

Scalar Implicatures and Beyond

Theory, Acquisition, and Processing

By

Daniele Panizza

Scalar Implicatures and Beyond: Theory, Acquisition, and Processing

By Daniele Panizza

This book first published 2024

Ethics International Press Ltd, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2024 by Daniele Panizza

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical photocopying, recording or otherwise, without the prior permission of the copyright owner.

Print Book ISBN: 978-1-80441-665-5

eBook ISBN: 978-1-80441-666-2

Table of Contents

Preface	vii
 Chapter 1: Theories of Implicatures.....	1
1.1 Paul Grice and the birth of implicatures	1
1.2 Types of implicatures.....	6
1.3 Implicatures in post-Griceans: The Relevance Theory.....	10
1.4 Neo-Griceans, grammar, lexicon, and scalar implicatures	14
1.5 Entailment patterns and embedded implicatures (Chierchia, 2004)	22
1.6 Ignorance inferences and embedded implicatures in complex sentences (Sauerland, 2004).....	29
1.7 Scalar implicatures as exhaustification.....	34
 Chapter 2: The comprehension and processing of scalar implicatures	46
2.1 Predictions on how scalar implicatures are derived and their empirical assessment	46
2.2 The frequency of scalar implicature derivation and its cognitive cost.....	51
2.3 The processing of scalar implicatures in real-time	59
2.4 Numerals, scalar diversity, and ad-hoc implicatures.....	69
 Chapter 3: The acquisition and development of scalar implicatures	78
3.1 Theories of scalar implicatures, cognitive science, and cognitive development.....	78
3.2 Are children more logical than adults? Studies on the acquisition of scalar implicatures.....	82

3.3 Tracking children's derivation of scalar implicatures with neuroscientific methods.....	86
3.4 The acquisition of scalar implicatures triggered by numerals	90
3.4 The acquisition of scalar implicatures generated by disjunction	96
3.5 The interpretation and processing of indirect scalar implicatures in children.....	103

**Chapter 4: Scalar implicatures in complex sentences redux:
Evidence for an exhaustification-based approach.....116**

4.1 Exhaustification and the problem of clausal implicatures (again).....	116
4.2 A new look to scalar implicatures in disjunctions: An exhaustification-based account.....	118
4.3 Scalar implicature in conditionals: The case of conditional perfection	126
4.4 Exhaustification and context: Ad-hoc implicatures and other issues	133
4.5 A final puzzle: Indirect implicatures, the question under discussion, and intonation.....	138

Preface

This book is the culmination of my two-decade-long journey dedicated to the study of scalar implicatures. It began as the topic of my master's thesis, and I continued to explore it in my doctoral thesis and through numerous experimental and theoretical works. The present volume is an enriched version of my doctoral dissertation, updated with more recent experimental works and ideas. Many of these works, carried out by me, collaborators, and other scholars in the field, were not published when I defended my PhD, about fifteen years ago.

More than two decades have passed since some of the most influential articles have been published in contemporary pragmatics, within the domain of neo-Gricean approaches, Relevance Theory, grammatical theories, and other accounts based on alternative semantics. In these years, experimental pragmatics has been a flourishing environment: a multitude of experimental studies on pragmatic processing have been conducted, and implicatures are still a hot topic, with hundreds of papers published every year and works presented at the most important conferences across the fields of linguistics, psychology, philosophy, and neuroscience. Scalar implicature is still one of the most debated topics in modern theoretical and experimental linguistics, and I believe this is so for two main reasons: First, there is no consensus on the basic facts about implicature derivation and processing. Theoretical approaches are plagued with problems, experimental outcomes are irreconcilable, and the whole idea of scalar implicature sometimes seems ineffable. I heard countless times the following statement uttered by colleagues and friends in the field: "I am fed up with scalar implicatures; I want to work something else!". Yet, many scholars, including me, are obsessed with this topic. This leads us to the second reason underlying the reputation and attractiveness of scalar implicatures: they are genuinely and inherently a multidisciplinary phenomenon that

allows the study of the human mind through language, reasoning, thought, and online performance.

This last consideration brings me to illustrating the structure of the present book, which tackles the study of scalar implicatures from a multidisciplinary perspective across theory, comprehension, acquisition, and processing of pragmatic inferences. The first chapter illustrates the classic view of Paul Grice on implicatures, followed by more recent developments introduced by Gricean scholars and contemporary linguists and philosophers. The second chapter is devoted to the comprehension of scalar implicatures investigated using experimental means. The third chapter overviews the most influential studies on the acquisition of these inferences in children. The fourth chapter includes some new ideas that address several problems about implicature computation, and it presents some novel hypotheses and proposals on how we may tackle these issues.

This book offers a comprehensive survey of the most classic and influential works on scalar implicatures, and it is written with informal prose accessible to readers who do not possess strong expertise in empirical science or knowledge of formal semantics/pragmatics. As such, it is suited to a broad range of students, linguists, philosophers, and psychologists who begin to take an interest in this topic. However, it is also addressed to scholars who wish to integrate their knowledge about the theory, acquisition, and processing of scalar implicatures. It is a concise compendium of the most relevant literature to date. Finally, the last chapter hints at what type of problems linguists and cognitive scientists who investigate these phenomena have to cope with, and its goal is to offer one viable perspective on how to work out some possible solutions.

