entailment

one

implicature

ano qualum

negation

irgendein

nessuno

existential

give a damn

khoii

free choice items

ever

antinegativity

concord

until

universality

un qualihe

epistemic indefinites

quantification   degrees of freedom

ek bhii

GENNARO CHIERCHIA

Logic in Grammar

Polarity, Free Choice, and Intervention

Logic in Grammar

OXFORD STUDIES IN SEMANTICS AND PRAGMATICS

GENERAL EDITORS
Chris Barker, *New York University*, and Chris Kennedy, *University of Chicago*

PUBLISHED
1 *Definite Descriptions*
Paul Elbourne

2 *Logic in Grammar*
Polarity, Free Choice, and Intervention
Gennaro Chierchia

IN PREPARATION
Weak Island Semantics
Márta Abrusán

Subjectivity and Perspective in Truth-Theoretic Semantics
Paul Lasersohn

A History of Formal Semantics
Barbara Partee

Numerically Qualified Expressions
Christopher Cummins

Measurement and Modality
Daniel Lassiter

Logic in Grammar

Polarity, Free Choice, and Intervention

GENNARO CHIERCHIA

OXFORD
UNIVERSITY PRESS

Contents

Acknowledgements

I am grateful to Chris Barker and Chris Kennedy for their editorial guidance. I acknowledge with gratitude three very thorough anonymous reports. Chris Kennedy also gave me extensive and detailed comments that earn him the title of 'honorary fourth referee'. Moreover, since this project has been going on for more than it should have, I got extensive feedback from many different people. I list some of them in what follows (with more specific pointers in the body of the text), even though this is insufficient to express my debt and gratitude toward them: Maria Aloni, Adriana Belletti, Ivano Caponigro, Carlo Cecchetto, Noam Chomsky, Luka Crnič, Veneeta Dayal, Anamaria Fălăuş, Danny Fox, Jon Gajewski, Irene Heim, Angelika Kratzer, Daphne Liao, Giorgio Magri, Clemens Mayr, Andreea Nicolae, Daniele Panizza, Hazel Pearson, Orin Percus, Luigi Rizzi, Jacopo Romoli, Uli Sauerland, Philippe Schlenker, Greg Scontras, Benjamin Spector, and Hedde Zeijlstra. *Gratias vobis ago*. I also have many sizeable intellectual debts to ideas and works in the literature that I hope will be made clear in the course of this book. But I must mention at the outset that much of what I try to do here is to work out an idea that I have seen most clearly articulated in Manfred Krifka's work.

By the time this volume comes out, it will be thirty years since my dissertation, from which I'd like to quote the following passage:

The person to whom I owe the most is my wife Isa. Since we met, 15 years ago, she always was near me, carrying the burden of staying optimistic during various difficult times and keeping me up, often against myself, with her love. Now we also have little Gabriel to keep us up, especially at night.   (Amherst, September 1983)

Gabe has just turned 30; and Tommaso and Matteo, who came later, are young adults, living their challenging and interesting lives. Isa's wise and loving presence throughout was crucial in getting each one of us where we are. I have put her through many more trials since 1983, and those words are more true and heartfelt today than ever; certainly the only lasting legacy of my dissertation.

<div align="right">

Cambridge, Mass.
July 8, 2012

</div>

Abbreviations

A	Alternative
AA	Anti-additive, Anti-additivity
API	Affective Polarity Item
ATV	Anti-Total Variation
DA	Domain alternative
DE	Downward Entailing
DN	Double Negation
DP	Determiner Phrase
EA	Enriched Assertion
EI	Epistemic Indefinite
EPP	Extended Projection Principle
FC	Free Choice
FCI	Free Choice Item
IE	Innocently Excludable, Innocent Excludability
LF	Logical Form
MC	Modal Containment
MH	Modal Horizon
NC	Negative Concord
NP	Negative Polarity
NPI	Negative Polarity Item
OS	Ordinary Scalar, Ordering Source
PIE	Proto-Indo-European
PP	Positive Polarity
PPI	Positive Polarity Item
PS	Polarity Sensitive, Polarity Sensitivity; Proper Strengthening; Polarity System
PSI	Polarity Sensitive Item
SA	Scalar Assertion
SDE	Strawson Downward Entailing
SE	Scale Economy
SI	Scalar Implicature

SNPI	Strong NPI
UE	Upward Entailing
UG	Universal Grammar
VP	Verb Phrase
WSC	Wide-Scope Constraint

This page intentionally left blank

Introduction

For the purposes of the present work, we can think of language as a structure-building facility (i.e. a computational device that forms complex expressions out of a lexicon), interfaced with the ability to use such structures in drawing inferences (i.e. a natural logic). One of the general goals of this work is to investigate this interface. A well-known example of how syntax and logic interact is constituted by the system of Polarity Sensitive Items (PSIs), a.k.a the Polarity System. Here is a generalization pertaining to it that one is likely to encounter in textbooks and introductory lectures. Focus on the following inferential patterns. If I tell you that every time I ate pizza, I got sick, you won't have any trouble in figuring out that if I ever ate pepperoni pizza, I got sick. This is a "subset inference" (because the set of pepperoni pizza eaters is a subset of the set of pizza eaters). By the same token, if I tell you that over the past month, every time at lunch I ate pepperoni pizza, you'll thereby know that for the last month at lunch I ate pizza. This is a "superset inference". In drawing these inferences we are crucially tapping the semantics of *every*. Structures of the form [Every NP VP (Verb Phrase)] give rise to subset inferences in the NP part (the restriction) and superset inferences in the VP part (the scope).

(1) a. every pizza makes me sick → every pepperoni pizza makes me sick
    b. every time at lunch I eat pepperoni pizza → every time at lunch I eat pizza

Licensing superset inferences is a property that the scope of *every* shares with positive sentences (I ate pepperoni pizza → I ate pizza). Licensing subset inferences is a property the restriction of *every* shares with negative sentences (I didn't eat pizza → I didn't eat pepperoni pizza). A context that licenses subset inferences, like negation or the restriction of *every*, is called 'Downward Entailing' (DE). These

inferential properties are acquired spontaneously as part of figuring out what *every* and negation mean (= how they are used in natural reasoning). Now, it has long been known that languages have classes of items that are somehow sensitive in their distribution to precisely this kind of inferential properties. *Ever* is a case in point:

(2) a. Every [person who **ever** ate pepperoni pizza from that place] [got sick]
     b. *Every [person who eats my pepperoni pizza] [will **ever** get sick]
     c. Every [person who eats my pepperoni pizza] [won't **ever** get sick]

It looks as though *ever* wants to be in DE environments, like the restriction of *every* (2a) or the scope of negation (2c), and is ungrammatical in non-DE environments, like the scope of *every* (2b). *Ever* is a typical *Negative* Polarity Item (NPI). *Any* works in parallel ways with respect to the structures in (2), but has a broader distribution, to be reviewed shortly. The existence of PSIs has been the object of intense study over the past 30 years. What immediately catches everybody's curiosity is why there should be a class of items that is restricted in its distribution to contexts that share the abstract property of being DE. What aspects of the grammar of *ever* or *any* makes them act that way? This question will be one of our central concerns.

Besides sharing with *ever* the desire to be in negative contexts, *any* also admits a class of uses that *ever* does not tolerate:

(3) a. Pick any cake
     b. You may have any of these cakes
     c. *ever pick any cake
     d. *You may ever have any of these cakes

Those exemplified in (3a,b), where *any* is in the scope of imperatives or modals of possibility, are prototypical Free Choice (FC) uses. Uttering (3a,b) says that you are free to choose the relevant cake. The existence of items like *any* that have both Negative Polarity and Free Choice uses is also very widespread across languages, and understanding why constitutes a second central objective of our investigation. The main question in this connection is why and how Polarity Sensitivity and Free Choice phenomena are related. It is not obvious that items that want to be in DE environments should also often want to be in FC environments.

What makes FC phenomena particularly interesting is that they don't arise just with 'special' items like English *any* (or German *irgendein* or Italian *qualsiasi*), with a restricted distribution and morphemes specifically marked for FC uses. FC effects also arise with ordinary disjunction and, for that matter, plain vanilla indefinites like *some* or the indefinite article *a*.

(4) a. Pick a cake
     b. Have wine or beer
     c. You may pick a cake
     d. You may have wine or beer

In all these cases it seems that I am giving my hearer permission to choose freely. If I say that you may pick a cake and add nothing else, I cannot then protest if you pick the apple pie rather than the cheesecake. As we will see, the Free Choice interpretation of disjunction/plain indefinites in (4) does not readily derive from their lexical semantics: the semantics of the imperative/modal and the way they are put together. This makes them challenging. So in other words, FC phenomena are analytically challenging and affect quite generally a wide class of items, some of them marked with designated morphology, others not so marked. We will argue that FC phenomena have a strong empirical unity (i.e. they are induced in specific environments, and have an identical semantic import across such environments). The natural conjecture is that this is so because they are based on a unitary mechanism. Understanding the unity of FC phenomena is a third key goal of the present inquiry. Some authors have argued that the FC effect with disjunction and plain indefinites is an implicature, rather than an entailment. Implicatures are generally understood to be optional phenomena rooted in conversational dynamics rather than in principles of syntax and compositional semantics. If FC effects across ordinary indefinites and items with a dedicated morphology (*any*/*irgendein*) are empirically uniform, then the FC behavior of the latter should be viewed as an implicature as well.

