

Cambridge Textbooks in Linguistics

# Pragmatics and Grammar

Mira Ariel

CAMBRIDGE

CAMBRIDGE

[www.cambridge.org/9780521550185](http://www.cambridge.org/9780521550185)

This page intentionally left blank

## Pragmatics and Grammar

When using language, many aspects of our messages are left implicit in what we say. While grammar is responsible for what we express explicitly, pragmatics explains how we infer additional meanings. The problem is that it is not always a trivial matter to decide which of the meanings conveyed is explicit (grammatical) and which implicit (pragmatic). *Pragmatics and Grammar* lays out a methodology for students and scholars to distinguish between the two. It explains how and why grammar and pragmatics combine together in natural discourse, and how pragmatic uses become grammatical in time. This textbook introduces students to a major topic within current pragmatics research, and discusses prominent questions addressed by scholars interested in grammar/pragmatics relations. Based on natural linguistic examples, providing more persuasive data, it addresses how we should tease grammar and pragmatics apart in controversial cases.

MIRA ARIEL is a Professor in the Linguistics Department at Tel Aviv University, Israel.



CAMBRIDGE TEXTBOOKS IN LINGUISTICS

*General editors:* P. AUSTIN, J. BRESNAN, B. COMRIE, S. CRAIN,  
W. DRESSLER, C. EWEN, R. LASS, D. LIGHTFOOT, K. RICE,  
I. ROBERTS, S. ROMAINE, N.V. SMITH.

## Pragmatics and Grammar

*In this series:*

- D. B. FRY *The Physics of Speech*  
R. A. HUDSON *Sociolinguistics* Second edition  
A. J. ELLIOT *Child Language*  
P. H. MATTHEWS *Syntax*  
A. RADFORD *Transformational Syntax*  
L. BAUER *English Word-Formation*  
S. C. LEVINSON *Pragmatics*  
G. BROWN and G. YULE *Discourse Analysis*  
R. HUDDLESTON *Introduction to the Grammar of English*  
R. LASS *Phonology*  
B. COMRIE *Tense*  
W. KLEIN *Second Language Acquisition*  
A. J. WOODS, P. FLETCHER and A. HUGHES *Statistics in Language Studies*  
D. A. CRUSE *Lexical Semantics*  
A. RADFORD *Transformational Grammar*  
M. GARMAN *Psycholinguistics*  
G. G. CORBETT *Gender*  
H. J. GIEGERICH *English Phonology*  
R. CANN *Formal Semantics*  
J. LAVER *Principles of Phonetics*  
F. R. PALMER *Grammatical Roles and Relations*  
M. A. JONES *Foundations of French Syntax*  
A. RADFORD *Syntactic Theory and the Structure of English: A Minimalist Approach*  
R. D. VAN VALIN, JR, and R. J. LAPOLLA *Syntax: Structure, Meaning and Function*  
A. DURANTI *Linguistic Anthropology*  
A. CRUTTENDEN *Intonation* Second edition  
J. K. CHAMBERS and P. TRUDGILL *Dialectology* Second edition  
C. LYONS *Definiteness*  
R. KAGER *Optimality Theory*  
J. A. HOLM *An Introduction to Pidgins and Creoles*  
G. G. CORBETT *Number*  
C. J. EWEN and H. VAN DER HULST *The Phonological Structure of Words*  
F. R. PALMER *Mood and Modality* Second edition  
B. J. BLAKE *Case* Second edition  
E. GUSSMAN *Phonology: Analysis and Theory*  
M. YIP *Tone*  
W. CROFT *Typology and Universals* Second edition  
F. COULMAS *Writing Systems: An Introduction to their Linguistic Analysis*  
P. J. HOPPER and E. C. TRAUGOTT *Grammaticalization* Second edition  
L. WHITE *Second Language Acquisition and Universal Grammar*  
I. FLAG *Word-Formation in English*  
W. CROFT and A. CRUSE *Cognitive Linguistics*  
A. SIEWIERSKA *Person*  
A. RADFORD *Minimalist Syntax: Exploring the Structure of English*  
D. BÜRING *Binding Theory*  
M. BUTT *Theories of Case*  
N. HORNSTEIN, J. NÚÑES and K. GROHMANN *Understanding Minimalism*  
B. C. LUST *Child Language: Acquisition and Growth*  
G. G. CORBETT *Agreement*  
J. C. L. INGRAM *Neurolinguistics: An Introduction to Spoken Language Processing and its Disorders*  
J. CLACKSON *Indo-European Linguistics: An Introduction*  
M. ARIEL *Pragmatics and Grammar*

# Pragmatics and Grammar

---

MIRA ARIEL

*Tel Aviv University*



CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521550185](http://www.cambridge.org/9780521550185)

© Cambridge University Press 2008

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2008

ISBN-13 978-0-511-42109-9 eBook (Adobe Reader)

ISBN-13 978-0-521-55018-5 hardback

ISBN-13 978-0-521-55994-2 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.



For the love of my life



# Contents

<i>List of tables</i>	<i>page</i> xii
<i>Preface</i>	xiii
<i>Acknowledgments</i>	xv
<i>How to use this book</i>	xvi
<i>Transcription conventions</i>	xvii
<b>1 Introduction: Grammar, pragmatics, and what's between them</b>	<b>1</b>
1.1 On inferring	3
1.2 Generating implicatures	9
1.3 Distinguishing between codes and inferences	11
1.3.1 Why distinguish codes from inferences?	12
1.3.2 How can we distinguish codes from inferences?	16
1.4 Distinguishing between types of inferences	19
1.5 The challenges of a code/inference division of labor	24
<b>PART I: Drawing the grammar/pragmatics divide</b>	<b>25</b>
Introduction	25
<b>2 Distinguishing the grammatical and the extragrammatical: referential expressions</b>	<b>27</b>
2.1 Code or inference? Existential presuppositions of definite descriptions	28
2.1.1 Focusing on commitment	31
2.1.2 Focusing on context satisfaction	36
2.1.3 Encoded Givenness and inferred commitment	39
2.2 Code or inference? The referential marking scale	44
2.3 Grammatical or extragrammatical? Discourse function and discourse profile	53
2.3.1 Preferred Argument Structure as discourse profile	55
2.3.2 Genre-related uses of definite descriptions and pronouns: discourse function and discourse profile	62
<b>3 Distinguishing codes, explicated, implicated, and truth-compatible inferences</b>	<b>68</b>
3.1 <i>And</i> -associated inferences	71
3.1.1 The semantics/pragmatics division of labor in interpreting <i>and</i>	72