My credo, akin to the original spirit of Grice, is that understanding language and communication is a highly interactive process that nonetheless requires an implicit mastery of symbolic and logical-like representations. Scalar implicatures are one paradigmatic example

of this process: an informational space in which speakers' knowledge, intentions, and communicative strategies consistently interact with informativity and logical strength. The former are conceived as extra-linguistic creatures, while the latter are understood as abstract objects on which Boolean-like algorithms operate.

In conclusion, there are many people that I'd like to thank who have provided me with unconditioned help and support during these years. The first of list is my mentor, Gennaro Chierchia, who drove me into studying scalar implicatures and always assisted me with patience and human warmth. I'm also indebted to Lyn Frazier, Maria Teresa Guasti, Remo Job, Chuck Clifton, Stephen Crain, Rosalind Thornton, Francesco Vespignani, Hedde Zeijlstra, Clemens Mayr, Yasutada Sudo, and Jacopo Romoli, who supported me during all these years, helped me in difficult times, and taught me an infinite amount of things, as well as my collaborators on the works that I discuss in this book, such as Jesse Snedeker, Yi Ting Huang, Karoliina Lohiniva, Maik Thalman, Greta Mazzaggio, Francesca Foppolo, Luca Surian, and Anna Notley.

All the mistakes included in this book are mine.

Chapter 1

Theories of Implicatures

1.1 Paul Grice and the birth of implicatures

In the last fifty years, since Paul Grice wrote *Logic and Conversation* (1975), *implicature* has become a central notion in the study of pragmatics across the fields of philosophy of language and those belonging to cognitive sciences such as linguistics, psychology, and neuroscience. It is not so common in the history of language sciences that a single construal has drawn the interest of so many scholars across theoretical domains, inspiring thousands of publications and leading to the claim, shared by many, that Grice is the father of modern pragmatics. Given that the present volume is devoted to studying scalar implicatures, one kind of Gricean pragmatic inferences, the present chapter starts by illustrating Grice's groundbreaking intuitions that changed the study of language and human communication.

Let us first introduce a basic definition of implicature.

- (1) An implicature is an inference that enriches the meaning of a proposition.

The definition in (1) includes two core properties of implicatures. First, it makes their inferential nature explicit: an implicature's informational content comes from some deductive reasoning whose outcome provides novel information that the original proposition did not directly encode, constituting the trigger and the premise of such deduction. Second, propositions like utterances spoken in a dialogue or sentences written in a text are subject to pragmatic enrichment; their interpretation goes beyond the literal meaning of the lexical items composing the sentence. Grice's fundamental intu-

ition about how linguistic communication unfolds is that the amount of information that is carried by a sentence uttered by a speaker in a given language goes beyond what is strictly coded in the meaning of the words included in the utterance itself. Implicatures are responsible for most, if not all, of this informational content, which does not overlap with the so-called “plain” sentence meaning. Implicatures add some novel content responsible for pragmatic enrichment, they are voluntarily conveyed by the speaker and inferred by the addressee. This idea was not entirely new, indeed it had been entertained by other philosophers that laid the foundations of modern pragmatics, such as John Austin and John Searle (cf. Levinson, 2000). However, the pioneer contribution of Grice’s works was to outline the direction in which a modern scientific theory could account for implicatures. Grice provided us with the notion of implicature, the tools, and fundamental insights to explain why they exist, how they work, and how they can be reproduced in a formal system.

The Gricean starting point highlighted a problem with language meaning concerning a specific class of words. He noticed that there exist discrepancies between the meaning of logical operators, such as \neg , \wedge , \vee , \rightarrow , $\forall x$, $\exists x$, $\neg\exists x$, and that of their counterparts in the natural language, words like *not*, *and*, *or*, *if*, *all*, *some*, and *the*, respectively. For instance, consider the sentence in (2), which includes the existentially quantified constituent “a woman”:

(2) I am meeting **a woman** this evening.

If someone utters sentence (2), there is a strong belief in the listener that the woman the speaker is talking about is not his girlfriend, his mother, his sister, or anyone known by both the speaker and the listener. While such a meaning appears to be trivial and immediately available to a listener of sentence (2) or the reader of this book — i.e., it does not require any particular effort to be understood — it is challenging to account for it within a linguistic formal framework. The problem regards how to encode in the semantics of the existen-

tial quantifier *a*, or the semantics of the Noun Phrase “a woman”, the information that this person is not in a close relationship to the speaker of the utterance or known to the agents of the conversation. Clearly, the indefinite *a* does not always convey this interpretation, as shown by the example in (3):

(3) What **a woman** my mother is!

Notice that in a sentence (3), the constituent “a woman” actually refers to the speaker’s mother. Let us take the first-order logic existential operator ($\exists x$) as the basic meaning of the existential quantifier *a*. Something, thus, must happen during the reading of (2) such that it leads us to conclude that the reference to the noun phrase “a woman” is not the speaker’s mother, while in sentence (3) it is. If we encode in the lexicon the meaning of “a woman” as a female individual *x* such that *x* is not the mother of the speaker, such a definition would fail to capture the meaning of “a woman” in (3). However, there is more to this puzzle. Imagine that the speaker of sentence (2) continues his utterance by adding: “... she is my mother!”. By doing so, he overrides the immediate intuition that the woman he is talking about is not his mother by explicitly stating that she is.