The notion of implicature and its role in grammar is currently at the center of great controversy. Some researchers, including yours truly, have argued that at least some implicatures (namely the scalar ones) are rooted in the compositional part of the grammatical system, rather than in the post-compositional, conversationally driven part, as Grice had originally proposed. I wish I could stay away from such disputes, for the purposes of investigating the Polarity and Free Choice systems.

And there are indeed parts of what I am going to propose that are independent (or can be easily made independent) of any claim regarding the nature of scalar implicatures and scalar reasoning. But there are also aspects of the investigation of the Polarity System that simply cannot be excised from some investigation of the implicature system. In particular, if FC phenomena are uniform and display points of contact with scalar reasoning, we would simply be depriving ourselves of the possibility of understanding what is going on by excluding what has traditionally been seen as an implicature within our purview. So, some understanding of scalar reasoning and scalar implicatures simply has to be part of our inquiry (a fourth main goal).

Let me illustrate a further way in which implicatures may enter into an analysis of the Polarity System. A phenomenon that appears to be closely tied to polarity is Negative Concord, illustrated by the Italian examples in (5):

(5) a.   Nessuno non ha protestato
          N-one not has protested
          'everybody protested'

     b.   Non ho detto niente a nessuno
          (I) not have said N-thing to N-one
          'I haven't said anything to anybody'

In (5a) both *nessuno* and *non* contribute a separate negation to the interpretation of the sentence, which therefore winds up having a "double negation" reading, as per the gloss. However, in (5b) there is only one semantically active negation, again as per the English gloss. So *nessuno* seems to contribute a semantic negation in (5a) but not in (5b), where it looks more like an agreement marker. Items like *nessuno* that display this kind of behavior are called "N-words". There is controversy on whether N-words are NPIs or not. We will side with those who argue that they are. But the point I want to make here is a different one. When N-words are made up of a quantificational element, like *nessuno* = NEG + *a/one*, it is always the logically weakest member of a scale (in this case, *uno* 'a/one') that is recruited. There are no N-words for numerals higher than one (something like NEG-two) or quantifiers like *most* or *every* (something like NEG-all). The question is why. To emphasize the puzzling nature of this restriction, consider that Italian has an N-word composed from the additive particle *anche* 'also', namely *neanche* = NEG-also, combining with a

variety of Noun Phrases, including Number Phrases (as in (6a)); but N-also usually cannot combine with numerals bigger than *one*. Something like (6b) is ungrammatical:

(6)   a.   Non ho visto neanche un professore di matematica
          (I) not have seen N-also a professor of mathematics
          'I didn't even see one math professor'
      b.   *Non ho visto neanche due professori di matematica
          (I) not have seen N-also two professors of mathematics

The ban in (6b) is not due to some kind of weird morphosyntactic restriction on *neanche*, but seems to relate to the position of the numeral on the scale, as the following example illustrates:

(7)   quel problema non lo risolverebbero neanche due professori di matematica
      that problem not it-would solve N-also two math professors
      'Not even two math professors would solve that problem'

What makes (7) acceptable is construing *two math professors* as the lowest number of math professors that stands a chance of solving the problem. In other words, if the pragmatically salient scale of numerals is the regular one (viz. <one, two,...>), then *neanche* can combine only with *one*. If the pragmatically salient scale has some other granularity (<n, n+a,...>), then *neanche* can associate only with the smallest amount on such a scale. The point of this example is to show the robustness of the present generalization on N-words. Non-lowest amounts on a scale cannot be N-words, even when there is a construction which would make that in principle possible like *neanche*. Now there is an extensive and insightful literature on negative concord and N-words; we will incorporate several key ideas stemming from such literature in our proposal on N-words. But there is no sign of a robust explanation of this striking constraint.

   Here is my hunch as to why this generalization holds. N-words want to be in a DE (or, in fact, in an even more strongly negative) environment. But when embedded in a DE environment, intermediate quantifiers on a scale tend to generate positive implicatures. Positive implicatures disrupt the DE character of the environment and N-words do not tolerate that. Let me elaborate on the various pieces of this intuition. Consider the dialogue in (8):

(8)   Q: How many assignments will you have graded by noon?
      A: No more than ten
      Implicature: I will have graded by noon a certain number of
      assignments

The inference in (8) is a typical quantity implicature. It is usually
inferred from an answer like A above, but doesn't always have to be
there. A in (8) will not be a lie if by noon no assignment has been
graded. The way in which such inferences are drawn in a Gricean
setting is by crucially exploiting the relevant scale, in the case at hand
something like

(9)   no   more   than → no   more → no more than [large
      [small amount]   than ten   amount]

An utterance such as A in (8) brings to salience a scale such as the one
in (9), ordered by logical entailment as shown there: if by noon the
speaker won't have graded more than ten assignments, s/he certainly
won't have graded more than any amount larger than ten. Now
choosing *no more than ten* is compatible with the speaker not grading
more than some smaller amount, say five. But if the speaker had
thought that that was a possibility, s/he would have said so (being
cooperative, etc.). Hence, s/he must believe that some relevant stronger
statement (like *I won't have graded more than five assignments*) is false.
This brings about the implicature that the speaker expects that at least
some of the stronger statements are false, and hence expects to have
graded some (vaguely specified) numbers of assignments by noon. So
the answer winds up being interpreted as *I will have graded some but
fewer than ten assignments by noon*. Propositions of this sort, which
have a positive and a negative component to them, do not license
subset inferences. Consider:

(10)   Some but no more than five students will eat pizza

Sentence (10) of course does not entail that some but no more than five
students will eat pizza with anchovies. So here is why there is no N-two
or N-ten word. N-words want to be in a purely negative environment,
that licenses subset inferences. But in such contexts, N-two/N-ten
would automatically generate a positive implicature, and the context
would no longer be a DE one (it would no longer legitimate subset
inferences). So the positive implicature gets in the way, "intervenes"
in the licensing of (hypothetical) middle-of-the-scale N-words. In the

absence of a more compelling story, this sounds like a plausible explanation for why there is no N-two/N-ten, one that extends readily to other quantifier scales (like <N-some, N-most, N-every>). Of course there are many things to spell out carefully here, and we will try to do so. But the main reason for going through this example at this early stage is to illustrate how scalarity and Negative Polarity licensing may well interact in surprising and perhaps unexpected ways. This motivates us further to studying them together. Intervention effects on NPIs or FCIs abound, as the literature amply documents. Our fifth main goal is to study them, in pursuit of the hypothesis that scalar reasoning may play a major role in intervention.

A key protagonist of our inquiry into the Polarity System is the phenomenon of "silent exhaustification". One often runs into examples of the following sort:

(11)   Yesterday, I went to the party, greeted everybody, hugged Mary and Sue, and left

Ask yourself how you interpret the third sentence of the narrative in (11), namely *I hugged Mary and Sue*. Notice that such sentence *per se* is compatible with my having hugged other people as well, as it readily comes out in other contexts. Suppose for example that Mary and Sue have contracted salmonella and I want to know who hugged them and is therefore at risk. If you hugged Mary, Sue, and John, you would nonetheless come to me and say 'I did hug Mary and Sue'. However, this is not the way in which the sentence is understood in the context of the speech in (11). There we understand 'I hugged Mary and Sue' in an exhaustive manner, as equivalent to *I hugged only Mary and Sue*. This phenomenon of "covert exhaustification" is extremely pervasive. As we form sentences, we spontaneously consider alternatives. And, depending on the context, we sometimes mean what we say exhaustively with respect to such alternatives, as in (11). We hypothesize that covert exhaustification plays a central role in both NPI and FCI licensing as well as in implicature calculation. The "alternative+exhaustification" framework we will use has many points of contact with current theories of focus.

Getting now to specifics, the main items we will study in the above setting are:

(12)   a.   Universal Free Choice items: English *any*, Italian *qualunque*
               [free choice items with a universal quantificational force]

b. Existential Free Choice items: German *irgendein*, Italian *uno qualunque*
    [free choice items with an existential quantificational force]
c. Epistemic Indefinites: Italian *un qualche*, Romanian *vreun*
    [indefinites associated with an ignorance implication about the identity of their referent]
d. Weak NPIs: English *ever*, Italian *mai*
    [items restricted to (roughly) DE contexts]
e. N-words: Italian *nessuno*, *neanche*
    [items that participate in negative concord]
f. Strong NPIs: English *in weeks*, punctual *until*
    [items limited to strongly negative context]
    (e.g. anti-additive ones – cf. Chapter 1, Section 1.4.3)
g. Emphatic NPIs and 'minimizers': English *give a damn*, *lift a finger*, Hindi *koii bhii*

The reason for studying this fairly large sample of Polarity Sensitive and FC items, which, as we shall see, differ so much in their behavior, is to investigate to what extent they form a system, and how tight/loose such a system is. We will try to make a case that all the items in (12a–g) have the same meaning and draw from the same alternative pool. Their alternatives and meaning are, in fact, the same as for ordinary indefinites like *some* or *a*. The latter are generally viewed as existential indefinites, which form scales with other quantificational elements (e.g. <some, many, most, every>, <a/one, two,...>), and also come with "domains of quantification" over which they range. This strong thesis on the nature of the PSI in (12) would explain, among other things, why *any* comes from the Proto Indo-European word for *one* with an adjectival ending tacked on (lit. something like *oney*) and why its German cousin *einig*- is a plain vanilla indefinite: *one, any, einig* are underlyingly the same thing.