---

3.1.2	What pragmatic status?	78
3.1.2.1	Explicated or implicated?	78
3.1.2.2	Explicated, implicated, and truth-compatible inferences	84
3.2	The upper bound on <i>most</i>	89
3.2.1	Implicating the upper bound	91
3.2.2	Explicating the upper bound	94
3.2.3	Encoding the upper bound	100
3.3	The coded, the inferred, and the truth-compatible	107
<b>PART II: Crossing the extralinguistic/linguistic divide</b>		<b>111</b>
	Introduction	111
<b>4</b>	<b>Grammar, pragmatics, and arbitrariness</b>	<b>117</b>
4.1	Is grammar motivated or arbitrary?	118
4.2	Multiple motivations per form, or, are they?	123
4.3	Motivations competing over forms	129
4.4	Forms competing over motivations	135
4.5	Unmotivated grammaticizations?	139
<b>5</b>	<b>All paths lead to the salient discourse pattern</b>	<b>149</b>
5.1	External forces behind grammar	150
5.1.1	Grammar and our cognition	151
5.1.2	Grammar and sociocultural norms	156
5.1.3	Grammar and pragmatic inferencing	159
5.2	External forces propose, discourse disposes	168
5.2.1	Discourse versus cognition	169
5.2.2	Discourse as arbiter	177
5.3	Discourse proposes, the salient discourse pattern disposes	180
5.3.1	The salient discourse pattern	185
5.3.2	Motivated salient discourse patterns	189
5.3.3	Relatively arbitrary salient discourse patterns	193
5.3.4	A possible mechanism for crossing the grammar/pragmatics divide	205
<b>6</b>	<b>The rise (and potential fall) of reflexive pronouns</b>	<b>212</b>
6.1	A structural account of reflexive pronouns	214
6.1.1	Motivating the structural pattern	216
6.2	The grammaticization path of indicating marked coreference	219
6.2.1	Before grammaticization: indicating marked coreference with old means	219
6.2.2	The grammaticization of emphatics into reflexives	222
6.2.3	The nonunitary nature of the grammaticization of reflexives	226
6.3	Reflexives: grammar or pragmatics?	233
6.3.1	Reflexives and coreference between non-co-arguments	235
6.3.2	Reflexives and coreference between co-arguments	238
6.3.3	Markedness differences among reflexive forms	241

---

6.3.4	Grammar, and pragmatics too	245
6.4	The potential demise of reflexive pronouns	250
<b>PART III: Bringing grammar and pragmatics back together</b>		<b>257</b>
Introduction		257
<b>7</b>	<b>Grammar/pragmatics interfaces</b>	<b>261</b>
7.1	The widening gap between ‘what is said’ and a basic-level meaning	261
7.2	A minimalist ‘what is said’	268
7.3	A maximalist ‘what is said’ (truth-conditional pragmatics)	276
7.3.1	Defining ‘what is said <sub>max</sub> ’	277
7.3.2	Distinguishing explicated and implicated inferences	282
7.4	A mini–max ‘what is said’	292
7.5	‘What is said’ as less than semantic meaning	295
7.6	Syncretic approaches to ‘what is said’	297
7.7	In conclusion: the grammar/pragmatics divide and interfaces	306
<i>References</i>		309
<i>Author index</i>		337
<i>Subject index</i>		342

# Tables

2.1	The percentage of various expressions out of all anaphoric expressions (adapted from Biber <i>et al.</i> , 1999)	<i>page</i> 64
2.2	The referential distance of various referring expressions (adapted from Biber <i>et al.</i> , 1999)	64
3.1	The discourse status of the upper bound on <i>most</i> according to the implicature, explicature, and circumbounded accounts	101
7.1	Distinguishing explicated and implicated inferences	292

# Preface

*Pragmatics and Grammar* is neither about pragmatics, nor about grammar. It is about the complex relationship between grammar and pragmatics. Grammar is defined as a set of codes, and pragmatics as a set of nonlogical inferences derived on the basis of these codes. Here are some of the main questions we seek to elucidate.

First, what does pragmatic inference do that grammar cannot? In other words, what aspects of the interpretation of an utterance should be seen as pragmatic rather than semantic? How much of the interpretation should we attribute to pragmatics? Should the role of grammar in this process be minimized or maximized? Given that the concept of pragmatic inference may involve more than one subtype, we have to ask which kind of pragmatic inference best accounts for any given pragmatic interpretation. Some inferences constitute part of the information explicitly conveyed, in that they contribute, along with the coded meanings, to the truth conditions of the proposition expressed. If so, once we've ascertained that a putative interpretation should be analyzed as a pragmatic inference, we must also determine whether the inference is best viewed as a conversational implicature or as part of the proposition expressed. The classification carries cognitive and interactional implications.

We also address a puzzle which is frequently overlooked in discussions of the grammar/pragmatics interface. The fact is that our current grammar is very often our pragmatics (of the past) turned grammatical. If grammar and pragmatics are absolutely distinct from each other, as suggested by standard analyses, how can we account for grammaticization and semanticization? There must be some way for pragmatic interpretations and distributional patterns to penetrate through the grammar/pragmatics divide to become part of the grammar. We explain how this penetrability is possible while maintaining the validity of the grammar/pragmatics division of labor.

Finally, we examine various synchronic levels at which the grammar/pragmatics interface operates. In addition to the level of conveyed meaning (a representation of the linguistic meaning augmented by all pragmatic inferences), researchers have identified a basic-level meaning which is not as maximal as the conveyed meaning, yet not as minimal as the linguistic meaning. This is the representation we focus on, and once again we ask how minimal/maximal it should be. As we shall see, both minimalist and maximalist basic-level representations play a role in accounting for how language works for communication. We

end by noting that the same basic-level synchronic interpretation may also be responsible for the diachronic grammaticization and semanticization.

There is a narrative development to this textbook. The first chapter introduces the issues and the terms needed for later analyses. Part I splits linguistic acts into separate grammatical (encoded) and pragmatic (inferred) components. Part II presents evidence for an intimate association between the two. Finally, part III brings codes and inferences back together, as we consider interface levels where codes and inferences combine.



# Acknowledgments

The research that led to this book was conducted at Tel Aviv University and at UC Santa Barbara. I deeply appreciate the feedback I received on parts of this book from colleagues at both universities, Rachel Giora, Tal Siloni, and Sandra Thompson especially. It's also a pleasure to thank François Recanati and Guy Deutscher, whose comments were most helpful to me.

One person deserves special mention, Tanya Reinhart, my advisor, my colleague, my friend. Tanya read the chapter on reflexives, gave me constructive comments, generously reassuring me that it was quite all right for me to criticize her theory. How sad it is for me, indeed for all of us, to no longer have her.

Several generations of bright and enthusiastic students at Tel Aviv University read various drafts of this textbook. Their questions and challenges made classes a fun time and helped me clarify my ideas to myself, and hopefully to the readers too. I thank Ofir Zussman for his meticulous bibliographical searches, Tamar Holoshitz of Harvard University for her help with the production of the example appendix for this book (available online), and Aviah Morag for help with the index. I thank Jack Du Bois for permission to use the Longman Spoken American Corpus (LSAC), which was compiled under his direction at UC Santa Barbara, and for early access to portions of the Santa Barbara Corpus of Spoken American English (SBC).

I am honored to have had Bernard Comrie as my series editor. His care(fulness) and wide knowledge saved me from some embarrassing mistakes, and his wisdom and patience made my book into the textbook I wanted it to be. Andrew Winnard has been a most helpful and encouraging editor, and Sarah Green, Jodie Barnes, and Kay McKechnie extremely professional, as well as considerate and patient with me. Last, I owe a great debt to Jack Du Bois for all the sound advice he gave me, for listening to me talk about this book to no end, and above all, for his love and friendship.

# How to use this book

*Pragmatics and Grammar* lays out the issues regarding both the division of labor between grammar and pragmatics, and the levels in which grammar and pragmatics interface. The topic holds significance for philosophers and for linguists, especially for those with an interest in semantics, pragmatics, grammaticization and semanticization, and functional explanations for language. The book asks which phenomena should be classified as grammatical and which as pragmatic; if classified as pragmatic, what type of inference should be assumed; how pragmatic inference turns into grammatical code (grammaticization); what the minimal relevant basic meaning in discourse ('what is said') is; whether grammar is pragmatically motivated or arbitrary; and other related questions. This book offers a unified approach to the diverse questions in pragmatics and grammar by framing the issues in terms of codes versus inferences, as well as codes accompanied by inferences. Combining reanalysis of the standard problems with a new approach to naturally occurring discourse examples of language use, the book is innovative in grounding pragmatic, semantic and grammatical issues in empirical, sometimes statistical, evidence drawn from spoken corpus research.

