designed to facilitate representations of this type, and the frequency of numbers exhibiting 2.5-ness in particular might in turn reflect this.

Furthermore, the inclusion of percentages could also bias the range of values used towards the range 0-100, as greater percentages than this do not always make sense (e.g. '110% of the votes'). This contrasts with the cardinal case in general, where crucially there is no *a priori* restriction on the range of values that may be exhibited. Jansen and Pollmann's model includes terms in n^{-1} and n^{-2} which are continuous, and therefore the underlying curve of their model does not exhibit stepwise behaviour at 100 (or any other point). Hence, the inclusion of percentages might result in Jansen and Pollmann's model being underfitted, and render it inappropriate for addressing the question they wish to discuss.

In passing, I would also query Jansen and Pollmann's decision to posit a term in n^{-2}. Over all positive integers, the sum of n^{-2} converges to $\pi^2/6$ ($\approx$ 1.64). This sum is almost entirely constituted of its first few terms[30], which are 1, 0.25, 0.111, 0.067, and 0.04. Only 5.8% of the total is contributed by terms with $n > 10$. It follows that the coefficient of n^{-2} in the regression analysis is largely determined by the frequencies of numerals less than 10, and thus does not tell us anything useful about the frequency distribution of larger numbers. In fact, the significance of n^{-2} as a predictor could be argued to arise largely because it enables the line to be fitted more accurately for just a couple of small values of n. Hence ascribing significance to a term in n^{-2} may reflect the analysts' decision to posit this type of term rather than the existence of a significant pattern relating number magnitude and frequency. By contrast, n^{-1} does not converge to a finite sum over the positive integers, and consequently any coefficient for n^{-1} must be somewhat appropriate for larger n to constitute a good fit.

---

[30] Jansen and Pollmann (2001) actually exclude n=1 from consideration, but the contribution of small n to the remaining sum is still vastly predominant.

The above objections notwithstanding, it seems entirely plausible that the model proposed by Jansen and Pollmann is fundamentally correct. However, some of these considerations suggest that a more selective approach to the corpus data would be advantageous. Furthermore, as our hypotheses are drawn from argumentation based upon the assumption that we are dealing with cardinal quantities, we should ideally restrict our attention to these quantities when evaluating our hypotheses.

Within the domain of cardinal numbers, one remaining issue is whether items such as '12 million' are treated as numbers in their own right, or as an instance of '12' (or indeed 'million'). Jansen and Pollmann's take on this is to count '12 million' as an instance of '12', but to count '12,000,000' as a number in its own right, even though this latter is presumably read as '12 million'. This arbitrary decision calls attention to a significant structural property of the number system that we use, namely that larger and less round numbers in the system are built from concatenations of smaller and rounder numbers. Moreover, our understanding of these numbers seems to proceed in the same way (hence we can understand novel numbers compositionally). Therefore, an instance of e.g. 'twenty-seven' could be said to rely upon the concepts of 'twenty' and 'seven'. It seems natural further to suppose that these more primitive numeral concepts are activated in some way both in production and comprehension, although we cannot demonstrate that with confidence (see chapter 7 for further discussion).

In interpreting corpus data, then, we must decide whether we wish to consider these 'incidental' usages as evidence of the cognitive availability of the numerals used in this way. Here I prefer not to take this step, for two main reasons. First, we avoid double counting: Jansen and Pollmann only consider numbers up to 1000, but we take a broader outlook, and would wish to think of '12 million' as an instance of '12 million' first and foremost. Second, when we consider 'more than' and similar quantifiers, we note that the '12' of '12 million' does not behave like a cardinal quantity on its own, even when '12 million' is being used in a cardinal way. For instance, 'more than 12 million' is not the same as 'at least 13 million'. Therefore, we aim in the first instance to identify and discuss numerals separately without reference to their compositional structure.

6.4. Corpus evidence for the predictions on quantifier usage

With the subsidiary goal in mind of identifying cardinal usages of individual numerals in corpora, we now turn to evaluating the predictions articulated in the preceding sections.

Prediction #1: Superlative quantifiers will occur less frequently than the corresponding comparative quantifiers.

To verify this prediction, we need to show that 'more than' is more common than 'at least' and 'fewer/less than' more common than 'at most', when these take cardinal complements. From a corpus perspective, this is not entirely straightforward, because we wish to exclude both non-numerical complements ('more than happy', 'at least satisfactory', etc.) and non-cardinal numerical complements ('more than 25%'). In order to obtain solely the cardinal uses we are looking for, I examine the corpus for instances of '<Quantifier> * <Noun>', where * is a wildcard and <Noun> denotes a common noun selected for its propensity to be quantified by expressions of the type we are interested in. I use the nouns 'people', 'men', 'women', 'cars' and 'houses'.

Searching the British National Corpus (BNC) via the BNCweb interface provided by Lancaster University (http://corpora.lancs.ac.uk/BNCweb/home.html), we obtain the frequencies shown in Table 28.

Table 28: Frequencies for some Q*N sequences in the BNC

Quantifier	'…people'	'…men'	'…women'	'…cars'	'...houses'
More than	474	55	51	10	14
Less than	27	25	3	1	4
Fewer than	16	4	2	2	3
At most	1	0	0	0	1
At least	376	27	20	4	1

Inevitably these results include some instances which are not of the type we are interested in: one such instance is 'Suppliers are more than sales people'. Furthermore, these do not include instances of 'Q * * N' strings, such as 'less than one hundred people'. However, as this applies equally to all quantifiers, there is no reason to suppose that the omission of these should bias the comparison.

It is possible, however, that the non-numerical quantifier responses are unevenly distributed across quantifiers, and therefore we need to consider these as we evaluate the hypotheses

under test. Of the first 50 instances of 'more than * people', 45 (90%) are of the cardinal quantifying type required. Of the first 50 instances of 'at least * people', 45 (90%) are of the type we require, with an additional instance of 'at least some people'. There is therefore no evidence of difference between these distributions. Based on this sample, the 99% confidence interval for these percentages is (74.7%, 96.9%). This implies that, with 99% probability in each case, the number of quantifying instances of 'more than * N' exceeds that of 'at least * N' for each N under test, assuming that 'people' is a representative case of N (74.7% of the uncounted instances of 'more than' exceeds 96.9% of the uncounted instances of 'at least' in each case).

Therefore, even without recourse to a full count, we can be confident that these data are sufficiently representative. Thus the greater frequency of 'more than' over 'at least' in every one of these conditions supports the hypothesis that 'more than' is preferred.

For 'less than', 16 of the 27 forms are quantifiers, although 3 of these are modified by 'no', which rules them out from consideration in this discussion. Nevertheless, it seems clear that 'less than' and 'fewer than', whether taken together or separately, substantially outnumber usages of 'at most', which on the above evidence is seldom encountered in quantifying contexts of this type.

Therefore, the corpus data support the first prediction: usage of comparative quantifiers is more widespread than that of superlative quantifiers. However, we note that the above methodology has limitations which would need to be addressed to permit subtler distributional issues to be resolved.

Prediction #2: Negated quantifiers will occur less frequently than the corresponding non-negated quantifiers.

To test this hypothesis using corpora, we also need to select quantifiers to test. For some quantifiers, such as 'exactly', negation is grammatical. For others, such as 'about' and 'approximately', it appears that explicit negation is ungrammatical, so we do not expect to see these forms in corpora. For other quantifiers, such as 'at least' and 'at most', negation is apparently marginal in grammaticality. These observations cohere with the general outlook of this thesis, as these can naturally be accounted for in terms of informativeness (in the case of the ungrammatical examples) or quantifier simplicity (in the case of the ungrammatical examples). Nevertheless, we wish to test our specific current hypothesis with forms that do

admit explicit negation: the absence of forms from the grammar is a sufficient explanation for their absence from corpora, without positing that markedness considerations are at play, even though we might argue that markedness underpins some of these forms' absence from the grammar.

In the case of 'more than' and 'fewer/less than', explicit negation is grammatical, but in two distinct ways: with 'no' and with 'not'. Nouwen (2010) argues that the forms 'no more than' and 'no fewer/less than' are not merely negations of 'more than' and 'fewer/less than', but convey additional communicative effects, in terms of the speaker's attitude towards the quantity being discussed. However, given that negation is marked, the constraint-based model admits the counter-argument that these effects are pragmatic in origin. Hence for the moment I assume that 'no more than' and 'not more than' are both truth-conditionally negations of 'more than', and exhibit greater complexity than the positive form. Thus I predict that both these forms will conform with prediction #2, and – taken together or separately – will occur less frequently than their non-negated counterpart. The same goes, *mutatis mutandis*, for 'fewer than'.

Note that, although we are primarily concerned with numerical expressions here, the prediction would appear to generalise to non-numerical quantifiers, assuming that complexity is also a factor in these environments. Of these, the most straightforward to test in corpora are 'all' and 'many', as their negations appears to be distributed identically to the positive forms, although we note that this negation can be expressed in other ways. We restrict our attention to the partitive ('…of the…') to exclude idiomatic uses of 'all' ('at all', etc.). These data are also presented below (Table 30).

Broadly, we expect the complexity of negation to apply to all quantificational contexts, not just to cardinal settings. Therefore, for convenience, we will examine the rates of co-occurrence with numerals 10, 20 and 50. As the BNC codes numerals either as digits or text, we count both separately. Results of this are given in Table 29. Note also that counts of the positive quantifiers include their negations – figures excluding these are given in parentheses.