If virtually everything in the Polarity System is the same as plain indefinites, how come NPIs and FCIs are so much more restricted in their distribution than *a/some*, and moreover so different from each other? We conjecture that the reason resides in a very small range of choices on how alternatives are used. For one thing, with plain indefinites, such a use (in say, an exhaustification process) is totally context-dependent: we exhaustify when we perceive it as appropriate to the conversational goals. With NPIs and FCIs, we *have* to exhaustify: meaning always has to be pitted against the alternatives. If you wish,

this is a typical process of "grammaticization": a phenomenon that is typically context- and goal-driven becomes a "rule of grammar". On top of this, there are probably a small number of slightly different ways of 'exhaustifying' that, say, look at different components of meanings (entailments, presuppositions, implicatures); there are also probably slightly different natural subclasses of the alternative pool that different lexical items may latch onto. These few choices may determine extremely diverse behaviors and possibly explain variation within the Polarity System, even if we are really dealing with one and the same creature. As it turns out, strong and emphatic NPIs (namely 12f,g), though not strictly identical to weak indefinites, are also closely related to them (e.g. a low point on a lexical or pragmatic scale).

Summing up, the scope of the present inquiry is the study of:

(13) a. Negative Polarity Items (NPIs);
b. Free Choice Items (FCIs) and their relation to NPIs;
c. the unity of FC phenomena across FCIs and disjunction/plain indefinites;
d. scalar reasoning and scalar implicature;
e. interference and intervention phenomena between the Polarity System and scalarity;
f. the items in (12) as part of a unified system.

Setting these goals will sound to you grand and immodest. I can certainly tell you that pursuing them nearly killed me. I did have the feeling of having gotten into an enterprise I could not pull through: too complex, too many variables, and (as it touches upon so many central aspects of grammar) too "polarizing", to use a pun by Chris Barker. Never in my now not-so-short linguistic life have I received so much criticism (and, if undoubtedly more rarely, so much support) before getting even close to finishing the project. At the same time, so much good work has piled up on these topics over the past thirty years: fascinating analyses of specific items, based on beautiful insights that expose aspects of the spontaneous logicality of language; developments of important frameworks, like alternative-based semantics; new debates on old questions like implicatures and presuppositions. It simply had to be brought together in one single logical space—a space where one can try out conceptually simple, purely generative machines for the Polarity System, and where the seemingly reckless claim that there is really one Polarity Sensitive Item can be put to the test.

I present to you, with many doubts, the outcome of this attempt. If I could, I would rewrite the whole thing from scratch. But that too would make no sense. My suggestion is that you start with the conclusions. If they intrigue you, try the rest. With luck, this work might provide a partial existence proof that the claim of the radical unity of the Polarity System is sensible and testable: with lots of generalizations and problems, some old and some maybe new; based on a computational device simple in its conception, but hard to work out in its details; and with many choices which could or maybe should have been different. The one thing I come out of this most firmly convinced of is that logically based notions like entailment, contradictoriness, and analyticity do play a central role in grammar. Syntax interfaces with a logical apparatus to the point that things that 'feel' syntactically ill-formed really owe their status to their logical properties. One of my hopes is that this study may help put into sharper focus some of these foundational issues: is logic part of grammar? Is grammar part of logic? Are they independent computational systems?

1

The Spontaneous Logicality
of Language

1.1 Polarity Sensitivity

A number of linguistic phenomena take place only in environments with a distinctive "negative" or "modal" character. The cover term "Polarity Sensitive" (PS) is used to refer to such a (seemingly highly diverse) set of phenomena. In the present chapter I try to give an idea of their scope and of the reasons why they are both interesting and revealing of central aspects of how language works, in particular of how grammar and logic are intertwined. I will begin by illustrating how polarity affects the readings of ordinary disjunction and a number of other items. Then, I consider the distribution of morphemes like *ever* or *any* in English. It turns out that these two phenomena (a purely interpretive one and a distributional/syntactic one) are responsive to structurally similar logical factors. We develop an initial hypothesis on why this might be so.

1.1.1 Interpretations of or

As is well known, the interpretation of *or* oscillates between an exclusive and an inclusive one.

(1)   Background: what will be the future departmental hires?
   a. If everything will go well, we'll hire either Mary   [exclusive]
       or Sue
   b. If we hire either Mary or Sue, everything will go   [inclusive]
       well

Sentence (1a) indicates an intention/expectation to hire one of two candidates, with no expectation of hiring them both. This constitutes a typical *ex*clusive construal of disjunction. On the other hand, sentence

(1b) says that if we hire Mary or Sue we will be lucky; but it also seems to convey that if we wind up hiring them both, we will be (all the more) in good shape, a typical *in*clusive use. Notice in connection with (1b) that it might in fact be the case that Mary and Sue are fierce intellectual adversaries, so much so that hiring them both would in fact lead to pain and disruption of departmental life. But unless this is common knowledge or made explicit by the speaker, an innocent hearer will tend to understand *or* in (1b) *in*clusively as conveying that a hypothetical (if unlikely) double hire would indeed constitute luck.

One might think that these preferences in how *or* is interpreted in (1a) vs. (1b) have to do with world knowledge, familiarity with academic scripts, etc. And indeed it would be highly surprising if such factors played no role. However, world knowledge and other purely pragmatic factors cannot be the whole story. Notice that the sentences in (1) are constructed out of the very same lexical material, modulo swapping antecedent and consequent. Hence, it is highly unlikely that the role of world knowledge, familiarity with specific scripts, etc. change in choosing (1a) over (1b) as responses to the background question in (1). Such sentences are likely to be appropriate to contexts that are, broadly speaking, quite similar in terms of the kind of general facts, patterns, conventions, etc. that are tapped. Hence the preference we perceive for an exclusive construal in (1a) vs. an inclusive construal in (1b) cannot be determined (solely) by contextually relevant encyclopedic knowledge and the like. The interaction of such knowledge with the meaning of *if*, and whether *or* occurs in the consequent (as in (1a)) vs. the antecedent (as in (1b)) of an *if*-clause must be playing a crucial and direct role in shaping our preferences.

In the next paragraphs, I would like to make a case that the interpretation of *or* (and a class of related words—the so called "scalar terms") is indeed sensitive to structural factors of this type (i.e. whether the scalar term occurs in the antecedent vs. consequent of an *if* clause).[1] Then I will try to formulate a hypothesis as to why this might be so. Such a hypothesis, if correct, will make manifest what I mean by "spontaneous logicality" of language. This will lead us

---

[1] The argument to be developed is a simplified and updated version of the one proposed in Chierchia 2006 (but cf. also Atlas and Levinson 1981 and in response to Gazdar 1979 Horn 1989 as important precursors). The update on Chierchia (2006) concerns some experimental results, to be briefly summarized in the text to follow, that weren't available at the time.

directly into the question of Polarity Sensitivity and its manifestations, which the rest of the book is devoted to.

Numbers are useful in illustrating the issue further. Like *or*, they admit two main interpretations, an 'at least' vs. an 'exact' one. These alternative interpretive choices can be brought up in ways fully parallel to (1):

(2)   Background: how much will they pay us for shoveling their driveway?
   a.   If we are in luck, they'll pay us $20
   b.   If they'll pay us $20, we are in luck

Sentence (2a) indicates an expectation of getting $20 and not (significantly) more. On the other hand, (2b) doesn't convey that if we get $20 *and no more*, we are in luck. That is not what the utterer of (2b) is likely to want to communicate. The intended sense of (2b) is something like: if they pay us $20 or possibly (if unexpectedly) more, we are in luck. Again, world knowledge (e.g. the fact that more money is better than less money, the going rates for occasional snow shoveling, etc.) remains constant across the hypothetical replies in (2a) vs. (2b) to the background question in (2). So our spontaneous preference for the exact construal of *$20* in the consequent (i.e. main clause) of the conditional sentence in (2a) vs. the 'at least' construal in the antecedent of the conditional in (2b) cannot be due just to such knowledge.

Since at least Grice (1989), this oscillation in meaning of words like *(either)/or* and (for some authors) the numerals (i.e. number words such as *one*, *two*, *three*) is viewed as a pragmatic enrichment phenomenon.[2] It is useful to review, in a very preliminary way, the main aspects of Grice's ideas, something we will take up in greater detail in Chapter 2. According to Grice, when we use a sentence with *or*, we do so while considering a set of relevant alternatives that might have been used in its stead. For example, if a situation calls for the use of sentence (3a), sentence (3b) might also be relevant.

(3)   a.   We will hire (either) Mary or Sue
       b.   We will hire (both) Mary and Sue

---

[2] This issue is much discussed under the rubric of Scalar Implicature. Standard references on SIs include Grice (1989), Horn (1989), Levinson (1983; 2000), Sperber and Wilson (1986).