The book is intended for upper-level undergraduates, graduate students, and researchers who have an interest in pragmatics, grammar, and the connection between them. It is suitable for a course on the grammar/pragmatics interface, on pragmatics, and on grammaticization/semanticization. While it should preferably be used in the sequence presented for a grammar/pragmatics interface course, courses with a more specialized focus can make a more selective use of the parts, each of which is designed to stand on its own. A course in pragmatics or semantics can focus on chapters 1, 2, 3, 4, and 7. A course specializing in grammaticization or semanticization can begin with chapter 1, proceed directly to part II (chapters 4, 5, and 6) and end with part III. Of course, I hope that practitioners of any of these fields will explore the full range of issues in the rest of the book. Researchers writing on the grammar/pragmatics division of labor should offer accounts which can also enable the crossing of pragmatic inferences into the grammar in a grammaticization process (discussed in part II), and similarly, researchers focusing on grammaticization/semanticization should argue for the grammaticization of some pragmatic phenomenon based on a solid grammar/pragmatics distinction (here offered in part I).

# Transcription conventions

The following conventions are used in most of the transcribed examples in this book, i.e. those taken from conversations in the Santa Barbara Corpus of Spoken American English (SBC, see Du Bois and Engelbretson, 2004, 2005; Du Bois *et al.*, 2000; Du Bois, Chafe *et al.*, 2003). Transcriptions have been slightly simplified for ease of reading.

## Santa Barbara Corpus of Spoken American English

SYMBOL	MEANING
JILL:	speaker label
New line	intonation unit*
...	pause, medium or long (untimed)
..	pause, short (less than 0.2 seconds)
@	laugh (one symbol per pulse)
@you're @kidding	laughing words
[ ]	overlapping/simultaneous speech
[ <sub>2</sub> ]	overlapping speech (2nd pair)
.	final intonation
,	continuing intonation
?	appeal/question intonation
—	truncated intonation unit (em dash)
wor—	truncated/cut-off word (en dash)
(H)	breathe (in)
(Hx)	exhale
(TSK)	click (alveolar)
(COUGH)	vocalisms (various)
###	unintelligible (one symbol per syllable)
#you're #kidding	uncertain hearing of words
((WORDS))	transcriber comment
<VOX>	voice of another

*Note:* In most cases, speaker names have been changed to preserve anonymity.

\* When the intonation unit is too long, it is carried over to the next line and indented using a hanging indent.

**Other symbols and abbreviations:**

??	unacceptable string
~	invented example
ACC	accusative
F	feminine
M	masculine
PL	plural
PRS	present tense
SG	singular
morpheme-morpheme	bound morpheme boundary
morpheme.morpheme	two glosses per single object-language element
morpheme:morpheme	a morphological division indicated only in the gloss
morpheme=morpheme	clitic boundary
1, 2, 3	first, second, third person

**Other sources commonly used:**

The Longman Spoken American Corpus (LSAC)

Lotan 1990: A Hebrew transcript of a conversation between an Israeli businessman and several income tax clerks (all males)

*Note:* Where the original non-English expressions are not crucial for the point being made, only English free translations of the examples are cited in order to facilitate reading. The original examples and their glosses can be found at [www.cambridge.org/9780521559942](http://www.cambridge.org/9780521559942).

# 1 Introduction: Grammar, pragmatics, and what's between them

Pragmatics has been notoriously hard to define. Or rather, it has proven quite impossible to reconcile between the patterning of phenomena assumed to be classical pragmatic topics (deixis and reference, speech acts, conversational and conventional implicatures, presuppositions, functional syntax) and the common set of definitions for pragmatics (most notably, context dependency, inferentiality, nontruth conditionality and others). In order to resolve the delimitation problem of the field we are forced to first abandon the expectation that all the definitional criteria converge on classifying some phenomenon as pragmatic (or as grammatical). In other words, we cannot expect that any given pragmatic phenomenon will simultaneously meet all the criterial definitions for pragmatics (and vice versa for grammatical phenomena). For example, while deixis is pragmatic in that it is context-dependent, it cannot meet the nontruth-conditionality criterion (for it contributes a truth-conditional meaning).

In addition, we must give up on what I have elsewhere called the topical approach to pragmatics, which assumes that all aspects of some phenomenon (e.g. of deixis, of presupposition, etc.) uniformly belong in pragmatics, or else, that all of them uniformly belong in grammar (see Ariel, forthcoming and chapter 2). Any specific instance of language use is neither wholly grammatical nor wholly pragmatic. To pick deixis again, it combines grammatical aspects (there is a grammatically specified difference between *I* and *this*) with pragmatic aspects (pinning down who the speaker is, what object *this* denotes). Hence, instead of struggling to find just the right set of definitions which would include all and only the canonical list of pragmatic topics (a mission impossible – see Levinson, 1983), we must choose one criterion to define pragmatics. We cannot even expect it to apply to all (aspects) of the topics on the classical pragmatic list (see Ariel, forthcoming). Perforce, a single criterion will offer a consistent division of labor between the grammatical and the pragmatic.

Now, if we can only choose one criterion for drawing a consistent and coherent grammar/pragmatics division of labor, which criterion should we opt for? This book, along with researchers such as Sperber and Wilson (1986/1995), adopts the code/inference distinction as the basis for the grammar/pragmatics division of labor. Why pick the code/inference distinction (as opposed to any other criterion, e.g. truth conditionality)? The simple answer is that we cannot afford not to adopt this distinction, given the nature of grammar. Whatever else grammar may be, there can be no controversy about it consisting of a set of codes. The essence of grammar is a set of conventional associations correlating specific forms with their

obligatory or optional, rule-governed positioning, meaning, and distributional patterns. Whether or not it is in addition devoted, for example, to truth-conditional interpretations remains to be seen.

At the same time, grammar obviously falls quite short of meeting our communicative needs. There is a consensus today about the underdeterminacy of grammar, i.e. the fact that our coded messages never exhaust the meaning we intend to convey (see especially Carston, 2002a, and see also Levinson, 2002a: 8). This is where pragmatics comes in, enriching our encoded messages with pragmatic, i.e. plausible inferred interpretations (as opposed to formal, logical deductions). Grammar and pragmatics always go together. You can't have one without the other for effective communication.

This book is about the complex relationship between grammar and pragmatics, that is, between codes and inferences involved in human communication. The relationship is not unidimensional. It has a few facets, and each one of them needs to be examined. Naturally enough, researchers interested in the grammar/pragmatics division of labor have focused on the complementarity between the two competencies. Such research seeks to establish the precise borderline between codes and inferences. Indeed, this is one important facet of the relationship between grammar and pragmatics, and part I is devoted to this topic. We try to resolve a number of intriguing questions, all of which have the same format: given a certain correlation between some linguistic form and some interpretation or use conditions, should it be grammar or should it be pragmatics that accounts for it? In other words, is the correlation encoded or is it derived by plausible inferences? Now, up till recently, all inferences were automatically seen as conversational implicatures. Relevance theoreticians have proposed that not all pragmatic inferences are implicatures, however. If so, every interpretation we view as a pragmatic inference needs to be further classified as to which type of inference it is.

But there is much more to explore about the relationship between grammar and pragmatics. Codes and inferences must make contact in diachronic change. Research by functional and historical linguists (e.g. Bybee *et al.*, 1994; Traugott and Dasher, 2002) has convincingly demonstrated that, historically, pragmatic inferences routinely become grammatical conventions (via the processes of grammaticization and semanticization). In fact, most, if not all, of our synchronic grammar is the product of our diachronic extralinguistic regularities, which have turned grammatical. If so, any analysis of the grammar/pragmatics divide must also meet the challenge of the penetrability of this divide: grammar is not only distinct from pragmatics, it is also its product. Part II addresses this facet of the relationship between grammar and pragmatics.