Table 29: Frequencies for 'Q # of the' in the BNC

Quantifier	'...10'	'...ten'	'...20'	'...twenty'	'...50'	'...fifty'
Exactly	5 (5)	17 (17)	2 (2)	3 (3)	4 (4)	2 (2)
Not exactly	0	0	0	0	0	0
More than	221 (192)	208 (162)	355 (336)	174 (154)	298 (284)	122 (115)
No more than	26	38	15	12	8	6
Not more than	3	8	4	8	6	1
Fewer than	21 (18)	17 (10)	18 (15)	11 (7)	10 (9)	5 (5)
No fewer than	3	7	3	4	1	0
Not fewer than	0	0	0	0	0	0
Less than	112 (103)	104 (92)	91 (83)	51 (44)	73 (68)	28 (28)
No less than	4	6	5	3	2	0
Not less than	5	6	3	4	3	0

Table 30: Frequencies for Q (partitive) in the BNC

Quantifier	Frequency
All of the	1970
Not all of the	82
Many of the	6057
Not many of the	11

These data appear to support the hypothesis under test. Each positive quantifier outnumbers its explicitly negated equivalents in every context examined. In the case of the quantifiers in Table 30, and for 'exactly', it could be argued that there is an asymmetry between the positive and negative case with respect to informativeness, the negations being less informative (in the sense of ruling out possibilities, as discussed in section 2.4.1) than the

positive statements. However, in the case of the comparative quantifiers in Table 29, positive and negative quantifiers are ostensibly equally informative, with each giving rise to entailment relations, negation merely reversing the entailment direction. Hence, the dominance of positive statements in frequency does appear to support the claim that quantifier complexity is a factor in determining usage.

Prediction #3: With reference to cardinal usages, round numbers are more frequently used than non-round numbers of comparable magnitudes.

This prediction is specific to cardinalities, so we need to restrict our corpus study to these; however, it is not specific to quantifier type, and therefore we do not need to consider the term before the numeral. Arbitrarily I will choose 20 and 50 as round numbers (these are expressed by single words and can unambiguously located in the corpus) and compare their frequencies with those of their immediate neighbours 19, 21, 49 and 51. In order to restrict the focus to cardinalities, I consider instances of numeral + noun for the same nouns used earlier. These frequencies are shown in Table 31. Here I tabulate digital ('50') and word ('fifty') uses separately to avoid any ambiguity as to the comparisons being performed, although I offer no theory as to the motivation for using one option instead of the other (nor any defence of the BNC's policy on transcribing numbers). Note that these figures include instances of, for instance, 'one hundred and fifty' among the instances of 'fifty'; however, this applies equally for all the numerals under test.

Table 31: Frequencies of '# N' in the BNC

Numeral	'…people'	'…men'	'…women'	'…cars'	'...houses'
19	31	5	3	2	0
Nineteen	6	2	2	0	0
20	96	14	10	7	7
Twenty	77	25	7	4	5
21	25	1	2	0	0
Twenty(-)one	9	3	2	1	1
49	5	0	1	0	0

Numeral	'…people'	'…men'	'…women'	'…cars'	'...houses'
Forty(-)nine	3	0	0	0	2
50	69	18	9	4	4
Fifty	59	20	5	1	2
51	8	0	1	0	0
Fifty(-)one	1	0	0	0	0

The above data appear clearly to support the hypothesis under test. Across all nouns considered, the counts for round numbers are higher than the counts for the adjacent non-round numbers at both the 20 and the 50 level. We could consider this to constitute a total of 20 pairwise comparisons (20 vs. 19, 20 vs. 21, 50 vs. 49, 50 vs. 51 for each of the five nouns), in which case the sign test shows that this pattern of difference is highly significant (20 comparisons, round numbers preferred in each; p < 0.001). Therefore, restricting our attention to cardinal uses of numerals, it appears that the pattern attested by Jansen and Pollmann (2001) for general numeral usage still arises, as predicted.

Prediction #4: The numeral complements of numerically-quantified expressions are preferentially round.

To test this hypothesis, we need to restrict our domain of enquiry to specific expressions such as 'about', 'approximately', 'more/fewer/less than', 'at least/most'. However, we also need to restrict our attention to cardinal quantities, on the grounds that the notion of roundness is dependent on the assumption that we are dealing with cardinals. In a more general setting of quantification, any integer could be considered a round quantity, as discussed earlier. However, we cannot apply the method used above, as the frequencies of e.g. 'about 50 people' are too low in the BNC to permit this type of quantitative comparison.

Therefore, in the first instance, I instead consider a sample of these quantified expressions, and consider whether the sample usages reflect a preference for round numbers. To do this, I search for expressions of the form 'there are Q' with numerical complements, and count how many of the first 50 numerical complements are round. This approach aims to elicit exclusively cardinal usages. I use the notion of *k*-ness to quantify this roundness, and

specifically consider whether the numbers have 2-ness, 2.5-ness, 5-ness and 10-ness. I label numbers with one of these types of *k*-ness 'degree 1', those with two types 'degree 2', and so on. I consider single-digit numbers separately as their potential to exhibit *k*-ness depends on their magnitude. The results of this are presented in Table 32. For reference, the total frequency count for each quantifier is also included: an asterisk denotes that this count includes non-numerical instances of quantification.

Table 32: Frequencies for 'there are Q' in the BNC, and roundness of their numerical complements

Quantifier	Total freq.	Single digit (%)	Non-round (%)	Degree 1 (%)	Degree 2 (%)	Degree 3 (%)	Degree 4 (%)
About	172*	12	20	18	16	22	12
Approximately	26	4	35	8	23	15	15
More than	115*	18	10	2	22	14	34
Fewer than	10	50	0	0	0	30	20
Less than	11	27	0	9	0	18	45
At least[31]	162*	70	10	2	6	4	8
At most	0	-	-	-	-	-	-

I hypothesised that these usages would disproportionately involve the use of round numbers. In order to obtain a baseline against which to assess this claim, we need to consider the roundness of numbers in general. Table 33 shows the extent to which the numbers from 1-100 exhibit roundness as defined and calculated above.

---

[31] In this case, 'at least' is more frequent than 'more than', contrary to the data presented in Table 28. Note however that we are restricting our attention to the existential 'There are Q', and it could be argued to make perfect sense from a numeral priming perspective that 'There are Q N' preferentially uses a Q that accommodates the possibility that 'There are exactly N'. However, I shall not attempt to argue this claim in detail here.

Table 33: Roundness distribution of numerals 1-100

Quantifier	Total freq.	Single digit (%)	Non-round (%)	Degree 1 (%)	Degree 2 (%)	Degree 3 (%)	Degree 4 (%)
- (numerals)	100	9	72	9	5	2	3

Across the set of quantifiers tested, the frequency of non-round numeral complements is much lower than would be expected if the numerals were selected at random. To quantify this, we note that 72 of the numbers 1-100 are non-round. If we take $p=0.72$ as the probability of selecting a non-round number, then – for a sample size of 50 – the 95% confidence interval on the number q of non-round numbers used is $29 \leq q \leq 42$). Across these quantifiers we examined, the greatest attested proportion of non-round numerical complements is 35%, corresponding to fewer than 20 such usages[32]. This clearly indicates a statistical preference for round numbers in cardinal contexts.

Hence, the numerical quantifiers examined here preferentially combine with round numbers, when we restrict our enquiry to cardinal uses. This supports the hypothesis under discussion. Given the global preference for round numbers already documented above, this is unsurprising; in the following sections, we attempt to refine this hypothesis further.

Prediction #5: The peaks of numeral usage at round numbers are higher for comparative quantifiers than for unmodified numerals.

Prediction #6: The peaks of numeral usage at round numbers are higher for superlative quantifiers than for unmodified numerals.

To test these predictions, we need to compare the preference for round numbers in the comparative and superlative cases, documented above, with the preference that we surmise

---

[32] In practice this is an oversimplification, as the assumption that each numeral from 1-100 is equally likely to be used is not realistic: smaller numerals are expected to occur with greater frequency. Nevertheless, assuming that this frequency term is in n^{-1}, the above argument should go through. Note that, excluding numerals 1-9 from consideration, the greatest frequency of non-round usage was 35/96 tokens = 36.5%. By contrast, across the range 11-100, entirely non-round numerals (11, 13, 17, 19, 21, 22, 23, 24, …) will still account for over 50% of predicted usage, even if we posit that frequency is inversely proportional to magnitude. Across a larger range, non-round numerals will account for a larger proportion still, as their distribution becomes denser. Lack of space precludes detailed discussion of what would constitute an optimal set of assumptions for the purpose of evaluating the hypothesis under discussion.

applies in the specific case of unmodified numerals. However, we have not yet attempted to quantify this latter preference: we have demonstrated a preference for round numbers across quantifying contexts in general, but not shown that this applies to unmodified numerals in particular. Therefore we need to establish data for this category of usage. Note that, if there is no preference for the use of round numbers in the case of unmodified numerals, then hypotheses 5 and 6 are true, as comparative quantifiers and superlative quantifiers have been shown preferentially to take round complements.

Paralleling the procedure performed in testing hypothesis 4, I proceed by examining the first 100 instances of 'there are' that take cardinal numerical complements, and counting how many of these complements exhibit each degree of roundness. In total, there are 39,955 instances of 'there are' in the BNC; a subset of 5000 was sampled from these. I examined the first 100 instances of 'there are' with immediate numerical complements in this subset (which required use of the first 971 tokens of this sample). The distribution of the roundness of the resulting complements was as shown in Table 34: for ease of comparison, the corresponding figures for comparative and superlative quantifiers are also repeated there.

Table 34: Frequencies for 'there are Q' in the BNC, and roundness of their numerical complements, including bare numeral case

Quantifier	Total freq.	Single digit (%)	Non-round (%)	Degree 1 (%)	Degree 2 (%)	Degree 3 (%)	Degree 4 (%)
- (none)	39,955*	78	11	7	1	1	2
More than	115*	18	10	2	22	14	34
Fewer than	10	50	0	0	0	30	20
Less than	11	27	0	9	0	18	45
At least	162*	70	10	2	6	4	8
At most	0	-	-	-	-	-	-

With respect to our first hypothesis, the bare numerals differ significantly in distribution from the comparative quantifiers. Notably, only 11% of the bare numeral sample consists of round numbers greater than 9, while the corresponding figure is 72% for 'more than', 50% for 'fewer than' and 63% for 'less than'. If we exclude the single digit numerals, then half of the

bare numeral sample is non-round, compared to only 12.2% of the 'more than' instances and none of the 'fewer than' and 'less than' instances. Therefore, these data support prediction #5: the preference for round numbers as complements for comparative quantifiers is indeed stronger than the preference for them in bare numeral contexts.