The example above illustrates how effective Grice’s implicatures are in helping us solve such interpretation puzzles. The inference “that woman is not speaker’s mother” in (2) is an implicature: it is a pragmatic inference that the speaker has invited the listener to endorse. However, this inference is not strictly part of the lexical content of the expression “a woman” and can be canceled without any sense of contradiction, as shown in the previous paragraph. This move allows us to keep existential quantification as the true meaning of the quantifier *a* while still being able to explain why the constituent “a woman” in sentence (2) can exclude the speaker’s mother from its reference. Grice’s idea, as exposed in *Logic and Conversation*, is that the gap between the basic meaning of terms like *a* and the interpretation we convey by using such terms in a typical communica-

tive exchange is filled in by rational and conversational machinery, composed by one principle and four maxims in which "... anyone who cares about the goals that are central to conversation/communication (such as giving and receiving information, influencing and being influenced by others) must be expected to have interest" (Grice, 1975, p. 49). The Gricean Cooperative Principle and the four maxims are implicit rules followed by the agents of a conversation, and they are reported below.

Cooperative Principle:

- Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Maxim of Quantity:

- Make your contribution as informative as is required.
- Do not make your contribution more informative than is required.

Maxim of Quality:

- Do not say what you believe to be false.
- Do not say that for which you lack adequate evidence.

Maxim of Relation:

- Be relevant.

Maxim of Manner:

- Avoid obscurity of expression and ambiguity.
- Be brief and orderly.

The Gricean Cooperative Principle and the four maxims are

universal: Speakers and hearers of any language observe them. It is from following the maxims and the principle that implicatures are derived. During any communication exchange, there is a tacit and mutual agreement that all conversation agents observe the maxims. One exception occurs in a lying situation, namely when the speaker is not telling the truth and voluntarily violates the Maxim of Quality, with the hearer unaware of that. Another issue is when the conversation participants know the speaker violates one or more maxims. For example, he might opt out from the Cooperative Principle by stating: "I can't say more, my lips are sealed"; he may be faced with a clash (he cannot be informative as required without having adequate evidence of what he is saying) or he may blatantly fail to fulfill it (see Grice, 1975, for plenty of examples of this sort).

The universal nature of the maxims and the principle originates from the behavior of rational, cooperative agents. The maxims are not rules of good behavior, nor is their essence descriptive. They are implicit rules that humans seem to observe while they communicate. They have a normative character. Grice is the first to say that his formulation of the maxims is far from definitive. Whatever model of human communication might come out in the future, though, it will have to account for what the maxims and the principle succeed in predicting: the generation of conversational implicatures. Grice further states that they are not likely to be learned like a good habit of not telling lies. They are not a standard practice that most speakers follow, but something *reasonable* for us to follow, not something quasi-contractual, but a characteristic of all the talk exchanges. Thus, echoing the Kantian categories, Grice indicated the existence of a rational cognitive apparatus in charge of dealing with talk, maxims, and implicatures.

1.2 Types of implicatures

According to Grice, there are two classes of implicatures: conventional and conversational implicatures. Conventional implicatures refer to inferences intrinsically tied to certain words and cannot be effortlessly canceled. Levinson (2000) observes that Grice did not provide many examples of conventional implicatures. A paradigmatic example of such inferences might be the contrastive interpretation introduced by adversative connectives like *but*. While at the semantic level, the meaning of *but* shares the same truth conditions of conjunction (i.e., p *but* q is true if both p and q are true), it further introduces an additional inference according to which q is not expected to hold given p . Consider the sentence in (4).

(4) Mary loves chocolate, *but* she rarely eats it.

The first clause states that Mary is fond of chocolate, which suggests that she often eats it, whereas the second sentence denies this expectation. Notice that this inference cannot be canceled; it is parasitic on using the word *but*. Indeed, conventional implicatures do not depend on the Gricean maxims.

Conversational implicatures consist of the most famous and more intensely studied type of Gricean inferences, derived from exploiting the maxims and, unlike conventional implicatures, are defeasible; that is, they can be canceled without any sense of contradiction. An example of a conversational implicature is the inference triggered by sentence (2).

The author provides a sketch of the cognitive process of “working out” a conversational implicature q by uttering a proposition p : “he said that p ; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that q , he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that q is

required; he has done nothing to stop me thinking that *q*; he intends me to think, or is at least willing to allow me to think, that *q*; and so he has *implicated* that *q*" (Grice, 1975, p. 50). If this process cannot be tracked back, we are not dealing with an implicature, or it is conventional.

Conversational implicatures are, in turn, subdivided into two categories: Particularized Conversational Implicatures (PCIs) and Generalized Conversational Implicatures (GCIs). PCIs are conveyed by saying something on a particular occasion in virtue of unique features of the context. A classic example of a PCI is the inference suggested by the following sentence, used as a recommendation letter for a job.

(5) The candidate has beautiful handwriting.

If this is all the speaker has to say about the candidate she is asked to write a letter for, the effect is that sentence (5) invites the reader to derive the following inference:

(6) The candidate is not very good.

The implicature in (6) arises because the writer of the recommendation letter could have said much more about the candidate, but she decided not to do so. From the assumption that the writer is following the Maxim of Quantity (say enough), she is telling the truth (Maxim of Quality), and she is saying something relevant to the purpose of recommending someone (Maxim of Relation), the reader infers that there was not much to say about the candidate, whence the inference in (6). In the case of PCIs, the context plays a pivotal role. Consider a situation in which sentence (5) was written in a recommendation for a scribe job; in such a context, it would be a good recommendation, and the implicature in (6) would not arise.