The final chapter of this book (part III) brings together these two very different approaches and research traditions. We discuss grammar/pragmatics interfaces, i.e. representational levels where the two combine. Synchronically, we have at least two such levels, conveyed meanings and basic-level meanings. We hardly touch on the first type of representation, despite the fact that it consists of the linguistic meaning plus all the conversational implicatures generated by the speaker (but see

section 1.2 below). The reason is that here, although semantics and pragmatics combine, they remain interactionally independent of each other. There is also no controversy regarding this interface level. We focus on the basic, minimal meaning level which the speaker is necessarily seen as committed to. This interactionally significant (synchronic) meaning is an integrative level, where codes and inferences are intimately woven together to form one proposition. Here codes and inferences are in fact so well integrated that it's not clear that they are separable (for cognitive and discoursal purposes). We consider proposals as to how to define this meaning representation, namely, minimally (i.e. with as few inferences as possible), or maximally (incorporating a substantial inferential contribution). The conclusion we reach is that this hybrid-level meaning is minimalist (*à la* Grice) on some occasions and maximalist (*à la* Sperber and Wilson) on other occasions.

Now, although this basic-level interpretation has been proposed as a synchronic, real-time representation, we end the book by entertaining the possibility that it also serves as the input for semanticization and grammaticization. In other words, it is possible that the synchronic basic-level grammar/pragmatics hybrid representation is not only the significant meaning level in terms of the ongoing interaction, it is also the diachronic arena where pragmatics may turn grammar. After all, linguistic change must occur in real-time discourse. If this is true, the most significant grammar/pragmatics interface level both synchronically and diachronically is the basic interactional interpretation, where codes and inferences co-mingle, regardless of the fact that they are arrived at via quite distinct cognitive processes.

Let's begin our introduction now, which will prepare the ground for the whole book. We set out from the now well-accepted assumption that we always communicate by combining codes (grammar) with inferences (pragmatics). We introduce the basic facts about inferencing in section 1.1. We briefly outline the Gricean mechanism for generating inferences (conversational implicatures) in section 1.2, and we define and distinguish between codes and inferences in section 1.3. Section 1.4 introduces pragmatic inferences other than conversational implicatures. We end with section 1.5, where we introduce the two challenges facing research into the grammar/pragmatics relationship: drawing the code/inference distinction on the one hand, and accounting for the process whereby inferences cross over and become codes on the other hand.

## 1.1 On inferring

We get more for our words when they are embedded in natural discourse context. Here are two examples where it is quite clear that we need to read between the lines:

- (1) a. LEWINSKY<sub>1</sub>: ... See my mom's big fear is that he's ((President Clinton – MA))  
going to send somebody out to kill me.  
((PART OMITTED))

TRIPP: Oh, my God. Don't even say such an asinine thing. **He's not that stupid. He's an arrogant . . . but he's not that stupid.**

LEWINSKY<sub>2</sub>: Well, you know, **accidents happen.**

(Nov. 20, 1997: *New York Times*, Oct. 3, 1998)

b. REBECCA<sub>1</sub>: So you can testify to two of [em].

RICKIE<sub>1</sub>: [Yeah].

REBECCA<sub>2</sub>: That's why I had you come up,  
**becau[se],**

RICKIE<sub>2</sub>: [Yeah],

REBECCA<sub>3</sub>: . . . um,  
. . . that's great. (SBC: 008)

While the above exchanges are natural enough, and easy to interpret (for their participants, at least), there is something strange about each of them. (1a) seems to involve irrelevant responses: why is Clinton's intelligence (Tripp's response) relevant to the fear of Lewinsky's mother that he will have her killed (a topic raised by Lewinsky<sub>1</sub>)? And what is the relevance of the occurrence of accidents (Lewinsky<sub>2</sub>'s response) to Clinton's intelligence (discussed by Tripp)? It seems that while there is a connection between Lewinsky's and Tripp's contributions above, it is not explicitly stated. It is inferred. Tripp intends to convey that since Clinton is intelligent he won't have Lewinsky killed, because killing her would be a stupid act (probably since he would be caught and get into even deeper trouble). Lewinsky then retorts that he may indeed have her killed, except it would be made to look like an accident (so he may not be caught). We similarly need to enrich Rebecca<sub>2</sub>'s contribution in (1b), since her sentence is cut off in the middle. What is the reason that Rebecca (a prosecuting attorney) had Rickie (a sexual abuse victim preparing to testify against her sexual abuser) come up to her office? Given the specific context, we can assume that it must be the fact that Rickie can testify to two cases of sexual abuse by the defendant, as opposed to other witnesses who "only" experienced one sexual abuse from him. In the light of what has already been said in the exchange, this reason is quite obvious, and hence need not be explicitly mentioned. The addressees can easily infer it. The exchanges in (1) are quite typical of natural interaction. Communication begins with the coded message, but it never ends there. Inferences are an inherent part of it.

The most basic goal of pragmatic theories is to provide an account for how we go about interpreting such everyday exchanges. This is the goal informing Grice (1989), as well as Horn (1984 and onwards) and Sperber and Wilson (1986/1995 and onwards), all following Grice. We here briefly present only Grice's pragmatic theory, because our purpose in this chapter, indeed in this book, is served by demonstrating that there is at least one pragmatic theory which can account for additional interpretations we get people to infer when we speak. In fact, all pragmatic theories are in this sense Gricean. They all assume that every act of communication is actually **inferential**, because the addressee is required to **infer** the speaker's intention from whatever evidence is available to him (the linguistic



code constituting only one important source of information).<sup>1</sup> At the same time, this chapter also serves as an introduction for later chapters, where disagreements between theorists will be discussed. These revolve around the role of particularized and generalized conversational implicatures, and the Relevance-theoretic concept of explicature, all here introduced.

Natural discourse is a **cooperative** activity. It's not a random collection of independent utterances (see Mann and Thompson, 1986). What makes discourse coherent? What expectations do we have from interlocutors' utterances? The idea is that whatever those expectations may be, they cannot be fulfilled by our coded messages exclusively. It is only when we (also) consider inferred interpretations that we can ultimately explain how discourse works. In order to produce and interpret appropriate pieces of discourse, speakers must rely on the many inferences which accompany those codes. This is why codes and inferences are so intimately connected with each other. And this is why discourse coherence and inferencing are co-dependent on each other as well. Let's examine briefly Grice's (1975, 1989) proposals on how to account for both the nature of discourse and the mechanism responsible for the derivation of pragmatic inferences.

No doubt, one of the most important features of human discourse is that we assume that speakers' utterances are somehow **relevant** to us. This is why we are led to interpret the following two statements as related, in fact, as constituting a contrast:

- (2)           The father was appointed chief of staff, the daughter refused to enlist (in the army).           (Originally Hebrew, *Hair* headline, April 25, 2002)

Each statement by itself is quite irrelevant to the addressees. The readers of the newspaper need not be informed that someone's father has been appointed chief of staff. They know the man by name, and they know he was appointed. The fact that some daughter refused to serve in the army is equally irrelevant. However, with the two pieces of information taken together, the headline becomes highly relevant: it turns out that the Israeli chief of staff himself (who had spoken about the importance of serving in the army) has a daughter who refused to serve. To see that we indeed expect discourse to be relevant, note what happens when Darryl can't see Pamela's utterance as relevant:

- (3)           PAMELA: I guess it's j– looking at my mother,  
                   too,  
                   I n— —  
                   (Hx)  
               DARRYL: . . . **What does that have to do with** why you're reading a book  
                                   on death?           (SBC: 005)

<sup>1</sup> Indeed Tomasello (1999) argues forcefully that children's ability to view others as intentional agents (an ability which emerges at nine to twelve months) is a prerequisite for language, and something that animals lack.