With respect to our second hypothesis, we have no data for 'at most' and therefore can only compare the bare numeral sample with our 'at least' data. By inspection, these data are quite similar in both cases, with 70% occurrence of single-digit numerals in the 'at least' case (compared to 78% in the bare numeral case). In the case of the remaining data, the distribution is again similar, although there is a slight preference for round numbers in the 'at least' case. It appears that the two distributions cannot be distinguished in a statistically significant way based on these samples. In fact, as there are only 162 counts of 'there are at most' in the BNC, it is likely to be impossible to distinguish the distributions even given a full survey of the corpus. Therefore, the corpus data do not support our final hypothesis: there is no evidence that the preference for round numbers as complements for superlative quantifiers is stronger than the preference for them in bare numeral contexts.

6.5. Discussion

In this chapter, I discuss six predictions derived from the constraint-based model, with particular reference to the constraints on quantifier simplicity and numeral salience. I then test these using the BNC, carefully restricting attention to the types of quantification about which each prediction is made. As a programme of study, this builds upon the work by Jansen and Pollmann (2001) by restricting its domain of enquiry to particular forms of numerical quantification.

Broadly, I follow Jansen and Pollmann (2001) i.a. in documenting a preference for round numbers. I further show that this applies in the case of cardinalities, a point that addresses a lacuna in Jansen and Pollmann's argumentation. I document a preference for simple quantifiers, using simplicity metrics that are *ad hoc* but that are well-founded in the literature. I further demonstrate an interaction between roundness and quantifier complexity in accordance with a novel prediction of the constraint-based model.

Five of the six hypotheses under test are borne out by these corpus data; for the sixth, the data are inconclusive. This is the claim concerning the precise behaviour of superlative quantifiers with round numbers. I surmise that this behaviour may also be influenced by

numeral priming, which is posited as a separate constraint in our model. It appears that, in corpora, superlative quantifiers are seldom used except where the numeral has particular significance (e.g. where it is a critical level for some criterion: 'if there are at least…', or when it represents best knowledge of a developing situation: 'at least 14 people were injured'). In terms of the constraint-based model, superlative quantifiers occur predominantly when numeral priming is being satisfied. This would account for why the distribution of the numerical complements of superlative quantifiers closely matches the distribution of the numerals in general: presumably the more often a numeral is mentioned, the more often it can exert a priming effect. However, as numeral priming is a faithfulness constraint, we cannot easily pursue this line of enquiry further through corpus study.

In conclusion, then, the corpus-testable predictions derived from the constraint-based model, under some plausible assumptions about the complexity of certain quantifiers, are borne out. I interpret this as support for the markedness constraints in the model (quantifier simplicity and numeral salience), and for the way in the model treats the interactions between constraints.

7. OVERVIEW AND OUTLOOK

In this concluding chapter, I briefly summarise the content of the preceding work, and then discuss some of the issues connected with potential future developments of the model proposed in this thesis.

7.1. The story so far

The goal of this thesis has been to propose an empirically-grounded account of the interpretation and use of numerically quantified expressions, to explain extant experimental findings in terms of this account, and to verify empirically some of the novel predictions arising from it.

In chapter 2, I motivate this approach by arguing that speaker behaviour is constrained by various competing considerations, and that no unified account of these has been offered that can yield testable predictions. With particular reference to numerical quantification, I suggest that the speaker's choice of expression can be treated as the solution to a problem of multiple constraint satisfaction, and thus modelled in a constraint-based framework such as Optimality Theory (OT). I specify the construction of such a model, and populate it with constraints motivated either by findings in the existing empirical literature or with intuitions supported by novel experimental data.

In chapter 3, I discuss how the proposed framework treats the interactions between these constraints, both under the assumptions of classical OT and under other sets of assumptions. Then, working broadly within the classical set of assumptions, I discuss how the model can be used to yield testable predictions, applying this to two simple examples of usage: one concerning the use of explicit approximation, as discussed by Krifka (2009), and one concerning priming effects in the correction of false and underinformative statements, relevant for the task employed by Katsos and colleagues (Katsos et al. 2011, i.a.).

In chapter 4, I move on to considering the case of comparative and superlative quantifiers. I offer an alternative to the modal account of superlative quantifier semantics proposed by Geurts and Nouwen (2007), arguing instead that superlative quantifiers possess additional representational complexity, and that this could yield pragmatic enrichments resulting in the meaning documented by Geurts et al. (2010). I document this representational complexity empirically. Having explored how this account serves to account for the existing data, I then embed this within the constraint-based model, showing that the quantifier simplicity

constraint predicts that superlative quantifiers will yield precisely these pragmatic enrichments. I argue that this approach is preferable in detail to the pragmatic account posited earlier in the chapter, in that it makes no unsupported assumptions about the internal structure of superlative quantifiers. Under this approach, the process of pragmatic enrichment undergone by the superlative quantifiers is seen as an example of the type of inference predicted by the constraint-based model, rather than requiring any specific theoretical treatment.

In chapter 5, I discuss the question of pragmatic enrichments arising from comparative and superlative quantifiers, such as to provide a bound on their interpretation. Accounts in the literature have argued that such expressions are immune from this type of scalar implicature; however, the constraint-based model predicts that these should arise under the appropriate conditions of numerical salience. In a series of experiments, these predictions are shown to be borne out, and hearers are shown to derive pragmatically restricted interpretations from such expressions. Moreover, in accordance with a prediction apparently unique to this constraint-based model, it is shown that the strength of this inference is attenuated if the numeral concerned is previously mentioned in the discourse context. This appears to constitute evidence in favour of the model's utility as a generator of predictions about quantifier usage and interpretation, and encourages related lines of speculation about the licensing conditions for certain categories of numerically quantified expression.

Finally, in chapter 6, I test additional predictions arising from this model concerning the distribution of round numerals. A preference for the use of these numbers is predicted by the numeral salience constraint in the model proposed in this thesis, but this is derived from the corpus study of Jansen and Pollmann (2001), which did not restrict its attention to cardinal uses of number. I show that this preference does apply to cardinal uses, and moreover that it is particularly manifest in the choice of numerical complements for comparative quantifiers. This bears out another prediction of the constraint-based model, and endorses the treatment of these quantifiers proposed in earlier chapters.

In the following section I consider the extent to which these findings bear out the constraint-based account, and the nature of the further work that would be required to prove or disprove the broad validity of this model.

7.2. Evidential basis for the constraint-based model

Although the experimental findings in this thesis are consistent with the constraint-based model, a question that arises is whether this constitutes strong evidence in favour of this approach. Over the course of the last few chapters, I show that this model is useful as a generator of predictions about numerical quantifier usage, and that at least some of these predictions are both non-obvious and experimentally supported. However, even allowing that the constraints are individually motivated and experimentally supported, their interaction could be treated within other models. Therefore, it could be argued that these results do not strongly favour the precise OT proposal discussed here, if we were to compare it with potential alternative accounts using different formalisms.

Testing the validity of this specific model would require a novel approach. I have argued that, under this account, individual differences between constraint rankings are predicted to underlie individual differences in usage preference. Therefore, the process for testing the approach could be sketched as follows.

- Obtain the constraint ranking for the individual speaker under investigation, by empirical means.

- Use this ranking to predict the speaker's behaviour in a range of elicitation contexts differing from those used to obtain the ranking.

- Collect the speaker's behavioural data in these contexts and assess the validity of the predictions.

From a classical OT perspective, obtaining a full constraint ranking for an individual speaker would be a difficult but not insuperable challenge. The ideal way to do this would be to establish that speaker's preference between pairs of constraints, by determining whether they prefer to satisfy constraint A or constraint B when all others are controlled for. Although a set of six constraints admits 720 possible rankings, there are only 21 pairwise comparisons that can be made between these constraints, so the speaker's ranking could in principle be completely determined through no more than 21 elicitation tasks. The resulting constraint ranking could then be used to predict the speaker's preference in more complex situations in which more than two constraints are simultaneously in play.

Nevertheless, this approach would require the development of a series of contexts in which only two constraints are relevant in the selection process, which may not be feasible for all pairs of constraints. If we adopt a more liberal approach in which we allow three constraints to be in effect for a given test item, the choice of outcome does not generally determine a ranking for all three constraints. It then becomes difficult to calculate precisely how many elicitation tasks are necessary to obtain a full constraint ranking, although this is clearly achievable in principle in no more than 21 tasks.

If we wish to establish the constraint rankings for a more liberal form of OT, such as stochastic OT, the above task becomes even more laborious. Under these conditions, we cannot reliably determine the relative ranking of two constraints based on a single observation. The precise number of such observations required to make that determination will depend upon how much variability we posit that the constraint ranking is permitted to exhibit. Furthermore, in this case, as in the other cases above, we need to consider the possibility of speaker error in their responses to individual elicitation tasks as we attempt to establish the constraint rankings.

In connection with the themes of variability and error, we also need to consider how the model is to be evaluated. We could not dismiss the model based merely on its failure to predict all data fully and accurately. In effect, we would wish to consider whether this model outperforms a baseline. However, as there does not appear to be a fully-specified general theory with which this model is in competition, the relevant baseline might vary from task to task. A sensible approach might be to ask how this model compares to a statistical model tailored to the task in hand, although of course the computation of this might itself be a laborious operation.

In this thesis, I have not attempted to establish complete constraint rankings for individuals. Whether this is a project worth undertaking, in practice, is an open question. If we are to consider the constraint-based model as an account of actual behaviour at a psycholinguistic level, we would need to demonstrate that such an approach is viable. However, if this model is merely to be considered as a descriptively adequate account of observable performance and a source of hypotheses about the interaction of the factors that influence the speaker's choice of utterance, this line of enquiry may be an unnecessary diversion.