GCIs, instead, rely upon the use of a specific form of words in an utterance (e.g., the indefinite *a* in sentence (2)), which under normal

circumstances triggers such implicatures. For instance, every time a speaker utters a simple Noun Phrase like “a woman” in a sentence such as (2), he triggers the following reasoning in the listener. Assuming that the speaker is cooperative (Principle of Cooperation) and telling the truth (Maxim of Quality), that he is providing relevant information to the conversation (Maxim of Relation) in a concise way (Maxim of Manner), and he is conveying enough information for the purpose of the communication (Maxim of Quantity), he invites the listener to draw the implicature that the woman he is talking about is not his mother, or anyone known to the listener, unless further information is provided.

Although GCIs are more systematic than PCIs, for they bear a particular relation with the specific words that are responsible for their derivation, according to Grice, all the conversational implicatures are drawn by a cognitive system, probably a reasoning system, which deals with the communicative routines between the speaker and the hearer. Because of this, the study of implicatures historically constituted one of the main topics of pragmatics of natural languages. Indeed, the main goal of pragmatics is to investigate the influence of the *context* and the *use* of language in a real conversation. These two terms are central to the study of implicatures and often recur in all the Gricean and post-Gricean works on this topic.

The factory mark of a conversational implicature is that its perceiver must entertain some work of intention recognition. This feature is called *calculability*. Even in the case of a GCI, in which the inference is non-detachable by the words uttered, nor does it depend upon a particular utterance context, the same pragmatic process is at work. *Non-detachable* means that “it is not possible to find a way of saying the same thing, which simply lacks the implicature in question, except where some special feature of the substituted version is itself relevant to the determination of an implicature” (Grice, 1975, p. 58). Another critical aspect of the conversational implicatures is their *cancellability*. An implicature can be canceled or blocked by

uttering, after its trigger, something that overrides the meaning of the implicature without any sense of contradiction, as exemplified by sentence (3).

As mentioned at the beginning of this chapter, Grice claims that implicatures account for the informational gap between the basic meaning of words, defined as *what is said*, and the meaning speakers convey in a real conversation. The additional meaning component carried by pragmatic inferences is called *what is implicated*. Let us remark that these are two different types of meaning. The first one, e.g., the lexical meaning of the existential quantifier *a*, is what in the tradition of theoretical semantics is called truth-conditional meaning because it determines the truth conditions of a sentence, that is, the circumstances under which a sentence is true. The basic meaning of a proposition does not fade away after pragmatic enrichment due to the addition of an implicature, as witnessed by the fact that sentence (6) does not contradict sentence (5). The second one, what is implicated, is a non-truth-conditional meaning. It is a portion of informational content that is not encoded in nor follows from the truth-conditional meaning of the original proposition. Grice attempts to frame the implicated meaning within the framework of cognitive science theories by maintaining that it results from a process of intention recognition, whose output is not strictly coded into the signal emitted by the speaker. The sum of what is said and what is implicated amounts to the whole communicated meaning: *what is meant*. Grice followers (Sperber and Wilson, 1989; Atlas and Levinson, 1981; Horn, 1972; among others) will try to pursue the same goal. While maintaining the tenets of the Gricean proposal, they will redefine the four maxims into a set of heuristics, which communicative agents need to narrow the spectrum of all the possible ways to enrich the coded meaning of a proposition. They are the premises, taken for granted by every participant of any verbal communication, on which the deductive inference system can work to yield its output, which is an implicature.

1.3 Implicatures in post-Griceans: The Relevance Theory

Post-Gricean scholars took two divergent routes. Some of them have maintained the view that a mighty reasoning and intentional system generate implicatures but collapsed the four maxims into only one maxim of Relevance (Sperber and Wilson, 1986; Carston, 1998). This approach stems from the theoretical framework Sperber and Wilson advanced, called Relevance Theory (RT). Some other theorists (the so-called neo-Griceans: Gazdar, 1979; Atlas and Levinson, 1981; Horn, 1989, 2004) retain the original Gricean apparatus but restructured the four maxims into four cognitive heuristics. I will introduce the neo-Gricean framework in the next section.

Sperber and Wilson's proposal builds on the core concept of *Relevance*. The authors define Relevance as "a potential property not only of utterances and other observable phenomena but of thoughts, memories and conclusions of inferences. In relevance-theoretic terms, any external stimulus or internal representation that provides input to cognitive processes may be relevant to an individual at some time. [...] Utterances raise expectations of Relevance not because speakers are expected to obey a cooperative principle and maxims or some other specifically communicative convention, but because the search for Relevance is a basic feature of human cognition" (Wilson & Sperber, 2004, p. 608). These words attest to a central departure from Grice's original apparatus. The cognitive principle of Relevance is pervasive; it is not restricted to verbal exchange but regards any kind of communication. The need for specific maxims or sub-principles to account for the speaker's expectations or intentions is ruled out. Instead, expectations and intentions constitute basic features of human cognition. The authors admit a resemblance between their principle of Relevance and the Gricean maxim of Relation ("be relevant"), from which it is explicitly inspired. However, they point to the fact that Relevance is not a maxim in the Gricean sense. Instead, Relevance is the property of a stimulus to yield a *positive cognitive*

effect, which in turn is a worthwhile difference to the individual representation of the world. The most important positive cognitive effect is the *contextual implication* (the counterpart of the implicature for the RT), which is a conclusion deducible from the linguistic input and context together. From this idea, it follows that context is always required to compute an implicature. The authors linked the theoretical notion of Relevance to the psychological concept of cognitive effort. The greater the positive cognitive effect achieved, the greater the Relevance of the input. The greater the processing effort expended, the lower the Relevance. Thus, pragmatic enrichment results from some cognitive assessment between the cost (cognitive effort) and the benefit (cognitive effect) of deriving a contextual implication.