Darryl's question shows that the default assumption is that interlocutors produce relevant utterances. Since he fails to find relevance, he asks about it. So, an important requirement on speakers is that they be relevant. This is what Grice's maxim of Relation ("Be relevant!") is about.

Next, let's examine Grice's maxim of Quantity. We expect utterances to be **informative**. Here's a case where Joanne finds one of Lenore's utterances (the one marked bold) not informative enough. She then asks for an elaboration:

- (4) JOANNE: ... (H) He's got iron,  
with his mul[tiples].  
LENORE: [Well,  
I] have iron too,  
but th— some of it isn't absorbable.  
this is very absorb[able iron].  
JOANNE: **What do you mean absorbable.**  
LENORE: (H) it's good for your anemia. (SBC: 015)

Of course, Joanne knows what *absorbable* means (semantically). But she can't get enough relevant information from Lenore, presumably because she is missing some background information that Lenore is taking for granted, something to the effect that absorbability is an important issue for the effectiveness of iron pills. In this sense, Lenore's contribution was not informative enough, and hence is not quite appropriate.

A third maxim proposed by Grice is the maxim of Quality. We naturally assume that our interlocutors tell us the **truth**. This is why when Pete in (5) realizes he's about to say something he is not sure is true (that salad spinners did not exist when they were growing up), he stops short, and asks his friends whether it's true or not:

- (5) PETE: I don't think they ever —  
.. **did they actually exist back then?** (SBC: 003)

Grice's maxim of Quality instructs speakers not only to tell only the truth, but also to avoid saying that which they lack evidence for. Pete is abiding by this maxim.

Finally, according to Grice's maxim of Manner, utterances should be constructed in an optimal style or **manner**. They should be brief and clear. Here's a case where the speaker was not clear enough (in the turn marked bold), and caused a problem for his addressee. She has to ask him about his intended message therefore:

- (6) W: M!  
M: I'm coming!  
W: Watch out along the way.  
M: Why?  
W: **Sandy!**  
M: ((Looks around, doesn't see Sandy, the dog))  
What?  
She vomited? (Aug. 5, 2006)

While on other occasions it may be enough to use an NP utterance and trust the addressee to complete the missing propositional information (see *Iraq* standing for 'the American war and occupation of Iraq' in (13) below), this was not a successful act in (6). This is then a violation of Grice's maxim of Manner.

The problems we detected in the discourses in (3)–(6) provide evidence for how proper discourse usually proceeds. The four Gricean maxims mentioned above characterize what Grice (1975, 1989) suggested are cooperative communicative principles. According to Grice, natural discourse reflects the application of the Cooperative Principle, which instructs speakers to "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" (Grice, 1975: 45). This cooperative (some prefer rational – see Kasher, 1976) behavior translates into abiding by the four maxims we've mentioned: **Quantity** (informativeness), **Relation** (relevance), **Quality** (truthfulness), and **Manner**. Speakers are expected to provide just the right amount of information (neither too much nor too little – Quantity), the information should be relevant (Relation) and true (Quality), and it should be optimally phrased, namely, the utterance should be as brief and clear as possible (Manner). As we saw in the above examples, whenever any one of these maxims was violated (or about to be violated), there was some breakdown in the communication, and speakers hastened to query about it, so as to somehow remedy the problem. I should add, however, that examples (3)–(6) were not at all easy to find. Most of the time, discourse does proceed rather smoothly, according to the Gricean maxims above.<sup>2</sup>

Now, what's all this got to do with the question of codes and inferences? A lot, according to pragmatists. The argument is that we simply cannot comply with the cooperative principle, nor with the four maxims, by subjecting just our coded, explicit messages to these principles. As already exemplified in (1), we must take into account inferences in order to see how speakers go about cooperating with each other. If we restrict ourselves to examining only codes, we will find quite a few violations of the Gricean proposal, ones which are not accidental lapses (performance errors), like the exceptional examples in (3)–(6). The violations we are about to examine below (section 1.2) do not cause any communication problems. This is so because they are intended, and they give rise to inferences which then show that speakers were cooperative after all. Inferences serve discourse when codes fail. Still, why is it that we don't communicate by codes alone? Why don't we make sure that our codes fully abide by the four Gricean maxims? Wouldn't it be clearer and more efficient to communicate this way? Why divide up our communicative task between two modes? For example, why didn't the newspaper choose as its headline the Hebrew counterpart of 'The father was appointed chief of staff **but his** daughter refused to enlist (in the army)' (instead of (2))?

<sup>2</sup> See Grice (1975, 1989) for a detailed outline of the theory.

First, explicating everything we wish to convey to the addressee would considerably slow us down. Following Levelt (1993), Levinson (2000a) has convincingly argued that our production mechanisms are far too slow to allow for an efficient, fully explicit mode of communication. Many pieces of information which constitute an integral part of our messages are better left to inferencing, because human beings are very apt at drawing inferences. Actually pronouncing out loud these assumptions would take a long time and effort on the speaker's part, and at the same time would waste the addressee's time in superfluous decoding. Many inferences are faster for the addressee to compute than for the speaker to form a verbalizing plan and to articulate with words. Here's a relevant example (and see again (1)):

- (7) HAROLD: that . . . really hot tap danc[er]<sub>i</sub>.  
 JAMIE: [Oh] that kid<sub>i</sub>.  
 MILES: . . . **He<sub>i</sub> was actually here two weeks ago,**  
**and I missed him<sub>i</sub>.** (SBC: 002)

In fact, what Miles is conveying to his interlocutors is that 'He, i.e. **the tap dancer**, was actually here, i.e. **in this town**, two weeks ago **for a public performance**, and I missed **that performance with** him.' But Miles can skip stating explicitly that the dancer gave a performance (see how much longer the explicit version is), because we can very easily infer it. The same applies to the deictic (*here*) and the anaphoric expressions (*He*, *him*, and see again (1b)). Leaving part of the message to inference can be an efficient step, then.

In addition to effort saving as in (7), speakers have a variety of reasons for preferring an implicit over an explicit mode for some interpretations. There is an interactional difference between explicit and implicit messages. Consider the following:

- (8) a. The Americans know Netanyahu, who is actually Benjamin Nitai.  
 (Originally Hebrew, Reshet Bet radio, Jan. 11, 2006)  
 b. She had an uncle who predicted that she will be a musician and a genius.  
 And she always said, well, at least I became a musician.  
 (Originally Hebrew, Reshet Bet radio, Mar. 17, 2006)

The implicit message in (8a) is that Israeli Knesset Member (and former prime minister) Netanyahu is not quite a patriotic Israeli, possibly, that he cannot be trusted as a leader. In order to arrive at this conclusion, we need to rely on a host of background assumptions: as is well known (to Israelis), years before he became a politician, Netanyahu had left Israel and lived in the US for quite a number of years, where he changed his name from the Israeli-sounding *Binyamin Netanyahu* to the American-sounding *Benjamin Nitai*. On the assumption that the name one adopts testifies to the national identity one aspires to, there is possibly an even stronger message here, that Netanyahu is not actually loyal to Israel but, rather, to the US. No wonder Knesset Member Ronnie Bar-On (the speaker in (8a)) prefers an implicit over an explicit mode. As we point out below, one can always deny

what one did not say explicitly.<sup>3</sup> Next, consider (8b). Being a musician certainly does not preclude being a genius. Yet, it's clear that Drora Chavkin (the *she* cited in the example) implied that she was not a genius. But she didn't quite say it. There is a difference between explicitly stating some information and conveying it by triggering an inference (a self-humorous effect in this case). So, the answer to our question of why rely on inferences rather than on decoding is that (a) inferences save on time and effort (sometimes), and (b) they are indirect.