Assume now that the basic meaning of *or* is the more general one (i.e. the inclusive one); such a basic meaning might be pragmatically enriched through the addition to *p or q* of a silent *and not both*. The enrichment in question, which has come to be known as a Scalar Implicature (SI), may be prompted by conversational maxims. More specifically, cooperative speakers tend to give all and only information that relevantly pertains to the issue under discussion (the maxim of Quantity). Thus, if a well-informed member of the department volunteers (3a), rather than the logically stronger and hence more informative (3b), in answer to the background question in (1), it must be because the speaker believes (3b) to be false. This leads the hearer to conclude that the exclusive construal must be what is meant, in such a case. However, if the truth (or falsity) of the conjunctive alternative in (3b) is irrelevant to the goals of the conversation, or if the speaker somehow signals his ignorance on the matter, then there would be good reasons not to run the reasoning sketched above, and the inclusive interpretation of the basic form would emerge. This constitutes a seemingly plausible account of the oscillation in interpretation of *or* (an account that potentially extends to numbers and other scalar terms).[3]

Adopting, for the time being, this familiar Gricean view, what examples like (1) and (3) seem to show is that adding the implicature to the consequent of a conditional seems to be easier than adding it to the antecedent, in minimally different contexts.[4] The word *easier* here is not casually chosen. We are dealing with a preference that obtains if everything else is equal. It is possible, by changing the context, to induce the readings that I am alleging to be dispreferred.

(4)   I know that we have only one position. But Sue and Mary are a couple. They are seeking jobs in the same place. So it is just wrong to think that if we hire Mary *or* Sue, everything will go well.

---

[3] Other major scalar terms, besides the coordinating particles *or/and* and the numerals include the modals *must/can/might*, the positive quantifiers *some/many/most/all*, the negative quantifiers *no/few/not all*. These are known as "Horn scales," after Horn's (1989) seminal discussion. The argument in the text is meant to apply to all scalar terms.

[4] 'Adding' in the present context should be understood as 'embedding'. That implicatures can be embedded has been observed and discussed extensively in e.g. Levinson (2000). How embedded implicatures may arise remains the object of controversy, a controversy we will pick up in Ch. 2.

Viewed in the context provided in (4), particularly with focal stress on *or*, the conditional in (1b) is interpreted as follows: don't expect that if we hire Mary or Sue and not both, everything will be OK. So in (4), disjunction has to be construed exclusively within the antecedent of a conditional. What this example shows is that an exclusive interpretation of *or* in the antecedent of a conditional is possible, sometimes even necessary. But such interpretation must somehow be induced by making explicit background assumptions that lead to it. In contrast, embedding an exclusive construal of *or* in the consequent of a conditional does not seem to require much priming, as (1a) and (2a) seem to show.

Summing up our discussion so far, I have suggested that the interpretation of disjunction (or of a numeral) is sensitive to where it occurs in a structure. If it occurs in the consequent of a conditional, the exclusive interpretation is relatively easier to get than with respect to a minimally different sentence in which *or* (or a numeral) occurs in the antecedent. "Interpretation" is used here rather loosely as referring to the basic compositional meaning enriched by whatever (possibly pragmatic) inference/implicature may be tacked onto that. "Minimally different" is to be understood as illustrated in the examples in (1) and (2). Given that the lexical material, background question, world knowledge, etc. are alike across the pairs in (1) and (2), we conjecture that the observed preference must be due to the interaction of the meaning of *or* with the position in the structure in which *or* occurs.

The next step is to show that this phenomenon is general and involves a wide class of structures. Consider the following minimal pairs:

(5) a. Everyone here either likes Mary or likes Sue and will write to the dean
     b. Everyone here who either likes Mary or likes Sue will write to the dean

Check it against your intuitions: to get an exclusive interpretation of *or* seems to be fairly easy in (5a), much less so in (5b); sentence (5b) seems to be saying that if there is somebody who happens to like both Mary and Sue, that person too will write to the dean. Once again, the two sentences in (5) differ minimally with respect to the lexical material employed. Sentence (5a) contains the conjunction *and*, while sentence (5b) contains the relative pronoun *who*. This simple change precipitates a significant difference in syntactic structure:

(6)   a.

```
                        IP
              ┌─────────┴─────────┐
             DP                   VP
            ╱  ╲           ┌───────┼───────┐
           D    NP        VP      conj     VP
           │    │         │        │        │
```

[Every one here] [[either likes Mary or likes Sue] and [will write to the dean]]

b.

```
                              IP
                   ┌──────────┴──────────┐
                  DP                      VP
            ┌──────┴──────┐               │
           D              NP              │
           │              │               │
```

every [one here who either likes Mary or likes Sue] [will write to the dean]
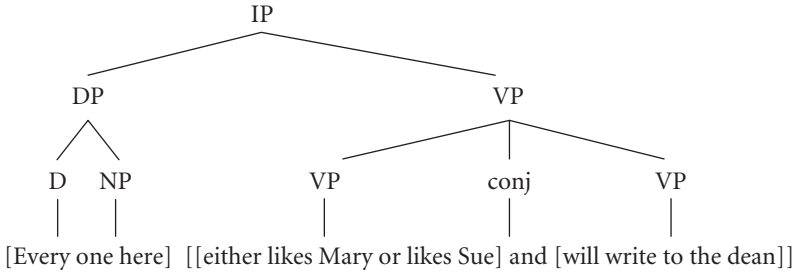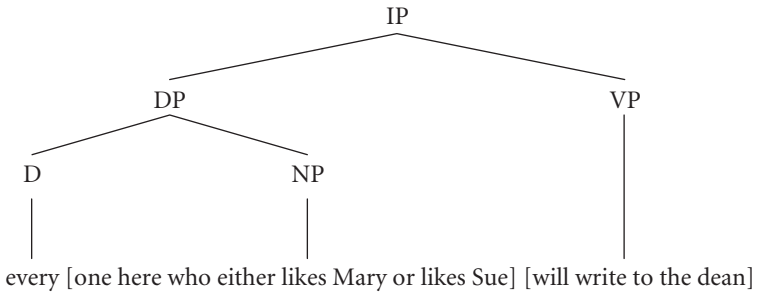
Disjunction occurs within the main VP in (6a) and within the subject in (6b). This syntactic difference corresponds to a semantic one. Assuming the standard analysis of quantifiers like *every* as relations between sets, summarized in (7), this is equivalent to saying that disjunction occurs within the second argument of *every* in (6a) (a.k.a. the scope) and within its first argument (a.k.a. the restriction) in (6b).

(7)   a.   i.   every (one) (either likes Mary or likes Sue and will write
                to the dean)
          ii.   {x: x is a person} ⊆ {x: x likes Sue or Mary and x will
                write to the dean}
      b.   i.   every (one who either likes Mary or likes Sue) (will write
                to the dean)
          ii   {x: x is a person and x likes Mary or Sue} ⊆ {x: x will
                write to the dean}

If intuitions are reliable, an exclusive interpretation of *or* is more natural in (6a) with respect to (6b).

But are intuitions reliable? Intuitions about grammaticality, contradictions, etc. are fairly reliable and are commonly used in standard linguistic practice. What about intuitions about *preferred* interpretations? How can we be sure that we are dealing with a general tendency and not with properties of specific examples, subjective biases, etc.?

Maybe when it comes to judgments about *preferences*, we can't rely just on intuitions, even when they seem relatively robust. Judgments about preferences may well be a case where experimental testing is called for. As it turns out, some experimental results that speak to this issue are now available. Let me briefly describe one such set of experiments (even though we can do so only in a highly incomplete way within the limit of the present work).

1.1.2 An experiment on how scalar terms are interpreted

Panizza et al. (2009) have tested sentences involving numerals isomorphic to those in (5), namely:

(8) a. In my neighborhood every girl has two older brothers and wants to have a little sister
     b. In my neighborhood every girl who has two older brothers wants to have a little sister

A list of such examples (duly mixed with appropriate fillers, etc.) was accompanied by an off line questionnaire that would ask subjects whether the example was talking about *exactly* two or *at least* two brothers. Subjects chose the "exact" reading significantly more frequently in sentences like (8a) with respect to sentences like (8b) (55% vs. 27%, p <.0001). This off line effect was further confirmed by strong on line effects in a follow-up experiment in which sentences like those in (8) were compared with sentences of the following sort.

(9) a. In my neighborhood, every girl has two brothers and wants a third one
     b. In my neighborhood, every girl who has two brothers wants a third one

The key difference between (8) and (9) is the following. The continuation *wants a little sister* in (8) is neutral with respect to the exact vs. the at least reading, in the sense that it is consistent with both. On the other hand, the continuation in (9) *wants to have a third* is not neutral. It is only compatible with the exact construal of the numeral *two* in the

first clause. In other words, such a continuation forces the implicature to be there (if the exact construal of numerals is indeed an implicature). Now, *two* in (9b) occurs in the first argument of every, while in (9a) it occurs in the second argument, and we expect therefore that having the 'exact' implicature in (9b) should be harder than in (9a). The experiment consisted in recording eye movements of the participant, with the expectation that the more marked character of (9b) vs. (9a) should result in a processing penalty; such a penalty would manifest itself in the regression pattern of eye movements towards the region of the numeral, once the continuation had been read (a general and consistent effect of processing difficulties). What was found is that indeed the reading pattern of (9a) was strikingly similar to that of (8a), while the pattern of (8b) vs. (9b) was diametrically opposed. Participants made more frequent second pass fixations to the numeral in (9b) with respect to (9a) (and (8a)), a typical symptom of a processing penalty. Thus the off line questionnaire results seem to be confirmed by sizeable on line effects.