The Cognitive Principle of Relevance states that human cognition tends to be geared toward maximizing Relevance. However, an extra layer of intention is needed to obtain inferential communication. The *informative intention* is to inform the audience of something. The *communicative intention* is to inform the audience of one's informative intention. Understanding in a typical communication exchange is only achieved if the communicative intention is fulfilled. In order to specify the role of Relevance in a communicative exchange, the authors stated the Communicative Principle of Relevance: every ostensive (i.e., intentional) stimulus conveys a presumption of its *optimal Relevance*. The concept of optimal Relevance defines the threshold over which the communicative effort must dwell. An ostensive stimulus is relevant enough to be worth the audience's processing effort, and it is the most relevant concerning the communicator's abilities and preferences: it entirely relies upon the addressee's expectations, effort, and Relevance.

Along with the RT framework, the comprehension procedure follows a path of least effort in computing cognitive effects. This procedure can be exemplified as follows: test interpretative hypotheses (disambiguation, reference resolutions, implicatures, and the like) in order

of accessibility, and stop when your expectations of Relevance are satisfied. This algorithm corresponds to the relevance-theoretic recipe to derive pragmatic implicatures:

1. The listener constructs appropriate hypotheses about explicit content via decoding, reference disambiguation, and pragmatic enrichment. Such hypotheses are defined as *explicatures*.
2. The listener applies the same procedure concerning the intended contextual assumption (*implicated premises*).
3. Hypotheses about contextual implications can be drawn; these are called *implicated conclusions*, which consist of the counterpart of Gricean implicatures.

Nonetheless, the authors warn that the interpretation path can follow a different order. Given that comprehension is an online process, it may be the case that background expectations and anticipatory processes start computing implicated premises and conclusions while the explicit meaning of the sentence is yet to be worked out. However, since the explicit content must warrant the implicated conclusions, this is a good reason to treat lexical narrowing (explicatures and enrichment) as falling on the explicit rather than the implicit side. GCIs, then, are more likely to be treated as explicatures whereas PCIs are part of the implicated content.

Let us see how the RT algorithm works in the two examples of implicatures I illustrated in the previous paragraphs. As for the reading of the expression “a woman” in sentence (2), as it refers to women unknown to the listener, the RT predicts that this meaning comes about as an explicature. This prediction comes from the interpretation of a referential Noun Phrase, which is a case of reference disambiguation (who is the woman the speaker is talking about?) and, possibly, a case of lexical narrowing, whereby the unspecified meaning of *a* is narrowed down to a more specific interpretation. In this case, the conversational context, e.g., the shared knowledge

between the speaker and the listener, does not play a significant role. Instead, the indefinite *a* generally leads to this meaning when used in such a sentential environment, which resembles the definition of a GCI. In order to obtain a contextual implication, what an implicature is within the framework of the RT, more cognitive work is needed. For instance, let us assume that the speaker wants to tell the listener that he is going on a date. The listener knows that the speaker is single (implicated premise). After enriching the proposition in (2) with the explicature “the speaker will meet a woman that is not known to me”, the listener may wonder why what he uttered was relevant to their mutual knowledge. That is, the listener is engaged in recognizing the speaker’s communicative intention. From the assumption that the speaker is single and that he ostensibly communicated to the listener about the upcoming meeting, the listener quite effortlessly infers that he is going on a date. Nevertheless, the speaker may withdraw this contextual implication just as a standard implicature by adding further information. For instance, the speaker may add the following statement:

(7) It’s not what you are thinking! She is just an old friend.

The utterance in (7) denotes that the speaker knew about the contextual implication the listener may have drawn, considering the contextual premises, and wanted to warn the listener from endorsing such a conclusion. This example highlights the communication interplay during a typical conversational exchange, imbued with communicative intention and intention recognition. It also illustrates one advantage of separating explicatures from implicatures, with the former being triggered by lexical content or linguistic form vs. the latter resulting from proper cognitive reasoning.

One of the problems with RT is due to the departure from Gricean maxims. Considering GCIs as explicatures can help distinguish them from contextual implications, but it does not offer a systematic way to predict their derivation, which was the main goal of the four

maxims. If we eliminate them, we throw away a predictive mechanism that accounts for most of the GCIs, and scalar implicatures, illustrated in the next section, are one very well-studied case of such inferences. Secondly, the calculation of GCIs needs to access the proper meaning of the proposition (what is said) or at least a part of it. How will such implicatures be derived if the implicit hypotheses are built before the explicit content is available? If, by contrast, the implicit hypotheses need pragmatic enrichment to be completed, GCIs must be generated later, in a so-called post-semantic stage. This view is endorsed by most of the linguists and psychologists advocating the RT (e.g., Noveck, 2001, 2004). As we will observe in the following sections, there are many reasons why global and post-semantic derivation of implicatures appears problematic, both from theoretical and psycholinguistic perspectives.