## 1.2 Generating implicatures

We need a “science of the unsaid,” as Levinson (2000b) calls inferential pragmatics theories. Grice's (1975, 1989) idea was that we use the same four maxims that inform our cooperative behavior in general as guiding principles in inference drawing as well. Interlocutors' working assumption is that cooperative speakers do abide by the maxims. If so, should they seem to violate one of the maxims, and blatantly so, rather than take it as a breakdown in the communication (as is done in the atypical (3)–(6)), addressees assume that there was a special speaker communicative intention behind the maxim flouting. That communicative intention is a pragmatic inference, based on what was said explicitly and contextual assumptions the speaker intends the addressee to consider in the computation of the inference. Those inferences which fall under the communicative intention of the speaker are what Grice termed **conversational implicatures** ((8a) triggers a particularized conversational implicature, (8b) a generalized conversational implicature – see below). Let's consider a few examples to see how the mechanism of speaker-generated implicatures works.

Consider (8) again. We can now explain why it is that speakers generated the specific implicatures we noted above. In both cases we have violations of the maxim of Quantity. The speaker in (8a) provides too much information. While vaguely relevant to the topic, Netanyahu's American name constitutes superfluous information at the current stage of the discourse. Once the addressee figures that the speaker is being cooperative, and that he has a specific intention in being too informative, the way is paved for deriving the implicature that the added information is relevant in that if Netanyahu has an American name, he may have an American identity and loyalty. Similarly, in (8b), where Chavkin provides too little information (confirming the prediction about her becoming a musician, but remaining silent on the question of her being a genius). Assuming that this is no accident (or performance error), the speaker is seen as avoiding the second question, thereby implicating nonconfirmation.

<sup>3</sup> Note that unlike the inferred conclusion specified above, that Netanyahu cannot be trusted, the background assumptions leading to this inference are not spelled out not because the speaker wants to be indirect about them, but because he doesn't want to articulate that which can be more easily inferred (the efficiency motivation alluded to before).

Ironical interpretations are more often than not implicatures generated because Quality (truthfulness) has been flouted:

- (9) S: I'm going upstairs now to the regional committee's office . . . to see what's going on. **Only nine years** they've had it, no less ((it = S's application for a building permit)). (Lotan 1990: 4)

Obviously, for a builder, as S is, to wait nine years to get a building permit is not a short time. He doesn't really mean '**only** nine years,' which conventionally implicates that one would have expected it to take longer. Since Quality was breached, but the speaker is taken as cooperative, he is taken to intend to implicate something, in this case, that he means quite the opposite of a short, negligible waiting time: nine years is much too long for the committee to delay its decision about a building permit.

Next, the following is a case where relevance is flouted, the speaker, a job candidate anxious to get the job, generating an implicature, which renders her reply relevant after all (and see again (1a)):

- (10) Boss: You have small children. How will you manage long hours?  
HD: I have a mother. (Originally Hebrew, June 14, 1996)

Surely, the fact that HD has a mother does not seem a relevant answer to the boss's question about how she will be able to stay at work after regular hours. But since we assume that HD is cooperative, we see the flouting as an indication that she is generating an implicature, in this case, that 'her mother will take care of her children when she needs to stay late at work.'<sup>4</sup>

Manner violations too can serve as a basis for implicature generation, as can be seen in (11a):

- (11) a. Let us look at the **racial, or rather, racist** themes in the argument for population control.  
(Pohlman, *Population: A clash of prophets*; Du Bois, 1974; ex. 8)  
b. ~Let us look at the **racist** themes in the argument for population control.

(11a) is a case of a repair, where the speaker corrects himself. Supposedly, he said *racial* when he really meant *racist*. When we compare the original (a) version with the contrived (b) version, it's quite clear that (a) is not as brief as (b). Now, in spontaneous conversation, we might see this as a performance error, and not attribute any significance (implicatures) to the repair. But since the example is taken from a carefully edited written source, argues Du Bois (1974), we cannot ignore the fact that the writer here chose the longer over the shorter version. We then attribute to the writer additional implicated interpretations, perhaps that he is not comfortable in asserting an unmitigated version of his strong term (*racist*).

<sup>4</sup> We are here glossing over the need to access just the right contextual assumptions for generating the intended implicature. For example, in the case of (10) one could theoretically render HD's utterance relevant by assuming that her mother will come to replace her at work when she has to work late. Of course, this is not what she implicates.

Or, consider the following example. Readers are advised that the Hebrew counterpart of 'lower one's profile' has a very salient collocational meaning: 'get a lower military health rating from a doctor (so as to avoid serving in a fighting unit in the army)':

- (12)      bou            le=horid      profil!  
              come:2<sub>PL</sub>      to=lower      profile!  
              'Come lower your profile!'

(Hebrew, ad for Gym Center in Tel Aviv, summer 2005)

The salient interpretation (a lower health rating) does not make any sense in a gym ad, however. If anything, one would expect to get a higher health profile following workouts at the gym. In order to see the ad as relevant, we have to re-process the utterance, and come up with a different interpretation, that the profile we need to reduce (flatten, rather) is our body profile, our fat etc., not our army health rating. So the writers here use an ambiguous term, whose automatically accessible interpretation ('Lower your military health rating') is precisely the one not intended by them. Note, moreover, that the intended reading ('Come get in shape') is rather difficult to arrive at actually, since it's not simply the literal, compositional meaning of the expression used. (12) manifests an innovative, unconventional use. Just as in English, in Hebrew too we routinely use 'profile' to denote specifically one's facial (rather than body) profile. Why not say then something like 'Come get in shape/reduce your weight' instead, either one of which is as short, but much clearer and easier to process? The addressors here clearly flout the Manner maxim. It's not in vain, of course. The surprising interpretative change we suddenly realize we need to perform is an added aesthetic effect here (as it is in some jokes). It's hard to explicate this implicature in conceptual terms. Laurence Horn (p.c.) proposes something like, "We're so clever – notice how we got you to misprocess the utterance before forcing you into the correct analysis." Be that as it may, it's clear that (12) has an added value over a straightforward 'Come get in shape,' and this added pragmatic effect justifies the flouting of the Manner maxim. In sum, conversational implicatures are often generated when the encoded meaning seems to violate some Gricean maxim. They then render the speaker cooperative after all. We cannot overestimate the important role of inferences in creating coherent discourses.

### 1.3 Distinguishing between codes and inferences

Now, so far we've taken codes and inferences simply in their pretheoretical sense. In fact, this is not too different from the way we will use the terms in this book. A linguistic phenomenon is governed by a code if the form–function correlation involved is **conventional** rather than **derived** by a plausible inference/implicature, based on the relevant context. In the latter case it is pragmatic. But actually, we should first understand why we need to draw this distinction (section 1.3.1). We then discuss how to apply the code/inference distinction (section 1.3.2).