Based upon the results presented in the previous chapters, I feel more confident in arguing for the model as a means of generating non-obvious predictions about the nature of numerical

quantifier usage, particularly given that these findings do not depend upon the accuracy of the model for their own validity. Hence, even though I entertain some scepticism towards the view that this model might be psychologically realistic, there remains a case for developing the model further, with a view to making it a better performance model, a theoretically sounder treatment of the effects of individual constraints, and a more accurate generator of predictions. Moreover, it may be worth considering whether this model can give rise to interesting predictions in other domains of usage. I discuss some of these issues in the following sections.

7.3. Informativeness and the nature of numerical representations

In all preceding discussions of the informativeness constraint, I have construed this in terms of excluding, versus failing to exclude, possibilities that the speaker knows not to be the case. This account was admittedly unsatisfactory in certain particulars, which are intuitively clear on closer scrutiny. Because it is stipulated that only quantifiers of the same type can compete with one another in informativeness, the constraint as defined in section 2.4.1 cannot account for the selection, for instance, of 'more than n' versus 'about p' versus 'between n and m'. To the extent that we are interested in constructing a psychologically realistic theory, it is also unsatisfactory that constraint violations may arise from the failure to convey information that the speaker does not knowingly consider relevant. For instance, 'more than 20' incurs violations by comparison with 'more than 30', but it is not necessarily the case that any of the possibilities excluded by the latter (21, 22, 23, …) are individually of interest to the speaker or the hearer. Introspectively, it seems more likely that 'more than 30' merely tends to convey a range of values that the speaker considers more appropriate for the communicative purpose than that conveyed by 'more than 20'.

The above limitations arise in part from the tacit assumption of an oversimplified landscape of probability in the mind of the speaker. In this account, each numeral is regarded simply as 'possible' or 'impossible' as a value for the quantity under discussion. More realistically, we might think of the speaker's attitude towards a quantity as being represented by a probability distribution over numerals. In this case, the certainty of a particular value would be represented as a distribution with probability 1 at this value and 0 elsewhere, while a typical uncertainty context would be represented as a normal distribution, or some perturbation of this, with total probability 1 across all numerals.

Against this landscape, we might instead evaluate informativeness in terms of the accuracy with which this distribution is communicated by the speaker. We might suppose that discourse participants adopt some default distribution of probabilities, given a particular quantity expression. Then the speaker can evaluate informativeness by computing the distance between the distribution they have in mind for the quantity under discussion and the distribution they consider to be default for the quantity expression that is being evaluated for use in the utterance. This could be done by summing the magnitudes of the difference in probabilities across all numeral values. In order for communication to be efficient, we would have to assume that the speaker's default distribution for a given expression closely matched that of the hearer, which should be a reasonable assumption given that both have acquired these distributions by exposure to similar data.

An account of this type would naturally encompass the data discussed in chapter 5 of this thesis. By default, the hearer would interpret 'more than 60' in accordance with their default distribution for this expression, which might naturally range from 61-80 or 61-70. This would follow from the assumption that, whatever the actual underlying distribution, 'more than 60' was among the best matches from the speaker's perspective. That is, the distribution was not one that would be best expressed by '60', 'about 70', or similar: and if there was a more suitable expression (say, 'more than 61'), it would be blocked by other considerations such as numeral salience. Given prior activation of the numeral '60', the speaker might be biased towards saying 'more than 60' even if some other utterance was a better match for the probability distribution they had in mind. The hearer would therefore have to adjust the strength of their inference accordingly. In short, the account would proceed along precisely the same lines as in chapter 5, but the broadened notion of informativeness would enable different types of numerical expression to be considered within the same tableau.

I should also add that this entire thesis has been concerned with the use of cardinal quantities, which in itself is an inevitable limitation given the original definition of informativeness. Under the more sophisticated definition proposed here, it should be possible to generalise the model to all categories of numerical quantification. It is impossible to evaluate the relative informativeness of 'more than 2 metres' versus 'more than 3 metres' using the metric introduced in section 2.4.1, because there are infinitely many possibilities included by the former and excluded by the latter. However, a probability distribution works perfectly well in the continuous case: the process described in the preceding paragraphs goes through in

exactly the same way, substituting integration for addition. Therefore an account of this type would enable us to take the major step of generalising away from the narrow discrete case.

While I believe this proposal has significant explanatory advantages over the version of the constraint used in this thesis, it is necessarily speculative. In particular, there does not appear to be any plausible way of demonstrating that the landscape of probabilities in the mind of a speaker (or a hearer) actually exists in anything like the form proposed here. As the probabilities are not consciously available to the individual, examining these experimentally is problematic given our current state of knowledge. In the absence of any empirical data concerning the probability distributions associated with expressions, the only way to obtain these is to stipulate them based on intuitions or introspection, which gives us arbitrarily many degrees of freedom and is clearly unsatisfactory from a modelling perspective. Nevertheless, as a thought experiment I feel that this approach has value and might contribute towards a more accurate picture of the way we represent quantities and indeed probabilistic information more generally.

7.4. Gradient priming effects

Just as numeral probabilities have hitherto been collapsed into the two categories 'possible' and 'impossible', so numerals and quantifiers in the discourse have been collapsed into the categories 'primed' and 'unprimed'. Again this intuitively seems to be an oversimplification. It seems quite possible in principle that both numerals and quantifiers might exhibit some measure of priming effects on the strength of prior mention or activation in the discourse by some other means, and that these effects might vary in intensity between zero and some maximal level.

Adopting a similar line of argument to that taken in the preceding section, it would be possible to construe a pattern of activation as prevailing over all numerals and quantifiers. Then the level of violation of a priming constraint could again be measured by the distance between the pattern of the proposed utterance and the existing priming profile.

In the applications considered so far, I have also assumed that the utterances either fully satisfy a given priming preference or fully fail to do so. This assumption could also be relaxed. In the case of quantifiers, we could posit that quantifiers with a particular entailment direction partially satisfy priming requirements for all other quantifiers of that entailment direction: so, for instance, if 'more than' is primed, 'at least' violates quantifier priming to a

lesser degree than 'about'. This opens up another possible way to account for the problematic case of choosing between a single-bounded and double-bounded quantifier, as discussed in section 2.4.1. It remains to be seen whether such an account is well-founded experimentally: it might be possible to document conventional priming effects between certain quantifiers, although I have not made any attempt to do so as yet.

The case of numeral priming is potentially more interesting in this regard, given the way in which large numbers are constructed by concatenating smaller ones. It seems plausible that a numeral might exert some priming effect in this context: (102) is an artificial example.

(102)   A. We can sell two hundred tickets.
        B. More like two thousand.

We could think of an example like this as one in which A's utterance primes both 'two' and 'hundred'. B's utterance then constitutes a partial satisfaction of numeral priming, matching the small numeral but not the large one. Again, this is a speculative notion in that I have no empirical data concerning the priming relationships between distinct numerals of this kind. However, in principle, relationships of this type could be empirically verified (if they exist). This extension would further broaden the explanatory power of the model by permitting numeral priming effects to manifest themselves across different orders of magnitude.

7.5. Extension to other domains of usage

Given that the model under discussion uses functionally motivated constraints, and therefore is intended to be founded upon fundamental properties of human interaction, it seems natural to ask – why should this model be restricted to handling numerical quantification? If it works at all, should it not have broader applicability to natural language?

The domain of numerical quantification appears a promising testbed for a model of this type, inasmuch as it is possible to posit easy ways to quantify the extent of constraint violation. The naïve version of informativeness proposed in section 2.4.1 is a case in point. We can readily see how 'more than 21' is more informative than 'more than 20' and less informative than 'more than 22'; it is far less clear how we could compare the informativeness of items such as 'Jane is tall' versus 'Jane is blonde'.

Nevertheless, broadly I would argue that such an approach ought to be feasible, and indeed that some kind of interdisciplinary and quantitative approach is essentially necessary in

achieving a full explanation of language usage and interpretation. I do not wish to argue that extending a constraint-based account of the type described here is necessarily the best way of going about this, but it does have certain features that would appear to be desiderata for such a proposal. For one thing, it is speaker-oriented: it is crucial that the speaker's decision can be accounted for wholly in terms of knowledge that the speaker possesses. For another, it is based upon factors that are individually shown to be contributory to utterance selection. Finally, it is tractable from the speaker's and the hearer's perspective, although this is not an issue to which I have devoted specific attention in this thesis.

We could generalise this model to a certain extent by broadening the definitions of some of the constraints. Numeral priming could be regarded as a special case of a more general lexical priming constraint, and correspondingly quantifier priming could be thought of as a species of syntactic priming. Similarly, numeral salience seems to have a clear lexical analogue, and quantifier salience could be identified with a constraint prohibiting syntactic complexity. The remaining pair of constraints, informativeness and granularity, would then correspond to constraints over the content that is to be conveyed by the speaker. Granularity would, for instance, require the speaker to make predications that were of the appropriate category level. Informativeness would, for instance, require that the speaker provide sufficient detail to uniquely identify referents.

Given the complexity of such a system, I shall not attempt to develop the above sketch further within this thesis. Notably, there does not appear to be a sufficiently well-established body of research into the necessary individual constraints to enable such a system to be built. As discussed, the numeral domain is perhaps simpler and already well-studied, permitting a numeral salience constraint to be spelled out with relative precision. The related domain of granularity appears similarly tractable, and the rich entailment relations provide ready access to some workable definition of informativeness (*pace* section 7.3). It has also been possible for researchers to establish the relative complexity of the members of the small closed class of quantifiers. In sum, a similar formalism to that proposed here could straightforwardly be stated for language use in general, but the possibility of deriving clear, testable and substantial predictions from it appears a distant prospect at present.

Nevertheless, I feel that the preceding chapters go some way towards illustrating the usefulness of a model of usage and interpretation based upon multiple constraint satisfaction. This approach provides a quantitative means by which utterances can be evaluated and

pragmatic enrichments drawn, and appears in at least some cases to approximate the results and perhaps even the methods employed by speakers in selecting utterances and hearers in interpreting them. Such an approach, applied to more general linguistic settings, might one day prove equally fruitful in generating testable predictions about speaker behaviour and thus throwing additional light on the nature of language use.

BIBLIOGRAPHY

Almor, A. and Nair, V. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass*, 1: 84-99.