In conclusion, while the RT may be suitable for defining a mechanism whereby PCIs and contextual inferences can be computed, it leads to significant loss of predictive power as soon as we move towards more systematic types of implicatures. Insofar as general cognitive processes enrich the meaning of what is said, the effort requested by a general cognitive system to handle the generation of those more systematic and less context-dependent inferences such as GCIs seems too demanding. In the next paragraph, I will review the main aspects of the neo-Gricean framework, which has often opposed the RT, for it provides a revision of the Gricean maxims instead of attempting to substitute them with a unique cognitive principle.

1.4 Neo-Griceans, grammar, lexicon, and scalar implicatures

In this paragraph, I will focus on the so-called neo-Gricean theories of implicatures; I will introduce scalar implicatures, the main topic of the present work, and discuss the relationship between such inferences and the grammar. Let us first spend a few words about

the relationship between pragmatics and the grammar of natural languages. Following the Chomskyan tradition (Chomsky, 2002), one way to conceive grammar is by stating that it is the set of derivational rules that transform a sensorial input, a string of spoken or written words, into a conceptual representation. In modern linguistic frameworks, such rules can be resumed in *merge* and *apply* (cf. Chierchia, 2013). The former comes from the latest Chomskyan proposal, the minimalism program (Chomsky, 2014), and it can be defined as a recursive binary operation that integrates words and constituents into larger units. We can think of the latter as the semantics counterpart of merge; it comes from the seminal work of Richard Montague (Montague, 1970), and it constitutes the functional application of a logical rule. It takes as input a formula of a defined type (basic types are *e* for individual, *t* for truth value, and *w* for possible word or circumstance) and returns as output a formula of another type. Semantic composition proceeds following the syntactic tree (either bottom-up or top-down, depending on the directionality of interpretation) until the last node is interpreted. The final output is the truth conditions of the proposition.

Both these operations, merge and apply, are argued to operate in an automatic fashion. Namely, they are blind to external considerations like speaker intentions and relevant contextual knowledge. They are the basic compositional and recursive operations through which lexical elements are combined into phrases. Under the classic Gricean view, pragmatics processes are post-semantics or post-compositional since they operate on interpreted propositions. This idea is illustrated in the model that Levinson called modularism *à la* Grice-Fodor-Chomsky (Levinson, 2000) because syntax, semantics, and pragmatics are seen as blind modules, each serving as an input for the other. Under this framework, implicatures undergo pragmatics rules and algorithms that have to wait for the grammatical analysis to be completed in order to produce a meaningful output; hence, they are post-compositional processes. Along with an alternative

view, implicatures may be derived within the grammar, handled by the same operations (merge and apply) responsible for generating an interpretable output from a string of text.

A promising environment to investigate the relation between grammar and pragmatic inferences is that of scalar implicatures. Such inferences are derived by exploiting the Gricean Maxim of Quantity and, more precisely, the first submaxim: make your contribution as informative as required. Scalar implicatures are upper-bounding GCIs associated with scalar values. This terminology refers to the fact that scalar triggers, such as the indefinite *some*, are semantically provided with a lower bound, such as existential quantification, but lack an upper bound. To understand this terminology, consider the sentence below.

(8) John ate some of the cookies on the table.

The quantifier *some* in sentence (8) ensures that there is at least a cookie that John ate, i.e., the lower bound, but leaves open the possibility that John ate all the cookies on the table: this possibility represents the lack of an upper bound. We will see in the following paragraphs that a scalar implicature is responsible for assigning to sentence (8), and more specifically to the reference of the quantified phrase “some of the cookies on the table”, an upper bound.

Let us first provide some historical background to scalar implicatures. One century before Grice, Sir William Hamilton individuated two different types of *some*, namely the *indefinite some* (“some and maybe all”) and the *semi-indefinite some* (“some but not all”). A few years later, as Horn reports (Horn, 1990), Stuart Mill recognized that the inference from “some” to “some but not all” belongs to an extra-logical domain. According to Grice, then, scalar implicatures are driven by the tacit and mutual knowledge, shared by the speaker and addressee obeying the conversational maxims and the Cooperative Principle, that the speaker uttered a logically weaker sentence

because he was not in the position to utter a stronger one. Such a stronger proposition is exemplified in (9)

(9) John ate all of the cookies on the table.

The Gricean reasoning unfolds as follows. The listener knows that the speaker obeys the Cooperative Principle and the conversational maxims. Hence, the listener infers that the speaker must have uttered a proposition that was informative enough for the purpose of communication (Maxim of Quantity) and for which she did have enough evidence (Maxim of Quality). If the speaker was in the position to utter a more informative proposition, such as (9), she should have done so. Because the speaker decided to utter (8) instead of (9), it follows that the sentence (9) does not hold.

Again, consider (8), which contains the scalar item *some*. Sentence (10) represents the proposition (8) strengthened via a scalar implicature, derived as the negation of (9). The technical term *scalar strengthening* stands for the fact that the new proposition (10) containing the original one (8) plus the implicated meaning, the denial of (9), is logically stronger than the proposition without the implicature, for sentence (10) logically entails sentence (8).