### 1.3.1 Why distinguish codes from inferences?

If both codes and inferences are subject to the same cooperative principle, as we've seen above, why do linguists insist on distinguishing between the two? The claim is that grammatical and pragmatic competencies involve different cognitive processes, which in turn entail different discourse roles. For example, both decoding and inferring are required in interpreting ambiguous expressions. But the processes are quite different. The addressee first needs to access all the linguistic meanings of the ambiguous word, e.g. (i) 'round object ...' and (ii) 'dancing party' for *ball*, a decoding process, and only then can he select the one appropriate sense intended by the speaker in the specific context, relying on pragmatic inferencing. The same is true for cases where we need to construct implicit and/or innovative interpretations based on the coded meaning and contextual assumptions (e.g. in cases of novel metaphors and ironies). There is ample psycholinguistic evidence for the distinction between coded and inferred meanings, between the working of lexical accessing and the working of inferences based on contextual assumptions (see Giora, 1997; Swinney, 1979). But it is not just psycholinguistically that we detect this distinction. Natural language interactions too testify to the very same distinction between the coded and the inferred. We have noted above that communication efficiency and a special discourse status motivate our relying on implicatures rather than only on codes in order to convey our messages. What is it that characterizes inferences?<sup>5</sup>

The most important feature of inferences is that they are implicit. Being implicit, they function differently from explicitly stated messages (see (8) again). Since they are not explicit, they can always be denied or canceled, that is, the speaker is not necessarily committed to their content. The point is that this noncommittal need not even be marked, and is not perceived as self-contradiction on the speaker's part. Consider the following. In each case, some cancellation is involved. Not surprisingly, in each it's an implicit aspect of the interpretation that is canceled:

- (13) a. On Tuesday night, host Jim Lehrer asked Gen. Myers: "Do you consider Iraq a success from your point of view?" The general replied: "I do **now**, I do. I mean, **I don't know why I said now. I do, absolutely; I think it's a success**"  
([www.truthout.org/docs2005/071305A.shtml](http://www.truthout.org/docs2005/071305A.shtml))
- b. W: So you baked this with the chicken?  
M: Well,  
**Separately.**  
W: I meant in the same oven. (Aug. 9, 2006)

Myers' initial response (*I do now* in (13a)) implicates that he didn't think so in the past (*now* constitutes too much information), and by implication that he may not consider the situation in Iraq such a great success. He hastens to cancel this

<sup>5</sup> Throughout this book, unless specified otherwise, the term *inferences* means specifically 'pragmatic inferences.'



implicature. And consider (13b), where the relevant inference concerns what it means to *bake this* (potato dumplings) *with the chicken*. M took it to mean 'bake the two **in the same pan**.' This is why she intends W to infer from her *separately* that she baked them in separate pans. W explains that he meant that M baked them at the same time, in the same oven, but not necessarily in the same pan. The inference drawn by M is rather easily canceled here with no feeling that W's first utterance, for which he asks for confirmation, is false.

Other characteristics distinguishing conversational implicatures from encoded meaning according to Grice are separateness from the basic, explicit meaning ('what is said'), nondetachability, calculability, and indeterminacy. Consider (10) again. As HD told her addressees when reporting on the exchange with the boss she cites, the truth was that she indeed had a mother, but her mother in fact never helped her with the children. Did HD say a false thing to the boss, then? She certainly generated an implicature which is false ('my mother will help me with the children'). But her explicit message ('I have a mother') was perfectly true. What is the status of false implicatures? Grice proposed that they have no effect on the truth conditions of the proposition made. In other words, the fact that the implicature is false does not render HD's proposition false.<sup>6</sup> This is so because 'what is said' (the explicit message) and inferences based on 'what is said' (the implicature) are independent of each other.

Next, an interpretation is nondetachable if it's not dependent on some specific form. Under this definition, conversational implicatures are interpretations which would have been generated had the speaker used a different phrasing with a similar enough semantic content. For example, what if HD were to say 'My mother lives right next door to my house,' instead of saying 'I have a mother' (in (10))? Most probably, the same implicature ('my mother will take care of my children when I have to stay late at work') would have been generated (in the same context). This is so because both of these responses are equally irrelevant to the question she is asked, and they lead to the same conclusion when we attempt to see the speaker as cooperative. So, inferences are nondetachable from certain contents, and they are independent of specific forms (except for Manner floutations, of course).

Next, unlike codes, inferences are at least to some extent indeterminate. For example, what is the implicature generated by HD in (10) exactly? Is it, 'my mother will take care of my children when I'm at work late'? or is it, 'my mother helps me out whenever I need help'? It's hard to judge, and it doesn't much matter, actually. Either formulation justifies the speaker's flouting of Relation. Two other features, not originally noted by Grice, but very much in the spirit of his proposal, are that conversational implicatures are universal and reinforceable. To see the first point, note that (10) is originally a conversation which took place in Israel in Hebrew. Yet, English speakers have no problem deriving the relevant implicature from the English translation. This is because no specific linguistic form is

<sup>6</sup> See section 7.6 for a somewhat different view and findings.

involved in the triggering of the inference. Provided we share a set of contextual assumptions (e.g. young children cannot be left alone, mothers help their daughters), since the maxims are universal and so are our inferential abilities, speakers of any language should be able to generate the same implicatures.<sup>7</sup>

Finally, since inferences are implicit, they may be reinforced explicitly without causing the speaker to sound redundant. Consider:

- (14) He passed away in the arms of a woman, not his ((wife)).  
(Originally Hebrew, *Hair*, Mar. 13, 2003)

First, consider only the first part of the utterance, preceding the comma. Note that even if the writer had not added ‘not his ((wife))’ we would have understood (by implicature) that ‘it’s not the man’s wife,’ because if she were his wife the writer would have said so (it’s relevant to indicate the relation between referents). But now, if it’s the case that the writer here generates a conversational implicature to the effect that ‘the woman is not the man’s wife’ anyway, why does he go on to also assert this piece of information? Isn’t it redundant, given that it is implicated anyway? This is another difference between codes and inferences. Reinforceability, as Sadock (1978) called the ability of implicatures to be asserted explicitly without bringing about a redundancy effect, characterizes inferences, but not codes.<sup>8</sup>

Next, let’s examine a few cases where the interlocutors themselves “argue for” the importance of distinguishing between codes and inferences. As argued in Ariel (2002b), only the (coded) semantic meaning can be imposed by (uncooperative) interlocutors in an inappropriate context. Note the following:

- (15) a. A municipal regulation determines that a piece of property which remains **empty** is exempt from city tax for six months. A, a lawyer who has moved into a new office, was astonished when city hall inspectors refused to declare his old office as **empty** because of a few chairs left behind. “What is **empty**,” he wondered, “when the floor tiles have been pulled out?” . . . The director of the city income department explained . . . that **formally, empty is empty**, and it’s possible to say that even if there is a rag in the office, it is considered **full**.

(Originally Hebrew, reported in the magazine *Tel Aviv*, May 14, 1993)

- b. BEN: What are you doing, criticizing me?

GUS: No, I was just . . .

BEN: You’ll get a swipe round your ear hole if you don’t watch your step.

GUS: **Now look here**, Ben . . .

BEN: **I’m not looking anywhere!**

(Pinter, *The Dumb Waiter*, pp. 15–16, quoted from Yus Ramos, 1998: 87)

<sup>7</sup> This may appear not to be the case when there are cultural differences. For example, in a culture where young mothers are not allowed to walk on their own at night, the implicature in (10) might be ‘my mother will come to escort me home when I work late.’ But the universality claim only pertains to the inferential mechanism, not to the background assumptions, of course. The idea is that the implicature is language-independent if it’s based on the same content (of linguistic utterances and contextual assumptions).

<sup>8</sup> What is crucial to us at this point is that one can reinforce implicatures by asserting them explicitly. But of course, this reinforcing is also used to generate further Manner implicatures. The repeated message ‘not his ((wife))’ in (14) underscores the special (embarrassing) circumstances of the man’s death.