Arnold, J. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23: 495-527.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4: 159-219.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17: 189-216.

Blutner, R. (2006). Embedded implicatures and optimality theoretic pragmatics. In T. Solstad, A. Grønn and D. Haug (eds.), *A Festschrift for Kjell Johan Sæbø: in partial fulfilment of the requirements for the celebration of his 50th birthday*. Oslo.

Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences*, 21: 43-58.

Branigan, H. P., Pickering, M. J. and Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, 75: B13-25.

Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, 25(2): 93-139.

Bresnan, J. and Aissen, J. (2002). Optimality and functionality: objections and refutations. *Natural Language and Linguistic Theory*, 20: 81-95.

Breheny, R., Katsos, N. and Williams, J. (2006). Are scalar implicatures generated by default? *Cognition*, 100: 434-63.

Bultinck, B. (2005). *Numerous meanings: the meaning of English cardinals and the legacy of Paul Grice*. London: Elsevier.

Büring, D. (2007). The least 'at least' can do. In C. B. Chang and H. J. Haynie (eds.), *Proceedings of WCCFL 26*. Somerville, MA: Cascadilla Press. 114-120.

Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.

Carston, R. (1998). Informativeness, relevance, and scalar implicature. In R. Carston and S. Uchida (eds.), *Relevance theory: applications and implications*. Amsterdam: Benjamins. 179-236.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In Belletti, A. (ed.), *Structures and Beyond*, Oxford: Oxford University Press. 39-103.

Cummins, C. and Katsos, N. (2010). Comparative and superlative quantifiers: pragmatic effects of comparison type. *Journal of Semantics*, 27: 271-305.

Cummins, C., Sauerland, U. and Solt, S. (submitted). Granularity and scalar implicature in numerical expressions.

Davies, C. and Katsos, N. (2010). Over-informative children: production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, 120 (Special Issue on Asymmetries in Child Language): 1956-72.

de Graaff-Hunter, J. (1955). Various determinations over a century of the height of Mount Everest. *Geographical Journal*, 121: 21-6.

Dehaene, S. (1997). *The Number Sense*. New York: Oxford University Press.

Dekker, P. and van Rooij, R. (2000). Bi-directional Optimality Theory: an application of game theory. *Journal of Semantics*, 17: 217-42.

Engelhardt, P. E., Bailey, K. G. D. and Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54: 554-73.

Evans, J. St. B. T., Newstead, S. and Byrne, R. M. J. (1993). *Human Reasoning: the Psychology of Deduction*. Hove: Psychology Press.

Fox, D. and Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy*, 29: 537-86.

Garagnani, M., Wennekers, T. and Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27: 492-513.

Gennari, S. P. and MacDonald, M. C. (2006). Acquisition of negation and quantification: insights from adult production and comprehension. *Language Acquisition*, 13: 125-68.

Geurts, B. (2006). Take 'five': the meaning and use of a number word. In Vogeleer, S. and Tasmowski, L., *Non-definiteness and Plurality*. Amsterdam: John Benjamins. 311-30.

Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind and Language*, 24: 51-79.

Geurts, B., Katsos, N., Cummins, C., Moons, J. and Noordman, L. (2010). Scalar quantifiers: logic, acquisition, and processing. *Language and Cognitive Processes*, 25: 130-48.

Geurts, B. and Nouwen, R. (2007). 'At least' et al.: the semantics of scalar modifiers. *Language*, 83: 533-59.

Gibbs, R. W. and Bryant, G. A. (2008). Striving for optimal relevance when answering questions. *Cognition*, 106: 345-69.

Givón, T. (1979). *On Understanding Grammar*. New York: Academic Press.

Gordon, P., Grosz, B. and Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17: 311-47.

Gregg, K. R. (2003). The state of emergentism in second language acquisition. *Second Language Research*, 19: 95-128.

Grice, H. P. (1975). Logic and Conversation. In Cole, P. and Morgan, J. L. (eds.), *Syntax and Semantics*, Vol. 3. New York: Academic Press. 41-58.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Gundel, J., Hedberg, N. and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language and Cognitive Processes*, 69: 274-307.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics*, 17: 63-98.

Hanna, J. E., Tanenhaus, M. K. and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49: 43-61.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hendriks, P. and de Hoop, H. (2001). Optimality Theoretic semantics. *Linguistics and Philosophy*, 24: 1-32.

Horn, L. R. (1972). On the semantic properties of logical operators in English. UCLA dissertation, distributed by Indiana University Linguistics Club, 1976.

Horn, L. R. (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Schiffrin, D. (ed.), *Meaning, Form and Use in Context (GURT '84)*. Washington DC: Georgetown University Press. 11-42.

Horn, L. R. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, 61(1): 121-74.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Horton, W. S. and Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59: 91-117.

Jansen, C. J. M. and Pollmann, M. M. W. (2001). On round numbers: pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics*, 8: 187-201.

Just, M. A. and Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, 10: 244-53.

Katsos, N. (2007). *Experimental investigations on the effects of structure and context on the generation of scalar implicatures*. PhD thesis, University of Cambridge.

Katsos, N. (2008). The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature, *Synthese*, 165: 358-401

Katsos, N., Andrés Roqueta, C., Estevan, R. A. C. and Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, 119: 43-57.

Katsos, N. and Bishop, D. V. M. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition*, 120: 67-81.

Katsos, N. and Cummins, C. (2010). Pragmatics: from theory to experiment and back again. *Language and Linguistics Compass*, 4/5: 282-95.

Katsos, N. and Smith, N. (2010). Pragmatic tolerance and speaker-comprehender asymmetries. In Franich, K., Iserman, K. M. and Keil, L. L. (eds.), *Proceedings of the 34*[th] *Annual Boston Conference in Language Development*. Somerville, MA: Cascadilla Press. 221-32.

Krahmer, E., van Erk, S. and Verleg, A. (2003). Graph-based generation of referring expressions. *Computational Linguistics*, 29(1): 53-72.

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis*, 25: 209-57.

Krifka, M. (1999). At least some determiners aren't determiners. In K. Turner (ed.), *The Semantics/Pragmatics Interface from Different Points of View*, Current Research in the Semantics/Pragmatics Interface Vol. 1. Oxford: Elsevier. 257-92.

Krifka, M. (2009). Approximate interpretations of number words: a case for strategic communication. In Hinrichs, E. and Nerbonne, J. (eds.), *Theory and Evidence in Semantics*. Stanford: CSLI Publications. 109-132.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: CUP.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8: 339-59.

McCarthy, J. J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: CUP.

McCarthy, J. J. and Prince, A. (1994). The emergence of the unmarked: optimality in prosodic morphology. In González, M. (ed.), *Proceedings of the North East Linguistics Society 24*, Amherst, MA: GLSA. 333-79.

McMillan, C., Clark, R., Moore, P., Devita, C. and Grossman, M. (2005). Neural basis for generalized quantifiers comprehension. *Neuropsychologia*, 43: 1729-37.

McMillan, C., Clark, R., Moore, P. and Grossman, M. (2006). Quantifiers comprehension in corticobasal degeneration. *Brain and Cognition*, 65: 250-60.

Montague, R. (1970). Universal grammar. *Theoria*, 36: 373-98.

Moxey, L. M. and Sanford, A. J. (2000). Communicating quantities: a review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, 14: 237-55.

Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition*, 93: 1-41.

Nadig, A. S. and Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13(4): 329-36.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4: 39-40.

Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics*, 3: 1-41.

Noveck, I. A. and Reboul, A. (2008). Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12: 425-31.

OED (1989). *Oxford English Dictionary*, 2nd ed. Oxford: Oxford University Press.

Partee, B. (1986). Noun-phrase interpretation and type-shifting principles. In Groenendijk, J., de Jong, D. and Stokhof, M. (eds.), *Studies in discourse representation theory and the theory of generalized quantifiers*. Dordrecht: Reidel. 115-44.

Pickering, M. J., and Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27: 169-226.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. In Yoon, J.-H. and Kathol, A. (eds.), *OSUWPL Volume 49: Papers in Semantics*. Columbus, OH: Ohio State University Department of Linguistics.

Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, 32: 3-23.

Smolensky, P. (1986). Information processing in dynamical systems: foundations of harmony theory. In Rumelhart, D. E., McClelland, J. L. and the PDP Research Group, *Parallel Distributed Processing: Explorations in the microstructure of cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press/Bradford Books. 194-281.

Smolensky, P. (1995). On the structure of the constraint component CON of UG. Rutgers Optimality Archive, ROA-86.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 27: 720-31.

Solt, S. (2010). On the expression of proportion: most and more than half. Presentation at 84th Annual Meeting of the Linguistic Society of America, Baltimore, MD.

Sperber, D. and Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43: 155-67.

Szymanik, J. and Zajenkowski, M. (2010). Comprehension of simple quantifiers: empirical evaluation of a computational model. *Cognitive Science*, 34: 521-32.

van Deemter, K. (2006). Generating referring expressions that involve gradable properties. *Computational Linguistics*, 32: 195-222.

Van der Henst, J. B., Carles, L. and Sperber, D. (2002). Truthfulness and relevance in telling the time. *Mind and Language*, 17: 457-66.

Van der Henst, J. B. and Sperber, D. (2004). Testing the cognitive and communicative principles of relevance. In Noveck, I. and Sperber, D. (eds.), *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan. 141-69.

Viethen, J. and Dale, R. (2008). Generating relational references: what makes a difference? In *Proceedings of the 2008 Australasian Language Technology Workshop*. Hobart: CSIRO Tasmania ICT Centre. 160-8.

Wilson, D. and Sperber, D. (2002). Truthfulness and relevance. *Mind*, 111: 583-632.

Zondervan, A. (2007). Effects of Question Under Discussion and focus on scalar implicatures. In Kluck, M. E. and Smits, E. J. (eds.), *Proceedings of the Fifth Semantics in the Netherlands Day (SiN V).* 39-52.