(10) John ate some of the cookies on the table, but he didn't eat all of them.

How do we obtain (9) as an alternative to (8)? Critically, logical strength alone cannot be the answer to this question. Indeed, infinitely many other propositions are logically stronger than (8), and their negation does not lead to a meaning such as the one carried by (10). Consider, for instance, sentence (11):

(11) John ate some of the cookies on the table and some apples from the chest.

Sentence (11) is logically stronger than (8) because if (11) is true,

it follows that (8) must be true as well. Thus, sentence (11) might have served as an alternative to (8), just like sentence (9). However, sentence (8) plus the negation of (11) leads to a different interpretation than (10), namely, “it is not true that John ate some of the cookies on the table and some apples from the chest”, that is, he only ate some of the cookies on the table. Clearly, this interpretation is not what we want as a result of the scalar strengthening of (8).

Philosophers of language like Paul Grice struggled to solve this problem satisfactorily until Larry Horn’s proposal (cf. Horn, 1972). Horn states that to derive a scalar implicature, we must have a *unilateral* (or asymmetric) *entailment scale*. The scale <some, all> meets this requirement because *all* unilaterally entails *some*, i.e., “John ate all of the cookies” entails “John ate some of the cookies”, but the converse does not hold. Other scales that potentially trigger scalar implicatures are adverbials (<sometimes, always>), logical operators (<and, or>), modals (<may, must>), determiners (<the, a>), and so on. As Horn formulated this model, he became well aware of a potential problem: How can we avoid an uncontrolled proliferation of the scales? For instance, if a boy tells a girl “you are really nice” does he mean “you are really nice but not beautiful”? It does not seem so. Horn solved this problem by maintaining that scalar implicatures are triggered by the so-called *Horn scales*, which are unilateral entailment scales that must undergo further constraints: The items of the scale must belong to the same semantic family and have the same degree of lexicalization. For this reason, <really nice, beautiful> is not a well-formed horn scale, whereas <nice, beautiful> might be, under some circumstances (imagine an art director grumbling to a model: “Hmmm.. you are nice but... not beautiful”). Even though it is not entirely clear which exact lexical items give rise to a Horn scale and under what circumstances, the main point, according to Horn’s proposal, is that the components forming a Horn scale are *lexically given*. In other words, the lexicon encodes these scales.

Furthermore, for any positive Horn scale there exists a corresponding

negative one. The negative scale of $\langle \text{some}, \text{all} \rangle$ is $\langle \text{not all}, \text{not some} (= \text{none}) \rangle$. Here is another interesting observation about scalar implicatures: In a negative scale, strength relations are reversed. Namely, once negated, a weak element of a positive scale becomes a strong element of the correspondent negative scale, and vice-versa. Such an observation was made, indeed, way before Horn by Aristotle in his square of opposition. The square of opposition is exemplified by the two Latin words “AffIrmo” and “nEgO”, which mean *to affirm* and *to deny* in Latin, and the so-called A, E, I, O corners, which intersect through two dimensions: universality (universal vs. particular) and polarity (positive vs. negative). Universals (A and E corners, exemplified by the quantifiers *all* and *no*) are strong positive and negative elements, which bear a *contrariety* relation among them. For instance, “All humans are intelligent” is a contrary statement to “No humans are intelligent”, being the quantifiers *all* at the A corner and *no* at the E corner. Particulars (I and O corners, exemplified by the quantifiers *some* and *not all*) are weak positive and negative elements, which bear a *sub-contrariety* relation. “Some humans are intelligent” is the sub-contrary to “Not all humans are intelligent”. First of all, the negation of any element of a corner returns its *contradictory*, that is, the element lying on the opposite corner (A vs. O, E vs. I). E.g., “All humans are intelligent” contradicts “Not all humans are intelligent”. More importantly for our purposes, Horn observes that a scalar implicature is an inference that goes from one sub-contrary to the other: “Some humans are intelligent” conversationally implicates “Not all humans are intelligent” and vice-versa. Instead, contraries are the strongest elements of the square and do not give rise to scalar implicatures.

At the beginning of this section, I anticipated that grammar comes into play within the neo-Gricean framework theory of implicatures, so how is it so? Two principles governing the distribution of scalar implicatures, logic entailment and the lexical source of the scales, are closely related to grammar. The Gricean idea of strong and weak

assertions indeed captured the phenomena of logic entailment and strength relations between alternative assertions. Nevertheless, they received a more explicit and procedural definition from Horn and the neo-Gricean scholars. Grammar and formal semantics are crucial in dealing with entailment phenomena. If we assume that the human linguistic system is equipped with a proof-theoretic device that deals with Logic Forms and logic relations – also referred to as a deductive system (cf. Reinarth, 2006) – we wind up having a powerful processor whose task is to properly interpret and compare logical formulas, compute relationships among them and logical inferences. Entailment relations, e.g., can be explicitly identified within the meta-language of first-order logic.