We see, therefore, that the intuitions discussed above in connection with examples like (1) or (2) come out supported by carefully controlled experimental testing.[5] This encourages us to conclude that we are dealing with a fairly robust empirical generalization: embedding an implicature in the first argument of *every* (or in the antecedent of a conditional) seems to be harder than doing so in its second argument (or in the consequent of a conditional) within the setting of a minimally different sentence.

1.1.3 Downward Entailing structures

So far we have compared the occurrence of scalar terms in the antecedent vs. consequent of conditionals and in the restriction vs. scope of *every*. But the list of such contrasts is much longer. Here is a taste of its reach (cf. Chierchia 2004 for more extensive discussion).

(10)   Basic vs. comparative forms of adjectives
      a.   John is angry at either Sue or at Mary   [exclusive]
      b.   John is angrier than either Sue or Mary   [inclusive]

---

[5] Other important experimental work on embedded scalar implicature can be found in Chemla and Spector (forthcoming).

(11)   Episodic vs. generic sentences
    a.   An Italian or a Frenchman ordered white wine   [exclusive]
       yesterday
    b.   An Italian or a Frenchman always orders white   [inclusive]
       wine

You should find it much easier to get an exclusive construal in the a-cases than in the b-cases above. I will not discuss these examples in detail (their relevance will become clear in subsequent chapters of this work). I would like instead to focus on one final type of structure which is, I think, particularly revealing of where all this is going.

(12)   a.   Somebody in the department intends to hire either Mary or
       Sue
    b.   Nobody in the department intends to hire either Mary or
       Sue
    c.   John intends to hire Mary or Sue
    d.   Molly does not intend to hire Mary or Sue

The most natural construal of *or* in (12a), an affirmative sentence whose subject is the positive quantifier *somebody*, is exclusive. The most natural one for sentence (12b), which is negative, is inclusive: such a sentence conveys that hiring Mary or Sue or both is not a viable option for anybody in the department. The same goes for the simple affirmative sentence in (12c) with respect to the minimally different negative sentence in (12d).

Before drawing the consequences of these observations, it should be noted that even within negation an exclusive construal is sometimes possible:

(13)   We are aware that Sue and Mary can only be hired together. We discussed it at length. And at this point, nobody wants to hire Mary *or* Sue, anymore; we all agreed to hire them both.

Here we first set up, via contextual background, the exclusive construal of *or*; then we embed the result under a negative quantifier. However, this reading has to be really forced by the context.[6]

---

[6] Horn (1989) labels these examples as cases of "metalinguistic" negation. I believe that most such cases, in which focus plays a key role, can be analyzed as instances of embedded implicatures. We will return to the relevant issues in Ch. 2.

We have so far provided a list of structures where the exclusive interpretation of *or* (and other scalar terms) appears to be disfavored with respect to minimally different structures.

(14)   Distribution of the **exclusive** interpretation of *or*.

	a. easy	b. hard
i.	Consequent of a conditional	Antecedent of a conditional
ii.	Second argument of *every*	First argument of *every*
iii.	Scope of a positive quantifier	Scope of a negative quantifier
iv.	Positive sentences	Negative sentences
v.	Basic forms of adjectives	Comparative forms of adjectives
vi.	Episodic sentences	Generic sentences

These columns can be made much longer; but what we have is enough to make our point. Why bother with such lists? The task of identifying contexts that disfavor the exclusive construal of *or* may seem daunting and perhaps also boring. While this would not be unprecedented in the life of a linguist, our current predicament is much more favorable.

First, coming up with the contrasts in (14) becomes easy once you see the connection with the behavior of a class of expressions that has attracted much attention over the past thirty years or so—the class of expressions we will focus on in the rest of this work. The "hard" column corresponds to environments in which elements like *any* or *ever*, also known as Polarity Sensitive Items (PSIs) are grammatical. The "easy" column corresponds to cases in which such elements are ungrammatical. Here is a relevant sample set (for brevity's sake, I limit my sample to rows (i)–(iii) in (14)):

(15)   *Any*
  (i) a. *If you are hungry, there are any cookies in the oven
     b. If there are any cookies in the oven, you won't go hungry
  (ii) a. *Everyone still has any cookies in the oven
     b. Everyone who still has any cookies in the oven should turn the oven off
  (iii) a. *Somebody brought any cookies
     b. Nobody brought any cookies

(16) *Ever*
> (i) a. *If you visit Sienna, you should ever try homemade ricciarellis
>   b. If you ever try homemade ricciarellis, you will become addicted
> (ii) a. *Everyone ever ate homemade ricciarellis
>   b. Everyone who ever ate homemade ricciarellis became addicted
> (iii) a. *Somebody ever tried homemade ricciarellis
>   b. Nobody ever tried homemade ricciarellis

The (a) versions of the sentences in (15) and (16), which are all deviant, correspond to the (a) "easy" column in (14); the (b) versions, which are all grammatical, correspond to the (b) "hard" column in (14). It is as if PSIs explicitly probed for us where an exclusive interpretation of *or* is to be *dis*favored. One may well wonder why this is so; but that there is a likely correspondence between these two seemingly unrelated phenomena (licensing of PSIs and ease of access to an exclusive interpretation of *or*) seems to be a fact.

One of the early breakthroughs of modern semantics has been the discovery of what distinguishes the environments of the (a) column in (14) from those in the (b) column—a discovery primarily due to Fauconnier (1975) and Ladusaw (1979; 1992), and further articulated in Zwarts (1996). The relevant property has to do with entailment patterns. The environments in column (a) give rise to "superset" inferences; those in column (b) to "subset" ones, as the following examples illustrate:

(17) $\{x : x$ eats pizza with anchovies$\} \subseteq \{x: x$ eats pizza$\}$

	a. superset inferences	b. subset inferences
(i)	If I am depressed, I eat pizza	If I eat pizza, I get sick
	$\Uparrow$	$\Downarrow$
(i')	If I am depressed, I eat pizza with anchovies	If I eat pizza with anchovies, I get sick
(ii)	Everyone eats pizza	Everyone who eats pizza gets sick
	$\Uparrow$	$\Downarrow$
(ii')	Everyone eats pizza with anchovies	Everyone who eats pizza with anchovies gets sick

(iii) Somebody ate pizza   Nobody ate pizza
                  ⇑   ⇓
(iii') Somebody ate pizza with   Nobody ate pizza with
         anchovies   anchovies

The direction of the arrow indicates what entails what. In column (a), the primed rows logically entail the unprimed rows (i.e. the inference goes from a smaller set to the larger ones). Contexts that give rise to this pattern have been dubbed Upward Entailing (UE). In column (b), the situation is reversed: the unprimed rows entail the primed ones (i.e. the inference goes from a larger set to the smaller ones). Contexts with this property have come to be known as Downward Entailing (DE). The match between being DE and licensing PSIs is remarkable, and has consequently attracted an intense discussion.[7] Equally striking is the match between being DE and disfavoring an exclusive interpretation of disjunction. The notion of being DE can be viewed as a generalized semantic notion of "being negative" (regardless of how such negativity manifests itself in the syntax or morphology of the relevant construction). Thus the distribution of items like *ever/any* and the distribution of implicatures appear to be sensitive to the presence of negativity in this precise semantic sense. This is why they constitute prime examples of Polarity Sensitivity in natural language.

Every language has items with a distribution similar to that of *ever/any*. In every language, I conjecture, the interpretation of *or* and other

----

[7] The DE character of some of the contexts discussed here is controversial. This applies in particular to the antecedent of conditionals. Consider:

(a)   If you put sugar in the coffee it tastes good
(b)   If you put sugar and salt in the coffee it tastes good

Clearly, (a) does not entail (b), even though the set of people who put sugar and salt in the coffee is a subset of the set of people who put sugar in the coffee. This phenomenon, known as "failure of strengthening of the antecedent," has been widely discussed in the literature and will be taken up in Ch. 4. The basic line we will adopt, following Heim (1984) and von Fintel (2001), is that the antecedent of conditionals *is* downward entailing once one takes their context dependency into account. Similar considerations apply to generics, which are closely related to conditionals. A sentence like (c) can be roughly analyzed as in (d):

(c)   An Italian always orders white wine
(d)   For any x, if x is Italian [and x feels like it] x orders white wine

Cf. on this Carlson and Pelletier (1995), Chierchia (1995a), and Lahiri (1998: sect. 5). Finally, for a discussion of how comparatives may give rise to DE contexts, cf. e.g. Hoeksema (1983), von Stechow (1984), Kennedy (2005).

scalar items works just as in English. The questions to ask are: Why do these correspondences hold? Why do languages have such properties so pervasively? This is the main issue addressed by the present book. I will begin to do so by offering a preliminary account of the sensitivity of exclusive interpretation to polarity.