Appendix A. Sample materials for Experiment 1 (section 2.4.5.2)

Figure 6: Visual display for five-item case (cars)

Figure 7: Visual display for two-item case (balls)
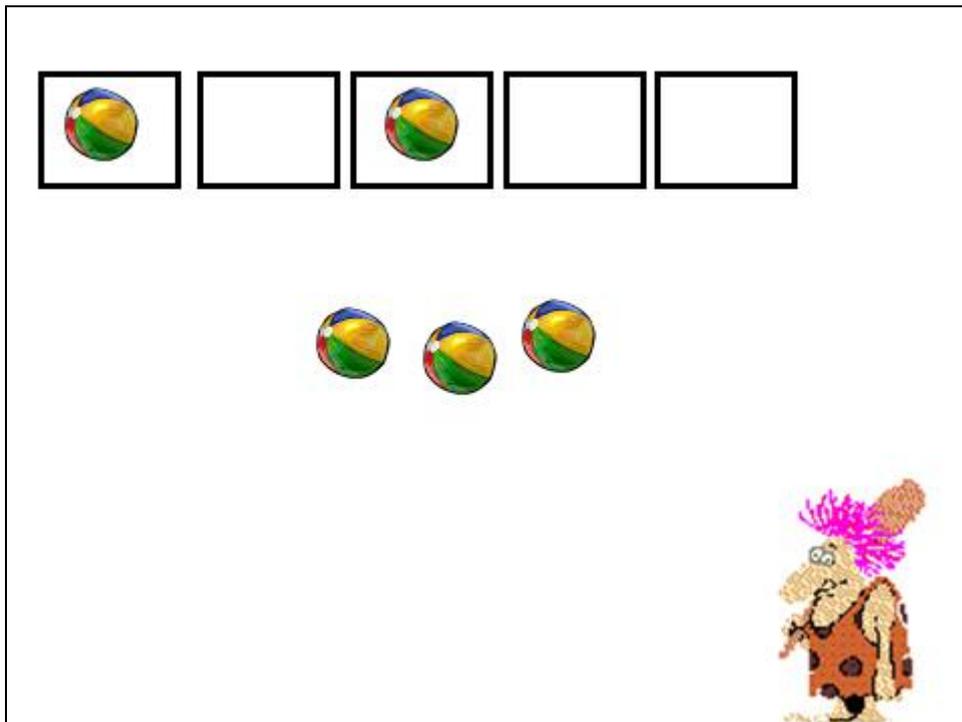
Figure 8: Visual display for no-item case (pens)
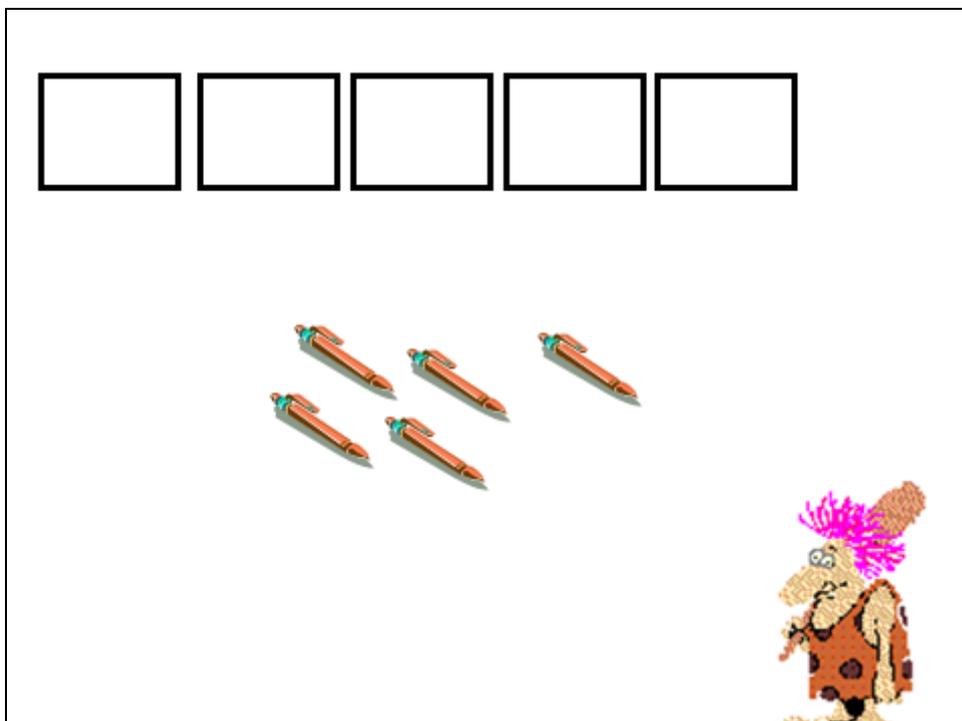
Appendix B. Sample materials for Experiment 2 (section 3.3.2.1)

Figure 9: Display for 'There are Q shoes in each box', n=4
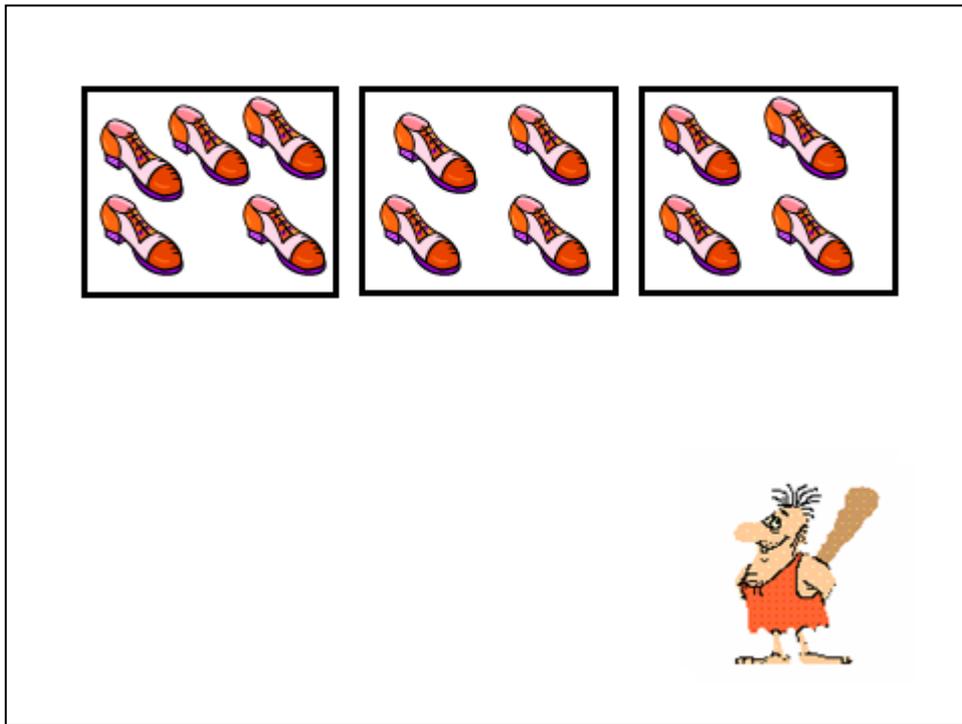
Figure 10: Display for 'There are Q clocks in each box', *n*=2
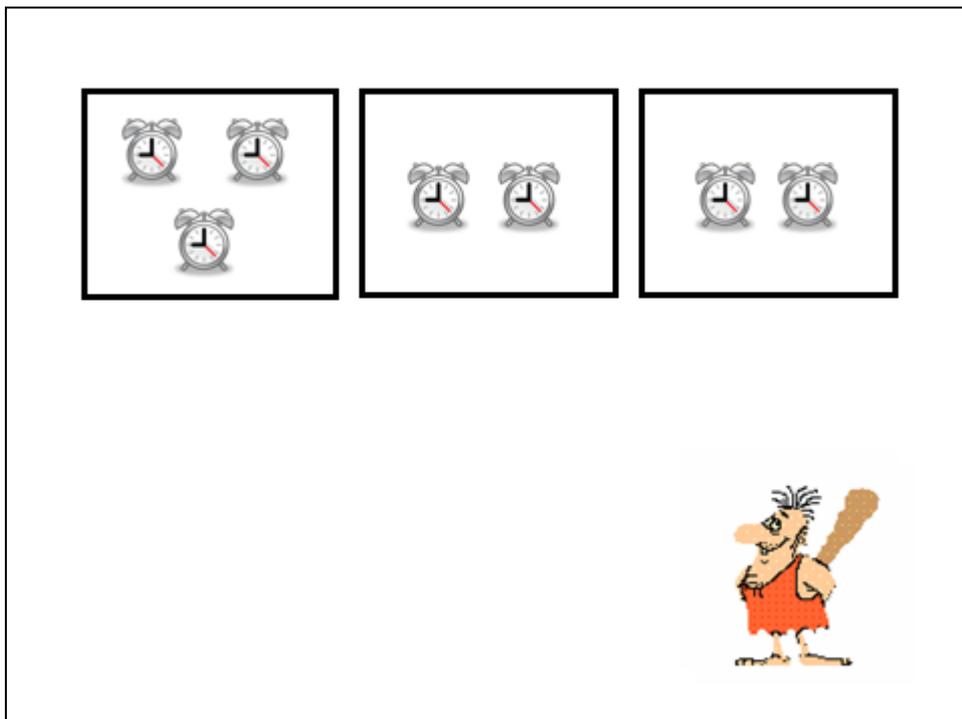
Appendix C. Test conditions for Experiment 3 (section 4.6.1)

Prompt	Display	Correct response
A > 2	A	False
	AA	False
	AAA	True
	AAAA	True
A ≥ 3	A	False
	AA	False
	AAA	True
	AAAA	True
A < 3	A	True
	AA	True
	AAA	False
	AAAA	False
A ≤ 2	A	True
	AA	True
	AAA	False
	AAAA	False
A = 3	AA	False
	AAA	True
	AAAA	False
B > 2	B	False
	BB	False
	BBB	True
	BBBB	True
B ≥ 3	B	False
	BB	False
	BBB	True
	BBBB	True
B < 3	B	True
	BB	True
	BBB	False
	BBBB	False
B ≤ 2	B	True
	BB	True
	BBB	False
	BBBB	False
B = 3	BB	False
	BBB	True
	BBBB	False