The hypothesis as to which Horn scales are lexically given calls into play the capacity of the interpreter to read the lexical entry of scalar items and automatically activate the relevant alternatives. Such capacity overlaps with the proper work of grammar, whereby each lexical item is assigned derivational (syntactic) and interpreting (semantic) rules. Although these considerations are not devoid of problems – see the observation shared by many scholars that the context must activate the scalar alternatives (cf. Levinson, 2000, for discussion) – they require a heavy commitment by the grammar in generating scalar implicatures. The systematic behavior of scalar implicatures (and GCIs in general) led Levinson to introduce, in the existing division of the levels/layers of communication, a new layer assigned to GCIs. According to such division, the bottom layer is the sentence-meaning level that returns the literal interpretation of a sentence. We can identify such level with the Gricean *what is said* and its output with a first sketch of the truth conditions of the proposition. The topmost layer is the speaker-meaning, which corresponds to the Gricean *what is communicated*: The whole informational content conveyed by the speaker, including implicatures, intentions, and the non-truth-conditional meaning enriching the literal interpretation of the proposition. In between, Levinson (2000) introduced

the *utterance-meaning* level: A layer in which systematic and default inferences, such as GCIs, occur. This level borders with both the bottom and topmost sides of the interpretation. It is in contact with the truth-conditional content, which serves as input, and the high-level pragmatic reasoning system, which receives its output.

Within this layer, the neo-Gricean heuristics operate: They are an explicit revision of the Gricean maxims. Instead of being conceived as rules, they are *inferential heuristics*. They are systematic, steady, and *default* pragmatic routines that we employ whenever we utter or hear a sentence. They are much different from the unique relevance-theoretic cognitive principle since they manifest a close link with grammar, which is responsible for the general inference patterns displayed by GCIs. For this reason, above all, they are default rules. Thus, they are in direct opposition with the cognitive principle of Relevance, which is variable from context to context and from utterance to utterance and which, in turn, triggers a wide range of implicatures until the hearer's relevance expectation is satisfied.

Nonetheless, neo-Gricean heuristics are yet pragmatic inferential rules, and thus, by definition, they lay outside the grammar. The neo-Gricean framework deals with a wider range of GCIs, such as the Levinsonian I-implicatures (or R-implicatures for Horn, cf. Levinson, 2000) that are default inferences towards a stereotype, or clausal implicatures, which are derived by comparing entire propositions instead of lexical units. In conclusion, the neo-Gricean framework provides a well-formed apparatus that is able to deal on the one side with the grammar and its principles and on the other side with certain non-truth-conditional features of scalar implicatures like defeasability and context dependence. All this works while maintaining the Gricean division between conventional vs. conversational implicatures and GCIs vs. PCIs. The special status of GCIs is explicitly captured by positing the existence of special cognitive heuristics that can interact systematically between grammar and communication.

In the next paragraphs, I will discuss some problems of the neo-Gricean theoretical apparatus, which led to the development of a new approach to scalar implicatures called the Grammatical View. One such problem regards the notion of *cancellability*. While, according to Grice and the neo-Griceans, conversational implicatures are always cancelable without any sense of contradiction, there are some cases in which such inferences are obligatory and cannot be defeated. One such case is when the scalar trigger is on focus. Consider the next examples:

(12) Did you read the papers I gave you yesterday?

(13) I **READ** some of them, in fact I read them all.

(14) ? I read **SOME** of them, in fact I read them all.

Imagine that the speaker of such sentences pronounced the words typed in boldface with focus intonation. In this case, sentence (13) is an acceptable answer to the question in (12), whereas sentence (14) is very awkward, if not ungrammatical. Scalar implicatures are obligatory in association with focus, although they carry the same interpretation as when they occur without focus.

1.5 Entailment patterns and embedded implicatures (Chierchia, 2004)

At the end of the previous chapter we have noticed that sometimes scalar implicatures are obligatory, that is, cannot be felicitously canceled. A possible way to defend the neo-Gricean approach is to claim that scalar implicatures are obligatory under focus due to the semantics of focus itself. That is, we could see exclusive focus as a semantic/prosodic counterpart of the Gricean implicature, which provides a grammaticalization of a pragmatic algorithm. While this is certainly a reasonable move, one may wonder how it is possible to encode in the grammar certain notions that are parasitic on the

rational nature of human communication — i.e., the Gricean Principle of Cooperation and the Maxims. One feature of scalar implicatures, which is systematic and very semantic in nature, is that of informativity, defined as the logical strength of scalar alternatives. Let us recall that according to the Gricean and neo-Gricean tenets, scalar implicatures arise from the utterance of a logically weaker proposition and, consequently, the fact that the speaker voluntarily avoided uttering a logically stronger alternative. This core property of scalar implicatures leads to the expectation that if a scalar alternative is not logically stronger than the utterance, an implicature should not arise.

This expectation is precisely the predicted behavior of scalar implicatures under downward entailing operators. The original observation that scalar implicatures should not arise in downward entailing contexts is due to Chierchia (2004). A downward entailing function, also known as downward monotonic, is a function that licenses inferences from a set to a subset. Clausal negation (15), the restriction of universal quantifier (16), the antecedent of conditionals (17), and covert negative predicates (18) are examples of downward entailing operators, that is, lexical functors that generate a downward monotonic context. Since the set of elements defined by the predicate *Marlboro* (i.e., Marlboro cigarettes) is a subset of the set of *cigarettes*, we may conclude that if an individual smokes Marlboros, he also smokes cigarettes. Interestingly, in downward entailing contexts, this logical relationship is reversed. Consider the next examples that include the downward monotonic operators mentioned above:

(15) a. John did not smoke a cigarette

b. John did not smoke a Marlboro

(16) a. Every boy who smoked a cigarette is happy

b. Every boy who smoked a Marlboro is happy