1.1.4 Maximize Strength

The distribution of the interpretation of scalar terms may have a very simple and arguably compelling explanation in terms of how logical strength affects information content. Whatever the mechanism behind SIs, their effect is typically that of strengthening an assertion, i.e. making it more informative. Consider the basic case again:

(18) a. We will hire Mary or Sue
      b. Situation 1: we hire Mary but not Sue
          Situation 2: we hire Sue but not Mary
          Situation 3: we hire Mary and Sue
      c. We will hire Mary or Sue [and we will not hire Mary and Sue]
      d. Situation 1: we hire Mary but not Sue
          Situation 2: we hire Sue but not Mary

Imagine someone asserting (18a). On its inclusive interpretation, sentence (18a) is compatible with the three situations shown in (18b), and it would normally indicate that speaker is uncertain as to which of these holds. However, running the Gricean reasoning, and banning evidence to the contrary, the hearer will take the speaker to be conveying, along with the assertion, the silent negation of its relevant alternatives, namely (18c), where the implicature is bracketed to indicate that it is unspoken. Sentence (18c) is logically stronger than (18a): situation 3 is now ruled out and a greater degree of informativeness is thereby achieved. In other words, SIs might be viewed as a mechanism to optionally gain (in contexts in which the speaker is cooperative, well-informed, etc.) more information from a given (alternative activating) sentence.

   Consider now what happens under negation (and, in fact, under any DE context). Negation takes you from a set of situations to its complement. Thus for example the negation of (18a) takes you from the set of situations in (18b) to that in (19b):

(19) a. We won't hire Mary or Sue
      b. Situation 4: we don't hire Mary and we don't hire Sue
      c. We won't hire Mary or Sue but not both
      d. Situation 3: we hire Mary and Sue
      e. Situation 4: we don't hire Mary and we don't hire Sue

Sentence (19a) is meant as the negation of the inclusive disjunction; hence it is incompatible with all the situations in (18b), leaving Situation 4 in (19b) as the only option. Imagine now negating exclusive *or* (i.e. negating (18c)). This negation might be expressed roughly as in (19c) and takes us from the situations in (18d) to their complement, namely those in (19d–e). In other words, negating exclusive *or* is compatible with two states of affairs: our hiring neither Sue nor Mary and our hiring both of them. The upshot is that the negation of inclusive *or* is more informative than the negation of the exclusive one. This is a completely general logical fact. If A *entails* B (i.e. if A is logically stronger, and hence more informative than B) and C is a DE environment, then C(A) *is entailed by* C(B). DE items (like negation) reverse logical strength/informativeness.

   Perhaps this is the right place to remark that throughout this work I will be using "logical strength" and "degree of informativeness" as synonymous. If ϕ asymmetrically entails ψ, then ϕ is more informative than ψ, simply because ϕ must be true in a proper subset of the situations(/worlds) in which ψ is true. In other words, ϕ rules out more "live options" than any logically weaker proposition ψ. So *John ate a sandwich* is compatible with fewer situations than, say, *John ate*, while *John didn't eat* is compatible with fewer situations than *John didn't eat a sandwich*, etc.

   Notice now that negating exclusive *or* is tantamount to first adding the implicature to the embedded clause and then negating the result. This holds true of all DE environments (like the antecedent of conditionals and the first argument of *every*). Thus the bottom line seems to be the following: adding an implicature to a simple sentence (or to a sentence embedded within a UE context) leads to something stronger than the original. It therefore carries with it a gain in informative content and a reduction of uncertainty. Adding an implicature within a DE context leads to weakening the original. This results in loss of information and a state of greater uncertainty. It is then not so surprising that loss of information is a dispreferred option. Weakening an assertion (through the addition of an implicature) is not impossible.

But we do so only when the (pragmatically unenriched) sentence would overtly clash with the context at hand (as in (4) or (13) above).

So the distribution of implicatures seems to reduce to something conceptually quite simple:

(20) Maximize Strength
      Don't add an implicature if it leads to weakening, unless you have to

This would explain the distribution of implicatures we have observed (i.e. the difference between (14a) and (14b)). The explanation relies on simple, if abstract, logical facts. What is remarkable is that such behavior arises spontaneously in us. Children, without explicit formal instruction, eventually converge on the relevant pattern of interpretation of scalar terms across DE vs. UE contexts. This means that they must be endowed with a device which computes the relevant entailment patterns and their impact on information content. If not, then how else are we going to explain the observed preference in interpretation?

Let us briefly review the form of the argument just sketched. We observe that certain ambiguities in interpretation (inclusive vs. exclusive *or*, at least vs. exact numerals, etc.) fall into a pattern (which can be confirmed through controlled experiments). Exclusive interpretations are dispreferred (though possible) within DE contexts, with respect to minimally different UE ones. Next we observe the elementary logical fact that adding a Scalar Implicature within a DE context leads to logical weakening of its implicature-less counterpart (= increase of uncertainty). We conjecture that this option is dispreferred. We tend not to add implicatures when they lead to weakening/loss of information content, unless the context forces us to do so. Functionally speaking, implicatures are primarily a strengthening device; it would be highly confusing to resort to them with equal ease both for strengthening and weakening purposes. But then, if this is on the right track, we must be endowed with logical capacities that enable us to perform the relevant computations in an automatic and unconscious way: the spontaneous logicality of language.

1.2 Why are there Polarity Sensitive Items?

Scalar terms like *or* are not restricted in their syntactic distribution. Their interpretation, however, is restricted in highly specific ways.

In particular, the strengthening via an SI of the interpretation of *or* occurs preferentially in UE contexts. The explanation for this restriction rests on a general logical property of UE vs. DE contexts. If α is logically stronger than β, embedding them within a UE context preserves strength, while embedding them within a DE context reverses it. Now languages have words like *ever* and *any* (and, as we shall see in due course, many other constructions) that are constrained *in their syntactic distribution* in a way that has a striking overlap with the restriction on the interpretation of scalar items. Why? It is very tempting to look for reasons relevantly similar to those we have found for the behavior of scalar items. Perhaps some natural aspect of the meaning of Polarity Sensitive words enables them to do their thing in DE contexts and leads to a short-circuit in non-DE contexts, by virtue of the logical properties of the latter. This aspect of their meaning should have to do with logical strength and degree of informativeness, for that is what the DE/non-DE contrast seems to affect. In the present section I will begin to sketch an approach that draws out this parallel between scalarity and PSIs, in an informal manner. The objective of this preliminary discussion is to illustrate why an account that relies on this parallelism might be desirable. The rest of the present work will be devoted to testing viability of such an account vis-à-vis a broader range of constructions.

1.2.1 Polarity Sensitive Items and their alternatives

The ideas I will present here draw most directly from the work of a number of scholars over the past fifteen years or so: Kadmon and Landman (1993), Krifka (1995), and Lahiri (1998). Mine can be viewed as an attempt to push as far as possible the spirit, if not the letter, of the account developed by these authors. The approach I will sketch is most directly linked to Krifka's proposals.

As noted in Section 1.1, *ever/any* are ungrammatical in plain UE contexts:

(21) a. *There are any cookies left
      b. *There ever were cookies in this house
      c. I doubt that there are any cookies left
      d. I doubt that there ever were cookies in this house

Sentences like (21a,b) sound almost like word salad. Embedded under a DE verb like *doubt* (or in the antecedent of a conditional), the very same sentences become perfect.[8] The question is why.

As it turns out, the answer should be sought in the way these words behave "under stress". *Ever/any* in and of themselves are pretty lame, in the sense that they can be used in out-of-the-blue contexts, with no special stress condition or emphasis of any sort:

(22) a. There wasn't anybody at the door
       b. Did you ever taste this?

Nonetheless, if one puts any such sentences next to a variant thereof without *any/ever*, one clearly perceives a difference in strength/emphasis. Compare the unprimed vs. the primed version of the following sentences:

(23) a.   I do not like republicans
      a'. I do not like any republican
      b.   I will never vote for a republican
      b'. I will never vote for any republican
      c.   I do not vote republican
      c'. I do not ever vote republican

Informally speaking, the difference seems to be one of emphasis or of tolerance of exceptions. The primed versions of the sentences in (23) feel more emphatic (or less exception-tolerant) when compared with the unprimed versions. We cannot readily point to any truth-conditional difference between the members of each pair. Yet, if you had to pick one of them as the strongest, which one would you pick? This small fact is, I think, the tell-tale that reveals what is special about Polarity Sensitive Items. Proclivity to emphasis or exception intolerance is quite clearly what separates *any* from, say, *a* in (23b,b').

---

[8] The distribution of *any* is broader than that of *ever*. For example, *any* is grammatical in the scope of a modal of possibility:

(a)   You may taste any dish
(b)   *You may ever taste that dish

Possibility modals are positive (non-DE) elements. The use of *any* exemplified in (a) is often referred to as a 'Free Choice' use. We will begin to discuss it in Ch. 2. In the present section we will focus on the contexts within which both *any* and *ever* are grammatical.

This intuition is reinforced in situations where contrastive stress is required. The dialogue in (24i–iv), modeled on an example by Kadmon and Landman, is natural; the one in (24v–viii) is not:

(24)   Dialogue I
     i.   Speaker A: Do you have an egg?
     ii.   Speaker B: No.
     iii.   Speaker A: Maybe a pickled one?
     iv.   Speaker B: I don't have ANY egg
    Dialogue II
     v.   Speaker A: Do you have any egg?
     vi.   Speaker B: No.
     vii.   Speaker A: Maybe a pickled one?
     viii.   Speaker B: *I don't have AN egg[9]

Note that the reason why (24viii) is deviant is not because the indefinite article cannot be contrastively stressed, for in the appropriate circumstances, it can:

(25)   John didn't buy AN egg; he bought EVERY egg in the store

Going back to the examples in (23), our perception of the contrasts there is as clear as it is puzzling, precisely because of the truth-conditional equivalence of the elements involved. In non-contrastive environments, members of the pairs in (23) are perfectly interchangeable. The unprimed versions of the sentences in (23) are, uncontroversially, negative existential statements. Abstracting away from irrelevant details, their truth conditional import can be expressed roughly as follows.