- c. A: When you come home,  
       I'll have the food ready **on the table**.  
 B: I'd rather have plates.  
 A: What?  
 B: I'd rather have plates.  
 A: (LAUGH) (Jul. 27, 2005)

The city hall clerks (15a), and Ben (15b) are not only uncooperative interlocutors, they are downright annoying. They each refuse to interpret some linguistic string (*empty* and *look here* respectively) as it was intended by the speaker. Note, however, that their uncooperative interpretations are fully backed up by the semantic meanings of these expressions. In (15a) the city hall clerks refuse to adapt (loosen up) the concept of 'empty' to the specific context (where *empty* should be interpreted as 'relevantly empty, a few chairs or a rag not counting as relevant'). They stick to an unenriched (strict) coded meaning instead. In (15b) Ben refuses to select the intended linguistic meaning (this is a case of ambiguity). In (15c) B is being playful, but what he does is in principle similar to the noncooperative clerks (in a) and Ben (in (15b)). When A says she'll have 'the food ready on the table,' she intends that the addressee take into consideration the culturally available assumption that one does not serve food directly on the table, and that of course, the food will be served on plates. But since this is not an explicitly communicated assumption, B (a wise-guy speaker according to Ariel, 2002b) can ignore it. He couldn't equally ignore an explicit aspect of A's message, saying, e.g. *??I'd rather have food*. The point is that while these interlocutors are quite uncooperative, since their choice is consistent with the (or a) semantic meaning of the expression, their utterances cannot be dismissed. So, codes can survive incompatible contexts. What about inferences?

Inferred interpretations cannot be imposed in an incompatible context. Consider the following:

- (16) The goblet (= trophy) is red and it is ours.  
       (Originally Hebrew bumper sticker, spotted Feb. 26, 2007)

A red goblet is a goblet which is red in color. But in this context, red is the color associated with a certain soccer team, rather than the goblet's actual color. The bumper sticker on the car with many positive references to the specific soccer team is interpreted as 'the goblet is inherently associated with the specific team.' Outside this exceptional context, say, in a context where the literal meaning of *red* is appropriate, this special meaning cannot arise, and not even a wise-guy interlocutor can get away with it. We cannot impose a potentially inferred interpretation ('the goblet belongs with the specific team') in a context which is incompatible with it. We can only get away with a contextually inappropriate interpretation if it's supported by some semantic meaning. So wise-guy (im)possibilities provide us with an interactional piece of evidence for the differential role of codes and inferences, as well as with a method of teasing the two apart. The code/inference distinction is not only psycholinguistically real, it is also discourse-real.

### 1.3.2 How can we distinguish codes from inferences?

So far, it looks as though the concepts of coded and inferred meanings are clearly distinguishable. But actual interpretations are not always easily classified as encoded or as inferred. We now come to aspects of the relationship between grammar and pragmatics which lie at the heart of this book. Note that once we assume, as indeed we must, that speakers' use of language makes crucial use of our pragmatic/inferential abilities in deriving additional, implicit meanings, we could perhaps make a case for pragmatics accounting for what is potentially accounted for grammatically. After all, interpretations do not come to us with grammar/pragmatics labels. Hence, in principle, many interpretations could be analyzed either as pragmatic or as semantic. How can we decide?

According to Grice's Modified Occam's Razor Principle, "Senses are not to be multiplied beyond necessity" (Grice, 1989: 47). In other words, we should opt for a maximal application of pragmatic inferences at the expense of grammar (semantics), made minimal. Thus, if we can explain some interpretation (or use conditions) as either a code or as an inference, we are supposed to prefer the latter. The advantage linguists see in such analyses is that they are economical. Since we need to assume pragmatic inferences in any case (accounting for inferences supplementing the codes as in section 1.2 above), if we can use the same mechanism for phenomena which would otherwise require the positing of extra grammatical facts, our analysis is genuinely more economical. The assumption usually is that it is to the speakers' advantage that grammar be minimized, since a big grammar is a burden on our limited memory capacity. So let's consider where we can and where we cannot apply Modified Occam's Razor in order to distinguish between encoded and inferred interpretations.

Contrast (17a) with (17b):

- (17) a. A: Trent wants me to make reservations at the Cattle Company.

**Have you eaten there yet?**

- B: No, but I understand it's real good. (LSAC)

- b. He is my man **and** he done me wrong

(<https://tspace.library.utoronto.ca/html/1807/4350/poem35.html>)

A's direct question in (17a) is whether there has been an occasion in which B ate at the Cattle Company. Indeed, B answers this direct question (negatively). However, given the background for A's question (A's first utterance), B infers that A actually wants to know whether the Cattle Company is a good place to eat at (an indirectly inferred question). He answers this indirect question too (positively, it would seem). Now, could the indirect question in (17a) be given a grammatical account? There seems to be no way to analyze *have you eaten there?* as (also) encoding 'is this a good place to eat at?' The latter interpretation is too ad hoc and not closely associated with any of the specific morphemes used. Moreover, we can easily account for it as an inference (see again section 1.2). An inferential account seems inevitable in this case.

But what about interpretations which are quite consistently associated with specific linguistic expressions? Deciding between a semantic versus a pragmatic

analysis is not invariably so easy. This is true for (17b). It is quite clear that the speaker of (17b) intends us to interpret the line in the famous ballad of "Frankie and Johnny" as conveying a certain relation between the two conjuncts. The fact that 'he is my man' is quite relevant to the fact that 'he done me wrong.' There is a contrast between the states of affairs described in the two conjuncts. *And* in (17b) is interpreted something like 'and despite that.' In cases such as these, it is less obvious that the conveyed contrastive interpretation is pragmatic. A possible account could specify a few meanings for natural language *and*, one of them being 'and despite that' perhaps.

In some cases, then, the main question we need to resolve is whether the interpretation at hand should be viewed as encoded (grammatical) or as inferred (pragmatic). There is indeed more controversy in the field regarding what should be relegated to grammar and what to pragmatics in such nonobvious cases. Should we analyze (17b) as semantically encoding 'He is my man' and 'he done me wrong,' and in addition as used by the speaker to also generate a conversational implicature that 'the fact that he done me wrong contrasts with the fact that he is my man'? Or should we analyze it as semantically encoding 'he is my man and despite that he done me wrong'? Modified Occam's Razor principle supports an inferential status for the contrast interpretation here. We analyze this case, as well as a variety of other cases, in part I, sometimes allotting the use/interpretation to grammar, sometimes to pragmatics.

Here's another case where it's not so easy to draw a grammar/pragmatics division of labor:

- (18) As long as I can remember I have always wanted kids and when we were talking about getting married over eight years ago, eight years or something I said I'm not willing to marry you unless you allow me to have kids 'cause he really didn't want me to have any, **the truth is** they terrify him. (LSAC)

*The truth is* constitutes "superfluous" information in (18). After all, Quality instructs us to only tell the truth. It seems quite straightforward to explain this violation as indicating a speaker intention to generate a conversational implicature to the effect that she feels uncomfortable in imparting the information modified by *the truth is* ('they terrify him'). The speaker is only telling her addressee that *they terrify him* because she's obliged to say so by Quality. If so, the relevant interpretation (that the information is socially dispreferred) is pragmatically inferred.

But then, consider the next example, uttered by a bilingual Hebrew/Arabic speaker:

- (19) M: efshar le=daber im IM?  
possible to=talk with IM?  
'Can I talk to IM?'  
H: **be=emet**, hi lo ba=bayit.  
in=truth, she not in.the=home  
'Really, she's not home.' (Phone conversation, Feb. 3, 1982)

Now, we could easily explain the use of Hebrew *be-emet* 'truly' here in the same manner. H is providing a dispreferred answer, which she feels uncomfortable with.