Appendix D. Materials for Experiment 4 (section 4.8.1)

Antecedent	Consequent
The shopper took 3 carrier bags.	The shopper took at least 3 carrier bags.
The tennis player served 3 aces.	The tennis player served fewer than 4 aces.
Anna wrote at least 3 letters.	Anna wrote 3 letters.
The artist sketched fewer than 3 landscapes.	The artist sketched more than 3 landscapes.
There are 3 parks in the area.	There are more than 2 parks in the area.
Ellen likes at most 2 bands.	Ellen likes at most 3 bands.
There are at most 3 cards on the table.	There are 3 cards on the table.
There are 3 reasons to attend the talk.	There are fewer than 3 reasons to attend the talk.
The student attends 3 lectures.	The student attends at most 3 lectures.
The grocer stocks fewer than 3 brands of coffee.	The grocer stocks fewer than 4 brands of coffee.
The room has more than 3 windows.	The room has fewer than 3 windows.
Craig pours 3 cups of tea.	Craig pours more than 3 cups of tea.
There are fewer than 3 ornaments on the shelf.	There are more than 3 ornaments on the shelf.
There are at least 3 cities on the map.	There are 3 cities on the map.
There are 3 ways to get there.	There are fewer than 4 ways to get there.
There are 3 books on the shelf.	There are at least 3 books on the shelf.
The gardener plants 3 bushes.	The gardener plants fewer than 3 bushes.
The director manages at most 3 employees.	The director manages 3 employees.

Antecedent	Consequent
The tourist chose at most 2 postcards.	The tourist chose at most 3 postcards.
The company rents 3 offices.	The company rents more than 2 offices.
The farmer ploughed 3 fields.	The farmer ploughed more than 3 fields.
The jogger ran more than 3 laps of the track.	The jogger ran fewer than 3 laps of the track.
The guitarist played fewer than 3 songs.	The guitarist played fewer than 4 songs.
The architect designs 3 houses.	The architect designs at most 3 houses.
Steve owns at least 3 suits.	Steve owns 3 suits.
There are more than 3 websites on the subject.	There are fewer than 3 websites on the subject.
Mary told 3 stories.	Mary told fewer than 3 stories.
Jane has 3 children.	Jane has at least 3 children.
There are 3 candidates in the election.	There are at most 3 candidates in the election.
There are at most 2 buttons on the jacket.	There are at most 3 buttons on the jacket.
There are fewer than 3 people in the meeting.	There are fewer than 4 people in the meeting.
Bob ate at most 3 biscuits.	Bob ate 3 biscuits.
The waitress carries fewer than 3 glasses.	The waitress carries more than 3 glasses.
There are 3 stripes on the blouse.	There are more than 3 stripes on the blouse.
Ed bought 3 CDs.	Ed bought more than 2 CDs.
The family owns 3 cars.	The family owns fewer than 4 cars.

Appendix E. Materials for Experiment 5 (section 4.8.2)

1.	Amy speaks at most three languages. In fact, she speaks exactly three.

2.	Joan does at most five jobs. In fact, she does exactly five.

3.	Mary does at most four jobs. Specifically, she does exactly four.

4.	Ian knows at least three celebrities. Specifically, he knows exactly four.

5.	Ian writes more than five papers. In fact, he writes exactly four.

6.	Mary wins some of her matches. In fact, she wins none of them.

7.	Sophie makes some of her dresses. Specifically, she makes half of them.

8.	Sue writes more than four papers. Specifically, she writes exactly three.

9.	Brian has at most four children. In fact, he had exactly four.

10.	Ian writes at least four papers. In fact, he writes exactly four.

11.	Michael sends more than four texts. In fact, he sends exactly three.

12.	John drives at most three cars. Specifically, he drives exactly two.

13.	Brian has at least three children. In fact, he has exactly four.

14.	Amy trusts some of her friends. Specifically, she trusts half of them.

15.	Amy speaks at least five languages. In fact, she speaks exactly four.

16.	James plays at most five instruments. In fact, he plays exactly six.

17.	Jane owns at least five DVDs. In fact, she owns exactly five.

18.	John trusts some of his friends. In fact, he trusts all of them.

19.	Sophie makes some of her dresses. Specifically, she makes all of them.

20.	Sue drives at least three cars. In fact, she drives exactly three.

21.	Mary wins some of her matches. In fact, she wins half of them.

22.	Michael speaks at least four languages. Specifically, he speaks exactly three.

23.	Brian likes at most three operas. Specifically, he likes exactly four.

24.	Jane has fewer than four children. Specifically, she has exactly five.

25.	Ben sends at least four texts. Specifically, he sends exactly five.

26.	James wins some of his matches. Specifically, he wins half of them.

27.	Brian likes at least five operas. Specifically, he likes exactly four.

28.	Mary wins some of her matches. In fact, she wins all of them.

29.	Joan owns more than five DVDs. Specifically, she owns exactly four.

30.	Sophie knows at least four celebrities. In fact, she knows exactly five.

31.	Sue drives at most four cars. In fact, she drives exactly three.

32.	Mary does fewer than five jobs. Specifically, she does exactly six.

33.   Ben sends more than three texts. Specifically, he sends exactly two.

34.   Jane has at least five children. Specifically, she has exactly six.

35.   Sue writes at most five papers. Specifically, she writes exactly six.

36.   John trusts some of his friends. In fact, he trusts half of them.

37.   Joan does fewer than three jobs. In fact, she does exactly four.

38.   Clare makes some of her dresses. In fact, she makes half of them.

39.   Sue writes at least three papers. Specifically, she writes exactly three.

40.   Joan owns at least four DVDs. Specifically, she owns exactly four.

41.   Joan does at least four jobs. In fact, she does exactly three.

42.   Michael speaks at least four languages. Specifically, he speaks exactly three.

43.   Sue drives fewer than five cars. In fact, she drives exactly four.

44.   Clare plays at most four instruments. Specifically, she plays exactly five.

45.   James plays more than three instruments. In fact, he plays exactly four.

46.   Michael sends at least five texts. In fact, he sends exactly six.

47.   Mary does at least three jobs. Specifically, she does exactly two.

48.   Amy trusts some of her friends. Specifically, she trusts none of them.

49.   Brian has fewer than five children. In fact, he has exactly six.

50.   Ian knows at most four celebrities. Specifically, he knows exactly three.

51.   Ian writes at most three papers. In fact, he writes exactly four.

52.   James wins some of his matches. Specifically, he wins all of them.

53.   Clare plays fewer than three instruments. Specifically, she plays exactly four.

54.   Sophie knows at most five celebrities. In fact, she knows exactly four.

55.   Michael sends at most three texts. In fact, he sends exactly two.

56.   James plays fewer than four instruments. In fact, he plays exactly five.

57.   Ian knows fewer than five celebrities. Specifically, he knows exactly four.

58.   Sophie knows fewer than three celebrities. In fact, she knows exactly two.

59.   Jane has at most three children. Specifically, she has exactly three.

60.   Clare plays more than five instruments. Specifically, she plays exactly six.

61.   John drives at least five cars. Specifically, he drives exactly five.

62.   Clare makes some of her dresses. In fact, she makes none of them.

63.   Sophie makes some of her dresses. Specifically, she makes none of them.

64.   Ben likes more than five operas. In fact, he likes exactly six.

65.   Michael speaks at most five languages. Specifically, he speaks exactly five.

66.   Jane owns fewer than four DVDs. In fact, she owns exactly three.

67. James wins some of his matches. Specifically, he wins none of them.

68. Brian likes more than four operas. Specifically, he likes exactly five.

69. Ben likes at least three operas. In fact, he likes exactly two.

70. Joan owns fewer than three DVDs. Specifically, she owns exactly two.

71. Clare makes some of her dresses. In fact, she makes all of them.

72. John drives fewer than four cars. Specifically, he drives exactly three.

73. Ben sends at most five texts. Specifically, he sends exactly four.

74. Amy trusts some of her friends. Specifically, she trusts all of them.

75. Ben likes at most four operas. In fact, he likes exactly five.

76. John trusts some of his friends. In fact, he trusts none of them.

77. Jane owns more than three DVDs. In fact, she owns exactly two.

78. Amy speaks at least five languages. In fact, she speaks exactly four.

Appendix F. Materials for Experiment 6 (section 4.8.3)

1. 'If the shopper took at least 3 carrier bags, they will get a discount. The shopper took 3 carrier bags.'

    Does the speaker think the shopper will get a discount?

2. 'If the tennis player served fewer than 4 aces, they will be disappointed. The tennis player served 3 aces.'

    Does the speaker think the tennis player will be disappointed?

3. 'If Anna wrote 3 letters, she will be satisfied. Anna wrote at least 3 letters.'

    Does the speaker think Anna will be satisfied?

4. 'If the artist sketched more than 3 landscapes, he will stop working. The artist sketched fewer than 3 landscapes.'

    Does the speaker think the artist will stop working?

5. 'If there are more than 2 parks in the area, the residents will be happy. There are 3 parks in the area.'

    Does the speaker think the residents will be happy?

6. 'If Ellen likes at most 3 bands, she probably prefers films to music. Ellen likes at most 2 bands.'

    Does the speaker think Ellen probably prefers films to music?

7. 'If there 3 cards on the table, the trick can begin. There are at most 3 cards on the table.'

    Does the speaker think the trick can begin?

8. 'If there are fewer than 3 reasons to attend the talk, I won't go. There are 3 reasons to attend the talk.'

    Does the speaker plan to attend the talk?

9. 'If there are at most 3 buttons on the jacket, it will look fine. There are at most 2 buttons on the jacket.'

Does the speaker think the jacket will look fine?

10. 'If the student attends at most 3 lectures, they will have plenty of free time. The student attends 3 lectures.'

Does the speaker think the student has plenty of free time?

11. 'If the grocer stocks fewer than 4 brands of coffee, I'll go to the supermarket. The grocer stocks fewer than 3 brands of coffee.'

Does the speaker intend to go to the supermarket?