(26)   a. $\neg \exists x \in D[\text{republican}(x) \wedge \text{I vote for } x]$
        It is not the case that there are x [x are people around here] such that x are republicans and I vote for x
    b. $\neg \exists t \in T \ \exists x \in D[\text{republican}(x) \wedge \text{I will vote for } x \text{ at } t]$
        It is not the case that there is some time t [t a part of my adult life span] and some person x [x someone I could vote for] such that x is a republican and I vote for x at t
    c. $\neg \exists t \in T[\text{I vote republican at } t]$

---

[9] Sentence (24viii) would improve if we added the emphatic element *single*:

(a)   I don't have A SINGLE egg.

But *a single egg* can act as a sort of (emphatic) polarity element; we will discuss in Ch. 3 how this is possible.

In giving these representations, I am assuming here that existential statements (like all quantified statements) come with a contextually supplied domain variable (a predicate, if you wish) that determines the range of the quantifier. I notate such a variable as "D" for individuals and "T" for time intervals (and give informal glosses of how such a variable might be employed in (26a,b)).[10] Now the truth conditions of the primed sentences in (23) appear to be just the same as in (24). This is to be expected if we assume, as most would, that *any* is simply an existential item akin to *a* or *some*, and *ever* an existential item akin to *sometimes*. But then, how on earth can the primed sentences ever come to be perceived as stronger, less exception-prone, etc. when contrasted with their unprimed counterparts?

There actually is a reasonable way to differentiate between such sentences: it is through their domains. Let me illustrate how through a simplified example.

(27) a. There aren't [$_{NP, D}$ cookies] left[11]
          $= \neg \exists x \in D[\text{cookies}(x) \wedge \text{left}(x)]$
          There aren't things [in the usual places in the kitchen] which are cookies and are left (uneaten).
      b. There aren't [$_{DP, D'}$ any cookies] left
          $= \neg \exists x \in D'[\text{cookies}(x) \wedge \text{left}(x)]$
          There aren't things [in the usual or unusual places in the kitchen] which are cookies and are left (uneaten).
      c. $D \subseteq D'$

Typically, D-variables of quantifiers are independent of one another (and we register this observation by using two distinct D-variables in

---

[10] When relevant I notate the domain variable as a subscript (on the Det or on the DP/ NP) in the syntax. For example:

(a)   There aren't any$_D$ cookies left
(b)   There aren't [$_{DP, D}$ any cookies left]
(c)   There aren't [$_{DP}$ [$_{NP, D}$ cookies left]]

See Westerstahl (1998) for a classic treatment of quantifiers as three place relations between a domain and two sets. See Stanley and Szabo (2000) for a discussion of how the domain variable may be encoded in syntax.

[11] I am assuming that bare plurals in contexts like (27a) are existentially quantified. The general approach to bare plurals I am assuming is discussed in Chierchia (1998).

(27a,b)); their value has then to be retrieved from the context. Now, if the domain associated with *any cookies* in (27b) is broader than the one associated with *cookies* in (27a) (so as to include, say, more marginal specimens of the relevant kind, such as frozen cookies, or so as to include cookies stored in unusual places), then (27b) becomes stronger/less exception-tolerant than (27a). Given these considerations, the fact that *any* is contrastively perceived as stronger than *a/some* may be due to a greater proclivity of the former to go for wider domains. It is not easy, however, to pursue this intuition in a way that is (i) compositional and (ii) compatible with the observation that emphasis or strengthening is by no means a necessary condition on the use of PSIs. What we would like to say is that when we compare a sentence with *any* with a similar one without *any*, the domain associated with *any* cannot be smaller than those of its comparison class. In other words, the distinguishing property of *any* seems to be a constraint on the *alternatives* it can be contrasted with. All this leads us to a consideration of how alternatives enter into characterizing the behavior of an item. We should, therefore, look into some version of alternative-based semantics, which in fact has been playing an increasingly important role in the treatment of a wide number of phenomena, including questions, focus, and (as we saw in Section 1.1) scalar implicatures. A brief and elementary excursion into the theory of focus will provide us with all the background we need.

1.2.2 Focus

Focus is a pervasive and complex phenomenon. One typical effect of focus is its association with focus-sensitive operators. The phenomenon is illustrated in (28):

(28)   a.   John only introduced SUE to Bill
       a'. The only person John introduced to Bill is Sue
       b.   John only introduced Sue to BILL
       b'. The only person John introduced Sue to is Bill

Focal stress (marked in caps) drives what *only* associates with in (28), and different associations may result in different truth conditions. In what follows we are going to sketch Rooth's (1985; 1992) classic theory of these phenomena (cf, Reinhart 2006 for developments). Rooth develops a bidimensional semantics in which an expression ϕ besides

its normal semantic value $\|\phi\|$, also has a focus value $\|\phi\|^{\text{F}}$, constituted by sets of values of the same type as ϕ (e.g. if $\|\phi\|$ is a proposition, $\|\phi\|^{\text{F}}$ is a set of propositions).[12] Focus-sensitive operators like *only* use alternatives to enrich the meaning of an expression. To see how this works, consider a simpler example than those in (28), with focal stress on the object.

(29) a. John only kissed [PAUL and SUE]
     b. ALT: {John kissed Paul, John kissed Sue,..., John kissed Paul and Sue, John kissed Paul and Mary,..., John kissed Paul and Mary and Sue,...}

The focal alternatives to (29a) are obtained by replacing the focused object (which denotes a plural individual comprising Paul and Sue) with other things of the same type (i.e. plural or singular individuals drawn from some contextually salient domain). Thus, the focal alternatives to (29a) have the shape in (29b). The contribution of *only* in (29a) is to state that the asserted proposition *John kissed Paul and Sue* is the only true member of the relevant set of alternatives (i.e. the only true member of set in (29b)). The way this takes place is as follows. *Only* is a binary operator. It takes two arguments: a set C of propositions and a proposition, along the lines illustrated in (30):

(30) a. Only$_{\text{C}}$ [John kissed Paul and Sue]
     b. Only$_{\text{C}}$ $\phi = \phi \wedge \forall p \in C[\ p \rightarrow \phi \subseteq p]$[13] (where $\subseteq$ = *entails*)

---

[12] Both $\|\phi\|$ and $\|\phi\|^{\text{F}}$ are built up compositionally. In particular, the definition of $\|\phi\|^{\text{F}}$ employs pointwise functional application, building on ideas developed in Hamblin (1958; 1973). The compositional expansions of sets of alternatives has come to be known as Hamblin semantics. Such ideas have been extended in e.g. Hagstrom (1998) and applied to the theory of PSI in Kratzer and Shimoyama (2002). We will discuss the latter approach in Chs 3 and 5.

[13] The standard semantics for *only* maintains, at least since Karttunen and Peters (1979), that the prejacent ϕ in Only$_{\text{C}}[\phi]$ is presupposed:

(a) Only$_{\text{C}}$ = $\lambda p\ \lambda w: p_w.\ \forall p \in C[\ p_w \rightarrow \phi \subseteq p]$

This function is defined relative to p and w only if p is true in w (which I notate as p_w). Whenever defined, it says that p (and its entailments) are the only true members of C. In a lambda term of the form $\lambda a{:}\phi.\psi$, the formula ϕ included between the column and the dot expresses the presupposition of the λ-term (i.e. the condition that a has to satisfy in order to be in the domain of $\lambda a{:}\phi.\psi$). See Heim and Kratzer (1998) for details. The presuppositional character of *only* will become relevant at several junctures in this work. For relevant discussion, see e.g. Horn (1996).

c. John kissed Paul and Sue $\wedge$ $\forall p \in$ ALT[p $\rightarrow$ John kissed Paul and Sue $\subseteq$ p]
Any member of the set of alternatives ALT which is true must be entailed by 'John kissed Paul and Sue'
(= Any member of the set of alternatives which is not entailed by the assertion must be false)

The structure (30a) can be viewed as the Logical Form of (30a) (obtained either by raising the adverb *only* or by reconstructing the subject in its VP-internal position). The first argument of *only* is the subscripted variable C ranging over alternatives (i.e. a set of propositions); its second argument is the prejacent sentence in square brackets. The role of focus, according to Rooth, is simply that of supplying the value for C. The lexical meaning of *only* is spelled out in (30b). When applied to (30a), the result is (30c). *Only* applied to *John kissed Paul and Sue* gets us back to something that states that such a proposition is true and, moreover, adds to it an exhaustivity clause: the proposition *John kissed Paul and Sue* and its entailments are the *only* true members of the set of active alternatives. In particular, propositions like *John kissed Paul* are entailed by the prejacent *John kissed Paul and Sue*, hence they have to be true; but propositions like *John kissed Mary* or *John kissed Mary and Paul* are not entailed by the prejacent, and hence they must be false. The end result is that (29a) winds up saying that John kissed Paul and John kissed Sue and performed no other kissing.

It is worth noticing (as it will later become relevant) that sometimes focus-sensitive operators like *only* can be silent or covert. A case in point is in answer to questions such as in (31).

(31) a. Who did John kiss?
      b. John kissed Paul and Sue
      c. John only kissed Paul and Sue
      d. O_C[John kissed Paul and Sue]

Normally, if one answers (31a) with (31b), such an answer is understood exhaustively, i.e. it is understood as (31c). One way of accounting for this phenomenon, proposed originally in Groenendijk and Stokhof (1984), is in terms of a phonologically null counterpart of *only* (which I notate as O in (31c)); such a phonologically null operator O is, let us say, freely available and can be used to exhaustify a statement, when

appropriate. Question–answer pairs such as those in (31) are a case in point, but besides question–answer cases, there are many other contexts where covert exhaustifications seem to come into play:

(32)   a.   A: John is fond of every new student
            B: No; he is fond of PAUL and SUE
      b.   Yesterday, John eventually decided to show up at the party. He walked in, grabbed a drink, greeted everybody, kissed Paul and Sue and then left

In (32a), we have a case of contrastive focus. Clearly, B's utterance is understood as exhaustified. Similarly, we understand the discourse in (32) as conveying the information that John kissed *only* Paul and Sue (even if *Paul and Sue* need not be contrastively stressed). These interpretations come about without resorting to an overt *only*. I assume that they are obtained by simply utilizing the covert counterpart of *only* that seems to be necessary in question–answer pairs; then in (32a) the relevant alternatives are obtained in the usual way (i.e. via focusing), while in (32b) they are determined through contextual clues (perhaps going through an 'abstract', i.e. not phonologically realized, focal marking on [$_{DP}$ Paul and Sue]).

While the existence of a covert counterpart of *only* strikes me as plausible enough on the basis of the evidence we have just reviewed, we should not jump to the conclusion that O and *only* have identical behavior and distribution. Covert operators virtually never have the same distribution as their overt counterpart. For example, the operator *Gn* involved in the interpretation of generic sentences is often characterized as a covert counterpart of quantificational adverbs like *typically* or *generally*. But it clearly does not have the same distribution as the latter (for discussion see Krifka et al. 1995; Chierchia 1995a). By the same token, the null subject *pro* posited for pro-drop languages is a null counterpart of overt (typically, reduced/clitic) pronominal elements of such languages, but it differs greatly in its distribution from the latter. Covert determiners never have quite the same distribution as their overt counterparts. How O and *only* may differ in their distribution has to be probed through empirical research (and several differences will emerge in the course of the present work).

With this final caveat, we have all we need from focus semantics:

(i)   Focus activates alternatives must be operated over by alternative sensitive operators.
(ii)   A common way of doing so is via the exhaustifying adverbial *only*.
(iii) Sometimes, exhaustification can take place without an overt *only* (i.e. covertly).[14]

1.2.3 Deriving the distribution of any/ever

This simple sketch of focus semantics provides us with a way of formulating a constraint on *any* that may readily capture its attitude towards domains. Imagine, in particular, that *any* carries an inherent focal feature F (one that may remain phonologically unrealized); such a focal feature signals that *any* associates with a set of alternatives constrained as follows:

(33)   a. $\|\text{There aren't any}_{F,D} \text{ cookies left}\| = \neg \exists x \in D[\text{cookies}(x) \wedge \text{left}(x)]$
       b. $\text{ALT} = \{ \neg \exists x \in D'[\text{cookies}(x) \wedge \text{left}(x)]: D' \subseteq D\}$
       c. $\|\text{any}_{F,D}\| = \lambda P \lambda Q \exists x \in D\ [P(x) \wedge Q(x)]$
       d. $\|\text{any}_{F,D}\|^F = \{\lambda P \lambda Q \exists x \in D'\ [P(x) \wedge Q(x)]: D' \subseteq D\}$

The focal feature on *any* is the lexical property that distinguishes it from plain existentials. Such a feature is used to code the fact that the alternatives with which a statement involving *any* can be contrasted involve existentials with smaller domains (the lexical specification that affords what we want is given in (33c,d), a constraint on *any*'s alternatives).[15]

---

[14]   I assume that the meaning of O is identical to that of *only*, but is not presuppositional (cf. n. 13):

(a) $O_C = \lambda p \lambda w[\ p_w \wedge \forall p \in C[\ p_w \rightarrow \phi \subseteq p]]$

The reason for this is that a sentence like *He is fond of Paul and Sue* in example (32a) is analyzed as having O in its semantic representation and clearly has no special presupposition. This issue will become relevant in Ch. 4, where further differences between O and *only* will emerge.

[15]   This "abstract focus" idea is just a "stepping stone", to be modified in Ch. 3. We will argue there that the alternatives associated with *any* are the very same alternatives that are associated with its ordinary scalar counterparts *some/a*. We will see, in particular that ordinary indefinites like *some/a* give rise to "Free Choice" interpretations and that such interpretations may be explained by assuming that *some/a* activate subdomain alternatives. The key difference between *some* and *any* will turn out to be that the alternatives associated with *some* are active only when relevant, while the alternatives associated with *any* are always active.

This hypothesis on the meaning of *any* suffices to capture the observations made above concerning the contrast between *any* vs., say, *a*/*some*. For one thing, we can set the original D on an occurrence of *any* as we please. So it can be chosen in such a way as to make (34a) completely equivalent to (34b):

(34) a. There aren't any$_D$ cookies left
     b. There aren't [$_D$ cookies] left

This accounts for the fact that *any*, out of the blue, can be as lame as we like. At the same time, when it comes to comparing, (33a) can only be compared with existentials with a smaller domain, because of the lexical constraint in (33d). This may be at the basis of our intuitions about the pairs in (23). In particular, let us see what happens in situations of contrastive stress, such as those exemplified in (24) above. To see how an account of the contrast in (24) might go, let us briefly consider Rooth's proposal on contrastive focus, which is very simple. According to Rooth, a contrastively focused item must find *among its focal alternatives* some element in the surrounding discourse distinct from the assertion itself. Contrastive focus, in other words, acts like an anaphoric component that seeks an antecedent in the context. For example, the discourse in (35) is well formed, because the shape of the focal alternatives of the second sentence allows us to find a suitable antecedent in the first sentence; this is not so for (35c):

(35) a. John was wearing an old coat. Bill was wearing a NEW one
     b. ‖John was wearing an old coat‖ ∈ {Bill was wearing an A coat: A ⊆ ADJ}
     c. ?? John was riding a bike. Bill was showing off a NEW coat
     d. ‖John was riding a bike‖ ∉ {Bill was showing off an A coat: A ⊆ ADJ}

Let us now apply Rooth's approach to the case of (24), beginning with the deviant dialogue in (24v–viii) repeated here in simplified form:

(36) a. Speaker A: Do you have any$_{D_1}$ egg?
     b. *No I do not have AN$_{D_2}$ egg

The problem is that D-variables are covert elements. By definition, they cannot be stressed. Hence, stressing the indefinite article is not going to activate domain alternatives; it can only activate functions of the same

type as *a(n)*, i.e. positive quantifiers. Thus the alternatives activated by stressing *an* in (36) are going to look as follows:

(37)   {I do not have an$_{D_2}$ egg, I do not have two$_{D_2}$ eggs, . . . , I do not have every$_{D_2}$ egg}

Clearly, there is no way of construing D1 in (36a) in such a way as to make it a member of the set in (37), as Rooth's condition on contrastive focus would demand. If we set D1 = D2, then (36a) = (36b); hence no contrast arises. If we set D1 ≠ D2, then (38a) is not a member of the focal alternatives in (37), in violation of Rooth's anaphoric constraint. Let us compare (36) with the well-formed version of the dialogue in (24), also repeated below in simplified form:

(38)   a.   Speaker A: Do you have an$_{D_1}$ egg?
        b.   Speaker B: I do not have ANY$_{D_2}$ egg

It is reasonable within our set of assumptions to regard contrastive stress on *any* as a way of spelling out its inherent focal feature; our chief (and sole) stipulation on the lexical meaning of *any* is that such a focal feature is associated with subdomain alternatives. So while D-alternatives cannot in general be activated via stress, through *any* they can—as a consequence of its lexical semantics. So, following Rooth, for (38b) to satisfy the condition on contrastive stress, we must have the following:

(39)   a. $\|$ I don't have an$_{D_1}$ egg $\| \in \|$I don't have any$_{F,D_2}$ egg $\|^F$
        b. $\|$ I don't have an$_{D_1}$ egg $\| \in \{ \neg \exists x \in$ D'[egg(x) $\wedge$ I have x]: D' $\subseteq$ D2}

Given the constraint on *any*'s alternatives, (39a) is equivalent to (39b). This requires that the domain variable associated with *an egg*, marked as D1 in (38a), must range over some subset of the one associated with *any egg*, marked as D2 in (38b)—a condition that clearly can be satisfied. This readily explains the naturalness of the dialogue in (38) and the contrast with that in (36). We therefore derive the observation that, when contrastively stressed, *any* acts as a "domain widener" (though when it is not contrastively stressed it is interchangeable with a plain indefinite).

  Thus, the use of a lexical constraint on *any*'s alternatives seems to be a good tool to capture its greater proclivity (when compared with other existentials) for emphasis or intolerance for exceptions. In our discussion, we saw in particular how such a proclivity emerges in