12. 'If the room has fewer than 3 windows, it will be too dark. The room has more than 3 windows.'

Does the speaker think the room will be too dark?

13. 'If there are 3 candidates in the election, it is difficult to choose. There are more than 3 candidates in the election.'

Does the speaker think it is difficult to choose?

14. 'If the tourist chose at most 3 postcards, she will have cash to spare. The tourist chose at most 2 postcards.'

Does the speaker think the tourist will have cash to spare?

Appendix G. Materials for Experiment 7 (section 4.8.4)

Antecedent	Consequent
Jane has five cats but Eric has more than five cats.	Jane and Eric each have more than five cats.
Elaine has five children but Yvonne has at most four children.	Elaine and Yvonne each have at most five children.
Colin has four houses but Lucy has fewer than four houses.	Colin and Lucy each have at most three houses.
Elaine has five watches but Yvonne has more than five watches.	Elaine and Yvonne each have more than four watches.
Mary has five apples but Amy has fewer than five apples.	Mary and Amy each have fewer than five apples.
John has five children but Anne has at least six children.	John and Anne each have at least six children.
Dave has three apples but Richard has at most three apples.	Dave and Richard each have fewer than four apples.
Brian has five apples but Tom has fewer than five apples.	Brian and Tom each have fewer than six apples.
Sam has five suits but Judy has at least six suits.	Sam and Judy each have at least five suits.
Sam has five children but Judy has at least five children.	Sam and Judy each have more than five children.
Mary has five houses but Amy has at most four houses.	Mary and Amy each have at most four houses.
John has three watches but Ann has more than three watches.	John and Anne each have at least three watches.

Antecedent	Consequent
Mary has four apples but Amy has at least five apples.	Mary and Amy each have at least four apples.
Jane has three assistants but Eric has fewer than three assistants.	Jane and Eric each have fewer than four assistants.
John has four cats but Anne has fewer than four cats.	John and Anne each have fewer than four cats.
Mary has five suits but Amy has at most five suits.	Mary and Amy each have fewer than five suits.
Dave has three suits but Richard has more than three suits.	Dave and Richard each have more than two suits.
Brian has four suits but Tom has more than four suits.	Brian and Tom each have more than four suits.
Dave has three cats but Richard has at most two cats.	Dave and Richard each have at most two cats.
Colin has four hats but Lucy has fewer than four hats.	Colin and Lucy each have fewer than five hats.
Brian has five assistants but Tom has fewer than five assistants.	Brian and Tom each have at most five assistants.
Sam has three houses but Judy has at least four houses.	Sam and Judy each have at least four houses.
Sam has four hats but Judy has at most three hats.	Sam and Judy each have at most four hats.
Elaine has four cats but Yvonne has more than four cats.	Elaine and Yvonne each have at least five cats.

Antecedent	Consequent
John has three apples but Anne has fewer than three apples.	John and Anne each have fewer than three apples.
Brian has four watches but Tom has at most three watches.	Brian and Tom each have at most three watches.
Jane has three watches but Eric has at most two watches.	Jane and Eric each have at most three watches.
Colin has three hats but Lucy has more than three hats.	Colin and Lucy each have more than three hats.
Jane has four hats but Eric has at least four hats.	Jane and Eric each have more than three hats.
Dave has four children but Richard has at least five children.	Dave and Richard each have at least five children.
Colin has three houses but Lucy has at least four houses.	Colin and Lucy each have at least three houses.
Elaine has four assistants but Yvonne has more than four assistants.	Elaine and Yvonne each have more than four assistants.

Appendix H. Materials for Experiment 9 (section 5.4.2)

Version 1

Please read the following short dialogues, and answer the questions by filling in a value for each blank space, according to your opinion. Consider each dialogue separately. Assume that participant B is well-informed, telling the truth, and being co-operative in each case.

1.   A: We need to sell 60 tickets to cover our costs. How are the ticket sales going?
        B: So far, we've sold fewer than 60 tickets.

How many tickets have been sold? From …… to ……, most likely …….

2.   A: To win the election, the candidate had to convince people to vote for him. How many votes did he get?
        B: He got more than 77 votes.

How many votes did the candidate get? From …… to ……, most likely …….

3.   A: This photo album has space for 150 4x6 photos. How many photos of that size do you have?
        B: I have about 150 photos.

How many photos does B have? From …… to ……, most likely ……

4   A: We've invited 80 members and non-members to the reception. How much room is there in the hall?
        B: There's room for more than 80 people.

How many people is there room for in the hall? From …… to ……, most likely …….

5.   A: There are still copies of the agenda left. How many delegates are we waiting for?
        B: We're waiting for more than 200 delegates.

How many delegates are they waiting for? From …… to ……, most likely …….

6.	A: This display case holds CDs. How many CDs do you own?

	B: I own fewer than 80 CDs.

How many CDs does B own? From …… to ……, most likely …….

7.	A: The lecture hall can accommodate students taking an exam. How many students

	will be taking tomorrow's exam?

	B: About 130 students will be taking tomorrow's exam.

How many students will be taking the exam? From …… to ……, most likely …….

8.	A: I expect 93 cars to arrive between now and three o'clock. How many spaces are

	left in the car park?

	B: There are more than 93 spaces left in the car park.

How many spaces are left in the car park? From …… to ……, most likely …….

9.	A: We wanted 200 new signatures on the petition. How many new people signed it?

	B: Fewer than 200 new people signed it.

How many new people signed the petition? From …… to ……, most likely …….

10.	A: The new prison still has space for 110 prisoners. How many prisoners need to be

	transferred out of the old facility?

	B: About 110 prisoners need to be transferred.

How many prisoners need to be transferred? From …… to ……, most likely …….

11.	A: People have already paid their deposit for this holiday. How many seats are there

	on the plane?

	B: There are more than 60 seats on the plane.

How many seats are there on the plane? From …… to ……, most likely …….

12.　　A: We can hire a bus with 77 reclining seats for the excursion. How many people will be coming?

　　　　B: Fewer than 77 people are coming.

How many people are coming on the excursion? From …… to ……, most likely …….

13.　　A: We have space in the garden to plant tulips. How many tulip bulbs are in the bag?

　　　　B: There are about 50 tulip bulbs in the bag.

How many tulip bulbs are in the bag? From …… to ……, most likely …….

14.　　A: There are free tickets available for people who call the station today. How many people have called so far?

　　　　B: Fewer than 100 people have called so far.

How many people have called the station so far today? From …… to ……, most likely …….

15.　　A: We have enough budget left to print colour copies of the brochure. How many people have asked for one?

　　　　B: Fewer than 93 people have asked for a copy.

How many people have asked for a copy of the brochure? From …… to ……, most likely …….

16.　　A: We can provide 100 lunches for the guests. How many of them will stay for lunch?

　　　　B: More than 100 of them will stay for lunch.

How many of the guests will stay for lunch? From …… to ……, most likely …….

Please read the following short dialogues, and answer the questions by filling in a value for each blank space, according to your opinion. Consider each dialogue separately. Assume that participant B is well-informed, telling the truth, and being co-operative in each case.

1.   A: We need to sell tickets to cover our costs. How are the ticket sales going?
        B: So far, we've sold fewer than 60 tickets.

How many tickets have been sold? From …… to ……, most likely …….

2.   A: To win the election, the candidate had to convince 77 people to vote for him. How many votes did he get?
        B: He got more than 77 votes.

How many votes did the candidate get? From …… to ……, most likely …….

3.   A: This photo album has space for 4x6 photos. How many photos of that size do you have?
        B: I have about 150 photos.

How many photos does B have? From …… to ……, most likely ……

4   A: We've invited members and non-members to the reception. How much room is there in the hall?
        B: There's room for more than 80 people.

How many people is there room for in the hall? From …… to ……, most likely …….

5.   A: There are still 200 copies of the agenda left. How many delegates are we waiting for?
        B: We're waiting for more than 200 delegates.

How many delegates are they waiting for? From …… to ……, most likely …….

6.	A: This display case holds 80 CDs. How many CDs do you own?
	B: I own fewer than 80 CDs.

How many CDs does B own? From …… to ……, most likely …….

7.	A: The lecture hall can accommodate 130 students taking an exam. How many students will be taking tomorrow's exam?
	B: About 130 students will be taking tomorrow's exam.

How many students will be taking the exam? From …… to ……, most likely …….

8.	A: I expect cars to arrive between now and three o'clock. How many spaces are left in the car park?
	B: There are more than 93 spaces left in the car park.

How many spaces are left in the car park? From …… to ……, most likely …….

9.	A: We wanted new signatures on the petition. How many new people signed it?
	B: Fewer than 200 new people signed it.

How many new people signed the petition? From …… to ……, most likely …….

10.	A: The new prison still has space for prisoners. How many prisoners need to be transferred out of the old facility?
	B: About 110 prisoners need to be transferred.

How many prisoners need to be transferred? From …… to ……, most likely …….

11.	A: 60 people have already paid their deposit for this holiday. How many seats are there on the plane?
	B: There are more than 60 seats on the plane.

How many seats are there on the plane? From …… to ……, most likely …….

12.   A: We can hire a bus with reclining seats for the excursion. How many people will be coming?

       B: Fewer than 77 people are coming.

How many people are coming on the excursion? From …… to ……, most likely …….

13.   A: We have space in the garden to plant 50 tulips. How many tulip bulbs are in the bag?

       B: There are about 50 tulip bulbs in the bag.

How many tulip bulbs are in the bag? From …… to ……, most likely …….

14.   A: There are 100 free tickets available for people who call the station today. How many people have called so far?

       B: Fewer than 100 people have called so far.

How many people have called the station so far today? From …… to ……, most likely …….

15.   A: We have enough budget left to print 93 colour copies of the brochure. How many people have asked for one?

       B: Fewer than 93 people have asked for a copy.

How many people have asked for a copy of the brochure? From …… to ……, most likely …….

16.   A: We can provide lunches for the guests. How many of them will stay for lunch?

       B: More than 100 of them will stay for lunch.

How many of the guests will stay for lunch? From …… to ……, most likely …….